Ligand – Based virtual screening procedure for the prediction and the identification of novel β-amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial...

ARTICLE *in* EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY · FEBRUARY 2011

Impact Factor: 3.45 · DOI: 10.1016/j.ejmech.2010.11.029 · Source: PubMed

5 AUTHORS, INCLUDING:

Antreas Afantitis

49 PUBLICATIONS **952** CITATIONS

SEE PROFILE

Georgia Melagraki

50 PUBLICATIONS **967** CITATIONS

SEE PROFILE

Panayiotis Koutentis

University of Cyprus

120 PUBLICATIONS **1,434** CITATIONS

SEE PROFILE

Haralambos Sarimveis

National Technical University of Athens

146 PUBLICATIONS **2,395** CITATIONS

SEE PROFILE

European Journal of Medicinal Chemistry 46 (2011) 497–508

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Ligand - based virtual screening procedure for the prediction and the identification of novel β-amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks

Antreas Afantitis [a,*], Georgia Melagraki [b,*], Panayiotis A. Koutentis [b], Haralambos Sarimveis [c], George Kollias [d]

[a] Department of ChemoInformatics, NovaMechanics Ltd, John Kennedy Ave 62-64, Nicosia 1046, Cyprus
[b] Department of Chemistry, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
[c] School of Chemical Engineering, National Technical University of Athens, Athens, Greece
[d] Biomedical Sciences Research Center "Alexander Fleming", Athens, Greece

ARTICLE INFO

Article history:
Received 26 March 2010
Received in revised form
15 November 2010
Accepted 17 November 2010
Available online 24 November 2010

Keywords:
Alzheimer's disease
β-Amyloid inhibitors
Kohonen map
CP-ANN
QSAR
In silico virtual screening

ABSTRACT

In this work we have developed an *in silico* model to predict the inhibition of β-amyloid aggregation by small organic molecules. In particular we have explored the inhibitory activity of a series of 62 *N*-phenylanthranilic acids using Kohonen maps and Counterpropagation Artificial Neural Networks. The effects of various structural modifications on biological activity are investigated and novel structures are designed using the developed *in silico* model. More specifically a search for optimized pharmacophore patterns by insertions, substitutions, and ring fusions of pharmacophoric substituents of the main building block scaffolds is described. The detection of the domain of applicability defines compounds whose estimations can be accepted with confidence.

1. Introduction

Alzheimer's disease (AD) is a chronic, slowly progressive neurodegenerative disorder and is a very common form of dementia in the elderly [1]. Over the past years many efforts have been made to cure AD or stop its progression, however, there is still no effective treatment [2]. The β-amyloid peptide (Aβ) is produced by proteolytic cleavage of the amyloid precursor protein (APP) and plays a central role in the neuropathology of AD. As β-amyloid protein aggregation is present in Alzheimer's disease, recent efforts have focused on the identification of small organic molecules that can act as β-amyloid aggregation inhibitors [3–5]. A variety of synthetic methods have been proposed recently for the design of new molecules with enhanced activity [6–9].

In silico methods have emerged as a useful tool in the identification of novel compounds with improved characteristics [10–14]. Different regression or classification methods have been employed [15–18] for this purpose in an effort to minimize the time and cost associated with identifying new leads.

In this study we have developed a classification model using a recently published dataset of 62 *N*-phenylanthranilic acids that were explored as potent β-amyloid aggregation inhibitors [19]. A great variety of *in silico* methods [20–25] have emerged as effective tools to predict the activity of a new molecule prior to its actual synthesis. In this work we present the development of an accurate and robust classification model based on Kohonen maps (or Self Organizing Maps, SOMs) and Counterpropagation Artificial Neural Networks [26–28] (CP-ANNs). The validated *in silico* model combined with the selected molecular descriptors, which demonstrate discriminatory and pharmacophore abilities, has been applied for the investigation of the effects of various structural modifications on biological activity. Novel structures were estimated using the developed *in silico* model. The detection of the

* Corresponding authors.
E-mail addresses: afantitis@novamechanics.com (A. Afantitis), georgiamelagraki@gmail.com (G. Melagraki).

0223-5234/$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2010.11.029

domain of applicability defined the compounds whose estimations can be accepted with confidence.

2. Material and methods

2.1. Dataset

A series of 62 *N*-phenylanthranilic acids that act as β-amyloid aggregation inhibitors have been collected from the literature [19]. The inhibitors were tested with Beta Amyloid Self Seeding Radioassay (BASSR) and the experimental IC_{50} values are a product of the true affinity of the small molecule, the stoichiometry of binding, and the concentration of the target aggregation intermediate [19]. The compounds are shown in Tables 1–3.

2.2. Descriptors

For each compound we calculated a large number of descriptors that account for their chemical, physicochemical, electronic and quantum characteristics. Following the optimization of the compound geometries using the PM6 method included in MOPAC2009 [29–31], the descriptors were calculated using Chemistry Development Kit (CDK) [32] and MOPAC2009 [29–31] (Table 4). The PM6 method was chosen for the geometry optimizations since it offered a good balance between speed and accuracy [29–31]. A recent paper highlighted the quality of models obtained by PM6 method as similar to that of models based on B3LYP [33] (Density Functional Theory). After the optimization 172 input attributes were calculated for each compound including topological, structural, electronic and physicochemical descriptors. The β-amyloid aggregation inhibition data along with the corresponding full set of descriptor values were used in the variable selection procedure.

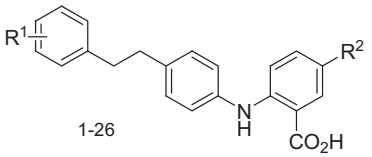
2.3. Separation into a training and a test set

The separation of the dataset into training and test sets was performed according to the popular Kennard and Stones algorithm [34]. The algorithm starts by finding two samples that are the farthest apart from each other on the basis of the input variables in terms of some metric, for example, the Euclidean distance. These two samples are removed from the original dataset and placed into the calibration dataset. This procedure is repeated until the desired number of samples has been reached in the calibration set. The advantages of this algorithm are that the calibration samples map the measured region of the input variable space completely with respect to the induced metric and that the test samples all fall inside the measured region.

2.4. Modeling methods

Variable selection techniques have become an apparent need in many chemoinformatics applications and different methods have been successfully applied as variable selection tools in QSAR problems [35,36]. Before running the model the most significant attributes among the 172 available were pre-selected for the training set using InfoGain variable selection and Ranker evaluator

Table 1
Dataset: ethyl-linked derivatives and 4-nitro substitution. Model predictions using CP-ANN.

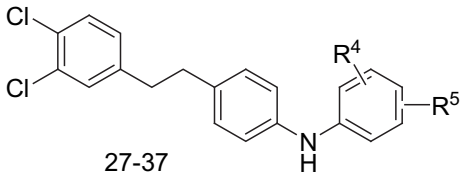


1-26

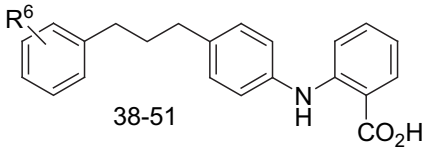
Id	R^1	R^2	BASSR IC_{50} (μM)	Class Threshold (>100 μM)	Training Data (Predicted)	Test Data (Predicted)	Class Weight Active	Class Weight Inactive
1[b]	4-MeO	H	>100	inactive	–	active[c]	0.75	0.25
2	*N*-DHIQ[a]	H	>100	inactive	inactive		0.50	0.50
3	H	H	60	active	inactive[c]		0.33	0.67
4	3,4-Me_2	H	>100	inactive	inactive		0.33	0.67
5	2-Cl	H	>100	inactive	inactive		0.33	0.67
6	3-Cl	H	>100	inactive	inactive		0.04	0.96
7	4-Cl	H	>100	inactive	inactive		0.03	0.97
8[b]	3-F	H	>100	inactive	–	inactive	0.04	0.96
9[b]	3,4-Cl_2	H	>100	inactive	–	inactive	0.01	0.99
10[b]	2,4-Cl_2	H	>100	inactive	–	inactive	0.03	0.97
11	3,5-Cl_2	H	>100	inactive	inactive		0.01	0.99
12	3,4-F_2	H	>100	inactive	inactive		0.01	0.99
13	4-F–3-F_3C	H	>100	inactive	inactive		0.00	1.00
14	3-Cl–4-Me	H	88	active	active		0.99	0.01
15	4-*N*(n–Bu)$_2$	NO_2	15	active	active		1.00	0.00
16	DHIQ[a]	NO_2	5	active	active		1.00	0.00
17	H	NO_2	70	active	active		0.52	0.48
18[b]	3,4-Me_2	NO_2	12	active	–	active	0.84	0.16
19	2-Cl	NO_2	3	active	active		1.00	0.00
20	3-Cl	NO_2	33	active	active		1.00	0.00
21	4-Cl	NO_2	15	active	active		1.00	0.00
22	3,4-Cl_2	NO_2	43	active	active		1.00	0.00
23	2,4-Cl_2	NO_2	17	active	active		1.00	0.00
24	3,4-F_2	NO_2	41	active	active		1.00	0.00
25	4-F–3-F_3C	NO_2	3	active	active		1.00	0.00
26	4-Cl–3-F_3C	NO_2	16	active	active		1.00	0.00

[a] *N*-decahydroisoquinoline.
[b] Test Set.
[c] Misclassified Compound.

