Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 410–422
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Bayesian-regularized genetic neural networks applied to the modeling of non-peptide antagonists for the human luteinizing hormone-releasing hormone receptor

Michael Fernández, Julio Caballero *

Molecular Modeling Group, Center for Biotechnological Studies, University of Matanzas, Matanzas, C.P. 44740, Cuba

Received 19 September 2005; received in revised form 19 February 2006; accepted 20 February 2006
Available online 28 February 2006

Abstract

Bayesian-regularized genetic neural networks (BRGNNs) were used to model the binding affinity (IC_{50}) for 128 non-peptide antagonists for the human luteinizing hormone-releasing hormone (LHRH) receptor using 2D spatial autocorrelation vectors. As a preliminary step, a linear dependence was established by multiple linear regression (MLR) approach, selecting the relevant descriptors by genetic algorithm (GA) feature selection. The linear model showed to fit the training set ($N = 102$) with $R^2 = 0.746$, meanwhile BRGNN exhibited a higher value of $R^2 = 0.871$. Beyond the improvement of training set fitting, the BRGNN model overcame the linear one by being able to describe 85% of test set ($N = 26$) variance in comparison with 73% the MLR model. Our non-linear QSAR model illustrates the importance of an adequate distribution of atomic properties represented in topological frames and reveals the electronegativities, masses and polarizabilities as the most influencing atomic properties in the structures of the heterocycles under analysis for having an appropriate LHRH antagonistic activity. Furthermore, the ability of the non-linear selected variables for differentiating the data was evidenced when total data set was well distributed in a Kohonen self-organizing map (SOM).
© 2006 Elsevier Inc. All rights reserved.

Keywords: QSAR analysis; Bayesian-regularized artificial neural network; Genetic algorithm; Self organizing maps; Non-peptide LHRH antagonists

1. Introduction

Luteinizing hormone-releasing hormone (LHRH), also known as gonadotropin-releasing hormone (GnRH), is a decapeptide (pGlu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-GlyNH_2), which was discovered by Schally et al. in 1971 [1]. It acts on the pituitary gland to stimulate the secretion of both luteinizing hormone and follicle-stimulating hormone. These gonadotropins, in turn, act on the reproductive organs, where they participate in the regulation of gonadal steroid production, spermatogenesis in male and follicular development in female.

Antagonists of LHRH, bind to their receptors in the pituitary gonadotrophs causing inhibition of gonadotropin release, which subsequently causes the suppression of sex steroids in mammals. This property of suppressing sex hormones renders the LHRH antagonists potentially useful in the treatment of endocrine-based diseases such as prostate cancer, breast cancer, endometriosis, uterine leiomyoma and precocious puberty [2].

Intense research has been focused on the development of potent and safe antagonists [3–5]. The relatively low potency and the adverse effects, due to histamine release, have been the main obstacles to their acceptance and clinical use. In this sense, peptide antagonists with low histamine-release properties were reported [6]. However, peptide antagonists still have their limitations such as poor oral bioavailability. Concluding the last decade, it was reported the first potent and orally active non-peptide LHRH receptor antagonist T-98475, a thieno[2,3-*b*]pyridine-4-one derivative [7]. As from this event, several classes of LHRH antagonists have been described in the literature [7–11]. Structure-activity relationship (SAR) studies lead to an increase of non-peptide LHRH antagonists reported in the last years in an endeavor to find superior drugs for clinical applications.

Computer simulation techniques potentially offer a further means to probe SAR results. Quantitative structure-activity relationship (QSAR) is one of the most embraced computa-

* Corresponding author. Tel.: +53 45 26 1251; fax: +53 45 25 3101.
E-mail addresses: jmcr77@yahoo.com, julio@fq.uh.cu (J. Caballero).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.02.005

tional approaches in drug design [12], which attempts to investigate and explain the relations between the molecular structure and the biological activity. A QSAR model is a tool for the comprehension of these two spaces in order to predict activities of as-yet untested molecules.

