Computers in Biology and Medicine 64 (2015) 148–154

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Searching therapeutic agents for treatment of Alzheimer disease using the Monte Carlo method

Mariya A. Toropova [a], Andrey A. Toropov [b,*], Ivan Raška Jr[c], Mária Rašková [c]

[a] Dipartimento di Scienze Farmaceutiche, Universita' degli Studi di Milano, Via L. Mangiagalli, 25, 20133 Milan, Italy
[b] IRCCS, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via La Masa 19, Milano 20156, Italy
[c] 3rd Department of Medicine - Department of Endocrinology and Metabolism, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, U Nemocnice 1, 12808 Prague 2, Czech Republic

ARTICLE INFO

Article history:
Received 9 May 2015
Accepted 20 June 2015

Keywords:
QSAR
Gamma-secretase inhibitor
Monte Carlo method
CORAL software
OECD principles

ABSTRACT

Quantitative structure - activity relationships (QSARs) for the pIC50 (binding affinity) of gamma-secretase inhibitors can be constructed with the Monte Carlo method using CORAL software (http://www.insilico.eu/coral). The considerable influence of the presence of rings of various types with respect to the above endpoint has been detected. The mechanistic interpretation and the domain of applicability of the QSARs are discussed. Methods to select new potential gamma-secretase inhibitors are suggested.

1. Introduction

Alzheimer's disease is a neurodegenerative disorder of the central nervous system accompanied by degradation of cognitive abilities and memory deterioration, together with a variety of neuropsychiatric symptoms, behavioral disturbances, and progressive impairment of daily life activities. Current pharmacotherapies are restricted to symptomatic interventions but do not prevent progressive neuronal degeneration. Therefore, new therapeutic strategies are needed to intervene with these progressive pathological processes [1].

Two major pathological hallmarks are characteristics of Alzheimer's disease: intracellular neurofibrillary tangles and extracellular amyloid plaques. The amyloid plaque is mainly comprised of an aggregated form of the 40–42 residue amyloid β-peptide (Aβ). The accumulation and deposition of Aβ eventually lead to neuronal damage and cell death. Reduction of Aβ by inhibition of γ-secretase may prevent the above neurotoxic events, representing an attractive strategy to reduce the probability of Alzheimer's disease [2].

Gamma-secretase inhibitors are possible therapeutic agents for treatment of Alzheimer disease and cancer [2]. The measure of therapeutic potential of different gamma-secretase inhibitors is their binding affinity [3]. There are a number of attempts to build up a model for the endpoint (binding affinity) by means of various approaches. For instance, in work [1], the partial least squares (PLS) regression and neural networks (NN) were utilized to build up a model for the endpoint. The aim of the present work is to develop quantitative structure – activity relationships (QSARs) related to the above-mentioned endpoint using the CORAL software [4]. In fact, the CORAL software is a tool to build up a model for arbitrary endpoint using the Monte Carlo technique. The comparison of the predictability of the above mentioned approaches (PLS and NN [1]) with the CORAL models can be interesting and useful from theoretical and practical point of view.

2. Method

2.1. Data

The binding affinity data (IC_{50} nM converted into negative decimal logarithm $pIC_{50} = -\log_{10} IC_{50}$) of 233 gamma-secretase inhibitors and their simplified molecular input-line entry system (SMILES) [4] were taken in the literature [2,5]. Three random splits into the training ($\approx 60\%$), calibration ($\approx 20\%$) and validation ($\approx 20\%$) sets of the above-mentioned 233 inhibitors were generated and examined in this work. The training set plays the role of builder of a model; the calibration set plays the role of preliminary critic of the model; and the validation set is the final estimator of the model.

* Correspondence to: Laboratory of Environmental Chemistry and Toxicology, IRCCS - Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy. Tel.: +39 02 39014595; fax: +39 02 39014735.
E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

http://dx.doi.org/10.1016/j.compbiomed.2015.06.019
0010-4825/

2.2. Optimal descriptors

The optimal descriptors which are using for QSPR/QSAR analyses [6–8] are mathematical functions of so-called correlation weights. In other words, they are "Descriptors of Correlation Weights" (DCW). The correlation weights are calculated using the Monte Carlo method: the numerical data on the correlation weights must give maximum correlation coefficient between the experimental endpoint values with the DCW for visible training set [6–8]. Two versions of the optimal descriptors calculated with the hydrogen filled graphs (HFGs) were examined:

$$DCW_1(T,N) = \sum_{k=1}^{NG} W(EC1_k) + \sum_{k=1}^{NG} W(PT2_k) + \sum_{k=1}^{NG} W(VS2_k) \quad (1)$$

$$DCW_2(T,N) = \sum_{k=1}^{NG} W(EC1_k) + W(C3) + W(C4) + W(C5) + W(C6) + W(C7) \quad (2)$$

