Computational Biology and Chemistry 65 (2016) 8–15

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Review Article

1,3-Oxazole derivatives as potential anticancer agents: Computer modeling and experimental study

Ivan Semenyuta[a,*], Vasyl Kovalishyn[a], Vsevolod Tanchuk[a], Stepan Pilyo[a], Vladimir Zyabrev[a], Volodymyr Blagodatnyy[b], Olena Trokhimenko[b], Volodymyr Brovarets[a], Larysa Metelytsia[a]

[a] Institute of Bioorganic Chemistry and Petrochemistry, National Academy of Sciences of Ukraine, Kyiv-94, 02660, 1 Murmanska Street, Ukraine
[b] P.L.Shupyk National Medical Academy of Postgraduate Education, Kyiv, 04112, 9 Dorohozhytska Street, Ukraine

ARTICLE INFO

Article history:
Received 13 April 2016
Received in revised form 5 July 2016
Accepted 20 September 2016
Available online 21 September 2016

Keywords:
Tubulin inhibitors
Drug design
Qsar
Molecular docking
Associative Neural Networks (ASNN)

ABSTRACT

Microtubules play a significant role in cell growth and functioning. Therefore inhibition of the microtubule assemblies has emerged as one of the most promising cancer treatment strategies. Predictive QSAR models were built on a series of selective inhibitors of the tubulin were performed by using Associative Neural Networks (ANN). To overcome the problem of data overfitting due to the descriptor selection, a 5-fold cross-validation with variable selection in each step of the analysis was used. All developed QSAR models showed excellent statistics on the training (total accuracy: 0.96–0.97) and test sets (total accuracy: 0.95–97). The models were further validated by 11 synthesized 1,3-oxazole derivatives and all of them showed inhibitory effect on the Hep-2 cancer cell line. The most promising compound showed inhibitory activity $IC_{50} = 60.2\ \mu M$. In order to hypothesize their mechanism of action the top three compounds were docked in the colchicine binding site of tubulin and showed reasonable docking scores as well as favorable interactions with the protein.

Contents

1. Introduction . . . 8
2. Materials and methods . . . 9
 2.1. Experimental data and descriptor generation . . . 9
 2.2. Associative neural networks . . . 9
 2.3. Search for an optimal descriptor set . . . 9
 2.4. Validation of QSAR models . . . 10
 2.5. Molecular docking . . . 10
3. Results and discussion . . . 10
 3.1. Results of QSAR modeling of tubulin inhibitor set . . . 10
 3.2. Chemistry . . . 11
 3.3. Biological evaluation . . . 12
 3.4. Molecular docking . . . 12
4. Conclusion . . . 14
 Conflict of interest . . . 14
 References . . . 14

1. Introduction

It is known that emergence of malignant tumors includes uncontrollable cell proliferation. Therefore majority of the chemotherapeutic agents affect the mitosis stage (Jordan and

* Corresponding author.
E-mail address: ivan@bpci.kiev.ua (I. Semenyuta).

http://dx.doi.org/10.1016/j.compbiolchem.2016.09.012
1476-9271/

Wilson, 2004). One of the molecular targets in the human body is a cellar protein tubulin and microtubule assemblies (microtubules) formed by it (Kavallaris, 2010). Cytotoxic compounds such as taxol, vinblastine, and colchicine as well as their numerous analogs inhibit the mitosis by linking with various areas of the tubulin and activating or inhibiting either the process of assembly of the microtubules (colchicine, alkaloids of periwinkle) or uncontrollable polymerization of the tubulin (taxans: taxol and taxotere) (Walsh and Goodman, 1999). Taxans due to their *in vivo* efficiency are widely used for the most widespread malignancies despite of difficulties and high cost of their synthesis (Launois et al., 2008). More available colchicine and its analogs possessing a significantly simpler structure are too toxic at therapeutic doses (Cocco et al., 2010). It is therefore urgent to identify novel tubulin-binding inhibitors with a novel mechanism of action and improved safety profile.

A sufficiently large number of chemical compounds – tubulin polymerization inhibitors that bind to tubulin using the colchicine binding site were synthesized and tested recently. These compounds include oxazole (Kaura et al., 2014), thiazole (Salehi et al., 2013), and benzimidazole (Chen et al., 2011) derivatives.

Here, we report QSAR studies, molecular docking, synthesis and anticancer activity of a set of known and novel 1,3-oxazole derivatives.

