QSAR Models for Nitrogen Containing Monophosphonate and Bisphosphonate Derivatives as Human Farnesyl Pyrophosphate Synthase Inhibitors Based on Monte Carlo Method

Authors
Parvin Kumar[1], Ashwani Kumar[2], Jayant Sindhu[3], Sohan Lal[1]

Affiliations
1 Department of Chemistry, Kurukshetra University, Kurukshetra, Haryana, India
2 Department of Pharmaceutical Sciences, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India
3 K.M. Govt. College, Narwana, Haryana, India

Key words
Human farnesyl pyrophosphate synthase, Phosphonates, QSAR, IIC, CORAL, SMILES

received 31.05.2018
accepted 01.07.2018

Bibliography
DOI https://doi.org/10.1055/a-0652-5290
Published online: 23.7.2018
Drug Res 2019; 69: 159–167

ISSN 2194-9379

Correspondence
Dr Parvin Kumar
Assistant Professor
Department of Chemistry
Kurukshetra University, Kurukshetra
136119 Haryana
India
Tel.: +91/829/5577 977, Fax: +91/1744/238277
parvinchem@kuk.ac.in; parvinjangra@gmail.com

Supporting Information for this article is available online at http://www.thieme-connect.de/products

ABSTRACT

Human farnesyl pyrophosphate synthase (hFPPS) is a well-settled therapeutic target and it is an enzyme of the mevalonate pathway which catalyzes the biosynthesis of the C-15 isoprenoid farnesyl pyrophosphate. QSAR studies by using Monte Carlo method for human farnesyl pyrophosphate synthase inhibitors has been carried out using balance of correlation technique with Index of ideality correlation. For construction of QSAR models, six random splits were prepared from the data of 73 phosphonates and hybrid optimal descriptors procured from graph (HFG) and SMILES based notations were employed. The developed QSAR models have robustness, good fitting ability, generalizability and internal predictive ability. The external predictive ability has been certified by testing various precedents. The values of R^2, IIC, Q^2 and ΔR^2_m for the best model are 0.9304, 0.9614, 0.9061 and 0.0861 respectively. The developed QSAR models met with the specified standards given in OECD guideline and applicability domain. The structural feature promoters for the end point increase and promoters for end point decrease have been extracted. The predicted pIC_{50} for the new proposed compounds have also been reported.

Introduction

For the drug development, farnesyl pyrophosphate synthase (FPPS) enzyme has been established as a molecular target [1]. FPPS (EC 2.5.1.10), a member of E-family of the prenyltransferases, is an important enzyme in the mevalonate pathway, the elite method of isoprenoid synthesis in animals, concerned with cholesterol biosynthesis and post-translational modification of signaling proteins [2–4]. The human farnesyl pyrophosphate synthase (hFPPS) catalyzes the biosynthesis of the C-15 isoprenoid farnesyl pyrophosphate (FPP) from dimethylallyl pyrophosphate (DMAPP) through geranyl pyrophosphate (GPP) via one by one abridgment of two isopentenyl pyrophosphate (IPP) units [5]. FPP and GGPP play an imperative function in a surfeit of cellular biological functions such as over expression of FPPS in fibroblasts also increases farnesylation of Ras signaling protein and activates the extracellular Ras/ERK signaling cascade [6]. Bisphosphonates, the stable analogues of inorganic pyrophosphate, are a class of drugs that have been used to treat several bone disease such as osteoporosis, post menopausal osteoporosis in elderly women, osteitis, Paget's disease of bone, multiple myeloma, osteogenesis imperfecta and similar diseases [7]. Literature survey reveals that N-containing bisphosphonates (NBps) target FPPS enzyme of the mevalonate pathway [8]. Many NBps drugs

like alendronate, risedronate and pamidronate have been revealed to inhibit FPPS enzyme [9–11].

Recently, J. Park et al. [12] discovered the biological importance of the allosteric pocket, near the IPP substrate binding site of the hFPPS active site, and exhibited that FPP product can bind to this allosteric pocket. This binding sealed the enzyme in an inactive conformation and therefore, suppling a feedback mechanism for regulating the intra-cellular levels of isoprenoid biosynthesis in vivo. To understand this mechanism, the thienopyrimidine-based monophosphonate (ThP-MP) or N-containing monophosphonates (NMps) have been designed and reported by J. Park et al. [1, 3] to get the essential minimal pharmacophore that is required for allosteric inhibition of hFPPS.

Therefore, many scientists are engaged to find out the liaison between the structure and the inhibition activities of phosphonate inhibitors [13–16]. The most successful method for constructing this liaison is Quantitative Structure Activity Relationship (QSAR) method, which is used to predict the drug activities and information for designing new potential drugs [17–21]. In the present era of drug design, the significance of quantitative structure activity relationship method is well recognized in view of the fact that QSAR know how to build the early forecast of activity-related characteristics of drug molecules and can eradicate the molecules with obnoxious properties [22, 23]. Literature survey shows that 3D-QSAR studies have been reported between NBPs and FPPS [24–31]. However, the majority of QSAR study was performed in non-mammalian species although Fernández et al. [32] and Liu et al. [33] reported the 3D-QSAR of N-BPs against hFPPS. It should be more reliable to take into account thienopyrimidine-based monophosphonate (ThP-MP) and N-BPs to study the QSAR of phosphonate.

Recent published papers reveal that the simplified molecular input line entry system (SMILES) is a substitute to classical QSAR methods and it can be used for the prediction of molecular structures with appropriate end point or activity [34–49]. In all the QSAR models, depending on Monte Carlo optimization method, the pertinent activity is treated as random event [50–53]. In the light of these facts and in continuation of our work [48, 54–63] on biological important heterocyclic compounds, we herein report Monte Carlo method based QSAR studies of 73 compounds i. e. thienopyrimidine-based monophosphonate (ThP-MP) and N-containing bisphosphonates N-BPs active against hFPPS using SMILES and graph optimal descriptors.

Method

The data set

A series of 73 phosphonate (NBPs and NMPs) derivatives with activity as hfpps inhibitors was retrieved from literature [3, 32, 33]. The structures of NBPs and NMPs were sketched using Marvin sketch [Marvin Sketch 6.0.4 Chem Axon Ltd. http://www.chemaxon.com] and converted into SMILES by OpenBabel [64]. The SMILES notations of NBPs and ThP-MP molecules are presented in **Table 1S** (supplementary information). The QSAR models were built up for six random splits (20-30 % of compounds were used in calibration set). All six splits were designed according to the following principles: i) the range of the endpoint (pIC_{50}) is evenly distributed for each subset ii) the total level of identity between all the splits is not more than 40 % which shows that splits are different (**Table 2S**, supplementary information).

