

An *In Silico* QSAR Model Study Using Electrophilicity as a Possible Descriptor Against *T. Brucei*

Ranita Pal, Indian Institute of Technology, Kharagpur, India

Goutam Pal, Indian Institute of Technology, Kharagpur, India

Gourhari Jana, Indian Institute of Technology, Kharagpur, India

Pratim Kumar Chattaraj, Indian Institute of Technology, Kharagpur, India

ABSTRACT

Human African trypanosomiasis (HAT) is a vector-borne sleeping sickness parasitic disease spread through the bite of infected tsetse flies (*Glossina* genus), which is highly populated in rural Africa. The present study constructed quantitative structure-activity relationship (QSAR) models based on quantum chemical electronic descriptors to bring out the extent to which the electronic factor of the selected compounds affects the HAT activity. Theoretical prediction of toxicity (pIC_{50}) of the series of heterocyclic scaffolds consisting 32 pyridyl benzamide derivatives towards HAT is investigated by considering all possible combinations of electrophilicity index (ω) and the square of electrophilicity index (ω^2) as descriptors in the studied models along with other descriptors previously used by Masand et al. A multiple linear regression (MLR) analysis is conducted to develop the models. Further, in order to obtain the variable selection on the overall data set having diverse functional groups, the analysis using sum of ranking differences methodology with ties is carried out.

KEYWORDS

Electrophilicity, Global Electronic Descriptor, Human African Trypanosomiasis, Multiple Linear Regression, QSAR, Sum of Ranking Differences

INTRODUCTION

Human African trypanosomiasis, commonly known as sleeping sickness is an endemic disease in the sub-Saharan African countries caused by two forms of the *Trypanosoma brucei* parasite, namely, *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense* (Sykes et al., 2012; Ferrins et al., 2014). According to the World Health Organisation, 98% of the reported cases are caused by the former species resulting in a more severe form of the disease-causing chronic infection, the symptoms of which mainly remain hidden until the central nervous system is seriously affected. *Trypanosoma brucei rhodesiense*, on the other hand, accounts for the remaining 2%, causing acute infection and showing symptoms in the early stages of the disease. This vector-borne disease is known to be

DOI: 10.4018/IJCCCE.20190701.oa1

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

transmitted by the bite of the infected tsetse fly and is very rarely identified in its first stage where drug administration is easier and less toxic compared to those required in the most commonly identifiable second stage (Barrett et al., 2011; Simarro et al., 2012; Seixas et al., 2013; Carvalho et al., 2014).

High-throughput screening *i.e.*, Alamar Blue based, 384-well HTS assay was applied by Sykes *et al.* (2012) to screen a huge dataset of 87,296 compounds against *T. brucei* where unfortunately pyridyl benzamide derivatives were excluded by the selection criteria. Later, Ferrins *et al.* (2014) showed the HAT activity of the aforementioned class of compounds suggesting their ability to treat the disease in its second stage. However, the search for experimental targets for such studies still possesses substantial challenges. This is where computational drug design and statistical approaches like quantitative structure-activity relationships (QSARs) come in handy as computational resources are easy to work with, cheaper and faster than conducting actual experiments. QSAR provides a statistical approach in understanding the correlation between structural features of a certain class of compounds and the biological behaviour exhibited by them.

Quantitative structure-activity relationships (QSARs) are of paramount importance which pay attention to the theoretical toxicity predictions in the highly complicated field of pharmaceutical sciences due to their potential in the assessment of various biological activities (e.g., drug activity, toxicity, etc.) and physicochemical features of bioactive molecules by evading the time consuming and cost-effective experiments. It is a very interesting aspect that molecular descriptors or chemical attributes are the fingerprints of the molecular or chemical structures and thus the choice of appropriate descriptors for a particular biological property/activity/toxicity has become quite a challenging task. Cheminformatics is the initial step in such studies that efficiently utilizes computer information techniques in the field of chemistry to search, extract and process relevant information of chemical compounds from huge databases. In cheminformatics analysis, machine learning techniques are also used in QSAR studies for chemical feature extraction and selection in multiple levels, characterization of the compounds by substructure fragments as well as a selection of the chemical descriptors. These approaches have wide applications in the chemical industry in modeling physicochemical properties of chemical compounds, performing virtual screening in pharmaceutical studies, and to predict pharmacodynamic and pharmacokinetic properties in computational drug designing. The field of cheminformatics specializes in handling huge datasets, performing direct data to knowledge mapping and thus allows an easy transition from quantum chemistry to biological activity. It can provide information that may be used in medicine, biology, and physics. Various forms of machine-readable information are included in chemical databases which may be used in clinical purposes. There are some limitations of cheminformatics when used in stereoisomers and analyzing tautomers.