Table 2
Dataset: modifications of ethyl-linked and propyl-linked derivatives. Model predictions using CP-ANN.



27-37

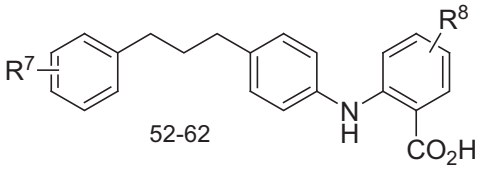


38-51

Id	R^4	R^5	R^6	BASSR IC_{50} (μM)	Class Threshold (>100 μM)	Training Data (Predicted)	TestData (Predicted)	Class Weight Active	Class Weight Inactive
27	H	4-CO_2H	–	>100	inactive	inactive		0.00	1.00
28	H	4-CO_2Me	–	>100	inactive	inactive		0.00	1.00
29[b]	H	5-CO_2H	–	>100	inactive	–	active[c]	0.51	0.49
30	4-F_3C	5-CO_2H	–	>100	inactive	inactive		0.06	0.94
31	4-F	5-CO_2H	–	>100	inactive	inactive		0.09	0.91
32	2-aza[a]	6-CO_2H	–	7	active	active		1.00	0.00
33	2-Me	6-CO_2H	–	>100	inactive	inactive		0.07	0.93
34	3-F_3C	6-CO_2H	–	27	active	active		0.88	0.12
35	2-F_3C	6-CO_2H	–	>100	inactive	inactive		0.12	0.88
36	5-F_3C	6-CO_2H	–	70	active	active		0.92	0.08
37	4-Et_2N	6-CO_2H	–	>100	inactive	inactive		0.08	0.92
38	–	–	4-Et_2N	3	active	inactive[c]		0.50	0.50
39[b]	–	–	4-MeO	30	active	–	active	0.75	0.25
40[b]	–	–	H	14	active	–	active	0.67	0.33
41[b]	–	–	4-Me	9	active	–	active	0.67	0.33
42	–	–	3,4-Me_2	6	active	active		0.67	0.33
43[b]	–	–	3-Br	2	active	–	active	0.67	0.33
44[b]	–	–	2-Cl	6	active	–	active	0.67	0.33
45[b]	–	–	3-Cl	4	active	–	active	0.67	0.33
46[b]	–	–	4-Cl	3	active	–	active	0.67	0.33
47[b]	–	–	3,4-Cl_2	25	active	–	active	0.51	0.49
48	–	–	2,4-Cl_2	2	active	active		0.67	0.33
49[b]	–	–	3,5-Cl_2	3	active	–	active	0.51	0.49
50[b]	–	–	3,4-F_2	8	active	–	active	0.51	0.49
51	–	–	4-F, 3-F_3C	5	active	active		0.81	0.19

[a] Substituted pyridine ring.
[b] Test Set.
[c] Misclassified Compound.

Table 3
Dataset: modifications of propyl-linked derivatives. Model predictions using CP-ANN.



52-62

Id	R^7	R^8	BASSR IC_{50} (μM)	Class Threshold (>100 μM)	Training Data (Predicted)	Test Data (Predicted)	Class Weight Active	Class Weight Inactive
52[a]	3,4-Cl_2	4-F	86	active	–	active	0.81	0.19
53	3,4-Cl_2	4-O_2N	>100	inactive	active[b]		0.52	0.48
54	3,4-Cl_2	4-Me	>100	inactive	active[b]		0.67	0.33
55[a]	3,4-Cl_2	2-aza	60	active	–	active	1.00	0.00
56	3,4-Me_2	4-O_2N	14	active	active		0.92	0.08
57	3,4-Me_2	4-F	>100	inactive	inactive		0.08	0.92
58	3,4-Me_2	2-aza	14	active	active		1.00	0.00
59[a]	3,4-Me_2	3-F	12	active	–	inactive[b]	0.33	0.67
60[a]	3,4-Me_2	4-Me	16	active	–	active	0.67	0.33
61[a]	4-MeO	3-F	14	active	–	active	0.75	0.25
62[a]	4-MeO	4-F	14	active	–	active	0.75	0.25

[a] Test Set.
[b] Misclassified Compound.

Table 4
Selected descriptors with Ranker.

Number	Descriptor
1	LUMO energy (LUMO)
2	WTPT-4
3	TopoPSA
4	HOMO energy (HOMO)
5	WTPT-5
6	nAtomp
7	MDEO-11

which are included in Weka [37,38]. InfoGainAttributeEval evaluates attributes by measuring their information gain with respect to the class [39]. It discriminates numeric attributes first using the Minimum Description Length (MDL)-based discrimination method. Information Gain then assigns a weight to each feature that is used for ranking the features. The specific method can treat presence/absence of a particular term as random variables and computes how much information about class membership is gained by knowing the presence/absence statistics. If the class membership is interpreted as a random variable C with two values, positive and negative, and a word is likewise seen as a random variable *T* with two values, present and absent, then using the information theoretic definition of mutual information, Information Gain can be defined as:

$$IG(t) = H(C) - H(C/T) = \sum_{\tau,C} P(C=c, T=\tau) In[P(C=c, T=\tau)/P(C=c)P(T=\tau)] \quad (1)$$

Ranker, is a ranking scheme for individual attributes. It sorts attributes by their individual evaluations and is used in conjunction with one of the single-attribute evaluators (*e.g.*, InfoGainAttributeEval) [37,38]. Furthermore, the selection procedure is very fast. *Ranker* not only ranks attributes but also performs attribute selection by removing the lower-ranking ones.

In order to develop the classification model we have used the Kohonen and CP-ANN toolbox [26,27] for MATLAB recently introduced by Ballabio *et al.* The algorithm [28] implemented in the toolbox is the one described by Zupan *et al.* Kohonen maps (or Self Organized Maps, SOMs) represent a two-dimensional grid of connected neurons, which are multi-dimensional vectors with a dimension equal to the number of descriptors. The learning of SOM is the projection from multi-dimensional space onto two-dimensional grid (array) of neurons. Kohonen maps have been used with success in the development of quantitative structure–activity relationship models [40–46].

A Kohonen top map is a representation of the space defined by the neurons where the samples are placed. Samples are randomly scattered in the space in a way that similar samples are placed on the same or adjacent neurons and dissimilar samples are placed far apart from each other. This representation allows the interpretation of data structure by analyzing sample position and their relationships. Based on the Kohonen top map we can interpret the sample relationship as well as the variable influence. For each variable encountered the neurons are colored based on the variable values and as a result the influence of each variable on the sample and the relationship between the variable and the sample distribution in the space is defined [26]. Variable interpretation is often considered as a difficult task for multi-dimensional data since the Kohonen map only plots all weights for a specific neuron or all neurons for a specific weight. For complex data a simple visual approach is considered insufficient and a method to investigate the variables in a global way is needed. Principal Component Analysis (PCA) on the Kohonen weights is considered as a trustable alternative. PCA is a variable tool that projects the data in a reduced hyperspace defined by the first significant components [27]. A matrix W that consists of N^2 rows and *J* columns, where *N* is the number of neurons on each side of the map and *J* is the number of variables, is produced so that each element of the matrix represents the weight of a variable in each neuron. PCA on the W matrix produces a loading matrix (*JXF*, where F is the number of significant components) and a score matrix (N^2X*F*). Using the loading and score plots the relationships between variables and neurons and therefore the relationships between variables and classes can be evaluated.

