Eur. J. Med. Chem. 36 (2001) 349–359

PII: S0223-5234(01)01233-8/FLA

Original article

Database mining applied to central nervous system (CNS) activity

Marco Pintore, Olivier Taboureau, Frédéric Ros, Jacques R. Chrétien*

Laboratory of Chemometrics and BioInformatics, University of Orléans, BP 6759, F-45067 Orleans Cedex 2, France

Received 12 October 2000; revised 8 March 2001; accepted 8 March 2001

Abstract – A data set of 389 compounds, active in the central nervous system (CNS) and divided into eight classes according to the receptor type, was extracted from the RBI© database and analyzed by Self-Organizing Maps (SOM), also known as Kohonen Artificial Neural Networks. This method gives a 2D representation of the distribution of the compounds in the hyperspace derived from their molecular descriptors. As SOM belongs to the category of unsupervised techniques, it has to be combined with another method in order to generate classification models with predictive ability. The fuzzy clustering (FC) approach seems to be particularly suitable to delineate clusters in a rational way from SOM and to get an automatic objective map interpretation.

Maps derived by SOM showed specific regions associated with a unique receptor type and zones in which two or more activity classes are nested. Then, the modeling ability of the proposed SOM/FC Hybrid System tools applied simultaneously to eight activity classes was validated after dividing the 389 compounds into a training set and a test set, including 259 and 130 molecules, respectively. The proper experimental activity class, among the eight possible ones, was predicted simultaneously and correctly for 81% of the test set compounds. © 2001 Éditions scientifiques et médicales Elsevier SAS

database mining / self-organizing map / fuzzy clustering / molecular diversity / CNS activity

1. Introduction

Combinatorial chemistry technologies [1, 2] and high throughput screening (HTS) [3, 4] are simultaneously used by all major pharmaceutical companies in order to isolate new active leads. These discovery strategies generate an impressive number of compound libraries. To maximize molecular diversity and to reduce overall costs, a pre-selection of the compounds to be screened should be envisaged. Indeed, a high molecular diversity increases the probabilities to find a new pharmacological lead by HTS, avoiding testing many redundant compounds. This objective can be achieved by designing and classifying large chemical libraries in order to get enhanced information content. With help of automated data classification methods, sets of molecular descriptors representing the compounds included in a chemical database are correlated to a given molecular property, and predictive models able to select untested molecules with required physicochemical characteristics or biological activities can be derived [5–7].

For chemical database mining (DBM) derived from the exploitation of molecular diversity, artificial neural networks (ANN) exhibit a powerful pattern recognition paradigm, able to analyze various types of data [8–10]. Self-Organizing Maps (SOM) [11], also known as Kohonen ANN, offer new means for DBM related to bio-active compounds [12–15]. Compound databases are represented in a hyperspace derived from their molecular descriptors and the purpose of SOM is to give a 2D map representation of the compound distribution in the hyperspace. This technique has been used to process huge amounts of data in a high dimension space [16]. However, it must be underlined that SOM belongs to the category of unsupervised techniques and it cannot be used directly for predictive aims.

To resolve this classification drawback, a possible solution consists in combining SOM with another technique, i.e. generating a hybrid system that pre-

* Correspondence and reprints.
E-mail address: jacques.chretien@univ-orleans.fr (J.R. Chrétien).

serves the user-friendly interpretation of Kohonen maps. The fuzzy set theory introduced by Zadeh in 1960 is worth investigating [17, 18]. It can provide very simple and interpretable solutions to classification problems within the context of imprecise categories. Fuzzy systems have been widely used in the field of process control where the idea is to convert human expert knowledge into fuzzy rules. More recently, several pattern recognition methods have been proposed to automatically generate fuzzy rules from numerical data [19]. This field of research is recent but very promising and it can be related successfully to the exploitation of chemical databases.

The aim of this work was to apply these DBM tools to the field of central nervous system (CNS) active compounds. CNS is a source of wide researches to generate new molecules in the care of diffuse neurology pathologies [20–24]. Compounds active on CNS receptors should readily permeate across the blood–brain barrier (BBB) [25, 26] and the CNS activity seems influenced by several physicochemical properties, such as molecular size and shape, lipophilicity, hydrogen-bonding and charge [27–29].

So far, predictive models of CNS activity have mainly been established through standard QSAR approaches on small series of a few dozen related compounds [30–33]. But the field of application of such QSAR approaches is clearly inappropriate if the aim is to design diverse combinatorial libraries. With this object, Bauknecht et al. [13] showed, by SOM, that two families of dopamine and benzodiazepine agonist compounds can still be distinguished if they are merged in a larger data set of 8323 compounds comprising a wide structural variety. Then, the projection of untested molecules in a self-organizing neural network could be used for searching for structural similarity and, consequently, new leads with CNS activity. But, SOM being an unsupervised method, this technique can be used only if a 'natural' regrouping of the compounds is evident and each 'cluster' is well separated from the others.

Finally, a recent paper of Ajay et al. [34] proposed the first solution to systematically design a CNS-active library, using an ANN trained by Bayesian method to classify a database including 15 000 active and 50 000 inactive compounds. The method shows a good ability to predict correctly which compounds present CNS activity, but was limited to active/inactive classification, neglecting the specific CNS mechanism, which is nevertheless the ultimate goal of a DBM strategy. Such an approach belongs to the first generation procedures.

The present paper aims to develop second generation procedures in order to go a step ahead in the virtual screening of new compounds. The proposed improvement consists in classifying the CNS-active compounds according to the different CNS receptors on which they could act. The classification problem is made more complex by the fact that the same chemical can be active on different receptors. Fuzzy techniques, which are based on the possibility to handle the 'concept of partial truth', should provide solutions to this problem, inserted in the context of 'imprecise categories'. Then, a hybrid system integrating SOM and fuzzy clustering (FC), just developed elsewhere for olfaction [35], was applied on a data set of 389 active molecules, acting on eight types of receptors in order to test the applicability and predictive ability of this new DBM procedure in medicinal chemistry.