In this present study, we have explored the effect of the electronic atmosphere of 32 pyridyl benzamide derivatives on its HAT activity against *T. brucei*. The dataset is taken from a study performed by Masand *et al.* (2016) where they had screened a large pool of descriptors via objective and subjective feature selection and created multiple QSAR models by randomly splitting the datasets into training and prediction sets. Here we have incorporated global electrophilicity index (ω) and the square of electrophilicity index (ω^2) as electronic descriptors by replacing those used by Masand *et al.* one or two at a time to obtain all possible combinations of QSAR models. Multiple linear regression (MLR) method is employed to construct the models, followed by comparing them using the Sum of Ranking Differences (SRDs) technique described by K. Héberger (2010; Kollár-Hunek and Héberger, 2013).

THEORETICAL BACKGROUND

The electronic parameter of any drug brings out information about the polarising effect of any electronegative center present in the molecule on its binding pattern with the protein (DNA). Thus including the electronic parameters in simulating these types of bonding mechanisms is of utmost importance. Global electrophilicity index (ω) can be computed within the domain of conceptual

density functional theory (CDFT) (Hohenberg and Kohn, 1964; Kohn and Sham, 1965) using quantum chemical properties of a molecule like chemical hardness (η) (Chattaraj and Parr, 1993; Pearson, 1997) and chemical potential (μ) employing Parr's definition (Parr and Yang, 1989) as:

$$\omega = \frac{\mu^2}{2\eta} \quad (1)$$

Parr also established a quantum chemical relation between chemical potential (μ) and electronegativity (χ) (Pauling, 1960; Sen and Jorgenson, 1987) using Iczkowski and Margrave's (1961) definition of χ as the negative of first derivative of energy (E) with respect to the number of electrons (N) at the constant external potential $v(\vec{r})$. Electronegativity (χ) and hardness (η) are defined as (Pauling, 1960; Parr and Pearson, 1983):

$$\chi = -\mu = -\left(\frac{\partial E}{\partial N}\right)_{v(\vec{r})} \quad (2)$$

and η was defined by Parr and Pearson (1983) as the second order variation of E with respect to N :

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\vec{r})} = \left(\frac{\partial \mu}{\partial N}\right)_{v(\vec{r})} \quad (3)$$

By applying finite difference approximation (Parr and Yang, 1989), μ and η can be expressed in terms of ionization potential (IP) and electron affinity (EA) at fixed $v(\vec{r})$ as:

$$\mu = -\frac{(IP + EA)}{2} \quad (4)$$

$$\eta = IP - EA \quad (5)$$

We have applied Koopmans' theorem for closed-shell molecules so as to avoid calculating E_N , E_{N+1} and E_{N-1} to get IP and EA and hence cutting down computational cost in the process:

$$\mu = \frac{E_{LUMO} + E_{HOMO}}{2} \quad (6)$$

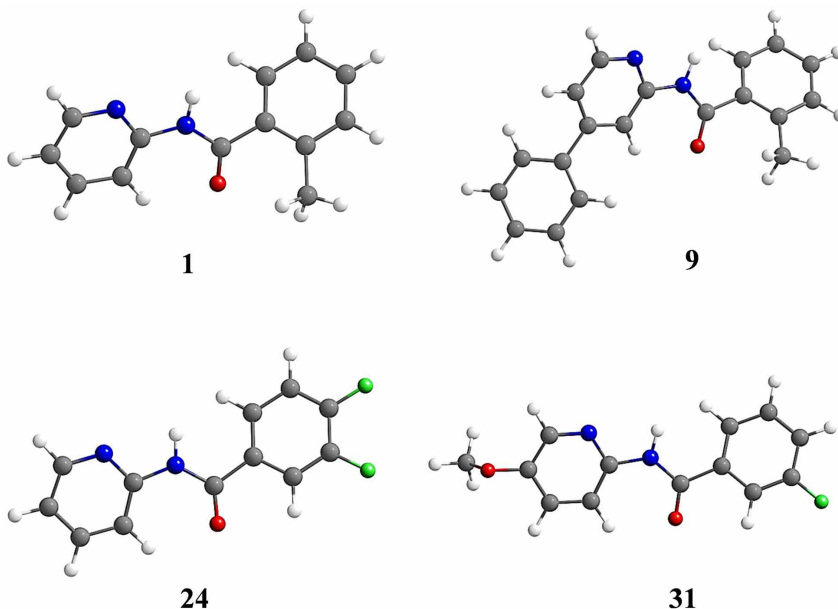
$$\eta = E_{LUMO} - E_{HOMO} \quad (7)$$

E_{HOMO} and E_{LUMO} being the energies of the highest occupied and lowest unoccupied molecular orbitals.

