Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfda-online.com

Special Invited Article

Amino substituted nitrogen heterocycle ureas as kinase insert domain containing receptor (KDR) inhibitors: Performance of structure–activity relationship approaches

Hayriye Yilmaz [a,b], *Natalia Sizochenko* [b,c], *Bakhtiyor Rasulev* [b], *Andrey Toropov* [d], *Yahya Guzel* [e], *Viktor Kuz'min* [c], *Danuta Leszczynska* [f], *Jerzy Leszczynski* [b,*]

[a] *Kayseri Vocational School, Biomedical Devices and Technologies, Erciyes University, 38039, Kayseri, Turkey*
[b] *Interdisciplinary Center for Nanotoxicity, Department of Chemistry and Biochemistry, Jackson State University, Jackson, MS, 39217, USA*
[c] *Odessa I.I. Mechnikov National University, Department of Chemistry, Dvoryanskaya Street, 2, 65082, Odessa, Ukraine*
[d] *Laboratory of Environmental Chemistry and Toxicology, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy*
[e] *Department of Chemistry, Faculty of Science, Erciyes University, 38039, Kayseri, Turkey*
[f] *Department of Civil and Environmental Engineering, Jackson State University, Jackson, MS, 39217, USA*

ARTICLE INFO

Article history:
Available online 1 April 2015

Keywords:
amino-substituted nitrogen heterocyclic ureas
descriptors
KDR inhibitors
QSAR
SiRMS
SMILES

ABSTRACT

A quantitative structure–activity relationship (QSAR) study was performed on a set of amino-substituted nitrogen heterocyclic urea derivatives. Two novel approaches were applied: (1) the simplified molecular input-line entry systems (SMILES) based optimal descriptors approach; and (2) the fragment-based simplex representation of molecular structure (SiRMS) approach. Comparison with the classic scheme of building up the model and balance of correlation (BC) for optimal descriptors approach shows that the BC scheme provides more robust predictions than the classic scheme for the considered pIC_{50} of the heterocyclic urea derivatives. Comparison of the SMILES-based optimal descriptors and SiRMS approaches has confirmed good performance of both techniques in prediction of kinase insert domain containing receptor (KDR) inhibitory activity, expressed as a logarithm of inhibitory concentration (pIC50) of studied compounds.

* *Corresponding author*. Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, Post Office Box 17910, Jackson, MI 39217, USA.
E-mail address: jerzy@icnanotox.org (J. Leszczynski).
http://dx.doi.org/10.1016/j.jfda.2015.03.001

1. Introduction

The kinase insert domain containing receptor (KDR), alternatively referred to as VEGFR-2, is a receptor for vascular endothelial growth factors (VEGFs). It functions as a key regulator of angiogenesis, the process by which new capillaries are created from preexisting blood vessels [1]. Accordingly, interruption of VEGFR-2 signaling by small molecule inhibitors to VEGFR-2 kinase domain has been shown to be an attractive strategy in the treatment of cancer. In recent years, a novel series of amino-substituted nitrogen heterocyclic urea derivatives has been reported as being essential inhibitors against KDR [2].

Quantitative structure–activity relationship (QSAR) methods are widely applied nowadays to find mathematical relationships between the chemical structure of a compound and its biological activity [3–17]. This technique was utilized here, based on experimental data available, and calculated theoretical descriptors, to perform an inhibitory activity study [6,10,17]. In the present study, the predictive QSAR models were developed for a set of amino-substituted nitrogen heterocyclic ureas for which the molecular structure is represented by simplified molecular input-line entry systems (SMILES) applying new techniques, such as the SMILES-based such as the SMILES-based optimal descriptors approach implemented in CORrelations And Logic (CORAL) (http://www.insilico.eu/coral), and the simplex representation of molecular structure (SiRMS) approach [18].

2. Materials and methods

2.1. Dataset

For prediction of inhibitory binding affinities (pIC_{50}, i.e., logarithm of the 50% effective concentration) the data on 63 amino-substituted nitrogen heterocyclic ureas were collected from existing literature [19].