where

$W(EC1_k)$ is the correlation weight for the presence of a given value of extended connectivity of the first order in HFG [6]; $W(PT2_k)$ is the correlation weight for the presence of a given number of paths of length 2 which started from the k-th vertex in HFG [6]; $W(VS2_k)$ is the correlation weight for the presence of a given valence shell value of second order [6]. The listed graph invariants are calculated with the adjacency matrix of the HFG. The adjacency matrix of a graph G with n vertices (which represent chemical elements) is the $n \times n$ matrix where the non-diagonal element a_{ij} is 1 if i-th and j-th vertices are connected (i.e. i-th and j-th atoms are connected by covalent bond); otherwise, the a_{ij} is zero. Fig. 1 contains the example of the adjacency matrix together with numerical values of the above graph invariants.

In the case of second version of the optimal descriptor calculated with Eq. (2) an other group of graph invariants were used. These are calculated in accordance with presence (absence) of three-members cycles (C3); four-members cycles (C4), five-members cycles (C5), six-members cycles (C6), and seven-members cycles (C7). The $W(C3)$, $W(C4)$, $W(C5)$, $W(C6)$, and $W(C7)$ are correlation weights for graph invariants related to cycles. Table 1 contains examples of listed graph invariants. In fact, each graph invariant is a molecular feature. The $EC1$, $PT2$, $VS2$ are examples of local molecular features (fragments), whereas $C3$ – $C7$ are global molecular features since they characteristics of a molecule totally.

Finally, the NG is the number of vertex in HFG; T and N are parameters of the Monte Carlo optimization. The optimization is aimed to give the maximal correlation coefficient between $DCW_1(T,N)$ or $DCW_2(T,N)$ and the endpoint for the training set.

The T is the threshold i.e. a coefficient to discriminate molecular features extracted from HFG into two categories (i) rare in

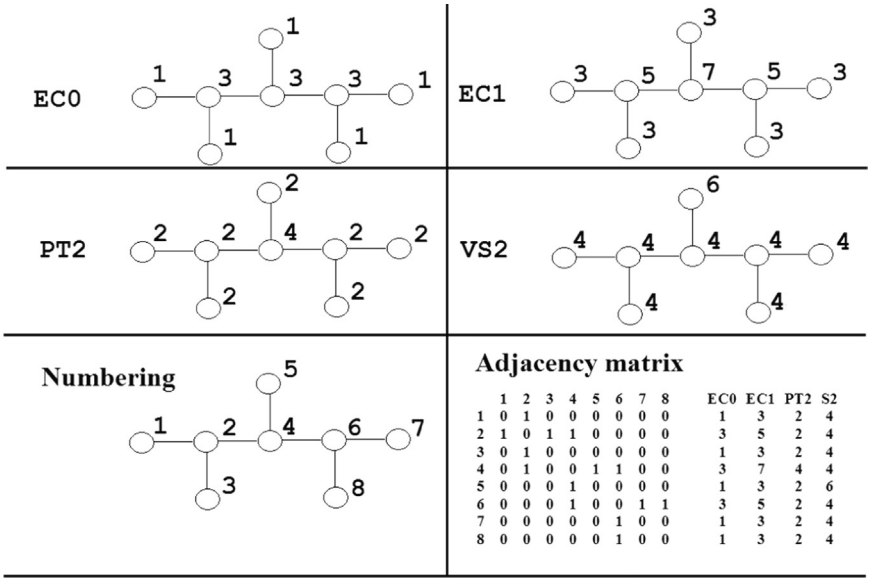

Fig. 1. An example of calculation of graph invariants with the adjacency matrix.

Table 1
An examples of codes which are used to represent graph invariants involved into the optimal descriptors calculated with Eq. (1) or with Eq. (2).

Twelve symbols' code	Comment
EC1-C...5...	The extended connectivity of first order equal to 5 for carbon atom
EC1-F...3...	The extended connectivity of first order equal to 3 for fluorine atom
EC1-S...9...	The extended connectivity of first order equal to 9 for sulfur atom
PT2-Cl..2...	The path of length 2 equal to 2 for chlorine atom
PT2-N...5...	The path of length 2 equal to 5 for nitrogen atom
VS2-C...7...	The valence shell of second order equal to 7 for carbon atom
VS2-H...9...	The valence shell of second order equal to 9 for hydrogen atom
C3.....1...	The presence of one ring with three members
C4....H.1...	The presence of one ring with four members containing at least one hetero atom (non carbon)
C5....H.2.	The presence of two rings with five members, at least one of these two rings contains hetero atoms.
C6. .A..3.	The presence of three rings with six members, at least one of these rings is aromatic one
C6...AH.4...	The presence of four rings with six members, at least one of these rings is aromatic, and at least one of these rings contains heteroatoms.