CrossMark
click for updates

QSAR modeling for predicting the antifungal activities of gemini imidazolium surfactants against *Candida albicans* using GA-MLR methods

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

[1]Department of Chemistry, Universitas Jenderal Soedirman, Purwokerto, Indonesia.
[2]Department of Chemistry, Universitas Gadjah Mada, Yogyakarta, Indonesia.

ABSTRACT

This report presents a Quantitative Structure–-Activity Relationships (QSAR) analysis of gemini imidazolium surfactants against *Candida albicans*. Mordred software is used to calculate various types of molecular descriptors. The data set contains 70 structures of gemini imidazolium surfactants and is divided into training set (75%) and test set (25%) to perform cross-validation step. Genetic algorithm technique combined with multiple linear regression method (GA-MLR) was used to investigate the correlation between molecular descriptors and antifungal activity of gemini imidazolium surfactants. As a result, the best GA-MLR model consisting of two topological descriptors (GATS4se and BalabanJ) exhibits good fitting and internal validation with $R^2 = 0.9073$, $Q^2_{LOO} = 0.8941$, and $Q^2_{LMO} = 0.8908$. Also, it was confirmed by the external validation procedure with $R^2_{test} = 0.8988$ and $RMSE_{test} = 0.3557$, indicating that the obtained model was robust, reliable, and strong to predict the antifungal activity of gemini imidazolium surfactants. The GA-MLR-QSAR could be a useful tool for the initial development and design of novel gemini imidazolium surfactant as antifungal agents.

INTRODUCTION

The genus *Candida* is responsible for about 80% of infectious fungi in a hospital environment and is a relevant cause of bloodstream infections (Kabir *et al.*, 2012). The most invasive species, which are responsible for severe cases of candidiasis, include *Candida albicans*, *Candida parapsilosis,* *Candida glabrata,* and *Candida krusei*. The most reported clinical pictures related to candidiasis are cutaneous-mucous, visceral, and allergic (Fisher-Hoch and Hutwahner, 1995; Pfaller and Diekema, 2007).

Antifungal agents are an option for treating oral candidiasis, but the availability of antifungal drugs is less than antibacterial agents because eukaryotic fungal organisms are the same as mammalian cells, which makes selecting suitable antifungal targets a problem (Mayer *et al.*, 2013). In addition,

several antifungal drugs become resistant, including fluconazole, ketoconazole, and itraconazole (Lewis *et al.*, 2012). Therefore, the development of new antifungal drugs to treat cases of candidiasis is important.

In recent years, cationic gemini surfactants that contain two head groups with two aliphatic chains connected by spacers have been exploited (Menger and Keiper, 2000; Rosen and Tracy, 1998). They exhibit lower cytotoxicity, better surface properties, and better able to bind negatively charged substance correspond to the monomer surfactants in the same conditions (Brycki *et al.*, 2017; Sharma *et al.*, 2017; Shukla and Tyagi, 2006). Cationic gemini surfactants also have good antimicrobial activity (Bao *et al.*, 2017; Tatsumi *et al.*, 2014). Several studies have shown that the antimicrobial activity of the cationic gemini surfactant imidazolium chloride depends on its structure, for example, the length of the alkyl groups (Rath and Bai, 2016). The hydrophobic portion (alkyl chains) of the imidazolium cationic gemini surfactant interacts with the cell membrane, inducing cell membrane damage leading to cell lysis and death (Dolezal *et al.*, 2016). The results of the research by Palkowski *et al.*

Corresponding Author
Ely Setiawan, Department of Chemistry, Universitas Jenderal Soedirman,
Purwokerto, Indonesia. E-mail: ely.setiawan @ unsoed.ac.id

(2014), Butorac *et al.* (2011), and Kamboj *et al.* (2012) show that imidazolium cationic gemini surfactant has potential as an antimicrobial agent.

The process of designing a new drug is a time-consuming, expensive, and very complex process. This is a challenge for researchers to find strategies and efforts that are effective and economical in producing new drugs. One of the strategies developed to design new drug compounds is the computer-aided drug design (CADD) approach (Yu and MacKerell, 2017). The relatively low cost of CADD compared to traditional drug discovery methods makes the CADD method attractive to save costs and time required in the development of new drug compounds (Shim and MacKerell, 2011).

