

# QSAR and Structure Based Molecular Docking of Angelicin Compounds: an Attempt to Drug Design against Swine Influenza

Muneer Alam and Sisir Nandi\*

Global Institute of Pharmaceutical Education and Research, Uttarakhand Technical University, India

## Abstract

There has been a great challenge of research and discovery of novel medicinal leads against swine influenza since 2009. Rational drug design utilizing pharmacoinformatics tools has been augmented now-a-days for *in-silico* screening of lead compounds prior to experimental synthesis, structural elucidation, biological evaluation and finally clinical trials to make the cost efficient drug design and discovery research. There is hardly any specific chemotherapeutics for the treatment against deadly swine influenza viral infection. Therefore, it is an urgent need to design and develop new anti-viral lead compounds active against swine influenza. Quantitative structure activity relationship (QSAR) has been used to develop models that correlate biological activity of angelicin compounds derived from published literature and their computed structural properties. The approach started by generation of a series of descriptors including topological, three dimensional, constitutional, functional groups and atom fragment indices respectively solely calculated from the compounds in the data set. In this study, data set consists of 53 angelicin compounds along with their inhibitory concentration 50% (IC<sub>50</sub>, μM) against H1N1 swine influenza virus. Genetic algorithm-multiple linear regression (GA-MLR) analysis technique has been used to generate a number of QSAR models. The models were validated statistically incorporating training and test set approaches. Finally, structure-based molecular docking study has been performed for interpretation of the mode of binding of the angelicin compounds toward H1N1 target. QSAR and molecular docking analysis of these congeners have not yet been reported. Therefore, this study has significant impact for designing of the highly active compounds in this series that are useful for the treatment of swine influenza. *In-silico* structure based docking model could be helpful for design and screening of congeneric compounds having mode of binding similarity.

**Keywords:** Swine influenza H1N1; Angelicin compounds; Structural descriptors; GA-MLR; QSAR; Molecular docking

## Introduction

Swine influenza is one of the viral family members of Orthomyxoviridae. It contains ribonucleic acid (RNA) as genetic component and has three serological types including A, B and C respectively of which A subtype is a very virulent and infectious which can attack to human [1]. This RNA virus contains hemagglutinin (HA) and neuraminidase (NA) in the cell surface glycoproteins which produces virulent effect. A number of 16 HA (H1 to H16) and 9 NA (N1 to N9) are known in viral glycoproteins. Therefore, a total number of  $16 \times 9 = 144$  distinct progenies could be produced by combinatorial combination of different types of HA and NA. These progenies may be exemplified by H1N1, H1N2...., H2N1, H2N2...H3N2....H3N3.. and so on. Currently, H1, H2 and H3 in combination with N1 and N2 are most frequently dormant strains in human. Remaining subtypes are to be zoonotic, causing disease mainly in fowl and nonhuman primates [2-4]. Swine influenza is originated in swine and easily picked up by wild aquatic birds and other animal species like birds, pigs, ferret, horses, seals, whales, mink, giant anteaters, cats and dogs in which infection is largely intestinal, waterborne and asymptomatic. Then the viral strains emit to the roaming environment of these animals via nasal secretions, saliva, cough, tears and intestinal diarrheal materials. The reassorted strains may infect to the humans adjacent proximity of these animals and cross-infection occurs by transmitting of viral genetic mutant drift [5] between humans and these animals. This may produce genetic re-assortment of the viral strains which become very dangerous. This genetically mutated drift can be speeded around the world and killed almost 100 million people in 1918 in Spain due to H1N1 whereas H2N2 engulfed a number of 4 million people in 1957 in Asia. H3N3 caused death of about one million people in Hong Kong in 1968, and in 2007 pathogenicity of H5N1 Avian influenza strain caused global Health threat. Currently in 2009, swine influenza (H1N1) outbreak in Mexico and other parts of the world has led to issuances of pandemic alertness by the WHO [6,7]. According to WHO approximately 526,060 cases of pandemic H1N1/09 infection and at least 6770 deaths were reported

all over the world by 15 November 2009. This has suddenly arisen in North America and spread rapidly in Europe, Asia and South Africa via human to human transmission within a very short span [8]. As the devastating impact of swine influenza is enormous, a renewed drug discovery effort worldwide is essential to counteract the disease more efficiently. From the literature search it was shown that there is no specific chemotherapeutic agent against swine influenza caused by H1N1 viable strain. Therefore researchers have been trying to develop anti swine influenza chemotherapeutics which are able to combat against swine influenza pandemic.

A series of 1H-1,2,3-triazole-4-carboxamide compounds were synthesized by Cheng et al. [9]. These compounds showed potent biological activity against various strain of H3N2 and H1N1 influenza virus as well as H5N1 (RG14) viral strain by inhibiting the nuclear polymerization. Lepri et al. synthesized a number of 2-aminothiophene-3-carboxamide derivatives and evaluated their biological activities against PA-PB1 interaction complex using Enzyme Linked Immuno Sorbent Assay (ELISA) technique. It was found that the viral polymerase consists of hetero trimetric complex consisting of PB1, PB2 and PA sub units which are essential for transcription and replication of viral m-RNA [10]. High throughput screening confirmed that the angular furocoumarins which is also called as angelicin can produce potent anti-viral activities. Further, Yeh and co-workers synthesized a number of angelicin derivatives and tested

**\*Corresponding authors:** Sisir Nandi, Division of Pharmaceutical Chemistry, Global Institute of Pharmaceutical Education and Research, Uttarakhand Technical University, Kashipur-244713, India, Tel: +91 7500458478; E-mail: [sisir.iicb@gmail.com](mailto:sisir.iicb@gmail.com)

**Received** July 04, 2016; **Accepted** July 19, 2016; **Published** July 26, 2016

**Citation:** Alam M, Nandi S (2016) QSAR and Structure Based Molecular Docking of Angelicin Compounds: an Attempt to Drug Design against Swine Influenza. J Bioanal Biomed 8: 048-057. doi:10.4172/1948-593X.1000152