Counterpropagation Artificial Neural Networks (CP-ANN) consists [47] of two layers: the input layer which is a Kohonen network and an associated output layer containing the values of the properties to be predicted (Fig. 1). The Counterpropagation neural network is an extension of Kohonen maps for classification purposes [28], which in addition to the Kohonen layer, it contains a set of output layers, called Grosberg layers. The number of Grosberg layers is equal to the number of classes. As far as the input layer is concerned, training does not differ from SOM, *i.e.,* the input variables determinate the arrangement of objects. When the arrangement is set the positions of objects are projected to the output layer where the weights are modified in such a way that the weights on projected positions correspond to the output values. In addition, the weights in the neighborhood are modified. In this way the response surface is constructed. This part of training is conducted considering the output values and therefore it is usually referred as the supervised part of training of Counterpropagation Artificial Neural Networks (CP-ANN). The prediction of new compounds is performed in two steps. During the first step the object is located into the input layer on the neuron with the most similar weights. In the second step, the position of that neuron is projected to the output layer, which gives the predicted output value.

Both Kohonen maps and CP-ANNs are increasing their uses related to different chemical issues and nowadays can be considered

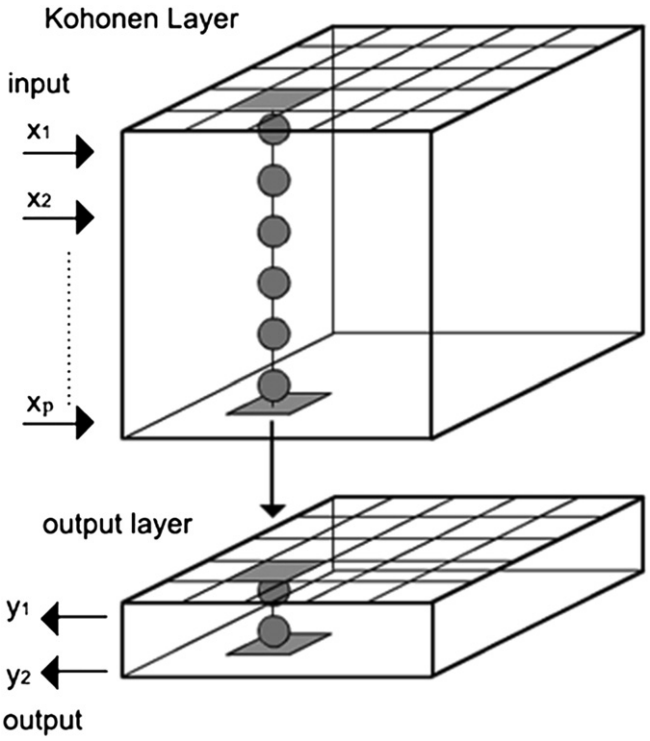


Fig. 1. Kohonen and Counterpropagation Artificial Neural Networks architecture (picture taken from www.disat.unimib.it/chm/).

Table 5
Confusion Matrix (Training Set − Cross-Validation 10 fold) SOM + CP-ANN model.

	Positive Predicted	Negative Predicted
Positive Observed (Active)	14	8
Negative Observed (Inactive)	5	13

as two important tools in chemometrics. One of the reasons of their success is their ability to solve both supervised and unsupervised problems, such as clustering and modeling of both qualitative and quantitative response. Furthermore, according to Vracko [47] in comparison to other neural networks Kohonen maps and CP-ANNs have transparent structure, *i.e.*, the results can be easily interpreted. An advantage is that one can follow the predictions analyzing individual descriptor layers in Kohonen maps and recognize the importance of individual descriptors. The comparison of plots of a descriptor layer and response surface shows the relationship between descriptor and the biological activity under study. The interpretation of results by visualization of Kohonen maps makes the method very appealing for solving drug discovery problems [26]. More specifically graphical representations of individual layers may indicate the roles of individual descriptors in the model. When a new compound is presented to the model it will be located on a defined position in the Kohonen network. Its mechanism of activity may be deduced from the mechanisms of neighboring compounds.

2.5. Validation methods

The validation of the proposed model was assessed by various validation techniques. In particular the proposed classification models were fully validated using the following measurements:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{5}$$

where: TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative.

Confusion Matrix is also given as shown below:

	Positive Predicted	Negative Predicted
Positive Observed (Active)	TP	FN
Negative Observed (Inactive)	FP	TN

Table 6
Confusion Matrix (Test Set) SOM + CP-ANN model.

	Positive Predicted	Negative Predicted
Positive Observed (Active)	16	1
Negative Observed (Inactive)	2	3

Table 7
Specificity, Sensitivity, Precision & Accuracy Statistics (SOM + CP-ANN & J48).

	Specificity	Sensitivity	Precision	Accuracy
Cross-Validation SOM + CP-ANN	0.72	0.64	0.74	0.68
Test Set SOM + CP-ANN	0.60	0.94	0.89	0.86
Cross-Validation J48	0.50	0.73	0.64	0.63
Test Set J48	0.80	0.47	0.89	0.55

2.6. Applicability domain

In order for an *in silico* model to be used for screening new compounds, its domain of application [48–52] must be defined and predictions for only those compounds that fall into this domain may be considered reliable. Similarity measurements were used to define the domain of applicability of the two models based on the Euclidean distances among all training compounds and the test compounds [53]. The distance of a test compound to its nearest neighbor in the training set was compared to the predefined applicability domain (APD) threshold. The prediction was considered unreliable when the distance was higher than APD. APD was calculated as follows:

$$\text{APD} = <d> + Z\sigma \tag{6}$$

Calculation of $<d>$ and σ was performed as follows: First, the average of Euclidean distances between all pairs of training compounds was calculated. Next, the set of distances that were lower than the average was formulated. $<d>$ and σ were finally calculated as the average and standard deviation of all distances included in this set. Z was an empirical cutoff value and for this work, it was chosen equal to 0.5.

3. Results and discussion

The available β-amyloid aggregation inhibition data along with the corresponding full set of 172 descriptor values were used in the variable selection procedure. The original dataset of 62 compounds was split according to the Kennard and Stones [34] algorithm into training and test set. 40 compounds constituted the training set whereas 22 compounds were left for external validation purposes.

In order to select the most significant descriptors we applied InfoGain variable selection and Ranker evaluator (included in WEKA platform) to the training set of molecules. Among the 172 available descriptors 7 were selected as most important to describe the β-amyloid aggregation inhibition. The selected descriptors are the following: LUMO energy (LUMO), WTPT-4, TopoPSA, HOMO energy (HOMO), WTPT-5, nAtomp, MDEO-11.

The descriptors are also presented in Table 4. The chemical meaning of the selected descriptors that will help in the interpretation of the results is briefly described as following:

Molecular orbital (MO) surfaces visually represent the various stable electron distributions of a molecule. According to Frontier Orbital Theory, the shapes and symmetries of the highest-occupied and lowest-unoccupied molecular orbitals (HOMO and LUMO) are crucial in predicting the reactivity of a species and the stereo- and regio- chemical outcome of a chemical reaction [54,55].

LUMO is a measure of electrophilicity of the molecule. Molecules with low LUMO energy values are more able to accept electrons than molecules with high LUMO energy values. The LUMO energy value can be increased (less negative) with the presence of electron-donating groups (EDG) such us amino, and alkoxy groups and decreased (more negative) with the presence of electron-withdrawing groups (EWG) such as halogens, cyano and nitro groups [54]. On the other hand, molecules with high HOMO

Table 8
Applicability domain of the SOM+CP-ANN model for the test set.

Compound	Distance (APD = 5.55)
1	5.54
8	0.00
9	0.01
10	0.01
18	0.04
29	0.13
39	5.54
40	0.03
41	0.02
43	0.02
44	0.02
45	0.01
46	0.03
47	0.06
49	0.06
50	0.05
52	0.06
55	0.11
59	0.06
60	0.08
61	5.54
62	5.53

(highest-occupied molecular orbital energy) values are able to donate electron density more easily than molecules with low HOMO energy values [54] The HOMO energy value can be increased with the presence of electron-donating groups (EDG) such us NMe_2, NH_2, NHEt, and OMe and decreased with the presence of electron-withdrawing groups (EWG) such as halogens and cyano and nitro groups.