2. Materials and methods

2.1. Compound selection

From the RBI database [36, 37], focused on receptor-based classification and signal transduction in nervous system, all peptides and enzymatic inhibitors were suppressed, keeping only 389 molecules acting as selective receptor agonists or antagonists. This data set was divided into eight classes, according to the receptors involved: (i) adenosines and purinergics; (ii) adrenergics; (iii) histaminergics; (iv) cholinergics; (v) dopaminergics; (vi) glutamatergics; (vii) serotonergics; and (viii) GABAergics. As the number of purinergic molecules in the data set is very low and as their structure is very similar to those of the adenosines molecules, these two classes were grouped together.

2.2. Molecular descriptors

The chemical data sets were distributed in a 205 multidimensional hyperspace derived from a selected set of 205 different molecular descriptors. This set of 205 descriptors includes topological, physicochemical and electronic parameters [38–40]. In virtual screening, gen-

eral descriptors have proved a good compromise for data mining in large databases in terms of efficiency. The advantage of these descriptors is their ability to take into account not only the main structural features of each molecule, but also their global behaviors.

Molar refractivity, molar volume, molecular weight and van der Waals volume were used as size descriptors. The shape features of the molecules were characterized by Kappa shape indices, which account for the ramification degree, the oblong character, etc. The degree of branching and cyclization in the compounds were described by 49 molecular connectivity indices (χ), whereas the number of rings, count of paths and clusters were taken into account by 17 subgraph count indices. Furthermore, seven information contents descriptors were used as well as the Wiener index (W), centric index (C), Balaban index (J), Gutman index ($M2$), Platt number (F). A lipophilicity descriptor represented by the octanol–water partition coefficient ($\log P_{\mathrm{oct/water}}$) was calculated using the Hansch and Leo method [41]. Another descriptor was derived from the electronegativity of molecules ($E_{\mathrm{M}}^{\mathrm{S}}$) by the Sanderson method [42].

Finally, the electrotopological state (E-state) indices were considered in the computations [40]. This family of atom-level molecular descriptors defines the electron density at each atom or hybrid group in a molecule, and the contribution of these electrons in the intermolecular interactions. The indices are computed by combining the electronic character and the topological environment for each atom in the molecule.

2.3. Self-organizing maps (SOM)

SOM [11], also known as Kohonen Neural Networks, is a non-linear mapping technique which gives a 2D space representation of a given set of points from a multidimensional space derived from a large series of molecular descriptors. Each point of this set is related to a SOM node. The co-ordinates of these nodes are represented by the weighted connections between the input layers and SOM. Two close points of the original data set occupy the nearest node in the SOM layer, in terms of Euclidean distance. Hence, the neighboring points of the original descriptor space remain neighbors on SOM. Training SOM consists in rearranging the Kohonen layer nodes by gradually adjusting their weights. A useful option, 'interpolation', allows us to convert the discrete coordinates of the Kohonen layer nodes into continuous coordinate values, averaging the discrete coordinates of the three nodes closest to the raw data point. A detailed description of the algorithm used in this study may be found in Ref. [43].

The following parameters were used in data processing: number of columns = 10, number of rows = 10, coefficient γ for bias calculations = 0.01, number of iterations for the training phase = 50 000, coefficient β for frequency calculations = 0.001, with the activation of the interpolate option. The calculations were performed using proprietary software.

2.4. Fuzzy clustering

In a Kohonen map, if a 'natural' regrouping of the compounds is evident and if each cluster is well separated from the others, a visual examination can be sufficient to analyze the data organization. But, generally, personal subjectivity can influence the conclusions related to the compound distribution. However, the power of SOM method is going beyond a simple map examination.

A hybrid system consists here in linking SOM to another technique. It aims to build a tool that keeps the user-friendly visual aspect of SOM and, at the same time, offers an automatic objective map interpretation. In this system, the coordinates of SOM projections of the points included in the molecular descriptor hyperspace constitute the input of the new classification technique. The FC approach [17, 18], amidst several methods that can be considered [44, 45], seems to be particularly suitable to delineate clusters from SOM in a rational way.

An FC algorithm, where clusters are derived from fuzzy sets [17], can be considered as a generalization of the traditional cluster procedures. Such clusters are derived assigning to each compound a number between zero and one called the membership degree of the object. A compound is defined by the degree of membership to each cluster, while a cluster can be characterized from the list of associated molecules with the highest membership degrees. In the proposed procedure, the fuzzy C-means algorithm [46] has been designed to produce the FC of the data sets.

The fundamental role of cluster analysis is to achieve the self-partitioning of the data and to look for compounds in the database which can be consid-

ered as close enough to be identified as a cluster. Each cluster can be identified as a set of compounds close to each other with regard to the molecular descriptors provided. A synthetic representation can be based on a clear separation of the clusters. Therefore, instead of trying to inspect all the compounds in the database to understand and analyze their chemical properties, it is only required to select typical compounds representing each cluster.

A detailed analysis of the FC method is beyond the aim of this article, but more explanation can be found in a previous paper [35]. The following parameters were used in the data processing of the data set of 389 compounds: space dimension = 2; tolerance = 0.001; number of iterations = 50; threshold = 0.69; number of clusters = 40.

The threshold value was selected in order to have no more than three activities for a given cluster and to derive a predictive model with a relevant selectivity. The number of clusters was chosen in order to maximize the prediction power of the model; sensibly worse prediction scores were got with a lower number of clusters, whereas values superior to 40 did not yield significant improving.

Calculations were performed using proprietary software.

2.5. Genetic algorithm

To select, amidst the 205 descriptors, the best parameters for separating and assessing the data set compounds, a method based on genetic algorithm (GA) [47, 48] procedure was used. This technique, in fact, allows us to decrease the number of descriptors thus increasing the computational speed.

GA, inspired by population genetics, consists of a population of individuals competing on 'only the fittest survive' basis. Each individual, or chromosome, represents a trial solution of the problem to solve. In the context of descriptor selection, the structure of the chromosome is very simple. Each descriptor is coded by a bit (0 or 1) and represents a component of the chromosome: 0 defines the absence of the descriptor, 1 defines its presence. The algorithm proceeds in successive steps called generations. During each generation, the population of chromosomes evolves by means of a 'fitness' function [49], which selects them by standard crossover and mutation operators [50]. The crossover phase takes two chromosomes and produces two new individuals, by swapping segments of genetic material, i.e. bits in this case. Within the population, mutation removes the bits affecting a small probability.