The identity percentage of the six splits has been confirmed by the reported method [65]. Four sets namely training, invisible-training, calibration, and validation sets were made from all the six splits. Each set has its sole responsibility. The roll of the different sets for developing a QSAR model are: (a) The training set (Train) is designer of the model; (ii) invisible-training (invTrain) set is surveyor of the model, this set sense and prevent the process of overtraining; (iii) the calibration set (Calib) is a specialist and have the authority to declare that the model is ready; (iv) the validation set (Vali) is the reviewer of actual predictive potential of the model.

Index of ideality of correlation (IIC) used to build up predictive model

The balance of correlations is a technique described in the literature [66–69]. The crux of this technique is building up of a model via the Monte Carlo optimization of the following target function (TF):

$$TE = R_{training} + R_{invisible\text{-}training} - | R_{training} - R_{invisible\text{-}training} | \times Const \quad (1)$$

The $R_{training}$ and $R_{invisible-training}$ are correlation coefficients between observed and calculated values of an endpoint for the training and invisible training sets, respectively. The Const is an empirical constant which is usually fixed [69].

In the present manuscript, modified target function (TF_m) for the balance of correlation has been used:

$$TF_m = TF + IIC \quad (2)$$

The index of ideality of correlation (IIC) can be an alternative of traditional correlation coefficient [66, 67]. Literature survey shows that a QSPR/QSAR model without IIC for the prediction of endpoint has some possible defects of error [68, 69].

The index of ideality of correlation (IIC) is calculated with the following formula:

$$IIC = r_{calibration} \times \frac{\min(^{-}MAE_{calibration}, ^{+}MAE_{calibration})}{\max(^{-}MAE_{calibration}, ^{+}MAE_{calibration})} \quad (3)$$

The $r_{calibration}$ is the correlation coefficient value between experimental and calculated values of an endpoint for the calibration set.

MAE is mean absolute error which can be determined with the following equation:

$$^{-}MAE_{calibration} = \frac{1}{^{-}N} \sum_{k=1}^{^{-}N} |\Delta_k| \quad \Delta_k < 0, \quad (4)$$

^{-}N is the number of $\Delta_k < 0$

$$^{-}MAE_{calibration} = \frac{1}{^{+}N}\sum_{k=1}^{^{+}N}\left|\Delta_k\right| \quad \Delta_k \geq 0, \qquad (5)$$

^{+}N is the number of $\Delta_k \geq 0$

The quality of prediction (Δ_k) for one substance from a set can be estimated as the following:

$$\Delta_k = observed_k - calculated_k$$

The optimal descriptor

Descriptors procured from either SMILES or molecular graph can be applied to symbolize the molecular structure. Literature survey reveals that "hybrid" demonstration of the molecular structure, i. e., by SMILES along with molecular graph, can grant a better model demonstrated by higher statistical quality than the model which is predicted by only SMILES or molecular graph[48],[70]. The hybrid optimal descriptor DCW, adopted for generating QSAR models for the pIC50, was determined as per the following equation:

$$^{Hybrid}DCW(T,N_{epoch}) = {}^{SMILES}DCW(T,N_{epoch}) + {}^{Graph}DCW(T,N_{epoch}) \qquad (6)$$

Where T is threshold and N_{epoch} is number of epochs used in monte carlo method optimization. Threshold helps in exclusion of use of rare molecular features.

SMILES based descriptors were calculated using following equation:

$$\begin{aligned} ^{SMILES}DCW(T,N_{epoch}) = {}^{SMILES}&\sum CW(S_k) + \sum CW(SS_k) \\ &+ \sum CW(SSS_k) + CW(BOND) \\ &+ CW(NOSP) + CW(HALO) \\ &+ CW(HARD) + CW(PAIR) \end{aligned} \qquad (7)$$

Where, S_k, SS_k and SSS_k represents local smile attributes while BOND, NOSP, HALO, HARD and PAIR represent global SMILES attributes and exhibit presence of double, triple or stereochemical bond, presence of nitrogen, oxygen, sulphur and phosphorus, presence of halogen as well as pairing between these attributes[71]. Optimal descriptors based on graph were calculated using following equation:

$$\begin{aligned} ^{GRAPH}DCW(T,N_{epoch}) = &\sum CW(EC0_k) + \sum CW(EC1_k) \\ &+ \sum CW(EC2_k) + \sum CW(EC3_k) \\ &+ \sum CW(VS2_k) + CW(NNC_k) \\ &+ CW(C3) + CW(C4) \\ &+ CW(C5) + CW(C6) \end{aligned} \qquad (8)$$

Where, $EC0_k$, $EC1_k$, $EC2_k$ are Morgan's extended connectivity indices for k^{th} vertex respectively. $VS2_k$ is the sum of vertex degrees which take place at topological distance 2 relative to k^{th} vertex. NNC is the nearest neighbor code. The C3, C4, C5 and C6 are descriptors for presence of three-member (C3), four-member (C4), five-member (C5) and six-member rings (C6) respectively. In this study, hydrogen filled graphs (HFG) have been used.

CORALSEA software was used to prepare the QSAR models. The best model was scrutinized using the procedure given in literature[70]. The most prognostic combination of threshold (T) and number of epochs (N_{epoch}) for the six splits was executed from values 1-10 for T and 1-50 for N_{epoch}. The model was built using balance of correlation technique with index of ideality of correlation (IIC) of Monte Carlo method. The weight of dr was 0.1. Start step of the optimization was 0. 5 * CW(SA). Precision of the optimization was 0.1 * CW(SA) and weight of IIC was 0.1. Here, CW(SA) is weight of structural attribute (SA) at the start.

Possessing numerical information on above CW, DCW (T, N_{epoch}) can be computed for molecules of training and test set. These data can be used for calculation of pIC_{50} according to following ▶ **Eq. (9)**:

$$pIC_{50} = C0 + C1^{*Hybrid}DCW(T,N_{epoch}) \qquad (9)$$

Validation of QSAR model

The most important objective of any QSAR modeling is to establish a sturdy model competent to predict the idiosyncrasy of new molecules in an objective, reliable and precise manner [49, 72]. Three methods are cited in the literature for evaluation of sturdiness and reliability of developed model. These are: (a) internal validation or cross-validation using the training set compounds (b) external validation using the test set compounds (c) data randomization or Y-scrambling.

