Research in Biotechnology, 5(6): 15-23, 2014 ISSN: 2229-791X
www.researchinbiotechnology.com

Regular Article

Hydrazinecarboxamide or Hydrazinecarbothioamide bearing small molecules as Dual inhibitor of Ras protein and Carbonic anhydrase enzyme as potential anticancer agent – A MLR approach based on docking energy

Supriyo Saha[1]*, Mrityunjoy Acharya[1]

[1]Department of Pharmacy, GRD (PG) IMT, Dehradun (UK), 248009, India
Correspondence author Email id: supriyo9@gmail.com

By using Multiple Linear Regression analysis we have developed two statistical equations using Biological Activity (Dock score) and a set of 1D, 2D descriptors and validated using Golbraikh and Tropsha acceptable model and Euclidean applicability domain analysis. In case of carbonic anhydrase inhibitor best equation is Biological Activity (Dock score)) = 5.53658(+/-0.08083) +0.24208(+/-0.0151) nRing -4.70972(+/-0.02089) nRotB +2.60876(+/-0.02375) phia +0.06039(+/-0.00067) Bac [r^2 :0.99998, r^2 adjusted :0.99993 F :20451.00423 (DF :4, 2)]. In case of Ras protein inhibitor best equation is Biological Activity (Dock score) = -5.93253(+/-0.21871) +0.1608(+/-0.00942) AMR -0.00625(+/-0.00538) CrippenMR -0.02941(+/-0.00181) TopoPSA -0.00468(+/-0.00132) MW [F : 140.88876 (DF :4, 2), Q2 :0.9459, r^2 :0.7257]. In both the cases with minimum diversity shown in between observed and predicted value. So to develop Hydrazinecarboxamide or Hydrazinecarbothioamide bearing small molecules as Dual inhibitor of Ras protein and Carbonic anhydrase enzyme as potential anticancer agent, this set of equation was highly effective.

Keywords: Multiple Linear Regression analysis; Ras protein; Carbonic anhydrase enzyme; Golbraikh and Tropsha acceptable model; Applicability Domain.

Over expression of Ras protein and human carbonic anhydrase cause critical progression of cancer by the conformation changes in between two regions as Switch-I and Switch-II in case of Ras protein mutation. Ras family small GTPases assume two interconverting conformations, "inactive" state 1 and "active" state 2, in their GTP-bound forms. Here, to clarify the mechanism of state transition, we have carried out x-ray crystal structure analyses of a series of mutant H-Ras and M-Ras in complex with guanosine 5'-(beta,gamma-imido)triphosphate (GppNHp), representing various intermediate states of the transition. Crystallization of H-RasT35S-GppNHp enables us to solve the first complete tertiary structure of H-Ras state 1 possessing two surface pockets unseen in the state 2 or H-Ras-GDP structure. Moreover, determination of the two distinct crystal structures of H-RasT35S-GppNHp, showing prominent polysterism in the switch I and switch II regions, reveals a pivotal role of the guanine nucleotide-mediated interaction between the two switch regions and its rearrangement by a nucleotide positional change in the state 2 to state 1 transition (Shima *et al.,* 2010). The molecular docking study for Ras oriented structure was performed using M-ras receptor (derived from *Mus musculus* organism, it is a single chain

peptide contains 183 amino acids), PDB Id: 3KKP procured from protein data bank. As we know the protein carbonic anhydrase family CA IX was associated with cancer progression. CA IX is composed of four domains: it has an N-terminal proteoglycan domain, a CA catalytic domain, a transmembrane region, and a short cytoplasmic tail. In contrast tumors originating from CA IX-positive tissues, such as stomach, tend to have lowered expression of CA IX (Saha *et al.,* 2014). Whereas carbonic anhydrase oriented docking was done by Carbonic anhydrase XII protein (derived from *Homo sapiens*, it is a four identical chain polymer which contains 263 amino acids in each chain) PDB Id: 4KP5, procured from protein data bank. In our previous work by *in silico* way, we developed a series of Hydrazinecarboxamide or Hydrazinecarbothioamide bearing small molecules (R1-R9) as dual inhibitor of Ras protein and carbonic anhydrase enzyme and performed molecular docking study which was validated using cluster analysis with Kobe 2601 as standard, where we concluded that all the designed ligand were in same clustering in case of Carbonic anhydrase enzyme but in case of Ras protein only R2, R4, R8, R9 were in same voxel with standard Kobe 2601 (Yap, 2011). Now we take an another step ahead, to develop a QSAR model using as set of 1D and 2D descriptor generated from Padel Descriptor (Subrata *et al.,* 2014).

