

## Structural requirements of some derivatives based on natural alkaloid lycorine for their dengue inhibitory activity to accelerate dengue drug discovery efforts

Monika Awasthi, Sk. Abdul Amin, Vijaya Shukla, Sanskar Jain, Umesh Kumar Patil and Shovanlal Gayen\*

Laboratory of Drug Design and Discovery, Department of Pharmaceutical Sciences,  
Dr. Harisingh Gour University (A Central University), Sagar 470003, Madhya Pradesh, India

Received 12 May 2016; Revised 08 June 2016

Dengue virus (DENV) is one of the life threatening problems in tropical countries including India. During the last few decades, 50-100 million dengue viral infections were reported specially in Asia and Pacific region. In search of novel DENV inhibitor, the present study tried to correlate the DENV inhibitory data with a series of natural alkaloid lycorine derivatives. A statistically robust QSAR model is generated [ $R = 0.944$ ;  $R_A^2 = 0.865$ ;  $Q^2 = 0.741$ ;  $R_{Pred}^2 = 0.51593$ ;  $cR_p^2 = 0.826$ ] with five descriptors such as *GATS2c*, *ZMIC3*, *T\_N\_O\_7*, *Quadrupole1* and *nHsNH2*. The reliability of the model was confirmed by Golbraikh and Tropsha acceptable model criteria. The QSAR result suggests that the hydroxyl group as well as linear aliphatic at R<sub>1</sub> position is very important for the biological potency, whereas, amino and ether substituent at the R<sub>2</sub> is detrimental. The study also revealed that bulky esterification at the R<sub>2</sub> is not tolerable. Additionally, lead molecule showed good absorption, distribution, metabolism, excretion and toxicity (ADMET) properties. The findings may be useful to design potent lycorine-based anti-dengue compounds in future.

**Keywords:** ADMET, Dengue, Lycorine, MLR, Natural alkaloid, NS4A, QSAR.

**IPC code; Int. cl. (2015.01)**– A61K 36/00

### Introduction

Dengue fever is currently a life threatening problem in tropical zone countries like India<sup>1-4</sup>. The causative agent for the dengue fever is RNA virus which belongs to the family Flaviviridae. This dengue virus (DENV) is transmitted by the mosquito *Aedes aegypti*. Dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) are severe and life threatening disease, especially in children<sup>3</sup>. Therefore, there is an urgent need to discover small molecules as drugs for the treatment. Different non-structural proteins of the virus (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) are excellent drug targets for the disease as they are important for the replication of the virus<sup>5-8</sup>.

Nowadays, drug discovery from herbal or natural source is in much demand. There are approximately 60 % of natural drugs developed as anti-proliferative agents and 25 % of drug as antiviral agents<sup>9,10</sup>. These agents are estimated in the form of various crude

extract, which are extracted from glycosidal, alkaloidal, flavonoides, and carbohydrate source in the form of essential oil or purified compounds. Glycosides, alkaloids, and flavonoids are groups of secondary metabolites while carbohydrate is a primary metabolite. Various flavanoids and alkaloids possess antiviral activities. One of the alkaloids named lycorine that is obtained from plants of Amaryllidaceae family has various biological activities such as antiviral activity against different viral infections like Yellow fever virus, polio virus, vaccinia virus, and herpes virus by suppressing RNA replication of virus and suppressing its translation process<sup>9,10</sup>. It showed antiviral action against dengue and related West Nile virus also. It has been reported that the lycorine derivatives may target dengue non-structural protein NS4A<sup>9</sup>. NS4A protein is one of the promising drug targets, but unfortunately, till date no crystal structure of NS4A is reported. So, the structure-based modeling is restricted in order to find out potent DENV inhibitor. Therefore, ligand-based modeling may serve as an important alternative approach to accelerate the dengue inhibitor designing process.

In this study, quantitative structure activity relationship (QSAR) modeling has been performed on

\*Correspondent author

Tel.: +91-8827680414

E-mail: shovanlal.gayen@gmail.com

Supplementary tables and figures are available online only.

some lycorine based derivatives having anti-dengue activities as a part of our rational drug design<sup>11-16</sup> program to identify the necessary structural features for the anti-dengue activity.

## Materials and Methods

### Dataset and biological activity

The anti-dengue activities of a series of lycorine derivatives (Table 1) were selected from Wang *et al.* in order to develop a chemometric tool<sup>9</sup>. The *in vitro* activity values in logarithmic scale ( $pEC_{50}$ ) are considered as dependent variable<sup>9</sup>.