Despite the considerable interest toward non-peptide LHRH antagonists, no QSAR data have been reported to date. To gain insights into the structural and molecular requirements influencing this activity, we describe herein the QSAR analysis of a dataset of 128 heterocyclic ring derivatives including thieno[2,3-*b*]pyridine-4-ones, thieno[2,3-*d*]pyrimidine-2,4-diones, imidazo[1,2-*a*]pyrimidin-5-ones and benzimidazole derivatives recently reported with both multiple linear regression (MLR) and Bayesian-regularized genetic neural network (BRGNN) approaches. 2D autocorrelation descriptors were used for encoding the structural information of the studied compounds. Optimum variable subsets of descriptors were selected using linear and non-linear genetic algorithm (GA) searches. Both MLR and BRGNN techniques were used for modeling the observed activity of the training set (102 compounds). The adequacies of the models were examined by means of their statistic significances, the statistics of leave-one-out (LOO) cross-validation experiments and the prediction of a test set (26 compounds), which represents the 20% of total data set. In addition, the capacity of the selected variables for differencing the data was evaluated by means of the unsupervised training of Kohonen self-organizing maps (SOMs).

2. Materials and methods

2.1. Spatial autocorrelation approach

The binding of a substrate to its receptor is dependent on the shape of the substrate and on a variety of effects such as the molecular electrostatic potential, polarizability, hydrophobicity and lipophobicity. Therefore, in a QSAR study the strategy for encoding molecular information must in some way, either explicitly or implicitly, account for these physicochemical effects. Furthermore, usually data sets include molecules of different size with different numbers of atoms, so the structural encoding structures must allow comparing such molecules [13].

Autocorrelation vectors have several useful properties. First, a substantial reduction in data can be achieved by limiting the topological distance, *l*. Second, the autocorrelation coefficients are independent of the original atom numberings, so they are canonical. And third, the length of the correlation vector is independent of the size of the molecule [13].

For the autocorrelation vector, H-depleted molecular structure is represented as a graph G and physicochemical properties of atoms as real values assigned to the vertices of G. These descriptors can be obtained by summing up the products of certain properties of two atoms, located at given topological distances or spatial lag in G. Three spatial autocorrelation vectors were employed for modeling the antagonistic activity.

Broto–Moreau's autocorrelation coefficient [14]:

$$\mathrm{ATS}(l, p_k) = \sum_i \delta_{ij} p_{ki} p_{kj} \tag{1}$$

Moran's index [15]:

$$\mathrm{MATS}(l, p_k) = \frac{N}{2L} \frac{\sum_{ij} \delta_{ij}(p_{ki} - \bar{p}_k)(p_{kj} - \bar{p}_k)}{\sum_i (p - \bar{p}_k)} \tag{2}$$

and Geary's coefficient [16]:

$$\mathrm{GATS}(l, p_k) = \frac{(N-1)}{4L} \frac{\sum_{ij} \delta_{ij}(p_{ki} - \bar{p}_k)(p_{kj} - \bar{p}_k)}{\sum_i (p_{ki} - \bar{p}_k)} \tag{3}$$

where $\mathrm{ATS}(l, p_k)$, $\mathrm{MATS}(l, p_k)$ and $\mathrm{GATS}(l, p_k)$ are Broto–Moreau's autocorrelation coefficient, Moran's index and Geary's coefficient at spatial lag l, respectively; p_{ki} and p_{kj} are the values of property k of atom i and j, respectively; $\bar{p}_k$ is the average value of property k and $\delta(l, d_{ij})$ is a Dirac-delta function defined as:

$$\delta(l, d_{ij}) = \left\{ \begin{array}{ll} 1 & \text{if } d_{ij} = l \\ 0 & \text{if } d_{ij} \neq l \end{array} \right\} \tag{4}$$

where d_{ij} is the topological distance or spatial lag between atoms i and j.

Spatial autocorrelation measures the level of interdependence between properties, and the nature and strength of that interdependence. It may be classified as either positive or negative. In a positive case all similar values appear together, while a negative spatial autocorrelation has dissimilar values appearing in close association [15,16]. In a molecule, Moran's and Geary's spatial autocorrelation analysis tests whether the value of an atomic property at one atom in the molecular structure is independent of the values of the property at neighboring atoms. If dependence exists, the property is said to exhibit spatial autocorrelation. Moreau and Broto first applied autocorrelation function to the topology of molecular structures [14]. The autocorrelation vectors represent the degree of similarity between molecules.

A data matrix is generated with the spatial autocorrelation vectors calculated for each compound. Afterwards, dimensionality reduction methods were employed for selecting the most relevant vector components for building MLR and BRGNN models.

Dragon [17] computer software was used for calculating unweighted and weighted Broto–Moreau, Moran and Geary 2D-autocorrelation vectors. As weighting properties we tried atomic masses, atomic van der Waals volumes, atomic Sanderson electronegativities and atomic polarizabilities. Autocorrelation vectors were calculated at spatial lags l ranging from 1 up to 8.