Material and Methods

To construct a QSAR model, a set of theoretical and constructive descriptors were calculated by use of PaDEL-descriptor: an open source software (Yap, 2011) and QSAR Model was constructed by use of MLR Plus Validation Tool (Roy *et al.,* 2011). By using PADEL we were calculate 1875 descriptors include Ghose-Cripen Log Ko/w, Ghose-Crippen molar refractivity, Sum of the atomic polarizabilities (including implicit hydrogens), Wildman-Crippen LogP and MR, Wildman-Crippen MR, Eccentric Connectivity Inde A topological descriptor combining distance and adjacency information, H Bond Acceptor Count Number of hydrogen bond acceptors, McGowan characteristic volume, Wiener Polarity Number. All the explanations of relevant descriptors were enlisted in Table 1. A descriptor represents a quantitative property depends on the molecular structure. Theoretical descriptors are advantageous due to its free from uncertainty of experimental measurement and can be calculated for compounds before synthesis. Theoretical descriptors employed in this QSAR study to model as dual inhibitor of Ras protein and Carbonic anhydrase enzyme as potential anticancer agent.

Table 1. List of relevant descriptor with explanation

SN	Abbreviation descriptors	Explanation of descriptors
1.	apol	Sum of the atomic polarizabilities (including implicit hydrogens)
2.	Crippen MR	Crippen's molar refractivity
3.	McGowan Volume	McGowan characteristic volume
4.	VABC	Van der Waals volume calculated
5.	nRing	No of Ring
6.	nRotb	No of Rotatable Bonds
7.	Phia	Kappa flexibility Index
8.	Bac	Balaban Centric Index
9.	AlogP	Ghose-Crippen LogKow
10.	Crippen LogP	Crippen's LogP
11.	XLogP	XLogP
12.	AMR	Molar refractivity
13.	TopoPSA	Topological Polar Surface Area
14.	Wpol	Weiner polarity number
15.	MW	Molecular Weight

The chemometric tool

For the development of QSAR equation two methods were implemented; (1) Stepwise regression (2) multiple linear regressions with factor analysis as pre processing factor analysis for variable selection (FA-MLR).

Stepwise regression

Multi step linear equation, a multistep equation was built by step by step. The basic procedure involved: (i) identifying an initial model (ii) repeating the previous step by altering descriptor or variable combination to achieve better f and r2 value. (iii) calibrate the equation by justify the values in between observed and predicted values (Roy *et al.,* 2014).
The stepwise MLR was performed using statistical software SPSS and it was judged by parameters as explained variance (r^2a), correlation coefficient (r), standard error of estimate (s) and variance ratio (F) at a specified degree of freedom (DF). All accepted mlr equation had regression level significant at 95 and 99% levels. the generated qsar equation was validated by leave one out or loo method using minitab software and different parameters like cross validation r^2 (q^2), standard deviation based on press (S_{PRESS}) and standard deviation of error of prediction (SDEP) (Ambure *et al.,* 2014*;* Ambure *et al.,* 2013*;* Martin *et al.,* 2012*;* Koutsoukas *et al.,* 2014).

FA-MLR

In this case a final statistical tool was used to develop a QSAR relation, factor analysis as a data pre processing step to identify the important factor to identify the important variables contributing the response variable by avoiding colinearities value. The data matrix is first standardized and correlation matrix and subsequently reduced correlation matrix. An eigen value problem is then solved and the factor pattern can be obtained from the corresponding eigen vectors. The main objectives are to display multidimensional data in space of lower dimensionality with minimum loss of information (explaining > 95% of variance of data matrix) and to extract the basic features behind the data with ultimate goal of interpretation (Dalai *et al.,* 2006).

