Environmental Science

AIMS Environmental Science, 12(3): 419–434.
DOI: 10.3934/environsci.2025019
Received: 20 August 2024
Revised: 06 January 2025
Accepted: 08 May 2025
Published: 19 May 2025

https://www.aimspress.com/journal/environmental

Research article

QSAR study and theoretical investigation on the lethality of halogenated aliphatic hydrocarbons toward *Aspergillus* (*A.*) *Nidulans*

Jabir H. Al-Fahemi[1,*], Faten A. Aljiffrey[1], Elshafie A. M. Gad[2] and Mahmoud A. A. Ibrahim[3,4,5,*]

[1] Department of Chemistry, Faculty of Science, Umm Al-Qura University, Makkah 21955, Saudi Arabia
[2] Petrochemicals Department, Egyptian Petroleum Research Institute, Cairo, Egypt
[3] Computational Chemistry Laboratory, Chemistry Department, Faculty of Science, Minia University, Minia 61519, Egypt
[4] Department of Engineering, College of Engineering and Technology, University of Technology and Applied Sciences, Nizwa 611, Sultanate of Oman
[5] School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4000, South Africa

* **Correspondence:** Email: jhfahemi@uqu.edu.sa, m.ibrahim@compchem.net.

Abstract: The prediction of *Aspergillus* (*A.*) *nidulans* toxicities ($\log 1/D_{37}$) for a set of 55 halogenated aliphatic hydrocarbons (HAHs) was thoroughly investigated using density functional theory (DFT) computations. Different multiple linear regression (MLR)methods were employed to assess the reliability of the proposed quantitative structure-activity relationships (QSAR) model. The obtained E_{LUMO}, E_{gap}, molecular polarizability (α), and molar refractivity (MR) values offered informative indications in determining the toxicity of the HAHs. A promising three-descriptor linear model was constructed using 41 molecules as a training set; then, the model was validated on the remaining 14 molecules. Statistical comparisons between these models and others quoted from the literature were presented. Furthermore, the potential causes of the outlier molecules in the proposed QSAR models were explored. The most preferable interactions were obviously noticed within the 1-bromo-2-methylpropane···α-glucan complex, followed by 2-bromo-2-methylpropane···α-glucan and 2-chloro-2-methylpropane···α-glucan complexes. Compared to other analogs, the higher number of bond paths and bond critical points within the 1-bromo-2-methylpropane···α-glucan complex highlighted its high preferability.

Keywords: QSAR; DFT; MLR; Descriptors;QTAIM

1. Introduction

Halogenated aliphatic hydrocarbons (HAHs) are of growing concern because of their effectual toxic and carcinogenic features [1,2]. Nevertheless, HAHs are commonly utilized as industrial and home solvents, chemical synthesis intermediates, and have a range of other applications. Within the literature, HAHs were addressed with a potent versatility to react and covalently bind with DNA, thus causing genetic damage in versatile experimental organisms.

In this regard, developing new models capable of promptly screening for potential hazards of any conceivable substance in the aquatic habitat has become an essential demand. Among the developed models, there has been a continued interest in estimating the toxicity data via quantitative structure-activity/-toxicity relationships (QSAR/QSTR) [3–6]. Furthermore, the QSAR model attempts to link molecular structures to biological endpoints such as toxicity [7]. These findings are considerably distinct from the well-known *in vivo* data, which would be attributed to a number of toxic action mechanisms. In contrast to *in vivo* observations, *in vitro* toxicity data are easy to collect and often have a direct mechanistic significance. Consequently, the QSAR concept is one of the ligand-based drug design strategies that has effectively led to the development of new drug candidates for a wide range of diseases [8].

Based on the literature, the QSAR can be mechanistically meaningful by using suitable molecular descriptors to mediate toxicity [9]. Except for the toxic effects caused by receptor binding, toxicity might be described as an outcome of the chemical compound's capacity to reach and covalently interact at the active site [5]. Considering a common biophysical mechanism, variations in the chemical structure of a set of known compounds are linked to toxicity alterations. Thence, the QSAR is utilized to extrapolate into other compounds.

As a point of departure, Crebelli et al. [3] presented a QSAR model based on the molar refractivity (*MR*) and energy gap (E_{gap}) for the prediction of 55 haloalkenes towards *Aspergillus* (*A*.) *nidulans*. This regression model was characterized by R^2 and *F* values of 0.64 and 45.5, respectively, and outlined the effectual role of the electrophilicity in predicting the aneugenic potential of the aliphatic hydrocarbons. Furthermore, the toxicity of 52 halogenated hydrocarbons toward *A*. *nidulans* was well-characterized by Trohalaki et al. using descriptors, including polarizability (α), lowest unoccupied molecular orbital energy (E_{LUMO}), and molecular volume (*V*) [4]. The most preferential model was detected with an R^2 value of 0.79 and an *F* value of 0.59, thereby utilizing the B3LYP/6-31G** and HF/6-31G** levels of theory. Afterward, the lethal effect of 55 aliphatic molecules using descriptors as the logarithm of the octanol-water partition coefficient (log *P*) and the E_{LUMO} was addressed by Cronin et al. [10]. Subsequently, this two-descriptors-based regression model was characterized by R^2 and standard error values of 0.615 and 0.413, respectively.

While QSAR models are widely used, there is a lack of predictive models which specifically target the toxicity of HAHs, and this gap necessitates further research. Therefore, the current study aims to construct QSAR models to predict *A. nidulans* toxicity of 55 halogenated aliphatic

hydrocarbons. A possible source of outlier molecules in the developed models is proposed along with the predictions.

2. Materials and methods

2.1. Data set

The utilized data on *A. nidulans* toxicity ($\log 1/D_{37}$) of 55 halogenated aliphatic hydrocarbons were previously measured . In particular, the data set splitting was performed using random division as the standard method. Based on the literature [11], the rational division methods produced better statistical outcomes for the test sets compared to models that used random division; however, the predictive powers of random and rational models are similar. Therefore, the selected halogenated hydrocarbons were randomly partitioned into two sets, namely, training and validation sets, which involve 41 and 14 molecules, respectively. The latter set was herein devoted to thoroughly confirming the accuracy of the developed QSTR model.