The total number of computed descriptors was 96. Descriptors with constant values were discarded. For the remaining descriptors pairwise correlation analysis was performed in order to reduce, in a first step, the colinearity and correlation between descriptors. The procedure consists on the elimination of the descriptor with lower variance from each pair of descriptors with the modulus of the pair correlation coefficients higher than a predefined value $R_{\max}$ (0.90). Afterwards, the number of remained descriptors was 53.

Table 1
Experimental and predicted activities of thieno[2,3-*b*]pyridine-4-ones derivatives from the training set

Comp[a]	R_1	R_2	R_3	R_4	$Log(10^6/IC_{50})$		
					Exp.	MLR	BRGNN
1 (T98475)	iPrCONH	BnN(Me)	iPr	2,6-F	6.699	6.144	6.384
2	MeO	MeNH	Et	2-OMe	2.699	3.241	3.285
3	MeO	BnN(Me)	Et	2-OMe	4.155	4.029	3.867
4	MeO	BnN(Me)	Et	2,6-F	5.222	4.676	4.503

[a] Compounds **1**–**4** are from Ref. [7].

2.2. Dataset: source and prior preparation

A data set of 128 non-peptide LHRH antagonists was collected from reports of two groups [7,11,18–21]. All antagonistic activities (IC_{50}) were tested as described in Ref. [7] where IC_{50} (nM) is the antagonist concentration required to inhibit the specific binding of [^{125}I]leuprorelin to human LHRH receptor from Chinese hamster ovary (CHO) cells by 50%. Since the same testing method was applied, the uncertainty associated with the pharmacological assays carried out in different laboratories must not significantly influence our QSAR models.

A subset of 26 compounds out of the total 128 compounds was randomly chosen as a prediction set and was used for external validation of all the models that were developed in this study. The external prediction set was chosen in such a way as to comprise the general characteristics of the data set. The compounds in the external prediction set were never used during the model development process but were reserved to validate potential models. For the development of MLR and BRGNN models, the training sets included all the remaining 102 compounds.

The chemical structures of the compounds used in this study are represented in Fig. 1. Experimental $log(10^6/IC_{50})$ values of compounds used for building the models are informed in Tables 1–5, meanwhile experimental $log(10^6/IC_{50})$ values of compounds used in the external validation process are shown in Table 6. The 3D structures of the studied compounds were geometrically optimized using the semi-empirical quantum-chemical method PM3 [22] implemented in the MOPAC 6.0 computer software [23].

2.3. Genetic Algorithm variable selection

Since 53 molecular descriptors were available for QSAR analysis and only a subset of them is statistically significant in terms of correlation with biological activities, deriving an optimal QSAR model through variable selection needs to be addressed. Following the Occam's Razor [24], we selected just the variables that contain the information that is necessary for the modeling but nothing more. In this sense, linear and non-linear GA searches have been carried out in order to build the linear and non-linear models.

GAs are governed by biological evolution rules [25]. They are stochastic optimization methods that have been inspired by evolutionary principles. The distinctive aspect of a GA is that it investigates many possible solutions simultaneously, each of which explores different regions in parameter space [26]. The first step is to create a population of N individuals. Each individual encodes the same number of randomly chosen descriptors. The fitness of each individual in this generation is determined. In the second step, a fraction of children of the next

Table 2
Experimental and predicted activities of thieno[2,3-*d*]pyrimidine-2,4-diones derivatives from the training set

Comp[a]	R_1	R_2	$Log(10^6/IC_{50})$		
			Exp.	MLR	BRGNN
6	EtCON(Me)	3-MeOPh	4.699	5.048	4.473
7	MeO	cHex	3.155	4.554	4.161
8	MeO	Bn	3.398	4.546	4.046
9	MeO	Ph	4.301	4.721	4.385
10	MeO	2-MeOPh	3.699	4.819	3.913
11	MeO	3-MeOPh	4.699	4.066	4.220
12	MeO	2-ClPh	4.097	4.664	4.375
13	MeO	4-ClPh	3.699	4.546	4.635
14	H_2N	3-MeOPh	6.000	4.657	5.115
15	EtCONH	Ph	6.523	5.832	6.265
16	iPrCONH	Ph	6.398	5.965	6.323
17	PhCONH	Ph	6.000	6.153	6.687
18	iPrCONH	3-MeOPh	6.523	5.301	6.713
19	MeNHCONH	Ph	7.000	6.292	6.597
20	EtNHCONH	Ph	7.000	6.360	6.606
21	iPrNHCONH	Ph	6.699	6.431	6.500
22	HONHCONH	Ph	6.398	6.630	6.229
23	EtONHCONH	Ph	6.699	5.956	6.192
24	iPrONHCONH	Ph	6.222	6.098	6.190

[a] Compounds **6**–**24** are from Ref. [19].

