QSPR MODELS FOR PREDICTING CRITICAL MICELLE CONCENTRATION OF GEMINI CATIONIC SURFACTANTS COMBINING MACHINE-LEARNING METHODS AND MOLECULAR DESCRIPTORS

Ely Setiawan[1,2,*], Mudasir[1], and Karna Wijaya[1]

[1,2,3]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Indonesia
[1]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Indonesia
[1]Austrian-Indonesian Centre for Computational Chemistry (AIC), Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Indonesia
*E-mail : : ely.setiawan@mail.ugm.ac.id
Mobile No.: +62-896-0561-5152
Address for Postal Correspondance: Austrian-Indonesian Centre for Computational Chemistry (AIC) FMIPA UGM, Sekip Utara Bulaksumur 21 Depok, Sleman Yogyakarta Indonesia 55281

ABSTRACT

A data set of 231 diverse gemini cationic surfactants has been developed to correlate the logarithm of critical micelle concentration (cmc) with the molecular structure using a quantitative structure-property relationship (QSPR) methods. The QSPR models were developed using the Online CHEmical Modeling environment (OCHEM). It provides several machine learning methods and molecular descriptors sets as a tool to build QSPR models. Molecular descriptors were calculated by eight different software packages including Dragon v6, OEstate and ALogPS, CDK, ISIDA Fragment, Chemaxon, Inductive Descriptor, SIRMS, and PyDescriptor. A total of 64 QSPR models were generated, and one consensus model developed by using a simple average of 13 top-ranked individual models. Based on the statistical coefficient of QSPR models, a consensus model was the best QSPR models. The model provided the highest $R^2 = 0.95$, $q^2 = 0.95$, RMSE = 0.16 and MAE = 0.11 for training set, and $R^2 = 0.87$, $q^2 = 0.87$, RMSE = 0.35 and MAE = 0.21 for test set. The model was freely available at https://ochem.eu/model/8425670 and can be used for estimation of cmc of new gemini cationic surfactants compound at the early steps of gemini cationic surfactants development.

Keywords: critical micelle concentration, gemini cationic surfactant, machine learning, OCHEM, QSPR

INTRODUCTION

Over the past two-decade, a new type of surfactants, gemini surfactants, have been increasingly investigated to many researchers for wide-range applications. These surfactants consist of two hydrophobic tails each attached to a hydrophilic head group connected by a linkage/spacer group[1]. The interest of gemini surfactants was gaining more attention due to the current report which found out that these surfactants exhibit exceptional surface and bulk properties, including unusual micelle structures, low Kraft point, high surface reduction efficiency, better wetting ability, and low critical micelle concentration (cmc) values comparable with corresponding single-chain surfactants[2,3].

Gemini cationic surfactants, which containing two positively charged headgroups, are generally used for most of the gemini surfactants research studies. The positively charged head groups are

quaternary ammonium, immidazolium, pyridinium, esterified quaternaries, etc[4]. These surfactants have been increasingly used in advances technology domains such as capping agents for the synthesis of nanoparticles and nanorods, enhanced oil recovery process, the fabrication of high-porosity materials, corrosion inhibitors, genetic science and pharmaceutical applications including gene delivery, drug delivery, and antimicrobial activity[5,6,7].

The most important parameter for measuring the surface activity of a surfactant is the determination of critical micelle concentration (cmc) values. The cmc is the concentration of surfactants when the micelles form in the solution[8]. Considered the ideal surfactant, the cmc values should be as low as possible. The cmc is the most important parameter of surfactant in solution; because it can be correlated with industrial characteristics of surfactant such as viscosity, foam stability, detergency, and dispersion ability[9,10]. The cmc value depends on many factors including temperature, pH, pressure, additive presence, and molecular structure of surfactants[11]. Usually, the cmc of surfactants is determined using experimental techniques, such as conductometry, fluorescence emission spectroscopy, cyclic voltammetry, NMR spectroscopy, tensiometry, etc[12].

Although the determination of cmc has been made very accurately in experimental methods, quantitative relationship analysis between molecular structure and physicochemical properties have been gaining as key methodologies to predict cmc of surfactants[13]. This technique known as quantitative structure-property relationships (QSPR) has already shown its predictive potential for a broad spectrum of properties. QSPR models are mathematical relationships that are defined by molecular descriptors between the molecular structure and the target property[14]. Several models such as multiple linear regressions (MLR) or artificial neural networks are used for developing the QSPR model[15].