2.2. Quantum chemical calculations and selected descriptors

Herein, all executed quantum chemical calculations were performed using the Amsterdam density functional program package (ADF 2010.02) [12,13]. Geometry optimization was first carried out for each molecule using the quadruple-ξ Slater basis set with four polarization functions (STO-QZ4P) in conjunction with the generalized gradient approximation (G_{GA}) within the PW91 exchange and correlation functional [14,15]. The inner electrons within the atomic shells were treated as a frozen-core approximation to accelerate the computations. In the realm of the DFT calculations, various quantum chemical parameters were calculated, including the energies of molecular orbitals (E_{HOMO} and E_{LUMO}) along with the energy gap ($E_{gap} = E_{LUMO} - E_{HOMO}$). Afterward, HyperChem (version 8.0) was employed to compute various physical properties, including the surface area grid (S) octanol-water partition coefficient ($\log P$), molecular volume (V), molar refractivity (MR), and molecular weight (M). Additionally, the molecular ovality (O) [16] was computed as a fundamental indicator to represent how the molecular shape approaches a sphere, which can be defined as follows:

$$O = S/4\pi\left(\frac{3V}{4\pi}\right)^{2/3} \quad (1)$$

2.3. Statistical analysis

For the statistical analysis, a principal component analysis (PCA) and multiple linear regression (MLR) were adopted [17–19]. To study the correlations among the variables, the PCA was executed using XLSTAT. The PCA method was used to study the individual correlation between the $\log 1/D_{37}$ and the molecular descriptors and excluded any variables that had a trivial effect on the $\log 1/D_{37}$ value (low R value). The MLR was used to generate linear models where no descriptor was excluded, and was performed using SPSS [20]. Due to the vast number of molecular descriptors, a stepwise multiple linear regression approach was employed to find the pertinent parameters based on the forward-

selection and the backward-elimination procedures. Statistical outliers were established as those compounds with definite standardized residuals greater than 2. The feature of the developed MLR models was judged using the correlation coefficient (R^2), the standard error of the estimate (s), the adjusted R^2 (R^2_{adj}), and Fisher's criterion (F).

2.4. Electrostatic potential (ESP) analysis and quantum theory of atoms in molecules (QTAIM) Calculations

The versatility of 1-bromo-2-methylpropane (**1Br**), 2-bromo-2-methylpropane (**2Br**), and 2-chloro-2-methylpropane (**2Cl**) to attractively interact with α-glucan (**G**) was thoroughly assessed and comparatively investigated. The geometry optimization calculations were first executed for the **1Br**/**2Br**/**2Cl**/**G** monomers and the **1Br**···/**2Br**···/**2Cl**···**G** complexes at the M06-2X/6-31+G* level of theory [21]. An electrostatic potential (ESP) analysis was carried out for the optimized systems to extract the molecular electrostatic potential (MEP) maps. The electron density envelope with a value of 0.002 au was employed to generate the MEP maps, as previously reported in the literature [22].

Relying on the optimized complexes, the interaction (E_{int}) energies were computed. The counterpoise correction method was invoked to eliminate the basis set superposition error (BSSE) [23], which is illustrated as follows:

$$E_{\text{int}} = E_{\mathbf{1Br/2Br/2Cl\cdots G}} - \left(E_{\mathbf{1Br/2Br/2Cl}\ \text{in complex}} + E_{\mathbf{G}\ \text{in complex}}\right) + E_{\text{BSSE}} \quad (2)$$

where $E_{\mathbf{1Br/2Br/2Cl\cdots G}}$, $E_{\mathbf{1Br/2Br/2Cl}\ \text{in complex}}$, and $E_{\mathbf{G}\ \text{in complex}}$ refer to the energies of the complex, **1Br/2Br/2Cl,** and the **G** structures pertinent to their coordinates in the optimized complexes.

Moreover, the quantum theory of atoms in molecules (QTAIM) calculations were carried out to deeply recognize the studied interactions by extracting the bond paths (BPs) and the bond critical points (BCPs). All executed DFT computations were performed using the Gaussian 09 package [24]. The QTAIM analysis was conducted by utilizing the Multiwfn 3.7 [25] and visualized by Visual Molecular Dynamics (VMD) [26] packages.

3. Results

3.1. QSAR for the Lethality of Halogenated Aliphatic Hydrocarbons to A. nidulans

In order to select the molecular descriptors and to build the QSAR models, the PCA and MLR methods were applied. The correlation coefficients (R) between the observed *Aspergillus* (*A*.) *nidulans* toxicities ($\log 1/D_{37}$) of 41 halogenated aliphatic hydrocarbons and the molecular descriptors are gathered in Table 1. The PCA, including the correlation between the *A*. *nidulans* toxicities and the molecular descriptors, is illustrated in the biplot depicted in Figure 1. The data clusters were depicted by exhibiting the scores for the first components (F1) versus the second ones (F2).

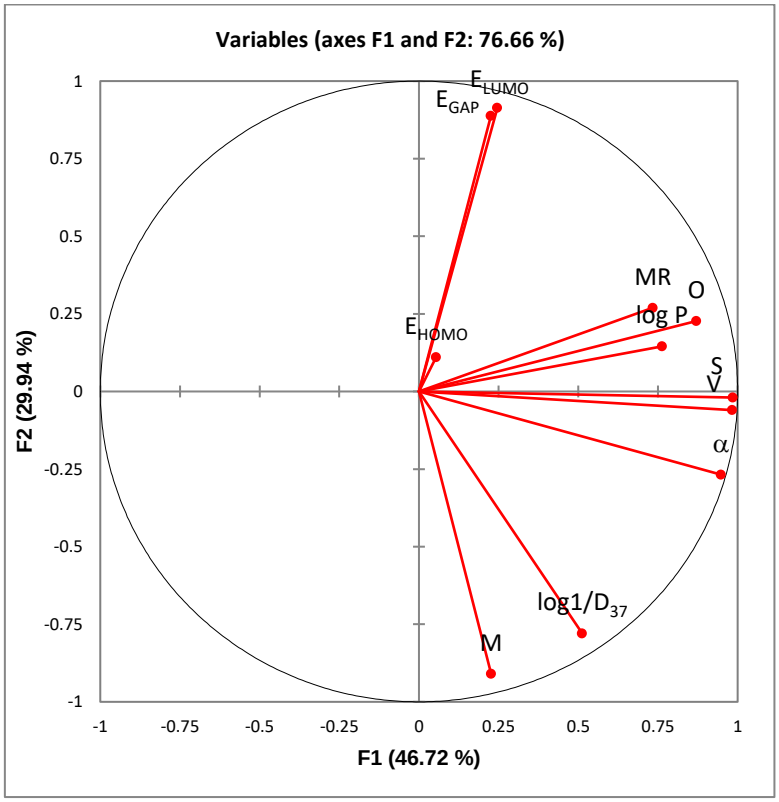

Figure 1. Correlation circle of log1/D_{37} and the computed descriptors.