### Generation and selection of descriptors

In order to sketch 2D structures of these lycorine derivatives, ChemDraw Ultra 8.0 software was used<sup>17</sup>. Then, those structures were converted to their corresponding 3D structures and subsequently, energy minimization process was carried out by batch energy minimization method with RMS gradient (convergence criterion) of 0.01 kcal/mol using a MMFF94s (Merck molecular force field) force field and charge in the VLife MDS platform<sup>18</sup>. Finally, various physico-chemical like electronic, thermodynamic, and spatial; structural descriptors, like retention index (chi), chain path count, path count, atomic valence connectivity index (chiV), element count, path cluster, semi-empirical, estate number, partition coefficient, molecular weight, molecular refractivity, and topological index; and Baumann alignment-independent (AI) descriptors were calculated by using VLife MDS 4.4 software<sup>18</sup>. Additionally, a pool of 2D PaDEL-descriptors (constitutional, topological, electronic, thermodynamic, geometric, and autocorrelation descriptors) using open source "PaDEL-descriptor" software was calculated<sup>19</sup>. These descriptors were taken as an independent tool for QSAR model development. From these large dataset descriptors having constant value (variance < 0.0001) were eliminated initially from our study to reduce redundancy<sup>13</sup>.

### Test set and training set selection

The splitting of dataset is a very important factor as per as the justification of the chemometric model is concerned<sup>11</sup>. The chemical nature of the training set compounds has influence in the predictive ability of a model<sup>14</sup>. A number of methods are there to divide the dataset. Y-based ranking method is one of the frequently used methods which divide dataset based on the biological activity in a non-random way<sup>20</sup>.

In this study, the division of whole dataset was done by Y-based ranking method into two sets i.e., test set and training set. The division was performed in the 1:4 ratio, which means that each of the fourth compound was selected for the test set and the rest were taken in the training set.

### Correlation analysis

Prior to correlation analysis, dataset thinning was performed by using Data Pre-treatment tool<sup>21</sup> keeping variance cut of 0.0001, correlation cut-off 0.70. Then, correlation studies were carried out in STATISTICA 7 software<sup>22</sup> by taking biological activity ( $pEC_{50}$ ) as dependent parameter, whereas descriptors values were considered as independent variable. The pairs of descriptors having intercorrelation more than 0.6 were not considered in order to avoid any error and decrease redundancy. Finally, model was constructed with five variables by maintaining the recommended ratio of the number of predictor parameters to number of data point of 1:5<sup>(Ref. 22)</sup>.

### Validation of the model

After correlation analysis, the MLR models were generated and validated by statistical parameters including correlation coefficient ( $R^2$ ), adjusted square correlation ( $R_A^2$ ), variance ratio ( $F$ ) standard error of estimation ( $SEE$ ), LOO cross-validated  $R^2$  ( $Q^2$ ), standard deviation error of prediction ( $SDEP$ ), and predicted residual sum of square ( $PRESS$ )<sup>11-13,20</sup>.

External validation of test compounds was judged through predicted  $R^2$  ( $R_{pred}^2$ ) and  $r_m^2$  (modified  $r^2$ ) of constructed QSAR model<sup>13</sup>. The external validation parameter  $R_{pred}^2$  was calculated by using the following equation:

$$R_{pred}^2 = 1 - \frac{\sum (y_i - \bar{y}_i)^2}{\sum (y_i - y_{mean})^2}$$

where,  $y_i$  and  $\bar{y}_i$  are the actual and predicted activity of the  $i^{\text{th}}$  molecule in the training set, respectively, and  $y_{mean}$  is the average activity of all molecules in the test set.

To calculate the  $r_m^2$  ( $r_m^2(test)$ ) metrics, the following equation was used:

$$r_m^2(test) = r^2 \times \left( 1 - \sqrt{r^2 - r_0^2} \right)$$

where,  $r^2$  and  $r_0^2$  are the squared correlation coefficients between the actual and predicted values of the test set compounds with and without intercept respectively.

Table 1—Anti-dengue activities ( $pEC_{50}$ ) of lycorine derivatives (1-34)