Ligand-based drug design and structure-based drug design are mainly two categories of CADD methods (Yu and MacKerell, 2017). The ligand-based drug design utilizes information on the physicochemical properties of several experimentally known active compounds as a basis for designing new compounds. Among these methods, there is a Quantitative Structure-Activity Relationships (QSAR) analysis that investigates the structures and molecular properties through chemoinformatic methods (Cherkasov *et al.*, 2014; Kubinyi, 1995; Roy *et al.*, 2015).

Considering the above, this work aimed to conduct a QSAR study based on gemini imidazolium surfactants synthesized by Palkowski *et al.* (2014) as a potential antifungal agent against *C. albicans*. Thus, through the most important physical-chemical parameters (descriptors), the models can be obtained that help in planning the synthesis of new gemini imidazolium surfactant with better antifungal activity.

MATERIALS AND METHODS

Dataset

Pałkowski *et al.* (2014) synthesized 70 gemini imidazolium chlorides, with antifungal activity against *C. albicans*. The antifungal activity values made available in [minimum inhibition concentration (MIC), the lowest concentration of surfactant which inhibits the growth of microorganisms, in mol/L] were converted into their respective pMIC (–log MIC). The total dataset of molecules was divided into training set of 75% and test set of 25% based on the diversity of antifungal activity (Table 1).

Descriptors calculation

Initially, all geometries were drawn and optimized using the Hartree–Fock method (*ab initio*) with basis set 3–21G which is implemented in the Gaussian 09 software. As a next step, the Gaussian output files were used by Mordred software to calculate various classes of descriptors such as constitutional, topological, and WHIM (Weighted Holistic Invariant Molecular) descriptors (Moriwaki *et al.*, 2018). Then filtering is done by eliminating descriptors that have constant and highly correlated values. The filtered descriptors are then used to build the QSAR models.

QSAR modeling and validation

The QSAR models were developed by a combination of genetic algorithms with multiple linear regression (GA-MLR) methods using the QSARINS software (Gramatica *et al.*, 2013). The selection of variables was carried out using the GA technique;

in this way, consistent models are obtained through an optimization process that considers the value of statistical parameters such as the correlation coefficient and standard deviation (Rogers and Hopfinger, 1994). Then, the descriptors generated based on these parameters were correlated with antifungal activity through the MLR method for the construction of the QSAR models.

The validation of statistical models is an important stage in the design of drugs based on QSAR techniques because it is guaranteed that the equations obtained have predictive power and are sufficiently reliable to be able to describe the structural changes associated with biological activity (Kiralj and Ferreira, 2009). In view of this, the model was evaluated by fitting criteria [R^2, the squared correlation coefficient of training set; $RMSE_{train}$, *Root Mean Square Error* in fitting; Friedman lack of fit (LOF)], internal validation (Q^2_{LOO}, *Leaving-One-Out* (LOO) cross-validation coefficient; Q^2_{LMO}, *Leaving-Many-Out* (LMO) cross-validation coefficient; $RMSE_{cv}$, *Root Mean Square Error* of cross-validation), and external validation (R^2_{test}, the squared correlation coefficient of test set; $RMSE_{test}$, *Root Mean Square Error* in test set; MAE_{tr}, *Mean Absolute Error* in test set). Also, the applicability domain (AD) was assessed using the Williams plot (Gramatica, 2007).

RESULTS AND DISCUSSION

The aim of this study was to find the correlation between structural parameters of gemini imidazolium surfactants and the antifungal activity against *C. albicans*. Based on the basic principles of QSAR analysis, the structural parameters of a compound are expressed by the molecular descriptors. To obtain those molecular descriptors, first, we have drawn the 3D model of each gemini imidazolium surfactant and then optimized the geometry by employing Hartree–Fock (ab initio) method with 3-21G basis set. We used the equilibrium geometry as input on Mordred software to calculate the molecular descriptors. As result, 835 molecular descriptors were generated.