Table 1. Experimental pIC_{50} , ω , ω^2 along with the substituents for all the compounds in the dataset are presented corresponding to the structural skeletons in Figure 1

Sl. No.	R ₁	R ₂	Expt. pIC_{50}	ω	ω^2
1	H	2-Me	5.5190	0.37	0.14
2	4-CN	2-Me	5.6380	0.60	0.36
3	4-Me	2-Me	5.6780	0.36	0.13
4	4-Cl	2-Me	5.6780	0.44	0.19
5	4-Br	2-Me	5.7450	0.44	0.19
6	4-F	2-Me	6.2920	0.43	0.18
7	4-C≡CPh	2-Me	5.7700	0.47	0.22
8	4-C≡CCH ₂ -iPr	2-Me	5.9590	0.40	0.16
9	4-Ph	2-Me	5.6990	0.40	0.16
10	5-OMe	2-Me	5.4200	0.35	0.12
11	6-CH=CH(CH ₂) ₂ CH ₃	2-Me	5.2220	0.31	0.10
12	6-NH ₂	2-Me	5.3370	0.30	0.09
13	H	H	5.0420	0.40	0.16
14	H	2-Et	5.5620	0.36	0.13
15	H	2-Me, 3-F	6.0510	0.41	0.17
16	H	3-F	5.1350	0.46	0.21
17	H	2-Me, 3-Cl	6.2010	0.40	0.16
18	H	2-Me, 3-Br	5.2440	0.41	0.16
19	H	2,3-diMe	5.4560	0.34	0.11
20	H	2-Me, 4-F	5.7210	0.38	0.15
21	H	2-Me, 4-Cl	5.9590	0.43	0.19
22	H	2-Me, 4-Br	5.9590	0.44	0.19
23	H	2,4-diMe	6.0090	0.36	0.13
24	H	3-F, 4-F	6.3870	0.47	0.22
25	4-Me	3-F	6.0810	0.48	0.23
26	4-Me	3-F, 4-F	6.0460	0.46	0.21
27	4-Cl	3-F	6.3280	0.53	0.28
28	4-Cl	3-F, 4-F	6.2760	0.54	0.30
29	4-F	3-F	7.0000	0.52	0.27
30	4-F	3-F, 4-F	6.7210	0.53	0.29
31	5-OMe	3-F	5.6380	0.43	0.19
32	5-OMe	4-F	5.5850	0.39	0.15

Figure 2. Geometries of a few representative compounds from the dataset provided in Table 1 optimized at the HF/6-31G(d) level of theory



$$pIC_{50} = 4.8523(\pm 0.3676) - 0.9802 (\pm 0.4082) * GATS8c + 0.1240(\pm 0.0734) * RDF40p + 0.0552(\pm 0.0099) * RDF55s \quad (1.1)$$

$$pIC_{50} = 6.4521(\pm 1.2282) - 3.1455 (\pm 2.3630) * E1s + 0.0545(\pm 0.0492) * RDF40m + 0.0489(\pm 0.0111) * RDF55s \quad (1.2)$$

where, Geary autocorrelation of lag-8/weighted by atomic charges, Radial distribution function-040/weighted by relative polarizabilities and Radial distribution function-055/weighted by relative I-state are abbreviated as GATS8c, RDF40p, and RDF55s respectively. E1 and RDF40m refer to 1st component accessibility directional WHIM index/weighted by relative I-state, and Radial distribution function - 040/weighted by relative mass.

As mentioned above, these descriptors were selected by screening a pool of descriptors generated using PaDEL 2.21. However, this descriptor calculator does not evaluate the electrophilicity index of a compound, which has been proven to be an effective quantum chemical descriptor in predicting toxicity (Parthasarathi et al., 2004) and biological activities (Roy et al., 2007). Thus, exploring the electronic environment of the chemical compounds in this regard makes sense. The electronic aspect of the compounds is taken care of in this study in the form of global electrophilicity index (ω) and its square (ω^2). These descriptors could have been used in addition to the ones used by Masand *et al.* to obtain QSAR models with better predictive ability, but in that case, over-fitting would have been a high possibility. To avoid that, we have developed separate models by replacing each of the above descriptors (in model 1.1 and 1.2) once with ω and again with ω^2 , followed by replacing two of them with both ω and ω^2 . All such possible combinations are done for models 1.1 and 1.2.

Now the division of set in our study is done by distributing the compounds in three sets (A, B and C, provided in Table 2) among which two of them are combined to form the training sets and the remaining is the test set. Thus, there are three such cases, *viz.*, Case 1 contains set A as test and sets

Table 2. Division of 32 compounds taken for this study into three sets A, B and C. The numbers correspond to the compound numbers as presented in Table 1

	Set A	Set B	Set C
Compound numbers ^a	29	30	24
	28	6	27
	17	25	15
	8	23	26
	21	22	7
	9	20	5
	3	4	2
	14	32	31
	1	19	10
	11	18	12
	16	13	