Compound no	Smile format	Biological activity ( $EC_{50}$ )	Biological activity ( $pEC_{50}$ )
1	<chem>c12cc3c(cc1CN1[C@@H]4[C@@H]2[C@@H]([C@H](C=C4CC1)OC(=O)C)OC(=O)C)OCO3</chem>	25.4	4.6
2	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)O)OC(=O)C)OCO3</chem>	0.4	6.4
3	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H]([C@H](C=C4CC1)OC(=O)CC)OC(=O)C)OCO3</chem>	28.7	4.5
4	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)OC(=O)CCCC)OC(=O)C)OCO3</chem>	120.1	3.9
5	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H]([C@H](C=C4CC1)OC(=O)CCCC)OC(=O)C)OCO3</chem>	150.2	3.8
6	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)OC(=O)CCCC(=O)O)OC(=O)C)OCO3</chem>	21.0	4.7
7	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)OC(=O)CNC(=O)OC(C)(C)C)OC(=O)C)OCO3</chem>	48.2	4.3
8	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H]([C@H](C=C4CC1)OC(=O)C(C)C)OC(=O)C)OCO3</chem>	72.4	4.1
9	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H]([C@H](C=C4CC1)OC(=O)c1cccc1)OC(=O)C)OCO3</chem>	150.3	3.8
10	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H]([C@H](C=C4CC1)OC(=O)c1ccc(cc1)[N](=O)O)OC(=O)C)OCO3</chem>	108.4	4.0
11	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)OC(=O)c1ccc(cc1)Cl)OC(=O)C)OCO3</chem>	90.6	4.0
12	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@@H](C=C4CC1)ON1C(=O)c2c(C1=O)cccc2)OC(=O)C)OCO3</chem>	181.0	3.7
13	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H]([C@@H](C=C4CC1)ON)OC(=O)C)OCO3</chem>	111.9	3.9
14	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)C)OCO3</chem>	1.8	5.7
15 <sup>a</sup>	<chem>O[C@@H]1[C@H]2c3cc4c(cc3CN3[C@@H]2C=C[C@@H]1O)CC3)OCO4</chem>	0.8	6.1
16	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H](C(=O)C=C4CC1)O)OCO3</chem>	0.5	6.3
17	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]2[C@@H](C=C4CC1)O2)OCO3</chem>	40.6	4.4
18	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@@H]([C@H](C=C4CC1)N=[N]=N)O)OCO3</chem>	220.4	3.7
19	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)N)O)OCO3</chem>	198.6	3.7
20	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)OC)O)OCO3</chem>	31.4	4.5
21	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)n1ccc(n1)O)OCO3</chem>	229.8	3.6
22	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@@H](C=C4CC1)n1cc(n1)CCCCO)OCO3</chem>	127.2	3.9
23	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)NC(=O)CCCC)O)OCO3</chem>	100.3	4.0
24	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)OC(=O)C)O)OCO3</chem>	19.6	4.7
25	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@H](C=C4CC1)OC(=O)CCCC)O)OCO3</chem>	28.4	4.5
26	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H]([C@@H](C=C4CC1)O[Si](C(C)(C)C)O)OCO3</chem>	36.0	4.4
27	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)CC)OCO3</chem>	8.2	5.1
28	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)CCC)OCO3</chem>	8.9	5.1
29	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)CCCC)OCO3</chem>	3.6	5.4
30	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)CCCC)OCO3</chem>	5.0	5.3
31	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)C(C)(C)C)OCO3</chem>	30.2	4.5
32	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)c1cccc1)OCO3</chem>	39.6	4.4
33	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)c1ccc(cc1)Cl)OC3</chem>	58.1	4.2
34	<chem>c12cc3c(cc1CN1[C@H]4[C@H]2[C@H](C(=O)C=C4CC1)OC(=O)c1ccc(cc1)OC)OC3</chem>	121.4	3.9

<sup>a</sup>parent molecule - lycorine

The internal cross-validated  $R^2$  ( $Q^2$ ), external predicted  $R^2$  ( $R_{pred}^2$ ) and modified  $r^2$  ( $r_m^2$ ) value should be greater than 0.50 for acceptance of a QSAR model<sup>23,24</sup>. Additionally, Y randomization test was carried out and the obtained  $cR_p^2$  value should be higher than that of 0.50 for acceptability<sup>13</sup>. The robustness of the QSAR equation was judged by performing this Y-randomization test by using following equation:

$$cR_p^2 = R \times \sqrt{R^2 - R_r^2}$$

where,  $R^2$  is the coefficient of the non-random model and  $R_r^2$  is the squared average correlation coefficient of the randomized models<sup>13</sup>.

#### Applicability domain and VIF

Applicability domain is considered as a region in chemical space that is made through model response and model descriptor<sup>24</sup>. All compounds falling in this space were defined to show good prediction and they are considered as reliable. Variation inflation factor is a determining parameter of intercorrelation of all descriptors to build a model<sup>12</sup>. The VIF value is obtained from  $1/1 - r^2$  where,  $r^2$  is known as multiple correlation coefficients i.e., the contribution of one descriptor regressed on the other descriptors. If VIF obtained as greater than that of 10 that means one descriptor may be hidden by other descriptors<sup>13</sup>.