Furthermore, with numerous experimental data and their physicochemical properties parameters, various QSPR models have been developed for cmc prediction of cationic surfactants[16,17,18]. However, the QSPR models for cmc prediction of gemini cationic surfactants based on structurally diverse of their hydrophilic head groups (quaternary ammonium, pyridinium, imidazolium, pyrrolidinium, and piperidinium) have not been reported. In the present work, we proposed a universal QSPR models for prediction of cmc of gemini cationic surfactants based on structurally diverse of their hydrophilic head groups. We used OCHEM - Online CHEmical Modeling environment platform (http://www.ochem.eu), as the main tool to build QSPR models[19]. The OCHEM platform allows combining several machine-learning methods with molecular descriptors sets. The developed models can be published through OCHEM website, which helps people to use them for cmc prediction of gemini cationic surfactants.

EXPERIMENTAL

Material and Methods

Data Preparation

The cmc dataset of gemini cationic surfactant were collected from multiple publications[20,21,22,23,24,25]. This dataset containing 231 compounds of gemini cationic surfactants including with their cmc value. These data were randomly divided into 183 compounds for a training set and 48 compounds for a test set (validation set). The cmc of gemini cationic surfactants were converted into a negative form of logarithm cmc (pCMC). Structural of gemini cationic surfactants were drawn by using MarvinSketch software and converted to canonical

SMILES (Simplified Molecular Input Line Entry System) format. The SMILES string of gemini cationic surfactants along with corresponding cmc value were then uploaded to the OCHEM database.

Procedures

Model development

We developed the QSPR models with eight machine-learning methods, including associative neural networks (ASNN)[26], multiple linear regression analysis (MLRA)[27], k nearest neighbors (kNN)[28], a library of support vector machine (LibSVM)[29], fast stepwise stagewise multivariateple linear regression (FSMLR)[30], deep neural network (DNN)[31], random forest regression (RFR)[32], and partial least squares (PLS)[33]. We used optimized parameters setting of each machine-learning method provided by OCHEM platform.

Molecular descriptors calculation and pruning

The OCHEM platform provides several software packages to calculate many different descriptor sets[34,35]. In this study, eight descriptor sets were used to build QSPR models, including OEstate and ALogPS, CDK, Chemaxon, Dragon v6, ISIDA Fragment, Inductive Descriptor, Mordred, and PyDescriptor. Details about the descriptors can be found elsewhere. Corina was used to optimizing 3D structures of gemini cationic surfactants. A simple pairwise correlation method was used as a filtering method to each descriptor sets before they were used as an input for the machine-learning logarithm. The number of selected descriptors after the filtering step is shown in Table-1.

Table-1. Number of selected descriptors for development QSPR models

Descriptor sets	Number of descriptors
AlogPS, OEstate	45
CDK	113
Dragon	850
ISIDA Fragment	71
Mordred	508
Inductive descriptors	33
PyDescriptors	519

Model validation and evaluation

A five-fold cross-validation method was used to evaluate the accuracy and robustness of the models internally[36]. While externally, the models were evaluated by external validation set (test set). To avoid incorrect estimation of the models due to over-fitting by the variable selection, the OCHEM platform repeats the cross-validation step for all steps of model development. Furthermore, the validation set (test set) was used to confirm the quality of the models.

Model performance

In this work, all QSPR models were evaluated with the squared correlation coefficient (R^2), the cross-validated coefficient (q^2), the root mean square error (RMSE), and the mean absolute error (MAE). The QSPR models are considered the value of $R^2>0.6$, $q^2>0.5$, lower RMSE and MAE value can be used to assess the properties of new compounds[37].

Applicability domain estimation

The critical problem that inhibits the application of the QSPR model is reliability in predicting new compounds, although the model has good predictive accuracy for the training set were used to create the model and test set compounds. This is why every QSPR model should have a defined applicability domain (AD), which is the chemical space that the prediction would be considered to be reliable[38]. The OCHEM platform can detect AD of all developed QSPR models for each new compound. In this study, distance to model (DM) was used as AD determination, and to measure the DM we used the CONSENSUS-STD (standard deviation of predictions of the ensemble of models in the consensus model)[35]. This DM technique provides the best results by separating molecules with low prediction accuracy and molecules with high prediction accuracy. This technique uses a threshold value of 95% of the compounds from the training set to determine the qualitative AD of the model.