Results

Developed QSAR equation against Ras Protein and Carbonic anhydrase enzyme

In the case of Ras protein inhibition the generated three equations were as

Biological Activity (Dock score)) = -6.61703(+/-0.20959) +0.15383(+/-0.01137) AMR -0.01286(+/-0.1212) nRing -0.02296(+/-0.00233) TopoPSA -0.05856(+/-0.01106) WPOL…… **Equation: 1**

F : 91.00721 (DF :4, 2), Q2 :0.69289, r^2 :0.98274

Biological Activity (Dock score)) = -6.471(+/-0.40715) +0.15346(+/-0.00943) AMR -0.00364(+/-0.00865) CrippenMR -0.02263(+/-0.00237) TopoPSA -0.05453(+/-0.01375) WPOL……**Equation: 2**

F : 98.54278 (DF :4, 2), Q2 :0.73137, r^2 :0.97941

Biological Activity (Dock score)) = -5.93253(+/-0.21871) +0.1608(+/-0.00942) AMR -0.00625(+/-0.00538) CrippenMR -0.02941(+/-0.00181) TopoPSA -0.00468(+/-0.00132) MW……..**Equation:3**

F : 140.88876 (DF :4, 2), Q2 :0.9459, r^2 :0.7257

Among the Equation:1-3, Equation : 3 was showing the best conformer based upon its Ypred and Yobs value

In the case of carbonic anhydrase enzyme activity, the generated equations were:
Biological Activity (Dock score)) = -7.87717(+/-7.80136) -0.65618(+/-1.20125) apol -0.7098(+/-0.40227) CrippenMR +66.27327(+/-78.92457) McGowan Volume -0.20109(+/-0.30684) VABC……**Equation: 1**
r2 value 1.0

Biological Activity (Dock score)) = 5.53658(+/-0.08083) +0.24208(+/-0.0151) nRing -4.70972(+/-0.02089) nRotB +2.60876(+/-0.02375) phia +0.06039(+/-0.00067) Bac……....**Equation: 2**
r2 value 0.99998

Biological Activity (Dock score)) = 6.14566(+/-3.71428) +1.13452(+/-0.38947) ALogP +1.84788(+/-0.36869) CrippenLogP -2.80718(+/-0.83199) nRotB -0.78859(+/-0.15731) XLogP……..**Equation: 3**
r2 value 0.99998

Biological Activity (Dock score)) = 5.82767(+/-0.09732) +0.01599(+/-0.00108) CrippenMR -4.78986(+/-0.02498) nRotB +2.46711(+/-0.0301) phia +0.06222(+/-0.00081) Bac……...**Equation:4**
r2 value 0.99997

Biological Activity (Dock score)) = 5.56053(+/-0.14225) +0.61115(+/-0.06703) McGowan Volume -4.67135(+/-0.0353) nRotB +2.40721(+/-0.05262) phia +0.06018(+/-0.00117) Bac.....**Equation: 5**
r2 value 0.99993

Validation of derived QSAR equation
In case of Ras protein inhibitor:
Validation and cross validation parameter below

Internal Validation Parameters	Leave-One-Out(LOO) Result
SEE :0.07554	Q2 :0.9459
r^2 :0.99646	PRESS :0.17456
r^2 adjusted :0.98939	SDEP :0.15791
F :140.88876 (DF :4, 2)	

The equation: 3 was further validated using Golbraikh and Tropsha acceptable model criteria's :

Q^2	0.9459	Passed	(Threshold value Q^2>0.5)
r^2	0.7257	Passed	(Threshold value r^2>0.6)

Unless the r2 and Q2 value the statistical F value was 140.88876 which correlate with the integrity of the equation.
From the train and test subset, the applicability domain was tested using Euclidean program which showed that only compound R7 showed maximum diversity and no compounds were outside the applicability domain. Results are showed in Table 3.

Some External Validation Parameters (After Scaling)	Golbraikh and Tropsha acceptable model criteria's If Q^2>0.5, then check whether it follows other three condition

rm^2 : -0.00002	r^2=0.00007 Failed (Threshold value r^2>0.6)
reverse rm^2 : 0.00003	|r0^2-r'0^2|=0.10451Passed (Threshold value |r0^2-
average rm^2 :0.00001	r'0^2|<0.3)
delta rm^2 :0.0	K=0.91968, [(r^2-r0^2)/r^2] = Passed
	(Threshold value: [0.85<k<1.15)

This data from External Validation Parameters was failed because due to data of R1 was out layered, so is the data of R1 was remove from the external validation then the results were below