2.2. Computational details

2.2.1. CORAL approach

There are three options for the selection of optimal descriptors in CORAL: (1) graph based; (2) SMILES based; and (3) hybrid descriptors which are calculated using both graph and SMILES approaches [20–23]. There are two classes of graph invariants which are available in CORAL: vertices and Morgan vertices degrees. In the case of hydrogen-suppressed graphs (HSGs) and hydrogen-filled graphs, vertices are representations of the chemical elements, such as carbon, nitrogen, oxygen, etc. In the case of graphs of atomic orbitals, vertices represent electronic structures i.e. atomic orbitals such as $1s^1$, $2s^2$, $2p^5$, $3d^{10}$, etc. [24].

The optimal graph-based descriptor based on so-called correlation weights (DCW) is calculated as the following:

$$^{Graph}(\text{Threshold}, N_{epoch}) = \sum CW(A_k) + a\sum CW(^0EC_k) + \beta\sum CW(^1EC_k) + \gamma CW(^2EC_k) + \delta CW(^3EC_k) \quad \text{Equation 1}$$

Three topological invariants of the molecular graphs were involved in current study: vertex degree (EC0); extended connectivity of first order (EC1); and extended connectivity of second order (EC2) [25].

The optimal SMILES-based descriptor based on correlation weights:

$$^{SMILES}(\text{Threshold}, N_{epoch}) = a\sum CW(S_k) + \beta\sum CW(SS_k) + \gamma\sum CW(SSS_k) + \delta CW(\text{PAIR}) + x\,CW(\text{NOSP}) + y\,CW(\text{HALO}) + z\,CW(\text{BOND}) \quad \text{Equation 2}$$

S_k, SS_k, and SSS_k are representations of molecular fragments, for example if SMILES = Clc1ccccc1 then s_k = (Cl, c, 1, c, c, c, c, 1); ss_k = (Clc, c1, cc, cc, cc, cc, c1); sss_k = (Clc1, c1c, ccc, ccc, ccc, ccc, cc1). *PAIR, NOSP, HALO,* and *BOND* are global SMILES attributes which are calculated with SMILES. These global attributes provide the possibility of carrying out an additional discrimination of substances into separated classes: for example nitrogen, oxygen, sulphur, and phosphorus (NOSP); fluorine, chlorine, and bromine (HALO) [24]. The BOND attribute is related to presence/absence of three categories of chemical bonds: double, triple, and stereospecific. The coefficients *a*, β, γ, x, y, and z can be either 1 or 0. One (1) indicates that the SMILES attribute is involved in the calculation of the descriptor of correlation weights (DCW) (Threshold) and zero (0) indicates that the SMILES attribute is not involved. Combinations of values of different attributes provide the possibility of defining various versions of SMILES based optimal descriptors [20].

CORAL software can be also used to build up a hybrid model which is calculated with SMILES-based and GRAPH-based descriptors:

$$^{Hybrids}(\text{Threshold}, N_{epoch}) = {}^{Graph}(\text{Threshold}, N_{epoch}) + {}^{SMILES}(\text{Threshold}, N_{epoch}) \quad \text{Equation 3}$$

The graph- and SMILES-based models are mathematical functions of the threshold and the number of N_{epoch} of the Monte Carlo optimization. The most predictive combination of T and N_{epoch} values for a split of data can be found by analyzing results of the calculations for several different splits of data in the training and test sets.

2.2.2. SiRMS approach

In addition to the above mentioned approaches, the SiRMS technique [18] was also applied to calculate fragmentary 2D descriptors (fragments of the size 2–5). In the framework of SiRMS, any molecule can be represented as a system of different simplexes (fragments of fixed composition and topology). In previous studies this method provided good results for solving different "structure–activity" problems [26–29].

In the current study a 2D level of molecule representation was utilized to generate simplex fragments. During the first step, the connectivity of atoms in simplex, atom type, and bond nature were considered. For each property the range is created with four to seven intervals. In this study all atoms were divided into groups corresponding to their atomic refraction (A < 1.5 < B < 3 < C < 8 < D), partial charges