Topological Polar Surface Area (TopoPSA) is defined as the part of the surface area of the module associated with nitrogens, oxygens, sulfurs, and the hydrogens bonded to any of these atoms [56,57]. Polar Surface Area is a descriptor that correlates well the passive molecular transport through membranes and allows the prediction of transport properties of drugs. Molecules with a polar surface area of greater than 140 $Å^2$ are usually believed to be poor at permeating cell membranes [54,57].

Weighted path descriptors (WTPT) were introduced by Randic. The scope of these descriptors is to identify a molecule by a single real number with the aim to obtain highly discriminatory power. WTPT-4 and WTPT-5 characterize the sum of path lengths starting from oxygens and nitrogens, respectively. Descriptor nAtomp indicates the number of atoms in the largest π chain of the molecules [54,56].

Molecular distance-edge (MDE) vector is based on the two most fundamental structural variables, one for distance between atoms in the molecular graph and another for edges of the adjacency in the graph. MDEO-11 calculates the molecular distance-edge between all primary oxygens [54,56].

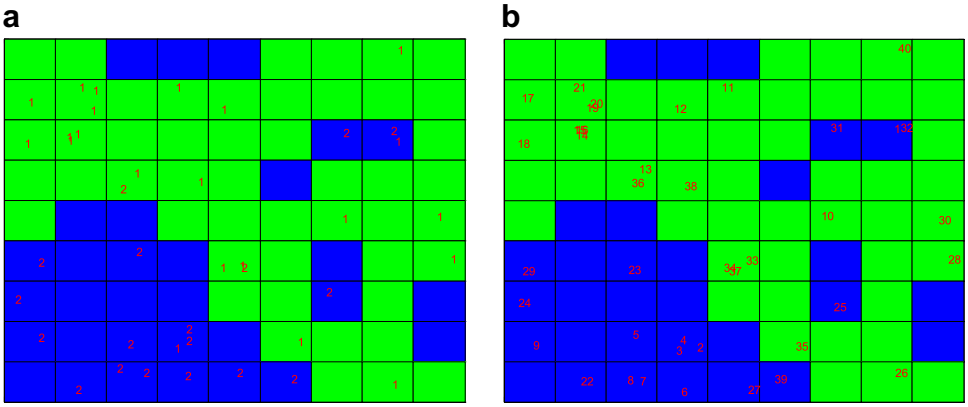
The molecular descriptors used in the model encode information about the structure, branching, electronic effects, and polarity of the modules and thus implicitly account for cooperative effects between functional groups. The proposed model aims to help researchers design novel chemistry driven molecules with desired biological activity.

The above mentioned subset of descriptors was used to develop a predictive model using the Kohonen map and CP-ANN methodology. The model parameters were set to 9 for neurons and 100 epochs for 10 − fold cross-validation. The order of instances was randomized before applying the cross-validation procedure.

The experimental values and the predictions for both training and test examples are presented in Tables 1–3. The confusion matrix for the cross-validation method and model predictions on the external test set are presented in Tables 5 and 6. Validation of the models was performed using the techniques mentioned in the previous section. The significance, accuracy and robustness of the models are illustrated from the corresponding statistics. In particular, the application of the 10 fold cross-validation method produced the following statistics: precision = 74%, sensitivity = 64%, specificity = 72%. Accordingly, by applying the model to the external test set, the following statistical results were obtained: precision = 89%, sensitivity = 94% and specificity = 60%. These statistical results are summarized in Table 7.

The applicability domain was defined for all compounds that constituted the test sets as described in the Materials and Methods section. Since all validation compounds fell inside the domain of applicability, all model predictions for the external test set were considered reliable (Table 8).

The validation of the classification performance using the measurements described above is very important for the evaluation of the model's performance. It is also useful to have an insight into the model which can be achieved by the visualization tools described below. The visual inspection of the Kohonen map helps in the analysis and interpretation of the data structure, the existence of cluster and outliers, the relationship between samples and the influence of each variable.



Class 1= active, Class 2= inactive

Ordinal numbers of training compounds

Fig. 2. Kohonen top map with normal boundary condition for the training set (Green area = active small molecules, Blue area = inactive small molecules). (a) Class 1 = active, Class 2 = inactive. (b) Ordinal numbers of training compounds. (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

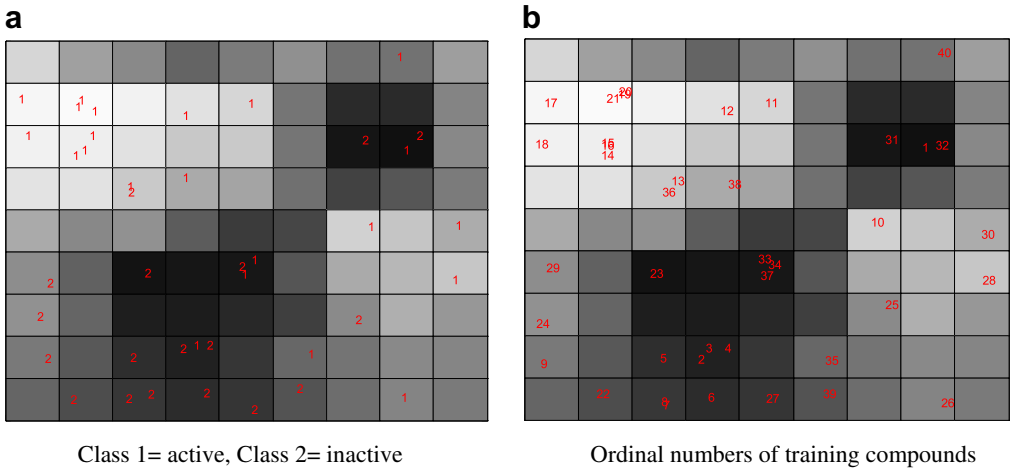


Fig. 3. Kohonen top map with normal boundary condition. Each sample is labelled on the basis of its class; neurons are colored on the basis of Kohonen weight of variable 1 (LUMO energy). (a) Class 1 = active, Class 2 = inactive. (b) Ordinal numbers of training compounds. (for interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

First a Kohonen top map with normal boundary conditions has been produced using the toolbox [26] (Fig. 2). The Kohonen top map has 9 neurons and each sample is labelled on the basis of its class. Class 1 is assigned to actives and Class 2 is assigned to inactives. As it can be seen from Fig. 2 there is a good discrimination between different classes.

Moreover, the Kohonen top map in Fig. 3 illustrates neurons that are colored with a gray scale on the basis of the weight of variable LUMO energy. Coloring of neurons is characteristic of the variable's value. Lighter colors are indicative of low values whereas high values are represented by darker regions. As shown from the coloring of the figure variable LUMO energy is highly discriminative between the two classes.

Fig. 4 shows the analysis of the average weights and standard deviations for all Kohonen neurons corresponding to Class 1 (actives) or Class 2 (inactives). Average Class 1 weight values are significantly higher than Class 2 values for the following descriptors: WTPT-4, TopoPSA, WTPT-5, nAtomp and MDEO-11, while the opposite happens for LUMO energy. As far as HOMO energy is concerned, the average weight values for the two classes almost coincide.