RESULTS AND DISCUSSION

Results of QSPR model development

In this study, the training set of 183 gemini cationic surfactants was used for developing QSPR model. In the QSPR model development step, 64 QSPR models were developed using a combination of eight machine-learning algorithms and eight molecular descriptor sets. Thus, this can count as one of the most comprehensive up to date studies to predict cmc of gemini cationic surfactants. As a result, 13 top-ranked models were built with good quality and high accuracy for predicting cmc of gemini cationic surfactants. The statistical coefficient of these QSPR models have high R^2 = 0.90-0.94, q^2 = 0.88-0.94, and low RMSE = 0.18-0.24, MAE = 0.13-0.18.

Table-2. Statistical coefficient of QSPR models built from combination machine-learning algorithms and molecular descriptors set

Method	Descriptors set	Training set				Test set			
		R^2	q^2	RMSE	MAE	R^2	q^2	RMSE	MAE
ASNN	AlogPS, OEstate	0.93	0.93	0.20	0.14	0.85	0.85	0.38	0.24
ASNN	CDK	0.93	0.92	0.20	0.13	0.83	0.82	0.41	0.25
ASNN	Dragon	0.93	0.93	0.19	0.13	0.84	0.83	0.40	0.26
ASNN	ISIDA Fragment	0.93	0.92	0.20	0.13	0.85	0.85	0.38	0.24
ASNN	Inductive Descriptors	0.90	0.89	0.24	0.15	0.86	0.86	0.37	0.22
ASNN	Mordred	0.94	0.94	0.18	0.13	0.82	0.81	0.42	0.25
LibSVM	AlogPS, OEstate	0.93	0.93	0.19	0.14	0.85	0.85	0.37	0.28
LibSVM	CDK	0.91	0.90	0.23	0.14	0.86	0.86	0.36	0.22
LibSVM	Dragon	0.94	0.93	0.19	0.13	0.89	0.89	0.33	0.20
LibSVM	ISIDA Fragment	0.92	0.92	0.21	0.15	0.85	0.84	0.39	0.28
LibSVM	Mordred	0.93	0.93	0.20	0.14	0.88	0.88	0.33	0.18
LibSVM	PyDescriptors	0.92	0.92	0.20	0.14	0.89	0.88	0.33	0.19
DNN	PyDescriptors	0.90	0.89	0.24	0.18	0.80	0.78	0.45	0.28
Consensus		0.95	0.95	0.16	0.11	0.87	0.87	0.35	0.21

Several studies have shown that the consensus model provides better predictive ability compared to the individual models[39,40]. We decided to develop a consensus model using a simple average of 13 top-ranked individual models. The statistical coefficient of the consensus model showed better than the individual model (Table-2). It has R^2 = 0.95 and q^2 = 0.95, which is higher than all

the individual models, and the value of RMSE = 0.16 and MAE = 0.11, which smaller than all individual models. The graphical plot of the experimental and predicted pCMC for the training set and test set of gemini cationic surfactants using the consensus model is shown in Fig-1.