^aas represented in Table 1

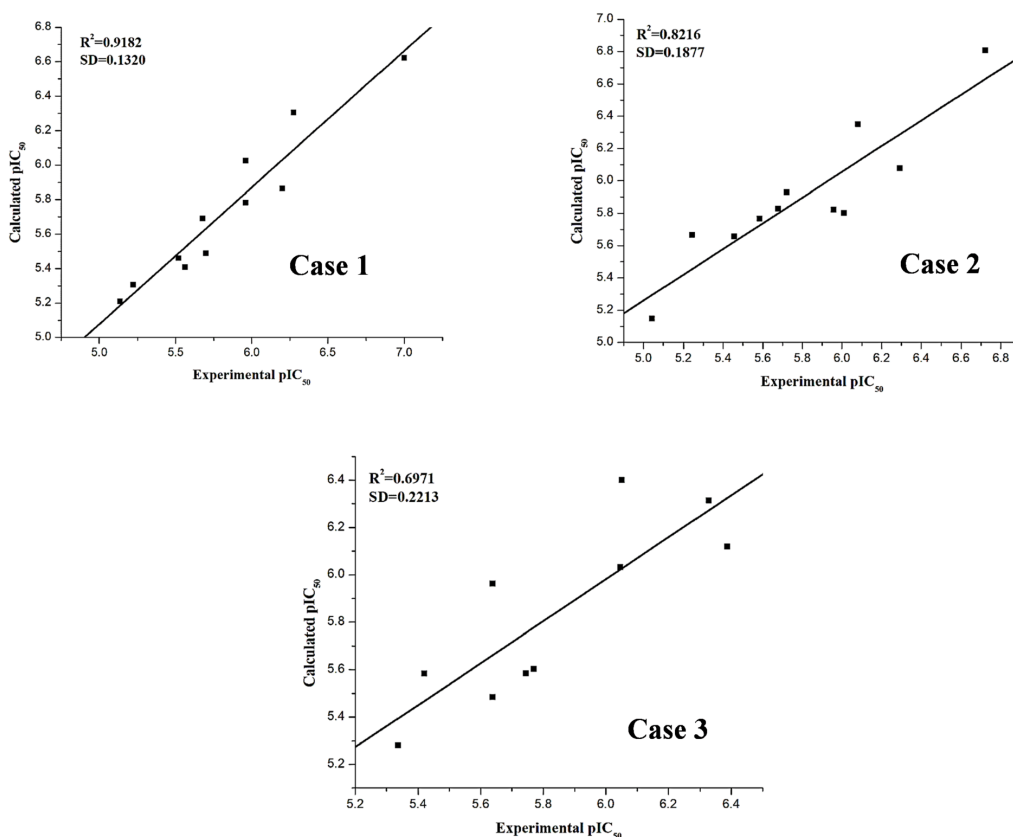
B & C as training, *Case 2* contains set B as test and A & C as training, and *Case 3* contains the only remaining combination *i.e.*, set C as test and A & B as training. The division of the dataset into 3 subsets is, however, not done randomly. Rather, the pIC_{50} values are used as a guide in this regard. The division is done in such a way to keep the average value of pIC_{50} of the compounds almost similar in each set. The QSAR models are built by simple multiple linear regression (MLR) technique where the coefficient of regression (R^2) and the standard deviation (SD) bring out their predictive ability (Table 3). For relatively small datasets, models built on the entire undivided datasets are required to compare and assess the ones built after dividing the dataset into training and test sets (Masand et al., 2015).

Table 3. R^2 and SD values obtained from MLR analysis on the undivided as well as three combinations of the divided dataset on 21 different QSAR models