According to the data displayed in Figure 1, the molecular weight (MW) and the *A. nidulans* toxicities (log1/D_{37}) were somewhat near to each other and were highly related (R = 0.756). This significant correlation coefficient demonstrated that the MW had a prominent influence on the *A. nidulans* toxicities. Furthermore, the surface area grid (S) and molecular volume (V) were superimposed on each other, thereby outlining the prominent positive correlation between these molecular descriptors with R = 0.997. Additionally, the biplot revealed that the S, V, log P, α, and O had positive correlations with log1/D_{37} (0.70 〉 R 〉 0.25).

Table 1. Correlation coefficients matrix of log 1/D_{37} and the computed parameters.

Variables	log1/D_{37}	E_{HOMO}	E_{LUMO}	E_{gap}	S	V	log P	MR	MW	α	O
log 1/D_{37}	1										
E_{HOMO}	0.019	1									
E_{LUMO}	−0.529	0.338	1								
E_{gap}	−0.561	−0.206	0.851	1							
S	0.536	0.090	0.256	0.216	1						
V	0.563	0.079	0.223	0.187	0.997	1					
log P	0.247	−0.202	0.183	0.303	0.672	0.667	1				
MR	0.027	0.107	0.372	0.327	0.650	0.654	0.621	1			
MW	0.756	−0.138	−0.753	−0.705	0.250	0.297	−0.041	−0.008	1		
α	0.694	0.013	−0.001	−0.008	0.934	0.951	0.666	0.650	0.468	1	
O	0.297	0.122	0.422	0.371	0.890	0.852	0.612	0.555	− 0.049	0.715	1

However, E_{LUMO} and E_{gap} had negative correlations with log1/D_{37} and were considered as important descriptors. Notwithstanding the importance of E_{HOMO} and *MR* in various QSAR studies [27,28], the E_{HOMO} and *MR* values were herein neglected due to their trivial effect on the log1/D_{37} value, which was attributed to their low correlation coefficients ($R \langle 0.03$). After excluding E_{HOMO} and *MR*, the preferred regression model was constructed using a backward method based on three descriptors (See Eq 3). Particularly, considerable correlations were found between the MW and both E_{LUMO} and E_{gap}, with *R* values of −0.753 and −0.705, respectively.

$$\log 1/\mathrm{D}_{37} = -3.630 - 9.175E_{\mathrm{LUMO}} - 10.052\,E_{\mathrm{gap}} + 0.016S \quad (3)$$

$$n = 41,\ R^2 = 0.806,\ R^2_{adj} = 0.790,\ s = 0.332,\ F = 51.261$$

where *n* represents the number of halogenated aliphatic hydrocarbon molecules.

Toward investigating the statistical significance of the selected descriptors in the considered model, the *t*-test values were calculated for each descriptor. Furthermore, the *t*-test is defined as the regression coefficient for the relevant descriptor divided by its error. The highest *t*-test value identifies the most important descriptors, where a molecular descriptor is considered significant when its t-test value exceeds $|2|$.

In model 1 (See Eq 3), the *t*-test values were –2.697, –2.871, and 9.589 for E_{LUMO}, $E_{gap,}$ and *S*, respectively. The obtained *t*-test values showed that all three descriptors played an important role in the regression models, with the surface area grid being the most relevant. It is worth mentioning that the positive values of the regression coefficient outlined that the molecular descriptors in the QSTR model provided a positive behavior towards the toxicity values. Meanwhile, negative values suggested that the log1/D_{37} values diminished by augmenting the value of the computed descriptor. From Eq 3, the log1/D_{37} value was disclosed to decline when the E_{LUMO} and E_{gap} values increased. Notwithstanding, the enhanced *S* value leads, in turn, to an increase in the value of log1/D_{37}. These results were consistent with the results obtained using the PCA. The statistical results of Eq 3 for the validation set (14 molecules) were poor and were characterized by ($R^2 = 0.443$, $s = 0.2449$, $F = 9.5267$).

As mentioned earlier, the E_{HOMO} and *MR* were found to be essential descriptors for the prediction of various biological activities. Consequently, the two aforesaid descriptors were considered, which gave rise to using all the calculated descriptors in the construction of the adopted QSAR models, bearing in mind that these two insignificant descriptors (E_{HOMO} and *MR*) could account for the residual. Consequently, the preferred model was obtained based on three descriptors, which are described as follows:

$$\log 1/\mathrm{D}_{37} = -1.242 - 9.769E_{\mathrm{gap}} - 0.047MR + 0.329\alpha, \quad (4)$$

$$n = 41,\ R^2 = 0.912,\ R^2_{adj} = 0.904,\ \mathrm{s} = 0.224,\ F = 127.106.$$

The *t*-test values for E_{gap}, *MR,* and α were –7.080, –7.157, and 15.137, respectively, thus indicating the significant contributions of the three considered descriptors to the log1/D_{37} value, with a highly appreciated role for the molecular polarizability descriptor. According to Eq 4, the log1/D_{37} decreases along with increasing the E_{gap} and *MR*. On the other hand, increasing the values of the molecular polarizability causes a boost in the value of the log1/D_{37}. These results were in good agreement with those obtained using the PCA. As an essential issue, E_{gap} plays decisive roles in the

chemical reactivity, biological activity, hydrophobicity, and electrophilicity of chemicals pertinent to the living cell activity and the associated mechanistic interactions. In that spirit, the E_{gap} is considered one of the most important parameters that can inversely affect the chemical reactivity.

The physical justification for using molar refractivity and polarizability in the developed QSAR models might be understood by considering that most of the biochemical process occurs in aqueous media. Therefore, the *MR* and α of aqueous toxic solutions are informative descriptors, which are useful in QSTR studies. As judged by the four statistical criteria (R^2, R^2_{adj}, s, and F values), Eq 4 is statistically superior to Eq 3. Table 2 compares the observed $\log 1/D_{37}$ values for 41 molecules to the predictions generated by Eq 4. Furthermore, Figure 2 visually depicts the relationship between these expectations and the experimental results of $\log 1/D_{37}$.