#### ADMET predictions

In order to understand the absorption, distribution, metabolism, excretion, and toxicity (ADMET) profile of one of the lead molecule (compound 2) of this series, we calculated *in silico* ADMET prediction. The ADMET properties were calculated from admetSAR tool<sup>25</sup>. Prior to entry of a molecule in the clinical trials, it is better to know the ADMET profile initially as the experimental evaluations require money and high work load.

#### Results and Discussion

Using the stepping criteria based on F value (F = 4 for inclusion; F = 3.99 for exclusion), the following best equation was derived with five variables.

$$pEC_{50} = 18.43108 (\pm 1.336) - 7.14225 (\pm 0.93281) \\ GATS2c - 0.17168 (\pm 0.02513) ZMIC3 - 1.52564 \\ (\pm 0.18082) T\_N\_O\_7 - 0.018 (\pm 0.00474) \\ Quadrupole1 - 1.30672 (\pm 0.32572) nHsNH2$$

$$n_{Train} = 26; R = 0.944; R^2 = 0.89194; R_A^2 = 0.865; F \\ (5, 20) = 33.017; SEE = 0.296; Q^2 = 0.741; PRESS =$$

$$4.17989, SDEP = 0.40095; Q^2_{L3O} = 0.737; r_{m(L3O)}^2 = \\ 0.7461; \Delta r_{m(L3O)}^2 = 0.15277; n_{Test} = 08; R_{Pred}^2 = 0.51593; \\ r_{m(test)}^2 = 0.54229; \Delta r_{m(Test)}^2 = 0.103; r_{m(Overall)}^2 = \\ 0.69003; \Delta r_{m(Overall)}^2 = 0.09227; cR_p^2 = 0.826.$$

The model (1) explains 86.5 % and predicts 74.1 % of variances of the DENV inhibitory activity and variance ratio ( $F = 33.017$ ) at specified degrees of freedom ( $df = 20$ ). The correlation matrix among descriptors used to develop QSAR model (1) is given in the Supplementary Table 1. This model successfully passed the  $p$ -statistics ( $p < 0.05 = 95$  % significance) as demonstrated in the Supplementary Table 2.

The model is robust enough as guided by the Y-randomization  $cR_p^2$  values more than 0.5 ( $cR_p^2 = 0.826$ ). As a result of the lowest  $p$ -value (Supplementary Table 2) and the acceptable cross-validated  $Q^2$  such as Leave-one-out (LOO) and leave-3-out method (L3O) suggested that an optimum fit was found. The values of all the statistical parameters being within the acceptable limit reflect the internal and external predictive potential of the developed model (1). Although the model (1) has a high cross-validated  $R^2$  ( $Q^2_{(LOO)} = 0.741$ ;  $Q^2_{(L3O)} = 0.737$ ) it may not imply that its predictive ability is high. To ensure the reliable predictive ability of QSAR model, Golbraikh and Tropsha<sup>23</sup> formulated a set of criteria for evaluation. Interestingly the model (1) successfully passed the Golbraikh and Tropsha acceptable model criteria<sup>23</sup> (Table 2).

The values of the distance score, mean distance score, and normalized mean distance score for each compounds in the model (1) is given in the Supplementary Table 3. The test of the applicability

Table 2—Golbraikh and Tropsha acceptable model criteria of model (1)

Parameter	Threshold value	Model (1)
$Q^2$	$Q^2 > 0.5$	0.74144
$r^2$	$r^2 > 0.6$	0.68347
$ r^2_o - r^2_o' $	$ r^2_o - r^2_o'  < 0.3$	0.1403
$k$	$0.85 < k < 1.15$	0.97402
$k'$	$0.85 < k' < 1.15$	1.01757
$(r^2 - r^2_o) / r^2$	$(r^2 - r^2_o) / r^2 < 0.1$	0.00001

where, (i)  $Q^2$  is the cross-validated correlation coefficient, (ii) The parameter  $r^2$  is the squared correlation coefficient between the observed vs predicted response values of the test set compounds, (iii) while  $r^2_o$  and  $r^2_o'$  represent the corresponding values for regression through origin (observed vs. predicted and predicted vs. observed, respectively), and (iv) the slope of the regression lines through the origin are assigned by  $k$  and  $k'$ .

domain by the Euclidean distance approach is graphically represented in the Supplementary Fig 1.

The observed, LOO-predicted, and activities of all compounds of model (1) are given in Supplementary Table 4. The residual plot between predicted and experimental values of train set (light blue box) and test set (red box) compounds for the model (1) is shown in Supplementary Fig 2 and the Observed vs. LOO-predicted activities of training and test set compounds for model (1) was plotted graphically (Fig 1).