Some External Validation Parameters (After Scaling)	Golbraikh and Tropsha acceptable model criteria's If Q^2>0.5, then check whether it6 follows other three condition
rm^2 : 0.4687	r^2=0.62478 Passed (Threshold value r^2>0.6)
reverse rm^2 : 0.53727	|r0^2-r'0^2|=0.0698 Passed (Threshold value |r0^2-
average rm^2 :0.50298	r'0^2|<0.3)
delta rm^2 :0.06857	K=0.97489 [(r^2-r0^2)/r^2] = Passed
	(Threshold value: [0.85<k<1.15)

In case of Carbonic anhydrase inhibitor:
Validation and cross validation parameter below

Internal Validation Parameters	Leave-One-Out(LOO) Result
SEE :0.01407, r^2 :0.99998,	Q2 :0.99899, PRESS :0.01635,
r^2 adjusted :0.99993	SDEP:0.04833
F :20451.00423 (DF :4, 2)	

The **equation 2** was further validated using Golbraikh and Tropsha acceptable model criteria's :

Q^2	0.99899Passed	(Threshold value Q^2>0.5)
r^2	0.72605Passed	(Threshold value r^2>0.6)
|r0^2-r'0^2|	0.07521Passed	(Threshold value |r0^2-r'0^2|<0.3)

Unless the r2 and Q2 value the statistical F value was 20451.00423 which correlate with the integrity of the equation.
From the train and test subset, the applicability domain was tested using Euclidean program which showed that only compound R2 showed maximum diversity and no compounds were outside the applicability domain. Results are showed in Table 2.
Further, by using xternal validation parameter with training set activity value 5.5 create a correlation using
r^2 :0.91325, r0^2 :0.90814, reverse r0^2:0.91298, rm^2 :0.84796, reverse rm^2 :0.89805, average rm^2 :0.87301,delta rm^2 :0.0501, rmsep:0.45278, rpred^2 :0.99897, Q2f1 :0.99897, Q2f2 :0.90395.

Some External Validation Parameters (After Scaling)	Golbraikh and Tropsha acceptable model criteria's If Q^2>0.5, then check whether it follows other three condition
rm^2 :0.82518	r^2=0.91325 Passed (Threshold value r^2>0.6)
reverse rm^2 :0.89602	|r0^2-r'0^2|=0.00483Passed (Threshold value |r0^2-
average rm^2 :0.8606	r'0^2|<0.3)
delta rm^2 :0.07084	K=0.98914, [(r^2-r0^2)/r^2] = Passed
	(Threshold value: [0.85<k<1.15]

Table 2. Comparison data of Yobs and Ypred results in between MLR equation against Carbonic Anhydrase enzyme and Ras Protein

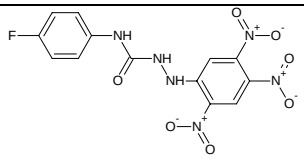
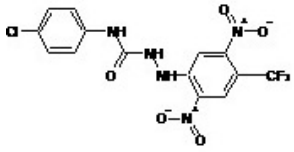
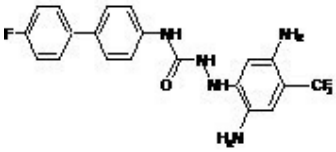
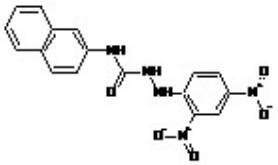
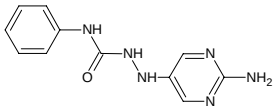
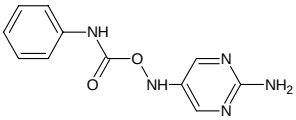
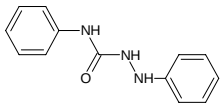
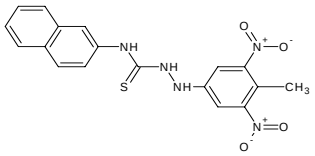
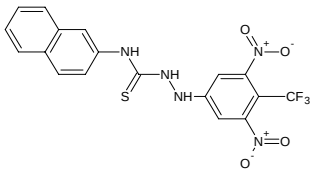
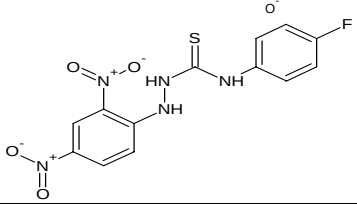
S N	Structure Code	Structure	Carbonic Anhydrase enzyme		Ras Protein	
			Yobs	Ypred	Yobs	Ypred
1.	R1		-9.00	-10.29	-5.76	-9.54
2.	R2		-9.11	-9.12	-7.32	-7.29
3.	R3		-7.27	-7.27	-7.01	-7.02
4.	R4		-10.87	-10.86	-9.04	-9.04
5.	R5		-6.5	-6.62	-6.99	-7.06
6.	R6		-6.62	-6.62	-6.6	-7.92
7.	R7		-6.98	-6.98	-6.94	-6.88
8.	R8		-9.42	-8.79	-7.54	-7.51
9.	R9		-10.03	-10.04	-7.59	-7.61
10.	Kobe 2061		-9.24	-9.24	-7.55	-7.9