GAs are very effective for exploratory search, applicable to problems where little knowledge is available, but they are not particularly suitable for local search. In the latter case, they are combined to a stepwise approach in order to reach local convergence [44]. Stepwise approaches are quick and adapted to find a solution in 'promising' areas already identified.

To evaluate the fitness function, a specific index was derived by using the FC method. Furthermore, to prevent over-fitting and a poor generalization, a cross validation procedure was included in the algorithm during the selection procedure, randomly dividing the database into training and test sets. The fitness score of each chromosome is derived from the combination of the scores of the training and test sets.

The following parameters were used in the data processing of the data set of 389 compounds:

(i) Fuzzy parameters — weighting coefficient = 1.5, tolerance convergence = 0.01, number of iterations = 30, cluster number = 10.

(ii) Genetic parameters — chromosome number = 10, chromosome size = 205 (number of descriptors used), crossover point number = 1, percentage of rejections = 0.1, percentage of crossover = 0.8, percentage of mutation = 0.05, time off (10,100), number of generations = 10, ascendant coefficient = 0.02, descendant coefficient = −0.02.

Calculations were performed using proprietary software.

3. Results and discussion

The GA procedure was applied on a data set of 389 compounds, allowing to select the 17 best descriptors. This set of descriptors, reported in *table I*, shows a wide diversity, including molecular connectivity indices, E-states relative to the heteroatoms and hydrogen bonding, size descriptors and general descriptors such as the number of chemical elements included in a molecule.

The descriptors selected, with the help of GA, were used to derive the Kohonen map represented in *figure 1*. Four wide regions associated with adenosine, adrenergic and glutamate receptors and

three restricted zones pertaining to dopamine, acetylcholine and GABA receptors can be isolated. However, the map globally shows many regions in which two or more activity classes are mixed. The weak separation between the different classes evidences two main aspects reported in literature [51]: (i) the same chemical can be active on different receptors; (ii) for a given receptor, agonist and antagonist compounds can show opposite therapeutic properties.

As no univocal relation between structures and activity classes can be clearly defined by SOM, it is necessary to develop an objective automated tool that allows the transfer from an unsupervised classification to a supervised one. In order to get this new classification and to take into account the multiactivity character of several compounds, a FC model was derived from the Kohonen map.

Table I. Descriptor selection by genetic algorithm.

Descriptor symbol	Descriptor definition
xvch7	Valence chain index (molecular connectivity index)
dxvp3	Valence difference index (molecular connectivity index)
SsCH3	Sum of all (–CH3) E-state values in a molecule
SdssC	Sum of all (= C <) E-state values in a molecule.
SssNH	Sum of all (–NH–) E-state values in a molecule.
SssO	Sum of all (–O–) E-state values in a molecule.
SHCHnX	Hydrogen bond between SH groups and halogeno-alkyl groups.
Gmin	Minimum E-state value in molecule
SHBint2	Hydrogen bond referred to S atom
NHBint3	Hydrogen bond referred to N atom
Redundancy	Information index $R = 1 - [i/\log(\mathrm{nvx})]$; i = Information content based on atom classes; nvx = Number of graph vertices, hybrid groups or non-hydrogen atoms.
chi 3c	Cluster and path/cluster index
chi 5p	Path index
NRC	The number of N, O S and H atoms connected with one another.
MR	Molar refractivity
log P	Lipophylicity coefficient (distribution in system *n*-octanol/water)
MMES	Electro-negativity index

To validate the model, the data set of 389 compounds was randomly divided into a training set of 259 molecules and a test set of 130 molecules. Then, the hybrid system SOM/FC was applied on the training set molecules; its graphical representation is given in *figure 2*. The clusters, whose values have been determined considering the number of compounds in the data set, are visualized on the SOM. Each compound is associated with a vector of 40 components, representing the degrees of membership of cluster number one, number two and so on. Then, each molecule can be examined through its coordinates in the Kohonen map and according to its membership of the fuzzy sets defined by the clusters. To express things differently, each cluster reflects the relevance level of the Kohonen map to separate the different properties. Homogeneous clusters define regions with highly specific compounds, whereas an increasing number of classes in a cluster reflects a higher ambiguity in the attribution of a compound to a given receptor type.

In order to examine the robustness of the SOM/FC proposed model, the x/y coordinates for the test set compounds were calculated in the same conditions as for the training set compounds; then their degrees of membership for each cluster were derived. The results, reported in *table II*, were compared with the experimental activities attributed to each compound. For 81% of the test set compounds, the clusters associated with the highest membership degrees include the experimentally attributed classes.

Furthermore, for several compounds, it is possible to explain the reasons for which the predicted activities did not agree with the experimental data. For example, compound **66** (cimetidine) and **58** (DP6,7-ADTN) are defined in the RBI database as histaminergic and dopaminergic compounds, respectively. But, in literature, these compounds are also known [52, 53] to exhibit an adrenergic activity, in accordance with the predicted results.