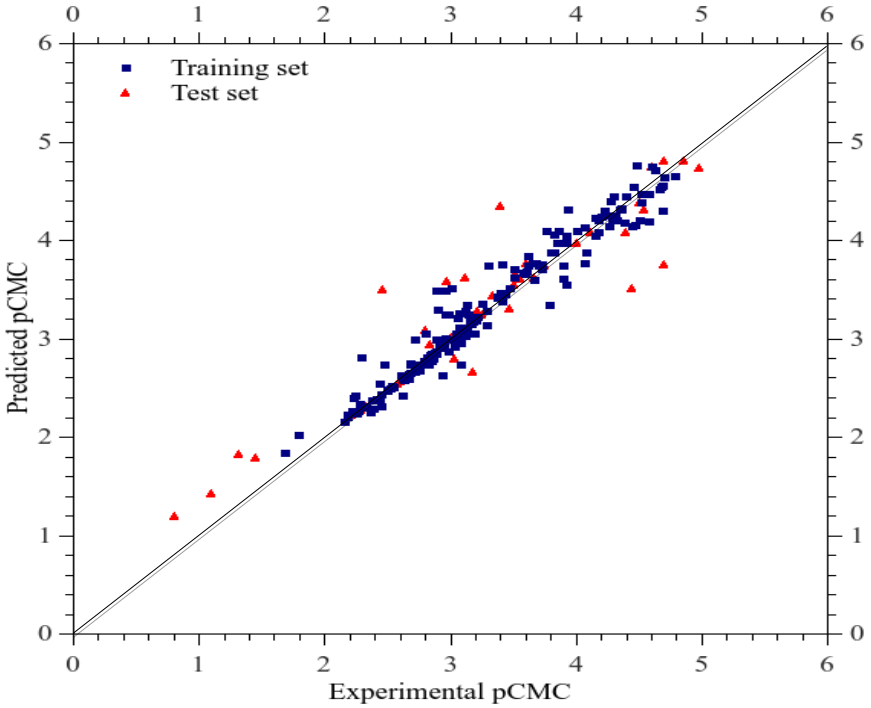

Fig.-1. Scatter plot of the experimental versus predicted pCMC using consensus model

Applicability Domain

Fig.-2. shows the standard deviation of an ensemble of consensus model (CONSENSUS-STD) used as a distance to model (DM) by William plot. The plot shows that almost all compounds in the test set were undercover the AD, except 5 compounds. These results prove that the QSPR models have a limitation on the prediction accuracies. However, the consensus model can become a useful tool to predict cmc of gemini cationic surfactants.

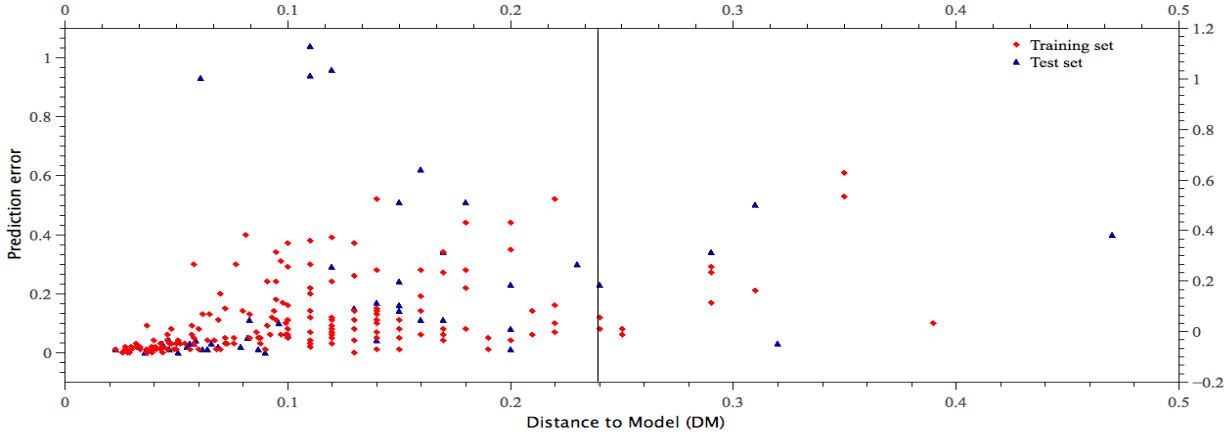

Fig.-2. William plot with CONSENSUS-STD as a distance to model (DM) for the consensus model. Vertical line is a threshold value of 95% of selected compounds from the training set.

Interpretable QSPR Models

To understand the structural descriptors that contributing to cmc value of gemini cationic surfactants, linear models were built using multiple linear regression (MLR) with different descriptors set, which can be shown in Table-3. The best linear model developed by MLR is model 2, which combination of MLR methods with CDK descriptors. Although the squared correlation coefficient (R^2) of model 2 is about 16% lower compared to the consensus model, the molecular descriptors of gemini cationic surfactants that affecting the cmc value could be explained easily with this model.

