

RESEARCH

Open Access

QSAR analysis on tacrine-related acetylcholinesterase inhibitors

Kai Y Wong¹, Andrew G Mercader², Laura M Saavedra³, Bahareh Honarparvar^{4*}, Gustavo P Romanelli³ and Pablo R Duchowicz^{2*}

Abstract

Background: The evaluation of the clinical effects of Tacrine has shown efficacy in delaying the deterioration of the symptoms of Alzheimer's disease, while confirming the adverse events consisting mainly in the elevated liver transaminase levels. The study of tacrine analogs presents a continuous interest, and for this reason we establish Quantitative Structure-Activity Relationships on their Acetylcholinesterase inhibitory activity.

Results: Ten groups of new developed Tacrine-related inhibitors are explored, which have been experimentally measured in different biochemical conditions and AChE sources. The number of included descriptors in the structure-activity relationship is characterized by 'Rule of Thumb'. The 1502 applied molecular descriptors could provide the best linear models for the selected Alzheimer's data base and the best QSAR model is reported for the considered data sets.

Conclusion: The QSAR models developed in this work have a satisfactory predictive ability, and are obtained by selecting the most representative molecular descriptors of the chemical structure, represented through more than a thousand of constitutional, topological, geometrical, quantum-mechanical and electronic descriptor types.

Keywords: Alzheimer's disease, Acetylcholinesterase Inhibitor, QSAR theory, Tacrine, Validation

Background

Alzheimer's Disease (AD) is a neurodegenerative process characterized by a progressive memory loss, decline in language skills and other cognitive abilities [1]. It is common among the elderly, affecting around 7% of the population above 65 years old [2]. Currently, this is an incurable disease without an effective therapeutic approach [3]. However, there exists a palliative strategy which enhances cholinergic transmission, as defined by the cholinergic hypothesis [4,5]. Patients with AD experience a selective loss of cholinergic neurons in the brain and decreasing levels of Acetylcholine, a neurotransmitter [6]. The Acetylcholinesterase enzyme (AChE) is responsible for terminating impulse signaling at cholinergic synapses by catalyzing the hydrolysis of Acetylcholine [7].

Tacrine (Additional file 1, compound 1) is an example of Acetylcholinesterase inhibitor (AChEI). It has been synthesized more than forty years ago, and in 1993, it has become the first drug to be approved by the US Food and Drug Administration as a form of treatment for AD [8]. Although Tacrine is the oldest palliative drug designed based on the cholinergic hypothesis, new Tacrine derivatives are still being designed to treat AD. It has been experimentally demonstrated that bis-Tacrine congeners display enhanced inhibitory activity towards AChE compared to Tacrine [9-11].

Quantitative Structure-Activity Relationships (QSAR) are mathematical frameworks which link molecular structures of compounds to their biological activities in a quantitative manner [12-14]. Although these tend to be statistical models rather than deterministic ones based on fundamental physical laws, the QSAR approach highlights parallelisms between structure and potency. Such knowledge allows researchers to pick out the most promising structures from a much larger pool of potential compounds, making drug design more rational by

* Correspondence: Honarparvar@ukzn.ac.za; pabloducho@gmail.com

⁴School of Pharmacy and Pharmacology, University of KwaZulu-Natal, 4001 Durban, South Africa

²Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diag. 113 y 64, Sucursal 4, C.C. 16, 1900 La Plata, Argentina

Full list of author information is available at the end of the article

reducing the number of expensive, time consuming experiments.

Among the QSAR research carried out on Tacrine derivatives, in 2006 Akula et al. [15] have published 3D-QSAR studies on bis-tacrine compounds by using molecular docking scores, in addition to Comparative Molecular Field Analysis (CoMFA). The Sybyl [16] and Mopac [17] programs are both used in the optimization of structures and molecular alignment. The docking scores are used as molecular descriptors along with the steric and electrostatic field values obtained from CoMFA, and 16 molecules are set aside in the training set. The structure-activity model is validated on a test set having 3 molecules. The proposed models provide insights into the inhibitory mechanism and could help designing new inhibitors. In another study of the same year, Fernández et al. [18] have applied Bayesian-Regularized Genetic Neural Networks (BRGNNs) to 136 Tacrine analogues. Here, the Bayesian-regularization avoids overtraining, while the Genetic Algorithm (GA) approach allows exploring an ample pool of 3D-descriptors generated by the Dragon software [19]. The resulting model is evaluated by averaging multiple validation sets generated as members of diverse-training set Neural Network Ensembles (NNEs). When considering 40 assembled members, the NNE provides reliable statistics.

In 2007, Jung et al. [20] have built QSAR models using variable selections based on Multivariable Linear Regression (MLR) approaches: Genetic Algorithms (GA) and Simulated Annealing (SA). The authors compile a set of 80 structurally heterogeneous compounds composed of 11H-indeno-[1,2-b]-quinolin-10-ylamine derivatives, thio-pyranoquinolines, pyranoquinolines and benzonaphthyrindines, tacrine-E2020 hybrids, bis-tacrine congeners, and tacrine-hurprine heterodimers, placing 68 Tacrine derivatives in the training set and leaving 12 in the test set. The molecules are geometrically optimized using the Titan Pro software [21], and their molecular descriptors are calculated using PreADME on the web [22] and BioMedCACHe [23]. The best model is obtained by SA-MLR with greater explanatory and prediction capability. The authors suggest important roles for hydrophobic and electrostatic interactions on increasing the structure's AChE activity. They also argue opposite effects for hydrophilic and topological features of molecules.

Saracoglu et al. [24] have performed in 2008 QSAR analyses of AChEIs related to Tacrine and 11 H-Indeno-[1,2-b]-quinolin-10-ylamine tetracyclic Tacrine analogues. The Electron-Topological Method (ETM) is applied with the ETM software [25] on a training set of 44 compounds, which we consider it as a valuable QSAR tool as this technique takes into account both geometrical and electronic characteristics of the molecules. Based on pharmacophores and anti-pharmacophores calculated as sub-

matrices containing spatial and quantum chemistry characteristics, a system for the activity prognostication is developed. Some molecular fragments specific for active and inactive compounds are also revealed.