The values of all the statistical parameters being within the acceptable limit reflect the statistical potential of the developed model (1).

#### Descriptors contribution in model (1)

The descriptor *GATS2c* belongs to 2D-autocorrelation descriptor which has the highest contribution (Fig 2) towards DENV inhibitory activity

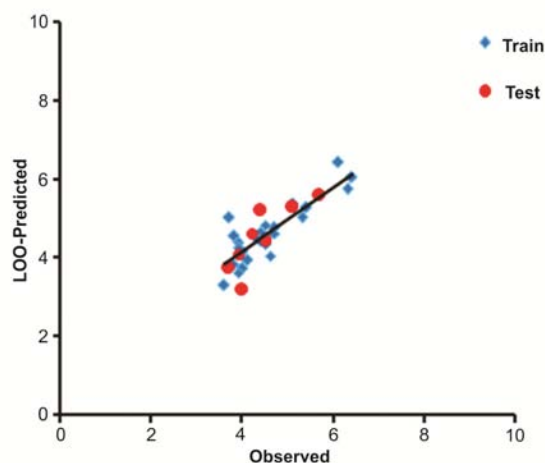


Fig. 1—Graphical representation of Observed (Obs) versus LOO-predicted (Pred) activities of training and test set compounds for model (1)

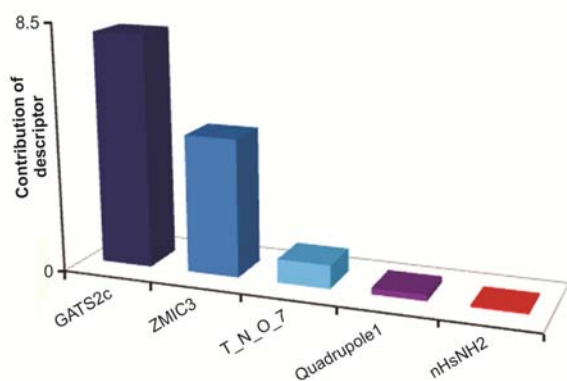


Fig. 2—Contribution of the Descriptors for model (1)

as suggested by the model (1). This distance type function is calculated for each atom of the compound by considering charges separation. Here, the compounds with high *GATS2c* values show poor biological activity as this descriptor contributed negatively towards biological activity data (Table 3). This means that *GATS2c* values should be considered at the time of designing new lycorine derivatives. This result can be explained by showing compounds 32-34 (having aromatic group at the  $R_1$ ) and compounds 27-30 (having linear aliphatic group at the  $R_1$ ) where upon decreasing of *GATS2c* values from aromatic to linear aliphatic substituent at the  $R_1$  influenced the outcome of DENV inhibitory values. Interestingly, this descriptor also explained that hydroxyl substitution at the  $R_2$  (compound 15; *GATS2c* = 1.054) is tolerable, whereas, ether substitution at the  $R_2$  is detrimental (compound 17; *GATS2c* = 1.247). Thus, upon increasing *GATS2c* values lead to loss in potency.

Next descriptor is *ZMIC3*, which also contributed negatively to the DENV inhibitory activity. Higher values of the Z-modified information content index are found in compounds having bulkier ester group at  $R_2$  position (compounds 3-11) and these compounds show low/moderate biological potency. This result can be understood by comparing compounds 7, 12 (bulkier ester group substitution) and compounds 2, 15, 16 (simple hydroxyl group) where the presence of bulkier ester group substitution at  $R_2$  position (bulkier ester group > simple hydroxyl group) would result in lesser biological activities as *ZMIC3* values are increased. Therefore, bulky esterification is not tolerable at  $R_2$  position.

The negative contribution of the descriptor *T\_N\_O\_7* indicates that the presence of nitrogen atoms separated from oxygen atoms by seven bond distance would result in decrease of biological activity. The compound 23 having maximum number of nitrogen atoms separated from oxygen atoms by seven bond distance is poorly active. Interestingly, no nitrogen atom separated from oxygen atoms by seven bond distance is found in case of best active compound 2 ( $EC_{50}$  = 0.4  $\mu$ M) and similar active compounds 15 ( $EC_{50}$  = 0.8  $\mu$ M) and 16 ( $EC_{50}$  = 0.5  $\mu$ M).