The notion of belonging by FC is less precise than usual statistics but leads to interesting results, chiefly when considering the necessity to search for new leads in large databases of bioactive compounds. Deeper validation of such a procedure

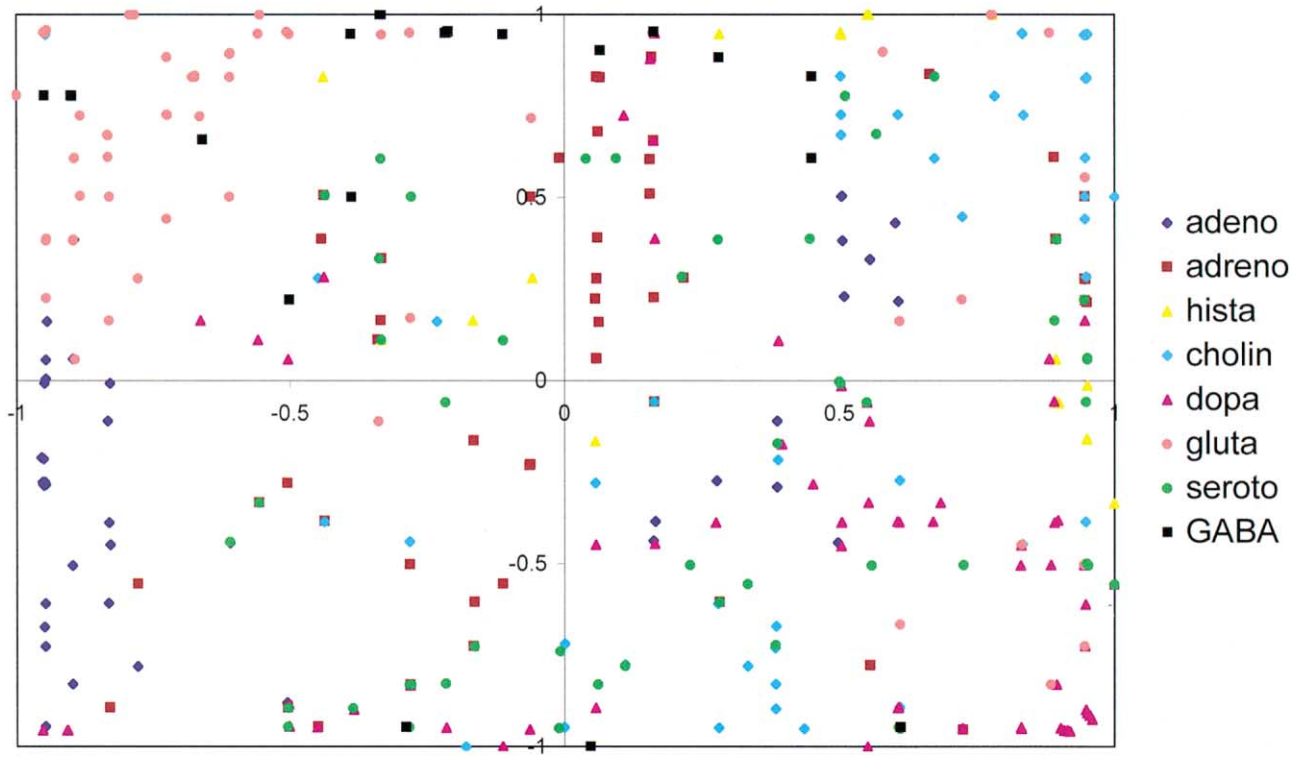

Figure 1. Kohonen SOM (10×10, interpolated map). Projection of a data set of 389 compounds extracted from the RBI database. The 17 descriptors selected by GA computations were taken into account for the molecular diversity analysis.

would require a specific validation index. Work is underway to fulfill this requirement. However, the objective here was only to show that coherent and robust results can be obtained through a hybrid system combining SOM and fuzzy logic (FL).

4. Conclusions

At present, in combinatorial chemistry, there is a requirement to maximize the efficiency in search for new leads and to reduce the cost associated to synthesizing and screening very large number of compounds. Then, more efficient tools allowing to design and to classify large chemical libraries with enhanced information contents become a must. Hybrid systems, derived from combining SOM and FL methods, constitute interesting approaches to virtual screening of chemical libraries.

In fact, these hybrid systems SOM/FL allow: (i) to examine the molecular diversity in a chemical database, giving a 2D representation of the compound distribution in the hyperspace of molecular descriptors; (ii) to build models, linking the structures of the compounds with the biological properties to be modeled and suitable to predict the activity values for new molecules external to the training data set.

More precisely, a hybrid system SOM/FC was applied here on a data set of 389 molecules active on CNS, representing eight classes of receptors. A large number of molecular descriptors was tested and an optimal reduced set was selected by a new GA procedure. The SOM 2D projection show either some specific regions associated with a unique receptor type, or many zones in which two or more activity classes are mixed. The ability of the hybrid system to model the eight activity classes was validated after dividing the 389 compounds into a training set and a test one,

