Chemometrics and Intelligent Laboratory Systems 146 (2015) 385–395

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Classification of estrogen receptor selective compounds with benzopyranskeleton using counterpropagation artificial neural networks optimised by genetic algorithms

Ying Cui [a,b,*], Jibin Shen [c], Yaxiao Li [a,b], Jianyu Li [a], Ting Zhang [d,**]

[a] Department of Medicine Chemistry, Logistics College of Chinese People's Armed Police Forces, Tianjin 300309, China
[b] Tianjin Key Laboratory of Occupational and Environmental Hazards Biomarkers, Tianjin 300309, China
[c] Department of Pharmacy, Jiangxi Provincial Corps Hospital of Chinese People's Armed Police Force, Nanchang, China
[d] Department of Pharmacy, Xiangya Hospital, Central South University, Changsha 410008, China

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form 10 June 2015
Accepted 12 June 2015
Available online 18 June 2015

Keywords:
QSAR
CP-ANN
Estrogen receptor selectivity
Benzopyran derivatives

ABSTRACT

A counter propagation artificial neural network (CP-ANN) was applied to classify the estrogen receptor selectivity of 94 benzopyrans. Molecules were represented by topostructural, topochemical, geometrical and quantum chemical descriptors. A Kohonen network was used for the rational division of the dataset into training and testing sets and for the selection of the variable, which was further reduced by correlation coefficient analysis for model construction. The most suitable network architecture of the CP-ANN was chosen using a genetic algorithm optimisation procedure for global optimisation and to avoid chance results caused by random initialisation. The optimisation procedure was developed by taking into considerable account the validation of the multivariate models. Both the percentage of correctly assigned samples for calibration and internal validation were used to generate simultaneously predictive and not overfitted models. The resulting model had a non-error rate for the training and testsets as high as 98.2% and 93.4%, respectively. It was shown that CP-ANN is a powerful tool for modelling the structure–ER selectivity relationships of the compounds considered.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Estrogen receptors (ERs), members of the nuclear hormone receptor superfamily, mediate the activity of estrogen in many different systems, including the reproductive, skeletal, cardiovascular, and central nervous systems [1,2]. However, ER stimulation of some tissues can increase the risk of cancer, particular in the female breast and uterus [3]. Thus, ER has been a target of pharmaceutical agents for hormone replacement in menopausal women and in those with uterine and breast cancers. In 1996, two different ER subtypes, the products of different genes, were discovered. These subtypes are known as ERα and ERβ [2,4]. Studies have shown that the two subtypes have different functions and distributions in certain tissues [5,6]. Molecules that selectively activate ERβ hold promise for the treatment of certain cancers, endometriosis, inflammatory and cardiovascular diseases [7]. Additionally, these molecules have a profound effect on brain development regulation and estrogen-induced promotion of neurogenesis and memory, in conjunction with reduced feminising effects [8].

Although high similarities exist in the ligand binding domain between the two subtypes, ERs constitute a serious impediment to the design and development of compounds that show high ERβ selectivity. The mobility and plasticity of the ER ligand binding domain (LBD) allow compounds of extraordinary structural diversity to mimic natural estrogen agonists or antagonists to bind to the ER subtypes [9–12]. Therefore, extensive efforts are being made to develop an estrogen receptor modulator that will selectively bind to ERβ while promoting positive estrogen effects for therapeutic purposes.

Isoflavones, such as daidzein (1a, Fig. 1), genistein (1b, Fig. 1), and the clover coumarin coumestrol (2), long known to be estrogenic (i.e., phytoestrogens), were among the first compounds noted to be ligands with a selective affinity for ERβ, a property shared by the daidzein enteric metabolite (S)-equol (3) [13,14]. Isocoumarins and their analogues, structurally related compounds of isoflavones, are high-affinity ligands that show considerable selectivity for ERβ in terms of binding affinity as well as strikingly high ERβ selectivity in terms of potency in gene transcription assays [15]. It has been observed that these derivatives bear the same functionality of benzopyrone. Furthermore, other benzopyrone derivatives have also been reported to exhibit ERβ selectivity activity [16].

The key issue in the design of new selective ER ligands is to explore the properties of chemical structure in combination with its ability to induce a pharmacological response as a consequence of receptor binding.

* Correspondence to: Y. Cui, Department of Medicine Chemistry, Logistics College of Chinese People's Armed Police Forces, Tianjin 300309, China. Tel.: +86 22 84876778.
** Corresponding author. Tel.: +86 731 5381599.
E-mail addresses: flyingyting2000@gmail.com (Y. Cui), flyingyting@aliyun.com (T. Zhang).

http://dx.doi.org/10.1016/j.chemolab.2015.06.007
0169-7439/© 2015 Elsevier B.V. All rights reserved.

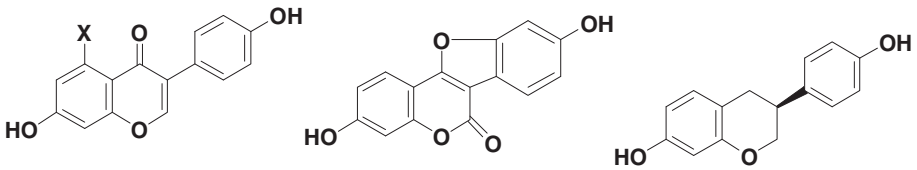

Fig. 1. Structure of various ERβ selective compounds. 1a. daidzein, 1b. genistein, 2. coumestrol, 3. (S)-equol.

