European Journal of Pharmaceutical Sciences 111 (2018) 215–225

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

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

Binding affinity toward human prion protein of some anti-prion compounds — Assessment based on QSAR modeling, molecular docking and non-parametric ranking

Strahinja Kovačević*, Milica Karadžić, Sanja Podunavac-Kuzmanović, Lidija Jevrić

University of Novi Sad, Faculty of Technology Novi Sad, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia

ARTICLE INFO

Keywords:
Biological activity
Chemometrics
In silico ADMET
Molecular docking
Prions
Sum of ranking differences

ABSTRACT

The present study is based on the quantitative structure-activity relationship (QSAR) analysis of binding affinity toward human prion protein ($huPrP^C$) of quinacrine, pyridine dicarbonitrile, diphenylthiazole and diphenyloxazole analogs applying different linear and non-linear chemometric regression techniques, including univariate linear regression, multiple linear regression, partial least squares regression and artificial neural networks. The QSAR analysis distinguished molecular lipophilicity as an important factor that contributes to the binding affinity. Principal component analysis was used in order to reveal similarities or dissimilarities among the studied compounds. The analysis of *in silico* absorption, distribution, metabolism, excretion and toxicity (ADMET) parameters was conducted. The ranking of the studied analogs on the basis of their ADMET parameters was done applying the sum of ranking differences, as a relatively new chemometric method. The main aim of the study was to reveal the most important molecular features whose changes lead to the changes in the binding affinities of the studied compounds. Another point of view on the binding affinity of the most promising analogs was established by application of molecular docking analysis. The results of the molecular docking were proven to be in agreement with the experimental outcome.

1. Introduction

The speculations about the nature of an infectious agent that causes serious damage of the brain tissue started in the second half of XX century (Ridley, 2001). The end of those speculations was the established "*prion hypothesis*" (Ridley, 2001; Prusiner, 1982). Generally speaking, the prions are considered to be the protein-structured infectious agents, with ability to cause different neurodegenerative lethal diseases. However, prions are small membrane-associated protein molecules, present in the cells of mammals, including humans, with the biological function that still remains controversial. In humans, these so-called cellular prions (PrP^C) are composed of 253 amino acids. They are copper-binding proteins that primarily have certain roles in copper (Cu^{2+}) metabolism (related to the prevention of oxidative stress) (Kretzschmar et al., 2001; Brown et al., 1997; Brown, 1999). The PrP^C prion contains three α-helixes and two short antiparallel β-sheets, where two longest α-helixes are connected with a single S—S bond. A prion becomes the infectious agent when its conformation is irreversibly changed into the conformation with higher β-sheet content, causing the formation of extracellular and intracellular agglomerates (Hyeon et al., 2015). This type of prions is labeled as a scrapie form — PrP^{Sc}. The PrP^{Sc} catalyzes misfolding of PrP^C into new PrP^{Sc} forms. Hence, the infective prion agent is actually protease-resistant PrP^{Sc} form. The agglomeration of the prion proteins causes the damage in the brain tissue inducing the cells death and finally the death of an individual. The conversion of PrP^C into PrP^{Sc} form induces the transmissible spongiform encephalopathies (TSEs) that are fatal neurodegenerative diseases, such as bovine spongiform encephalopathy (BSE),

Abbreviations: ADMET, absorption, distribution, metabolism, excretion and toxicity; ANN, artificial neural networks; ANNR, artificial neural networks regression; BBB, blood-brain barrier; BFGS, Broyden-Fletcher-Goldfarb-Shanno algorithm; BSE, bovine spongiform encephalopathy; CJD, Creutzfeldt–Jakob disease; CNS, central nervous system; GLN, glutamine; GLU, glutamic acid; GLY, glycine; GSA, global sensitivity analysis; HCA, hierarchical cluster analysis; HIA, human intestinal absorption; LV, latent variable; MDCK, Madin-Darby canine kidney cells; MLP, multi-layer perceptron; MLR, multiple linear regression; MMFF, molecular mechanics force field; NIPALS, nonlinear iterative partial least squares algorithm; PC, partition coefficient; PCA, principal component analysis; PLS, partial least squares; PLSR, partial least squares regression; PPB, plasma protein binding; PRESS, predicted residual sum of squares; PrP^C, prion protein (cellular); PrP^{Sc}, prion protein (scrapie); QSAR, quantitative structure–activity relationship; RMSE, root mean square error; SER, serine; SP, skin permeability; SPR, surface plasmon resonance; SRD, sum of ranking differences; SS, stepwise selection; TCR, Topliss-Costello rule; TSE, transmissible spongiform encephalopathy; TSS, total sum of squares; TYR, tyrosine; ULR, univariate linear regression; VC, variation coefficient; VIF, variance inflation factor; VIP, variable importance in projection

