CrossMark

ORIGINAL CONTRIBUTION

Molecular connectivity indices for modeling the critical micelle concentration of cationic (chloride) Gemini surfactants

Anna Mozrzymas[1]

Abstract The molecular connectivity indices were used to derive the simple model relating the critical micelle concentration of cationic (chloride) gemini surfactants to their structure. One index was selected as the best to describe the effect of the structure of investigated compounds on critical micelle concentration consistent with the experimental results. This index encodes the information about molecular size, the branches, and also the information about heteroatoms. The selected model can be helpful in designing novel chloride gemini surfactants.

Keywords Chloride gemini surfactants · QSPR · Cmc · Molecular connectivity indices

Introduction

The quantitative structure-property relationship (QSPR) studies use the statistical models to estimate the various properties of the chemical compounds from its molecular structure [1–14]. In QSPR studies, the structure is often represented by different structural descriptors. Among the various structural parameters applied to QSPR analysis, the topological indices are often used in modeling physical, chemical, or biological properties [5–14]. The first applications of topological indices in structure-property relationship studies was proposed by Wiener in 1947 [15] and later in the 1975 by Randic[16]. The generalization of the Randic index are the

Kier and Hall molecular connectivity indices [10]. The molecular connectivity indices contain considerable information about the structure of the molecule. Kier and Hall [10]widely described the information encoded by molecular connectivity indices especially on thetopological but also the electronic properties of the molecule.

Gemini surfactants consist of two hydrophobic tails and two hydrophilic heads connected by the spacer group. Due to the binding together of two conventional surfactant molecules by the spacer, these compounds have very good properties in aqueous solution. The cmc values of these surfactants are significantly lower than those of the corresponding monomeric counterparts.

In the previous paper [12], the QSPR study was performed to derive the model which relates the critical micelle concentration of gemini surfactants to their structure. The relationship was developed for a set of 21 cationic (bromide) gemini surfactants employing the molecular connectivity indices only. The previous model contains the second-order molecular connectivity index which, as was suggested in [12], probably encodes the information about the flexibility.

In the present study, the 4models were derived. The relationships were developed for a set of 23 cationic (chloride) gemini surfactants also employing only the molecular connectivity indices. Just as in the previous study [12], the present models were derived for the molecules of various structures, i.e., the effect of all groups of the molecule on cmc value was taken into account. The structure of the investigated compounds are quite different from the previous bromides. Also, the test compounds differ in structure from previously studied compounds. The present study confirms that the one-descriptor model which best estimates the cmc values is that which contains the second-order molecular connectivity index, but the further analysis showed that the model which contains the first-order valence molecular connectivity index

✉ Anna Mozrzymas
anna.mozrzymas@up.wroc.pl

[1] Department of Physics and Biophysics, Wrocław University of Environmental and Life Sciences, ul. Norwida 25, 50-375 Wrocław, Poland

🕿 Springer

better describes the changes of cmc values of cationic (chloride) gemini surfactants caused by structure modification.

Materials and methods

Dataset

The data set contains only gemini surfactants with chlorides as counterions. The compounds were chosen to contain gemini surfactants with medium and long spacer length. The chemical structures of the investigated compounds along with their abbreviations are presented in Fig. 1. The data set contains 23 compounds of training set and 2test compounds. The chemical

structures of the surfactants and the experimental values of cmc were taken from literature [17–21].
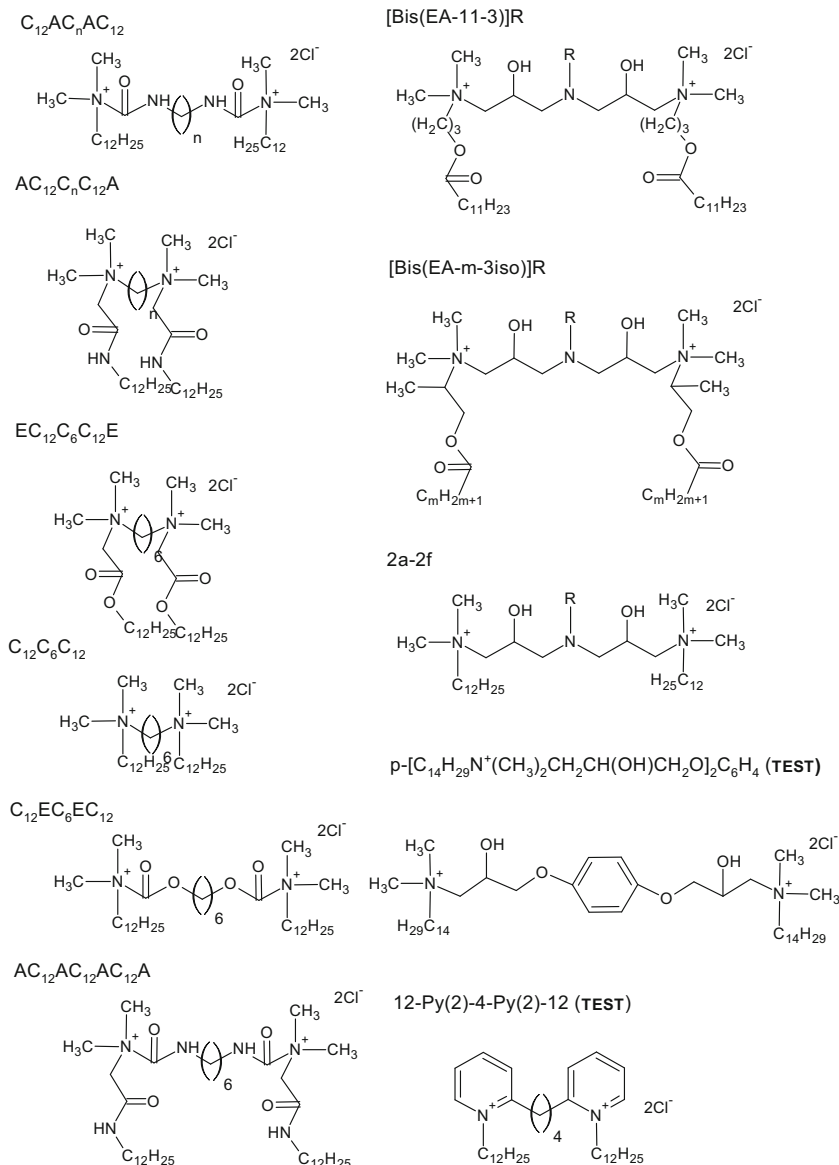
Molecular connectivity indices

Just as in the previous papers [11–13], the structure of the molecule is represented by Kier and Hall's molecular connectivity indices.

The mth order molecular connectivity index is defined [10] by

$$^{\mathrm{m}}\chi_k = \sum_{j=1}^{n_m} \prod_{i=1}^{m+1} (\delta_i)_j^{-0.5} \tag{1}$$

where δ_i is a connectivity degree, i.e., the number of non-hydrogen atoms to which the ith non-hydrogen atom is

Fig. 1 Structures of investigated compounds and their abbreviations

bonded; m is the order of the connectivity index; k denotes the type of the fragment of the molecule, for example: path (p), cluster (c), and path-cluster (pc); and n_m is the number of fragments of type k and order m.