Model No.	Generalized Regression Equations	Undivided		Case 1		Case 2		Case 3	
		R^2	SD	R^2	SD	R^2	SD	R^2	SD
1	$pIC_{50} = a + b * GATS8c + c * RDF40p + d * RDF55s$	0.8284	0.1960	0.9182	0.1320	0.8216	0.1877	0.6971	0.2213
2	$pIC_{50} = a + b * GATS8c + c * RDF40p + d * \omega$	0.3743	0.3742	0.4839	0.1872	0.3764	0.1546	0.2351	0.4378
3	$pIC_{50} = a + b * GATS8c + c * \omega + d * RDF55s$	0.7599	0.2318	0.8325	0.1768	0.7829	0.2063	0.5114	0.2415
4	$pIC_{50} = a + b * \omega + c * RDF40p + d * RDF55s$	0.7113	0.2542	0.7644	0.1983	0.7098	0.2283	0.4961	0.2529
5	$pIC_{50} = a + b * GATS8c + c * RDF40p + d * \omega^2$	0.3650	0.3770	0.4825	0.1827	0.3901	0.1447	0.2211	0.4098
6	$pIC_{50} = a + b * GATS8c + c * \omega^2 + d * RDF55s$	0.7592	0.2322	0.8323	0.1762	0.7853	0.2058	0.5817	0.2299
7	$pIC_{50} = a + b * \omega^2 + c * RDF40p + d * RDF55s$	0.7068	0.2562	0.7620	0.1977	0.7101	0.2301	0.4645	0.2575
8	$pIC_{50} = a + b * GATS8c + c * \omega + d * \omega^2$	0.3285	0.3877	0.4334	0.1888	0.2506	0.1882	0.1725	0.4421
9	$pIC_{50} = a + b * \omega + c * RDF40p + d * \omega^2$	0.3660	0.3767	0.4746	0.1781	0.2911	0.1800	0.1810	0.4594
10	$pIC_{50} = a + b * \omega + c * \omega^2 + d * RDF55s$	0.6836	0.2661	0.7637	0.2014	0.6962	0.2212	0.5163	0.2364
11	$pIC_{50} = a + b * E1s + c * RDF40m + d * GATS6m$	0.3056	0.3942	0.3991	0.1983	0.3199	0.2404	0.1159	0.2055
12	$pIC_{50} = a + b * E1s + c * RDF40m + d * \omega$	0.3647	0.3771	0.4949	0.1839	0.3196	0.1784	0.2793	0.3939
13	$pIC_{50} = a + b * E1s + c * \omega + d * GATS6m$	0.4847	0.3396	0.5936	0.2275	0.5415	0.2138	0.3369	0.3171
14	$pIC_{50} = a + b * \omega + c * RDF40m + d * GATS6m$	0.4758	0.3425	0.5824	0.2177	0.5019	0.2127	0.3263	0.3134
15	$pIC_{50} = a + b * E1s + c * RDF40m + d * \omega^2$	0.3571	0.3793	0.4997	0.1769	0.3241	0.1706	0.2135	0.4273
16	$pIC_{50} = a + b * E1s + c * \omega^2 + d * GATS6m$	0.4763	0.3424	0.5839	0.2266	0.5473	0.2083	0.2648	0.3481
17	$pIC_{50} = a + b * \omega^2 + c * RDF40m + d * GATS6m$	0.4666	0.3455	0.5708	0.2181	0.5105	0.2030	0.2518	0.3441
18	$pIC_{50} = a + b * E1s + c * \omega + d * \omega^2$	0.3421	0.3421	0.4673	0.1865	0.2898	0.1777	0.1339	0.4681
19	$pIC_{50} = a + b * \omega + c * \omega^2 + d * GATS6m$	0.4784	0.3417	0.5861	0.2301	0.4044	0.2680	0.1784	0.3648
20	$pIC_{50} = a + b * \omega + c * RDF40m + d * \omega^2$	0.3583	0.3790	0.4822	0.1844	0.2185	0.2091	0.1454	0.4405
21	$pIC_{50} = a + b * \omega + c * \omega^2$	0.3272	0.3813	0.4540	0.1813	0.2922	0.1765	0.1288	0.4612

As we can see from the R^2 and SD values provided in Table 3, the models with the descriptor combinations 1, 3, 4, 6, 7 and 10 show good correlation coefficients (R^2). For the rest of the models, the R^2 values lie below the threshold 0.6 which are statistically not relevant. Model 9 as reported in the study by Masand *et al.*, was based on multiple random divisions of the dataset. Here in this carefully sorted dataset splitting based on pIC_{50} values does not produce good correlation coefficients. However, in certain cases, the R^2 value does reach close to 0.6 *e.g.*, for the combinations 13, 14, 16, and 17. The model containing the combination of GATS8c, RDF40p, and RDF55s as descriptors for the undivided dataset shows excellent R^2 , and it is noted that removing RDF55s from the models drastically worsens the predictive ability of the models highlighting its importance. The efficiency with which a QSAR model predicts is reflected in the correlation coefficient between experimental and calculated activity (here pIC_{50}) of compounds that were not used for model development (*i.e.*, the test set). The correlation between the calculated and experimental pIC_{50} values for model number 1 for all cases are depicted in Figure 3.