Table 3. Applicability Domain Analysis of Carbonic Anhydrase enzyme and Ras Protein

Carbonic Anhydrase Enzyme		Ras Protein	
Training Set	**Test Set**	**Training Set**	**Test Set**
1	0.48779	0.02191	0.26105
0.06088	0.52579	0.09296	0.77305
0.06726	0.00571	0	0.15686
0.52579		0.7653	
0.61717		1	
0.48185		0.01277	
0		0.3183	

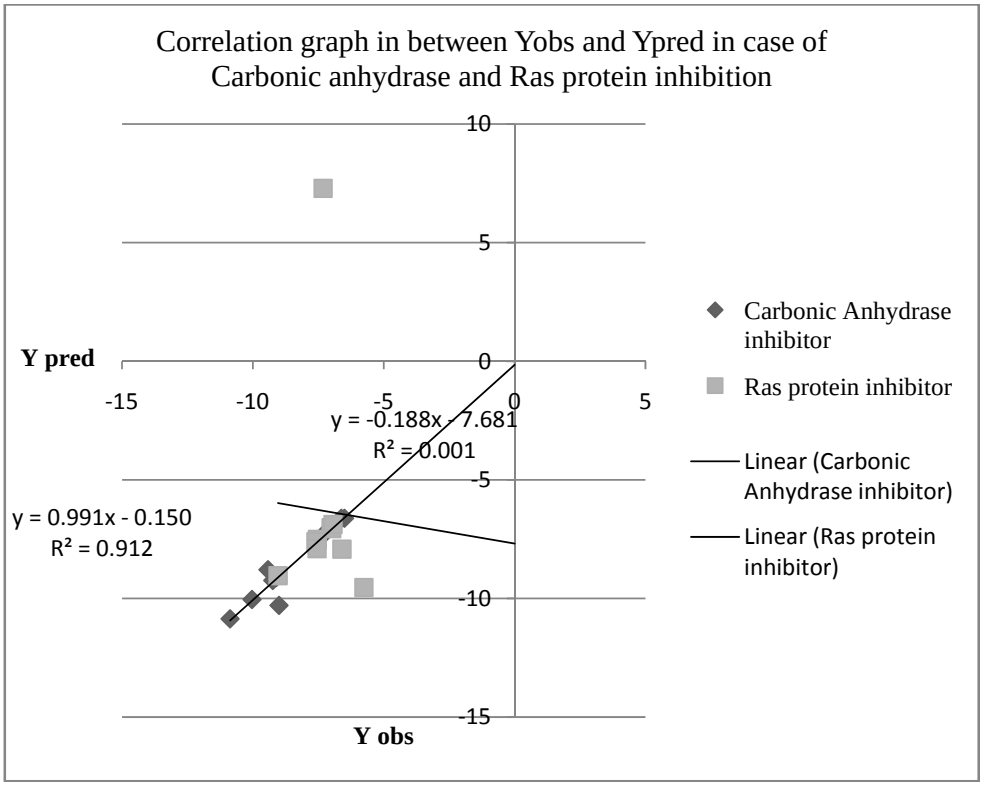


Fig.1. Correlation graph in between Yobs and Ypred in case of Carbonic anhydrase and Ras protein inhibition