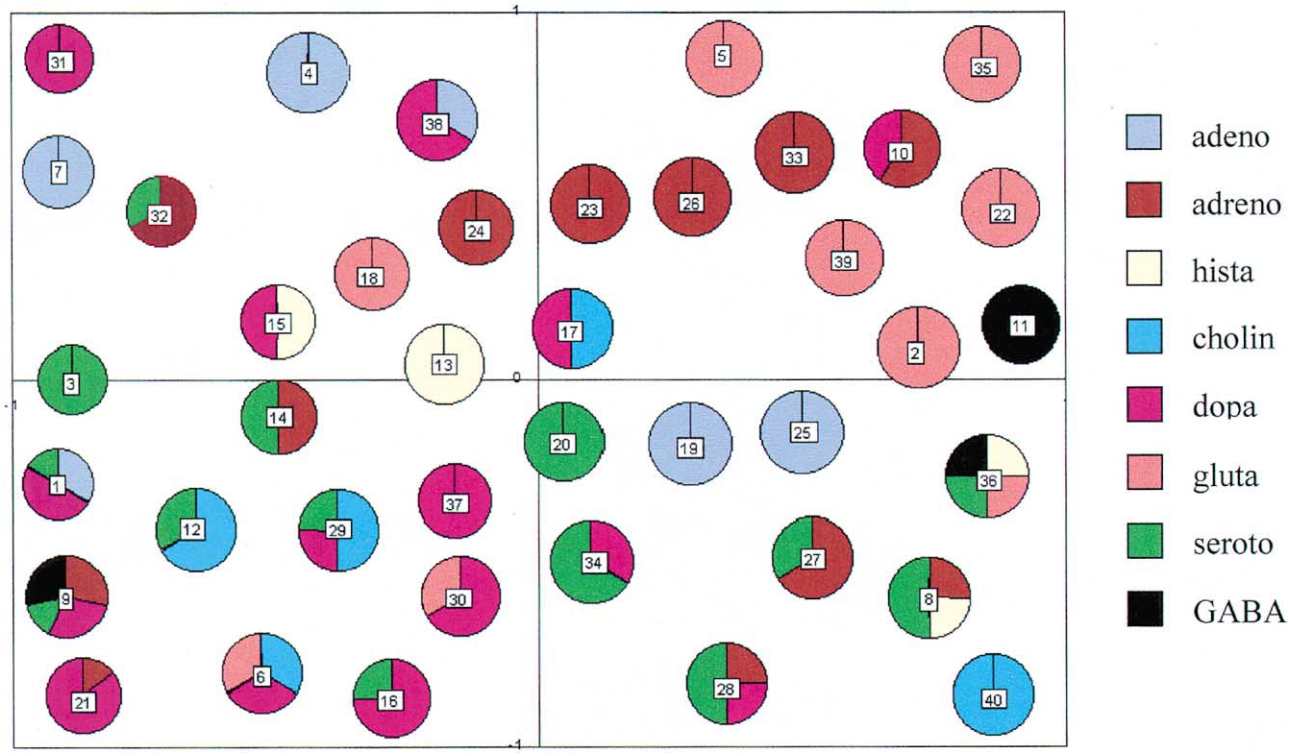

Figure 2. FC derived from the Kohonen SOM. The map is based on a training set including 259 compounds from the RBI database divided into 40 clusters with a 0.69 threshold.

including 259 and 130 molecules, respectively. For 81% of the test set compounds, the experimental activity class was predicted correctly. Furthermore, pseudo wrong predictions can be explained, in many cases, by a likely multiactivity behavior of the compounds which could prompt to revisit and/or to enlarge some previous biological results.

More generally, this paper promotes new trends in DBM and underlines two main points:

1. SOM, known as a projective technique, can be considered now as a supervised classification method, with help of FL.
2. The limitation related to binary classification of large CNS databases [34], on a drug-like/non-drug-like basis, is overcome with this opening example processing eight CNS activities.

These preliminary results show that the proposed methods SOM/FL are worth investigating more deeply and testing with larger chemical databases. Work is underway to apply the above DBM strategy, based on FL, to a larger database of CNS active compounds.

Acknowledgements

Our gratitude goes to Dr Keith Watling, Director of the RBI division, Sigma–Aldrich, Natick (MA), USA, and to Mrs Annie Blache, Engineer, Sigma–Aldrich, France, for RBI database facilities.

Table II. Three most important membership degrees related to 130 test set compounds; the associated clusters, shown in *figure 2*, are reported between brackets.

Compound	Experimental activity	Predicted membership degrees		
1	serotonergic	0.64 (Cl 34)	0.04 (Cl 30)	0.03 (Cl 19)
2	serotonergic	0.82 (Cl 8)	0.03 (Cl 36)	0.03 (Cl 27)
3	serotonergic	0.19 (Cl 12)	0.19 (Cl 1)	0.14 (Cl 3)
4	histaminergic	0.38 (Cl 27)	0.33 (Cl 36)	0.05 (Cl 36)
5	histaminergic	0.35 (Cl 8)	0.21 (Cl 36)	0.06 (Cl 27)
6	glutamatergic	0.57 (Cl 35)	0.09 (Cl 22)	0.07 (Cl 5)
7	glutamatergic	0.87 (Cl 5)	0.03 (Cl 35)	0.01 (Cl 33)
8	GABAergic	0.16 (Cl 27)	0.11 (Cl 36)	0.10 (Cl 25)
9	GABAergic	0.20 (Cl 36)	0.19 (Cl 11)	0.13 (Cl 2)
10	GABAergic	0.20 (Cl 36)	0.19 (Cl 11)	0.13 (Cl 2)
11	GABAergic	0.72 (Cl 11)	0.06 (Cl 2)	0.02 (Cl 36)
12	serotonergic	0.25 (Cl 12)	0.23 (Cl 1)	0.17 (Cl 9)
13	serotonergic	0.44 (Cl 28)	0.13 (Cl 27)	0.08 (Cl 8)
14	serotonergic	0.27 (Cl 28)	0.19 (Cl 27)	0.07 (Cl 34)
15	serotonergic	0.44 (Cl 28)	0.06 (Cl 16)	0.05 (Cl 27)
16	serotonergic	0.75 (Cl 27)	0.05 (Cl 8)	0.02 (Cl 28)
17	serotonergic	0.55 (Cl 36)	0.06 (Cl 2)	0.05 (Cl 8)
18	serotonergic	0.14 (Cl 12)	0.14 (Cl 3)	0.11 (Cl 1)
19	glutamatergic	0.70 (Cl 35)	0.13 (Cl 5)	0.02 (Cl 22)
20	glutamatergic	0.33 (Cl 35)	0.13 (Cl 5)	0.09 (Cl 10)
21	glutamatergic	0.14 (Cl 5)	0.08 (Cl 23)	0.08 (Cl 33)
22	glutamatergic	0.48 (Cl 5)	0.06 (Cl 35)	0.06 (Cl 33)
23	glutamatergic	0.40 (Cl 22)	0.14 (Cl 35)	0.08 (Cl 10)
24	glutamatergic	0.91 (Cl 35)	0.02 (Cl 5)	0.01 (Cl 22)
25	dopaminergic	0.81 (Cl 21)	0.04 (Cl 9)	0.02 (Cl 6)
26	dopaminergic	0.34 (Cl 38)	0.3 (Cl 24)	0.07 (Cl 18)
27	dopaminergic	0.29 (Cl 30)	0.25 (Cl 34)	0.09 (Cl 29)
28	dopaminergic	0.55 (Cl 6)	0.06 (Cl 16)	0.04 (Cl 9)
29	dopaminergic	0.93 (Cl 31)	0.02 (Cl 7)	0.01 (Cl 4)
30	dopaminergic	0.93 (Cl 31)	0.02 (Cl 7)	0.01 (Cl 4)
31	dopaminergic	0.46 (Cl 12)	0.08 (Cl 1)	0.05 (Cl 3)
32	dopaminergic	0.39 (Cl 9)	0.23 (Cl 21)	0.08 (Cl 12)
33	dopaminergic	0.73 (Cl 10)	0.05 (Cl 33)	0.04 (Cl 39)
34	dopaminergic	0.54 (Cl 21)	0.09 (Cl 9)	0.08 (Cl 6)
35	dopaminergic	0.26 (Cl 15)	0.12 (Cl 14)	0.09 (Cl 9)
36	dopaminergic	0.80 (Cl 16)	0.07 (Cl 6)	0.02 (Cl 30)
37	cholinergic	0.34 (Cl 19)	0.12 (Cl 25)	0.06 (Cl 27)
38	cholinergic	0.31 (Cl 40)	0.28 (Cl 8)	0.10 (Cl 36)
39	cholinergic	0.23 (Cl 40)	0.18 (Cl 28)	0.12 (Cl 8)
40	cholinergic	0.38 (Cl 17)	0.13 (Cl 13)	0.07 (Cl 20)
41	cholinergic	0.31 (Cl 36)	0.29 (Cl 8)	0.06 (Cl 27)
42	cholinergic	0.29 (Cl 28)	0.11 (Cl 8)	0.10 (Cl 27)
43	cholinergic	0.59 (Cl 28)	0.07 (Cl 27)	0.05 (Cl 8)
44	adrenergic	0.46 (Cl 28)	0.12 (Cl 27)	0.07 (Cl 8)
45	serotonergic	0.64 (Cl 34)	0.03 (Cl 19)	0.02 (Cl 20)
46	adrenergic	0.73 (Cl 10)	0.04 (Cl 33)	0.04 (Cl 39)
47	adrenergic	0.34 (Cl 33)	0.14 (Cl 26)	0.11 (Cl 11)
48	adrenergic	0.56 (Cl 36)	0.06 (Cl 2)	0.05 (Cl 5)
49	adrenergic	0.65 (Cl 28)	0.07 (Cl 27)	0.04 (Cl 8)
50	adrenergic	0.31 (Cl 10)	0.11 (Cl 22)	0.08 (Cl 5)
51	adrenergic	0.50(Cl 33)	0.24 (Cl 26)	0.05 (Cl 10)
52	adrenergic	0.28 (Cl 12)	0.16 (Cl 29)	0.04 (Cl 9)
53	adenosine/purinergic	0.76 (Cl 4)	0.03 (Cl 31)	0.03 (Cl 38)
54	adenosine/purinergic	0.85 (Cl 4)	0.03 (Cl 28)	0.01 (Cl 31)
55	adenosine/purinergic	0.99 (Cl 4)	0.00 (Cl 38)	0.00 (Cl 32)