Table 2. Observed and predicted values of $\log 1/D_{37}$ for the training set. Predictions were made using Eq 4.

Molecule name	$\log 1/D_{37}$ (Observed)	$\log 1/D_{37}$ (Predicted)	MAE*	Molecule name	$\log 1/D_{37}$ (Observed)	$\log 1/D_{37}$ (Predicted)	MAE*
Dichloromethane	–1.97	–1.96	0.17	1,3-Dichloropropene	–0.72	–0.83	0.16
Chloroform	–1.39	–1.36	0.18	1,1,3-Trichloropropene	–0.38	–0.43	0.16
Tetrachloromethane	–0.49	–0.82	0.18	3-Chloro-2-chloromethylpropene	–0.43	–0.49	0.17
1,2-Dichloroethane	–1.71	–1.64	0.18	1-Chloro-2-methylpropene	–1.10	–0.85	0.17
1,1,2-Trichloroethane	–1.03	–1.12	0.18	Chlorodibromofluoromethane	–0.49	0.07	0.17
1,1,1,2-Tetrachloroethane	–0.45	–0.56	0.18	Bromoform	–0.72	–0.58	0.14
Pentachloroethane	–0.23	–0.11	0.18	Bromochloromethane	–1.79	–1.60	0.14
Hexachloroethane	0.10	0.27	0.18	Bromotrichloromethane	0.10	–0.05	0.14
1,1,2-Trichloroethylene	–1.05	–0.68	0.19	Bromodichloromethane	–1.03	–1.02	0.14
Tetrachloroethylene	–0.08	–0.23	0.18	Chlorodibromomethane	–0.46	–0.78	0.15
1,1-Dichloroethylene	–1.40	–1.23	0.18	1-Bromo-2-chloroethane	–1.38	–1.26	0.13
1,2-Dichloropropane	–1.19	–1.21	0.18	1-Bromobutane	–0.75	–0.95	0.14
1,3-Dichloropropane	–1.02	–1.13	0.19	2-Bromo-1-chloropropane	–0.99	–0.84	0.13
1,2,3-Trichloropropane	–0.97	–0.73	0.19	1-Bromo-2-methylpropane	–1.26	–0.93	0.12
2-Chlorobutane	–1.28	–1.38	0.19	1-Bromo-4-chlorobutane	–0.52	–0.55	0.09
1,3-Dichlorobutane	–0.94	–0.76	0.19	1-Bromo-3-methylbutane	–0.49	–0.54	0.10
1-Chloro-2-methylpropane	–1.16	–1.32	0.19	Dibromodichloromethane	0.70	0.61	0.12
1-Chloropentane	–0.70	–0.95	0.19	Dibromomethane	–1.33	–1.39	0.13
1-Chlorohexane	–0.18	–0.56	0.19	Tetrabromomethane	2.00	1.79	0.16
1-Chlorooctane	–0.18	0.22	0.18	1,1,2,2-Tetrabromoethane	0.40	0.29	0.11
2,3-Dichloropropene	–0.43	–0.82	0.17				

* MAE stands for mean absolute error.

For a comparison purpose between the developed model (See Eq 4) and those constructed by Cronin et al. [10], the QSAR model was built for the whole data set, and the model was characterized

by R^2 = 0.897, s = 0.219, and F =148.766. The currently developed model was superior to those developed by Cronin et al., who reported R^2 = 0.615, s = 0.413, and F = 44.2 (Table S1).

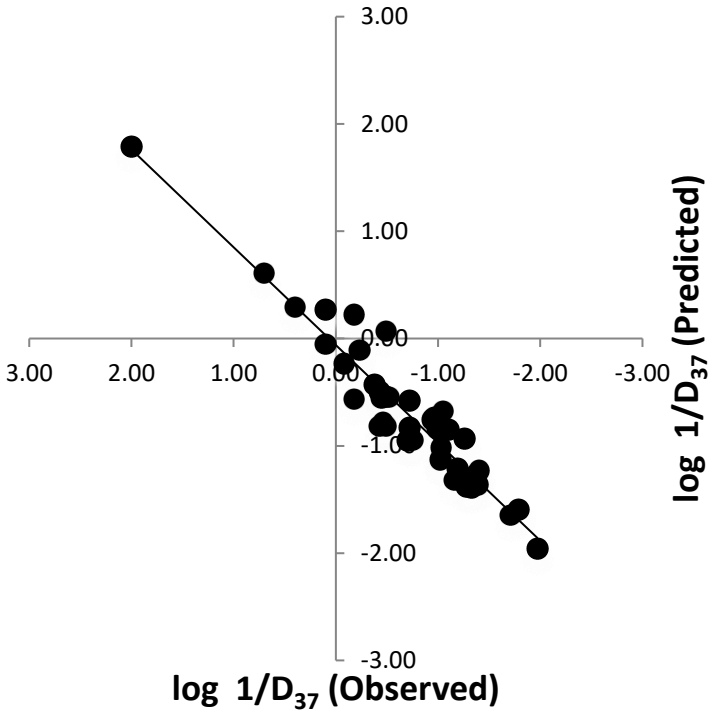

Figure 2. Observed values of log1/D_{37} for 41 molecules vis predictions made by Eq 4.

3.2. Validation set

The regression coefficients for the 41 molecules were adopted to compute the log1/D_{37} for 14 halogenated aliphatic molecules to assess the predictive performance of the three-variable QSTR model (See Eq 4). Table 3 contains the gathered data, which is graphically displayed in Figure 3.

According to Table 3, the results were labeled with an R^2 value of 0.575 and a standard error of 0.214. Therefore, 2-bromo-2-methylpropane (**2Br**) had an outlier behavior in the verification set (See Figure 3) with a standard residual of –2.167. The mathematical performance of the model was marginally improved by excluding this particular compound from the predictive equation (R^2 = 0.737, s = 0.168). Still, of course, there were no posterior justifications for reselections of the data set.