The variable selection process using the GA technique followed by the MLR method led to a GA-MLR model which consists of two molecular descriptors. The obtained model is shown in Eq. 1. Table 2 shows the statistical parameters of the obtained model. The antifungal activity (pMIC) prediction results for gemini imidazolium surfactants of this model are summarized in Table 1. The correlation graph between the predicted and experimental antifungal activity (pMIC) shown in Figure 1 also shows a slope close to 1. This means that the resulting model can provide a good level of prediction.

pMIC = 23.1793 + 0.628 * GATS4se – 0.7791 * BalabanJ.   Eq. 1

The coefficients of the molecular descriptors in Eq. 1 suggest that the 2D autocorrelation descriptor, namely, Geary coefficient of lag 4 weighted by Sanderson EN (GATS4se), and the topological of Balaban index (BalabanJ) are the most influence descriptors to the antifungal activity of imidazolium gemini surfactants. The positive coefficient of GATS4se descriptor indicates that an increase in GATS4se leads to an increase in the antifungal activity of gemini imidazolium surfactants. The topological of BalabanJ has a negative coefficient, which indicates that an increase in BalabanJ leads to a decrease in the antifungal activity of gemini imidazolium surfactants.

Table 1. Chemical structure and antifungal activities values (pMIC) of the gemini imidazolium surfactants against *C. albicans* (*test set compounds).
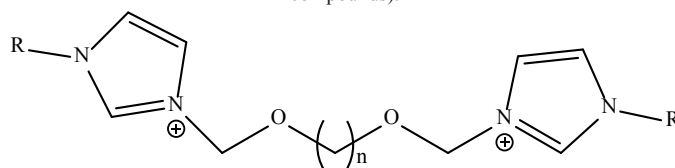
| Comp. ID	R	n	pMIC$_{(exp)}$	pMIC$_{(pred)}$	Comp. ID	R	n	pMIC$_{(exp)}$	pMIC$_{(pred)}$
1	CH_3	2	1.5279	1.6895	36	C_8H_{17}	4	4.0915	4.3753
2	C_2H_5	2	1.5626	1.3544	37	C_9H_{19}	4	4.1487	4.5123
3*	C_3H_7	2	1.7712	1.8827	38	$C_{10}H_{21}$	4	4.2676	4.5832
4	C_4H_9	2	2.1477	1.6895	39*	$C_{11}H_{23}$	4	4.4202	4.5172
5	C_5H_{11}	2	2.1647	1.6944	40	$C_{12}H_{25}$	4	4.4685	4.6542
6	C_6H_{13}	2	2.1968	1.9090	41	$C_{14}H_{29}$	4	4.5229	4.9282
7*	C_7H_{15}	2	2.2267	1.9799	42	$C_{16}H_{33}$	4	4.5229	5.0651
8	C_8H_{17}	2	2.4488	2.0773	43*	CH_3	5	4.5686	4.9036
9	C_9H_{19}	2	2.4821	2.3975	44	C_2H_5	5	4.6990	4.6346
10	$C_{10}H_{21}$	2	2.4821	2.5279	45	C_3H_7	5	4.7696	4.9598
11*	$C_{11}H_{23}$	2	2.4821	2.6715	46	C_4H_9	5	4.7696	4.5636
12	$C_{12}H_{25}$	2	2.4978	2.8283	47*	C_5H_{11}	5	4.7959	4.5685
13	$C_{14}H_{29}$	2	2.7867	3.1551	48	C_6H_{13}	5	4.8239	4.5734
14	$C_{16}H_{33}$	2	2.7867	3.5017	49	C_7H_{15}	5	4.8539	4.4413
15*	CH_3	3	2.7867	3.5583	50	C_8H_{17}	5	5.0000	4.5783
16	C_2H_5	3	2.8604	3.2892	51*	C_9H_{19}	5	5.0000	4.6493
17	C_3H_7	3	3.1331	3.5485	52	$C_{10}H_{21}$	5	5.0000	4.7863
18	C_4H_9	3	3.1475	3.3552	53	$C_{11}H_{23}$	5	5.0458	4.9233
19*	C_5H_{11}	3	3.1475	3.2941	54	$C_{12}H_{25}$	5	5.0458	4.8572
20	C_6H_{13}	3	3.1475	3.2990	55*	$C_{14}H_{29}$	5	5.0458	5.1312
21	C_7H_{15}	3	3.1612	3.3700	56	$C_{16}H_{33}$	5	5.0969	5.2682
22	C_8H_{17}	3	3.4622	3.4409	57	CH_3	6	5.0969	5.1116
23*	C_9H_{19}	3	3.7620	3.5779	58	C_2H_5	6	5.0969	4.9085
24	$C_{10}H_{21}$	3	3.7747	3.6489	59*	C_3H_7	6	5.0969	5.2338
25	$C_{11}H_{23}$	3	3.8013	3.7858	60	C_4H_9	6	5.0969	4.8376
26	$C_{12}H_{25}$	3	3.8013	3.9228	61	C_5H_{11}	6	5.0969	4.9085
27*	$C_{14}H_{29}$	3	3.9355	4.2628	62	C_6H_{13}	6	5.0969	4.7104
28	$C_{16}H_{33}$	3	3.9666	4.5368	63*	C_7H_{15}	6	5.1549	4.7814
29	CH_3	4	3.9747	4.6957	64	C_8H_{17}	6	5.3010	4.7153
30	C_2H_5	4	3.9830	4.2236	65	C_9H_{19}	6	5.3010	4.7863
31*	C_3H_7	4	3.9914	4.6859	66	$C_{10}H_{21}$	6	5.3979	4.9233
32	C_4H_9	4	4.0000	4.2896	67*	$C_{11}H_{23}$	6	5.3979	4.8572
33	C_5H_{11}	4	4.0269	4.2285	68	$C_{12}H_{25}$	6	5.6990	4.9942
34	C_6H_{13}	4	4.0655	4.2994	69	$C_{14}H_{29}$	6	5.6990	5.3342
35*	C_7H_{15}	4	4.0862	4.3044	70	$C_{16}H_{33}$	6	5.6990	5.4052