Leave-one-out (LOO) cross validation technique was used to develop models as an internal validation. Cross-validated Q^2 interpret the predictive ability of the model [73, 74]. Higher the value of Q^2 means better model prediction. The cross-validated Q^2 is stated as:

$$Q^2 = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - \bar{Y}_{train})^2} \qquad (10)$$

Where, Y_{obs} is observed property of the training set compounds, Y_{pred} is LOO-predicted property of the training set compounds and is mean observed property of the training set compounds. The predictive ability of model is considered as acceptable when Q^2 is greater than 0.5.

Similar methodology can be applied for external validation. The predictive ability of a model is determined by calculating Q^2_{ext} which is defined as:

$$Q^2_{ext} = 1 - \frac{\sum(Y_{obs\ (test)} - Y_{pred\ (test)})^2}{\sum(Y_{obs(test)} - \bar{Y}_{train})^2} \qquad (11)$$

Where, $Y_{obs(test)}$ is the observed property of the test set compounds, $Y_{pred(test)}$ is the predicted property of the test set compounds and $\overline{Y}_{train}$ is mean observed property of the training set compounds. The value Q^2_{ext} for an acceptable model should be greater than 0.5 test [75].

Y-randomization test was performed to inspect the robustness of the model. A parameter $^cR^2_p$ penalizes the model R^2 for a small difference between squared mean correlation coefficient (R^2r) of randomized models and squared correlation coefficient (R^2) of the non-randomized model [75]. The parameter $^cR^2_p$ is defined as:

$$^cR^2_p = R\sqrt{(R^2 - R^2_r} \quad (12)$$

For an acceptable QSAR model, the value of $^cR^2_p$ should be greater than 0.5.

Compliance with Applicability domain (AD) and OECD principles

Applicability domain (AD) is another important aspect of a built QSAR model. It gives the information related to the biological, structure and physiochemical properties to make the prophecy for new compounds. AD is especially significant because it is applied for the evaluation of the authenticity of the developed QSAR model. Compounds from the training set can be used to interpret the AD of the developed QSPR model. The predictions of a QSAR/QSPR model are more authentic when predicted molecules are within applicability. For defining AD, procedure reported in literature was applied [65].

To assist the reflection of a QSAR model for regulatory purposes [76], it should satisfy the five principal as: 1) a defined endpoint: hFPPS inhibitory activity as definite endpoint, 2) an unambiguous algorithm: Monte Carlo method as unambiguous algorithm, 3) a defined domain of applicability: percentage of molecular features with defined role as domain of applicability, 4) appropriate measures of goodness-of–fit, robustness and predictivity: High values of R^2, Q^2; $^cR^2_p$, $R^2_m(av)$ and ΔR^2_m metrics, as suitable measures of goodness-of-fit, robustness and predictivity 5) a mechanistic interpretation, if possible: List of molecular features responsible for increase and decrease of activity are applied for mechanistic interpretation.

Result and Discussion

Six random splits were generated from the data set presented in **Table 1S** (supplementary information). All sets were carefully prepared so the ranges are approximately equivalent for each sub-set. The hybrid optimal descriptor DCW was adopted for generating QSAR models. The outcomes of the applied methodology for defining AD reveal that maximum numbers of molecules are within the defined AD. So, all the studied compounds had typical behavior and all were incorporated in developing QSAR models.

The best QSAR models for six different splits with other statistical parameters of all six equations are given in ▶ **Table 1**.

The statistical parameters given in ▶ **Table 1** clearly indicate that the values of statistical criteria of all six equations are significant for each subset i. e. training, invisible- training, calibration and validation sets. The statistical trait of calibration set ($R^2 = 0.9304$ and $Q^2 = 0.9061$) for the equation of split 4 was distinguished best, therefore the QSAR model expressed by this was judged to be the preeminent model. The good fitting ability and good internal predictive power of the described model has been implied by the R^2 value (0.7380) for the training sets and Q^2 value of calibration set. The value of R^2 is 0.7602 for validation sets which shows the excellent external predictive ability of the developed QSAR model. The authenticity and robustness of the QSAR models are justified by less difference between R^2 and Q^2 values.

Y-randomization process was applied to check the chance correlation, in which Y values were scrambled in 1000 trials in ten separate runs (**Table 3S**, supplementary information). The resulting value of $^cR^2_p$ was found more than 0.5 which authenticate that the developed models are free from chance correlation [75].

The built QSAR model was checked for the external predictive power by applying the benchmarks proposed by Golbraikh and Tropsha [74], Roy and Roy [77] and Ojha et al. [78]. The values of different proposed benchmarks are displayed in **Table 4S** (supplementary information) and it can be seen that these values are within specified ranges. Therefore, it can be said that these QSAR equations have good external predictive ability. Calculated activity and applicability domain of all compounds along with different sets are exhibited in supplementary **Table 5S** (supplementary information). The procedure described in literature was used for the calculation of applicability domain and it is based on split defect [66].

The plots between observed versus calculated activity and residuals versus observed activity are shown in ▶ **Fig. 1**.

Mechanistic interpretation of developed QSAR models

The correlation weights (CW) of structural features (SAk) give the information about the structural attributes which are responsible for the increase and decrease of the endpoint. If CW(SAk) is positive in all three probes then this feature will enhance the value of endpoint and if CW(SAk) is negative in all three probes then it will reduce the value of endpoint. Based on these considerations, some graph based structural attributes with stable positive values (pIC50 enhancer) for best QSAR model 3 are EC0-O...1..., EC0-P...4..., EC1-O...4..., NNC-O...101., EC1-C...5..., NNC-P...413, EC2-C...10.., VS2-C...6... etc. Similarly, some graph based structural attributes with stable negative values (pIC50 decreasing) for this model are EC0-C...2..., EC1-C...4..., EC0-C...3..., EC0-N...2..., EC1-C...6..., NNC-C...220., VS2-C...9... etc.