Table 3
Experimental and predicted activities of imidazo[1,2-*a*]pyrimidin-5-ones derivatives from the training set

Comp[a]	R_1	R_2	$Log(10^6/IC_{50})$		
			Exp.	MLR	BRGNN
31	iPr	EtOCO	4.523	5.492	4.863
32	Et	EtOCO	5.000	5.305	5.922
33	cPr	EtOCO	5.301	5.466	5.733
34	Vinyl	EtOCO	5.523	5.205	5.549
35	Ph	EtOCO	4.699	5.670	5.251
36	3-furyl	EtOCO	5.523	5.556	4.942
37	EtNH	EtOCO	6.301	5.781	5.593
38	EtNH	iPrOCO	6.523	6.001	5.923
39	MeONH	iPrOCO	6.398	5.484	6.041

[a] Compounds **31**–**39** are from Ref. [18].

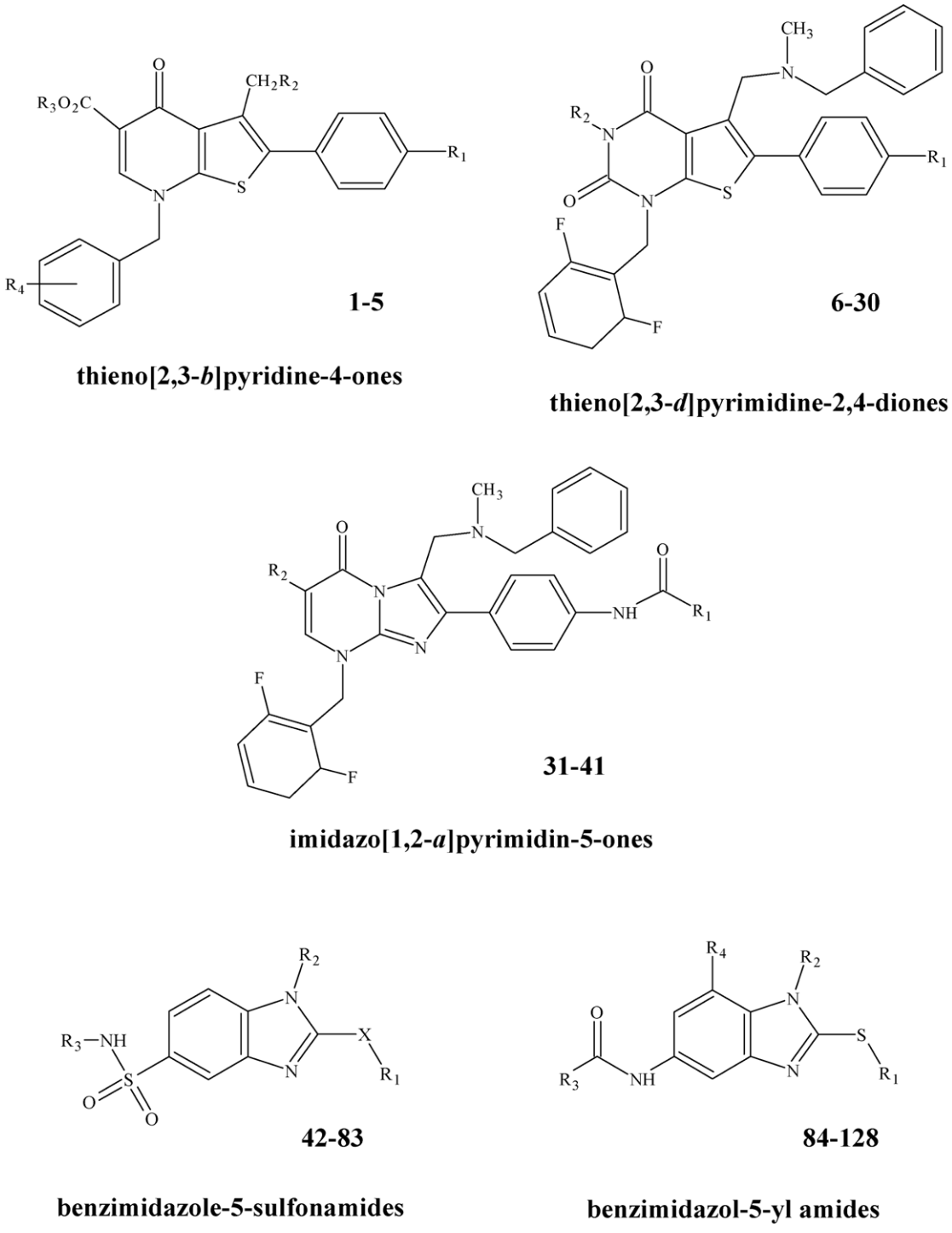

Fig. 1. General structure of non-peptide antagonists for the human LHRH receptor.

generation is produced by crossover (crossover children) and the rest by mutation (mutation children) from the parents on the basis of their scaled fitness scores. The new offspring contains characteristics from two or one of its parents.

Linear GA search was carried out exploring MLR models. In turn, neural network feature selection procedure, which extracts non-linear information from the data set was employed for data dimensionality reduction before network training.

The GA implemented in this paper is a version of the So and Karplus report [27] that was previously reported by our group [28] and was programmed within the MATLAB environment using the genetic algorithm and neural networks toolboxes [29]. We explored MLR models for the linear GA search (MLR modeling) and Bayesian-regularized artificial neural networks (BRANNs) with a simple architecture (two neurons in a sole hidden layer) for the non-linear GA search (BRGNN modeling). On both cases, the MSE of data fitting was tried

