International Journal of Chemoinformatics and Chemical Engineering
Volume 8 • Issue 2 • July-December 2019

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/IJCCE.20190701.oa1

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