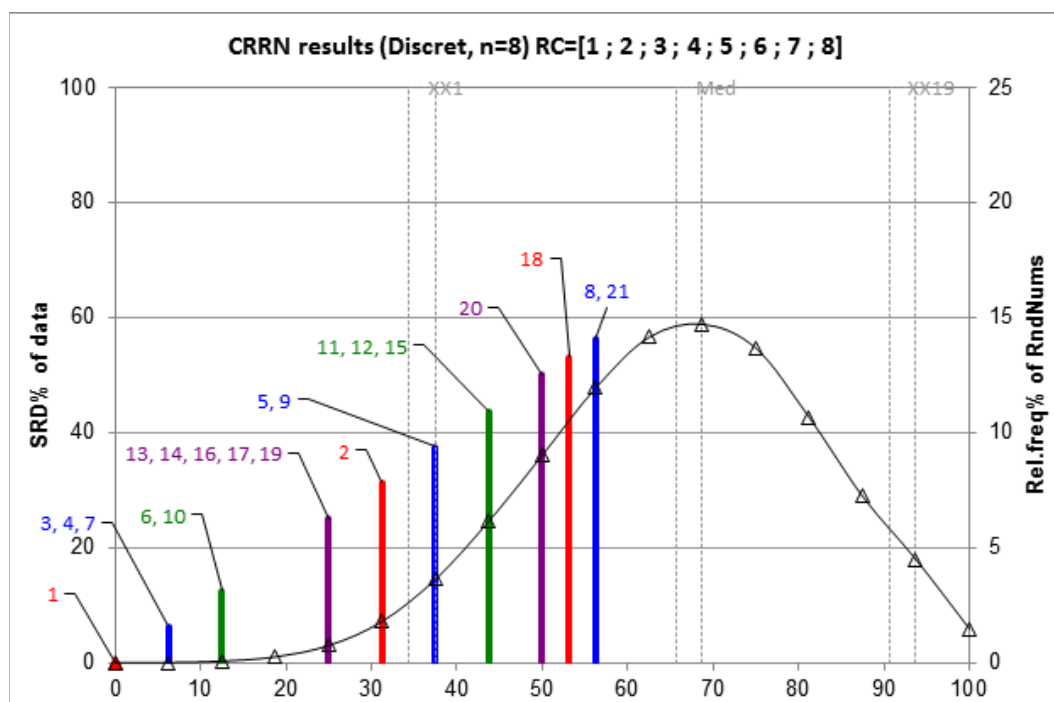
Figure 3. Plots of experimental versus calculated pIC_{50} along with their respective R^2 and SD values for the test sets of model 1



Further, we have employed the sum of ranking differences (SRDs) method as a model-comparison tool. In this method, relevant data need to be arranged in a matrix form where the rows and columns correspond to objects and variables, respectively. In this study, the objects refer to R^2 and SD values of the corresponding QSAR models (variables). The results of each model are ranked in increasing order, followed by taking a difference between the actual rank of the model results and that of the

reference or standard results (ideal). The sum of the absolute values of these differences is known as the SRD values which are used to compare the models. In the case where the ideal/reference ranking is unknown, the maximum/minimum/average value can be considered as ideal/golden standard depending upon the type of the objects. For example, in the present study, the maximum value of R^2 and minimum of SD are taken as reference/ideal in each row. This technique uses the concept of Occam's razor to evaluate the ranking: lower the value of SRD, better is the model. The graph (Figure 4) depicts the ranking of all the models used in this study, which also helps understand the extent of similarity between models from their relative position.

Figure 4. Evaluation of choice of descriptors using SRD with ties and subjected to the scaling between 0-100. Maximum R^2 and minimum SD values are used as golden standards. Scaled SRD values are plotted on the x-axis and left y-axis, the right y-axis shows the relative frequencies (black curve).



We have built 21 QSAR models on the undivided set and each of the three cases of divided sets as tabulated in Table 3. Thus our input matrix contains 21 columns of variables and 8 numbers of row (R^2 and SD for cases 1, 2, 3 and the undivided set). The program is so written such that if the nR (number of rows) is less than or equal to 8, it uses theoretical discrete distribution for the CRRN (Comparison of Ranks by Random Numbers) figure. From the graph, it can be easily inferred that the model with descriptors GATS8c, RDF40p and RDF55s having SRD value zero is the one with the best predictive ability, followed by models 3, 4, 7 and 6, 10 with SRD values of 6.25 and 12.50% respectively.

CONCLUSION

The current theoretical investigation of the QSAR model study is performed considering a dataset consisting of 32 pyridyl benzamide derivatives toward HAT activity using all possible combinations of

electronic factors (ω , ω^2) as easily computable descriptors in addition to the other descriptors against *T. brucei* as reported by Masand *et al.* The inclusion of electronic factors along with the screened descriptors on the developed models has a pretty good influence on predicting the biological activity of compounds. It is, however, noticed that when RDF55s is replaced with ω or ω^2 in model 1.1, the correlation coefficient (R^2) value decreases by a significant amount suggesting its importance in describing the HAT activity. The robustness of the predicted models has been analyzed not only in terms of R^2 (above the threshold value of 0.6) but also in the SRD-CRRN tool with ties to compare the efficacy of the models. CRRN integrated computational tool SRD helps us choose better models on the overall prediction of datasets consisting of dissimilar functionalities. The model constituted with descriptors GATS8c, RDF40p and RDF55s have the best predictive ability as reflected in the SRD graph (SRD value is zero) and models with number 3, 4, 7 and 6, 10 come next to it as the second and the third best predictive models having SRD values of 6.25 and 12.50% respectively.

ACKNOWLEDGMENT

PKC would like to thank the Editor-in-Chief, Professor Rama Rao Karri, Assistant Development Editor, Dr. Alexis Miller and IGI Global for kindly inviting us to contribute an article to the Special Issue of the International Journal of Cheminformatics and Chemical Engineering. He also thanks DST, New Delhi for the J. C. Bose National Fellowship. GJ and RP thank IIT, Kharagpur and CSIR respectively, for their fellowships. We would like to thank Professor K. Heberger for providing us with the links to the SRD-CRRN program to demonstrate our overall result.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interests regarding the publication of this article, financial, and/or otherwise.