A PCA on Kohonen weights [27] has also been contacted. According to Ballabio *et al.*, the weights of a Kohonen layer can be analyzed by means of Principal Component Analysis (PCA), in order to examine the relationship between variables and classes in a global way and not by examining one variable each time [27]. The Kohonen weights have values in the range of 0–1 and therefore no scaling is needed and PCA has been performed without data pretreatment. A data matrix has been created and PCA was calculated on centered matrix. In Figs. 5 and 6 score and loading plots of the first two components are presented. These first two components explain together 96.86% of the total information. In Fig. 5 each spot corresponds to a neuron of the CP-ANN model. Neurons assigned to Class 2 are all colored and placed at the bottom right side corner. Neurons assigned to Class 1 (actives) are mostly placed at the top left side corner or the lower right side. In the score plot (Fig. 5), each point represents a neuron of the previous CP-ANN model. Each neuron is colored with a gray scale on the basis of the output weight of class 2 (inactive): the larger the value of the output weight, the higher the probability that the neuron belongs to class 2 and the darker the color. Thus, it is easy to see that neurons assigned to class 2 are all clustered and placed at the bottom right side of the score plot. Then, comparing score and loading plots, one can evaluate how all the variables characterize

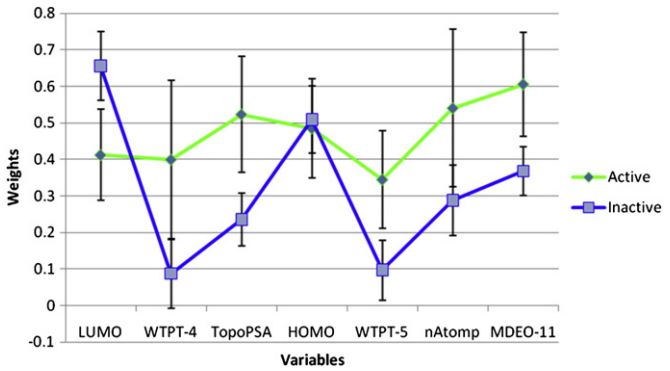


Fig. 4. Profiles of Kohonen weights (average values and standard deviations as error bars) for Class 1 (active) and Class 2 (inactive).

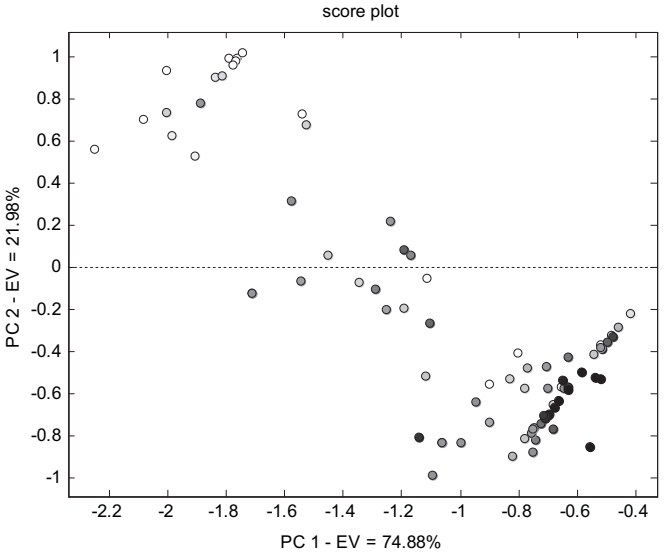


Fig. 5. Score plot of the first two principal components (explaining together 96.86% of the total information). A darker colour indicates a larger value of the output weight and a higher probability of the neuron belonging to inactive class.

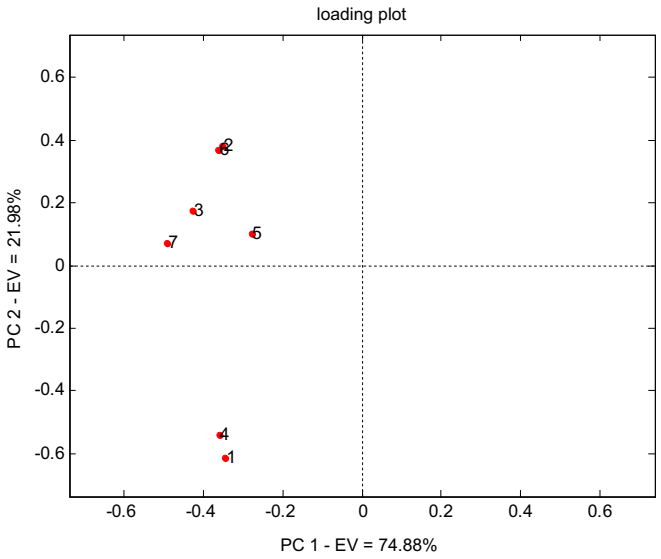


Fig. 6. Loading plot of the first two principal components (explaining together 96.86% of the total information). Each variable is labelled with its identification number and characterize a specific class. Active class: 2 (WTPT-4), 3(TopoPSA), 5(WTPT-5), 6 (nAtomp) and 7(MDEO-11). Inactive class: LUMO (1) and HOMO (4).

this specific class. As shown in Fig. 6, variables LUMO (1) and HOMO (4) are placed at the bottom left of the loading plot and thus is directly correlated to the Class 2(inactive). On the contrary loading such as 2(WTPT-4), 3(TopoPSA), 5(WTPT-5), 6(nAtomp) and 7 (MDEO-11) have positive loadings in the first component and thus samples of Class 2 will be characterized by small values of these variables.

To further validate the performance of the method that combines Kohonen maps and CP-ANNs, we compared it with a popular decision tree classifier, namely the J48 decision tree [37,38] (included in Weka). In order to classify a new item, the J48 method first needs to create a decision tree based on the attribute values of the available training data and identifies the attributes that discriminates the various instances most clearly. The specific classifier has the ability to do simultaneously variable selection and modeling. Among the 172 available descriptors 5 were selected as most important to describe the β-amyloid aggregation inhibition. The selected descriptors are the following: LUMO energy, BCUTw-1h (eigenvalue based descriptor noted for its utility in chemical diversity described by Pearlma), FPSA-3 (encodes surface area and partial charge information), ATSc3 (an autocorrelation Moreau-Broto descriptor)

Table 9
Confusion Matrix (Training Set – Cross-Validation 10 fold) J48.

	Positive Predicted	Negative Predicted
Positive Observed (Active)	16	6
Negative Observed (Inactive)	9	9

Table 10
Confusion Matrix (Test Set) J48.

	Positive Predicted	Negative Predicted
Positive Observed (Active)	8	9
Negative Observed (Inactive)	1	4

and finally TopoPSA (topological polar surface area). It is notable to mention that LUMO energy and Topological polar surface area (TopoPSA) have been selected from both algorithms. The produced decision tree is shown in Fig. 7. The confusion matrix for the cross-validation method and model predictions on the external test set are presented in Tables 9 and 10. The J48 decision tree model produced the following statistics after applying the 10 fold cross-validation method: precision = 64%, sensitivity = 73%, specificity = 50%. The respective statistics for validation with the external test set were: precision = 89%, sensitivity = 47% and specificity = 80%. In order to directly compare the two methods the statistical results of the J48 training method are also shown in Table 7.

3.1. In silico virtual screening

The proposed method, due to the high predictive ability [58,59] and simplicity, can be a useful aid to the costly and time-consuming experiments for the synthesis and the determination of the amyloid aggregation inhibition of *N*-phenylanthranilic acids. A virtual screening procedure [58,60] could be based on the proposed model. The design of novel active molecules by the insertion, deletion, or modification of substituents on different sites of the molecule and at different positions could therefore be guided by the proposed model.

To demonstrate the usefulness of the CP-ANN model we subsequently conducted a virtual screen to identify potential new active targets within the models domain of applicability threshold (*ca.* 5.55). To guide the selection of virtual molecules the descriptors used to construct the model were referred to where possible.