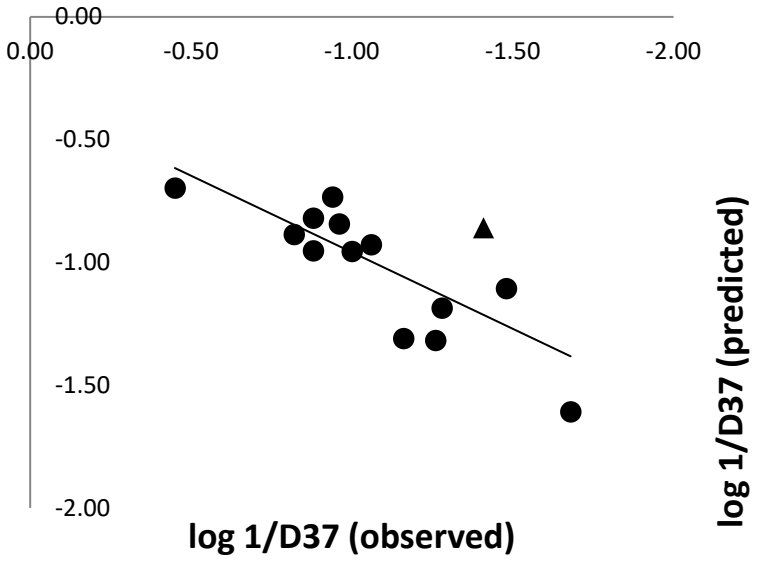

Figure 3. Observed values of log1/D_{37} for validation set versus predictions made by Eq 4. The outlier is represented as a triangle.

Table 3. Observed and predicted values of log1/D_{37} for the validation set. Predictions were made using Eq 4.

Molecule name	log1/D_{37} (Observed)	log1/D_{37} (Predicted)	MAE*
1,1-Dichloroethane	–1.68	–1.61	0.16
1,1,1-Trichloroethane	–1.00	–0.96	0.17
1,1,2,2-Tetrachloroethane	–0.45	–0.70	0.18
1,2-Dichloroethylene	–1.48	–1.11	0.17
2,2-Dichloropropane	–1.28	–1.19	0.15
1-Chlorobutane	–1.16	–1.31	0.16
2,3-Dichlorobutane	–0.94	–0.73	0.16
2-Chloro-2-methylpropane	–1.26	–1.32	0.15
1,1-Dichloropropene	–0.82	–0.89	0.17
3-Chloro-2-methylpropene	–0.88	–0.95	0.19
2-Bromobutane	–1.06	–0.93	0.21
1-Bromo-3-chloropropane	–0.88	–0.82	0.24
2-Bromo-2-methylpropane	–1.41	–0.86	0.33
1,2-Dibromoethylene	–0.96	–0.84	0.12

* MAE stands for mean absolute error.

In accordance with the strong statistical significance, the presented QSTR models had a substantial reliability and an internal extrapolative capacity. As illustrated in Figure 3, the observed log1/D_{37} values for the verification set were consistent with the projected values.

3.3. Further investigations on 2-Bromo-2-Methyl Propane

Furthermore, additional calculations were performed on 2-bromo-2-methyl propane (**2Br**) in order to clarify the potential causes for its outlier behavior in the QSTR models. Due to the structural similarities of 1-bromo-2-methyl propane (**1Br**) and 2-chloro-2-methyl propane (**2Cl**) to **2Br,** the same procedure was performed on **2Cl** and **1Br**. Several studies have claimed that the outlier molecules can emerge for a variety of reasons [29–33]. Outliers, for instance, may be caused by inaccurate molecular descriptor values, or they may suggest an error in the experimental toxicities. They could either be the result of an uncommon mechanism or a distinct binding type. First, a deep inspection of the values of the E_{gap} energy was performed to examine the effect of the stability of the HAHs on the value of log $1/D_{37}$. A stable molecule is indicated by an alkyl halide with a high value of E_{gap}, while a compound with a low E_{gap} value has a high effectiveness. Remarkable correlations (R = –0.561) were observed between the stability of the alkyl halide and its toxicity. More stable molecules turned out to have a high log $1/D_{37}$ value. These results were previously outlined using a PCA (See Figure 1). The E_{gap} for **2Br**, **1Br**, and **2Cl** were 0.187, 0.195, and 0.225 au, respectively. These values indicated that **2Br** had the lowest energy gap and, therefore, demonstrated the highest efficacy (i.e., lowest stability). On the other hand, **2Cl** turned out to be the most stable one. The log $1/D_{37}$ values of **2Br**, **1Br**, and **2Cl** were −1.41, −1.26, and −1.26, respectively. The identical results of log$1/D_{37}$ for the latter two molecules may suggest errors in the experimental toxicities' values.

3.3.1. Interactions of halide alkyls with α-glucan

The structural compositions of *A. nidulans* cell walls have been studied and reported in various studies [34–38]. The cell wall is necessary for fungi to survive in their natural habitat [34]. Fungi cell walls account for approximately one-quarter of the total fungal biomass [35], and approximately one-third of the fungal genome is engaged in cell wall formation and maintenance [36]. A carbohydrate investigation of the *A. nidulans* wall concluded that it contains roughly 40% α-glucan and *β*-glucan [34,37,38]. In this study, *α*-glucan was utilized to explore the binding interactions between the *A. nidulans* cell wall and the selected halide alkyls.

3.3.2. Electrostatic potential (ESP)analysis

An electrostatic potential (ESP) analysis was performed to unveil the nucleophilic and electrophilic regions over the molecular systems. By employing the ESP analysis, the molecular electrostatic potential (MEP) maps of the optimized monomers were extracted utilizing a 0.002 au electron density envelope. Figure 4 displays the MEP maps for the optimized geometries of **1Br**, **2Br**, **2Cl**, and **G**.