**Copyright:** © 2016 Alam M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

their biological activities against influenza A (H1N1) virus induced cytopathic effect in Madin-Darby canine kidney cell culture in low micro molar range. Some of these compounds were also found to be highly active against influenza A (H3N2) and influenza B viral strains. These compounds can produce its biological activities by inhibiting neuraminidase containing viral ribonucleoprotein (vRNP) complex [11]. Structure activity relationship studies were performed for these compounds against influenza A/WSN/33 (H1N1) strain. To explore biochemical mechanism of angelicin compounds at a molecular level it has to carry out quantitative structure activity relationship (QSAR) utilizing structural properties but such type of QSAR modeling for these derivatives are not yet done. Quantitative structure activity relationship is an interdisciplinary emerging research area with many interesting and practical applications to environmental protection, new drug discovery, and understanding the molecular basis of property/bioactivity/ toxicity of chemicals. In particular, various classes of graph invariants and molecular descriptors are being increasingly used in these methods [12-13]. Although studies in structure-activity relationships go back to antiquity since the times of Crum-Brown and Fraser [14], it is only in the recent times that one witnesses a vigorous study of it as an interdisciplinary area. Of all recent studies in quantitative structure activity relationship modeling which are interwoven with those of medicinal chemistry, of pharmacokinetics, drug design etc., the seminal account of progress in the subject for the application of chemoinformatics and prediction of biological activities may be found in the book written by Leach and Gillet [15] entitled "An Introduction to Chemoinformatics". One has to mention precursors to this book Waterbeemd [16], Gasteiger [17] and Roy et al. [18] which really grew out for chemometric methods in molecular design and strategy for QSAR development for many QSAR applications requiring the predictions of the property of the molecules and use of these predictions to evaluate, screen and help priorities of the synthesis of these candidates [15-18].

Therefore, it is our aim in the present study to formulate QSAR models under the frame work of various sets of descriptor including topological, three dimensional (3D), constitutional, functional group and atom center fragment indices, which are computed solely from the structure of angelicin compounds utilizing genetic algorithm-multiple linear regression (GA-MLR) methods which can predict essential structural invention responsible for producing biological activity. Finally, mode of actions of these derivatives was explored by molecular docking of the active ligand towards neuraminidase target and that docking model could be applied for the design, screening and synthesis of coumarin compounds supposed to have potent antiviral activities.

## Materials and Methods

### Biological activity data

Linear furobenzopyrones are called as psoralen having reported to induce bifunctional photodamage to the DNA of the cutaneous cells in a selective way, thus inhibiting DNA functions and result in inter-strand cross linkage and produce anti-viral activities. Angular furobenzopyrone or fused furocoumarin in which the furan ring is attached to the 6,7 positions of the coumarin ring is called as angelicin. In the present study, a series of 53 angelicin derivatives (Table 1) have been considered from the published literature. Various group substitutions including R=methyl, ethyl; R<sup>1</sup>=phenyl, methoxy phenyl; R<sup>2</sup>=Benzoyl have been incorporated in the given angelicin nucleus to generate a number of congeners [11]. Angelicin derivatives are capable to inhibit the virus-induced cytopathic effect (CPE) on MDCK

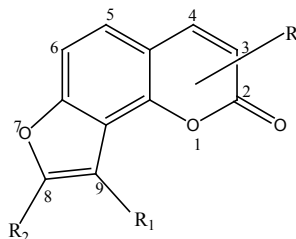
(Madin-Darby canine kidney) cells and the results expressed as IC<sub>50</sub> (concentration required to reduce CPE by 50% relative to the virus control). Activities in terms of IC<sub>50</sub> were converted into pIC<sub>50</sub> which was considered as dependent variables for further QSAR modeling.

### Structure optimization and descriptor calculation

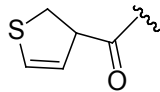
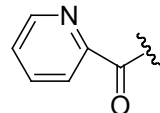
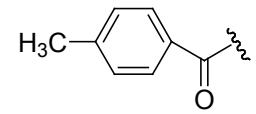
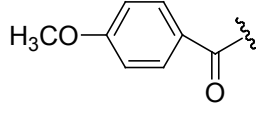
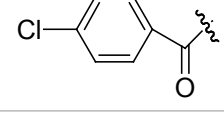
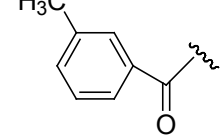
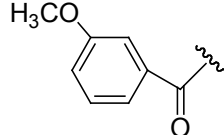
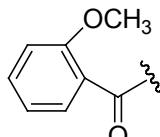
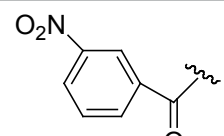
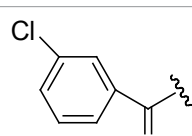
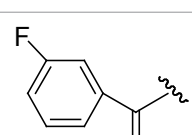
2D angelicin structures were drawn using Chemdraw [19]. The drawn structures were then converted into three-dimensional (3D) modules, and the 3D geometries of all compounds were fully optimized using MM2 force field using a value of 0.01 as dielectric constant considering Chem3D ultra. These energetically minimized stable confirmations were then taken into consideration for the computation of theoretical structural descriptor for further QSAR modeling. Theoretical molecular descriptors are the numerical representation of molecule, achieved by applying the principles of graph theory to molecular structure. It encodes molecular architecture and quantifies such aspects of molecular structure as size, shape, symmetry, complexity, branching, cyclist, stereo electronic character, etc. Structural descriptors can be categorized as topological, 3D, atom centered fragments, constitutional and functional groups respectively. In the present work, DRAGON software [20,21] is used for the computation of theoretical molecular descriptors. A number of 436 topological descriptors were calculated and after elimination, a number of 344 are remaining. A total number of 596 three dimensional (3D) descriptors, useful for our purpose, were calculated via DRAGON software, and before model development, these were reduced to 487. A total number of 131 Constitutional, Functional group and atom-centered fragments descriptors, useful for our purpose, were calculated and before model development, these were reduced to 108, and The reduction in the descriptors was due to keeping a constant value for, or nearly all, of the compounds, and for those that perfectly correlated (r=1.0) with other descriptors. The reduced sets of descriptors were then treated by genetic algorithm- multiple linear regressions (GA-MLR) algorithm for developing QSAR models. Table S1 represents different classes of molecular descriptors along with their symbols.