Table II. (Continued)

Compound	Experimental activity	Predicted membership degrees		
56	adenosine/purinergic	0.21 (Cl 4)	0.19 (Cl 7)	0.14 (Cl 31)
57	glutamatergic	0.12 (Cl 5)	0.09 (Cl 33)	0.08 (Cl 23)
58	dopaminergic	0.91 (Cl 23)	0.02 (Cl 26)	0.01 (Cl 17)
59	adenosine/purinergic	0.54 (Cl 4)	0.09 (Cl 38)	0.02 (Cl 32)
60	adenosine/purinergic	0.77 (Cl 4)	0.03 (Cl 38)	0.02 (Cl 32)
61	adenosine/purinergic	0.99 (Cl 4)	0.00 (Cl 38)	0.00 (Cl 32)
62	adenosine/purinergic	0.21 (Cl 4)	0.18 (Cl 38)	0.03 (Cl 32)
63	serotonergic	0.71 (Cl 34)	0.04 (Cl 30)	0.04 (Cl 37)
64	serotonergic	0.44 (Cl 9)	0.17 (Cl 21)	0.09 (Cl 9)
65	histaminergic	0.43 (Cl 28)	0.06 (Cl 16)	0.05 (Cl 34)
66	histaminergic	0.18 (Cl 27)	0.14 (Cl 19)	0.08 (Cl 34)
67	glutamatergic	0.76 (Cl 35)	0.08 (Cl 5)	0.02 (Cl 22)
68	glutamatergic	0.76 (Cl 35)	0.06 (Cl 5)	0.03 (Cl 22)
69	glutamatergic	0.47 (Cl 25)	0.12 (Cl 19)	0.04 (Cl 39)
70	GABAergic	0.50 (Cl 13)	0.07 (Cl 18)	0.05 (Cl 15)
71	GABAergic	0.81 (Cl 21)	0.03 (Cl 9)	0.03 (Cl 6)
72	GABAergic	0.72 (Cl 11)	0.06 (Cl 2)	0.02 (Cl 22)
73	serotonergic	0.99 (Cl 29)	0.00 (Cl 12)	0.00 (Cl 14)
74	serotonergic	0.37 (Cl 17)	0.13 (Cl 13)	0.07 (Cl 20)
75	serotonergic	0.94 (Cl 28)	0.01 (Cl 27)	0.00(Cl 34)
76	serotonergic	0.98 (Cl 32)	0.00(Cl 32)	0.00(Cl 3)
77	serotonergic	0.63 (Cl 3)	0.14 (Cl 1)	0.02 (Cl 9)
78	serotonergic	0.29 (Cl 12)	0.16 (Cl 29)	0.05 (Cl 30)
79	serotonergic	0.83 (Cl 16)	0.03 (Cl 6)	0.02 (Cl 30)
80	glutamatergic	0.34 (Cl 2)	0.27 (Cl 11)	0.07 (Cl 39)
81	glutamatergic	0.75 (Cl 5)	0.06 (Cl 35)	0.03 (Cl 10)
82	glutamatergic	0.41 (Cl 16)	0.09 (Cl 30)	0.06 (Cl 6)
83	glutamatergic	0.35 (Cl 8)	0.21 (Cl 36)	0.06 (Cl 27)
84	glutamatergic	0.53 (Cl 6)	0.10 (Cl 21)	0.06 (Cl 16)
85	glutamatergic	0.15 (Cl 5)	0.15 (Cl 33)	0.10 (Cl 26)
86	glutamatergic	0.87 (Cl 36)	0.03 (Cl 8)	0.01 (Cl 27)
87	glutamatergic	0.39 (Cl 5)	0.09 (Cl 33)	0.07 (Cl 10)
88	glutamatergic	0.42 (Cl 22)	0.15 (Cl 10)	0.06 (Cl 39)
89	GABAergic	0.24 (Cl 27)	0.11 (Cl 36)	0.10 (Cl 8)
90	dopaminergic	0.15 (Cl 34)	0.12 (Cl 30)	0.12 (Cl 28)
91	dopaminergic	0.49 (Cl 1)	0.12 (Cl 12)	0.09 (Cl 3)
92	dopaminergic	0.37 (Cl 16)	0.12 (Cl 30)	0.06 (Cl 34)
93	dopaminergic	0.12 (Cl 38)	0.10 (Cl 23)	0.07 (Cl 4)
94	dopaminergic	0.98 (Cl 1)	0.00 (Cl 12)	0.00(Cl 9)
95	dopaminergic	0.42 (Cl 16)	0.09 (Cl 30)	0.06 (Cl6)
96	dopaminergic	0.80 (Cl 1)	0.04 (Cl 3)	0.04 (Cl 12)
97	dopaminergic	0.96 (Cl 28)	0.00 (Cl 27)	0.00 (Cl 34)
98	dopaminergic	0.66 (Cl 1)	0.15 (Cl 3)	0.03 (Cl 12)
99	dopaminergic	0.84 (Cl 16)	0.03 (Cl 6)	0.02 (Cl 30)
100	dopaminergic	0.39 (Cl 9)	0.23 (Cl 21)	0.08 (Cl 12)
101	cholinergic	0.3 (Cl 37)	0.21 (Cl 20)	0.07 (Cl 13)
102	cholinergic	0.94 (Cl 40)	0.01 (Cl 8)	0.01 (Cl 28)
103	cholinergic	0.77 (Cl 12)	0.05 (Cl 9)	0.03 (Cl 1)
104	cholinergic	0.99 (Cl 29)	0.00 (Cl 12)	0.00 (Cl 6)
105	cholinergic	0.91 (Cl 6)	0.01 (Cl 9)	0.01 (Cl 12)
106	cholinergic	0.39 (Cl 6)	0.12 (Cl 21)	0.06 (Cl 16)
107	cholinergic	0.97 (Cl 17)	0.01 (Cl 13)	0.00 (Cl 24)
108	cholinergic	0.89 (Cl 38)	0.02 (Cl 24)	0.01 (Cl 18)
109	adrenergic	0.95 (Cl 21)	0.01 (Cl 9)	0.01 (Cl 6)
110	histaminergic	0.63 (Cl 28)	0.06 (Cl 27)	0.04 (Cl 34)
111	histaminergic	0.41 (Cl 16)	0.09 (Cl 30)	0.06 (Cl 6)