In the same manner, some pIC50 enhancer SMILES based attributes are (...O...(... (branched oxygen with branching), + + + + N---B2 = = (nitrogen with double bond), + + + + N---O = = = (nitrogen with oxygen), + + + + N---P = = = (nitrogen with phosphorous), + + + + O---B2 = = (oxygen with double bond), + + + + O---P = = = (oxygen with phosphorous), + + + + P---B2 = = (phosphorous with double bond), = ... (.....(double bond with branching), = ...(double bond), C..... (methyl group), O...(...(...(oxygen with double branching), O...(..... (oxygen with branching), O..... (oxygen), O... = ...(...(oxygen with double bond and branching), P...(.....(phosphorous with branching), P.....(phosphorous), C4.....0...(absence of four membered ring), C3.....0...(absence of three membered ring), BOND10000000(one

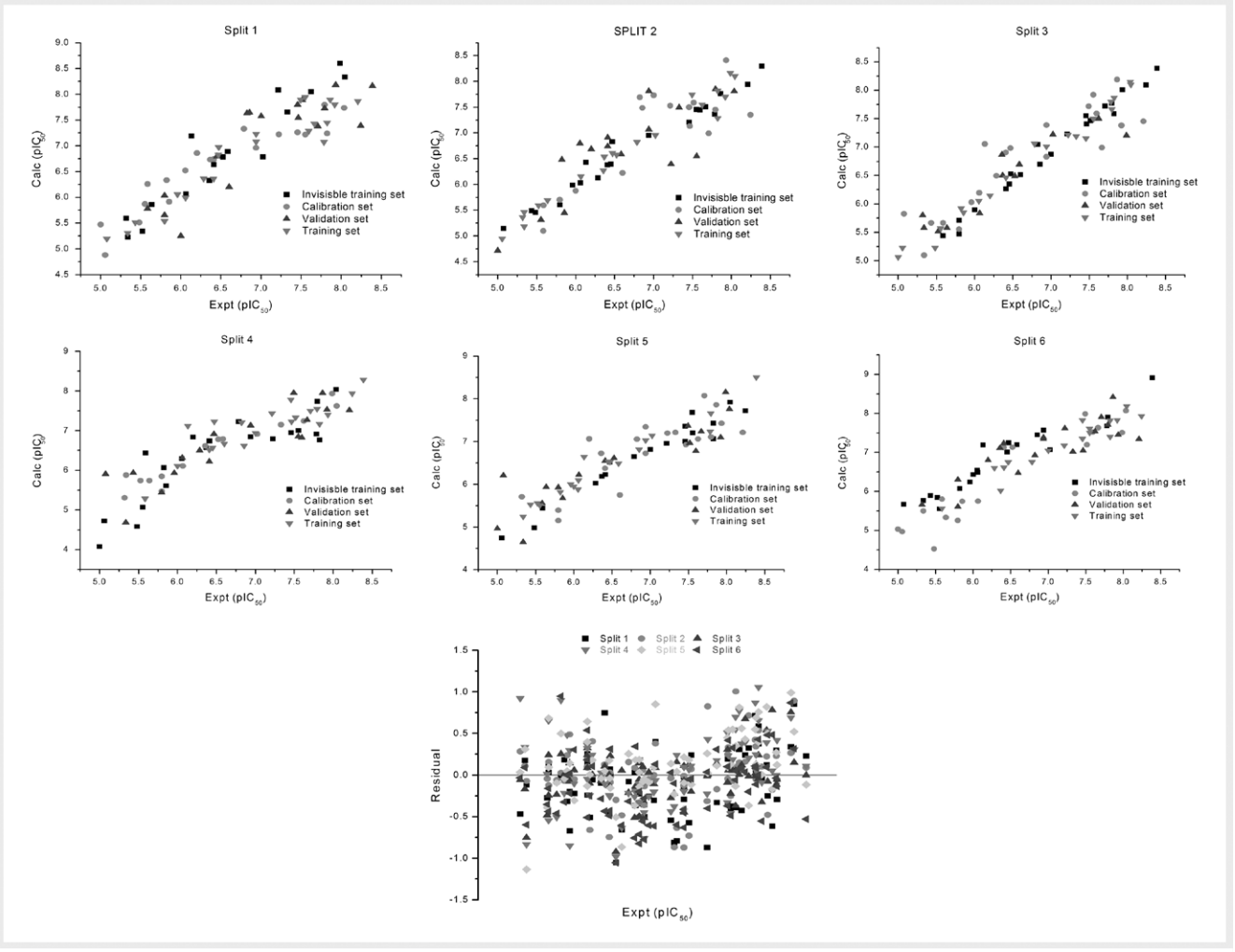
▶ **Table 1** Statistical parameters of all six QSAR Models

Split	SET	n	R^2	CCC	IIC	Q^2	s	MAE	F	R^2-Q^2	Equation for different Models
Split 1	Training	21	0.9019	0.9484	0.8634	0.8848	0.303	0.238	175	0.0171	**pIC_{50}** = 3.6970585 (±0.0365511) + 0.0163073 (±0.0002255) * DCW(3,27)
	InvTrain	16	0.9053	0.9002	0.2412	0.8835	0.451	0.339	134	0.0218	
	Calib	19	0.8667	0.9020	0.9309	0.8254	0.398	0.318	110	0.0413	
	Validation	17	0.7466	NC	0.8479	0.6801	0.486	0.398	44	0.0665	
Split 2	Training	20	0.9722	0.9859	0.6574	0.9662	0.173	0.126	628	0.006	**pIC_{50}** = -0.1563000 (±0.0563218) + 0.0282081 (±0.0002523) * DCW(2,48)
	InvTrain	20	0.9722	0.9805	0.9581	0.9676	0.189	0.143	629	0.0046	
	Calib	17	0.7386	0.8546	0.8404	0.6860	0.498	0.389	42	0.0526	
	Validation	16	0.6587	NC	0.6432	0.5872	0.544	0.433	27	0.0715	
Split 3	Training	20	0.9827	0.9913	0.8112	0.9787	0.134	0.096	1025	0.004	**pIC_{50}** = -1.5927099 (±0.0653207) + 0.0314089 (±0.0002405) * DCW(1,37)
	InvTrain	21	0.9828	0.9872	0.5259	0.9792	0.137	0.109	1083	0.0036	
	Calib	20	0.7924	0.8797	0.8902	0.7531	0.458	0.366	69	0.0393	
	Validation	12	0.8513	NC	0.6793	0.7840	0.344	0.243	57	0.0673	
Split 4	Training	20	0.7380	0.8492	0.7028	0.6888	0.400	0.307	51	0.0492	**pIC_{50}** = -1.3745288 (±0.2456229) + 0.0363736 (±0.0010521) * DCW(7,13)
	InvTrain	20	0.7361	0.8433	0.5890	0.6926	0.575	0.469	50	0.0435	
	Calib	15	**0.9304**	**0.9538**	**0.9614**	**0.9061**	**0.267**	**0.209**	**174**	**0.0243**	
	Validation	18	0.7602	NC	0.6703	0.6864	0.486	0.409	51	0.0738	
Split 5	Training	19	0.9483	0.9735	0.8766	0.9365	0.195	0.124	312	0.0118	**pIC_{50}** = 0.4692126 (±0.0724623) + 0.0249464 (±0.0003331) * DCW(2,17)
	InvTrain	18	0.9467	0.9341	0.3001	0.9335	0.342	0.279	284	0.0132	
	Calib	19	0.6500	0.7905	0.7903	0.5775	0.530	0.431	32	0.0725	
	Validation	17	0.8111	NC	0.6967	0.7591	0.493	0.356	64	0.052	
Split 6	Training	17	0.9103	0.9530	0.6677	0.8843	0.224	0.166	152	0.026	**pIC_{50}** = 1.8911561 (±0.1196966) + 0.0174180 (±0.0003871) * DCW(1,15)
	InvTrain	20	0.9070	0.8486	0.2262	0.8871	0.517	0.438	176	0.0199	
	Calib	17	0.8749	0.9289	0.9348	0.8376	0.428	0.322	105	0.0373	
	Validation	18	0.7417	NC	0.4430	0.6865	0.450	0.380	46	0.0552	