Discussion

In case of carbonic anhydrase inhibitor best equation is **Equation:2**, where the phia parameter shows parabolic relation in response to Biological Activity (Dock score), it means if we increase the kappa flexibility index value up to certain limit the Biological Activity (Dock score) parameter will increase after that it will decrease just maintain a bell shaped curve and subsequently if we minimize the number of rotatable bonds the Biological Activity (Dock score)) will be increase. Where phia represent attributes of molecular shape and are derived from counts of one-bond, two-bond, and three bond fragments. Phia indices measure structural properties that restrict a molecule's flexibility (**Table1**). In case of Ras protein inhibitor best equation is **Equation:3** where the AMR parameter shows a parabolic relation with Biological Activity (Dock score), it means if we increase the value of molar

refractivity up to a certain limit the value increase by maintaining a bell shaped curve and subsequently minimizing the topological polar surface area the Biological Activity (Dock score) will increase. Where AMR index contain a representation of the dispersive forces involved in the compound macromolecule interaction; and also topological information about the chemical structure of the ligand (**Table 1**). Then calculate the Yobs and Ypred value from the equation evolved and draw a correlation equation (Fig.1). From the correlation data it was clearly reveals that in case of carbonic anhydrase inhibitor development, equation is highly correlated with r2 value 0.912 and in case of development of Ras protein inhibitor the r2 value is 0.001 which was not good due to out layered value of R1 and R6.

Conclusion

This total work is an attempt to deliver a platform for the new research to develop a better Hydrazine carboxamide or Hydrazinecarbothioamide bearing small molecules as dual inhibitor of Ras protein and Carbonic anhydrase enzyme as potential anticancer agent. In future for the development of dual inhibitors of Ras and carbonic anhydrase AMR and phia is the most important physicochemical and topological parameter has to be taken under consideration.

Acknowledgement

We are highly thankful to Mr. Pravin Ambure, Research Scholar Jadavpur University.

References

Ambure P, Roy K. Advances in quantitative structure--activity relationship models of anti-Alzheimer's agents. 2014. Expert Opin Drug Discov. 9: 697-72.

Ambure P, Roy K. Exploring structural requirements of leads for improving activity and selectivity against CDK5/p25 in Alzheimer's disease: An in silico approach. 2013. RSC Advances. 4**:** 6702-6709.

Dalai KM, Leonard TJ, Roy K. Exploring QSAR of peripheral benzodiazepine receptor binding affinity of 2-phenylpyrazolo [1,5-a] pyrimidin-3-yl-acetamides using topological and physicochemical descriptors. 2006. I. J. Chem. 45B: 2497-2505.

Koutsoukas A, Paricharak S, Galloway WRJD, Spring DR, IJzerman AP, Glen RC, Marcus D, Bender A. How Diverse Are Diversity Assessment Methods? A Comparative Analysis and Benchmarking of Molecular Descriptor Space. 2014. J. Chem Inf Model. 54 (1): 230–42.

Martin TM, Harten P, Young DM, Muratov EN, Golbraikh A, Zhu H, Tropsha A. Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling?. 2012. J. Chem Inf Model. 52 (10): 2570–78.

Roy K, Kar S. The rm2 metrics and regression through origin approach: reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?') 2014. Eur. J Pharm Sci. 62: 111–114.

Roy K, Mitra I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. 2011. Comb Chem High Throughput Screen. 14(6): 450-74.

Saha S, Acharya M. Discovery of Hydrazinecarboxamide or Hydrazinecarbothioamide Bearing Small Molecules as Dual Inhibitor of Ras Protein and Carbonic Anhydrase

Enzyme as Potential Anticancer Agent Using Validated Docking Study and In-silico ADMET Profile. 2014. Res. J Pharm Biol Chem Sci. 5(3): 1884-93.

Shima F, Ijiri Y, Muraoka S, Liao J, Ye M, Araki M, Matsumoto K, Yamamoto N, Sugimoto T, Yoshikawa Y, Kumasaka T, Yamamoto M, Tamura A, Kataoka T. Structural basis for conformational dynamics of GTP-bound Ras protein. 2010. J. Biol Chem. 285(29): 22696-705.

Subrata P, Kunal R. Modeling bioconcentration factor (BCF) using mechanistically interpretable descriptors computed from open source tool "PaDEL-Descriptor". 2014. Environmental Science and Pollution Research. 21(4): 2955-2965.

Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. 2011. J. Comput Chem. 32(7): 1466-74.