Chemometrics and Intelligent Laboratory Systems 150 (2016) 23–28

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

QSAR study of pyrazolo[1,5-a]pyrimidine derivative inhibitors of Chk1

Jing Chen *, Miao Zhang, Qing Ma, Dongdong Qin, Liping Zhang, Xiaoquan Lu *

Key Laboratory of Bioelectrochemistry and Environmental Analysis of Gansu Province, College of Chemistry and Chemical Engineering, Northwest Normal University, Lanzhou, 730070, PR China

ARTICLE INFO

Article history:
Received 30 August 2015
Received in revised form 25 October 2015
Accepted 29 October 2015
Available online 6 November 2015

Keywords:
Checkpoint kinase 1
Quantitative structure–activity relationship
Particle swarm optimization
Support vector machine

ABSTRACT

Checkpoint kinase 1 (Chk1) is a serine/threonine kinase that plays a key role in the response to DNA-mediated cell injury. In this paper, the quantitative structure–activity relationship (QSAR) models were constructed to predict the activity of pyrazolo[1,5-a]pyrimidine derivatives of Checkpoint kinase 1 (Chk1) by using SVM and PSO-SVM methods. The root-mean-square errors (*RMSE*) of the training set and the test set for the PSO-SVM model were 0.0886 and 0.1803, respectively. For the SVM model, the values were 0.2185 and 0.4023, respectively. The results showed that the performance of the PSO-SVM model was better than the corresponding SVM model. Thus, it can be inferred that the PSO-SVM analysis will be a promising method and be spread to apply in the QSAR studies.

1. Introduction

Checkpoint kinase 1 (Chk1) is a serine/threonine kinase that has been involved in mediating cellular response to DNA damage. In the event of DNA damage, Chk1 is activated via ataxia-telangiectasia mutated (ATM) and ataxia-telangiectasia related (ATR) to cell-cycle arrest at S and G2 checkpoints, the activation of DNA repair, senescence or apoptosis [1]. Chk1 has been reported to play key roles in a variety of processes including cogenesis, tissue growth, metabolic stress, tumorigenesis and neuronal survival [2]. Chk1 is discovered and initially identified in fission yeast genus. Subsequently, the analog of Chk1 is also found in eukaryotic cells, such as mammals, insect, budding yeast and so on [3]. The damage of the cell genomic causes increased sensitivity of the cell toward ultraviolet light and ultimately progression into mitosis and cell death [4]. However, the cell division can rapidly lead to the cancer and tumors because the inhibition of Chk1 is regarded as a targeted approach to enhance the cytotoxicity of DNA-damaging agents toward tumor cells [5,6]. Hence, it has become more and more attractive as potential antitumor agents in the oncology field.

The inhibition of Chk1 is regarded as a potential therapy for cancer and tumor, which can potentiate the cytotoxicity of genotoxic therapies [7]. Thus, it has been extensively demonstrated in preclinical studies with Chk1 RNAi and small molecule Chk1 inhibitors [7,8]. In recent years, several checkpoint kinase inhibitors, such as PF00477736 [9], AZD7762 [10], CCT244747 [11] and SCH 900776 [12] have been described in literatures. In view of developing novel inhibitors, quantitative structure–activity relationship (QSAR) has been widely utilized in the fields of chemical/biological chemistry and related sciences, especially within the areas of the computer-aided drug design. Linear and nonlinear methods are used to build the QSAR model. Support vector machine (SVM) is one of nonlinear statistical methods that perform the nonlinear statistical modeling. The SVM, which is proposed by Heikamp and Bajorath [13], is mainly used for binary object classification and ranking. The advantages of SVM are to avoid local optima and over learning problems. Therefore, it has been used in conventional neural networks based on the structural risk minimization inductive principle, and the weighted SVM is already applied for constructing the prediction model of active compounds [14]. SVM is also a meaningful method in QSAR study.

In this study, the character parameters of chemical structure were selected by stepwise multiple linear regression (stepwise-MLR), including JGI4, G3s, R8u^{+} and RDF085e. The QSAR models were established by using SVM and PSO-SVM methods via a biological activity and the proposed parameters to predict the activities of pyrazolo[1,5-a]pyrimidine derivatives. The results showed that the predictive capability of PSO-SVM was better than the SVM model. Therefore, the PSO-SVM approach was more promising for predicting the activities of pyrazolo[1,5-a]pyrimidine derivatives. Moreover, the PSO technology will provide a way in the selection parameters of SVM. And, the method could provide much beneficial information to investigate different skeletons of Chk1 inhibitors in the future.