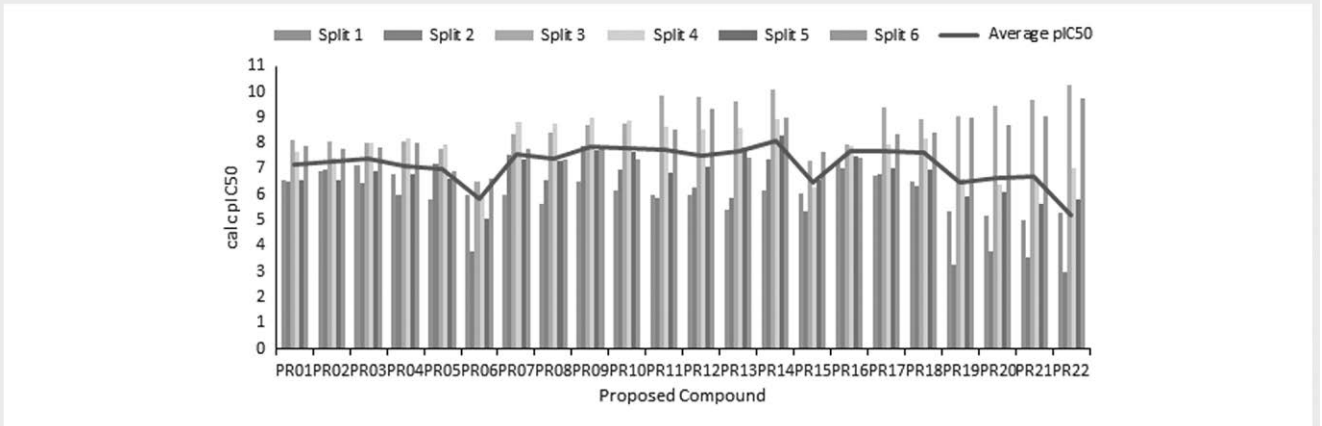
Where, n is number of cases, R^2 is the squared correlation coefficient, CCC is concordance correlation coefficient, IIC is Index of ideality of correlation, Q^2 is the leave-one-out (LOO) cross-validation coefficient, s is standard error of estimation, MAE is mean average error, F is the Fischer ratio and NC means not calculated.

double bond), 1...(....., HALO00000000 (absence of halogen), = ...O... (...(double bond with oxygen and branching), P...(...O...(phosphorous with branched oxygen) etc and some SMILES based promoters of pIC50 decrease are (...(.....(branching with further branching), (.....(branching), C...(.....(methyl with branching), O... = ...(oxygen with double bond), 1.....(presence of one ring), (...P...(...(branch having phosphorous with branching) etc. The structure generated using the promoters of increase in end point with their predicted pIC50 are given in **Table 6S** (supplementary information). By using the promoters of endpoint, we have developed some more hFPPS inhibitors. We have predicted the pIC_{50} by all QSAR models using their respective equation and then average pIC50 was calculated. On the basis of average calculated pIC50, compound 14 was found most potent by using model of split 3 but the difference between the maximum and minimum calculated pIC50 from all models is 3.8938. Compound 16 shows nearly constant value of calculated pIC50 and the difference between the maximum and minimum calculated pIC50 is 0.9059. The graphical representation of all the calculated pIC50 with average calculated pIC50 is shown in ▶ **Fig. 2**. The result of **Table 6S** (supplementary information) clearly shows that the bisphosphonates developed as hFPPS inhibitors are more potent inhibitors than monophosphonates. It has been reported in literature [1, 3, 7–12] that the interactions of the bisphosphonates in the active site of the HFFPS are highly conditioned by the protonation state of the functional groups but the method used in this work does not consider this aspect. It is a limitation of this work. Despite that the proposed QSAR models predicted the activity of reported bisphosphonates and monophosphonates accurately.

Two 3D QSAR model developed by comparative molecular field analysis (CoMFA) method for the some part of our dataset has been described by Fernández et al. [32] and Liu et al.[33]. The values for statistical features used by Fernández et al. [32] to examine the model were n = 20, R^2 = 0.943, Q^2 = 0.586, F = 63 and SEE = 0.11 and by Liu et al. [33] to examine the model were $n_{(training)}$ = 42, $R^2_{(training)}$ = 0.975, $n_{(test)}$ = 11 and $R^2_{(test)}$ = 0.0.753. In the reported QSAR models, monophosphonates were not used, but have used to develop a relationship between monophosphonate and bisphosphonates. So, we can say that the developed QSAR models in the present manuscript have better statistical quality as compared to the reported model in terms of internal as well as external prediction criteria. Also, the resent QSAR models are one parameter 2D QSAR models which are simple in interpretation and simple to apply.