The next descriptor *Quadrupole1*, similar to dipole moment defines the charge separation and the interactions with the receptor. Highest *Quadrupole1* value (52.907) is observed in compound 12, which is

Table 3—Definition of descriptors

Descriptor	Type	Definition	Contribution
<i>T<sub>N</sub>O<sub>7</sub></i>	Baumann alignment-independent (AI) descriptors	the count of nitrogen atoms separated from oxygen atoms by seven bond distance	Negative
<i>GATS2c</i>	Autocorrelation Descriptor	Geary autocorrelation - lag 2 / weighted by charges	Negative
<i>ZMIC3</i>	Information Content Descriptor	Z-modified information content index (neighborhood symmetry of 3-order)	Negative
<i>nHsNH2</i>	Electrotopological State Atom Type Descriptor	Count of atom-type H E-State -NH <sub>2</sub>	Negative
<i>Quadrupole1</i>	Dipole moment	Magnitude of first tensor of quadrupole moments.	Negative

Table 4—The ADMET property of compound 2 (calculated by admetSAR tools)

Entry	Parameter	Compound 2
1	Blood-Brain Barrier	+
2	Human Intestinal Absorption	+
3	Caco-2 Permeability	+
4	P-glycoprotein	Inhibitor
5	Renal Organic Cation Transporter	Inhibitor
6	CYP450 2C9 Substrate	Non-substrate
7	CYP450 2D6 Substrate	Non-substrate
8	CYP450 3A4 Substrate	Substrate
9	CYP450 1A2 Inhibitor	Inhibitor
10	CYP450 2C9 Inhibitor	Non- inhibitor
11	CYP450 2D6 Inhibitor	Inhibitor
12	CYP450 2C19 Inhibitor	Non- inhibitor
13	CYP450 3A4 Inhibitor	Non- inhibitor
14	Human Ether-a-go-go-Related Gene Inhibition	Weak inhibitor
15	Carcinogens	Non-carcinogens

one of the least active compounds. Interestingly, this descriptor explained that ketone group at R<sub>2</sub> position (compound 16) is more suitable than hydroxyl (compound 15) substitution at the R<sub>2</sub> and this finding positively support the SAR studies reported earlier {compound 16 (EC<sub>50</sub> = 0.5 μM; *Quadrupole1* = 16.927) > compound 15 (EC<sub>50</sub> = 0.8 μM; *Quadrupole1* = 22.7)}. Therefore, dipole moment should be considered to design novel derivatives.

The descriptor *nHsNH2* has very low contribution in this model (1). This descriptor is negatively correlated with the DENV inhibitory data i.e., increasing values of this parameter may diminish the activity of these series. The maximum number of atom-type H E-State -NH<sub>2</sub> is observed in compound 13 and 19, and they are poorly active. This result is in close agreement with the biological activity where amino derivatives such as compound 13 (EC<sub>50</sub> = 111.9 μM) and 19 (EC<sub>50</sub> = 198.6 μM) possess very poor biological activity.

#### ADMET predictions

An ideal drug should have less toxicity and good pharmacokinetic property. The risk of elimination of a drug like molecule in clinical trials may be reduced by initial assessment of absorption, distribution, metabolism, excretion and toxicity (ADMET) property. In a recent communication, our investigated compound 2 (one of the lead molecules) was taken for the prediction of ADMET profiles by admetSAR tool<sup>25</sup> and results are given in Table 4. Interestingly, compound 2 showed good absorption towards blood-brain barrier and epithelial cells of human intestine. Additionally, compound 2 is a substrate for CYP450 3A4 whereas non-substrate for CYP450 2C9. Under the toxicity category, compound 2 was non-carcinogenic and showed weak inhibitory action against human Ether-a-go-go-related gene.

#### Target implications of lycorine derivatives

Various non structural proteins of DENV are excellent drug targets. Lycorine derivatives target the C-terminal region (designated as 2K fragment) of non structural protein, NS4A and exert their anti-dengue action by suppressing the viral RNA replication<sup>27,28</sup>. Secondary structure prediction using Protter server<sup>28</sup> showed that the protein has four trans-membrane helices (Fig 3a) where the 2K fragment is the last one. It has been reported that valine to methionine substitution in 2K fragment confers resistance to lycorine<sup>27</sup>. Our model of NS4A showed that the lycorine derivatives may possibly act at the membrane surface of the intra-cellular region. Thus, the balances of charge, bulkiness as identified in our modeling study are very important features in order to interact with the target NS4A as we explored some structural features responsible for the dengue inhibitory activity from this QSAR study. The important structural requirements of lycorine derivatives are shown in the Fig 3b.