REFERENCES

- Barrett, M. P., Vincent, I. M., Burchmore, R. J., Kazibwe, A. J., & Matovu, E. (2011). Drug resistance in human African trypanosomiasis. *Future Microbiology*, 6(9), 1037–1047. doi:10.2217/fmb.11.88 PMID:21958143
- Carvalho, A. S. D., Salomão, K., Castro, S. L. D., Conde, T. R., Zamith, H. P. D. S., Caffarena, E. R., & Boechat, N. et al. (2014). Megazol and its bioisostere 4H-1, 2, 4-triazole: Comparing the trypanocidal, cytotoxic and genotoxic activities and their in vitro and in silico interactions with the Trypanosoma brucei nitroreductase enzyme. *Memorias do Instituto Oswaldo Cruz*, 109(3), 315–323. doi:10.1590/0074-0276140497 PMID:24676659
- Chattaraj, P. K., & Parr, R. G. (1993). Density functional theory of chemical hardness. In *Chemical Hardness* (pp. 11–25). Springer. doi:10.1007/BFb0036797
- Deschenes, L. A., & Vanden Bout, D. A. (2000). *Origin 6.0: Scientific Data Analysis and Graphing Software Origin Lab Corporation (formerly Microcal Software, Inc.)*. www.originlab.com.
- Ferrins, L., Gazdik, M., Rahmani, R., Varghese, S., Sykes, M. L., Jones, A. J., & Charman, S. A. et al. (2014). Pyridyl benzamides as a novel class of potent inhibitors for the kinetoplastid Trypanosoma brucei. *Journal of Medicinal Chemistry*, 57(5), 6393–6402. doi:10.1021/jm500191u PMID:24978605
- Fock, V. (1930). Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik*, 61(1-2), 126–148. doi:10.1007/BF01340294
- Frisch, M., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., ... Petersson, G.E. (2014). Gaussian~ 09 Revision D. 01.
- Hartree, D. (1928). The wave mechanics of an atom with a non-coulomb central field. part iii. term values and intensities in series in optical spectra. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(3), 426-437. doi:10.1017/S0305004100015954
- Hartree, D. (1929). The Distribution of Charge and Current in an Atom consisting of many Electrons obeying Dirac's equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 25(2), 225-236. doi:10.1017/S0305004100018764
- Héberger, K. (2010). Sum of ranking differences compares methods or models fairly. *Trends in Analytical Chemistry*, 29(1), 101–109. doi:10.1016/j.trac.2009.09.009
- Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review*, 136(3B), B864–B871. doi:10.1103/PhysRev.136.B864
- Iczkowski, R. P., & Margrave, J. L. (1961). Electronegativity. *Journal of the American Chemical Society*, 83(17), 3547–3551. doi:10.1021/ja01478a001
- Kohn, W., & Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A), A1133–A1138. doi:10.1103/PhysRev.140.A1133
- Kollár-Hunek, K., & Héberger, K. (2013). Method and model comparison by sum of ranking differences in cases of repeated observations (ties). *Chemometrics and Intelligent Laboratory Systems*, 127, 139–146. doi:10.1016/j.chemolab.2013.06.007
- Masand, V. H., Mahajan, D. T., Maldhure, A. K., & Rastija, V. (2016). Quantitative structure–activity relationships (QSARs) and pharmacophore modeling for human African trypanosomiasis (HAT) activity of pyridyl benzamides and 3-(oxazolo [4, 5-b] pyridin-2-yl) anilides. *Medicinal Chemistry Research*, 25(10), 2324–2334. doi:10.1007/s00044-016-1664-1
- Masand, V. H., Mahajan, D. T., Nazeruddin, G. M., Hadda, T. B., Rastija, V., & Alfeefy, A. M. (2015). Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model. *Medicinal Chemistry Research*, 24(3), 1241–1264. doi:10.1007/s00044-014-1193-8
- Parr, R. G., & Pearson, R. G. (1983). Absolute hardness: Companion parameter to absolute electronegativity. *Journal of the American Chemical Society*, 105(26), 7512–7516. doi:10.1021/ja00364a005
- Parr, R. G., & Yang, W. (1989). *Density-functional theory of atoms and molecules*. New York: Oxford University Press.

Parthasarathi, R., Subramanian, V., Roy, D. R., & Chattaraj, P. K. (2004). Electrophilicity index as a possible descriptor of biological activity. *Bioorganic & Medicinal Chemistry*, 12(21), 5533–5543. doi:10.1016/j.bmc.2004.08.013 PMID:15465330

Pauling, L. (1960). *The Nature of the Chemical Bond*. Cornell University Press.