▶ **Fig. 1** The plots between observed pIC50 versus calculated pIC50 and residuals versus observed pIC50.



▶ **Fig. 2** Graphical representation of calculated pIC50 from all models for all proposed compound.

Conclusion

Monte carlo optimization method using CORAL software has been used successfully for designing a statistically robust QSAR models for human farnesyl pyrophosphate synthase inhibitors. The best QSAR model described by DCW(7,13) optimal descriptor was obtained from split 4. In the mechanistic interpretation, the presence of nitrogen with double bond, a phosphate group and oxygen with branched double bond were found promoter of endpoint. The dis-

cussed QSAR models satisfy the OECD conditions for a good QSAR model. The outcomes of the applied methodology for defining AD reveal that maximum numbers of molecules are within the defined AD. Thus, this approach can be used for generation of new potential human farnesyl pyrophosphate synthase inhibitors.

Acknowledgements

The authors are grateful to Dr Andrey A. Toropov and Dr Alla P. Toropova for providing CORAL software.

Conflict of Interest

The authors declare no conflicts of interest.

References

[1] Park J, Rodionov D, De Schutter JW et al. Crystallographic and thermodynamic characterization of phenylaminopyridine bisphosphonates binding to human farnesyl pyrophosphate synthase. Plos One 2017; 12: e0186447.DIO: 10.1371/journal.pone.0186447.

[2] Leung CY, Langille AM, Mancuso J et al. Discovery of thienopyrimidine-based inhibitors of the human farnesyl pyrophosphate synthase-Parallel synthesis of analogs via a trimethylsilyl ylidene intermediate. Bioorgan Med Chem 2013; 21: 2229–2240

[3] Park J, Leung CY, Matralis AN et al. Pharmacophore Mapping of Thienopyrimidine-Based Monophosphonate (ThP-MP) Inhibitors of the Human Farnesyl Pyrophosphate Synthase. J Med Chem 2017; 60: 2119–2134

[4] Szabo CM, Matsumura Y, Fukura S et al. Inhibition of geranylgeranyl diphosphate synthase by bisphosphonates and diphosphates: A potential route to new bone antiresorption and antiparasitic agents. J Med Chem 2002; 45: 2185–2196

[5] Lin YS, Park J, De Schutter JW et al. Design and Synthesis of Active Site Inhibitors of the Human Farnesyl Pyrophosphate Synthase: Apoptosis and Inhibition of ERK Phosphorylation in Multiple Myeloma Cells. J Med Chem 2012; 55: 3201–3215

[6] Hooff GP, Peters I, Wood WG et al. Modulation of Cholesterol, Farnesylpyrophosphate, and Geranylgeranylpyrophosphate in Neuroblastoma SH-SY5Y-APP695 Cells: Impact on Amyloid Beta-Protein Production. Mol Neurobiol 2010; 41: 341–350

[7] Russell RGG. Bisphosphonates: The first 40 years. Bone 2011; 49: 2–19

[8] Tsoumpra MK, Muniz JR, Barnett BL et al. The inhibition of human farnesyl pyrophosphate synthase by nitrogen-containing bisphosphonates. Elucidating the role of active site threonine 201 and tyrosine 204 residues using enzyme mutants. Bone 2015; 81: 478–486

[9] De Schutter JW, Zaretsky S, Welbourn S et al. Novel bisphosphonate inhibitors of the human farnesyl pyrophosphate synthase. Bioorg Med Chem Lett 2010; 20: 5781–5786

[10] Dozier JK, Distefano MD. An enzyme-coupled continuous fluorescence assay for farnesyl diphosphate synthases. Anal BioChem 2012; 421: 158–163

[11] Dhar MK, Koul A, Kaul S. Farnesyl pyrophosphate synthase: A key enzyme in isoprenoid biosynthetic pathway and potential molecular target for drug development. N Biotechnol 2013; 30: 114–123

[12] Park J, Zielinski M, Magder A et al. Human farnesyl pyrophosphate synthase is allosterically inhibited by its own product. Nature Communications 2017; 8: 14132

[13] Gisselberg JE, Herrera Z, Orchard LM et al. Specific Inhibition of the Bifunctional Farnesyl/Geranylgeranyl Diphosphate Synthase in Malaria Parasites via a New Small-Molecule Binding Site. Cell Chem Biol 2018; 25: 185–193

[14] Fernandez D, Ramis R, Ortega-Castro J et al. New insights into human farnesyl pyrophosphate synthase inhibition by second-generation bisphosphonate drugs. J Comput Aid Mol Des 2017; 31: 675–688

[15] Fernandez D, Ortega-Castro J, Frau J. Theoretical study of the deposition and adsorption of bisphosphonates on the 001 hydroxyapatite surface: Implications in the pathological crystallization inhibition and the bone antiresorptive action. Appl Surf Sci 2017; 392: 204–214

[16] Rodriguez JB, Falcone BN, Szajnman SH. Approaches for Designing new Potent Inhibitors of Farnesyl Pyrophosphate Synthase. Expert Opin Drug Dis 2016; 11: 307–320

[17] Abdel-Hamid MK, Abdel-Hafez AA, El-Koussi NA et al. Quantitative structure-activity relationship (QSAR) studies on a series of 1,3,4-thiadiazole-2-thione derivatives as tumor-associated carbonic anhydrase IX inhibitors. J Enzyme Inhib Med Chem 2009; 24: 722–729

[18] Gao H, Katzenellenbogen JA, Garg R et al. Comparative QSAR analysis of estrogen receptor ligands. Chem Rev 1999; 99: 723–744

[19] Hadjipavlou-Litina D, Garg R, Hansch C. Comparative quantitative structure-activity relationship studies (QSAR) on non-benzodiazepine compounds binding to benzodiazepine receptor (BzR). Chem Rev 2004; 104: 3751–3794

[20] Hansch C, Hoekman D, Gao H. Comparative QSAR: Toward a Deeper Understanding of Chemicobiological Interactions. Chem Rev 1996; 96: 1045–1076

[21] Thareja S. Steroidal 5alpha-Reductase Inhibitors: A Comparative 3D-QSAR Study Review. Chem Rev 2015; 115: 2883–2894

[22] Agatonovic-Kustrin S, Morton DW, Celebic D. QSAR: an in silico approach for predicting the partitioning of pesticides into breast milk. Comb Chem High Throughput Screen 2013; 16: 223–232