However, because testing on in vitro assays may result in a tremendous financial cost and waste of time, there is an urgent need for rapid and cost-effective screening tools to detect and characterise agents with selective ER subtype binding affinity. The QSAR modelling of individual classes of chemicals has demonstrated good predictive ability and provided information concerning the mechanisms of action depending on the relevant properties or features of the chemicals [17].

Because previously confirmed experimental compounds provide an opportunity to understand the basis of subtype selectivity, the development of models for predicting the subtype selectivity would allow for the development of more potent and selective compounds for these important pharmaceutical targets. Thus, countable QSAR models are available for finding the subtype selective ligands through methods such as CoMFA [9,18], ANNs [19], MLR and PLSR [20]. 3D-QSAR techniques are generally considered to be the most effective means of predicting biological activity. However, they usually require an accurate structural superposition, which has proven to be the major bottleneck of these techniques [21,22]. Linear QSAR methods rely on a large number samples and exhibit limitations in complex model simulations [23]. Artificial neural network algorithms constitute a more flexible class of modelling techniques that can naturally model complex nonlinear systems in both classification and regression problems [24]. They also have several advantages over statistical techniques, e.g., they have the ability to continuously adapt to new data through the use of less rigid assumptions about the underlying data distribution, they allow models to be built without the knowledge of the actual modelling functions, and they provide useful information about the input and output variables that can be extracted from K-ANN and counter-propagation-ANN (CP-ANN).

In this study, a classification model is developed based on a data set consisting of 94 compounds within the benzopyran group. To maintain uniformity and minimise experimental error, the dataset was obtained from research groups, who adopted identical methodologies and experimental conditions. When dealing with classification issues, CP-ANNs are generally efficacious methods for modelling classes separated by non-linear boundaries [25,26]. This technique is able to extract the best molecular properties for the classification of compounds. Therefore, the main body of this article presents the data collected with respect to the structure of benzopyrans and reports the results of the CP-ANN model for the classification of the ERβ selective character of the compounds in the data set. Another objective of this pilot study was to obtain the best method for fast and reliable classifications of a 94 compound data set, a large number of samples. Based on this study, hypotheses on the possible nature of the ligand–target interaction are proposed.

2. Material and methods

2.1. Dataset

The dataset used in this study was adopted from the literature [10,15,16,27–31]. The compounds in the dataset belong to several families: flavones, isoflavones, chalcones, flavanones, coumarins, isocoumarins, and benzopyran (Fig. 2 and Table 1). The compounds were classified according to their selectivity for ERβ, which was calculated by determining the ratio of the binding affinities, IC_{50} for ERβ to IC_{50} for ERα, of humans in competitive radiometric binding assays. Based on the selectivity values, the data set was grouped into four classes: compounds with a selectivity ≥ 10 were assigned to class 1; compounds with a selectivity of 1–10 and 0.1–1 were assigned to class 2 and class 3, respectively; and compounds with a selectivity ≤ 0.1 were assigned to class 4.

2.2. Molecular descriptors

Molecular descriptors are quantitative representations of chemical structures and structural or physicochemical properties. The PaDEL descriptor software (National University of Singapore) was used to calculate the 2D and 3D molecular structure descriptors, which can adequately represent the structural characteristics of molecules [32]. Then, various measures were taken to eliminate the descriptors that did not contribute to the selectivity (uninformative descriptors). Overall, 228 descriptors remained after eliminating the descriptors with constant values or mostly zero values (>90%). The descriptor values were first autoscaled to range from zero to one, according to the following formula, enabling equal weighting of each descriptor regardless of its absolute value and maintaining the original distribution:

$$X_{ij}^{n} = \frac{X_{ij} - X_{j,\ min}}{X_{j,\ max} - X_{j,\ min}}$$

where X_{ij} and X_{ij}^{n} are the non-normalised and normalised j-th ($j = 1, \ldots, K$) descriptor values for compound i ($i = 1, \ldots, N$), respectively, and $X_{j,\ min}$ and $X_{j,\ max}$ are the minimum and maximum values for the j-th descriptor, respectively. Thus, for all descriptors, $\min(X_{ij}^{n}) = 0$ and $\max(X_{ij}^{n}) = 1$.

2.3. Data set split

Date set split ensures that the training set is represented in the model such that the predicted properties of the compounds in the testing set are within the statistical limits determined by such a procedure. This condition could be achieved by using the Kohonen ANN to divide the data set into two classes, based on the facts: i) the components are represented by structure descriptors; ii) the components that occupied the same neuron have similar Euclidean distances from each other. And in an other sentence, components with similar structure excite the same neuron in Kohonen map.

The Kohonen map is usually characterised as a squared toroidal space that consists of a grid of N^2 neurons, where N is the number of neurons on each side of the squared space. Each neuron contains as many elements (weights) as the number of input variables. The weights of each neuron are randomly initialised between 0 and 1 and updated based on the input vectors (i.e., samples) a certain number of times (called training epochs). The user must define both the number of neurons and epochs to train the map. In each training step, samples are presented to the network, one at a time. For each sample (xi), the most similar neuron (i.e., the winning neuron) is selected based on the Euclidean distance. Then, the weights of the r-th neuron (wr) are varied

Download English Version:

https://daneshyari.com/en/article/7563184

Download Persian Version:

https://daneshyari.com/article/7563184

Daneshyari.com