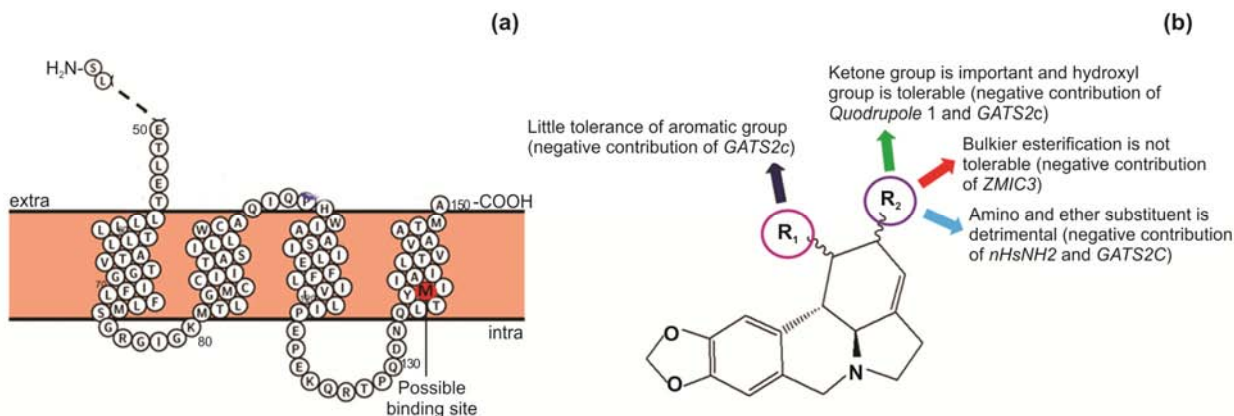


Fig. 3— a) Secondary structure of DENV NS4A (four trans-membrane helices where the 2K fragment is the last one) using Protter server<sup>28</sup> and b) Structural requirements of lycorine derivatives having DENV inhibitory activity

## Conclusion

In the present communication, multiple molecular modelling tools were applied to explore the structural and physico-chemical requirements of lycorine derivatives for their DENV inhibitory activity. The QSAR model was statistically significant and well validated. It may guide to design new lycorine analogues as anti-dengue agents. The QSAR results revealed that the hydroxyl group or linear aliphatic at R<sub>1</sub> position is suitable for the DENV inhibitory activity, whereas bulky esterification, amino and ether substituent at the R<sub>2</sub> is not acceptable. One of the lead molecules (Compound 2) showed acceptable ADMET properties as predicted from the admetSAR tool. Therefore, the present study has a potential to accelerate dengue drug discovery efforts.

## Acknowledgement

S.G. is thankful for the Start up grant from University Grants Commission, New Delhi (No.F.30-106/2015-BSR). The authors (M.A., S.A.A., and V.S.) are grateful to All India Council for Technical Education, New Delhi for awarding Post Graduate GPAT fellowship. The authors also thank the authority of the Department of Pharmaceutical Sciences, Dr. Harisingh Gour University, Sagar, India for providing facility.

## References

- Kim Y M, Gayen S, Kang C, Joy J, Huang Q, Chen A S, *et al.*, NMR Analysis of a novel enzymatically-active unlinked Dengue NS2B-NS3 protease complex, *J Biol Chem*, 2013, **288**, 12891-12900.
- Muhammad M, Kee L Y, Rahman N A and Yusof R, Antiviral action of flavanoid derived compound on dengue virus type-2, *Int J Biol Sci*, 2010, **6**(3), 294-302.
- Sarangi M K and Padhi S, Dengue and its physiotherapy: A review, *Int J Pharm Phytopharmacol Res*, 2014, **4**(1), 37-46.
- Abd Kadir S L, Yakoob H and Mohamed Z, Potential anti dengue medicinal plants: A review, *J Nat Med*, 2013, **67**(4), 677-689.
- Xie X, Gayen S, Kang C, Yuan Z and Shi P Y, Membrane topology and function of dengue virus NS2A protein, *J Virol*, 2013, **87**(8), 4609-4622.
- Li Y, Kim Y M, Zou J, Wang Q Y, Gayen S, Wong Y L, *et al.*, Secondary structure and membrane topology of dengue virus NS4B N-terminal 125 amino acids, *Biochim Biophys Acta*, 2015, **1848**(12), 3150-3157.
- Gayen S, Chen A S, Huang Q and Kang C, West Nile virus (WNV) protease and membrane interactions revealed by NMR spectroscopy, *Biochem. Biophys Res Commun*, 2012, **23**(4), 799-804.
- Kang C, Gayen S, Wang W, Severin R, Chen A S, Lim H A, *et al.*, Exploring the binding of peptidic West Nile virus NS2B-NS3 protease inhibitors by NMR. *Antiviral Res*, 2013, **97**(2), 137-144.
- Wang P, Li L F, Wang Q Y, Shang L Q, Shi P Y and Yin Z, Anti-dengue-virus activity and structure-Activity relationship studies of lycorine derivatives, *Chem Med Chem*, 2014, **9**(7), 1522-1533.
- Teixeira R R, Pereira W L, Oliveira A F, da Silva A M, de oliveira A S, da Silwa C C, *et al.*, Natural products as source of potential dengue antivirals, *Molecules*, 2014, **19**(6), 8151-8176.
- Amin S A, Adhikari N, Jha T and Gayen S, Exploring structural requirements of unconventional Knoevenagel-type indole derivatives as cytotoxic agents through comparative QSAR modeling approaches, *Can J Chem*, 2016, **94**, 637-644.
- Amin S A and Gayen S, Modeling cytotoxic activity of some pyrazolo-triazole hybrids using descriptors calculated from open source tool "PaDEL-descriptor", *J Taibah Univ Sci*, 2016, in press.
- Amin S A, Adhikari N, Jha T and Gayen S, Insight into the structural requirements of theophylline-based aldehyde dehydrogenase 1A1 (ALDH1A1) inhibitors through multi-QSAR modeling and molecular docking approaches, *Curr Drug Discov Technol*, 2016, **13**, 84-100.

