Computers in Biology and Medicine 39 (2009) 1089–1095

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

A combinatorial feature selection approach to describe the QSAR of dual site inhibitors of acetylcholinesterase

Ebrahim Barzegari Asadabadi [a], Parviz Abdolmaleki [a,*], Seyyed Mohsen Hosseini Barkooie [b], Samad Jahandideh [a,c,d], Mohammad Ali Rezaei [a]

[a] Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, P.O. Box: 14115/175, Tehran, Iran
[b] Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran
[c] Department of Medical physics, Shiraz University of Medical Sciences, Shiraz, Iran
[d] Center for research in medical physics and biomedical engineering, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Article history:
Received 3 May 2009
Accepted 12 September 2009

Keywords:
Acetylcholinesterase (AChE)
Dual (binding) site inhibitors
Feature selection
Artificial intelligence

ABSTRACT

Regarding the great potential of dual binding site inhibitors of acetylcholinesterase as the future potent drugs of Alzheimer's disease, this study was devoted to extraction of the most effective structural features of these inhibitors from among a large number of quantitative descriptors. To do this, we adopted a unique approach in quantitative structure–activity relationships. An efficient feature selection method was emphasized in such an approach, using the confirmative results of different routine and novel feature selection methods. The proposed methods generated quite consistent results ensuring the effectiveness of the selected structural features.

1. Introduction

Acetylcholinesterase (AChE) plays a vital role in the central and peripheral nervous systems. There is much evidence showing the relationship of the enzyme function and impairments with Parkinson's disease [1], Huntington disease [2], myasthenia gravis [3], schizophrenia [4], glaucoma [5], multiple sclerosis (MS) [6], and especially Alzheimer's disease (AD) [7,8]. In order to eliminate the deficiency in cholinergic function in the brain, the most promising AD therapeutic strategy has been the use of choline-like agents with the ability to enter the active site gorge of the enzyme and inhibit its cholinesterase activity. The enzyme, therefore, has long been an attractive target for the discovery of mechanism-based inhibitors for the treatment of AD. A large number of previous researches have been devoted to the extraction of new compounds with AChE inhibitory activity from different plant species [9–11], synthesis and assaying of new AChE inhibitors [12–14] and structure–activity relationship studies of AChE inhibitors [15,16].

Together with its cholinesterase activity, AChE is also known to accelerate the aggregation of beta-amyloid peptides during the early stages of AD. Amyloid fibrils form through a conformational change in the structure of amyloid precursor protein (APP) induced by peripheral anionic site of AChE. This feature of AChE, known as its non-cholinergic function, is demonstrated to be inhibited by the noncompetitive peripheral site binding inhibitors, rather than by the competitive active site ones [7].

Recently, a new generation of AChE inhibitors, namely dual binding site acetylcholinesterase inhibitors, has been designed with the ability to bind both the catalytic site of the enzyme, inhibiting its cholinesterase activity, and the peripheral site of the enzyme, inhibiting its promotion of amyloid fibril formation. A series of inhibitors with this mechanism of action was synthesized by Munoz-Ruiz and colleagues with much greater inhibitory activities compared with all previously synthesized and examined inhibitors [17]. The great ability of these compounds was demonstrated through both experimental and molecular dynamics simulation studies. However, what structural features of these compounds really determine their inhibition activity still remains unclear and obtaining knowledge of exact determinants of such a bioactivity seems extremely helpful and essential to design new effective AD drugs.

Due to the limited number of dual binding site acetylcholinesterase inhibitors, to date synthesized, a comprehensive structural study on these compounds has not been possible. Furthermore, the majority of previous structure–activity relationship studies on conventional AChE active site inhibitors have sought these relationships only among their visually detected structural properties without considering the quantitative aspects of the problem. Numerous mathematical modeling approaches, however, could help extract the most effective quantitative structural parameters, namely descriptors, which determine the

* Corresponding author. Fax: +98 21 88004484.
E-mail address: parviz@modares.ac.ir (P. Abdolmaleki).

0010-4825/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2009.09.003

inhibition activity of the molecules. Exploiting these capabilities, this study aims to develop quantitative models to describe the structure–activity relationships in dual binding site inhibitors of acetylcholinesterase, as well as to introduce the structural determinants of their bioactivity.

2. Materials and methods

2.1. Dataset

The dataset used in the present study was described by Munoz-Ruiz and colleagues [17], including 24 dual binding site AChE inhibitors with a unique core structure depicted in Fig. 1. As can be seen, the structure pattern contains a tacrine and an indole ring linked together by a hydrocarbon linker chain, allowing a dual mechanism of inhibition to block the active and peripheral sites of the enzyme, respectively. The experimental index of IC_{50} was also reported for every compound as a measure of inhibition (Table 1). The greater is this value, the weaker is the inhibition activity of the compound. In the classification stages, classifier models with the ability to solve two-class problems, which also were able to select the effective parameters were developed and trained. With this purpose, compounds were labeled as active for $IC_{50} < 2$ nM and inactive for $IC_{50} > 2$ nM (Table 1).

2.2. Database preparation

Structures of all compounds were drawn in HyperChem (Hypercube Inc.). Geometrical optimization was then carried out using the semi-empirical method of Austin Model 1 (AM1) [18] which used the conjugate gradient algorithm of Fletcher–Reeves for structure energy minimization. Optimized structures were used to generate different types of structural descriptors. These descriptors are categorized in groups of constitutional, topological, Gálvez topological charge indices [19], empirical, molecular walk counts [20], atom-centered fragments [21], 2D autocorrelations [22], BCUT [23], properties and functional group descriptors.