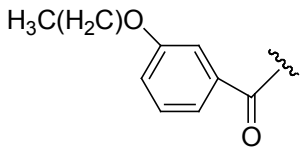
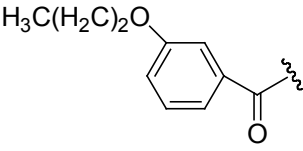
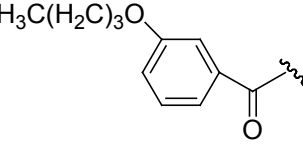
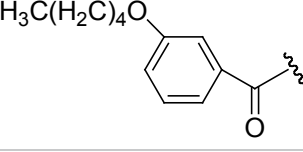
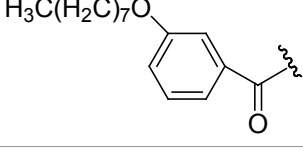
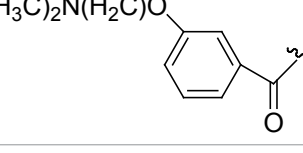
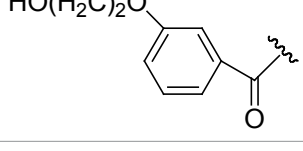
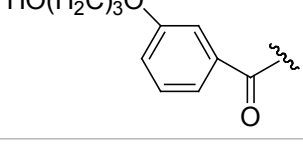
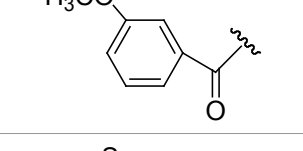
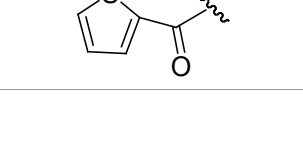
### Statistical analysis by GA-MLR

The reduced descriptor data of angelicin still contain a huge number of predictors which were then treated by genetic algorithm (GA) for selection of features [22] having significant impact on the anti H1N1 activity. Genetic algorithm is a random optimization tool useful for searching large probability space based on the principle of generation of new chromosome by mutation and crossover of parent gene. In this method gene is encoded by descriptor and each chromosome consists of combination of gene representing population consisting of combination of molecular descriptor. The fitness of each chromosome, which reflects the quality of the solution, is evaluated. The next step is the reproduction, which indicates creation of new chromosome (offspring) by selecting existing chromosomes through cross-over and mutation. The fitness of new chromosome is now evaluated and the cycle is repeated until optimal solutions or optimal fitness is attained [23]. Multivariate regression analysis (MRA), one of the oldest data reduction methodologies, continues to be widely used in QSAR, as it does not impose any restriction on the type and number of graphical invariants used in structure-property-activity studies. The ultimate goal of QSAR-based drug design is to find out which structural properties confer the drug highest potency or lowest toxicity. The drug's potency is here a dependent variable, and the structural properties, also called molecular descriptors, are the independent variables. The experimental signal that measures the potency could be, for example, the binding affinity of a drug candidate to its target protein.



Compound No.	R	R <sub>1</sub>	R <sub>2</sub>	IC <sub>50</sub> (μM)	p IC <sub>50</sub> (μM)
1*	4-CH <sub>3</sub>	CH <sub>3</sub>	Benzoyl	4.51	-0.654
2	4-CH <sub>3</sub>	H	Benzoyl	2.43	-0.385
3	4-CH <sub>3</sub>	n-C <sub>3</sub> H <sub>7</sub>	Benzoyl	0.29	0.537
4*	4-CH <sub>3</sub>	n-C <sub>6</sub> H <sub>13</sub>	Benzoyl	0.65	0.187
5	4-CH <sub>3</sub>	1-naphthyl	Benzoyl	0.88	0.055
6	4-CH <sub>3</sub>	2-naphthyl	Benzoyl	1.91	-0.281
7*	4-CH <sub>3</sub>	2-thienyl	Benzoyl	0.15	0.823
8	4-CH <sub>3</sub>	2-furanyl	Benzoyl	0.65	0.187
9	4-CH <sub>3</sub>	4-CH <sub>3</sub> -Ph	Benzoyl	0.82	0.086
10	4-CH <sub>3</sub>	4-CH <sub>3</sub> O-Ph	Benzoyl	10.08	-1.003
11	4-CH <sub>3</sub>	4-NO <sub>2</sub> -Ph	Benzoyl	0.76	0.119
12	4-CH <sub>3</sub>	4-Cl-Ph	Benzoyl	3.92	-0.593
13	4-CH <sub>3</sub>	4-Br-Ph	Benzoyl	0.46	0.337
14	4-CH <sub>3</sub>	3-Br-Ph	Benzoyl	0.08	1.096
15	4-CH <sub>3</sub>	2-Br-Ph	Benzoyl	0.73	0.136
16	4-CH <sub>3</sub>	3,5-diBr-Ph	Benzoyl	0.10	1.000
17	4-CH <sub>3</sub>	3-CH <sub>3</sub> -Ph	Benzoyl	0.10	1.000
18	4-CH <sub>3</sub>	3-NO <sub>2</sub> -Ph	Benzoyl	0.25	0.602
19	4-CH <sub>3</sub>	3-Cl-Ph	Benzoyl	1.41	-0.149
20	4-CH <sub>3</sub>	3-CN-Ph	Benzoyl	0.53	0.275
21	OC(O)CH=CHCH <sub>3</sub>	Ph	Benzoyl	20.82	-1.318
22	H	Ph	Benzoyl	17.6	-1.245
23	3-CH <sub>3</sub>	Ph	Benzoyl	2.79	-0.445
24*	5-CH <sub>3</sub>	Ph	Benzoyl	7.76	-0.889
25*	4-C <sub>2</sub> H <sub>5</sub>	Ph	Benzoyl	0.24	0.619
26*	4-C <sub>3</sub> H <sub>7</sub>	Ph	Benzoyl	0.47	0.327
27	3-CH <sub>3</sub> , 4-CH <sub>3</sub>	Ph	Benzoyl	0.29	0.537
28*				0.72	0.142
29	4-CH <sub>3</sub>	Ph		0.32	0.494
30	4-CH <sub>3</sub>	Ph		0.63	0.200
31	4-CH <sub>3</sub>	Ph		0.41	0.387
32*	4-CH <sub>3</sub>	Ph		0.07	1.154