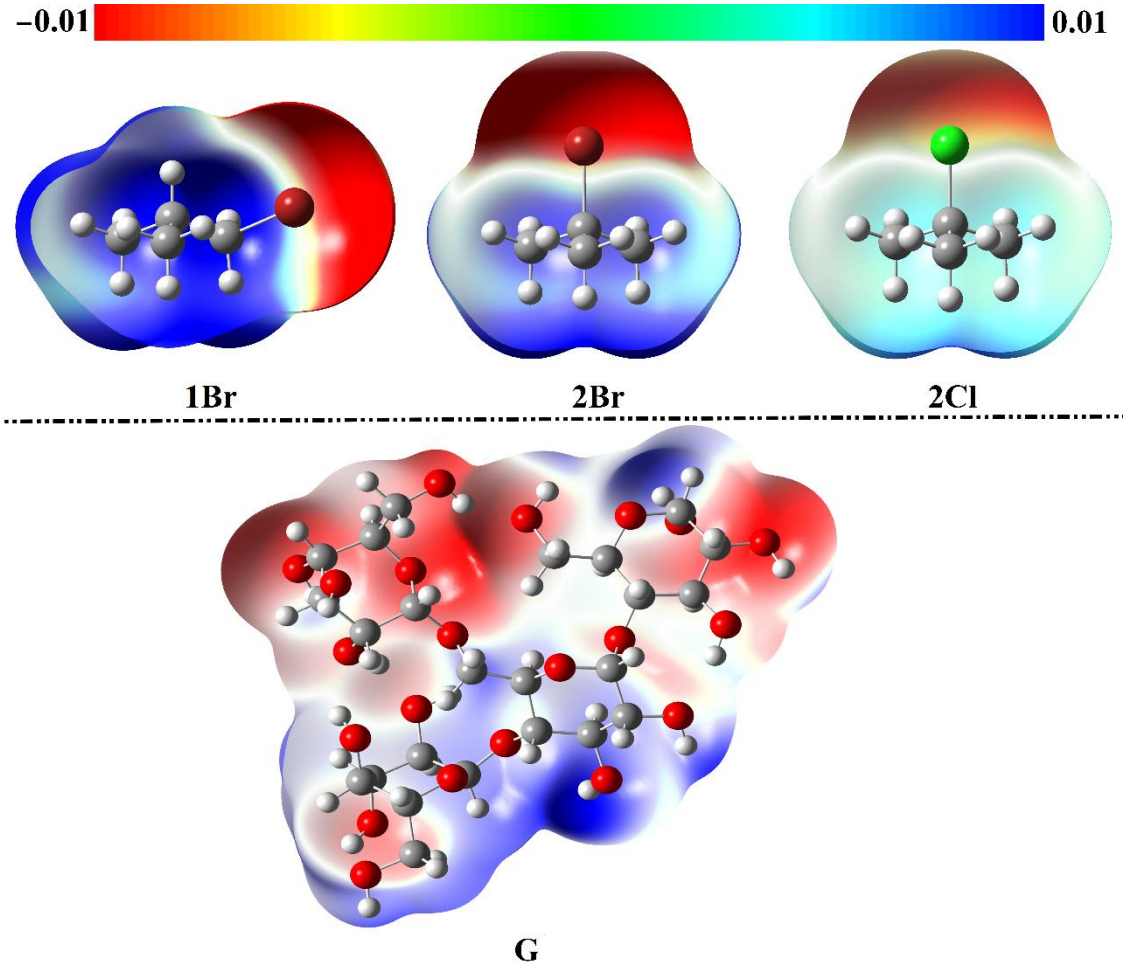

Figure 4. MEP maps of 1-bromo-2-methylpropane (**1Br**), 2-bromo-2-methylpropane (**2Br**), 2-chloro-2-methylpropane (**2Cl**), and α-glucan (**G**). The color scale is extended from the red (–0.01 au) to the blue (0.01 au) scope.

As depicted in Figure 4, the molecular surfaces with relative electrophilic and nucleophilic regions over the molecular surface of the studied systems were conspicuously observed by the presence of blue-coded surfaces (i.e., positive ESP) and red-coded surfaces (i.e., negative ESP), respectively. Notably, negative ESP regions were observed over the surfaces of the Br, Cl, and O atoms of the investigated molecules. On the other hand, the positive ESP areas were found around the surfaces of the H atoms. The obtained findings paraded the supreme penchant of the **1Br**, **2Br**, and **2Cl** molecules to attractively interact with the **G** molecule through hydrogen bonding interactions.

3.3.3. Energetic manifestations

The preferability of the **1Br**, **2Br**, and **2Cl** molecules to interact with the **G** molecule was herein thoroughly addressed through the **1Br**···/**2Br**···/**2Cl**···**G** complexes. First, geometry optimization calculations were performed. Using optimized complexes, the interaction energy (E_{int}) values were computed. Figure S1 shows the optimized geometries of the studied complexes within all the plausible configurations and their E_{int} values. Relying on the negative E_{int} values, the most preferable

configurations of the **1Br**···/**2Br**···/**2Cl**···**G** complexes were selected and gathered in Figure 5.

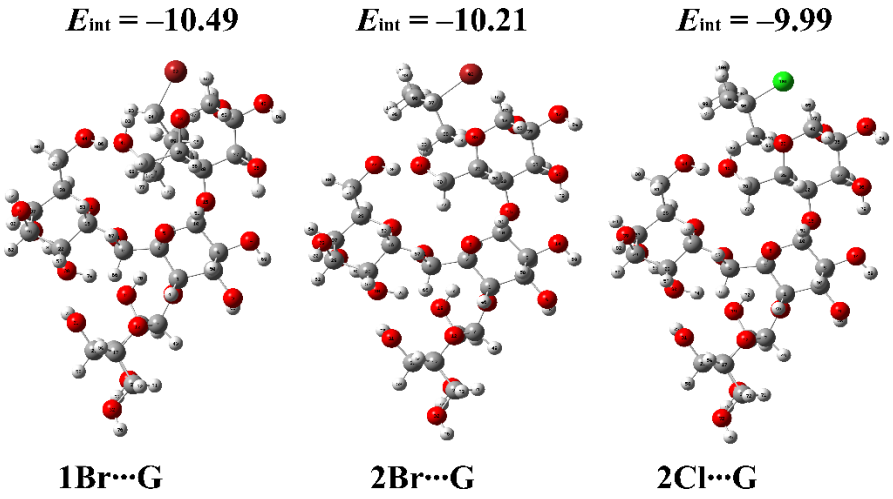

Figure 5. Optimized complexes within the most preferable configurations of the **1Br**···, **2Br**···, and **2Cl**···**G** complexes. Interaction energy (E_{int}) values are in kcal/mol.

As shown in Figure S1, the **1Br**, **2Br**, and **2Cl** molecules showed a superior potentiality to interact with the **G** molecule through the depicted configurations with respectable negative interaction energy values. Notably, the stability of the scouted complexes was generally explained as an upshot to the presence of numerous hydrogen bonds, which were found to vary in numbers and distances. It is worth noting that the preferable interactions within the studied complexes were ascribed to the **1Br**···**G** complex, followed by **2Br**···**G** and **2Cl**···**G** complexes (Figure 5). Numerically, E_{int} of the **1Br**···, **2Br**···, and **2Cl**···**G** complexes were –10.49, –10.21, and –9.99 kcal/mol, respectively.