The total 758 descriptors extracted from compounds were too many to be fitted to our models. In order to reduce this number, an objective feature selection was carried out in three filtration steps, as follows: firstly, descriptors with constant values for all compounds and/or those with constant values for more than 80% of compounds within the dataset were removed. Then, descriptors with correlation coefficient less than 0.25 with dependent variable IC_{50} were removed from the database. At the third stage, pairwise Pearson's correlations were calculated for the remained descriptors and if their correlation coefficient exceeded 0.85, the one with lower correlation with the target was removed. The number of descriptors was reduced to 60 after the mentioned filtration steps.

2.3. Feature selection using linear discriminant analysis (LDA)

As a powerful tool, linear discriminant analysis has been widely used in many application areas, such as medicine, biology, chemistry, finance, and engineering [24–26]. The basic theory of LDA is to classify the dependent variable by dividing an *n*-dimensional descriptor space into two regions that are separated by a hyperplane defined by a linear discriminant function. LDA can be used to build a predictive model of the group membership based on observed characteristics of each case. This procedure generates a discriminant function based on linear combinations of

Table 1
Structure, inhibition potency and activity class of dual binding site inhibitors.[a]

Compound	R1	R2	R3	X	Y	Z	IC_{50} (nM)	Activity class
3	H	H	H	$(CH_2)_5$	$(CH_2)_2$	CH	70	0
4	Cl	H	H	$(CH_2)_5$	$(CH_2)_2$	CH	4	0
5	Cl	H	H	$(CH_2)_6$	$(CH_2)_2$	CH	0.02	1
6	Cl	H	H	$(CH_2)_7$	$(CH_2)_2$	CH	0.06	1
7	Cl	H	H	$(CH_2)_8$	$(CH_2)_2$	CH	0.5	1
8	Cl	H	H	$(CH_2)_9$	$(CH_2)_2$	CH	4.4	0
9	Cl	H	H	$(CH_2)_{10}$	$(CH_2)_2$	CH	21.9	0
10	H	H	H	$(CH_2)_3NMe(CH_2)_3$	$(CH_2)_2$	CH	147	0
11	Cl	H	H	$(CH_2)_3NMe(CH_2)_3$	$(CH_2)_2$	CH	2.9	0
12	Cl	CN	H	$(CH_2)_6$	$(CH_2)_2$	CH	0.7	1
13	Cl	H	H	$(CH_2)_6$	(CH═CH)	CH	18	0
14	Cl	H	H	$(CH_2)_5$		CH	180	0
15	Cl	H	H	$(CH_2)_6$		CH	33	0
16	Cl	H	H	$(CH_2)_7$		CH	36	0
17	Cl	H	H	$(CH_2)_8$		CH	46	0
18	Cl	H	H	$(CH_2)_7$	CH_2	CH	0.2	1
19	Cl	Br	H	$(CH_2)_7$	CH_2	CH	0.6	1
20	Cl	H	H	$(CH_2)_5$	$(CH_2)_3$	CH	0.3	1
21	Cl	H	H	$(CH_2)_6$	$(CH_2)_3$	CH	0.5	1
22	Cl	H	Me	$(CH_2)_7$		CH	10.9	0
23	Cl	H	H	$(CH_2)_7$		N	95	0
27	Cl	H	H	$(CH_2)_5$	$(CH_2)_2O$	CH	1.5	1
28	Cl	H	H	$(CH_2)_6$	$(CH_2)_2O$	CH	0.7	1
29	Cl	H	H	$(CH_2)_7$	$(CH_2)_2O$	CH	3.0	0

[a] For substitution locations, see Fig. 1.

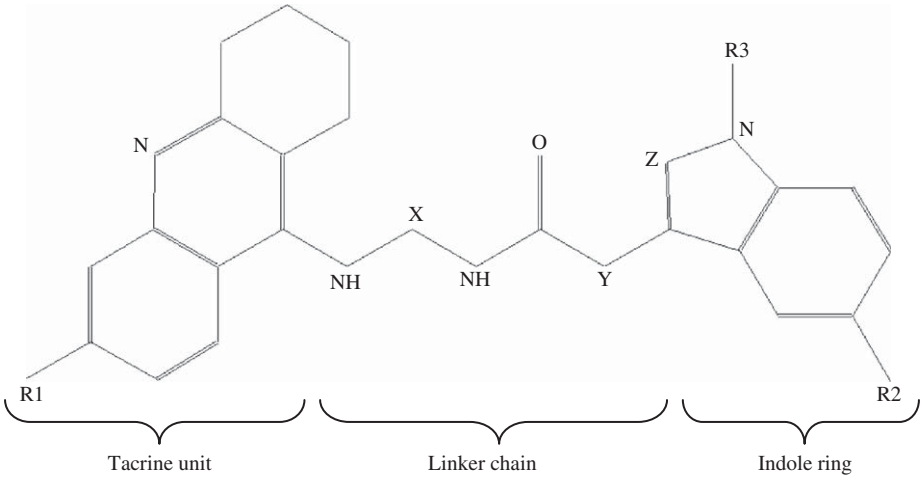

Fig. 1. The core structure of acetylcholinesterase dual binding site inhibitors.