33	4-CH <sub>3</sub>	Ph		0.15	0.823
34*	4-CH <sub>3</sub>	Ph		0.61	0.214
35	4-CH <sub>3</sub>	Ph		0.87	0.060
36	4-CH <sub>3</sub>	Ph		1.87	-0.271
37*	4-CH <sub>3</sub>	Ph		22.34	-1.349
38	4-CH <sub>3</sub>	Ph		0.26	0.585
39*	4-CH <sub>3</sub>	Ph		0.11	0.958
40	4-OCH <sub>3</sub>	Ph		0.64	0.193
41	4-OCH <sub>3</sub>	Ph		1.71	-0.233
42	4-CH <sub>3</sub>	Ph		0.26	0.585
43*	4-CH <sub>3</sub>	Ph		0.29	0.537

44	4-CH <sub>3</sub>	Ph		0.12	0.920
45*	4-CH <sub>3</sub>	Ph		0.15	0.823
46*	4-CH <sub>3</sub>	Ph		0.08	1.096
47*	4-CH <sub>3</sub>	Ph		0.14	0.853
48	4-CH <sub>3</sub>	Ph		7.79	-0.891
49*	4-CH <sub>3</sub>	Ph		5.28	-0.722
50*	4-CH <sub>3</sub>	Ph		6.09	-0.784
51*	4-CH <sub>3</sub>	Ph		2.65	-0.423
52	4-CH <sub>3</sub>	3-Br-Ph		0.09	1.045
53	4-CH <sub>3</sub>	3-Br-Ph		0.06	1.221

\*Test molecules

Table 1: Biological activity data of angelicin compounds.

GA-MLR has been incorporated in Nanobridges software [24] which is used in the present study to formulate QSAR models of angelicin against H1N1 swine influenza. For the QSAR modeling, training and test set approaches has been incorporated. Training set was used to build the QSAR model whereas test set molecules activity is predicted by the developed training model. In the present study 70% of the 53 molecule were considered as training set while rest 30% was taken as test set compounds as indicated by asterisk in Table 1. The division of training and test data has been carried out by Kennard stone method [25] for model validation. All representative points of the test set has been resembled to those in the training set in the multi-dimensional vector space as well as the representative points of the training set must be well distributed within the whole area occupied by the entire dataset

### Model validation

The quality of each model is denoted by  $R^2$  ( $R$  is the square root of multiple R-square for regression),  $Q^2_{\text{Loo}}$  (cross-validated  $r^2$ ) values for the training set, an external validation was performed by calculating predictive  $R^2$  ( $R^2_{\text{pred}}$ ) and the standard error of estimation, SEE, represents standard deviation which is measured by the error mean square, which expresses the variation of the residuals or the variation about the regression line. Thus standard deviation is an absolute measure of quality of fit and should have a low value for the regression to be significant [26].

$R^2$  and  $Q^2_{\text{Loo}}$  of a model are calculated by

$$R^2 = 1 - \frac{\sum (Y_{\text{obs}} - Y_{\text{calc}})^2}{\sum (Y_{\text{obs}} - \bar{Y})^2} \text{ and } Q^2_{\text{Loo}} = 1 - \frac{\sum (Y_{\text{obs}} - Y_{\text{pred}})^2}{\sum (Y_{\text{obs}} - \bar{Y})^2}$$

Where  $Y_{\text{obs}}$ ,  $Y_{\text{calc}}$  and  $Y_{\text{pred}}$  denote observed, calculated and predicted activity values, respectively, and  $\bar{Y}$  indicates mean activity value of training molecules.  $Q^2_{\text{Loo}}$  denotes predictive statistics which should be greater than 0.5. The validated QSAR's can identify the most significant contribution of the descriptor data modeled.

External validation of the model is carried out by calculating Predicted  $R^2$ .

$$R^2_{\text{pred}} = 1 - \frac{\sum (Y_{\text{pred test}} - Y_{\text{test}})^2}{\sum (Y_{\text{test}} - \bar{Y}_{\text{training}})^2}$$

where,  $Y_{\text{pred test}}$  and  $Y_{\text{test}}$  indicate predicted and observed activity values respectively of the test set compounds and  $\bar{Y}_{\text{training}}$  indicates mean of observed activity values of the training set. For a predictive QSAR model, the value of  $R^2_{\text{pred}}$  should be more than 0.5 [27].

### Structure based docking study of angelicin

In silico molecular docking is one of the sophisticated computational structure based drug design methods to study the ligand-receptor interaction which may produce energetically stable geometry of ligand-receptor complex having minimal interaction energy represented by different scoring functions such as dock score, piecewise linear potential score, potential of mean force score, and steric and electrostatic score. This score is used to predict the binding affinity of a ligand toward receptor. This utility allows screening a set of compounds for lead optimization [28,29]. Present molecular docking study has performed by using ArgusLab 4.0.1 dock engine [30]. It is free software which considers ascore as scoring function. The docking engine is allowed full flexibility of the ligand in side ridge active of the target receptor [31]. In the present study, all the 53 compounds have been docked inside the target cavity of neuraminidase. Compound 32 entitled as 4-Methyl-9-phenyl-8-(thiophene-2-carbonyl)-furo[2,3-h]

chromen-2-one represent highest active lead compounds in angelicin series and considered as an active ligand to predict the mode of binding towards the target cavity. As per the experimental evidence described by Yeh et al., the optimized lead 4-methyl-9-phenyl-8-(thiophene-2-carbonyl)-furo[2,3-h]chromen-2-one (compound number 32, IC50=70 nM (Table 1)) showed many-fold enhanced activity compared to the high throughput screening (HTS) compound 1. Also, compound 32 was found effective in case of influenza A (H3N2) and influenza B virus strains similar to approved anti-influenza drug zanamivir. This compound can produce its biochemical mechanism by inhibition of vRNP (viral-Ribonucleo protein). Neuraminidase H1N1 viral enzyme is responsible for releasing dangerous progeny particle after attachment to the host cell. Therefore, the crystal structure of neuraminidase was considered as a receptor in the presence study. It is worth to mention that structure of neuraminidase-ligand complex is known and one of the PDB Codes of Neuraminidase (PDB ID:3B7E) [32] was downloaded from the Brookhaven Protein Databank (<http://www.rcsb.org>). Then ligand receptor study was carried out by the "Dock ligand tool" of ArgusLab software. All water molecules and HET bound molecules except 1000 GOL which has been co-crystallized were deleted to prepare the target. The target contains a number of binding sites. Binding site of co-crystallized ligand 1000 GOL carries largest surface area and were treated as a reference to make the binding site for the ligand X-ray group. The generated binding cavity consists of all active residues having at least one atom within 3.5 Å from any atom in the co-crystallized ligand X-ray group. Comparative docking analyses of highly active ligands such 32, 53 and 52, intermediate active ligands including 1, 2 and 3 and lower active compounds such 10 and 37 towards target have been well discussed in next section.