[23] Aouidate A, Ghaleb A, Ghamali M et al. 3D QSAR studies, molecular docking and ADMET evaluation, using thiazolidine derivatives as template to obtain new inhibitors of PIM1 kinase. Comput Biol Chem 2018; 74: 201–211

[24] Martin MB, Sanders JM, Kendrick H et al. Activity of bisphosphonates against Trypanosoma brucei rhodesiense. J Med Chem 2002; 45: 2904–2914

[25] Szabo CM, Martin MB, Oldfield E. An investigation of bone resorption and Dictyostelium discoideum growth inhibition by bisphosphonate drugs. J Med Chem 2002; 45: 2894–2903

[26] Sanders JM, Gomez AO, Mao J et al. 3-D QSAR investigations of the inhibition of Leishmania major farnesyl pyrophosphate synthase by bisphosphonates. J Med Chem 2003; 46: 5171–5183

[27] Prokopenko V, Kovalishyn V, Shevchuk M et al. Design and synthesis of new potent inhibitors of farnesyl pyrophosphate synthase. Curr Drug Discov Technol 2014; 11: 133–144

[28] Ling Y, Sahota G, Odeh S et al. Bisphosphonate inhibitors of Toxoplasma gondi growth: In vitro, QSAR, and in vivo investigations. J Med Chem 2005; 48: 3130–3140

[29] Kotsikorou E, Oldfield E. A quantitative structure-activity relationship and pharmacophore modeling investigation of aryl-X and heterocyclic bisphosphonates as bone resorption agents. J Med Chem 2003; 46: 2932–2944

[30] C KMC Hudock MP, Zhang Y et al Inhibition of geranylgeranyl diphosphate synthase by bisphosphonates: A crystallographic and computational investigation. J Med Chem 2008; 51: 5594–5607

[31] Sanders JM, Ghosh S, Chan JM et al. Quantitative structure-activity relationships for gammadelta T cell activation by bisphosphonates. J Med Chem 2004; 47: 375–384

[32] Fernandez D, Ortega-Castro J, Frau J. Human farnesyl pyrophosphate synthase inhibition by nitrogen bisphosphonates: A 3D-QSAR study. J Comput Aided Mol Des. 2013; 27: 739–754

[33] Liu QZ, Wang SS, Li X et al 3D-QSAR, molecular docking, and ONIOM studies on the structure-activity relationships and action mechanism of nitrogen-containing bisphosphonates. Chem Biol Drug Des 2018; 91: 735–746

[34] Trinh TX, Choi JS, Jeon H et al. Quasi-SMILES-Based Nano-Quantitative Structure-Activity Relationship Model to Predict the Cytotoxicity of Multiwalled Carbon Nanotubes to Human Lung Cells. Chem Res Toxicol. 2018; 31: 183–190

[35] Stoickov V, Stojanovic D, Tasic I et al QSAR study of 2,4-dihydro-3 H-1,2,4-triazol-3-ones derivatives as angiotensin II AT(1) receptor antagonists based on the Monte Carlo method. Struct Chem 2018; 29: 441–449

[36] Prachayasittikul V, Worachartcheewan A, Toropova AP et al. Large-scale classification of P-glycoprotein inhibitors using SMILES-based descriptors. Sar Qsar Environ Res 2017; 28: 1–16

[37] Kumar A, Chauhan S. QSAR Differential Model for Prediction of SIRT1 Modulation using Monte Carlo Method. Drug Res 2017; 67: 156–162

[38] Kumar A, Chauhan S. Monte Carlo method based QSAR modelling of natural lipase inhibitors using hybrid optimal descriptors. Sar Qsar Environ Res 2017; 28: 179–197

[39] Kumar A, Chauhan S. Use of the Monte Carlo Method for OECD Principles-Guided QSAR Modeling of SIRT1 Inhibitors. Arch Pharm 2017; 350: e1600268. DOI: 10.1002/ardp.201600268

[40] Sokolović D, Stanković V, Toskić D et al. Monte Carlo-based QSAR modeling of dimeric pyridinium compounds and drug design of new potent acetylcholine esterase inhibitors for potential therapy of myasthenia gravis. Struct Chem 2016; 27: 1511–1519

[41] Veselinović JB, Veselinović AM, Toropova AP et al. The Monte Carlo technique as a tool to predict LOAEL. Eur J Med Chem 2016; 116: 71–75

[42] Veselinović AM, Veselinović JB, Toropov AA et al. In silico prediction of the β-cyclodextrin complexation based on Monte Carlo method. Int J Pharm 2015; 495: 404–409

[43] Toropov AA, Rallo R, Toropova AP. Use of Quasi-SMILES and Monte Carlo Optimization to Develop Quantitative Feature Property/Activity Relationships (QFPR/QFAR) for Nanomaterials. Curr Top Med Chem 2015; 15: 1837–1844

[44] Veselinovic AM, Milosavljevic JB, Toropov AA et al. SMILES-based QSAR model for arylpiperazines as high-affinity 5-HT1A receptor ligands using CORAL. Eur J Pharm Sci 2013; 48: 532–541

[45] Toropov AA, Toropova AP, Benfenati E et al. QSPR/QSAR Analyses by Means of the CORAL Software: Results, Challenges, Perspectives. Quantitative Structure-Activity Relationships. In Drug Design, Predictive Toxicology, and Risk Assessment. Roy K. (eds) Hershey, IGI Global; 2015: 560–585

[46] Nesmerak K, Toropov AA, Toropova AP et al. SMILES-based quantitative structure-property relationships for half-wave potential of N-benzylsalicylthioamides. Eur J Med Chem 2013; 67: 111–114

[47] Toropov AA, Toropova AP, Benfenati E. Additive SMILES-Based Carcinogenicity Models: Probabilistic Principles in the Search for Robust Predictions. Int J Mol Sci 2009; 10: 3106–3127

[48] Kumar P, Kumar A. Monte Carlo Method Based QSAR Studies of Mer Kinase Inhibitors in Compliance with OECD Principles. Drug Res (Stuttg) 2018; 68: 189–195

[49] Veselinović J, Veselinović A, Toropov A et al. Monte Carlo method based QSAR modeling of coumarin derivatives as potent HIV-1 integrase inhibitors and molecular docking studies of selected 4-phenyl hydroxycoumarins. Acta Fac Med Naiss 2014; 31: 95–103