The correlation matrix between selected descriptors shows that the correlation between GATS4se and BalabanJ is very low (Table 3). This indicates that there is no significant intercorrelation among the descriptors used in the development of the model. Additionally, the residual predicted pMIC using Eq. 1 versus the experimental value of pMIC is shown in Figure 2. All residual predicted pMIC values are located between 1 and −1, which indicates that Eq. 1 has good accuracy and reliability for predicting the antifungal activity of gemini imidazolium surfactants against *C. albicans*.

Based on the validation parameters of the model (Table 2), the model satisfies the requirements made by Golbraikh *et al.* (2002) and Roy *et al.* (2012). The values of R^2 (0.9073) and Q^2_{LOO} (0.8941) were reasonable, showing that the model was

Table 2. Validation parameters of model.

Fitting criteria		Internal validation		External validation	
R^2	0.9073	Q^2_{LOO}	0.8941	R^2_{test}	0.8988
$RMSE_{train}$	0.3455	Q^2_{LMO}	0.8908	$RMSE_{test}$	0.3557
LOF	0.1397	$RMSE_{cv}$	0.3694	MAE_{test}	0.2963
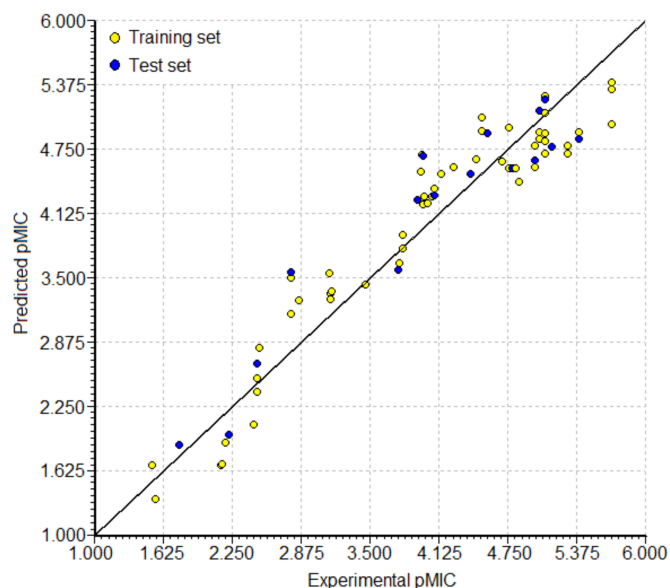
Figure 1. Plot of experimental and predicted value of pMIC for training and test set.
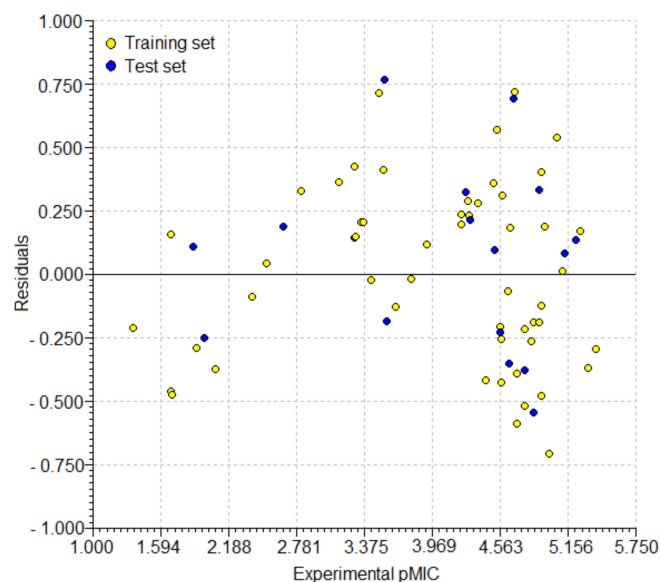
Figure 2. Plot of residuals of predicted pMIC determined using Eq. 1 versus experimental pMIC.