Table II. (Continued)

Compound	Experimental activity	Predicted membership degrees		
112	adrenergic	0.65 (Cl 3)	0.05 (Cl 1)	0.04 (Cl 32)
113	adrenergic	0.73 (Cl 23)	0.12 (Cl 26)	0.02 (Cl 33)
114	adrenergic	0.15 (Cl 3)	0.13 (Cl 12)	0.11 (Cl 1)
115	adrenergic	0.74 (Cl 23)	0.11 (Cl 26)	0.02 (Cl 33)
116	adrenergic	0.96 (Cl 28)	0.00 (Cl 27)	0.00 (Cl 8)
117	adrenergic	0.20 (Cl 32)	0.15 (Cl 3)	0.10 (Cl 7)
118	adrenergic	0.22 (Cl 32)	0.16 (Cl 3)	0.07 (Cl 7)
119	adrenergic	0.61 (Cl 6)	0.07 (Cl 16)	0.04 (Cl 21)
120	adenosine/purinergic	0.30 (Cl 19)	0.11 (Cl 20)	0.07 (Cl 37)
121	adenosine/purinergic	0.66 (Cl 19)	0.05 (Cl 25)	0.03 (Cl 20)
122	adenosine/purinergic	0.62 (Cl 32)	0.06 (Cl 7)	0.03 (Cl 38)
123	adenosine/purinergic	0.15 (Cl 38)	0.12 (Cl 4)	0.1 (Cl 32)
124	dopaminergic	0.76 (Cl 21)	0.04 (Cl 9)	0.04 (Cl 6)
125	adenosine/purinergic	0.22 (Cl 14)	0.14 (Cl 15)	0.07 (Cl 3)
126	adenosine/purinergic	0.30 (Cl 19)	0.11 (Cl 20)	0.07 (Cl 37)
127	adenosine/purinergic	0.65 (Cl 19)	0.05 (Cl 25)	0.04 (Cl 20)
128	adenosine/purinergic	0.15 (Cl 38)	0.09 (Cl 4)	0.04 (Cl 18)
129	adenosine/purinergic	0.79 (Cl 25)	0.04 (Cl 2)	0.04 (Cl 19)
130	adenosine/purinergic	0.91 (Cl 38)	0.02 (Cl 24)	0.01 (Cl 23)

References

[1] Lowe G., Chem. Soc. Rev. 24 (1995) 309–317.

[2] Fauchere J.L., Boutin J.A., Henlin J.M., Kucharczyk N., Ortuno J.C., Chemom. Intell. Lab. Syst. 43 (1998) 43–68.

[3] Kenny B.A., Bushfield M., Parry-Smith D.J., Fogarty S., Treherne J.M., Prog. Drug Res. 51 (1998) 245–269.

[4] Fox S., Farr-Jones S., Yund M.A., J. Biomol. Screening 4 (1999) 183–186.

[5] Bayada D.M., Hamersma H., van Geerestein V.J., J. Chem. Inf. Comput. Sci. 39 (1999) 1–10.

[6] Warr W.A., J. Chem. Inf. Comput. Sci. 37 (1997) 134–140.

[7] Van Hijfte L., Marciniak G., Froloff N., J. Chromatogr. B 725 (1999) 3–15.

[8] Chastrette M., Cretin D., Aidi C.E., J. Chem. Inf. Sci. 36 (1995) 108–113.