In a recent work of 2012, Chen et al. [26] have studied multi-target-directed AChEIs of Tacrine-Nimodipine dihydropyridines. They establish 3D-QSAR models using CoMFA and Comparative Molecular Similarity Index Analysis (CoMSIA) methods. The compounds employed are very potent and selective AChEIs, and show an excellent neuroprotective profile and a moderate Ca^{2+} channel blockage effect. A training set of 60 compounds is used, and the resulting models are validated on a test set of 12 compounds. The structures of the investigated ligands are built and optimized using Sybyl [16], while the lowest energy structures are used during the alignment. The partial atomic charges required to estimate the electrostatic interaction are computed by semi-empirical molecular orbital methods using Mopac [17] with an AM1 Hamiltonian. For the CoMSIA approach, descriptors of five physicochemical field properties are used to correlate with changes of ligands affinities, which explicitly define hydrophobic, hydrogen-bond donor and acceptor descriptors (in addition to the steric and electrostatic fields used in CoMFA).

It is notable that most published QSAR studies that analyze a higher number of Tacrine compounds involve inhibitory activities measured on AChE samples extracted from different sources [18,20]. In this work, we complement the research performed previously by considering quantitative structure-activity relationships for ten homogeneous classes of new developed Tacrine compounds, each class measured under different biochemical conditions and extracted from a different AChE source. Therefore, present QSAR study is specific for each data set. In such a situation of having a scarce amount of experimental data for each AChEI group, the combination of the MLR technique and the 'Rule of Thumb' [27] has proven to be appropriate for developing predictive multi-parametric QSAR models [28-31].

Methods

Experimental AChE inhibitory activity

The ten classes of new synthesized Tacrine-like inhibitors have been measured by varying the temperature, incubation time and number of tested concentrations, and using different sources of AChE. AChE inhibitory activity is spectrophotometrically evaluated by the Ellman method [32], using AChE from bovine (bAChE) or human erythrocytes (hAChE) and Acetylthiocholine iodide (0.53 or 0.13 mM for bAChE and hAChE, respectively) as substrate. AChE activity in *Electrophorus electricus* (EeAChE) is measured following the spectrophotometric Rappaport method [33] using purified AChE from Electric

eel (*E. electricus*) and Acetylcholine chloride (29.5 mM) as substrate. The inhibitory activity is presented as half of the maximal inhibitory concentration in molar units ($\log_{10}(IC_{50})$).

The data sets of Additional file 1 are composed as follows: Data set A includes 7 pyrano[2,3-*b*]quinolines, 6 benzonaphthyridines and Tacrine, as in reference [34], with measured values of bAChE (Additional file 2: Table S1-S7). Data set B includes 2 hydrochlorides of 6-chlorotacrine and tricyclic esters, Propidium iodide, 10 dihydrochlorides of the pyrano[3,2-*c*]quinoline-6-chlorotacrine and 6-chlorotacrine, with values of bAChE and hAChE (Additional file 2: Table S2) [35]. Data set C consists on 17 Tacrine-dihydropyridine hybrids, Tacrine and Propidium, presenting measured values of EeAChE and hAChE (Additional file 2: Table S3) [36]. Data set D is composed of 8 donepezil-Tacrine hybrid related derivatives, Tacrine and Donepezil, having measured values of bAChE (Additional file 2: Table S4) [37]. Data set E consists on 12 Tacrine-based dimers, *bis*(7)-Tacrine and Tacrine, presenting hAChE measured values (Additional file 2: Table S5) [38]. Data set F includes 12 Tacrine analogs from highly substituted 2-aminopyridine-3-carbonitriles and Tacrine, having measured values of bAChE and EeAChE (Additional file 2: Table S6) [39]. Data set G contains 18 Tacrine-based dual binding site Acetylcholinesterase inhibitors, Tacrine, 6-chloroTacrine, Propidium iodide and Donepezil, with known values of hAChE (Additional file 2: Table S7) [40].

Molecular descriptors

The molecular structures are drawn with the aid of the "Add Hydrogen and Model Build" modulus of HyperChem 6.03 for Windows [41]. The compounds are firstly pre-optimized with the Molecular Mechanics Force Field (MM+) procedure included in HyperChem, and the resulting geometries are further refined by means of the Semiempirical Method PM3 (Parametric Method-3) using the Polak-Ribiere algorithm and a gradient norm limit of $0.01 \text{ kcal.}\text{\AA}^{-1}$. We keep the R-configuration for all the molecules having a chiral Carbon atom when such configuration detail is not specified in the original data.

All the molecular descriptors are computed using the software Dragon [19]. This well-known descriptors database includes descriptor families such as: 0D-descriptors: constitutional indices; 1D-descriptors: functional group counts, atom-centred fragments, empirical descriptors, molecular properties; 2D-descriptors: topological indices, walk and path counts, connectivity indices, information indices, 2D-autocorrelations, Burden eigenvalues, Galvez topological charge indices; and 3D-descriptors: aromaticity indices, Randic molecular profiles, geometrical indices, radial distribution functions, 3D-MoRSE (3D-Molecular

Representation of Structure based on Electron diffraction) descriptors, WHIM (Weighted Holistic Invariant Molecular) descriptors, GETAWAY (Geometry, Topology and Atoms-Weighted Assembly) descriptors and charge indices [42]. Finally, five quantum-chemical descriptors not provided by the Dragon program are added to the pool: molecular dipole moment, total energy, homo-lumo energy, and homo-lumo gap ($\Delta_{\text{homo-lumo}}$). The resulting pool consists on $D = 1502$ structural descriptors.