For molecules with the heteroatoms, the valence connectivity degree has been defined [10] as

$$\delta_i^\nu = \frac{Z_i^\nu - h_i}{Z_i - Z_i^\nu - 1} \tag{2}$$

where Z_i^ν is the number of valence electrons in the ith atom, h_i is the number of hydrogen atoms connected to the ith atom, and Z_i is the number of all electrons in the ith atom.

The replacement δ_i by δ_i^ν defines the valence molecular connectivity index $^m\chi_k^\nu$.

An example of calculations of molecular connectivity indices for one of the investigated gemini surfactants is presented in Appendix 1.

The molecular connectivity indices contain considerable information about the molecule. The low-order molecular connectivity indices include information about atoms and molecular size while cluster and path/cluster molecular connectivity indices include structure information about branch point and branch point environment; the valence indices add information about heteroatoms [10, 22]. For example, $^0\chi^\nu$ index includes information about heteroatoms contained in the molecule, $^1\chi$ and $^1\chi^\nu$ indices contain the information about molecular volume and molecular surface area; additionally, the $^1\chi^\nu$ adds information about heteroatoms but $^3\chi_c^\nu$ index contains the information about the number of branches and their heteroatoms [10].

Statistics

The least squares method was used to generate the formula expressing the relationship between the logcmc and the molecular connectivity indices. In order to test the quality of the derived equation, three statistical parameters were used: a coefficient of determination (r^2), a correlation coefficient (r), a Fisher ratio (F), and a standard deviation (s). The best relationship is that which has possibly the highest values of r^2, r, and F and simultaneously the lowest value of s.

In the case of the simple linear least-squares model, the values of statistical parameters may be calculated using the following formulas [10]:

-the coefficient of determination : $r^2 = \dfrac{\sum \left(y_i(\text{cal}) - \overline{y}\right)^2}{\sum \left(y_i(\text{exp}) - \overline{y}\right)^2}$ (3)

where $y_i(\text{exp})$—the experimental value of the property, $y_i(\text{cal})$—the calculated value of the property and $\overline{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$,

- the correlation coefficient (r) can be obtained from Eq. (3) as a square root of the coefficient of determination. Notice that this definition of r, in agreement with ref. [10], does not correspond to the standard definition of Pearson's linear correlation coefficient, although it has a similar meaning.

-the Fisher ratio : $F = (n-2) \cdot \dfrac{r^2}{(1-r^2)}$ (4)

-the standard deviation of the fit :

$$s = \sqrt{\frac{\sum \left(y_i(\text{exp}) - y_i(\text{cal})\right)^2}{n-2}} \tag{5}$$

where n is the number of compounds in the data set,

-the residual for compound i : $\Delta_i = y_i(\text{exp}) - y_i(\text{cal})$ (6)

Results and discussion

The values of the molecular connectivity indices along with the experimental logcmc values for the training set are listed in Table 1.

Basing on the data contained in Table 1, the correlation formulas containing one index were derived (Step 1). All values of statistical parameters for the relationships obtained in the first step are listed in Table 2.

As follows from Table 2, the highest values of r and F and the lowest value of s are for the relationship containing the second-order molecular connectivity index ($^2\chi$). The inspection of data contained in Table 2 suggests that $^0\chi$, $^1\chi$, and $^0\chi^\nu$ indices also correlate well with the logcmc values. Table 3 shows the correlations between all the indices appearing in Table 1.

Two indices with $r \geq 0.97$ are highly correlated, those with $0.90 \leq r < 0.97$ are appreciably correlated, the indices with $0.50 \leq r < 0.89$ are weakly correlated, and those with $r < 0.50$ are not correlated. As follows from the correlation matrix, there are 12 pairs of highly correlated indices, among them, the pairs of $^0\chi$ and $^2\chi$ indices with value of correlation coefficient 0.997. Because of, the $^0\chi$ and $^2\chi$ indices carry similar structural information related to the changes of molecular structure, i.e., the values of these indices increase with the increase of atoms and branches in the molecule [10]; therefore, we can ignore the $^0\chi$ index in further considerations. The remaining indices which highly correlate with logcmc values, namely, $^2\chi$, $^1\chi$, and $^0\chi^\nu$ indices also highly correlate to each other (Table 3), but they contain somewhat different structure information;

Table 1 Experimental logcmcvalues [17–19] and values of molecular connectivity indices

Compound	logcmc[a]	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi_c$	$^4\chi_{pc}$	$^0\chi^\nu$	$^1\chi^\nu$	$^2\chi^\nu$	$^3\chi_c^\nu$	$^4\chi_{pc}^\nu$
bis(EA-11-3)C5	−4.31158	40.342595	25.83567	21.71634	3.77304	4.01125	37.06539	22.87411	17.76306	2.693502	2.6531996
bis(EA-11-3)C6	−4.34969	41.049702	26.33567	22.069896	3.77304	4.01125	37.772499	23.37411	18.11661	2.693502	2.6531996
bis(EA-11-3)C8	−4.39041	42.46392	27.33567	22.777	3.77304	4.01125	39.18671	24.37411	18.82372	2.693502	2.6531996
bis(EA-9-3iso)C6	−4.07469	38.54755	24.17711	20.76965	3.86803	6.47149	35.270345	21.22925	16.85702	2.77846	4.74458
bis(EA-9-3iso)C8	−4.10568	39.96176	25.17711	21.47676	3.86803	6.47149	36.68456	22.22925	17.56413	2.77846	4.74458
bis(EA-11-3iso)C5	−4.20204	40.66887	25.67711	21.83031	3.86803	6.47149	37.39167	22.72925	17.91768	2.77846	4.74458
bis(EA-11-3iso)C6	−4.22841	41.37598	26.17711	22.18386	3.86803	6.47149	38.09877	23.22925	18.27124	2.77846	4.74458
bis(EA-11-3iso)C8	−4.266803	42.79019	27.17711	22.89097	3.86803	6.47149	39.51299	24.22925	18.97834	2.77846	4.74458
$C_{12}AC_2AC_{12}$	−3.02687	28.53948	18.16987	14.92768	2.509202	4.85064	26.68149	16.39809	12.64503	1.80341	2.18671
$C_{12}AC_6AC_{12}$	−3.07058	31.367904	20.16987	16.34189	2.509202	4.85064	29.50991	18.39809	14.05924	1.80341	2.18671
$C_{12}AC_{12}AC_{12}$	−3.65758	35.61054	23.16987	18.46321	2.509202	4.85064	33.75255	21.39809	16.18056	1.80341	2.18671
$C_{12}C_6C_{12}$[b]	−2.88606	26.79899	17.32843	14.1066	2.41421	2.41421	26.69342	16.96798	13.544065	2.15934	2.15934
$C_{12}EC_6EC_{12}$	−3.11351	31.367904	20.16987	16.34189	2.509202	4.85064	29.32641	18.17658	13.82412	1.78666	2.03045
$AC_{12}C_6C_{12}A$	−3.46852	32.78212	21.11612	17.50965	2.99156	2.88963	30.92413	19.29044	15.11298	2.30368	2.00972
$AC_{12}C_{12}C_{12}A$	−3.79588	37.02476	24.11612	19.63097	2.99156	2.88963	35.16677	22.29044	17.23431	2.30368	2.00972
$EC_{12}C_6C_{12}E$	−3.49485	32.78212	21.11612	17.50965	2.99156	2.88963	30.74062	19.06893	14.85402	2.27719	1.97915
$AC_{12}AC_6AC_{12}A$	−3.60206	37.35103	23.95756	19.74494	3.08655	5.34987	33.74062	20.72055	15.62816	1.94775	2.09135
2a (R = C_2H_5)	−3.12843	30.82393	19.54797	16.55666	3.19569	3.25454	29.48265	18.27316	14.70763	2.57565	2.53284
2b (R = C_3H_7)	−3.204815	31.53104	20.04797	16.93708	3.19569	3.19475	30.18976	18.77316	15.11501	2.57565	2.48653
2c (R = C_4H_9)	−3.25104	32.23815	20.54797	17.29064	3.19569	3.19475	30.896865	19.27316	15.46856	2.57565	2.48653
2d (R = C_5H_{11})	−3.40561	32.94525	21.04797	17.64419	3.19569	3.19475	31.60397	19.77316	15.82212	2.57565	2.48653
2e (R = C_6H_{13})	−3.50169	33.65236	21.54797	17.997745	3.19569	3.19475	32.31108	20.27316	16.17567	2.57565	2.48653
2f (R = C_8H_{17})	−3.598599	35.06657	22.54797	18.70485	3.19569	3.19475	33.72529	21.27316	16.882775	2.57565	2.48653