[9] Bate A., Lindquist M., Edwards I.R., Olsson S., Orre R., Lansner A., De Freitas R.M., Eur. J. Clin. Pharmacol. 54 (1998) 315–321.

[10] Gini G., Lorenzini M., Benfenati E., Grasso P., Bruschi M., J. Chem. Inf. Comput. Sci. 39 (1999) 1076–1080.

[11] Kohonen T., Self-Organizing Maps, Springer, New York, 1995.

[12] Van Osdol W.W., Myers T.G., Paull K.D., Kohn K.W., Weinstein J., J. Natl. Cancer Inst. 86 (1994) 1853–1859.

[13] Bauknecht H., Zell A., Bayer H., Levi P., Wagener M., Sadowski J., Gasteiger J., J. Chem. Inf. Comput. Sci. 36 (1996) 1205–1213.

[14] Kireev D., Chrétien J.R., Bernard P., Ros F., SAR QSAR Environ. Res. 8 (1998) 93–107.

[15] Obu-Cann K., Tokukata H., Fujimura K., Yoshihara K., J. Surf. Anal. 5 (1999) 208–211.

[16] Varfis A., Versino C., Neural Netw. World 2 (1992) 813–834.

[17] Zadeh L.A., Classification and Clustering, Academic Press, New York, 1977, pp. 251–299.

[18] Sugeno M., Inf. Sci. 36 (1985) 59–83.

[19] Pedrycs W., Pattern Recogn. 23 (1990) 121–146.

[20] Pardridge W.M., J. Neurochem. 70 (1998) 1782–1792.

[21] Kuhar M.J., Life Sci. 62 (1998) 1573–1575.

[22] Paul S., Br. J. Psychiatry Suppl. 37 (1999) 23–25.

[23] Lloyd E.J., Andrews P.R., J. Med. Chem. 29 (1986) 453–462.

[24] Spector R., Pharmacology 40 (1990) 1–7.

[25] N'Goka V., Schlewer G., Linget J.M., Chambon J.P., Wermuth C.G., J. Med. Chem. 34 (1991) 2547–2557.

[26] Cai S.X., Kher S.M., Zhou Z.-L., Ilyin V., Espitia S.A., Tran M., Hawkinson J.E., Woodward R.M., Weber E., Keana J.F.W., J. Med. Chem. 40 (1997) 730–738.

[27] Conradi R.A., Burton P.S., Borchardt R.T., in: Pliska V., Testa B., van de Waterbeemd H. (Eds.), Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, 1996, pp. 233–252.

[28] Van de Waterbeemd H., Camenisch G., Folkers G., Chretien J.R., Raevsky O., J. Drug Target. 6 (1998) 151–165.

[29] Fischer H., Gottschlich R., Seelig A., J. Membr. Biol. 165 (1998) 201–211.

[30] Young R.C., Mitchell R.C., Brown T.H., Ganellin C.R., Griffiths C.R., Jones M., Rana K.K., Saunders D., J. Med. Chem. 31 Chem. 31 (1988) 656–671.

[31] Calder J.D., Ganellin C.R., Drug Des. Discovery 13 (1995) 123–131.

[32] Lombardo F., Blake J.F., Curatolo W.J., J. Med. Chem. 39 (1996) 4750–4755.

[33] Norinder U., Sjöberg P., Österberg T., J. Pharm. Sci. 87 (1998) 952–959.

[34] Ajay, Bemis G.W., Murcko M.A., J. Med. Chem. 42 (1999) 4942–4951.

[35] Ros F., Audouze K., Pintore M., Chrétien J.R., SAR QSAR Environ. Res. 11 (2000) 281–300.

[36] HTS Library of Pharmacologically Active Compounds, Catalogue No. SC001, Plate Maps and Structures, RBI©, Sigma Company, 1997.

[37] K.J. Watling, The RBI Handbook of Receptor Classification and Signal Transduction, RBI, Sigma-Aldrich Research Biochemicals Inc., 1998.

[38] Sabljic A., in: Karcher W., Devillers J. (Eds.), Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology1, Kluwer Academic Publishers, Dordrecht, 1990, pp. 61–82.

[39] Dearden J.C., in: Karcher W., Devillers J. (Eds.), Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology1, Kluwer Academic Publishers, Dordrecht, 1990, pp. 25–59.

[40] Kier L.B., Hall L.K., Molecular Structure Description — The Electrotopological State, Academic Press, New York, 1999.

[41] Hansch C., Leo A.W., Subtituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979, pp. 18–43.

[42] Sanderson R., Chemical Bonds and Bond Energy, Academic Press, New York, 1976, p. 15.

[43] Neural computing: A technology handbook for neuralworks professional II/Plus and Neuralworks explorer, NeuralWare Inc., Technical publications group, 1995.

[44] Ros F., Guillaume S., Rabatel G., Sevila F., J. Food Control 6 (1995) 37–43.

[45] Ros F., Guillaume S., Rabatel G., Sevila F., Bertrand D., J. Chemom. 11 (1997) 486–500.

[46] Cubes R.C., Jain A.K., in: Yovits M. (Ed.), Advances in Computers, Academic Press, New York, 1980, pp. 113–228.

[47] Haupt R.L., Haupt S.E., Practical Genetic Algorithms, Wiley, New York, 1998.

[48] F. Ros, M. Pintore, J.R. Chrétien, Chemometrics and Intelligent Laboratory Systems, 2001, in preparation.

[49] David L., Handbook of Genetic Algorithms, Van Nostrand Reinhold, New York, 1991.

[50] Kinnear K.E., Advances in Genetic Programming, MIT Press, Cambridge, 1994.

[51] RBI® Signaling Innovation, 1999 Catalogue, RBI/Sigma, Sigma-Aldrich® Co., 1999.

[52] Kondo S., Tashima Y., Morita T., Tohoku J., Exp. Med. 167 (1992) 163–166.

[53] Timmermans P.B., Mathy M.J., Thoolen M.J., de Jonge A., Wilffert B., van Zwieten P.A., J. Pharm. Pharmacol. 36 (1984) 772–775.