Table-3. Statistical coefficient of MLR QSPR models

MLR Model	Descriptors set	Training set				Test set			
		R^2	q^2	RMSE	MAE	R^2	q^2	RMSE	MAE
Model 1	AlogPS, OEstate	0.77	0.76	0.36	0.27	0.71	0.71	0.52	0.38
Model 2	**CDK**	**0.80**	**0.80**	**0.33**	**0.25**	**0.74**	**0.72**	**0.51**	**0.41**
Model 3	Chemaxon	0.75	0.75	0.37	0.29	0.69	0.68	0.55	0.43
Model 4	Dragon	0.71	0.70	0.40	0.30	0.79	0.76	0.47	0.36
Model 5	ISIDA Fragment	0.75	0.75	0.37	0.27	0.68	0.66	0.56	0.44
Model 6	Inductive Descriptors	0.77	0.76	0.36	0.27	0.78	0.76	0.48	0.35
Model 7	Mordred	0.74	0.73	0.38	0.28	0.68	0.65	0.57	0.42
Model 8	PyDescriptors	0.79	0.78	0.34	0.27	0.68	0.66	0.56	0.43

pCMC = -0.269 + 0.672*SC-3 + 0.34*XLogP - 18.5*RPCG - 0.724*ATSc1 - 1.24*ATSc2 + 0.163*RNCS+ 0.179*C1SP3 - 0.0928*ATSm1 + 0.0832*nAtomP + 18.3*tpsaEfficiency (1)

R^2 = 0.80, q^2 = 0.80, RMSE = 0.33, MAE = 0.25

Table-4. List of molecular descriptors involved in the best MLR model (model 2)

Symbol	Descriptor Class	Definition
SC-3	Chi Cluster	Simple cluster, order 3
XLogP	Constitutional	Prediction of logP based on the atom-type method
RPCG	CPSA	Relative positive charge – most positive charge / total positive charge
ATSc1	Autocorrelation	Broto-Moreau autocorrelation descriptor - lag 1 weight by charge
ATSc2	Autocorrelation	Broto-Moreau autocorrelation descriptor-lag 2 weight by charge
RNCS	CPSA	Relative negative charge surface area – most negative surface area
C1SP3	Carbon Type	Singly bound carbon to one other bound
ATSm1	Autocorrelation	Broto-Moreau autocorrelation descriptor-lag 1 weight by mass
nAtomP	Large Pi System	Number of atoms in the largest phi system
tpsaEfficiency	Topological	Polar survace area expressed as a ratio to molecular size

In the MLR model (Eq. 1) SC-3, XLogP, RNCS, C1SP3, nAtomP, and tpsaEfficiency descriptors indicate a positive contribution to the value of pCMC of gemini cationic surfactants. While RPCG, ATSc1, ATSc2 and ATSm1 have a negative contribution to the value of pCMC of gemini cationic surfactants. In other words, increasing SC-3, XLogP, RNCS, C1SP3, nAtomP

and tpsaEfficiency will increase pCMC and decreasing the RPCG, ATSc1, ATSc2 and ATSm1 will decrease pCMC of gemini cationic surfactants.

CONCLUSION

In conclusion, based on diverse hydrophilic head groups of gemini cationic surfactants dataset containing 231 gemini cationic surfactants, several QSPR models were developed with eight machine-learning algorithms and eight molecular descriptor sets using the OCHEM platform. Moreover, a consensus model was built based on the 13 top-ranked individual models. The consensus model showed the best models which provided provided the highest $R^2 = 0.95$, along with $q^2 = 0.95$, RMSE = 0.16 and MAE = 0.11 for training set, and $R^2 = 0.87$, $q^2 = 0.87$, RMSE = 0.35 and MAE = 0.21 for test set. The QSPR model was available at https://ochem.eu/model/8425670.

ACKNOWLEDGEMENT

We would like to thank Directorate Research and Community Service, Ministry of Research, Technology, and Hinger Education for supporting this research through a Doctoral Dissertation Research in 2019 with contract number 2884/UN1.DITLIT/DIT-LIT/LT/2019.

