BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 7(8)

2 D - QSAR studies on CYP26A1 inhibitory activity of 1-[benzofuran-2-yl-(4-alkyl/aryl-phenyl)-methyl]-1 H-triazoles

Madhu Yadav

Department of Computational Biology & Bioinformatics, Jacob School of Biotechnology & Bioengineering, Sam Higgonbottom Institute of Agriculture, Technology & Sciences, (Formerly Allahabad Agricultural Institute), Allahabad – 211007, Uttar Pradesh, India; Madhu Yadav - Email : madhuyadav2003@gmail.com.

Received November 19, 2011; Accepted December 05, 2011; Published December 21, 2011

Abstract:
The Quantitative Structure Activity Relationship (QSAR) study is performed over a set of 15, 4-alkyl/aryl-substituted 1-[benzofuran-2-yl-phenylmethyl]-1 H-triazoles derivatives. This study is based on the application of physicochemical parameters in QSAR. The parameters include (MR (molar refractivity), MW (molecular weight), Pc (parachor), St (surface tension), D (density), Ir (index of refraction) and log P (partition coefficient). The parameters describing physiochemical properties are used as independent variables and the biological activity (IC_{50}) is considered as dependent variable in multiple regression analysis. Different models were generated with high co-efficient of determination (R^2). The 2D-QSAR study identified compounds capable of inhibiting the metabolic breakdown of the retinoid (trans-retinoic acid (ATRA)) involved in the activation of specific nuclear Retinoic acid receptors (RARs). This study identifies R115866 as a potential inhibitor of the cytochrome P450 (CYP) mediated metabolism with increased RA levels for retinoid actions.

Keywords: Physiochemical property, biological activity, QSAR, multiple linear regression

Background:
In human, CYP26A1 mapped to chromosomes 10q23-q24 is expressed in the liver, heart, pituitary gland, adrenal gland, testis, brain and placenta. It is thought that the principle role of CYP26A1 is homeostatic for the regulation of intercellular ATRA (Anti Trans Retinoic Acid) steady state levels via a negative feedback loop (similar to CYP26A1 cholecalciferol metabolism). The enzyme is an important regulator of differentiation and a possible modulator of disease states in indirectly controlling ATRA and other retinoid concentrations.

Retinoic acid has been used in a number of clinical situations, especially oncology and dermatology. In Oncology ATRA has shown spectacular success in the treatment of acute promyelocytic leukaemia. Although other CYPs are also induced by ATRA, it is thought that CYP26A1 is likely to be the most important enzyme involved in its degradation **[1]**. ATRA regulates epithelial differentiation and growth through activation of specific nuclear retinoic acid receptors (RARs). We identified compounds capable of inhibiting the metabolic breakdown of the retinoid because high rate metabolism largely impairs the biological efficiency of RA.

QSAR studies are of importance in molecular biochemistry. It is essential that appropriate descriptors are employed, whether they are theoretical, empirical or derived from available experimental characteristics of structure to obtain significant correlation. QSAR represents an attempt to correlate structural or property descriptors of compound with activities. These physiochemical descriptors include parameters to account for hydrophobicity, topology, electronic parameters and steric effects. These are determined empirically by computational methods. Quantitative Structure Activity Relationship (QSAR) is currently being applied in many disciplines pertaining to drug design and environmental risk assessments **[2].**

An insight to structure activity relationship is established using molecular descriptors that effectively characterize molecular size, molecular branching or the variations in molecular shapes affecting the structure and its molecular activity. Structural invariants derived from molecular structure are important for the understanding of a particular structure/property activity relationship. Several graph theoretical invariants have been generalized to produce structure dependent descriptors. It should be noted that most biological activities are dominated by molecular size in such generalization.

The activities and properties are related using the general mathematical function, F: Biological activity = f [structure (physicochemical descriptors as structural parameters)]. Biological activity is often defined as a measure of log 1/C, ki, IC_{50}, EC_{50}, log k and Km as described elsewhere **[3-5].** The relationship is often not a mathematical expression derived by statistical or related techniques. The parameters describe structural and physiochemical properties as independent variables and the biological activities as dependent variable for usage in multiple linear regression (MLR) analysis. We describe the QSAR of 15 triazoles derivatives with relationship between structural and physiochemical parameters to known biological activity log IC_{50} values.