## Results and Discussion

### QSAR modeling

A number of QSAR models have been generated for angelicin compounds utilizing different sets of computed descriptor including topological, 3D, constitutional, functional group, atom centered fragments. Out of these models, results of four QSARs based on topological, 3D, topological coupled with 3D and combination of topological, constitutional, functional groups, and atom centered fragments have been discussed in terms of  $R^2$ ,  $Q^2_{\text{Loo}}$ , and predictive  $R^2$ . The QSAR results have been summarized in Table 2.

$R^2$  represent explained variance and  $Q^2_{\text{Loo}}$  represent internal validation predictive statistic whereas  $R^2_{\text{pred}}$  indicates external validation parameter of the develop QSAR models. 3D descriptor alone does not produce satisfactory inhibition of H1N1. Topological descriptor based model can produce 73.20% and 56.10% explained and predictive statistic variances of the H1N1 inhibitory activity of the training compounds while topological along with 3D descriptor based model can produce improved inhibition in terms of 79.80% and 72.70% explained and predictive statistic variances of the H1N1 inhibitory activity of the training compounds. Combination of topological, constitutional, functional groups, and atom centered fragments based model can produce 78.20% and 72.70% explained and predictive statistic variances of the H1N1 inhibitory activity of the training compounds. It was found that out of four QSAR models topological coupled with 3D descriptor based model can produce maximum external predictability in terms of 38.70% of the activity prediction. Topological as well as combination of topological, atom centered, constitutional, and functional group based model may produce all most equal test set predictability given as 30.00%. From comparative statistical validation parameters of four QSAR

Molecular descriptor class	QSAR model
3D	$pIC50 = -1.6385(+/-0.5902) + 0.0817(+/-0.0232) RDF055u + 0.8663(+/-0.484) E3m - 0.0132(+/-0.003) Vs - 0.7342(+/-0.3048) Mor17e + 4.2099(+/-0.935) R7m +$ (1) $N=35, R^2=0.617, Q^2_{Loo} = 0.535, R^2_{pred} = -0.453, PRESS=6.871, SE=0.441$
Topological	$pIC50 = -4.58664(+/-1.29814) + 3.45291(+/-1.27468) MATS5v + 2.18531(+/-0.69791) EEig11r + 0.04186(+/-0.00743) TIC1 - 0.03778(+/-0.00493) Eig1v - 0.37628(+/-0.35574) EEig13r$ (2) $N=35, R^2=0.732, Q^2_{Loo} = 0.561, R^2_{pred} = 0.305, PRESS=6.491, SE=0.369$
Topological+3D	$pIC50 = -5.6026(+/-2.06) - 0.7843(+/-0.5068) GATS5p + 3.5426(+/-1.1811) IVDE + 0.0297(+/-0.0047) TIC3 - 0.0436(+/-0.0051) Eig1v + 4.7074(+/-1.2038) H7v$ (3) $N=35, R^2=0.798, Q^2_{Loo} = 0.723, R^2_{pred} = 0.387, PRESS=4.092, SE=0.320$
Topological + constitutional +functional groups + Atom centered fragments	$pIC50 = -28.2307(+/-4.7059) + 2.8961(+/-1.0646) BEHp3 - 3.0933(+/-0.9962) MATS7m + 11.1004(+/-1.5598) Yindex + 0.9471(+/-0.1956) nR=Ct + 1.9944(+/-0.3094) IC3$ (4) $N=35, R^2=0.782, Q^2_{Loo} = 0.727, R^2_{pred} = 0.299, PRESS=4.028, SE=0.332$

**Table 2:** Different QSAR models along with the statistical quality parameters.

models, it was observed that topological coupled with 3D descriptor based model gives highest impact and acceptable internal validation whereas external predictability in terms of predicted  $R^2$  is found as beyond 0.5. Quality of the models can be improved by applying MAE (mean absolute error)-based criteria [33] and determining applicability domain of a (Q)SAR [34]. Therefore, further attempt has been made to improve the quality of this topological coupled with 3D descriptor based model by detecting outlier compound which is out of fit in the applicability domain (AD) of the response data. "The applicability domain of a QSAR is the physico-chemical, structural, or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds. The applicability domain of a (Q)SAR should be described in terms of the most relevant parameters, i.e., usually those that are descriptors of the model. Ideally the (Q)SAR should only be used to make predictions within that domain by interpolation not extrapolation" [34]. NanoBridges software incorporated "AD using Standardization approach" tool which was used in the present training and test data to identify outlier compounds are located outside the applicability domain of the built QSAR model [35,36]. It was shown that training molecule numbers 2 and 48 as well as test molecule numbers including 25 and 51 were predicted as outliers which do not belong to the zone of applicability domain. In a next attempt, compound numbers 2, 48, 25, and 51 have been omitted from the topological coupled with 3D descriptor data set and again QSAR was modeled which follows as

$$pIC50 = -6.00526(+/-2.04247) - 0.71306(+/-0.50401) GATS5p + 3.18407(+/-1.2205) IVDE + 0.03381(+/-0.00529) TIC3 - 0.04585(+/-0.00552) Eig1v + 5.69041(+/-1.33561) H7v$$

$$N=33, R^2=0.798, Q^2_{Loo}=0.719, R^2_{pred}=0.518, PRESS=3.734, SE=0.336 \quad (5)$$

Validation parameters of this model are quite acceptable and external predictability in terms of predictive  $R^2$  has increased to 0.518. Therefore, this model could be further used to predict biological activity of the newly designed congeneric compounds. Significant parameters included in models (4) and (5) were interpreted in Table 3. Mechanistic interpretation of the modeled descriptors including EEig11r, EEig13r, GATS5p and BEHp3 having positive contribution towards H1N1 inhibition is due to of resonance effect and atomic polarizabilities. Descriptor including TIC1, TIC3, IVDE, Yindex, and IC3 are 3D information indices encoding size, shape, and symmetry

and atom distribution of the inhibitors. Functional groups having higher van der Waal volumes are favorable for the activity which has been captured as 2D autocorrelation indices including MATS5v and H7v as well as functional group index like nR=Ct. Much more increase in the molecular volume, size and mass may produce detrimental effect on the biological activity which is given by Moran autocorrelation of lag 7 weighted by mass (MATS7m) with negative coefficient. These significant predictors could be considered for further generation of highly active angelicin congeners. So the study in this direction can increase the searching of hit rates and decrease in cost of drug design and development prior to experiment by producing potent chemotherapeutics against swine influenza.