3.3.4. Quantum theory of atoms in molecules (QTAIM) analysis

To gain further insight into the origin of the studied interactions, a quantum theory of atoms in molecules (QTAIM) analysis was executed for the optimized **1Br**···/**2Br**···/**2Cl**···**G** complexes within the most preferable configurations. In that spirit, the BPs and BCPs were generated and are illustrated in Figure 6.

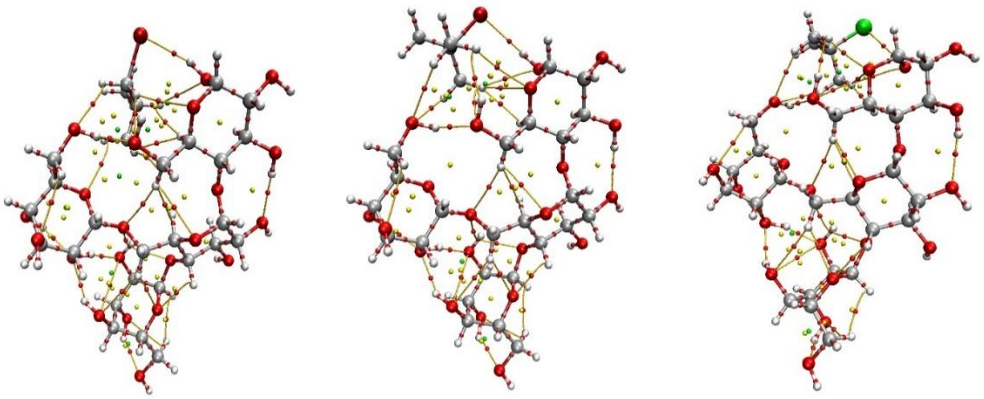

Figure 6. QTAIM plots of the most preferable configurations of the optimized **1Br**···, **2Br**···, and **2Cl**···**G** complexes.

As shown in Figure 6, the occurrence of interactions within the investigated complexes was assured by the presence of BPs and BCPs (i.e., hydrogen bonds). The higher number of the BPs and BCPs within the **1Br**···**G** complex over the other analogs could interpret its higher preferability in comparison to the others. This observation was in line with the interaction energy affirmations.

4. Conclusions

Aspergillus (*A*.) *nidulans* toxicities (log1/D_{37}) were precisely predicted for a set of 55 halogenated aliphatic hydrocarbons (HAHs) using DFT calculations. Besides, the individual correlations between a series of evaluated descriptors and the *A*. *nidulans* toxicities (log1/D_{37}) were unveiled. The importance of quantum-chemical parameters along with various physical properties, including the surface area grid (*S*), molecular volume (*V*), molar refractivity (*MR*), molecular weight (MW), and molecular ovality (*O*) were investigated to predict the *A*. *nidulans* toxicities of halogenated aliphatic hydrocarbons. A potential three-variable QSTR model was trained using 41 molecules and subsequently validated for a set of 14 compounds. Based on the E_{gap}, *MR,* and α, the linear model had correlation coefficient (R^2) values of 0.912 and 0.575 for the training and the validation sets, respectively. The generated QSAR model could be used for the log1/D_{37} predictions of HAHs provided in this paper; however, more research would be required to test more aliphatic compounds for an external validation. The most preferred interactions were clearly identified within the 1-bromo-2-methylpropane···α-glucan complex. The greater number of bond pathways and bond critical sites within the 1-bromo-2-methylpropane···α-glucan complex over other analogs affirmed its significant favorability.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

All authors have no conflict of interest to report.

References

1. Premkumar R, Hussain S, Koyambo-Konzapa SJ, et al. (2021) SERS and DFT studies of 2-(trichloroacetyl)pyrrole chemisorbed on the surface of silver and gold coated thin films: In perspective of biosensor applications. *J Mol Recognit* 34: e2921. https://doi.org/10.1002/jmr.2921
2. MacFarland HN (1986) Toxicology of solvents. *Am Ind Hyg Assoc J* 47: 704–707. https://doi.org/10.1080/15298668691390511
3. Crebelli R, Andreoli C, Carere A, et al. (1995) Toxicology of halogenated aliphatic hydrocarbons: structural and molecular determinants for the disturbance of chromosome segregation and the induction of lipid peroxidation. *Chem-Biol Interact* 98: 113–129. https://doi.org/10.1016/0009-2797(95)03639-3
4. Trohalaki S, Gifford E, Pachter R (2000) Improved QSARs for predictive toxicology of halogenated hydrocarbons. *Comput Chem* 24: 421–427. https://doi.org/10.1016/s0097-8485(99)00093-5
5. Gadaleta D, Vukovic K, Toma C, et al. (2019) SAR and QSAR modeling of a large collection of LD(50) rat acute oral toxicity data. *J Cheminformatics* 11: 58. https://doi.org/10.1186/s13321-019-0383-2
6. Li F, Wang P, Fan T, et al. (2024) Prioritization of the ecotoxicological hazard of PAHs towards aquatic species spanning three trophic levels using 2D-QSTR, read-across and machine learning-driven modelling approaches. *J Hazard Mater* 465: 133410. https://doi.org/10.1016/j.jhazmat.2023.133410
7. Wang QQ, Fan TJ, Jia RQ, et al. (2024) First report on the QSAR modelling and multistep virtual screening of the inhibitors of nonstructural protein Nsp14 of SARS-CoV-2: Reducing unnecessary chemical synthesis and experimental tests. *Arab J Chem* 17: 105614. https://doi.org/10.1016/j.arabjc.2024.105614
8. Oubahmane M, Hdoufane I, Delaite C, et al. (2023) Design of Potent Inhibitors Targeting the Main Protease of SARS-CoV-2 Using QSAR Modeling, Molecular Docking, and Molecular Dynamics Simulations. Pharmaceuticals.
9. Dearden JC, Cronin MT, Dobbs AJ (1995) Quantitative structure-activity relationships as a tool to assess the comparative toxicity of organic chemicals. *Chemosphere* 31: 2521–2528. https://doi.org/10.1016/0045-6535(95)00121-n
10. Cronin MT, Dearden JC, Duffy JC, et al. (2002) The importance of hydrophobicity and electrophilicity descriptors in mechanistically-based QSARs for toxicological endpoints. *SAR QSAR Environ Res* 13: 167–176. https://doi.org/10.1080/10629360290002316
11. Martin TM, Harten P, Young DM, et al. (2012) Does rational selection of training and test sets improve the outcome of QSAR modeling? *J Chem Inf Model* 52: 2570–2578. https://doi.org/10.1021/ci300338w