2. Materials and methods

2.1. Experimental data and descriptor generation

The bioassay data entries in PubChem (AID:2205) corresponding to the screening of anticancer activity of a very diverse set of compounds were used to construct a random dataset (Anon., 2015a,b,c,d,e,f,g,h). The data were uploaded into the Instant JChem (Anon., 2015a,b,c,d,e,f,g,h) database and ranked according to Dice Index (DI) values. All active tubulin inhibitors were included. Inactive compounds were selected by the Kennard–Stone design (Kennard and Stone, 1969) from the first 10,000 compounds as ranked by the DI score in order to form the most diverse subset. This way, the training set consisted of 1621 active and 1621 inactive tubulin inhibitors based on bioassay AID:2205. All molecules were converted to their standardized forms using the ChemAxon standardizer (Anon., 2015a,b,c,d,e,f,g,h). The 2D and 3D coordinates of atoms were recalculated, counterions and salts were removed from molecular structures, molecules were neutralized, mesomerized, and aromatized. Datasets were checked for structural duplicates. The 3D structures were designed using the ChemAxon standardizer from the SMILES notations available for each compound and stored in SDF format. 3200 descriptors were calculated as the initial set of variables using DRAGON (Anon., 2015a,b,c,d,e,f,g,h). Descriptors with values that greatly correlated with those of other descriptors (with a correlation coefficient $\geq$0.99) were omitted to avoid redundancy. As a result, 1314 molecular descriptors were selected.

2.2. Associative neural networks

Associative Neural Networks (ASNN) is a combination of an ensemble of Feed-Forward Neural Networks (FFNN) with a k-Nearest Neighbors (k-NN) method (Tetko, 2002). FFNN is a nonlinear supervised regression data fitting procedure. A traditional FFNN represents a memory-less approach, i.e. after training, the initial data are no longer needed and all the information necessary for predictions is stored within the neural network weights and architecture (Sandberg et al., 2001). The k-NN addition is method represents a memory-based approach (Dasarathy, 1991). The k-NN keeps in memory all the input data and their predictions are corrected based on a local approximation of the closest neighbors. The ASNN uses the k-NN method in the space of ensemble residuals. All compounds are represented as vectors of neural network predictions by the neural network ensemble. Correlation between such vectors is used by the k-NN as a measure of distance between the analyzed cases. Therefore, the ASNN performs the k-NN in the space of the ensemble residuals. As a result ASNN improves prediction by the biased correction over the FFNN ensemble (Tetko, 2002).

The FFNN was trained by the SuperSAB algorithm (Tollenaere, 1990). The neural networks had the number of inputs equal to the number of descriptors and five neurons in one hidden layer. There was also a bias neuron both on the input and on hidden layers. Weights were initialized with random numbers. ASNNs were used with two output neurons – the target values were assigned to 1 for active and 0 for inactive compounds. All neural networks had the same architecture. The ASNN ensemble included M = 200 networks. The possibility of data over-fitting was strictly controlled by the cross-validation techniques known as the Early Stopping over Ensemble (ESE) (Tetko et al., 1995). More details of the algorithm can be found in publications (Tetko et al., 1995; Tetko and Villa, 1997).

2.3. Search for an optimal descriptor set

Methods that deal with the sensitivity analysis estimate the level of changes in the model's output resulting from the changes of model's inputs. They are used to find a set of input descriptors, which produce the most accurate output values. Pruning methods implemented in ASSN were used as a selection tool. These methods are very efficient in QSAR studies (R Development Core Team, 2004; LeCun et al., 1990). Pruning algorithms introduce some measure of importance of the ASNN matrix weights by the so called sensitivities. These algorithms work similarly to stepwise multiple regression analysis excluding on each step one input parameter, which is considered non-significant. At each step, the model sensitivities to all weights and input nodes are estimated and the descriptor corresponding to the input neuron with the smallest sensitivities is deleted. Detailed descriptions of the used sensitivity method can be found in earlier publications (Chauvin, 1989; Tetko et al., 1996).

To evaluate the classification ability and to separately control the classification performance of the two classes, sensitivity (Sn), specificity (Sp), and overall accuracy (Ac) were calculated (Kovalishyn et al., 1998). Sensitivity is also called the true positive rate or positive class accuracy, while specificity is also called the true negative rate or negative class accuracy. A theoretical optimal prediction can achieve 100% sensitivity (i.e. predict all active molecules from the active group as active) and 100% specificity (i.e., does not predict any molecule from the inactive group as active).

$$Sn = TP/(TP + FN) \quad (1)$$

$$Sp = TN/(TN + FP) \quad (2)$$

$$Ac = (TP + TN)/(TP + FN + TN + FP) \quad (3)$$

where TP, FP, TN and FN denote number of true positives, false positives, true negatives and false negatives, respectively. In general, the overall accuracy Ac is a good measure of predictive ability of models if the number of active and inactive compounds is roughly equal.