[a] Thecmc values were measured in pure water at 25 °C

[b] For this compound the values of molecular connectivity indices were taken from [12]

especially, their values vary with changes in molecular structure. The first-order molecular connectivity index ($^1\chi$) decreases with the increase of branches, but the second-order molecular connectivity index ($^2\chi$) increases with the increase of branches, whereas the valence molecular connectivity index of zero order ($^0\chi^\nu$) encodes the information about heteroatoms [10, 22]. Thus, we keep these indices in the next considerations. These indices define models 1–3 in the first step. To these indices, the remaining indices were added separately (step 2). The

values of the correlation coefficients for this step (second step) are contained in Table 4.

Because the $^2\chi$ index alone gives $r = 0.982$, therefore the relationships with pair of indices $^1\chi$ and $^2\chi$ ($r = 0.982$) and also with the pair $^0\chi^\nu$ and $^2\chi$ ($r = 0.982$) indices can be ignored in the further investigations. Next, from Table 4, it follows that in the case of models 1 and 3, the values of the correlation coefficients did not change significantly, so for those models, the step by step process was ended. In the case of model 2, the values of the correlation coefficients are higher for the rela-

Table 2 Values ofstatistical parameters for Step 1

Index	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi_c$	$^4\chi_{pc}$	$^0\chi^\nu$	$^1\chi^\nu$	$^2\chi^\nu$	$^3\chi_c^\nu$	$^4\chi_{pc}^\nu$
r	0.978	0.976	0.982	0.864	0.517	0.975	0.957	0.946	0.656	0.636
F	465.827	414.171	563.629	61.856	7.658	403.639	231.364	177.729	15.832	14.268
s	0.104	0.110	0.095	0.252	0.429	0.111	0.144	0.163	0.378	0.387

Table 3 Correlation matrix

	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi_c$	$^4\chi_{pc}$	$^0\chi^\nu$	$^1\chi^\nu$	$^2\chi^\nu$	$^3\chi_c^\nu$	$^4\chi_{pc}^\nu$
$^0\chi$	1.000	0.997	0.997	0.848	0.590	0.992	0.974	0.952	0.607	0.650
$^1\chi$	**0.997**	1.000	0.992	0.816	0.551	0.993	0.982	0.952	0.574	0.598
$^2\chi$	**0.997**	**0.992**	1.000	0.881	0.568	0.991	0.970	0.959	0.658	0.664
$^3\chi_c$	0.848	0.816	0.881	1.000	0.473	0.846	0.799	0.860	0.903	0.767
$^4\chi_{pc}$	0.590	0.551	0.568	0.473	1.000	0.525	0.443	0.406	0.161	0.799
$^0\chi^\nu$	**0.992**	**0.993**	**0.991**	0.846	0.525	1.000	0.992	0.981	0.647	0.637
$^1\chi^\nu$	**0.974**	**0.982**	**0.970**	0.799	0.443	**0.992**	1.000	0.985	0.621	0.562
$^2\chi^\nu$	0.952	0.952	0.959	0.860	0.406	**0.981**	**0.985**	1.000	0.740	0.621
$^3\chi_c^\nu$	0.607	0.574	0.658	0.903	0.161	0.647	0.621	0.740	1.000	0.660
$^4\chi_{pc}^\nu$	0.650	0.598	0.664	0.767	0.799	0.637	0.562	0.621	0.660	1.000

The bold values mean high correlation

tionships which contain additionally $^3\chi_c$ or $^3\chi_c^\nu$ indices. The $^3\chi_c$ index encodes the information about the number of branches and their environment [10, 22]. The $^3\chi_c^\nu$ index adds information about heteroatoms. Thus, the relationship containing the $^3\chi_c^\nu$ index is richer in structural information than with the $^3\chi_c$ index. Furthermore, the addition others indices (step 3) did not change significantly the values of correlation coefficients therefore model 2 is now defined by the pair of indices $^1\chi$ and $^3\chi_c^\nu$.

The obtained formulas (models 1–3) are given below:

$$\text{Model 1 : logcmc} = -0.17261 - 0.184411 \cdot {}^2\chi \tag{7}$$

$$\text{Model 2 : logcmc} = 0.18447 - 0.14866 \cdot {}^1\chi - 0.19248 \cdot {}^3\chi_c^\nu \tag{8}$$

$$\text{Model 3 : logcmc} = 0.44266 - 0.12317 \cdot {}^0\chi^\nu \tag{9}$$

The statistical characteristics of the descriptors included in models 1–3 are shown in Appendix 2.

The plots of the experimental logcmc versus the logcmc calculated using Eqs. 7–9 are presented in Figs. 2, 3, and 4.

The comparisons of the experimental logcmc with the values calculated using Eqs. 7–9 presented in Figs. 2, 3, and 4 show that models 1–3 estimate the logcmc of compounds from the training set very well, and model 2 is slightly better than model 1 and better than Model 3. The values of coefficients of determination are equal to 0.964, 0.966, and 0.951 for models 1, 2, and 3, respectively.

The plots of residuals versus the experimental values of logcmc are shown in Figs. 5, 6, and 7.

The examination of the residuals (Figs. 5, 6, and 7) shows generally good agreement between the experimental and calculated values of logcmc. Most of the residuals are close to zero and only one residual for model 1 is slightly larger than $2s$.

The obtained models were used to estimate the logcmc values of other compounds, different from gemini surfactants from the training set. The values of the literature logcmc for test compounds are listed in Table 5.