Linear model search

In past years, the MLR technique has proved to be a multidisciplinary technique of valuable applicability for establishing predictive QSAR models by different research groups, by performing an exhaustive analysis of a pool containing a great number of structural descriptors [19,43]. Linear models are more general and can transparently reveal the effect on the activity being modeled when including/excluding molecular descriptors in the equation, thus making it possible to suggest cause/effect relationships by means of simple parallelisms. The main advantage of developing linear regression models, when compared to non-linear models, is the fact that linear models suffer in a lesser extend from the over-fitting (over-training) problem [44,45] because the MLR technique does not involve too many optimization parameters during the model building, just the regression coefficient for each molecular descriptor. Therefore, the MLR is considered as the best choice of descriptor selection method when few experimental activity values are available, as it is the case in present study: few data are available for each data set A-G.

It is our purpose to search the set \mathbf{D} having D descriptors, for an optimal subset \mathbf{d} of $d < D$ ones with minimum standard deviation (S), by means of the MLR technique. More precisely, we want to obtain the global minimum of $S(\mathbf{d})$ where \mathbf{d} is a point in a space of $D!/(D-d)!d!$ ones. Each point is a possible model of d descriptors as discussed below. Taking into account that a Full Search (FS) of optimal variables is impractical because it requires $D!/(D-d)!d!$ linear regressions, some time ago we have proposed the Replacement Method (RM) [46-51] that produces linear regression QSAR models that are quite close to the FS ones with much less computational work. The RM approaches the minimum of S by judiciously taking into account the relative errors of the coefficients of the least-squares model given by a set of d descriptors $\mathbf{d} = \{X_1, X_2, \dots, X_d\}$.

The procedure of the RM technique is as follows: choose d descriptors $\{X_1, X_2, \dots, X_d\}$ at random and do a linear regression. Choose one of the descriptors of this set, say X_b , and replace it by each of the D descriptors of the pool (except itself) keeping the best resulting set. Since one can start replacing any of the d descriptors in

the initial model, then a regression equation with d variables has d possible paths to achieve the final result; for example, the choice above will develop into path i . Next, choose the variable with greatest relative error in its coefficient (except the one replaced in the previous step) and replace it with all the D descriptors (except itself) keeping again the best set. Replace all the remaining variables in the same way by passing those replaced in previous steps. When finishing, start again with the variable having greatest relative error in the coefficient and repeat the whole process. Repeat this process as many times as necessary until the set of descriptors remains unchanged. At the end, we have the best model for the path i . Proceed in exactly the same way for all possible paths $i = 1, 2, \dots, d$, compare the resulting models, and keep the best one. Our numerical experiments show that in this way one obtains a model almost as good as the best one with much less than $D!/(D-d)!d!$ linear regressions when this combinatorial number is large. The RM gives models with better statistical parameters than the Forward Stepwise Regression procedure [52] and variants of the more elaborated Genetic Algorithms [53].

Model validation

The design of a properly validated model constitutes the most important step for every QSAR analysis, in order to generate predictive models that involve general applicability and that are not limited to only function correlatively. The Cross-Validation technique of Leave-One-Out (loo) is practiced [54]. The parameters R_{loo}^2 and S_{loo} (square of the correlation coefficient and standard deviation of Leave-One-Out) measure the stability of the model upon exclusion of molecules. According to the literature, R_{loo}^2 should be greater than 0.5 for a validated model.

The Y-Randomization procedure [55] is also applied in order to verify that the model is robust. This technique consists on scrambling the experimental property values in such a way that they do not correspond to the respective compounds. After analyzing 1000 cases of Y-Randomization, the standard deviation obtained (S^{rand}) has to be a poorer value than the one found by considering the true calibration (S).

Finally, we also apply the standard practice that consists on omitting from the complete molecular set some compounds which constitute the 'test set,' denoted here as 'test'. The main purpose of performing such a splitting is to assess whether the QSAR found have predictive capability for estimating the activity values on the independent test set compounds, that are not involved during the model fitting using the 'training set' compounds (denoted as 'train'). We select the molecules composing the training and test series as a previous step

to the model search, and this is done in such a way that both sets share similar qualitative structure-property characteristics.

Results and discussion

The application of the RM variable subset selection method on the total pool with $D = 1502$ molecular descriptors leads to the best linear models for the Alzheimer's data sets. In order to determine the number of descriptors to be included in the structure-activity relationship, we consider the 'Rule of Thumb,' which states that at least 5 or 6 training set molecules (N) should be present for each fitting parameter. In the following we present the best QSAR found for data sets A-G.

Data set A (bAChE):

$$\log_{10}IC_{50} = 1.345(\pm 2) - 1.051(\pm 0.3) \cdot Mor06v - 44.823(\pm 10) \cdot G1u$$
$$N = 12, \text{ range} = [-6.886, -4.678], R = 0.90, S = 0.29,$$
$$F = 18.36, p < 10^{-4}, R_{ij}^{max} = 0.43, o(> 2.S) = 0, R_{loo} = 0.84,$$
$$S_{loo} = 0.39, S^{rand} = 0.31 \quad (1)$$

Data set B (bAChE):

$$\log_{10}IC_{50} = -7.244(\pm 0.4) + 6.741(\pm 0.4) \cdot GATS2e - 4.476(\pm 0.4) \cdot R5e$$
$$N = 12, \text{ range} = [-8.242, -5.000], R = 0.98, S = 0.23,$$
$$F = 143.01, p < 10^{-4}, R_{ij}^{max} = 0.46, o(> 2.S) = 0,$$
$$R_{loo} = 0.96, S_{loo} = 0.37, S^{rand} = 0.30 \quad (2)$$

Data set B (hAChE):

$$\log_{10}IC_{50} = -18.159(\pm 0.8) + 101.916(\pm 6) \cdot Qmean$$
$$+ 35.567(\pm 6) \cdot R3m^+$$
$$N = 11, \text{ range} = [-8.153, -4.491], R = 0.99, S = 0.22,$$
$$F = 170.55, p < 10^{-4}, R_{ij}^{max} = 0.61, o(> 2.S) = 0, R_{loo} = 0.98,$$
$$S_{loo} = 0.27, S^{rand} = 0.27 \quad (3)$$