Table 3. Variable correlation matrix of model.

Descriptor	GATS4se	BalabanJ
GATS4se	1	
BalabanJ	0.0988	1

significant and robust to predict the antifungal activity of gemini imidazole surfactants. Consider that the values of the difference between R^2 and Q^2_{LOO} ($R^2–Q^2_{LOO}$ = 0.0132) are within the limit suggested by Kiralj and Ferreria (2009) which is an indication that the model does not have data overfitting. The low value of the LOF parameter (LOF = 0.1397) implies a good fit model with no current overfitting in the model.

Validation of the final model consists mainly of internal and external validation. The LOO and LMO cross-validation procedures were used for internal validation. According to Table 2, the model was stable and internally good for predictive results, because the values of Q^2_{LOO} and Q^2_{LMO} are comparable to R^2. To validate the accuracy of the prediction model, it has been confirmed that the statistical parameters contribute to the external validity of the model. The coefficient of determination value (R^2_{test} = 0.8988) and error parameters value ($RMSE_{test}$ = 0.3557 and MAE_{test} = 0.2963) of the test set showed that the obtained model is good and reliable.

After generating and evaluating the model, AD was employed to confirm that the obtained model can be considered reliable. Williams plot or leverage approach was used to measure the influence of descriptors on the model (Gramatica, 2007). The leverage value (h_i) shows the distance of a compound from the centroid of X, which is defined as

$$h_i = X_i \, (X^T X)^{-1} \, X_i^T,$$

where X is the characteristic matrix of the training set. The critical leverage value (h^*) is defined as

$$h^* = 3 \frac{(p+1)}{n},$$

where p is the number of descriptors in the model and n is the total number of compounds in the training set. As shown in Figure 3, all compounds lied within the domain of applicability which lower than the threshold leverage (h^* = 0.170). This indicated that no compounds in the dataset fell outside of the AD as an outlier.

The Y-scrambling method was employed as randomization tests to confirm that there was no random correlation between antifungal activity and selected descriptors (Rücker *et al.*, 2007). This criterion is shown by the average value of R^2Y_{scr} and Q^2Y_{scr}, which are both lower than R^2 and Q^2 of the model. In this work as shown in Figure 4, the values of R^2 and Q^2 of the model are higher than the R^2Y_{scr} and Q^2Y_{scr} values, which indicates that the model is not derived from random correlation.