The comparison of the experimental values of logcmc of the compounds used in the test with the values estimated using Eqs. 7–9 is shown in Figs. 2, 3, and 4. The agreement between predicted and experimental logcmc values

Table 4 Values of correlation coefficients for models 1–3 in step 2

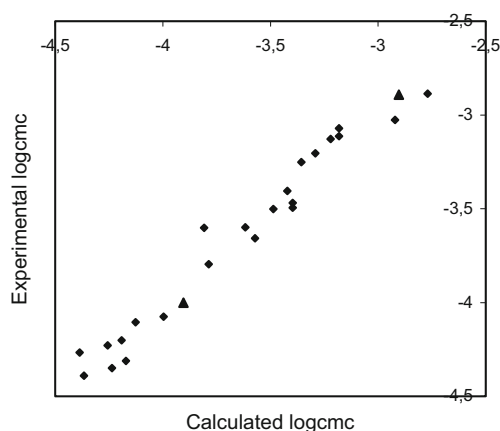| Indices	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi_c$	$^4\chi_{pc}$	$^0\chi^\nu$	$^1\chi^\nu$	$^2\chi^\nu$	$^3\chi_c^\nu$	$^4\chi_{pc}^\nu$
Model 1	0.982	0.982	–	0.982	0.983	0.982	0.982	0.982	0.982	0.982
Model 2	0.978	–	0.982	0.983	0.976	0.977	0.976	0.977	0.983	0.978
Model 3	0.979	0.977	0.982	0.978	0.975	–	0.978	0.976	0.976	0.975

Fig. 2 Plot of the experimental logcmc versus thatcalculated using Eq. 7 for training set (*rhomb*) ($r = 0.982$, $F = 563.629$, $s = 0.095$) and test compounds (*triangle*)
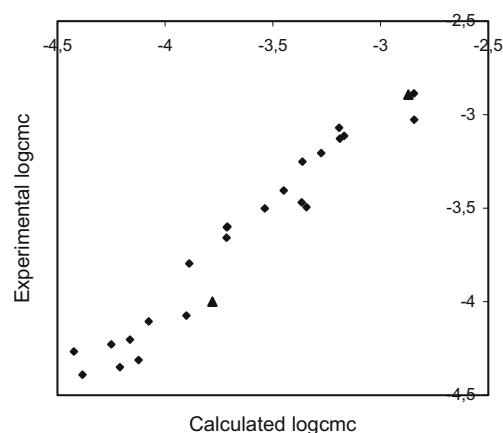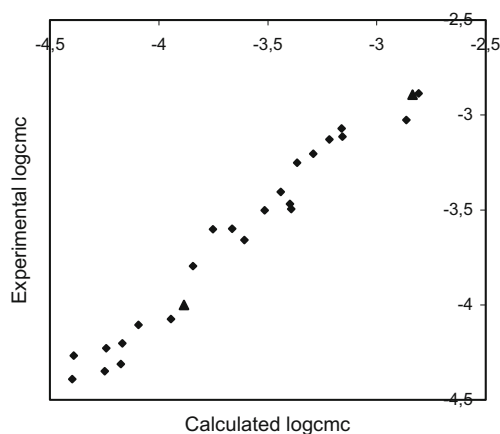

Fig. 4 Plot of the experimental logcmc versus thatcalculated using Eq. 9 for training set (*rhomb*) ($r = 0.975$, $F = 403.639$, $s = 0.111$) and test compounds (*triangle*)

of the test compounds is very good. The plots of residuals (Figs. 5, 6, and 7) confirm this agreement.

In brief, the best model in the first step is that which contains the second-order molecular connectivity index ($^2\chi$) (model 1). The second step shows that the relationship containing the first-order molecular connectivity index ($^1\chi$) and the third-order cluster valence molecular connectivity index ($^3\chi_c^\nu$) (model 2) estimates slightly better the values of the critical micelle concentration of cationic (chloride) gemini surfactants.

The second-order molecular connectivity index ($^2\chi$) appearing in model 1 does not differentiate heteroatoms;it represents two-bond terms within the molecule and its values depend on the isomers of the compound [10]. The values of $^2\chi$ index increase with the increase in length and branches of hydrocarbon

chains. The zeroth-order valence molecular connectivity index ($^0\chi^\nu$) appearing in model 3 relates to the atoms of the molecule, and it differentiates heteroatoms. The values of $^0\chi^\nu$ index increase with the increase in length and branches of hydrocarbon chains, and its values are smaller for the compounds containing in their structure heteroatoms in comparison with those of their hydrocarbon analogous compounds. The first-order molecular connectivity index ($^1\chi$) appearing in model 2 does not differentiate heteroatoms;it represents the one-bond terms within the molecule. The values of $^1\chi$ index depend on the isomers of the compound and, in this case, decrease with the increase in branches, but its values increase with the increase in length of hydrocarbon chains. The third-order cluster valence molecular connectivity index ($^3\chi_c^\nu$) appearing in model 2 represents

Fig. 3 Plot of the experimental logcmc versus thatcalculated using Eq. 8 for training set (*rhomb*) ($r = 0.983$, $F = 585.435$, $s = 0.093$) and test compounds (*triangle*)
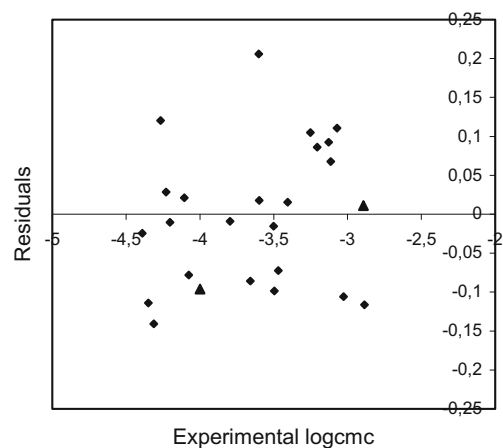

Fig. 5 Plot of residuals versus the experimental logcmc values for training set (*rhomb*) and test compounds (*triangle*) (model 1)
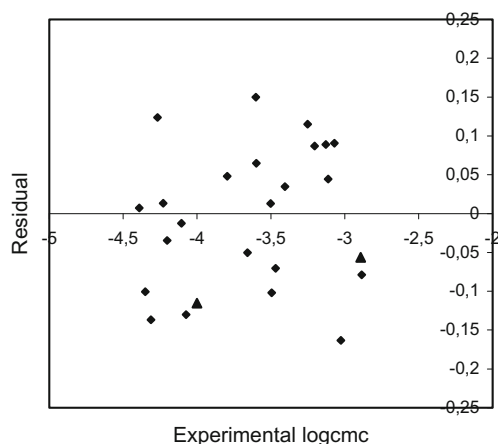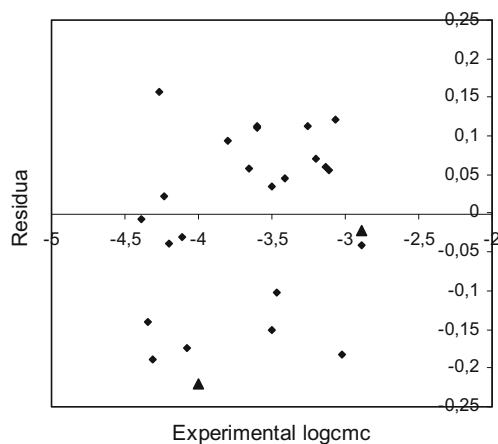
Fig. 6 Plot of residuals versus the experimental logcmc values for training set (*rhomb*) and test compounds (*triangle*) (model 2)

three-bond cluster terms within the molecule, and it differentiates heteroatoms. The values of $^3\chi_c^\nu$ index increase with the increase in branches of hydrocarbon chains, and its values are smaller for the compounds containing in their structure heteroatoms in comparison with those of their hydrocarbon analogous compounds. All models contain the molecular connectivity indices with negative coefficients, thus as their values increase, the cmc decreases.