* Corresponding author.
E-mail address: strahko@uns.ac.rs (S. Kovačević).

http://dx.doi.org/10.1016/j.ejps.2017.10.004
Received 23 July 2017; Received in revised form 15 September 2017; Accepted 3 October 2017
Available online 04 October 2017
0928-0987/

Creutzfeldt–Jakob disease (CJD), Gerstmann–Sträussler–Scheinker syndrome, fatal familial insomnia and kuru. The causes of the PrP^C prion misfolding can be various. Spontaneous change in conformation leads to sporadic prion diseases. However, in the familial form the mutation in the prion protein gene occurs. The induced (acquired) misfolding occurs when the abnormal prions from the environment reach the central nervous system (CNS). About 85% cases of CJD is sporadic, and 10–15% is familial or induced (Marković and Marković, 2004). The large amounts of misfolded isoforms of prions (PrP^{Sc}) can be observed in the brains of affected individuals. Because of their infectious nature and the resistance to physical and chemical inactivation, the prions can be considered to be the environmental pollutants, and ultimately the borderline between chemistry and biology (Ridley, 2001).

Unfortunately, there are still no effective therapeutics for TSEs treatment, however many *in vivo, in vitro* and *in silico* trials in the World are aimed to find the lead compounds. It has been determined that certain large molecules, *i.e.* congo red and suramin, and small molecules, such as quinacrine, can express anti-prion activity by inhibiting the conversion of PrP^C into PrP^{Sc} (Hyeon et al., 2015). As it was described in literature (Ghaemmaghami et al., 2009), quinacrine, as an antimalarial compounds, has failed to extend the lifespan of prion-infected animals, however quinacrine analogs with more potent anti-prion activity have been synthesized (Nguyen et al., 2011). The pyridine dicarbonitrile derivatives have been shown as significant PrP^C binders as well (Guo et al., 2008). Some synthesized analogs of 2,4-diphenylthiazoles and 2,4-diphenyloxazoles also expressed binding affinity to PrP^C proteins and have been presented as compounds of considerable interest for further development (Heal et al., 2007).

Chemometrics and computational modeling of molecules have become a significant factor in drug design workflow (Cherkasov et al., 2014; Vračko et al., 2010; Kovačević et al., 2014). In many cases chemometric tools have facilitated the search for lead compounds in many drug discovery spheres (Cherkasov et al., 2014; Coluccia et al., 2016; Paul et al., 2016; Kovačević et al., 2016a; Minovski and Šolmajer, 2010; Tenorio-Borroto et al., 2014). Quantitative structure–activity relationships (QSARs) are based on validated mathematical models and have certain advantage over classical structure–activity relationship (SAR) approach. The QSAR method is able to reveal some factors important for biological response, hidden in the form of molecular descriptors, which sometimes cannot be noticed by simple observation and comparison of the molecular structure and biological activity values. QSAR is based on many different mathematical and statistical methods, including univariate linear regression (ULR) and multiple linear relationships (MLR), principal component regression (PCR), partial least squares regression (PLSR), artificial neural networks regression (ANNR), hierarchical cluster analysis (HCA), principal component analysis (PCA), sum of ranking differences (SRD) (Cherkasov et al., 2014; Miller and Miller, 2010; Héberger and Kollár-Hunek, 2011). The present study is focused on the establishing the QSARs for three aforementioned groups of the compound with anti-prion activity and/or binding affinity toward human PrP^C ($huPrP^C$) protein: quinacrine analogs (the group I), pyridine dicarbonitrile analogs (the group II), diphenylthiazole and diphenyloxazole analogs (the group III). This study presents the possibility of the application of chemometric regression tools in prediction of binding affinity of the analyzed compound by using molecular descriptors. Eventually, the compounds with the highest binding affinity have been subjected to the molecular docking analysis in order to predict the basic molecular interactions between them and $huPrP^C$.