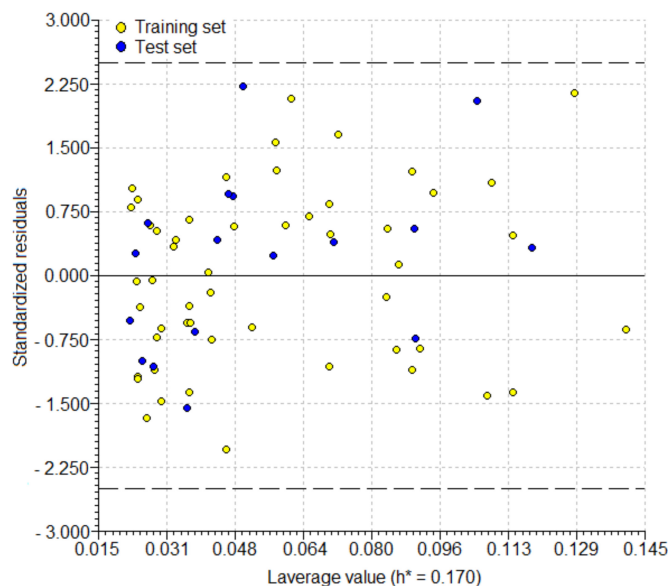**Figure 3.** William's plot of the QSAR model.

Figure 4. $Y_{scramble}$ plot of the obtained model.

CONCLUSION

It is concluded, through this work, that the GA-MLR analysis showed that the two highlighted descriptors play the role of antifungal activity of gemini imidazolium surfactants, namely, GATS4se (Geary coefficient of lag 4 weighted by Sanderson EN) and BalabanJ (topological of Balaban index). The obtained QSAR model is significant and robust, does not show random correlation, and has a strong predictive ability ($R^2 = 0.9073$, $Q^2_{LOO} = 0.8941$, and $R^2_{test} = 0.8988$). The AD indicates that most structures are adequately represented by the chemical space of the model. Thus, the values of the predicted activities can be considered reliable.

CONFLIC OF INTEREST

The authors declared that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work. All the authors are eligible to be an author as per the international committee of medical journal editors (ICMJE) requirements/guidelines.

ETHICAL APPROVALS

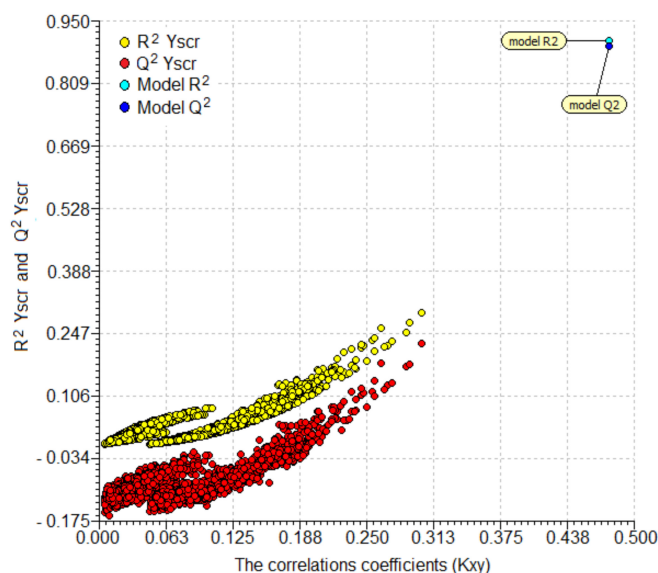This study does not involve experiments on animals or human subjects.

PUBLISHER'S NOTE

This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

REFERENCES

Bao Y, Guo J, Ma J, Li M, Li X. Physicochemical and antimicrobial activities of cationic gemini surfactants with polyether siloxane linked group. J Mol Liq, 2017; 242:8–15.

Brycki B, Koziróg A, Kowalczyk, I, Pospieszny, T, Materna, P, Marciniak, J. Synthesis, structure, surface, and antimicrobial properties of new oligomeric quaternary ammonium salts with aromatic spacers. Molecules, 2017; 22:1810.

Butorac RR, Al-Deyab SS, Cowley AH. Syntheses, structures, and antimicrobial activities of bis(imino) acenaphthene (BIAN) imidazolium salts. Molecules, 2011; 16:3168–78.

Cherkasov A, Muratov EN, Fourches D. QSAR modeling: where have you been? Where are you going to? J Med Chem, 2014; 57(12):4977–5010.

Dolezal R, Soukup O, Malinak D, Savedra RML, Marek J, Dolezalova M, Pasdiorova M, Salajkova S, Korabecny J, Honegr J, Ramalho TC, Kuca K. Towards understanding the mechanism of action of antibacterial N-alkyl-3-hydroxypyridinium salts: biological activities, molecular modeling and QSAR studies. Eur J Med Chem, 2016; 121:699–711.