2. Materials and methods

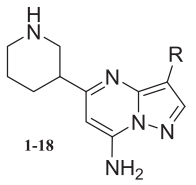
2.1. Experimental data

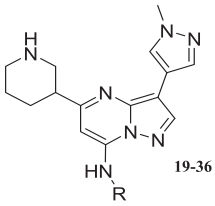
A total of 34 pyrazolo[1,5-a]pyrimidine derivatives based on Chk1 inhibitors were used for the QSAR studies. The structures and activity data of all compounds were taken from literature recently reported by

* Corresponding authors. Tel./fax: +86 0931 7971276.
E-mail addresses: jchen@nwnu.edu.cn (J. Chen), luxq@nwnu.edu.cn (X. Lu).

http://dx.doi.org/10.1016/j.chemolab.2015.10.014
0169-7439/

Table 1
The pyrazolo[1,5-α]pyrimidine derivative inhibitors of Chk1 structures and activity data.

A  1-18

B  19-36

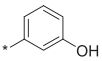
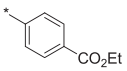
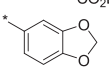
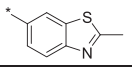
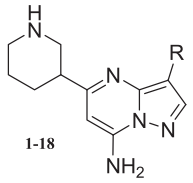
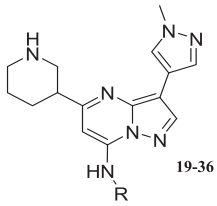
Compounds	R	pIC_{50}
1[a]	H	−1.5563
2		−1.2553
3		−1.6990
4[a]		−0.3222
5		−0.9731
6		−0.7076
7		−0.6532
8		−0.2305
9		0.5528
10[a]		0.5376
11[a]		0.2676
12		−0.0414
13		−0.1140
14[a]		0.0852
15		−0.3010
16[a]		−1.6990
17[a]		1.2218
18		1.1549
19	Me	0.8539
20	Et	0.3188
21[a]		−0.1761
22[a]		−0.6990
23		−0.8325
24[a]		0
25		0.0655
26[a]		0.3872
27		1.0605
28		1.2924
29		0.4089
30		0.2147

Table 1 (*continued*)

A  1-18

B  19-36

Compounds	R	pIC_{50}
31		−1.5315
32		−0.3424
33		1.5528
34[a]		1.5528
35		1.6778
36		2.0458

[a] Test set.

Labroli et al. [15], and listed in Table 1. In the process of QSAR studies, their biological activities pIC_{50} values were obtained by the half inhibitory concentration IC_{50}. The entire data were divided into the training set and the test set by the Kennard–Stone (KS) method. Then, the training set including 24 compounds was used to build the QSAR model, and the remaining 12 compounds were used to validate the model.

2.2. Descriptor calculation and selection

The chemical structures of the compounds were sketched by using the HyperChem software, and were optimized by the semi-empirical MP3 method until the root mean square gradient reached 0.001 kcal mol L^{-1}. All optimized structures were imported into the online E-Dragon software (http://www.vcclab.org/lab/edragon/start.html) to calculate molecular descriptors. Then, 1666 molecular descriptors were acquired in this program. In order to reduce the redundancy, the descriptors with constant or almost constant values were eliminated and some zero variables were also excluded in this work. Finally, the remaining 1352 molecular descriptors were treated as an original variable set. In order to build a reliable model, stepwise multiple linear regression method was used to select the most relevant descriptors from the pool of 1352 descriptors. For the purpose of avoiding co-linearity problems, some key descriptors were selected based on permutation and correlation matrices. Finally, four descriptors listing in Table 2 have been selected from the pool of original variable set. The contribution rate of each descriptor was shown in Fig. 1. As can be seen from Fig. 1, the contribution rate showed the relative importance of each descriptor comparing with the other descriptors.

2.3. Methods

A series of methods were used to construct QSAR model, such as multiple linear regression (MLR) [16], partial least squares (PLS) [17], artificial neural networks (ANN) [18], support vector machine (SVM) [19] and so on. In this section, the QSAR models were constructed by using SVM and PSO-SVM methods.

SVM is a novel type machine learning method based on the statistical learning theory, which can solve the practical problems of small samples and nonlinearity. Its basic idea is to map the original data into a higher dimensional feature space via a kernel function and then to