- 14 Mondal C, Halder A K, Adhikari N, Saha A, Saha K D, Gayen S, *et al.*, Comparative validated molecular modeling of p53-HDM2 inhibitors as antiproliferative agents, *Eur J Med Chem*, 2015, **90**, 860-875.
- 15 Debnath B, Gayen S, Samanta S, Basu A, Ghosh B and Jha T, QSAR study on some synthesized and biologically evaluated glutamine analogs as possible anticancer agents, *Indian J Chem Sec A*, 2006, **45**, 93-99.
- 16 Basu A, Gayen S, Samanta, S, Panda P, Srikanth K and Jha T, QSAR modeling of neonicotinoid insecticides for their selective affinity towards Drosophila nicotinic receptors over mammalian  $\alpha_4\beta_2$  receptors, *Can J Chem*, 2006, **84**, 458-463.
- 17 Chem 3D Pro Version 5.0 and Chem Draw Ultra Version 5.0 are software programs developed by Cambridge Soft Corporation, U. S. A.
- 18 VLife MDS 4.4, VLife Sciences Technologies Pvt. Ltd., Pune, 2015, available from <http://www.vlifesciences.com>.
- 19 Yap C W, Padel-Descriptor: An open source software to calculate molecular descriptors and fingerprints, *J Comput Chem*, 2011, **32**, 1466-1474.
- 20 Jana D, Halder A K, Adhikari N, Maiti M K, Mondal C and Jha T, Chemometric modeling and pharmacophore mapping in coronary heart disease: 2-arylbenzoxazoles as cholesteryl ester transfer protein inhibitors, *Med Chem Comm*, 2011, **2**, 840-852.
- 21 Ambure P, Aher R B, Gajewicz A, Puzyn T and Roy K, "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling, *Chemometr Intell Lab Syst*, 2015, **147**, 1-13.
- 22 STATISTICA version 7 is statistical software of Stat Soft Inc, Tulsa, USA.
- 23 Golbraikh A and Tropsha A, Beware of  $q^2$ ! *J Mol Graph Model*, 2002, **20**, 269-276.
- 24 Adhikari N, Halder A K, Saha A, Saha K D and Jha T, Structural findings of phenylindoles as cytotoxic antimitotic agents in human breast cancer cell lines through multiple validated QSAR studies, *Toxicol in Vitro*, 2015, **29**, 1392-1404.
- 25 Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, *et al.*, admet SAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties. *J Chem Inf Model*, 2012, **52**, 3099-3105.
- 26 Miller S, Kastner S, Krijnse-Locker J, Bu'hler S and Bartenschlager R, The Non-structural Protein 4A of Dengue Virus Is an Integral Membrane Protein Inducing Membrane Alterations in a 2K-regulated Manner, *J Biol Chem*, 2007, **282**, 8873-8882.
- 27 Zou G, Puig-Basagoiti F, Zhang B, Qing M, Chen Z, Pankiewicz K W, *et al.*, A single-amino acid substitution in West Nile virus 2K peptide between NS4A and NS4B confers resistance to lycorine, a flavivirus inhibitor, *Virology*, 2009, **384**, 242-252.
- 28 Omasits U, Ahrens C H, Müller S and Wollscheid B, Protter: interactive protein feature visualization and integration with experimental proteomic data, *Bioinform*, 2014, **30**(6), 884-886.