So, from Eqs. (7–9) and also from Table 1, it follows that as the number of methylene groups increases in the hydrocarbon chains, the cmc decreases. For example, for compound bis(EA-m-3iso)C6 (*m* = 9, 11), the experimental values of cmc are the following: 0.084 and 0.059 mM[], and the calculated values of cmc are the following: 0.101and 0.055 mM (model 1), 0.113and 0.057 mM (model 2), and 0.125and 0.056 mM (model

Fig. 7 Plot of residuals versus the experimental logcmc values for training set (*rhomb*) and test compounds (*triangle*) (model 3)

Table 5 Experimental logcmc values [20, 21] of test compounds

Compound[a]	Experimentallogcmc
12-Py(2)-4-(2)Py-12·2Cl$^-$	−2.89279
p-[C$_{14}$H$_{29}$N$^+$(CH$_3$)$_2$CH$_2$CH(OH)CH$_2$O]$_2$C$_6$H$_4$·2Cl$^-$	−4.0

[a] The structures of the compounds are presented in Fig. 1

3). Also, as the number of methylene groups increases in the spacer group then the experimental and also the calculated values of the cmc decrease. For example, for compound AC$_{12}$C$_n$C$_{12}$A (*n* = 6, 12), the experimental values of cmc are the following: 0.340and 0.160 mM[18], and the calculated values of cmc are the following: 0.401and 0.163 mM (model 1), 0.399and 0.143 mM (model 2), and 0.430and 0.129 mM (model 3). In the case of compounds bis(EA-m-3)R and bis(EA-m-3iso)R for *R* = 5, 6, 8, the experimental and also the calculated values of cmc decrease too with the increase in the alkyl chain length at the central nitrogen atom in the molecule. Thus, the increase in length of hydrocarbon chain and simultaneously in flexibility of this chain results in the decrease of cmc values.

The comparison of the compounds with straight and branched chains shows that the branches differently influence the calculated cmc values. For example, for compounds bis(EA-11-3)C8 and bis(EA-11-3iso)C8 using model 1, we obtain the following values of cmc: 0.043and 0.041 and the following using model 3: 0.041and 0.038 mM, whereas using model 2, we obtain 0.040and 0.041 mM, respectively. The experimental cmc values are 0.041 mM for compound bis(EA-11-3)C8 and 0.054 mM for compound bis(EA-11-3iso)C8 [17]. It means that the experimental value of cmc is higher for the compound bis(EA-11-3iso)C8; therefore, the cmc values calculated using model 2 are in good agreement with the experimental results. Some othergemini surfactants and the corresponding calculated values of logcmcare presented in Appendix 3. For the compounds presented in in Appendix 3, the cmc values which are calculated using models1 and 2 are smaller for the compounds with branched chains than for those with straight chains and the same number of atoms. Using model 3, the cmc values are smaller only for compounds with branched carbon chains but for compounds containing heteroatoms, the branches cause the higher cmc values. The result obtained for the compounds containing heteroatoms is in agreement with the experimental one [23]. That is, for chloride compounds C$_{12}$EO$_1$C$_{12}$ (0.5 mM at 20 °C) and C$_{12}$C$_4$(OH)C$_{12}$ (0.65 mM at 20 °C) [23].

The comparison of the heteroatom compounds with their hydrocarbon analogous compounds (Appendix 3) shows that the presence of heteroatoms in the molecules results in higher calculated, using Model 3, values of critical micelle concentration in comparison with its carbon analogous compounds. Model 2 differentiates heteroatoms only on branches but Model 1 does not differentiate heteroatoms. Some experimental results show higher values of critical micelle concentration of gemini surfactants containing in their structure heteroatoms in comparison with those of their hydrocarbon analogous compounds [18, 24–26]. That is, for example, for bromide compounds $C_{12}EO_2C_{12}$ (1.09 mM[24]) and $C_{12}C_8C_{12}$ (0.84 mM[25]) and also for $C_{12}7NHC_{12}$ (1.17 mM[26] and 1.21 mM[18]) and $C_{12}C_7C_{12}$ (0.9 mM[26]). Also, the theoretical results obtained for cationic (bromide) gemini surfactants with various spacer group only [13] show that the presence of heteroatoms in the spacer group results in higher value of cmc. Thus, model 3 better describes the effect of heteroatoms on cmc values.

In brief, the investigated models (models 1–3) show high correlations between logcmc and the molecular connectivity indices and statistically, the best models (models1–2) can be used to estimate the values of critical micelle concentration, but the description of the effect of the structure of investigated compounds on cmc values by those models is different. All models describe the cmc values very well if we take into account only the elongation of alkyl chains. In the case of branches and heteroatoms, these models differently describe cmc values and some results differ from the experimental ones. It suggests that another index will be better to describe the effect of the structure on critical micelle concentration of cationic (chloride) gemini surfactants. Because some experimental data show that the branched chains especially branched hydrocarbon chains [17, 23, 27], and also heteroatoms [24–26], cause the higher cmc values therefore the best index which will satisfactorily describe the effect of the chemical structure on cmc value is the first-order valence molecular connectivity index ($^1\chi^\nu$). The first-order valence molecular connectivity index ($^1\chi^\nu$) is similar to the first-order molecular connectivity index ($^1\chi$), but it includes heteroatom information. The values of $^1\chi^\nu$ index increase with the increase in length of hydrocarbon chains, and its values decrease with the increase in branches. This index differentiates heteroatoms and its values are smaller than the values of the $^1\chi$ index.

The formula containing the $^1\chi^\nu$ index is the following:

$$\text{Model 4}: \text{logcmc} = 0.56045 - 0.20443 \cdot {}^1\chi^\nu \quad (10)$$

The statistical characteristic of the selected descriptor is given in Appendix 2.

From Eq. 10, it follows that as the number of methylene groups increases in the hydrocarbon chains and also in spacer chain, the cmc decreases. For example, for compound bis(EA-m-3iso)C8 ($m = 9, 11$), the experimental values of cmc are the following: 0.078 and 0.054 mM[17], and the calculated values of cmc are the following: 0.104 and 0.040 mM. For compounds with different spacer lengths $AC_{12}C_nC_{12}A$ ($n = 6, 12$), the experimental values of cmc are the following: 0.340 and 0.160 mM[18], and the calculated values of cmc are the following: 0.414 and 0.101 mM, respectively. The comparison of the compounds with straight and branched chains shows that the calculated values are also in good agreement with the experimental results. The example are the compounds bis(EA-11-3)C8 and bis(EA-11-3iso)C8, for which the experimental cmc values are the following: 0.041 and 0.054 mM, and the calculated using model 4 cmc values are the following: 0.038 and 0.040 mM, respectively.

The plot of the experimental logcmc versus the logcmc calculated using Eq. 10 and the plot of residuals versus the experimental values of logcmc for training set and test compounds are shown in Figs. 8 and 9

The statistical parameters show that model 4 estimates logcmc values of investigated compounds lower than models 1–3, but comparison of the experimental and calculated values of cmc by means of the effect of the structural elements on cmc values shows that the values of critical micelle concentration calculated using model 4 are in good agreement with the experimental results. Some additional comparisons are presented in Appendix 3.