Data set C (EeAChE):

$$\log_{10}IC_{50} = -2.611(\pm 0.3) - 83.235(\pm 8) \cdot R5e^+$$
$$- 1.592(\pm 0.1) \cdot nNHR$$
$$N = 16, \text{ range} = [-7.284, -4.155], R = 0.97, S = 0.24,$$
$$F = 109.06, p < 10^{-4}, R_{ij}^{max} = 0.25, o(> 2.S) = 0, R_{loo} = 0.95,$$
$$S_{loo} = 0.32, S^{rand} = 0.50 \quad (4)$$

Data set C (hAChE):

$$\log_{10}IC_{50} = 16.021(\pm 4) - 52.448(\pm 10) \cdot X1A$$
$$N = 9, \text{ range} = [-6.959, -5.553], R = 0.90, S = 0.24,$$
$$F = 28.44, p < 10^{-4}, R_{ij}^{max} = 0, o(> 2.S) = 0, R_{loo} = 0.84,$$
$$S_{loo} = 0.30, S^{rand} = 0.25 \quad (5)$$

Data set D (bAChE):

$$\log_{10}IC_{50} = -5.510(\pm 0.1) - 1.263(\pm 0.1) \cdot Mor09u \\ - 32.160(\pm 2.0) \cdot R2m^+ \\ N = 9, \text{ range} = [-8.553, -5.522], R = 0.98, S = 0.18, \\ F = 98.81, p < 10^{-4}, R_{ij}^{\max} = 0.30, o(> 2.S) = 0, \\ R_{loo} = 0.95, S_{loo} = 0.33, S^{rand} = 0.26 \quad (6)$$

Data set E (hAChE):

$$\log_{10}IC_{50} = 5.385(\pm 2) - 6.669(\pm 1) \cdot BELm2 - 2.697(\pm 0.3) \cdot Mor27e \\ N = 12, \text{ range} = [-9.092, -6.342], R = 0.96, S = 0.29, \\ F = 51.70, p < 10^{-4}, R_{ij}^{\max} = 0.046, o(> 2.S) = 0, R_{loo} = 0.92, \\ S_{loo} = 0.39, S^{rand} = 0.46 \quad (7)$$

Data set F (bAChE):

$$\log_{10}IC_{50} = -17.546(\pm 2) + 1.510(\pm 0.2) \cdot AMW \\ + 0.247(\pm 0.2) \cdot MEcc \\ N = 11, \text{ range} = [-7.000, -4.000], R = 0.98, S = 0.25, \\ F = 21.33, p < 10^{-4}, R_{ij}^{\max} = 0.13, o(> 2.S) = 0, \\ R_{loo} = 0.96, S_{loo} = 0.36, S^{rand} = 0.39 \quad (8)$$

Data set F (EeAChE):

$$\log_{10}IC_{50} = -15.372(\pm 1) + 13.386(\pm 2) \cdot E2u \\ + 18.306(\pm 1) \cdot HATS6m \\ N = 11, \text{ range} = [-7.854, -4.523], R = 0.99, S = 0.19, \\ F = 150.88, p < 10^{-4}, R_{ij}^{\max} = 0.66, o(> 2.S) = 0, \\ R_{loo} = 0.97, S_{loo} = 0.28, S^{rand} = 0.34 \quad (9)$$

Data set G (hAChE):

$$\log_{10}IC_{50} = 66.887(\pm 6) + 0.014(\pm 0.0009) \cdot MPC09 \\ - 78.669(\pm 6) \cdot MATS1m - 6.087(\pm 0.3) \cdot RDF020m \\ N = 19, \text{ range} = [-9.569, -5.000], R = 0.98, S = 0.21, \\ F = 131.81, p < 10^{-4}, R_{ij}^{\max} = 0.65, o(> 2.S) = 3, \\ R_{loo} = 0.97, S_{loo} = 0.29, S^{rand} = 0.53 \quad (10)$$

In these equations, *range* stands for the range of experimental activity of the training set, *R* is the correlation coefficient, *S* is the standard deviation of the model, *F* is the Fisher parameter, *p* is the significance of the model, R_{ij}^{\max} denotes the maximum intercorrelation coefficient between descriptors, $o(>2.S)$ indicates the number of outlier molecules having a residual greater than two times the standard deviation, and R_{loo} and S_{loo} are the correlation coefficient and standard deviation obtained with the loo technique, respectively.

In most cases, it is appreciated that a single descriptor does not achieve enough accuracy for predicting the AD activities, while models based on two- or three- descriptors are acceptable for the number of training

molecules involved. When we plot the QSAR predicted $\log_{10}IC_{50}$ inhibitory activities as function of experimental values for each data set (A-G) in Additional file 3, a straight line trend is observed. It is also appreciated that Eqs. 1–10 predict reasonably well the experimental activities of the molecules that are members of the test set, and thus the models established tend to behave as predictive. In addition, both parameters R_{loo} and S_{loo} measure the stability of the developed QSAR upon inclusion/exclusion of compounds, and according to the specialized literature, R_{loo}^2 must be greater than 0.50 for obtaining a validated model [54].

Dispersion plots of residuals (residuals as function of predicted activities) for each QSAR are provided in Additional file 4: Figure S1A up to S1G with the purpose of demonstrating the validity of these MLR equations. Although some outliers are detected in some plots, having residuals exceeding the 2.S value, we decide to derive general models having applicability to any biomolecule without restrictions, and so we do not remove such molecules from the training set.