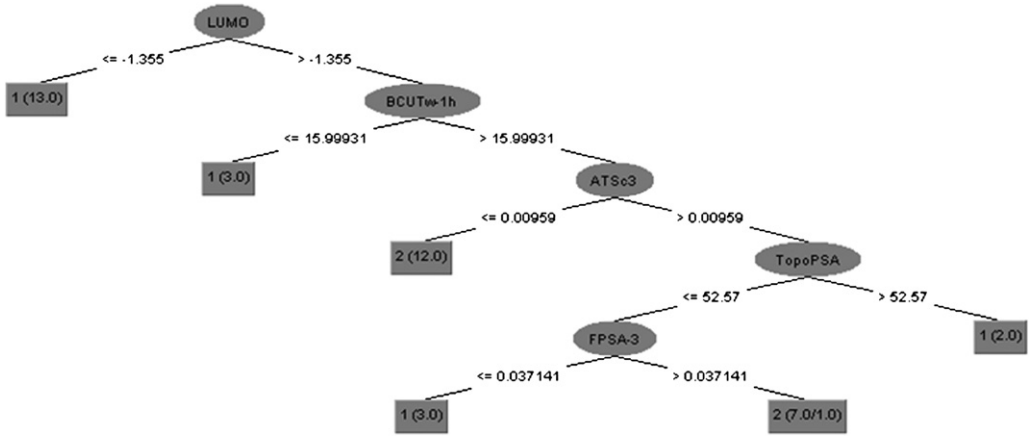
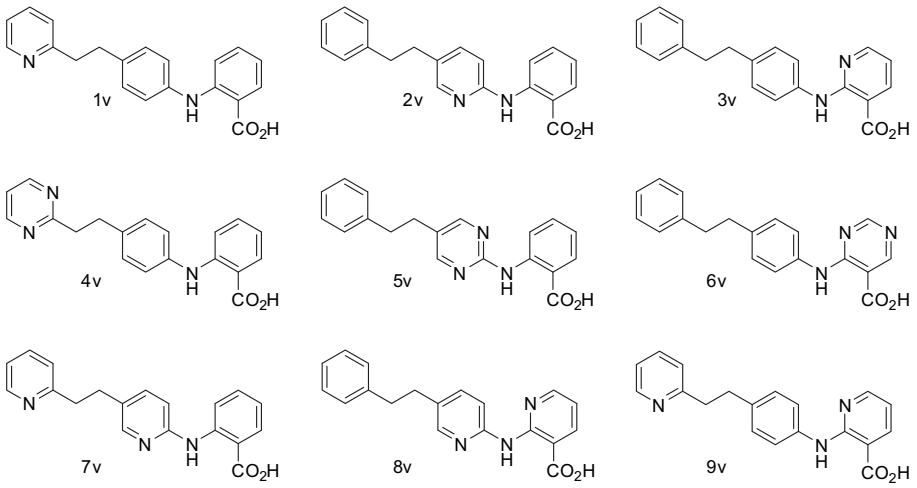


Fig. 7. Graphical representation of the J48 decision tree (Class 1 = active, Class 2 = inactive).

Table 11
Virtual screening, compounds **1v**–**9v**.



Id	Activity (Predicted)	Class Weight Active	Class Weight Inactive	Applicability Domain Distance (APD = 5.55)
1v	Active	0.79	0.21	0.41
2v	Active	0.79	0.21	0.30
3v	Active	1.00	0.00	0.20
4v	(Active)	1.00	0.00	13.24
5v	(Active)	1.00	0.00	13.27
6v	(Active)	1.00	0.00	13.25
7v	(Active)	1.00	0.00	13.26
8v	(Active)	1.00	0.00	13.26
9v	(Active)	1.00	0.00	13.25

(Active): Compound predicted active but falls outside the applicability domain.

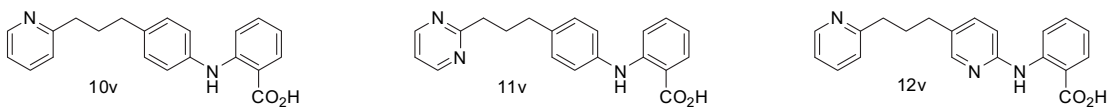
The primary objective of the *in silico* screen was to determine whether the developed *in silico* model could classify structures that are not included in the training or test sets, as active or inactive. The secondary objective was to identify which structural modifications could be tolerated using the domain of applicability. The ultimate role of the *in silico* screen was as a guide to the identification of the most promising new synthetic targets.

Rather arbitrarily we chose 2-(4-phenethylphenylamino)benzoic acid **3** as a starting point. This compound was known to have moderate activity (Table 1, ID **3**). According to the descriptors used to construct the model, more negative HOMO-LUMO energies were preferred for active compounds. As such, we replaced various arene CH's for N's to give compounds **1v**–**9v** (Table 11).

The model identified all compounds as active, however, only the predictions on the mono aza analogues **1v**–**3v** fell within the domain of applicability of the model (<5.55). Interestingly, the location of the N made little noticeable difference to the predicted activity and domain of applicability values. Increasing the length of the saturated alkyl spacer for selected aza substituted molecules improved slightly the outcome for the diaza derivatives, which remained active but well out of the domain of applicability. However, for the mono aza substituted analogue the longer propyl spacer was well tolerated (Table 12). The descriptor nAtomp also indicated that molecules with extended π chains could also show the desired activity. As such the structure 2-(4-phenethylphenylamino)benzoic acid **3** was modified to provide several new potential targets **13v**–**21v** (Table 13).

Among these two compounds, 2-(4-phenethylnaphthalen-1-ylamino)benzoic acid **17v** and 2-(4-phenethylphenylamino)-1-naphthoic acid **18v** were identified as active and within the domain of

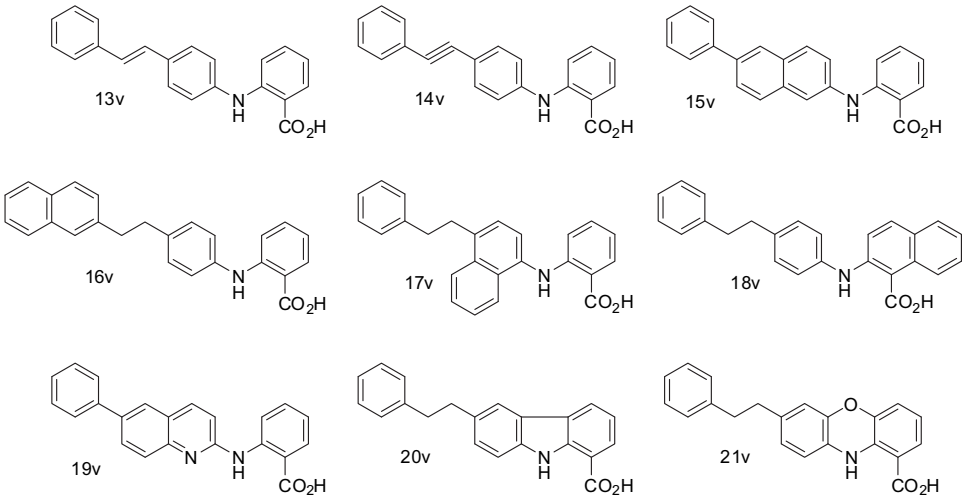
Table 12
Virtual screening, compounds 10v-12v.



Id	Activity (Predicted)	Class Weight Active	Class Weight Inactive	Applicability Domain Distance (APD = 5.55)
10v	Active	0.79	0.21	0.42
11v	(Active)	1.00	0.00	13.24
12v	(Active)	1.00	0.00	13.25

(Active): Compound predicted active but falls outside the applicability domain.

Table 13
Virtual screening, compounds **13v**–**21v**.

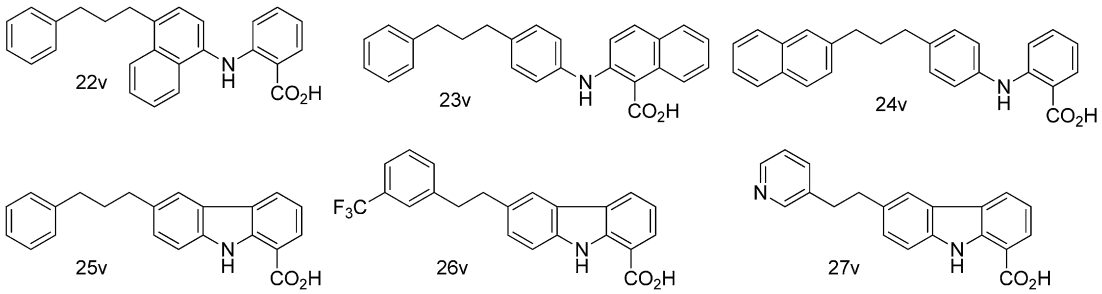


Id	Activity (Predicted)	Class Weight Active	Class Weight Inactive	Applicability Domain Distance (APD = 5.55)
13v	(Inactive)	0.50	0.50	8.00
14v	(Inactive)	0.50	0.50	8.00
15v	(Active)	0.92	0.08	10.00
16v	Inactive	0.33	0.67	0.02
17v	Active	0.75	0.25	2.00
18v	Active	0.57	0.43	4.00
19v	(Active)	0.92	0.08	10.00
20v	Inactive	0.09	0.91	3.38
21v	(Active)	0.57	0.43	5.83

(Active): Compound predicted active but falls outside the applicability domain (Inactive): Compound predicted inactive but falls outside the applicability domain.