The data contained in Appendix 3 show that the increase in the number of atoms by lengthening or by the increase of branches causes the decrease of the cmc value calculated using models1–4. If we take in to account the
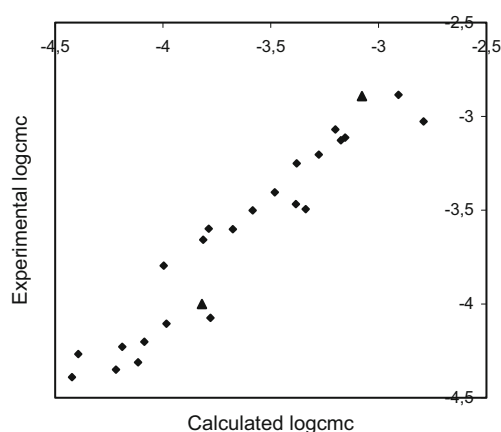

Fig. 8 Plot of the experimental logcmc versus that calculated using Eq. 10 for training set (rhomb) ($r = 0.957$, $F = 231.36$, $s = 0.144$) and test compounds (triangle)
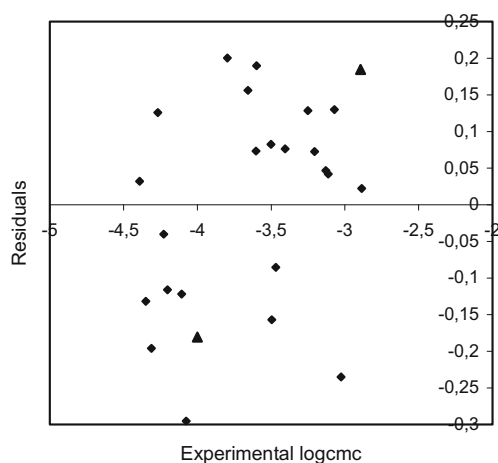
Fig. 9 Plot of residuals versus the experimental logcmc values for training set (*rhomb*) and test compounds (*triangle*) (model 4)

heteroatom compounds, the effect of branches is in good agreement with experimental results obtained for bromide compounds [28]. But in the case of the elongation of hydrophilic spacer, as is for compounds $C_{12}EO_nC_{12}$, the experimental results [24] show the opposite behavior. Maybe it is due to the fact that the length of hydrocarbon chains has the dominant effect on cmc values and, in consequence, on obtained models.

The experimental studies [18] show also that the cmc values of chloride gemini surfactants are higher than of bromides ones. Indeed, the experimental cmc values of $C_{12}C_6C_{12}$ gemini surfactant with bromides and chlorides as counterions are the following: 0.89 and 1.30 mM [18], respectively. But, using previous model [12] for bromide geminis and present (Eqs. 7–10) for chlorides ones, we obtain the following calculated values of cmc: 1.11 mM [12] and 1.70 mM (model 1), 1.56 mM (model 2), 1.43 mM (model 3), and 1.24 mM (model 4), respectively. So, both the experimental and the calculated values of cmc of cationic (chloride) gemini surfactants are higher than for the bromide ones and in the case of chloride surfactants, the best estimated value is for model 4.

It is worth to add that the test compounds (Table 5) differ in structure of spacer and head groups from the training set compounds, but also for those molecules, the agreement between predicted and experimental logcmc values is very good.

Conclusion

In the present work, the cationic (chloride) gemini surfactants with various structures were taken into account. All the models obtained confirm the experimental results that the length of alkyl chains plays the major role in micelle formation. The present study shows that

although the second-order molecular connectivity index correlates high with logcmc values of cationic gemini surfactants, the statistically lower correlation logcmc with the first order valence molecular connectivity index better describes the effect of the branches and heteroatoms on the critical micelle concentration of cationic (chloride) gemini surfactants. Because model 4 (Eq. (10)) has good prediction ability of investigated compounds, it can be used to predict the critical micelle concentration and in particular to design new cationic (chloride) gemini surfactants more active in micelle formation.

Appendix 1

To illustrate the calculation of the molecular connectivity indices, the 2a gemini surfactant (Table 1) was taken into account. The first step of calculations is to draw the structural formula of the molecule and to count the values of connectivity degrees [10]. The hydrogen atoms are suppressed in graphic structural formula. The structure along with the values of connectivity and valence connectivity degrees are shown in Fig. 10.
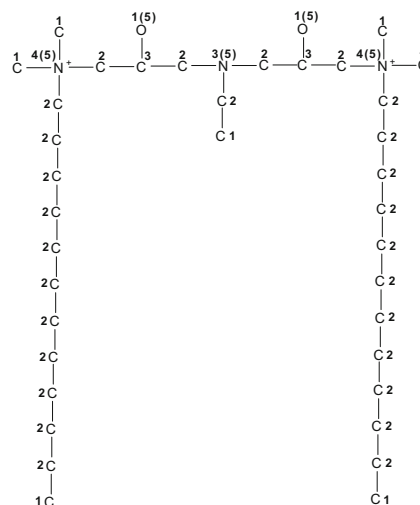

Fig. 10 Hydrogen-suppressed graphic structural formula of exemplary gemini surfactant and delta values

Next, the molecule is dissected into the appropriate fragments, for example: path, cluster, or path-cluster. The values of connectivity indices can be easily calculate using Eq. 1.

The calculations of molecular connectivity indices for exemplary gemini surfactant read:

$$^0\chi = \sum (\delta_i)^{-0.5} =$$
$$= 9 \cdot (1)^{-0.5} + 27 \cdot (2)^{-0.5} + 2 \cdot (4)^{-0.5} + 3 \cdot (3)^{-0.5} =$$
$$= 30.82393$$

$$^1\chi = \sum (\delta_i \times \delta_j)^{-0.5} =$$
$$= 3 \cdot (2 \times 1)^{-0.5} + 20 \cdot (2 \times 2)^{-0.5} + 4 \cdot (4 \times 1)^{-0.5} + 4 \cdot (2 \times 4)^{-0.5} + 7 \cdot (2 \times 3)^{-0.5} + 2 \cdot (3 \times 1)^{-0.5} =$$
$$= 19.54797$$

$$^2\chi = \sum (\delta_i \times \delta_j \times \delta_k)^{-0.5} =$$
$$= 2 \cdot (2 \times 2 \times 1)^{-0.5} + 18 \cdot (2 \times 2 \times 2)^{-0.5} + 4 \cdot (2 \times 2 \times 4)^{-0.5} + 8 \cdot (4 \times 2 \times 1)^{-0.5} + 2 \cdot (1 \times 4 \times 1)^{-0.5} +$$
$$+ 2 \cdot (4 \times 2 \times 3)^{-0.5} + 5 \cdot (2 \times 3 \times 1)^{-0.5} + 5 \cdot (2 \times 3 \times 2)^{-0.5} + 2 \cdot (3 \times 2 \times 3)^{-0.5} =$$
$$= 16.55666$$

$$^3\chi_c = \sum (\delta_i \times \delta_j \times \delta_k \times \delta_l)^{-0.5} =$$
$$= 4 \cdot (4 \times 2 \times 2 \times 1)^{-0.5} + 4 \cdot (4 \times 2 \times 1 \times 1)^{-0.5} + 2 \cdot (2 \times 3 \times 2 \times 1)^{-0.5} + (2 \times 3 \times 2 \times 2)^{-0.5} =$$
$$= 3.19569$$