Fisher-Hoch SP, Hutwagner L. Opportunistic candidiasis: an epidemic of the 1980s. Clin Infect Dis, 1995; 21:897–904.

Golbraikh A, Tropsha A. Beware of q²! J Mol Graph Model, 2002; 20:269–76.

Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci, 2007; 26:694–701.

Gramatica, P, Chirico N, Papa E, Cassani S, Kovarich S. QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. J Comput Chem, 2013; 34(24):2121–32.

Hoque I, Chatterjee A, Bhattacharya S, Biswas R. An approach of computer-aided drug design (CADD) tools for *in silico* pharmaceutical drug design and development. Int J Adv Res Biol Sci, 2017; 4:60–71.

Kabir MA, Hussain MA, Ahmad Z. *Candida albicans*: a model organism for studying fungal pathogens. ISRN Microbiol, 2012; 2012:1–12.

Kamboj R, Singh S, Bhadani A, Kataria H, Kaur G. Gemini imidazolium surfactants: synthesis and their biophysiochemical study. Langmuir, 2012; 28(33):11969–78.

Kiralj R, Ferreira MMC. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. J Braz Chem Soc, 2009; 20(4):770–87.

Kubinyi, H. QSAR: Hansch analysis and related approaches. Weinheim: Wiley-VCH, Germany, 1993.

Lewis RE, Viale P, Kontoyiannis DP. The potential impact of antifungal drug resistance mechanisms on the host immune response to Candida. Virulence, 2012; 3:368–76.

Mayer FL, Wilson D, Hube B. *Candida albicans* pathogenicity mechanisms. Virulence, 2013; 4:119–28.

Menger FM, Keiper JS. Gemini surfactants. Angew Chem Int Ed, 2000; 39:1906–20.

Moriwaki H, Tian YS, Kawashita N, Takagi T. Mordred: a molecular descriptor calculator. J Cheminform, 2018; 10(4).

Palkowski L, Krysinski J, Blaszczynski J, Slowinski R, Skrzypczak A, Blaszczak J, Gospodarek E, Wróblewska J. Application of rough set theory to prediction of antimicrobial activity of bis-quaternary imidazolium chlorides. Fundam Inform, 2014; 132:315–30.

Pfaller MA, Diekema DJ. Epidemiology of invasive candidiasis: a persistent public health problem. Clin Microbiol Rev, 2007; 20:133–63.

Rath EC Bai Y. Quantitative structure-activity relation study of quaternary ammonium compounds in pathogen control: computational methods for the discovery of food antimicrobials. Chem Inform, 2016; 2:1–7.

Rogers D, Hopfinger AJ. Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships. J Chem Inf Comput Sci, 1994; 34:854–66.

Rosen MJ, Tracy DJ. Gemini surfactants. J Surfactants Deterg, 1998; 1:547–54.

Roy K, Kar S, Das RN. 2015. A primer on QSAR/QSPR modeling. Springer International Publishing, Cham, Switzerland.

Roy K, Mitra I, Kar S, Ojha PK, Das RN, Kabir H. Comparative studies on some metrics for external validation of QSPR models. J Chem Inf Model, 2012; 52:396–408.

Rücker C, Rücker G, Meringer M. y-Randomization and its variants in QSPR/QSAR. J Chem Inf Model, 2007; 47(6):2345–57.

Sharma R, Kamal A, Abdinejad M, Mahajan RK, Kraatz HB. Advances in the synthesis, molecular architectures and potential applications of gemini surfactants. Adv Colloid Interface Sci, 2017; 248:35–68.

Shim J, Mackerell AD, Jr. Computational ligand-based rational design: role of conformational sampling and force fields in model development. Medchemcomm, 2011; 2:356–70.

Shukla D, Tyagi VK. Cationic gemini surfactants: a review. J Oleo Sci, 2006; 55:381–90.

Tatsumi T, Imai Y, Kawaguchi K, Miyano N, Ikeda I. Antimicrobial activity of cationic Gemini surfactant containing an oxycarbonyl group in the lipophilic portion against gram-positive and gramnegative microorganisms. J Oleo Sci, 2014; 63:137–40.

Yu W, MacKerell AD, Jr. Computer-aided drug design methods. Methods Mol Biol, 2017; 1520:85–106.