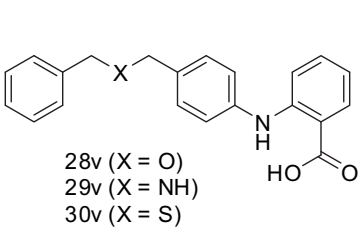
applicability (2.00 and 4.00, respectively). While 2-{4-[2-(naphthalen-2-yl)ethyl]phenylamino}-benzoic acid **16v** was clearly identified as inactive within the domain of applicability (0.02). Alternative fusions that extended the π chain length were also fruitful identifying a clear inactive molecule 6-phenethyl-9*H*-carbazole-1-carboxylic acid **20v** and an active 7-phenethyl-10*H*-phenoxazine-1-carboxylic acid **21v**, that fell marginally out of the domain of applicability (5.83). An attempt to bring the active prediction of compound **15v** within the domain of applicability by introducing an arene CH by N to give compound **19v** that should have more negative HOMO/LUMO values failed to provide an active molecule within the domain of applicability (Table 13).

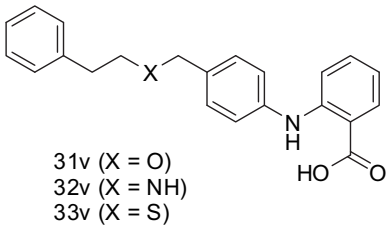
Table 14
Virtual screening, compounds **22v**–**27v**.



Id	Activity (Predicted)	Class Weight Active	Class Weight Inactive	Applicability Domain Distance (APD = 5.55)
22v	Active	0.57	0.43	4.00
23v	Active	0.57	0.43	4.00
24v	Inactive	0.33	0.67	0.02
25v	Inactive	0.09	0.91	3.38
26v	Active	0.53	0.47	3.42
27v	Active	1.00	0.00	3.76

Table 15
Virtual screening, compounds **28v**–**33v**.





Id	Activity (Predicted)	Class Weight Active	Class Weight Inactive	Applicability Domain Distance (APD = 5.55)
28v	(Active)	0.75	0.25	5.69
29v	Active	0.51	0.49	0.93
30v	(Active)	0.67	0.33	12.79
31v	(Active)	0.75	0.25	5.69
32v	Active	1.00	0.00	0.90
33v	(Active)	0.67	0.33	12.79

(Active): Compound predicted active but falls outside the applicability domain.

Similarly, switching from an ethyl to propyl linker made little difference to the predicted activities (Table 14). Nevertheless, by modifying the carbazole side chain the predicted activity could be switched from inactive to active within the domain of applicability. Introduction of pyridyl or trifluoromethylphenyl substituents was favourable, while increasing the length of the saturated linker was not (Table 14).

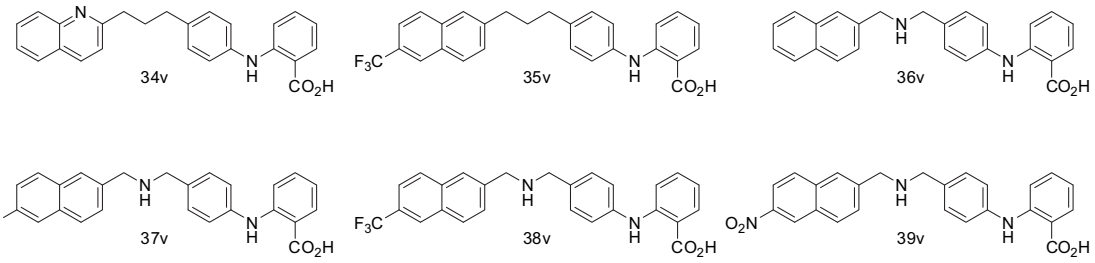
We then examined the effect of introducing heteroatom spacers in the saturated alkyl spacer. Oxygen, nitrogen and sulfur units were introduced into both propyl and butyl chains to give molecules **28v**–**33v** (Table 15). While all of the molecules were active only the nitrogen derivatives **29v** and **32v** were within the models domain of applicability.

By combining the heteroalkyl linkage with extended π-systems such as the naphthenes we were able to investigate the effects further. We focused here on the virtual molecule **24v** that had been predicted to be inactive and clearly within the model's domain of applicability. Our intension was to investigate the effects of minor modifications such as exchange of arene CH's for N, alkyl CH for N, and direct arene substitution (Table 16).

The simple replacement of an aryl or alkyl CH by N switched the activity providing active molecule **34v** and **36v**, respectively both within the domain of applicability. The introduction of a trifluoromethyl substituent on the naphthyl portion to give molecule **35v** led to an inactive prediction again within the domain, however, combining the introduction of a similar trifluoromethyl or even a fluoro group with the exchange of an alkyl CH for N led to active compounds **37v** and **38v** both within the model's domain of application.

It can be observed that the model acts to identify molecules predicted to be either active or inactive that could not be readily differentiated based on simple intuition. The domain of applicability provides a level of confidence that can be used to prioritise the selection of targets to be synthesized.

Table 16
Virtual screening, compounds **34v**–**39v**.



Id	Activity (Predicted)	Class Weight Active	Class Weight Inactive	Applicability Domain Distance (APD = 5.55)
34v	Active	0.79	0.21	0.35
35v	Inactive	0.00	1.00	0.07
36v	Active	0.51	0.49	0.92
37v	Active	1.00	0.00	0.89
38v	Active	1.00	0.00	0.87
39v	(Active)	0.55	0.45	7.38

(Active): Compound predicted active but falls outside the applicability domain.

4. Conclusions

A classification model for the prediction of β-amyloid aggregation of *N*-phenyl-anthranilic acid inhibitors was developed. After the calculation of a large number of descriptors, we selected the most significant for each compound that fully describe the characteristics responsible for the inhibition activity under study. Based on this dataset, we used Kohonen maps and CP-ANN methodology which resulted in the development of an accurate and reliable model that was fully validated using various cross-validation and external test prediction techniques. The results were interpreted based on the physical meaning of descriptors and by visualization of Kohonen maps. The method [61–63] can also be used to screen existing databases [64–66] or virtual combinations to identify derivatives with desired activity. In this scenario, the classification model will be used to screen out inactive compounds, while the applicability domain will serve as a valuable tool to filter out "dissimilar" combinations. An attempt in this direction was carried out. Synthesis of the proposed chemistry driven small molecules using the aforementioned virtual screening procedure and experimental evaluation of their biological activity will show if the method can be used as a general rational drug discovery tool.

Acknowledgements

This work was supported by funding from the Cyprus Research Promotion Foundation [Grant YGEIA/BIOS/0308(BIE)/13]. P.A.K thanks the following organizations in Cyprus for generous donations of chemicals and glassware: the State General Laboratory, the Agricultural Research Institute and the Ministry of Agriculture. Furthermore we thank the A.G. Leventis Foundation for helping to establish the NMR facility in the University of Cyprus. The authors thank Prof. Alexander Tropsha for helpful discussions.

Appendix. Supplementary material

Supplementary data related to this article can be found online at doi:10.1016/j.ejmech.2010.11.029.