2. Material and methods

2.1. The investigated compounds and their binding affinity toward $huPrP^C$

The investigated compounds have been divided into three groups. The group I includes 25 quinacrine analogs with basic phenyl residues at the 9-amino position (Nguyen et al., 2011). The group II covers 26 pyridine dicarbonitrile analogs (Guo et al., 2008), while the group III is consisted of 15 2,4-diphenylthiazole and 2,4-diphenyloxazole analogs (Heal et al., 2007). The binding affinity data was taken from the ChEMBL (European Molecular Biology Laboratory) database (Bento et al., 2014) and the corresponding literature (Nguyen et al., 2011; Guo et al., 2008; Heal et al., 2007). The structures of the investigated compounds are available in Supplementary data A (Table S1). The representative structures are given in Fig. 1. The data retrieved from the ChEMBL data base are presented in Supplementary data B. The binding affinity of the studied compounds was expressed in maximum response units (RU_{max}) and/or in normalized maximum response units (RU $\%_{max}$). The binding affinity of the group I is defined as the binding affinity to truncated $huPrP_{121–231}$ determined by the surface plasmon resonance (SPR) assay (Nguyen et al., 2011). The binding affinity of the group II and the group III is defined as the binding affinity to full length $huPrP^C$ determined by SPR assay (Guo et al., 2008; Heal et al., 2007).

2.2. In silico modeling of the molecular features — molecular descriptors calculations

The molecular descriptors have been calculated on the basis of 1D, 2D and 3D molecular structures. The 2D structures have been drown in MarvinSketch v.15.3.23 (ChemAxon software, http://www.chemaxon.com/) and ChemBioDraw Ultra v.12.0 (ChemBioOffice, 2012, http://www.cambridgesoft.com/) programs. The 3D molecular descriptors have been calculated on the basis of 3D molecular structures subjected to energy minimization by using molecular mechanics force field (MMFF) method — MM2. Full geometry optimization was carried out until the root mean square gradient reached a value smaller than 0.0001 kcal/Å mol. Lipophilicity, physico-chemical, topological and ADMET molecular descriptors have been calculated by using the following software: MarvinSketch v.15.3.23, ChemBioDraw Ultra v.12.0, ChemBio3D Ultra v.12.0, ALOGPS 2.1 (VCCLAB, n.d., http://www.vcclab.org/), Molinspiration online program (Molinspiration Cheminformatics, n.d., http://www.molinspiration.com/) and PreADMET online program (PreADMET Software, https://preadmet.bmdrc.kr/). The list of the calculated descriptors for all three groups of compounds is presented in Supplementary data C.

2.3. Chemometric methods

The present study is based on the definition of the possible relationships between the molecular structure described by molecular descriptors and binding affinity toward $huPrP^C$ protein. Therefore, the regression chemometric methods have been used: ULR, MLR, PLSR and ANNR. ULR is the simplest regression method used in this study, taking into account only one independent variable (molecular descriptor) in correlation with one dependent variable (binding affinity, RU_{max} or RU $\%_{max}$). In MLR, there are more than one independent variable which are not mutually significantly correlated (intercorrelated). The intercorrelation is checked by the variance inflation factor (*VIF*). It the *VIF* is equal or < 10, the multicollinearity can be neglected (Young et al., 2008). PLSR is a multivariate regression method applied when there is a considerable degree of correlation between the independent variables, but the extra weight is given to the variables highly correlated with the response variable (Miller and Miller, 2010). In PLS modeling, latent variables (LV) are calculated for dependent and independent variable matrices, and the relationship between them is estimated. PLS models are interpreted in a similar way as PCA and PCR models. Presenting the *X*-loadings and *Y*-loadings in the same plot allows to study the relationships between variables, including the relationships between *X* and *Y* variables (Esbensen, 2009). PLS approach is a very useful tool in detection of erroneous measurements (outliers) and interferences, and it enables the interpretation of many generic relationships between