The molecular descriptors appearing in Eqs. 1–10 are of different types and are briefly described in Table 1. We provide their numerical values in Additional file 2: Tables S1-S7. Constitutional descriptors are 0D-descriptors, independent from the molecular connectivity and conformation; they are the most simple and commonly used descriptors, reflecting the molecular composition of a compound. The topological indices derived from the Chemical Graph Theory [56] are obtained with a graph representation of the molecule, that is to say, its planar image, and provide only information on the structural composition and connectivity but nothing about its three dimensional or stereochemical aspects. In a molecular graph, the vertices are atoms weighted with different physicochemical properties such as mass, polarity, electronegativity, charge, etc. The 3D-MoRSE (3D-Molecule Representation of Structure based on Electron diffraction) descriptors [57] provide 3D-information from the molecular structure using a molecular transform derived from an equation used in electron diffraction studies. Several atomic properties can be taken into account, thus giving high flexibility to this representation of a molecule.

WHIM (Weighted Holistic Invariant Molecular Descriptors) descriptors are based on statistical indices calculated on the projections of atoms along principal axes [58]. The aim is to capture 3D-information regarding size, shape, symmetry and atom distributions with respect to invariant reference frames. In order to calculate them, a weighted covariance matrix is obtained from different weighting schemes for the atoms. The different structural variables introduced by Broto, Moreau, and Geary [59] account for bi-dimensional autocorrelations between atoms pairs in the molecule, and are defined in

Table 1 Notation and brief description of the molecular descriptors involved in the QSAR models for AD

Molecular descriptor	Type	Description
<i>Mor06v</i>	3D-MoRSE	3D-MoRSE - signal 06/weighted by atomic van der Waals volumes
<i>Mor09u</i>		3D-MoRSE - signal 09/unweighted
<i>Mor27e</i>		3D-MoRSE - signal 27/weighted by atomic Sanderson electronegativities
<i>G1u</i>	WHIM	1st component symmetry directional WHIM index/unweighted
<i>Eu2</i>		2nd component accessibility directional WHIM index/unweighted
<i>GATS2e</i>	2D- Autocorrelations	Geary autocorrelation of lag 2/weighted by atomic Sanderson electronegativities
<i>MATS1m</i>		Moran autocorrelation of lag 1/weighted by atomic masses
<i>R5e</i>	GETAWAY	R autocorrelation of lag 5/weighted by atomic Sanderson electronegativities
<i>HATS6m</i>		leverage-weighted autocorrelation of lag 6/weighted by atomic masses
<i>R3m⁺</i>		R maximal autocorrelation of lag 3/weighted by atomic masses
<i>R2m⁺</i>		R maximal autocorrelation of lag 2/weighted by atomic masses
<i>R5e⁺</i>		R maximal autocorrelation of lag 5/weighted by atomic Sanderson electronegativities
<i>nNHR</i>	Constitutional	number of secondary amines (aliphatic)
<i>AMW</i>		average molecular weight
<i>X1A</i>	Topological	average connectivity index chi-1
<i>MPC09</i>		molecular path count of order 09
<i>Qmean</i>	Charge	mean absolute charge (charge polarization)
<i>BELm2</i>	BCUT	lowest eigenvalue number 2 of Burden matrix/weighted by atomic masses
<i>MEcc</i>	Geometrical	molecular eccentricity
<i>RDF020m</i>	RDF	Radial Distribution Function - 2.0/weighted by atomic masses

order to reflect the contribution of a considered atomic property to the experimental observations under investigation. These indices can be readily calculated, *i.e.*: by summing products of atomic weights (employing atomic properties such as atomic polarizabilities, molecular volumes, etc.) of the terminal atoms of all the paths of a prescribed length.

The GETAWAY (GEometry, Topology, and Atom-Weights Assembly) type of descriptors [60] have been designed with the main purpose of matching the 3D-molecular geometry. These numerical variables are derived from the elements h_{ij} of the Molecular Influence matrix (**H**), obtained through the values of atomic Cartesian coordinates. The diagonal elements of **H** (h_{ii}) are called leverages, and are considered to represent the influence of each atom on the whole shape of the molecule. For instance, the mantle atoms always have higher h_{ii} values than atoms near the molecule center, while each off-diagonal element h_{ij} represents the degree of accessibility of the *j*th atom to interactions with the *i*th one. Charge descriptors are electronic descriptors defined in terms of atomic charges and used to describe electronic aspects both of the whole molecule and of particular regions, such as atoms, bonds, and molecular fragments. Charge descriptors can be considered among quantum chemical descriptors [61]. Electrical charges in the molecule are the driving force of electrostatic interactions, and it is

well known that local electron densities or charges play a fundamental role in many chemical reactions, physico-chemical properties and receptor-ligand binding affinity.

BCUT descriptors are the eigenvalues of a modified connectivity matrix, the Burden matrix (**B**) [62]. The ordered sequence of the *n* smallest eigenvalues of **B** has been proposed as a molecular descriptor based on the assumption that the lowest eigenvalues contain contributions from all the atoms and thus reflect the molecular topology. The BCUT descriptors are an extension of the Burden eigenvalues and consider three classes of matrices, whose diagonal elements account for atomic charge related values, atomic polarizability related values and atomic H bond abilities. Geometrical descriptors are a different kind of conformationally dependent parameters based on the molecular geometry. Reliable values are obtained if reliable conformations are previously calculated [42]. A Radial Distribution Function (RDF) [60] of an ensemble of atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of certain radius, also incorporating different atomic properties, in order to differentiate the contribution of each atom to the property under study.

The QSAR models presented in Eqs. 1–10 for data sets A-G result specific to each group of Tacrine derivatives used and to each AChE source. An attempt to model the complete set of molecules (combining data sets A-G)

with a single model would result in a deteriorated statistics. Furthermore, this attempt of increasing the number of experimental observations is not completely justified, as different biochemical conditions are employed for measuring the inhibitory potency. We believe that the QSAR obtained here should acceptably work in an applicability domain defined by the range of descriptor variations for each model, and would be valid for predicting structures whose experimental activities lay close to the experimental range of the training set.