[50] Toropova AP, Toropov AA. CORAL: QSAR models for carcinogenicity of organic compounds for male and female rats. Comput Biol Chem 2018; 72: 26–32

[51] Toropova AP, Toropov AA. CORAL: Binary classifications (active/inactive) for drug-induced liver injury. Toxicol Lett 2017; 268: 51–57

[52] Toropova AP, Toropov AA, Martyanov SE et al. CORAL: QSAR modeling of toxicity of organic chemicals towards Daphnia magna. Chemometr Intell Lab 2012; 110: 177–181

[53] Toropova AP, Toropov AA, Lombardo A et al. CORAL: QSAR Models for Acute Toxicity in Fathead Minnow (Pimephales promelas). J Comput Chem 2012; 33: 1218–1223

[54] Kumar P. A novel, facile, simple and convenient oxidative aromatization of Hantzsch 1, 4-dihydropyridines to pyridines using polymeric iodosobenzene with KBr. Journal of Heterocyclic Chemistry 2010; 47: 1429–1433

[55] Kumar P, Kuamr A, Mohan LJ et al. Heterocyclic Systems Containing Bridgehead Nitrogen Atom: Synthesis and Evaluation of Biological Activity of Imidazo [2, 1-b]-1, 3, 4-thiadiazolo [2, 3-c]-s-triazoles, s-Triazolo [3, 4-b]-1, 3, 4-thiadiazolo [3, 2-b] imidazo [4, 5-b] quinoxaline and bis-(s-Triazolo [3, 4-b]-1, 3, 4-thiadiazolo [3, 2-b] [imidazo [4, 5-b]-cyclohexane]-5a, 6a-diene). Bulletin of the Korean Chemical Society 2010; 31: 3304–3308

[56] Kumar P, Kumar A, Hussain K. Iodobenzene diacetate (IBD) catalyzed an quick oxidative aromatization of Hantzsch-1, 4-dihydropyridines to pyridines under ultrasonic irradiation. Ultrasonics sonochemistry 2012; 19: 729–735

[57] Kumar A, Jain S, Parle M et al. 3-Aryl-1-phenyl-1 H-pyrazole derivatives as new multitarget directed ligands for the treatment of Alzheimer's disease, with acetylcholinesterase and monoamine oxidase inhibitory properties. EXCLI Journal 2013; 12: 1030

[58] Kumar A, Goyal R, Kumar S et al. Estrogenic and anti-Alzheimer's studies of Zingiber officinalis as well as Amomum subulatum Roxb.: the success story of dry techniques. Med Chem Res 2015; 24: 1089–1097

[59] Kumar P, Bhatia R, Khanna R et al. Synthesis of some benzothiazoles by developing a new protocol using urea nitrate as a catalyst and their antimicrobial activities. Journal of Sulfur Chemistry 2017; 38: 585–596. DOI: 10.1080/17415993.2017.1334781

[60] Kumar P, Duhan M, Kadyan K et al. Synthesis of novel inhibitors of α-amylase based on thiazolidine-4-one skeleton containing pyrazole moiety and their configurational studies. MedChemComm 2017; 8: 1468–1476

[61] Kumar P, Kadyan K, Duhan M et al. Design, synthesis, conformational and molecular docking study of some novel acyl hydrazone based molecular hybrids as antimalarial and antimicrobial agents. Chemistry Central Journal 2017; 11: 115. DOI: 10.1186/s13065-017-0344-7

[62] Kumar R, Khanna R, Kumar P et al. Synthesis of Some 4-Quinolinyl Pyridines and their Antimicrobial and Docking Studies. Journal of Heterocyclic Chemistry 2017; 54: 2740–2747

[63] Kumar P, Duhan M, Kadyan K et al. Multicomponent Synthesis of Some Molecular Hybrid Containing Thiazole Pyrazole as Apoptosis Inducer. Drug Res (Stuttg) 2018; 68: 72–79

[64] O'Boyle NM, Banck M, James CA et al. Open Babel: An open chemical toolbox. J Cheminform 2011; 3: 33

[65] Toropov AA, Toropova AP, Benfenati E et al. The definition of the molecular structure for potential anti-malaria agents by the Monte Carlo method. Struct Chem 2013; 24: 1369–1381

[66] Toropova AP, Toropov AA. The index of ideality of correlation: A criterion of predictability of QSAR models for skin permeability? Sci Total Environ 2017; 586: 466–472

[67] Toropov AA, Toropova AP. The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? Mutat Res 2017; 819: 31–37

[68] Toropova AP, Toropov AA. CORAL: Monte Carlo Method to Predict Endpoints for Medical Chemistry. Mini Reviews in Medicinal Chemistry 2018; 18: 382–391

[69] Toropov AA, Carbó-Dorca R, Toropova AP. Index of Ideality of Correlation: new possibilities to validate QSAR: a case study. Struct Chem 2018; 29: 33–38

[70] Toropova AP, Toropov AA, Veselinovic JB et al. QSAR models for HEPT derivates as NNRTI inhibitors based on Monte Carlo method. Eur J Med Chem 2014; 77: 298–305

[71] Toropova AP, Toropov AA, Benfenati E et al. CORAL: quantitative structure-activity relationship models for estimating toxicity of organic compounds in rats. J Comput Chem 2011; 32: 2727–2733

[72] Roy K. On some aspects of validation of predictive quantitative structure-activity relationship models. Expert Opin Drug Discov 2007; 2: 1567–1577

[73] Roy PP, Leonard JT, Roy K. Exploring the impact of size of training sets for the development of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems 2008; 90: 31–42

[74] Golbraikh A, Tropsha A. Beware of q2!. J Mol Graph Model 2002; 20: 269–276

[75] Ojha PK, Roy K. Comparative QSARs for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection. Chemometrics and Intelligent Laboratory Systems 2011; 109: 146–161

[76] OECD. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models; 2014 https://doi.org/10.1787/9789264085442-en.

[77] Roy PP, Roy K. QSAR Studies of CYP2D6 Inhibitor Aryloxypropanolamines Using 2D and 3D Descriptors. Chemical Biology & Drug Design 2009; 73: 442–455

[78] Ojha PK, Mitra I, Das RN et al. Further exploring rm2 metrics for validation of QSPR models. Chemometrics and Intelligent Laboratory Systems. 2011; 107: 194–205