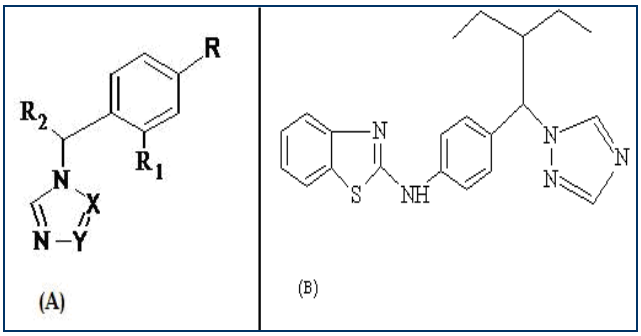


Figure 1 (A): Parent structure of 1-[benzofuran-2-yl-(4-alkyl/aryl-phenyl)-methyl]-1H-triazoles; **(B):** 2D structure of R115866 model compound.

Methodology:

The physiochemical properties used in the study are molar refractivity, molar volume, parachor, index of refraction, surface tension and polarizibility to co-relate with biological activity (log IC_{50} inhibitory concentration). The physiochemical properties were estimated using Chemsketch 5.0 (from ACD Lab) and multiple linear regressions are done by SPSS Software **[4, 6]**. The physicochemical parameters used in the study are described below.

Density (D):

Density is a steric parameter calculated by ACD Lab Chem Sketch Software. This parameter is related to the bulk and size of the substituent.

Molecular weight (MW):

Molecular weight descriptor has been used as a descriptor in systems such as transport studies where diffusion is the mode of operation. It is an important variable in QSAR studies pertaining to cross resistance of various drugs in multi-drug resistant cell lines.

Molar refractivity (MR):

Molar refractivity is generally considered to be a measure of overall bulkiness and is related to London dispersion forces using MR=4πNα/3; where N is Avogadro number and α is the polarizability of the molecule. It gives no information about shape and is generally scaled 0.1 and has been extensively used in QSAR.

Parachor (Pc):

This parameter is calculated by ACD Lab Chem Sketch Software (Chem Sketch). The Parachor may be defined as the molar volume of a liquid at a temperature at which its surface tension is unity. Parachor is obtained experimentally using Pc= (MW/d) γ ¼ where, Pc is a molar volume term that is corrected by surface tension (γ) rise to the power 1/4.

Indicator variables:

These are not QSAR parameters but are used to indicate the significance of any particular group or species at a particular substitution site in a given series of drugs.

Discussion:

QSAR was performed on a series of 15, 4-alkyl/aryl-substituted 1-[benzofuran-2-yl-phenylmethyl]-1 H-triazoles derivatives with the physicochemical parameters **(Table 3, see supplementary material),** the biological activity (pIC_{50}) is a measure of inhibitory activity indicators **(Table 1, see supplementary material)**. We have used Hansch analysis **[8]** for developing these models. The QSAR multiple regression analyses were performed with SPSS (version 14.0) package. Their activity data and the physicochemical parameters evaluated in the correlation are listed in **(Table 3, supplementary material).** Linear regression analysis was performed on present series of triazoles derivaties according to physiochemical properties given in **(**Table **2, See supplementary material)**. These parameters were found to be useful in QSAR based drug modeling **(please see supplementary material for modeling equations 1 to 7).** The effect of substituent indicator parameters $-I_6$ (CH2CH2CH2), + I_{20} (CH-(CH2CH3)2) at R **(see Figure 1a)**, IR_2 position was studied (see Table 2, supplementary material) in equations 2 to 7.

Analysis through univariate correlation produced very low correlation coefficient (see **equation 1** in **supplementary material**). Bivariate analysis gave a slightly improved value. Tetra and penta variate analysis results are encouraging with high correlation coefficient. The high value of F-test ratio and low value of standard error of estimation support reports described elsewhere **[7, 8].**

We estimated the log IC_{50} values with the observed ones **(Table 4, See supplementary material)** for comparison. Such correlations are graphically represented in **Figure 2.** In bivariate analysis, correlation coefficient are little higher. However, it is not sufficient to describe the structure activity relationship in a quantitative manner. Correlation coefficients show best result obtained using St (surface tension), D (density), Pz (polarizability) with I6 ($CH_2CH_2CH_3$) and I20 (CH-(CH2CH3)2). The mathematical model obtained from physiochemical properties are given in equations 5 to 7 (see supplementary material). Results are compared with observed

values of log IC_{50} **(Table 4, See supplementary material)**. Substitution effects are shown using the indicator parameters.