Conclusions

In this work, we present ten useful structure-activity models to analyze the activity of Acetylcholinesterase inhibitors. The proposed QSAR models highlight parallels between the molecular structure and the AD inhibitory activity. The importance of this study relies on the fact that new Tacrine related inhibitors are used, which have been measured under different experimental conditions and using different AChE sources. Despite of this limitation, such models show appropriate predictive capability, and the scarce amount of observations available on each data set is successfully analyzed by means of the linear methodology of the Replacement Method approach. Nevertheless, whenever newer experimental information on Tacrine-like compounds are available, further QSAR studies are encouraged on larger data sets.

Additional files

Additional file 1: Molecular structures of Tacrine derivatives analyzed. Structure 1 is Tacrine.

Additional file 2: Table S1. Experimental and predicted \log_{10}/C_{50} for bAChE inhibitory activity in data set A. Numerical values of the involved descriptors are also given. **Table S2.** Experimental and predicted \log_{10}/C_{50} for bAChE and hAChE inhibitory activities in data set B. Numerical values of the involved descriptors are also given. **Table S3.** Experimental and predicted \log_{10}/C_{50} for EeAChE and hAChE inhibitory activities in data set C. Numerical values of the involved descriptors are also given. **Table S4.** Experimental and predicted \log_{10}/C_{50} for bAChE inhibitory activity in data set D. Numerical values of the involved descriptors are also given. **Table S5.** Experimental and predicted \log_{10}/C_{50} for hAChE inhibitory activity in data set E. Numerical values of the involved descriptors are also given. **Table S6.** Experimental and predicted \log_{10}/C_{50} for bAChE and EeAChE inhibitory activities in data set F. Numerical values of the involved descriptors are also given. **Table S7.** Experimental and predicted \log_{10}/C_{50} for hAChE inhibitory activity in data set G. Numerical values of the involved descriptors are also given.

Additional file 3: Predicted \log_{10}/C_{50} as the function of experimental values for data set A-G. A: Data set A (bAChE); B: Data set B (bAChE); C: Data set B (hAChE); D: Data set C (EeAChE); E: Data set C (hAChE); F: Data set D (bAChE); G: Data set E (hAChE); H: Data set F (bAChE); I: Data set F (EeAChE); J: Data set G (hAChE).

Additional file 4: Figure S1A. Dispersion plot of residuals for data set A (bAChE). **Figure S1B.** Dispersion plot of residuals for data set B (bAChE). **Figure S1C.** Dispersion plot of residuals for data set B (hAChE). **Figure S1D.** Dispersion plot of residuals for data set C (EeAChE). **Figure S1E.** Dispersion plot of residuals for data set C (hAChE). **Figure S1F.** Dispersion plot of residuals for data set D (bAChE). **Figure S1G.** Dispersion plot of residuals for data set E (hAChE). **Figure S1H.**

Dispersion plot of residuals for data set F (bAChE). **Figure S1I.** Dispersion plot of residuals for data set F (EeAChE). **Figure S1J.** Dispersion plot of residuals for data set G (hAChE).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed equally to this work. All authors read and approved the final manuscript.

Acknowledgments

PRD, AGM and EAC thank the financial support provided by the National Research Council of Argentina (CONICET) PIP11220100100151 project. KW is grateful to a fellowship from the International Association for the Exchange of Students for Technical Experience (IAESTE). PRD, AGM, GPR and EAC are members of the Scientific Researcher Career of CONICET.

Author details

¹Department of Chemical Engineering, South Kensington Campus, Imperial College London, SW7 2AZ London, UK. ²Instituto de Investigaciones Físicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diag. 113 y 64, Sucursal 4, C.C. 16, 1900 La Plata, Argentina. ³Centro de Investigación y Desarrollo en Ciencias Aplicadas "Dr. J.J. Ronco" (CINDECA), Departamento de Química, Facultad de Ciencias Exactas, UNLP-CCT-CONICET, Calle 47N° 257, B1900AJK La Plata, Argentina. ⁴School of Pharmacy and Pharmacology, University of KwaZulu-Natal, 4001 Durban, South Africa.

Received: 10 May 2014 Accepted: 13 August 2014

Published online: 20 September 2014

References

- Goedert M, Spillantini MG: A Century of Alzheimer's Disease. *Science* 2006, **314**:777–781.
- McDowell I: Alzheimer's disease: insights from epidemiology. *Aging (Milano)* 2001, **13**:143–162.
- Schmidt R, Neff F, Lampl C, Benke T, Anditsch M, Bancher C, Dal-Bianco P, Reiskeker F, Marksteiner J, Rainer M, Kapeller P, Dodel R: Therapy of Alzheimer's disease: current status and future development. *Neuropsychiatr* 2008, **22**:153–171.
- Bartus RT, Dean RL, Beer B, Lipka AS: The cholinergic hypothesis of geriatric memory dysfunction. *Science* 1982, **217**:408–417.
- Gualtieri F, Dei S, Manetti D, Romanelli MN, Scapecchi S, Teodori E: The medicinal chemistry of Alzheimer's and Alzheimer-like diseases with emphasis on the cholinergic hypothesis. *Farmacologia* 1995, **50**:489–503.
- Alcala MDM, Vivas NM, Hospital S, Camps P, Muñoz-Torrero D, Badia A: Characterisation of the anticholinesterase activity of two new tacrine-huperzine A hybrids. *Neuropharmacology* 2003, **44**:749–755.
- Quinn DM, Feaster SR, Nair HK, Baker NA, Radic Z, Taylor P: Delineation and decomposition of energies involved in quaternary ammonium binding in the active site of acetylcholinesterase. *J Am Chem Soc* 2000, **122**:2975–2980.
- Racchi M, Mazzucchelli M, Porrello E, Lanni C, Govoni S: Acetylcholinesterase inhibitors: novel activities of old molecules. *Pharmacol Res* 2004, **50**:441–451.
- Carlier PR, Han YF, Chow ESH, Li CPL, Wang H, Lieu TX, Wong HS, Pang YP: Evaluation of short-tether Bis-THA AChE inhibitors. A further test of the dual binding site hypothesis. *Bioorg Med Chem* 1999, **7**:351–357.
- Pang YP, Quiram P, Jelacic T, Hong F, Brimijoin S: Highly potent, selective, and low cost bis-tetrahydroacminacrine Inhibitors of Acetylcholinesterase. Steps toward novel drugs for treating Alzheimer's disease. *J Biol Chem* 1996, **271**:23646–23649.
- Qizilbash N, Birks J, Lopez-Arrieta J, Lewington S, Szeto S: Tacrine for Alzheimer's disease. *Cochrane Database Syst Rev* 2000, **3**:CD000202.
- Hansch C, Leo A: *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*. Washington, D. C: American Chemical Society; 1995.
- Kubinyi H: *QSAR: Hansch Analysis and Related Approaches*. New York: Wiley-Interscience; 2008.
- Katritzky AR, Lobanov VS, Karelson M: QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem Soc Rev* 1995, **24**:279–287.