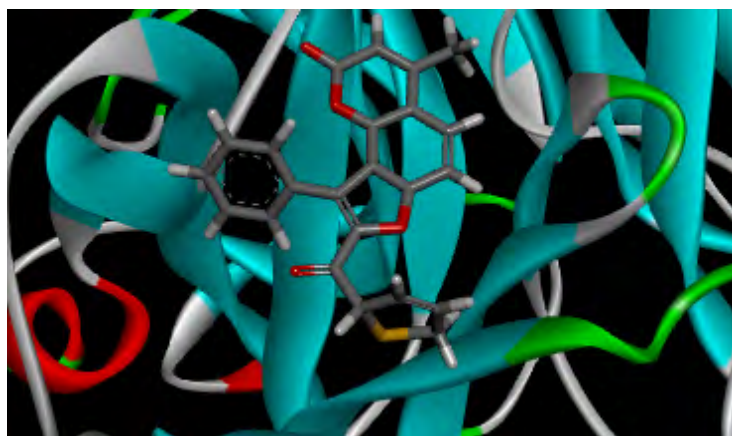
### Angelicin ligand-neuraminidase vRNP target docking

Docking analyses of angelicin- neuraminidase vRNP target revealed that angelicin core can produce Aromatic  $\pi$ -stacking, groups substituted in 8<sup>th</sup>, 9<sup>th</sup> and R are responsible for producing hydrophobic interactions whereas O-atom of angular furan can produce hydrogen bonding. The optimized ligands were docked into active binding cavity of neuraminidase target considering grid resolution (angle) of 0.4 degrees as default value. ArgusLab allows free rotation of the ligand inside the cavity so as to generate a number of 150 poses. The best complex pose with minimal interaction energy of has been taken into consideration for better explanation of mode of interaction between the ligand and active amino acid residues of the receptor protein. Comparative best docking complex pose of some highly active, intermediate active and lower active ligands were analyzed in Table 4. It was found that O-atom of angular furan of intermediate active and lower active compounds do not produce any hydrogen bonding whereas highly active ligand can produce same with TRP 760 which is responsible for the inhibition of Swine influenza viral ribonucleoprotein.

To predict the mode of binding of highly active ligand, interaction patterns of compound 32 has been explained in Figure 1. From the docking study (Figure 1) it was shown 4-methyl group can produce hydrophobic interaction with TYR 738. The fused benzene ring is surrounded by hydrophobic and aromatic residues including TRP 760. This moiety may contribute aromatic pi ( $\pi$ ) stacking interaction. At which electron delocalization may be produce and may enhance the H1N1 ribonucleo protein inhibitory activity. The study in this way may direct to design active congeneric compounds for the treatment of against swine influenza.

Sr. No.	Name of the descriptors	Coefficient	Interpretation
1	MATS5v	+3.45291	MATS5v(2D autocorrelations): Moran autocorrelation of lag 5 weighted by van der Waals volume
2	EEig11r	+2.18531	EEig11r :Eigenvalue 11 from edge .matrix weighted by resonance integrals
3	TIC1	+0.04186	TIC1(Information indices): Total Information Content index (neighborhood symmetry of 1-order
4	Eig1v	-0.03778	Eig1v-Leading eigenvalue from vander waals weighted distance matrix
5	EEig13r	-0.37628	EEig13r :Eigenvalue 13 from edge .matrix weighted by resonance integrals
6	TIC3	+0.0297	Information index: Total information content index (neighborhood symmetry of 3 order)
7	GATS5p	-0.7843	GATS5p(2D autocorrelations): Geary autocorrelation of lag 5 weighted by polarizability
8	IVDE	+3.5426	IVDE(Information indices): mean information content on the vertex degree equality
9	H7v	+4.7074	H7v(GETAWAY descriptors): H autocorrelation of lag 7 / weighted by van der Waals volume
10	BEHp3	+2.8961	Burden index: Highest Eigenvalue n.3 of burden matrix/weighted by atomic polarizability.
11	Yindex	+11.1004	Information indices):Balaban Y index
12	nR=Ct	+0.9471	Functional group counts):number of aliphatic tertiary C(sp2)
13	IC3	+1.9944	Information indices):Information Content index (neighborhood symmetry of 3-order
14	MATS7m	-3.0933	2D autocorrelations):Moran autocorrelation of lag 7 weighted by mass

Table 3: QSAR modeled parameters.