REFERENCES

[1] Kumar, V., Chatterjee, A., Kumar, N., Ganguly, A., Chakraborty, I., and Banerjee, M. *Carbohydrate Research.,* **397**, 37−45(2014), **DOI:** 10.1016/j.carres.2014.08.005

[2] S. Engin Ozdil, H. Akbar, M. Boz, *Journal of Chemical and Engineering Data*, **61**, 142−150(1997), **DOI:** 10.1021/acs.jced.5b00367

[3] B. Cai, J. F. Dong, L. Cheng, Z. Jiang, Y. Yang, X. F. Li, *Soft Matter.*, **9,** 7637−7646(2013), **DOI:** 10.1039/C3SM50916H

[4] A. Badhani, S. Singh, H. Sakai, M. Abe, H., Synthesis and Properties of Heterocyclic Cationic Gemini Surfactants, In Encyclopedia of Biocolloid and Biointerface Science 2V Set, Oshima (Ed) John Wiley & Sons, Inc., (2016) pp. 539−553, **DOI:** 10.1002/9781119075691.ch43

[5] H. Wang, T. Kaur, N. Tavakoli, J. Joseph, S. Wettig, *Physical Chemistry Chemical Physics,* **15**, 20510−20516(2013), **DOI:** 10.1039/C3CP52621F

[6] L. Casal-Dujat, M. Rodrigues, A. Yague, A. C. Calpena, D. B Amabilino, J. González-Linares, M. Borràs, L. Pérez-García, *Langmuir*., **28**, 2368−2381(2012), **DOI:** 10.1021/la203601n

[7] W. Wang, Y. Han, M. Tian, Y. Fan, Y. Tang, M. Gao, Y. Wang, *ACS Applied Materials and Interfaces*., **5,** 5709−5716(2013), **DOI:** 10.1021/am4011226

[8] R. Zana, J. Xia (Eds.), Gemini Surfactants: Synthesis, Interfacial and Solution-phase Behavior, and Applications (1st ed), Marcel Dekker Inc., New York (2004)

[9] R. Patel, M.U.H. Mir, U.K. Singh, I. Beg, A. Islam, A.B. Khan, *Journal of Colloid Interface Science*, **284**, 205−212(2016), **DOI:** 10.1016/j.jcis.2016.09.004

[10] M. Parray, N. Maurya, F.A. Wani, M.S. Borse, N. Arfin, M.A. Malik, R. Patel, *Journal of Molecular Structure*, **1175**, 49−55(2019), **DOI:** 10.1016/j.molstruc.2018.07.078

[11] Y. Zhang, W. Kong, P. An, S. He, X. Liu, *Langmuir*, **32**, 2311−2320(2016), **DOI:** 10.1021/acs.langmuir.5b04459