15. Akula NL, Lecanu J, Greeson V: **3D QSAR studies of AChE inhibitors based on molecular docking scores and CoMFA.** *Bioorg Med Chem Lett* 2006, **16**:6277–6280.
16. Tripos: <http://www.tripos.com/index.php?family=modules,SimplePage,,&page=SYBYL-X>.
17. *Stewart Computational Chemistry.* <http://www.openmopac.net>.
18. Fernández M, Carreiras MC, Marco JL, Caballero J: **Modeling of acetylcholinesterase inhibition by tacrine analogues using Bayesian-regularized Genetic Neural Networks and ensemble averaging.** *J Enzym Inhib Med Ch* 2006, **21**:647–661.
19. Todeschini R, Chemometrics M, QSAR Research Group: Bicocca: Dipartimento di Scienze dell'Ambiente e del Territorio e di Scienze della Terra, Università degli Studi di Milano; <http://michem.disat.unimib.it/chm/>.
20. Jung M, Tak J, Lee YN, Jung YG: **Quantitative structure-activity relationship (QSAR) of tacrine derivatives against acetylcholinesterase (AChE) activity using variable selections.** *Bioorg Med Chem Lett* 2007, **17**:1082–1090.
21. Wavefunction, Inc: <http://www.wavefun.com>.
22. BMDRC: http://www.bmdrc.org/04_product/01_preadme.asp.
23. Fujitsu Ltd. <http://www.cache.fujitsu.com/biomedcache/index.shtml>.
24. Saracoglu M, Kandemirli F: **The investigation of structure-activity relationships of tacrine analogues: electronic-topological method.** *Open Med Chem J* 2008, **2**:75–80.
25. Dimoglo AS, Shvets NM, Tetko IV, Livingstone DJ: **Electronic-topological investigation of the structure-acetylcholinesterase inhibitor activity relationship in the series of n-benzylpiperidine derivatives.** *Quant Struct-Act Rel* 2001, **20**:31–45.
26. Chen N, Liu CK, Zhao LZ, Zhang HB: **3D-QSAR study of multi-target-directed AChE inhibitors based on autodocking.** *Med Chem Res* 2012, **21**:245–256.
27. Tute MS: *History and Objectives of Quantitative Drug Design in Advances in Drug Research.* London: Academic Press; 1971.
28. Vicente E, Duchowicz PR, Benítez D, Castro EA, Cerecetto H, González M, Monge A: **Anti-T. cruzi activities and QSAR studies of 3-arylquinoline-2-carbonitrile di-N-oxides.** *Bioorg Med Chem Lett* 2010, **20**:4831–4835.
29. Mullen LMA, Duchowicz PR, Castro EA: **QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents.** *Chemom Intel Lab Syst* 2011, **107**:269–275.
30. Pasquale G, Romanelli GP, Autino JC, García J, Ortiz EV, Duchowicz PR: **Quantitative structure-activity relationships on Chalcone derivatives: mosquito larvicidal studies.** *J Agric Food Chem* 2012, **60**:692–697.
31. Duchowicz PR, Giraudo MA, Castro EA, Pomilio AB: **Amino acid profiles and quantitative structure-property relationship models as markers for Merlot and Torrontés wines.** *Food Chem* 2013, **140**:210–216.
32. Ellman GL, Courtney KD, Andres V Jr, Featherstone RM: **A new and rapid colorimetric determination of acetylcholinesterase activity.** *Biochem Pharmacol* 1961, **7**:88–95.
33. Rappaport F, Fischl J, Pinto N: **An improved method for the estimation of cholinesterase activity in serum.** *Clin Chim Acta* 1959, **4**:227–230.
34. Marco JL, de los Ríos C, Carreiras MC, Baños JE, Badía A, Vivas NM: **Synthesis and acetylcholinesterase/butyrylcholinesterase inhibition activity of new tacrine-like analogues.** *Bioorg Med Chem* 2001, **9**:727–732.
35. Camps P, Formosa X, Galdeano C, Muñoz-Torrero D, Ramírez L, Gómez E, Isambert N, Lavilla R, Badía A, Clos MV, Bartolini M, Mancini F, Andrisano V, Arce MP, Rodríguez-Franco MI, Huertas Ó, Dafni T, Luque FJ: **Pyrano[3,2-c]quinoline-6-chlorotacrine hybrids as a novel family of acetylcholinesterase- and beta-amyloid-directed anti-Alzheimer compounds.** *J Med Chem* 2009, **52**:5365–5379.
36. León R, de los Ríos C, Marco-Contelles J, Huertas O, Barril X, Luque FJ, López MG, García AG, Villarroya M: **New tacrine-dihydropyridine hybrids that inhibit acetylcholinesterase, calcium entry, and exhibit neuroprotection properties.** *Bioorg Med Chem* 2008, **16**:7759–7769.
37. Alonso D, Dorronsoro I, Rubio L, Muñoz P, García-Palomero E, Del Monte M, Bidon-Chanal A, Orozco M, Luque FJ, Castro A, Medina M, Martínez A: **Donepezil-tacrine hybrid related derivatives as new dual binding site inhibitors of AChE.** *Bioorg Med Chem* 2005, **13**:6588–6597.
38. Rizzo S, Bisi A, Bartolini M, Mancini F, Belluti F, Gobbi S, Andrisano V, Rampa A: **Multi-target strategy to address Alzheimer's disease: Design, synthesis and biological evaluation of new tacrine-based dimers.** *Eur J Med Chem* 2011, **46**:4336–4343.