References

[1] C. Haass, D.J. Selkoe, Nat. Rev. Mol. Cell Biol. 8 (2007) 101–112.
[2] F. Chiti, C.M. Dobson, Annu. Rev. Biochem. 75 (2006) 333–336.
[3] B.J. Blanchard, G. Konopka, M. Russell, V.M. Ingram, Brain Res. 776 (1997) 40–50.
[4] B.J. Blanchard, A.E. Hiniker, C.C. Lu, Y. Margolin, A.S. Yu, V.M. Ingram, J. Alzheimer's Dis. 2 (2000) 137–149.
[5] L.D. Estrada, C. Soto, Curr. Top. Med. Chem. 7 (2007) 115–126.
[6] S. Cellamare, A. Stefanachi, D.A. Stolfa, T. Basile, M. Catto, F. Campagna, E. Sotelo, P. Acquafredda, A. Carotti, Bioorg. Med. Chem. 16 (2008) 4810–4822.
[7] F. Campagna, F. Palluotto, A. Carotti, G. Casini, G. Genchi, Bioorg. Med. Chem. 7 (1999) 1533–1538.
[8] M. Catto, R. Aliano, A. Carotti, S. Cellamare, F. Palluotto, R. Purgatoriox, R.A. De Stradis, F. Campagna, Eur. J. Med. Chem. 45 (2010) 1359–1366.
[9] F. Leonetti, M. Catto, O. Nicolotti, L. Pisani, A. Cappa, A. Stefanachi, A. Carotti, Bioorg. Med. Chem. 16 (2008) 7450–7456.
[10] M.T.H. Khan, N. Biotechnol. 25 (2009) 331–346.
[11] S. Deswal, N. Roy, Eur. J. Med. Chem. 42 (2007) 463–470.
[12] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, Bioorg. Med. Chem. 15 (2007) 7237–7247.
[13] A. Afantitis, G. Melagraki, H. Sarimveis, H.O. Igglessi-Markopoulou, G. Kollias, Eur. J. Med. Chem. 44 (2009) 877–884.
[14] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, G. Kollias, O. Igglessi-Markopoulou, Mol. Divers. 13 (2009) 301–311.
[15] E.B. de Melo, M.M.C. Ferreira, QSAR Comb. Sci. 28 (2009) 1156–1165.
[16] E.B. de Melo, M.M.C. Ferreira, Eur. J. Med. Chem. 44 (2009) 3577–3583.
[17] Z. Gong, R. Zhang, B. Xia, R. Hu, B. Fan, QSAR Comb. Sci. 27 (2008) 1282–1290.
[18] A.A. Toropov, A.P. Toropova, E. Benfenati, Mol. Divers. 14 (2010) 183–192.
[19] L.J. Simons, B.W. Caprathe, M. Callahan, J.M. Graham, T. Kimura, Y. La, H. LeVine III, W. Lipinski, A.T. Sakkab, Y. Tasaki, L.C. Walker, T. Yasunaga, Y. Ye, N. Zhuang, C.E. Augelli-Szafran, Bioorg. Med. Chem. Lett. 19 (2009) 654–657.
[20] A. Nargotra, S. Sharma, J.L. Koul, P.L. Sangwan, I.A. Khan, A. Kumar, S.C. Taneja, S. Koul, Eur. J. Med. Chem. 44 (2009) 4128–4135.
[21] B. Xia, W. Ma, B. Zheng, X. Zhang, B. Fan, Eur. J. Med. Chem. 43 (2008) 1489–1498.
[22] A.A. Toropov, A.P. Toropova, E. Benfenati, Eur. J. Med. Chem. 44 (2009) 2544–2551.
[23] S. Durdagi, T. Mavromoustakos, M.G. Papadopoulos, Bioorg. Med. Chem. Lett. 18 (2008) 6283–6289.
[24] J. Li, P. Gramatica, Mol. Divers. (2010). doi:10.1007/s11030-009-9212-2.
[25] A. Nowaczyk, K. Kulig, B. Malawska, QSAR Comb. Sci. 28 (2009) 979–988.
[26] D. Ballabio, V. Consonni, R. Todeschini, Chemom. Intell. Lab. Syst. 98 (2009) 115–122.
[27] D. Ballabio, R. Kokkinofta, R. Todeschini, C.R. Theocharis, Chemom. Intell. Lab. Syst. 87 (2007) 78–84.
[28] J. Zupan, M. Novic, J. Gasteiger, Chemom. Intell. Lab. Syst. 27 (1995) 175–187.
[29] MOPAC2007, Stewart J.J.P., Stewart Computational Chemistry, Version 7.295W, http://www.openmopac.net
[30] J.J.P. Stewart, J. Mol. Model. 13 (2007) 1173–1213.
[31] J.J.P. Stewart, J. Mol. Model. 14 (2008) 499–535.
[32] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E.L. Willighagen, J. Chem. Inf. Comput. Sci. 43 (2003) 493–500.
[33] T. Puzyn, N. Suzuki, M. Haranczyk, J. Rak, J. Chem. Inf. Model. 48 (2008) 1174–1180.
[34] R.W. Kennard, L.A. Stone, Technometrics 11 (1969) 137–148.
[35] I. Kuzmanovski, M. Novic, M. Trpkovska, Anal. Chim. Acta 642 (2009) 142–147.
[36] F. Marini, A. Roncaglioni, M. Novic, J. Chem. Inf. Mod. 45 (2005) 1507–1519.
[37] Ian H. Witten, Eibe Frank, Data mining, practical machine learning tools and techniques Microsoft Research. in: Jim Gray (Ed.), second ed, The Morgan Kaufmann Series in Data Management Systems. Elsevier, 2005.
[38] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, SIGKDD Explor. 11 (1) (2009).
[39] X. Geng, T.-Y. Liu, T. Qin, H. Li, Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07, 2007, pp 407–414.
[40] K. Roy, I. Mitra, Expert Opin. Drug Deliv. 11 (2009) 1157–1175.
[41] E. Papa, J.C. Dearden, P. Gramatica, Chemosphere 67 (2007) 351–358.
[42] N. Fjodorova, M. Vračko, M. Tušar, A. Jezierska, M. Novič, R. Kühne, G. Schüürmann, Mol. Divers. (2009). doi:10.1007/s11030-009-9190-4.
[43] Z. Wang, A. Yan, Q. Yuan, J. Gasteiger, Eur. J. Med. Chem. 43 (2008) 2442–2452.
[44] I. Kuzmanovski, M. Novic, Chemom. Intell. Lab. Syst. 90 (2008) 84–91.
[45] A. Roncaglioni, M. Novic, M. Vracko, E. Benfenati, J. Chem. Inf. Comput. Sci. 44 (2004) 300–309.
[46] S. Mazurek, T.R. Ward, M. Novic, Mol. Divers. 11 (2007) 141–152.
[47] M. Vracko, Curr. Comput. Aided Drug Des. 1 (2005) 73–78.
[48] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, O. Igglessi-Markopoulou, G. Kollias, Mol. Divers. 14 (2010) 225–235.
[49] H. Liu, X. Yao, P. Gramatica, Comb. Chem. High Throughput Screen. 12 (2009) 490–496.
[50] E. Papa, S. Kovarich, P. Gramatica, QSAR Comb. Sci. 28 (2009) 790–796.
[51] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, O. Igglessi-Markopoulou, G. Kollias, Chem. Biol. Drug Des. 76 (2010) 397–406.
[52] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, QSAR Comb. Sci. 27 (2008) 432–436.
[53] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, J. Chem. Inf. Model. 46 (2006) 1984–1995.
[54] R. Todeschini, V. Consonni, R. Mannhold, in: H. Kubinyi, H. Timmerman (Eds.), Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[55] G. Melagraki, A. Afantitis, K. Makridima, J. Markopoulos, O. Igglessi-Markopoulou, H. Sarimveis, J. Mol. Model. 12 (2006) 297–305.
[56] J. Devillers, A.T. Balaban, Topological Indices and Related Descriptors in QSAR and QSPR. Gordon and Breach Science Publishers, The Netherlands, 1999.
[57] P. Ertl, B. Rohde, P. Selzer, J. Med. Chem. 43 (2000) 3714–3717.
[58] G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou, A. Alexandridis, Mol. Divers. 10 (2006) 213–221.
[59] A. Tropsha, A. Golbraikh, Curr. Pharm. Des. 13 (2007) 3494–3504.
[60] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, J. Comput. Aided Mol. Des. 21 (2007) 251–267.
[61] L.B. Salum, A.D. Andricopulo, Mol. Divers. 13 (2009) 277–285.
[62] P.A. Babu, D.J. Smiles, M.L. Narasu, K. Srinivas, QSAR Comb. Sci. 27 (2008) 1362–1373.
[63] S. Paliwal, A. Narayan, S. Paliwal, QSAR Comb. Sci. 28 (2009) 1367–1375.
[64] M. Gredičak, F. Supek, M. Kralj, Z. Majer, M. Hollósi, T. Šmuc, K. Mlinarić-Majerski, S. Horvat, Amino Acids 38 (2010) 1185–1191.
[65] F. Supek, M. Kralj, M. Marjanović, L. Šuman, T. Šmuc, I. Krizmanić, B. Žinić, Invest. .New Drugs 26 (2008) 97–110.
[66] M. Marjanović, M. Kralj, F. Supek, L. Frkanec, I. Piantanida, T. Šmuc, L. Tušek-Božić, J. Med. Chem. 50 (2007) 1007–1018.