[12] P. Mukerjee, K.J. Mysels, Critical Micelle Concentrations of Aqueous Surfactant Systems; National Bureau of Standards: Washington, DC (1971)
[13] J. Hu, X. Zhang, Z. Wang, *International Journal of Molecular Sciences*, **11**, 1020−1047(2010), **DOI:** 10.3390/ijms11031020
[14] P. Iswanto, E.V.Y. Delsy, E. Setiawan, F.A. Putra, *Molekul*, **14**, 78−83(2019), **DOI:** 10.20884/1.jm.2019.14.2.467
[15] D. Wang, Y. Yuan, S. Duan, R. Liu, S. Gu, S. Zhao, L. Liu, J. Xu, *Chemometrics and Intelligent Laboratory Systems*, **143**, 7−15(2015), **DOI:** 10.1016/j.chemolab.2015.02.009
[16] A. Mozrzymas, *Colloid and Polymer Science*, **295**, 75−87(2017), **DOI:** 10.1007/s00396-016-3979-3
[17] A.R. Katritzky, L.M. Pacureanu, S.H. Slavov, D.A. Dobchev, D.O. Shah, M. Karelson, *Compututer & Chemical Engineering*, **33**, 321−332(2009), **DOI:** 10.1016/j.compchemeng.2008.09.011
[18] G. Absalan, B. Hemmateenejad, M. Soleimani, M. Akhond, R. Miri, *QSAR Combinatorial Science*, **23**, 416−425(2004), **DOI:** 10.1002/qsar.200430872
[19] I. Sushko, S. Novotarskyi, R. Korner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, Baskin, II, V. A. Palyulin, E. V. Radchenko, W. J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-de-Sousa, Q. Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. TkachenkoI. V. Tetko, *Journal of Compututer-Aided Molecular Design*, **25**, 533–54(2011), **DOI:** 10.1007/s10822-011-9440-2
[20] K. Taleb, M. Mohamed-Benkada, N. Benhamed, S. Saidi-Besbes, Y. Grohens, A. Derdour, *Journal of Molecular Liquids*, **241**, 81−90(2017), **DOI:** 10.1016/j.molliq.2017.06.008
[21] A. Badhani, S. Singh, H. Sakai, M. Abe, Synthesis and Properties of Heterocyclic Cationic Gemini Surfactants, 2016, In Encyclopedia of Biocolloid and Biointerface Science 2V Set, H. Oshima (Ed) John Wiley & Sons, Inc., pp. 539−553, **DOI:** 10.1002/9781119075691.ch43
[22] J. Hao, P. Wang, Y. Zhang, Y. Zhang, *Journal of Surfactants and Detergents*, **19**, 915−923(2016), **DOI:** 10.1007/s11743-016-1850-7
[23] B. Brycki, A. Szulc, H. Koenig, I. Kowalczyk, T. Pospieszny, S. Gorka, *Comptes Rendus Chimie*, **22**, 386−392(2019), **DOI:** 10.1016/j.crci.2019.04.002
[24] L. Palkowski, J. Blaszczynski, A. Skrzypczak, J. Blaszczak, K. Kozakowska, J. Wroblewska, S. Kozuszko, E. Gospodarek, J. Krysinski, R. Slowinski, *Chemical Biology & Drug Design*, **83**, 278−288(2014), **DOI:** 10.1111/cbdd.12236
[25] V. Chauhan, S. Singh, T. Kaur, G. Kaur, *Langmuir*, **31**, 2956−2966(2015), **DOI:** 10.1021/la5045267
[26] I.V. Tetko, Associative Neural Network, 2008, In: Livingstone D.J. (eds) Artificial Neural Networks. Methods in Molecular Biology™, vol 458. Humana Press, pp. 185-202, **DOI:** 10.1007/978-1-60327-101-1_10
[27] M. Shahlaei, A. Madadkar-Sobhani, A. Fassihi, L. Saghaie, D. Shamshirian, H. Sakhi, **21**, 100−115(2012), **DOI:** 10.1007/s00044-010-9501-4
[28] G.W. Kauffman, P.C. Jurs, *Journal of Chemical Information and Computer Sciences*, **41**, 1553−1560(2001), **DOI:** 10.1021/ci010073h
[29] C. Chang, C. Lin, *ACM Transactions on Intelligent System and Technology*, **2**, 1−27(2011)

[30] K. Roy, S. Kar, R.N. Das, Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Elsevier Academic Press, Amsterdam; Boston (2015), **DOI:** 10.1016/C2014-0-00286-9
[31] Y. Xu, J. Ma, A. Liaw, R.P. Sheridan, V. Svetnik, *Journal of Chemical Information and Modeling*, **57**, 2490−2504(2017), **DOI:** 10.1021/acs.jcim.7b00087
[32] L. Breiman, *Machine Learning*, **45**, 5-32(2001), **DOI:** 10.1023/A:1010933404324
[33] I.V. Tetko, V.Y. Tanchuk, *Journal of Chemical Information and Computer Sciences*, **42**, 1136−1145(2002), **DOI:** 10.1021/ci025515j
[34] D. Hodyna, V. Kovalishyn, I. Semenyuta, V. Blagodatnyi, S. Regalsky, L. Metelytsia, *Computational Biology and Chemistry*, **73**, 127−138(2018), **DOI:** 10.1016/j.compbiolchem.2018.01.012
[35] S. Vorberg, I.V. Tetko, *Molecular Informatics*, **33**, 73-85(2014), **DOI:** 10.1002/minf.201300030
[36] I.V. Tetko, I. Sushko, A.K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Oberg, R. Todeschini, D. Fourches, A. Varnek, *Journal of Chemical Information and Modeling*, **48**, 1733−1746(2008), **DOI:** 10.1021/ci800151m
[37] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, *International Journal of Drug Design and Discovery*, **2**, 511−519(2011)
[38] K. Roy, S. Kar, P. Ambure, *Chemometrics and Intelligent Laboratory Systems*, **145**, 22−29(2015), **DOI:** 10.1016/j.chemolab.2015.04.013
[39] A.K. Halder, *SAR and QSAR Environmental Research*, **29**, 911−933(2018), **DOI:** 10.1080/1062936X.2018.1529702
[40] X. Cui, J. Liu, J. Zhang, Q. Wu, X. Li, *Journal of Applied Toxicology*, **39**, 1224−1232(2019), **DOI:** 10.1002/jat.3808