12. te Velde G, Bickelhaupt FM, Baerends EJ, et al. (2001) Chemistry with ADF. *J Comput Chem* 22: 931–967. https://doi.org/10.1002/jcc.1056
13. Fonseca Guerra C, Snijders JG, te Velde G, et al. (1998) Towards an order- N DFT method. *Theor Chem Acc* 99: 391–403. https://doi.org/10.1007/s002140050353
14. Perdew JP, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. *Phys Rev B* 45: 13244–13249. https://doi.org/10.1103/physrevb.45.13244
15. Perdew JP, Chevary JA, Vosko SH, et al. (1992) Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys Rev B* 46: 6671–6687. https://doi.org/10.1103/physrevb.46.6671
16. Bodor N, Gabanyi Z, Wong CK (1989) A New Method for the Estimation of Partition-Coefficient. *J Am Chem Soc* 111: 3783–3786. https://doi.org/10.1021/ja00193a003
17. Elmi Z, Faez K, Goodarzi M, et al. (2009) Feature selection method based on fuzzy entropy for regression in QSAR studies. *Mol Phys* 107: 1787–1798. https://doi.org/10.1080/00268970903078559
18. Goudarzi N, Goodarzi M, Chen T (2012) QSAR prediction of HIV inhibition activity of styrylquinoline derivatives by genetic algorithm coupled with multiple linear regressions. *Med Chem Res* 21: 437–443. https://doi.org/10.1007/s00044-010-9542-8
19. Cai Z, Zafferani M, Akande OM, et al. (2022) Quantitative Structure-Activity Relationship (QSAR) Study Predicts Small-Molecule Binding to RNA Structure. *J Med Chem* 65: 7262–7277. https://doi.org/10.1021/acs.jmedchem.2c00254
20. Nie NH, Bent DH, Hull CH (1970) SPSS: Statistical Package for the Social Sciences: McGraw-Hill.
21. Zhao Y, Truhlar DG (2008) Exploring the limit of accuracy of the global hybrid meta density functional for main-group thermochemistry, kinetics, and noncovalent interactions. *J Chem Theory Comput* 4: 1849–1868. https://doi.org/10.1021/ct800246v
22. Ibrahim MAA (2012) Molecular mechanical perspective on halogen bonding. *J Mol Model* 18: 4625–4638. https://doi.org/10.1007/s00894-012-1454-8
23. Boys SF, Bernardi F (2006) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol Phys* 19: 553–566. https://doi.org/10.1080/00268977000101561
24. Frisch MJ, Trucks GW, Schlegel HB, et al. (2009) Gaussian 09. Revision E01 ed. Wallingford CT, USA.: Gaussian09, Gaussian Inc.
25. Lu T, Chen F (2012) Multiwfn: a multifunctional wavefunction analyzer. *J Comput Chem* 33: 580–592. https://doi.org/10.1002/jcc.22885
26. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14: 33–38. https://doi.org/10.1016/0263-7855(96)00018-5
27. Al-Fahemi JH (2012) The use of quantum-chemical descriptors for predicting the photoinduced toxicity of PAHs. *J Mol Model* 18: 4121–4129. https://doi.org/10.1007/s00894-012-1417-0
28. Al-Fahemi JH (2013) Structural descriptors for the correlation of human blood: air partition coefficient of volatile organic molecules by QSPRs. *Struct Chem* 24: 2155–2161. https://doi.org/10.1007/s11224-013-0224-2

29. Kim KH (2021) Outliers in SAR and QSAR: 3. Importance of considering the role of water molecules in protein-ligand interactions and quantitative structure-activity relationship studies. *J Comput-Aid Mol Des* 35: 371-396. https://doi.org/10.1007/s10822-021-00377-7
30. Zhao L, Wang W, Sedykh A, et al. (2017) Experimental Errors in QSAR Modeling Sets: What We Can Do and What We Cannot Do. *ACS Omega* 2: 2805–2812. https://doi.org/10.1021/acsomega.7b00274
31. Kim KH (2007) Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers? *J Comput-Aid Mol Des* 21: 421–435. https://doi.org/10.1007/s10822-007-9126-y
32. Kim KH (2007) Outliers in SAR and QSAR: is unusual binding mode a possible source of outliers? *J Comput-Aid Mol Des* 21: 63–86. https://doi.org/10.1007/s10822-007-9106-2
33. Kim KH (2022) Outliers in SAR and QSAR: 4. effects of allosteric protein-ligand interactions on the classical quantitative structure-activity relationships. *Mol Divers* 26: 3057–3092. https://doi.org/10.1007/s11030-021-10365-6
34. Alam MK, van Straaten KE, Sanders DA, et al. (2014) Aspergillus nidulans cell wall composition and function change in response to hosting several Aspergillus fumigatus UDP-galactopyranose mutase activity mutants. *PLoS One* 9: e85735. https://doi.org/10.1371/journal.pone.0085735
35. Gastebois A, Clavaud C, Aimanianda V, et al. (2009) Aspergillus fumigatus: cell wall polysaccharides, their biosynthesis and organization. *Future Microbiol* 4: 583–595. https://doi.org/10.2217/fmb.09.29
36. de Groot PW, Brandt BW, Horiuchi H, et al. (2009) Comprehensive genomic analysis of cell wall genes in Aspergillus nidulans. *Fungal Genet Biol* 46 Suppl 1: S72–81. https://doi.org/10.1016/j.fgb.2008.07.022
37. Guest GM, Momany M (2000) Analysis of cell wall sugars in the pathogen Aspergillus fumigatus and the saprophyte Aspergillus nidulans. *Mycologia* 92: 1047–1050. https://doi.org/10.2307/3761470
38. Free SJ (2013) Chapter Two - Fungal Cell Wall Organization and Biosynthesis. In: Friedmann T, Dunlap JC, Goodwin SF, editors. *Advances in Genetics*: Academic Press. pp. 33–82.

AIMS Press