39. Samadi A, Valderas C, de los Ríos C, Bastida A, Chioua M, González-Lafuente L, Colmena I, Gandía L, Romero A, del Barrio L, Martín-de-Saavedra MD, López MG, Villarroya M, Marco-Contelles J: **Cholinergic and neuroprotective drugs for the treatment of Alzheimer and neuronal vascular diseases. II. Synthesis, biological assessment, and molecular modelling of new tacrine analogues from highly substituted 2-aminopyridine-3-carbonitriles.** *Bioorg Med Chem* 2011, **19**:122–133.
40. Camps P, Formosa X, Galdeano C, Gómez T, Muñoz-Torrero D, Ramírez L, Viayna E, Gómez E, Isambert N, Lavilla R, Badía A, Clos MV, Bartolini M, Mancini F, Andrisano V, Bidon-Chanal A, Huertas Ó, Dafni T, Luque FJ: **Tacrine-based dual binding site acetylcholinesterase inhibitors as potential disease-modifying anti-Alzheimer drug candidates.** *Chem Biol Interact* 2010, **187**:411–415.
41. *Hyperchem, 6.03 (Hypercube).* <http://www.hyper.com>.
42. Todeschini R, Consonni V: *Molecular Descriptors for Chemoinformatics, Volume 2.* Wiley-VCH: Weinheim; 2009.
43. aKatritzky Group: www.arkchem.ufl.edu/; bJurs Research Group, <http://research.chem.psu.edu/pcjgroup/>.
44. Livingstone ADJ, Manallack DT: **Statistics using neural networks: chance effects.** *J Med Chem* 1993, **36**:1295–1297.
45. Tetko IV, Luik AI, Poda GI: **Applications of neural networks in structure-activity relationships of a small number of molecules.** *J Med Chem* 1993, **36**:811–814.
46. Duchowicz PR, Castro EA, Fernández FM, González MP: **A new search algorithm of QSPR/QSAR theories: normal boiling points of some organic molecules.** *Chem Phys Lett* 2005, **412**:376–380.
47. Duchowicz PR, Castro EA, Fernández FM: **Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies.** *MATCH Commun Math Comput Chem* 2006, **55**:179–192.
48. Duchowicz PR, Fernández M, Caballero J, Castro EA, Fernández FM: **QSAR of non-nucleoside inhibitors of HIV-1 reverse transcriptase.** *Bioorg Med Chem* 2006, **14**:5876–5889.
49. Duchowicz PR, Talevi A, Bruno-Blanch LE, Castro EA: **New QSPR study for the prediction of aqueous solubility of drug-like compounds.** *Bioorg Med Chem* 2008, **16**:7944–7955.
50. Talevi A, Goodarzi M, Ortiz EV, Duchowicz PR, Bellera CL, Pesce G, Castro EA, Bruno-Blanch LE: **Prediction of drug intestinal absorption by new linear and non-linear QSPR.** *Eur J Med Chem* 2011, **46**:218–228.
51. Mercader AG, Duchowicz PR, Fernández FM, Castro EA: **Advances in the replacement and enhanced replacement method in QSAR and QSPR theories.** *J Chem Inf Model* 2011, **51**:1575–1581.
52. Draper NR, Smith H: *Applied Regression Analysis.* New York: John Wiley & Sons; 1981.
53. Mercader AG, Duchowicz PR, Fernández FM, Castro EA: **Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories.** *J Chem Inf Model* 2010, **50**:1542–1548.
54. Golbraikh A, Tropsha A: **Beware of q²!** *J Mol Graphics Modell* 2002, **20**:269–276.
55. Wold S, Eriksson L: *Statistical Validation of QSAR Results, Chemometrics. Methods in Molecular Design.* In Edited by van de Waterbeemd) H. Weinheim New York: VCH; 1995:309–318.
56. Trinajstić N: *Chemical Graph Theory.* Boca Raton (FL): CRC Press; 1992.
57. Schuur JH, Selzer P, Gasteiger J: **The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity.** *J Chem Inf Comput Sci* 1996, **36**:334–344.
58. Todeschini R, Gramatica P: **SD-modelling and prediction by WHIM descriptors. Part 5. Theory development and chemical meaning of WHIM descriptors.** *Quantitative Structure-Activity Relationships* 1997, **16**:113–119.
59. Moreau G, Broto P: **Auto-correlation of molecular structures, application to SAR studies.** *Nouv J Chim* 1980, **4**:757–764.
60. Consonni V, Todeschini R, Pavan M, Gramatica P: **Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies.** *J Chem Inf Model* 2002, **42**:693–705.
61. Streitwieser AJ: *Molecular Orbital Theory for Organic Chemists.* New York: Wiley; 1961.
62. Frank RB: **A chemically intuitive molecular index based on the Eigenvalues of a modified adjacency matrix.** *Quant Struct Act Relat* 1997, **16**:309–314.

doi:10.1186/s12929-014-0084-0

Cite this article as: Wong et al.: QSAR analysis on tacrine-related acetylcholinesterase inhibitors. *Journal of Biomedical Science* 2014 **21**:84.