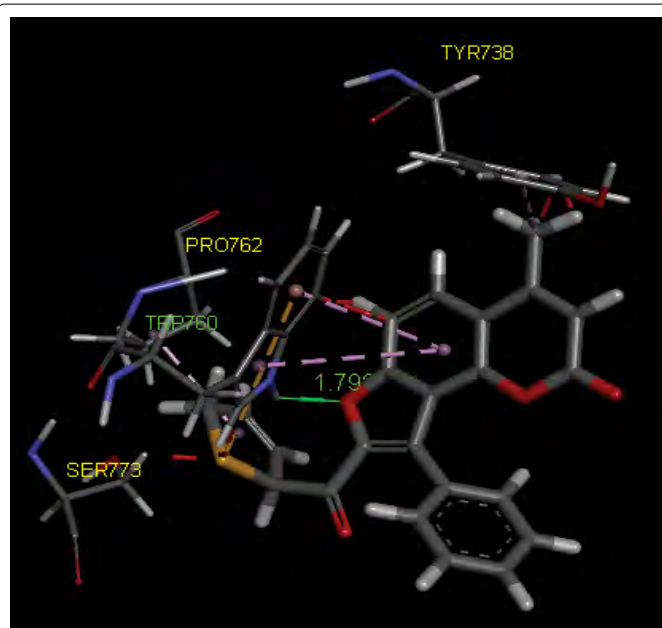


Highly active ligand docked inside the target cavity

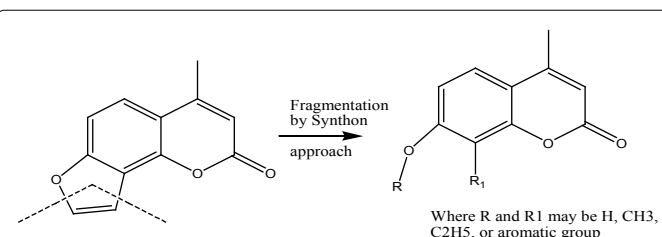
Comp. No.	Dock score Kcal/Mol	Detailed observations of ligand-receptor interaction				
		Angelicin core (Aromatic $\pi$ -stacking)	8 <sup>th</sup> substituents (R <sub>2</sub> ) (Hydrophobic)	9 <sup>th</sup> substituents (R <sub>1</sub> ) (Hydrophobic)	R (Hydrophobic)	O-atom of angular furan (Hydrogen Bonding)
32 (Highly active)	-7.360	TRP 760	PRO 762	-	TYR 738	TRP 760
53 (Highly active)	-7.540	TRP 760	ASP 739	TYR 738; PRO 479	-	TRP 760
52 (Highly active)	-9.167	TRP 760	VAL 774	PRO 762	PRO 479; TRP 743	-
1 (Intermediate active)	-9.820	TYR 738	TRP 743 ( $\pi$ -stacking); VAL 774(hydrophobic)	TRP 760	TYR 738; PRO 479	-
2 (Intermediate active)	-8.790	TYR 738; TRP 760	LEU 477	-	TRP 760; PRO 762	-
3 (Intermediate active)	-9.677	TRP 760; TRP 743; VAL 774 (hydrophobic)	PRO 479	PRO 479	TRP 760; TRP 743; VAL 774	-
10 (lower active)	-6.101	TYR 738; PRO 479 (alkyl hydrophobic)	-	TRP 760	TYR 738	-
37 (lower active)	-6.997	TYR 738 (alkyl hydrophobic)	VAL 774; TRP 760	-	TYR 738; PRO 479	-

Table 4: Comparative study of angelicin ligand-receptor interactions.

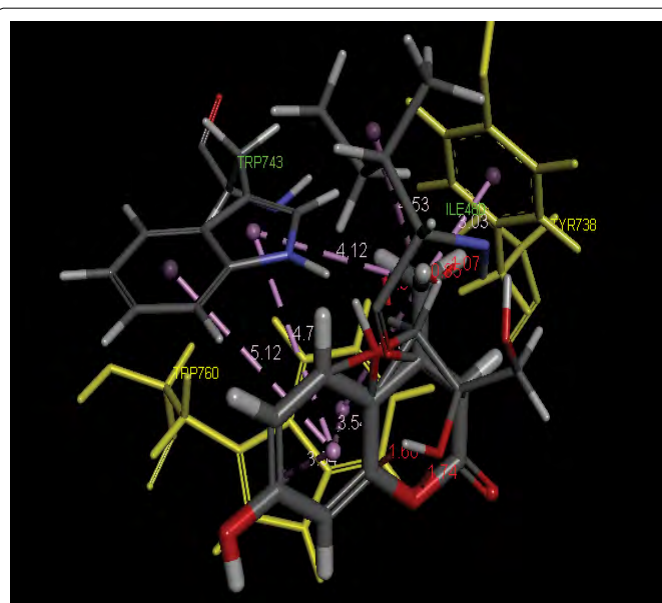




**Figure 1:** Best pose of highly active ligand 32 (stick model) docked inside neuraminidase target cavity represented by molecular surface. The active residues within 4 Å of the inhibitor are displayed. Green color dotted lines represent H-bonding.



**Figure 2:** Schematic Representation of fragmentation of angelicin.



**Figure 3:** Mode of interactions of 7-hydroxy-4-methyl-coumarin docked inside the same cavity of 3B7E.

## Conclusion

From Table 3, it was predicted that the modeled descriptors including EEig11r, EEig13r, GATS5p and BEHp3 contribute resonance effect and atomic polarizabilities. Therefore it was concluded that substitution of electron donating functional groups containing hetero atoms in the angelicin may increase the resonance integrals and polarizability to increase the inhibition of H1N1 neuraminidase. This is in good agreement with the docking findings where fused benzene ring is surrounded by hydrophobic and aromatic residues including TRP 760 that may contribute aromatic pi ( $\pi$ ) stacking interaction. The modeled descriptors including MATS5v and H7v represented three dimensional van der Waals volumes. Therefore, the functional groups with higher van der Waals volume are responsible for interacting H1N1 viral protein amino acid in the active binding cavity which may produce the biological activity. The capture of 3D information indices indicating size, shape, and symmetry and atom distribution has been proved by the significant descriptor including TIC1, TIC3, IVDE, Yindex, and IC3 respectively. Increase in the value of these descriptors may increase the H1N1 viral protein inhibition. Presence of nR=Ct functional group may increase size of the molecule and increase in number of aliphatic tertiary C may contribute hydrophobic interaction in cleft of H1N1 neuraminidase. This model may provide clues for future drug design to combat against swine influenza.

## Future scope: Docking based screening model

Angelicin can be fragmented by utilizing synthon approach [37] proposed by E. J. Corey who introduced the concept of a synthon in retrosynthetic analysis. By fragmentation of angelicin core (Figure 2), 7-hydroxy-4-methyl-coumarin template has been produced which was docked inside the same neuraminidase target cavity and energy of ligand receptor interaction was found as -6.242 kcal/mol.

The study of interaction pattern of 7-hydroxy-4-methyl-coumarin along with 3B7E is given in Figure 3. From this model it is clear that 4-methyl has interacted with TYR 738 and ILE 480 as well as TRP 743. Fused benzene ring has interacted with TRP 760, TRP 743 by aromatic pi-stacking interaction. So there is a similarity between mode of binding of these two docked models which showed TYR 738 and TRP 760 as common interactive residues. Therefore, these docking models could be used for the screening and designing of potent 7-hydroxy-4-methyl-coumarin congeneric compounds supposed to have antiviral activity.