$$^4\chi_{pc} = \sum (\delta_i \times \delta_j \times \delta_k \times \delta_l \times \delta_m)^{-0.5} =$$
$$= 2 \cdot (1 \times 4 \times 2 \times 2 \times 1)^{-0.5} + 4 \cdot (1 \times 4 \times 2 \times 2 \times 2)^{-0.5} + 2 \cdot (1 \times 4 \times 1 \times 2 \times 3)^{-0.5} + 6 \cdot (1 \times 4 \times 2 \times 2 \times 3)^{-0.5} +$$
$$+ 2 \cdot (1 \times 3 \times 2 \times 2 \times 3)^{-0.5} + 2 \cdot (2 \times 3 \times 2 \times 2 \times 3)^{-0.5} + (2 \times 3 \times 2 \times 2 \times 1)^{-0.5} =$$
$$= 3.25454$$

and

$$^0\chi^\nu = \sum (\delta_i^\nu)^{-0.5} =$$
$$= 7 \cdot (1)^{-0.5} + 27 \cdot (2)^{-0.5} + 2 \cdot (3)^{-0.5} + 5 \cdot (5)^{-0.5} =$$
$$= 29.48265$$

$$^1\chi^\nu = \sum (\delta_i^\nu \times \delta_j^\nu)^{-0.5} =$$
$$= 3 \cdot (2 \times 1)^{-0.5} + 20 \cdot (2 \times 2)^{-0.5} + 7 \cdot (2 \times 5)^{-0.5} + 4 \cdot (5 \times 1)^{-0.5} + 4 \cdot (2 \times 3)^{-0.5} + 2 \cdot (5 \times 3)^{-0.5} =$$
$$= 18.27316$$

$$^2\chi^\nu = \sum (\delta_i^\nu \times \delta_j^\nu \times \delta_k^\nu)^{-0.5} = 2 \cdot (2 \times 2 \times 1)^{-0.5} + 18 \cdot (2 \times 2 \times 2)^{-0.5} + 7 \cdot (2 \times 2 \times 5)^{-0.5} + 2 \cdot (1 \times 5 \times 1)^{-0.5} +$$
$$+ 9 \cdot (1 \times 5 \times 2)^{-0.5} + 8 \cdot (5 \times 2 \times 3)^{-0.5} + 2 \cdot (2 \times 2 \times 3)^{-0.5} =$$
$$= 14.70763$$

$$^3\chi_c^\nu = \sum (\delta_i^\nu \times \delta_j^\nu \times \delta_k^\nu \times \delta_l^\nu)^{-0.5} =$$
$$= 4 \cdot (5 \times 2 \times 2 \times 1)^{-0.5} + 4 \cdot (5 \times 2 \times 1 \times 1)^{-0.5} + 2 \cdot (5 \times 2 \times 2 \times 3)^{-0.5} + (5 \times 2 \times 2 \times 2)^{-0.5} =$$
$$= 2.57565$$

$$^4\chi_{pc}^\nu = \sum (\delta_i^\nu \times \delta_j^\nu \times \delta_k^\nu \times \delta_l^\nu \times \delta_k^\nu)^{-0.5} =$$
$$= 2 \cdot (1 \times 5 \times 2 \times 2 \times 1)^{-0.5} + 5 \cdot (1 \times 5 \times 2 \times 2 \times 2)^{-0.5} + 2 \cdot (1 \times 5 \times 2 \times 3 \times 1)^{-0.5} + 4 \cdot (1 \times 5 \times 2 \times 2 \times 3)^{-0.5} +$$
$$+ 4 \cdot (3 \times 5 \times 2 \times 2 \times 5)^{-0.5} + 2 \cdot (2 \times 5 \times 2 \times 2 \times 3)^{-0.5} =$$
$$= 2.53284$$

Appendix 2

The statistical characteristics of the descriptors contained in models1–4 are given in Table 6.

Table 6   Characteristics of descriptors

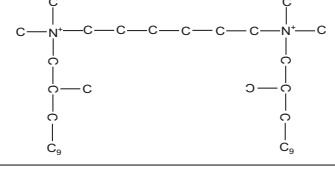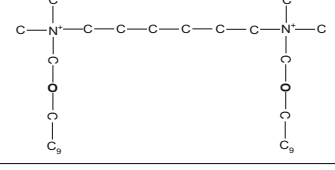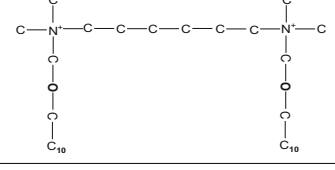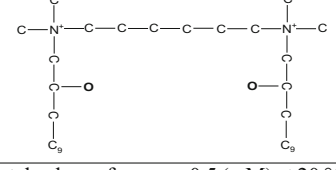| Model	Constant/descriptor	Coefficient	Standard error	t value	p value
1	Constant	−0.17261	0.14814	−1.165	0.257006
	$^2\chi$	−0.18411	0.00775	−23.741	0.000000
2	Constant	0.18447	0.16798	1.0982	0.285164
	$^1\chi$	−0.14866	0.00845	−17.587	0.000000
	$^3\chi_c^{\nu}$	−0.19248	0.06861	−2.806	0.010922
3	Constant	0.44266	0.20543	2.155	0.042935
	$^0\chi^{\nu}$	−0.12317	0.00613	−20.091	0.000000
4	Constant	0.56045	0.27897	2.009	0.057569
	$^1\chi^{\nu}$	−0.20443	0.01344	−15.211	0.000000
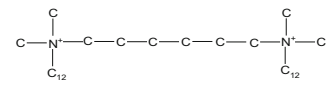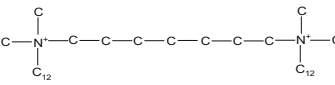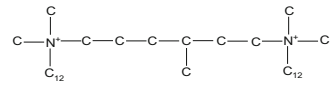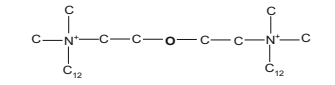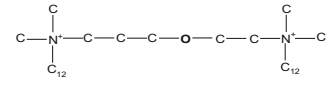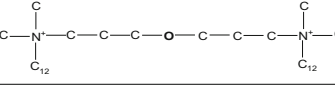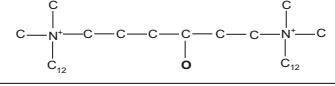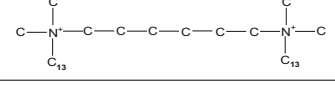
High absolute Student t values of the descriptors express that the regression coefficients of the descriptors are significantly larger than the standard error. Descriptors with p values below 0.05 (95 % confidence) are considered statistically significant [4].

As follows from Table 6, all the descriptors are statistically significant.

Appendix 3

The hydrogen-suppressed graphic structural formulas of some gemini surfactants, and the corresponding calculated, using Eqs. 7–10, logcmc values are contained in Table 7.