Equations suggest that the positive correlation coefficient of structural parameters D, St, IOR show a direct relationship with biological activity log IC_{50}. Parameters like Pz have negative correlation coefficient with inverse relationship for biological activity log IC_{50}. The positive sign for parameter D, St and IOR explains that as the bulkiness of the compound increases the potency of the drug likeliness to increase. The negative sign of the regression coefficient of Pz means less electronegative substituent for increased activity **[9].**

The equation 4 (supplementary material) suggests that the positive correlation coefficient of structural parameters Index of Refractivity (IOR) shows direct relationship with biological activity logIC_{50}. The positive correlation coefficient of Indicator IR_2 for substituent 15 shows positive impact on the biological activity quantitatively. The negative (-ve) correlation coefficient of indicator IR $CH_2CH_2CH_3$ at position 3 show inverse relationship with biological activity.

Predicted and residual values for the best model equation 4 are given in **(Table 4, See supplementary material)**. In this equation the F ratio value is much higher than the theoretical F value [$F_{3,11}$=33.052] indicating the statistical significance of this model equation. Predicted values are the calculated activities from the equation and the residual values are the difference between the observed biological activities and the calculated activities.

Conclusion:

The study shows that the biological activity log IC_{50} is structure specific in nature for a particular series of derivatives. Model equations suggest that the presence of $CH_2CH_2CH_3$ at R position show negative and substituent 15 at R_2 position have positive impact on the biological activity. Thus, compound R115866 shows the best correlation for further consideration. The name and molecular structure of compound is R115866 – N- [4-[2- ethyl – 1- (1H-1, 2, 4-triazol-1-yl) butyl] phenyl]-2-benzothiazolamine **(Figure 1b).**

References:

[1] Kim KH, J *Comput.-Aided Mol Des*. 2001 **15**: 649 [PMID: 11688945]
[2] Hansch CA, *Acct Chem Res*. 1969 **2**: 232
[3] Lipnick RL, *Trends Pharmacol. Scci*. 1986 **7**: 161
[4] Hatheway GJ *et al*. *J Med. Chem*. 1978 **21**: 563 [PMID: 671453]
[5] DesJarlais RL *et al. J Med Chem*. 1988 **31**: 722 [PMID: 3127588]
[6] Borman S. *Chem Eng News*. 1990 **68**: 20
[7] Hansch C *et al*. *American Chemical Society Washington*. 1995
[8] Marill J *et al*. *Curr Drug Metab*. 2003 **4**: 1[PMID: 12570742]
[9] www.netsci.org

Edited by P Kangueane
Citation: Yadav Bioinformation 7(8): 388-392 (2011)

BIOINFORMATION

open access

Supplementary material:

Model Equations:

Equation	Number	Observations	Inference
pIC_{50}= 14.057(±22.206) IOR- (18.189)	→ (1)	n=15, r=0.335, r^2 = 0.126, RA^2 = -0.059, Se = 0.923806, F = 1.870.	poor
pIC_{50} = 5.651 (±8.896) IOR + 3.296 (±0.827)I_{20} - (4.477)	→ (2)	n=15, r=0.938, r^2 = 0.880, RA^2= 0.860, Se= 0.356466, F=43.936	better
pIC_{50} = 12.609 (±23.565) IOR - (0.603) (±2.192)I_6 - (15.750)	→ (3)	n=15, r=0.389, r^2 = 0.058, RA^2= 0.010, Se= 0.947465, F=1.068.	QSAR
pIC_{50} = 4.383(±8.757) IOR + 3.286 (±0.797)I_{20} -(0.539)(±0.794) I_6 - (2.339)	→ (4)	n=15, r=0.949, r^2 = 0.890, RA^2= 0.860, Se= 0.339419; F=33.052.	good
pIC_{50} = 3.443E-02 (±0.0604) St + 3.319 (±0.771)I_{20}- (0.543)(±0.777) I_6 + 3.271	→ (5)	n=15, r=0.950, r^2 = 0.903, RA^2= 0.877, Se= 0.334404, F=34.161	good
pIC_{50} = 0.754 (±3.141) D + 3.364 (±0.809)I_{20} (0.584)(±0.813) I_6 + 5.117	→ (6)	n=15, r=0.944, r^2 = 0.892, RA^2= 0.862, Se= 0.353196, F=30.243	good
pIC_{50} = -1.69E-02 (±0.0641) Pz + 3.450 (±0.852)I_{20}- (0.663)(±0.819)I_6+5.607	→ (7)	n=15, r=0.945, r^2 = 0.892, RA^2= 0.863, Se= 0.352333, F=30.409	good

IOR = Index of Refraction; I = indicator parameters; n = number of data paint; r = Coefficient of regression; R^2A = Explained Variance; r^2 = Coefficient of determination; Se= Standard error of estimation; F = Ratio of mean square for regression to mean square for residual.