## Acknowledgement

Authors are sincerely thankful to GIPER for providing necessary research facilities. Sincere thanks to Dr. Kunal Roy, Professor, Drug Theoretics and Chemoinformatics Lab, Department of Pharmaceutical Technology, Jadavpur University, India for providing "NanoBridges" software. We are thankful to NIC Ljubljana for availing DRAGON software for the calculation of theoretical molecular descriptors. The authors sincerely acknowledge the valuable comments of the anonymous Reviewers that helped to improve the quality of the final manuscript.

## References

1. Kilbourne ED (2006) Influenza pandemics of the 20th century. *Emerging Infect. Dis.* 12: 9-14.
2. Yuen KY, Wong SS (2005) Human infection by avian influenza A H5N1. *Hong Kong Med J* 11: 189-199.
3. Buchy P, Mardy S, Vong S, Toyoda T, Aubin JT, et al. (2007) Influenza A/H5N1 virus infection in humans in Cambodia. *J Clin Virol* 39: 164-168.
4. Michaelis M, Doerr HW, Cinatl J (2009) Jr Of chickens and men: avian influenza in humans. *Curr Mol Med* 9: 131-51.
5. Vijaykrishna D, Poon LLM, Zhu HC, Ma SK, Li OTW, et al. (2010) Reassortment of Pandemic H1N1/2009 Influenza A Virus in Swine. *Science* 328: pp1529.

6. Fraser C, Donnelly CA, Cauchemez S (2009) Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* 324: pp1557.
7. Carrat F, Flahault A (2007) Influenza vaccine: the challenge of antigenic drift. *Vaccine* 25: 6852-6862.
8. <http://www.who.int/csr/don/2009>. Accessed 8 July 2011.
9. Cheng H, Wan J, Lin MI, Liu Y, Lu X, et al. (2012) Design, Synthesis, and in Vitro Biological Evaluation of 1H-1,2,3-Triazole-4-carboxamide Derivatives as New Anti-influenza A Agents Targeting Virus Nucleoprotein. *J Med Chem* 55: 2144-2153.
10. Lepri S, Nannetti G, Muratore G, Cruciani G, Ruzziconi R, et al. (2014) Optimization of Small-Molecule Inhibitors of Influenza Virus Polymerase: From Thiophene-3-Carboxamide to Polyamido Scaffolds. *J Med Chem* 57: 4337-4350.
11. Yeh JY, Coumar MS, Horng JT, Hui Yi, Shiao HY, et al. (2010) Anti-Influenza Drug Discovery: Structure-Activity Relationship and Mechanistic Insight into Novel Angelicin Derivatives. *J Med Chem* 53: 1519-1533.
12. Basak SC, Mills D, Gute BD (1997) Predicting bioactivity and toxicity of chemicals from mathematical descriptors: A chemicalcum-biochemical approach, in *Advances in Quantum Chemistry*. Elsevier, Academic Press 3: 39-80.
13. Dearden JC (2016) The History and Development of Quantitative Structure-Activity Relationships (QSARs). *International Journal of Quantitative Structure-Property Relationships* 1: 1-44.
14. Crum Brown A, Fraser TR (1968) On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases, derived from Strychia, Brucia, Thebaia, Codeia, Morphia and Nicotia. *Trans R Soc Edinburgh* 25: 151-203.
15. Leach AR, Gillet VJ (2003) *An introduction to chemoinformatics*. Kluwer Academic Publishers, Boston, London.
16. Waterbeemd HV (1995) *Chemoinformatics methods in molecular design*. VCH, Weinheim, New York.
17. Gasteiger J, Marsili M (1980) Iterative Partial Equalization of Orbital Electronegativity-A Rapid Access to Atomic Charges. *Tetrahedron* 36: 3219-3228.
18. Roy K, Kar S, Das RN (2015) *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. 1<sup>st</sup> Edition, Academic Press.
19. Mills N (2006) ChemDraw Ultra 10.0. *J Am Chem Soc* 128:13649-13650.
20. Todeschini R, Consonni V (2009) *Molecular descriptors for chemoinformatics, revised and enlarged edition, 2nd edn*. Wiley, Weinheim.
21. Todeschini R, Consonni V (2006) *Dragon software (version 5.4-2006)*. Milano, Italy.
22. De Campos LJ, De Melo EB (2014) Modeling structure-activity relationships of prodiginines with antimalarial activity using GA/MLR and OPS/PLS. *J Mol Graph Model* 54:19-31.
23. Broadhurst D, Goodacre R, Jone Rowland JJ, Kell BD (1997) Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Anal Chim Acta* 348: 71-86.
24. Ambure, Aher P, Gajewicz RB, Puzyn AT, Roy K (2015) "NanoBRIDGES" software, Open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems* 147: 1-13.
25. Kennard, Ronald W, Larry A Stone (1969) Computer aided design of experiments. *Technometric* 11: 137-148.
26. Nandi S, Bagchi MC (2014) QSAR modeling of 4-anilino-furo [2, 3-b]quinolines: an approach to anticancer drug design. *Med Chem Res* 23:1672-1682.
27. Roy PP, Roy K (2009) Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *Eur J Med Chem* 44:2913-2922.
28. Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. *J Med Chem* 44:1035-1042.
29. Nandi S, Bagchi MC (2010) 3D-QSAR and molecular docking studies of 4-anilinoquinazoline derivatives: a rational approach to anticancer drug design. *Mol Divers* 14:27-38.
30. <http://www.arguslab.com>.
31. Thompson MA, Zerner MC (1990) A theoretical examination of the electronic structure and spectroscopy of the photosynthetic reaction center from *Rhodospseudomonas viridis*. *J Am Chem Soc* 113: 8210-8215.
32. Xu X, Zhu X, Dwek RA, Stevens J, Wilson IA (2008) Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J Virol* 82: 10493-10501.
33. Roy K, Das RN, Ambure P, Aher RB (2016) Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom Intell Lab Sys*. 152: 18-33.
34. OECD (2007) *Guidance document on the validation of (quantitative) structure-activity relationships (Q)SARs Models*, ENV/JM/MONO(2007)2.
35. Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Sys*. 145: 22-29.
36. Gadaleta D, Mangiatordi GF, Catto M, Carotti A, Nicolotti O (2016) Applicability Domain for QSAR Models: Where Theory Meets Reality. *International Journal of Quantitative Structure-Property Relationships* 1: 45-63.
37. Corey EJ (1967) *Pure Applied Chemistry* 14: 30-37.