Chemometrics and Intelligent Laboratory Systems 148 (2015) 134–144

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

QSAR models and scaffold-based analysis of non-nucleoside HIV RT inhibitors

Bilal Nizami [a], Igor V. Tetko [b,c], Neil A. Koorbanally [d], Bahareh Honarparvar [a,*]

[a] School of Pharmacy and Pharmacology, University of KwaZulu-Natal, Durban 4000, South Africa
[b] Helmholtz-Zentrum München–German Research Centre for Environmental Health (GmbH), Institute of Structural Biology, Ingolstaedter Landstrasse 1, D-85764, Neuherberg, Germany
[c] A.M. Butlerov Institute of Chemistry, Kazan Federal University, Kremlyovskaya St. 18, 420008 Kazan, Russia
[d] School of Chemistry, University of KwaZulu-Natal, Private Bag X54001, Durban, 4000, South Africa

ARTICLE INFO

Article history:
Received 29 June 2015
Received in revised form 9 September 2015
Accepted 21 September 2015
Available online 28 September 2015

Keywords:
Non-nucleoside reverse transcriptase (NNRT) pyrimidine derivatives
HIV reverse transcriptase (HIV RT)
Quantitative structure–activity relationship (QSAR)
Matched molecular pair analysis (MMPA)
Molecular docking

ABSTRACT

A selection of 289 pyrimidine derivatives with anti-HIV RT activities as non-nucleoside HIV RT inhibitors (NNRTI) were studied. The associative neural network (ASNN) method was applied to develop a quantitative structure–activity relationship (QSAR) for anti-HIV RT activity. The calculated models were validated using the bagging approach. A consensus model with $R^2 = 0.87$ and RMSE = 0.5 was obtained from 10 individual models. Scaffold analysis and molecular docking of the compounds used in the QSAR model identified a potential chemical scaffold. The results showed that scaffold-based analysis of the QSAR model could be helpful in identifying potent scaffolds for further exploration than analyzing the overall model. Matched molecular pair analysis (MMPA) was applied in the QSAR model to characterize molecular transformations causing a significant change in the anti-HIV activity. The linear QSAR model was calculated to explore the structural features important for NNRTI activity. The results revealed that the activity of NNRT inhibitors is strongly dependent on their aromaticity and structural flexibility. The scaffold-based analysis of QSAR models with molecular docking and MMPA was found to be helpful in characterizing potential scaffolds for anti-HIV RT derivatives. The outcome of this study provides a deeper insight into the computer-aided scaffold-based design of novel molecules with HIV RT activities. It was also clearly shown that the consensus model's failure to correctly predict new chemical series could be due to the limitation of its applicability domain (AD). Redevelopment of models using new measurements can dramatically increase their AD and performance.

1. Introduction

According to a recent WHO estimate, 35.3 million people were living with HIV/AIDS worldwide in 2012 [1], with a significant number of these infections being resistant to antiretroviral therapies. HIV utilizes reverse transcriptase (RT), an enzyme that makes copies of cDNA from RNA, a process called reverse transcription. This makes RT an attractive target for anti-retroviral drugs like non-nucleoside reverse transcriptase inhibitors (NNRTIs) [2].

The higher rate of mutation in HIV strains, and the subsequent development of resistance to the NNRTIs, is a major issue in managing HIV infection. This highlights the need for rapid and rational development of NNRTIs. Pyrimidine derivatives were synthesized for decades and have been actively pursued as NNRTIs [3]. Two main series of pyrimidine derivatives are DABO (dihydro-alkoxy-benzyl-oxopyrimidine) and DAPY (di-aryl pyrimidine) [2]. Owing to NNRTI's importance in targeting HIV RT, QSAR studies have been used to understand the relationship between its structure and anti-HIV RT activity.

Quantitative structure–activity relationship (QSAR) is the mathematical modeling of chemical structures of compounds and their relationship with biological activity, and is actively used in drug design [4,5]. Knowledge of the relationship between structural properties of chemical compounds and their biological activity is crucial in optimizing lead molecules. The construction of QSAR models typically consists of two main steps: (i) calculation and representation of structural features (molecular descriptors) of the selected compounds; (ii) multivariate analysis for correlating molecular descriptors with observed activities (biological, physico-chemical and ADMET properties). Numerous methods, ranging from simple linear regression to complex machine learning algorithms are applied to explore structure–activity relationships. The linear regression models in general allow a relatively straightforward interpretation in terms of linear regression coefficients, provided that descriptors used in the equation are not correlated. However, the models obtained by machine learning methods, such as neural network and support vector machines, are more difficult to interpret due to the non-linear nature of the algorithms.

A recently developed matched molecular pair analysis (MMPA) approach has the capability to address the "black box" nature of QSAR models [6]. A matched molecular pair (MMP) is defined as a pair of molecules that differ by a minor structural change at a single point [7].

* Corresponding author at: School of Pharmacy and Pharmacology, University of KwaZulu-Natal, Durban 4001, South Africa. Tel.: +27 31 2608482.
E-mail address: Honarparvar@ukzn.ac.za (B. Honarparvar).

http://dx.doi.org/10.1016/j.chemolab.2015.09.011
0169-7439/