Pearson, R. G. (1997). *Chemical hardness*. Wiley-VCH. doi:10.1002/3527606173

Petersson, A., Bennett, A., Tensfeldt, T. G., Al-Laham, M. A., Shirley, W. A., & Mantzaris, J. (1988). A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *The Journal of Chemical Physics*, 89(4), 2193–2218. doi:10.1063/1.455064

Roy, D., Pal, N., Mitra, A., Bultinck, P., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2007). An atom counting strategy towards analyzing the biological activity of sex hormones. *European Journal of Medicinal Chemistry*, 42(11-12), 1365–1369. doi:10.1016/j.ejmech.2007.01.028 PMID:17416442

Seixas, J., Kennedy, A., & Lutje, V. (2013). Chemotherapy for second-stage Human African trypanosomiasis. *Cochrane Database of Systematic Reviews*, 2013. PMID:23807762

Sen, K., & Jorgenson, C. (1987). *Structure and bonding*, vol. 66: Electronegativity. Springer.

Simarro, P., Franco, J., Diarra, A., Postigo, J. R., & Jannin, J. (2012). Update on field use of the available drugs for the chemotherapy of human African trypanosomiasis. *Parasitology*, 139(7), 842–846. doi:10.1017/S0031182012000169 PMID:22309684

Sykes, M. L., Baell, J. B., Kaiser, M., Chatelain, E., Moawad, S. R., Ganame, D., & Avery, V. M. et al. (2012). Identification of compounds with anti-proliferative activity against *Trypanosoma brucei brucei* strain 427 by a whole cell viability based HTS campaign. *PLoS Neglected Tropical Diseases*, 6(11), e1896. doi:10.1371/journal.pntd.0001896 PMID:23209849

Ranita Pal is a 2nd year research scholar working under the supervision of Professor Pratim Kumar Chattaraj at Centre for Theoretical Studies, Indian Institute of Technology Kharagpur. She acquired her Bachelor of Science degree in Chemistry (Hons.) from Lady Brabourne College, Calcutta University in the year 2016. She pursued a Master's degree from School of Chemistry, University of Hyderabad and graduated in the year 2018. She secured an all India CSIR rank of 71 in National Eligibility Test (CSIR-NET) in December 2017, and AIR 73 in Graduate Aptitude Test in Engineering (GATE) in 2020. She has already co-authored 5 journal articles and a book chapter. Her research activities include electronic structure theory, hydrogen storage, chemical reactivity, binding of small molecules and isomerization processes, studies on Adaptive Natural Density Partitioning method (AdNDP), investigation of nc-2e bond, machine learning and QSAR/QSPR/QSTR analyses in the domain of theoretical and computational chemistry.

Goutam Pal is a 2nd year student at Indian Institute of Technology Kharagpur, pursuing his Master of Science degree in Chemistry. He acquired his Bachelor of Science degree in Chemistry (Hons.) from Santipur College, University of Kalyani in the year of 2017. He secured an all India rank of 469 in Indian Institute of Technology - Joint Admission Test (IIT-JAM) in the year of 2018. He joined the research group of Professor Pratim Kumar Chattaraj in the Department of Chemistry to work on this M.Sc. dissertation. His research activities include machine learning and QSAR/QSPR/QSTR analyses in the domain of theoretical and computational chemistry.

Gourhari Jana is a 5th-year research scholar working under Professor Pratim Kumar Chattaraj, Department of Chemistry, Indian Institute of Technology Kharagpur. He completed his B.Sc. degree in Chemistry (Hons.) from Midnapore College under Vidyasagar University in 2013. He received his M.Sc. degree from the same University in 2015. He ranked 65th (AIR) in Graduate Aptitude Test in Engineering (GATE-2016) and secured 50th rank in National Eligibility Test (NET) under CSIR Fellowship Scheme in June 2017. He has received "APC Memorial award" from ICS in 2018, "Research Excellence Award and Young Achiever Award" from InSc in 2019. He has already co-authored 30 journal articles including one book chapter. His research activities include electronic structure theory, machine learning and deep learning, QSAR/QSPR/QSTR analyses, quantum trajectory (ADMP, BOMD), Time-Dependent Density Functional Theory (TDDFT), Non-Linear Optical (NLO) properties, Adaptive Natural Density Partitioning (AdNDP) method, aromaticity, etc. in the domain of theoretical and computational chemistry.

Pratim Kumar Chattaraj is an Institute Chair Professor at the Indian Institute of Technology Kharagpur. His research interests include density functional theory, nonlinear dynamics, aromaticity in metal clusters, hydrogen storage, noble gas compounds, machine learning, chemical reactivity and quantum trajectories. He is a Fellow of The World Academy of Sciences and all three Indian Science Academics. He is a Sir J.C. Bose National Fellow. Several of his papers have become editors' choice/hot/most cited/most accessed/cover articles.