Table 7 Hydrogen-suppressed structural formulas of some gemini surfactants and calculated logcmc values

Compound		Model 1	Model 2	Model 3	Model 4
1	(structure)	-2.770	-2.807	-2.845	-2.908
2	(structure)	-2.835	-2.881	-2.932	-3.011
3	(structure)	-2.884	-2.921	-2.952	-2.989
4*	(structure)	-2.705	-2.733	-2.721	-2.720
5	(structure)	-2.770	-2.807	-2.808	-2.822
6	(structure)	-2.835	-2.881	-2.895	-2.924
7	(structure)	-2.884	-2.890	-2.884	-2.923
8	(structure)	-2.900	-2.956	-3.019	-3.113
9	(structure)	-3.006	-3.035	-3.060	-3.069
10	(structure)	-2.770	-2.807	-2.772	-2.736
11	(structure)	-2.900	-2.956	-2.946	-2.940
12	(structure)	-3.006	-2.974	-2.923	-2.939

[a] For this compound, the experimental values of cmc are 0.5 (mM) at 20 °C [23] and 2.2 (mM) at about 23 °C [29] and the calculated values of cmc using models1–4 are about 1.9 (mM) for 25 °C

References

1. Creton B, Nieto-Draghi C, Pannacci N (2012) Prediction of surfactants properties using multiscale molecular modeling tools: a review. Oil Gas Sci Tech 67:969–982. doi:10.2516/ogst/2012040

2. Yuan S, Cai Z, XuG JY (2002) Quantitative structure-property relationships of surfactants: prediction of the critical micelle concentration of nonionic surfactants. Colloid PolymSci 280:630–636. doi:10.1007/s00396-002-0659-2

3. Li X, Zhang G, Dong J, Zhou X, Yan X, Luo M (2004) Estimation of critical micelle concentration of anionic surfactants with QSPR approach. J MolStruct (THEOCHEM) 710:119–126. doi:10.1016/j.theochem.2004.08.039

4. Xu J, Zhu L, Fang D, Liu L, Wang L, Xu W (2013) Prediction of dielectric dissipation factors of polymers from cyclic dimmer structure using multiple linear regression and support vector machine. Colloid PolymSci 291:551–561. doi:10.1007/s00396-012-2743-6

5. Bortolotti M, Brugnara M, Della Volpe C, Maniglio D, Siboni S (2006) Molecular connectivity methods for the characterization of surface energetics of liquids and polymers. J Colloid Interface Sci 296:292–308. doi:10.1016/j.jcis.2005.09018

6. Wang Z-W, Feng J-L, Wang H-J, Cui Z-G (2005) Effectiveness of surface tension reduction by nonionic surfactants with quantitative structure-property relationship approach. J DispersSciTechnol 26:441–447. doi:10.1081/DIS-200054572

7. Zhen L, Liu K, Huang D, Ren X, Li R (2016) Structure-property relationship of sulfosuccinic acid diester sodium salt micelles: 3D-QSAR model and DPD simulation. J Dispersion SciTechol 37:941–948. doi:10.1080/01932691.2015.1073601

8. Roy K, Kabir H (2012) QSPR with extended topochemical atom (ETA) indices: modeling of critical micelle concentration of nonionic surfactants. ChemEngSci 73:86–98. doi:10.1016/j.ces.2012.01.005

9. Roy K, Kabir H (2012) QSPR with extended topochemical atom (ETA) indices, 3: modeling of critical micelle concentration of cationic surfactants. ChemEngSci 81:169–178. doi:10.1016/j.ces.2012.07.008

10. Kier LB, Hall LH (1986) Molecular connectivity in structure-activity analysis. Research Studies Press Ltd, Letchworth

11. Mozrzymas A, Różycka-Roszak B (2011) Prediction of critical micelle concentration of cationic surfactants using connectivity indices. J Math Chem 49:276–289. doi:10.1007/s10910-010-9738-7

12. Mozrzymas A (2013) Modelling of the critical micelle concentration of cationic Gemini surfactants using molecular connectivity indices. J SolutChem 42:2187–2199. doi:10.1007/s10953-013-0095-6

13. Mozrzymas A (2016) On the spacer group effect on critical micelle concentration of cationic gemini surfactants using molecular connectivity indices. Comb Chem High Throughput Screen 19:481–488. doi:10.2174/1386207319666160504095717

14. Wang Z, Li G, Zang X, Wang R, Lou A (2002) A quantitative structure-propertyrelationship study for the prediction of critical micelle concentration of nonionic surfactants. Colloids Surfaces A: PhysEng Aspects 197:37–45. doi:10.1016/S0927-7757(01)00812-3

15. Wiener H (1947) Structural determination of paraffin boiling points. J Am ChemSoc 69:17–20. doi:10.1021/ja01193a005

16. Randic M (1975) On characterization of molecular branching. J Am ChemSoc 97:6609–6615. doi:10.1021/ja00856a001

17. Wegrzyńska J, Chlebicki J, Maliszewska I (2007) Preparation, surface-active properties and antimicrobial activities of bis(ester quaternary ammonium) salts. J SurfactDeterg 10:109–116. doi:10.1007/s11743-007-1020-z

18. Han Y, Wang Y (2011) Aggregation behaviour of gemini surfactants and their interaction with macromolecules in aqueous solution. PhysChemChemPhys 13:1939–1956. doi:10.1039/c0cp01196g

19. Wegrzyńska J, Chlebicki J (2006) Preparation, surface-active properties and antielectrostatic properties of multiple quaternary ammonium salts. J SurfactDeterg 9:221–226. doi:10.1007/s11743-006-5000-5

20. Quagliotto P, Viscardi G, Barolo C, Barni E, Bellinvia S, Fisicaro E, Compari C (2003) Geminipyridinium surfactants: synthesis and conductometric study of a novel class of amphiphiles. J Org Chem 68(20):7651–7660. doi:10.1021/jo034602n

21. Ding Z, Hao A (2010) Synthesis and surface properties of novel cationic gemini surfactants. J Dispersion SciTechol 31:338–342. doi:10.1080/01932690903192580

22. Contrera JF, MacLaughlin P, Hall LH, Kier LB (2005) QSAR modeling of carcinogenic risk using discriminant analysis and topological molecular descriptors. Current Drug Discovery Technol 2:55–67. doi:10.2174/1570163054064684

23. Kim T-S, Kida T, Nakatsuji Y, Hirao T, Ikeda I (1996) Surface-active properties of novel cationic surfactants with two alkyl chains and two ammonio groups. JAOCS 73:907–911. doi:10.1007/BF02517994

24. Wettig SD, Li X, Verrall RE (2003) Thermodynamic and aggregation properties of gemini surfactants with ethoxylated spacers in aqueous solution. Langmuir 19:3666–3670. doi:10.1021/la0340100

25. Wettig SD, Verrall RE (2001) Thermodynamic studies of aqueous *m-s-m*gemini surfactants systems. J Colloid Interface Sci 235:310–316. doi:10.1006/jcis2000.7348

26. Akbar J, Tavakoli N, Marangoni DG, Wettig SD (2012) Mixed aggregate formation in gemini surfactant 1,2-dialkyl-sn-glycero-3-phosphoethanolamine systems. J Colloid Interface Sci 377:237–243. doi:10.1016/j.jcis.2012.03.048

27. Hu Z, Zhu H, Wang J, Cao D (2016) Surface activities of three anionic gemini surfactants derived from cyanuric chloride: effect of a branched hydrophobic chain. J SurfactDeterg 19:487–492. doi:10.1007/s11743-016-1812-0

28. Wettig SD, Nowak P, Verrall RE (2002) Thermodynamic and aggregation properties of gemini surfactants with hydroxyl substituted spacers in aqueous solution. Langmuir 18:5354–5359. doi:10.1021/la011782s

29. Laschewsky A, Lunkenheimer K, Rakotoaly RH, Wattebled L (2005) Spacer effect in dimeric cationic surfactants. Colloid PolymSci 283:469–479. doi:10.1007/s00396-004-1219-8