Table 1: Biological activity, physiochemical and Indicator parameters of compounds

S.No.	R	R_1	R_2	X	Y	pIC_{50}(μM)	IR	IR_2
1	CH_2CH_3	H		N	CH	5.346787	0	0
2	CH_2CH_3	H		CH	**N**	5.30103	0	0
3	$CH_2CH_2CH_3$	H		N	**CH**	4.30103	0	0
4	$CH(CH_3)_2$	H		N	CH	4.69897	0	0
5	$CH(CH_3)_2$	H		CH	N	4.30103	0	0
6	$C(CH_3)_3$	H		N	CH	5	1	0
7	$CH_2CH(CH_3)_2$	H		N	CH	4.39794	0	0
8	cyclohexyl	H		N	CH	5.30103	0	0
9	C_6H_5	H		N	CH	5.154902	0	0
10	C_6H_5	H		CH	N	5.045757	0	0
11	p-Cl-C_6H_5	H		N	CH	4.69897	0	0

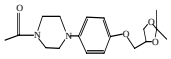
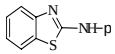
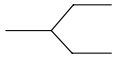
12	p-Cl-C$_6$H$_5$	H		CH	N	4.69897	0	0
13	p-Cl-C$_6$H$_5$	Cl		CH	CH	4.920819	0	0
14	p-H-C$_6$H$_5$	H		CH	CH	5.154902	0	0
15		H		N	CH	8.30193	0	1

Table 2: Physicochemical parameters of 15 triazoles derivatives

comp	Mol. weight	Molar refractivity	Molecular volume	Parachor	Index of Refraction	Surface Tension	density	Polarizability
1	303.357	91.18	250.2	654	1.649	46.6	46.6	36.14
2	303.357	91.18	250.2	654	1.649	46.6	46.6	36.14
3	317.384	95.79	266.2	692.6	1.638	45.7	45.7	37.97
4	317.384	95.6	265.4	685.1	1.639	44.4	44.4	37.9
5	317.384	95.6	265.4	685.1	1.639	44.4	44.4	37.9
6	331.41	100.56	285.9	731.1	1.62	42.7	42.7	39.86
7	331.41	100.21	281.4	723.7	1.63	43.7	43.7	39.72
8	357.448	107.25	286.8	761.2	1.67	49.6	49.6	42.51
9	351.4	107.25	286.8	761.2	1.67	49.6	49.6	42.51
10	351.4	107.25	286.8	761.2	1.67	49.6	49.6	42.51
11	385.845	111.85	296.1	790	1.679	50.6	50.6	44.34
12	385.845	111.85	296.1	790	1.679	50.6	50.6	44.34
13	531.437	138.58	379.6	1019.7	1.65	52	52	35.5
14	308.76	88.41	226.1	613.7	1.709	53.9	53.9	35.05
15	377.505	112.74	298.8	793.6	1.678	49.7	49.7	44.69

Table 3: Auto–correlation Matrix demonstrating physicochemical parameters with indicators

	MW	MR	MV	Pc	Ior	St	Density	Pz	I_6	I_{20}
Activity	0.1108508	0.173446	0.090809	0.122906	0.35377	0.277093	0.16759	0.3319	-0.234	0.928
MW	1	0.975411	0.951138	0.977554	0.13529	0.502336	0.6168282	0.155	-0.164	0.126
MR		1	0.975435	0.993565	0.11925	0.452125	0.5079479	0.34	-0.173	0.198
MV			1	0.991838	-0.10103	0.258493	0.350012	0.2563	-0.124	0.141
Pc				1	0.02005	0.37907	0.4564434	0.2643	-0.141	0.153
Ior					1	0.899101	0.7580828	0.3087	-0.232	0.234
St						1	0.9198325	0.1731	-0.189	0.142
Density							1	-0.016	-0.204	0.093
Pz								1	-0.148	0.393
I_6									1	-0.071
I_{20}										1

Table 4: Comparison between observed, predicted and their residual values

Observed values	Predicted values	Residual values
5.34679	4.88437	0.46242
5.30102	4.88437	0.41663
4.301	4.301	0
4.69897	4.83956	-0.14066
4.30102	4.83956	-0.53854
5	4.75442	0.24558
4.39794	4.79923	-0.40129
5.30102	4.97847	0.32255
5.15490	4.97847	0.17643
5.04575	4.97847	0.06653
4.69897	5.0188	-0.3208
4.69897	5.0188	-0.3208
4.92081	4.88885	0.03115
5.15490	5.15323	0.00077
8.301	8.301	0