Table 4
Experimental and predicted activities of benzimidazole-5-sulfonamides from the training set

Comp[a]	X	R_1	R_2	R_3	$Log(10^6/IC_{50})$		
					Exp.	MLR	BRGNN
42	S	Et	Bn	4-MeOBn	2.959	2.724	2.953
43	S	Et	Bn	4-NO_2Bn	2.959	3.057	2.783
44	S	Et	Bn	4-ClBn	3.553	3.088	3.501
45	S	Et	Bn	4-FBn	3.602	3.327	3.056
46	O	Et	Bn	4-FBn	3.523	2.933	3.204
47	S	Me	Bn	4-FBn	3.658	3.326	2.923
48	O	Et	3-ClBn	4-FBn	3.056	2.695	3.078
49	O	Et	3-thienyl	4-FBn	2.854	2.830	2.627
50	S	2-(4-morpholinyl)Et	Bn	4-FBn	3.796	3.748	3.850
51	O	Et	2-(4-morpholinyl)Et	4-FBn	2.678	2.253	2.816
52	S	Et	Bn	2-(4-morpholinyl)Et	2.000	2.003	2.330
53	O	2-(4-morpholinyl)Et	Bn	4-FBn	2.237	2.320	2.092
54	NH	2-(4-morpholinyl)Et	Bn	4-FBn	2.000	2.329	1.866
55	S	2-(BnNH)Et	Bn	4-FBn	3.174	4.289	3.410
56	S	(2-pyridinyl)Me	Bn	4-FBn	3.086	3.605	3.390
57	S	(4-pyridinyl)Me	Bn	4-FBn	3.495	3.493	3.317
58	S	2-(2-pyridinyl)Et	Bn	4-FBn	3.638	4.249	3.473
59	S	2-(3-pyridinyl)Et	Bn	4-FBn	3.921	4.054	3.390
60	CH_2	Et	Bn	4-FBn	2.000	2.258	2.710
61	CHMe	Me	Bn	4-FBn	3.409	2.544	3.074
62	CH_2	H_2NCH_2	Bn	4-FBn	2.000	2.681	2.974
63	CH_2	$tBuOCONHCH_2$	Bn	4-FBn	3.131	3.928	4.419
64	CH_2	$tBuCH_2CONHCH_2$	Bn	4-FBn	2.585	3.471	3.387
65	CH_2	$tBuNHCONHCH_2$	Bn	4-FBn	4.699	4.032	3.999
66	CH_2	2-(tBuNHCO)Et	Bn	4-FBn	3.398	3.710	3.719
67	CH_2	$iPrNHCONHCH_2$	Bn	4-FBn	3.824	3.797	3.768
68	CH_2	$tBuCH_2NHCONHCH_2$	Bn	4-FBn	3.959	3.462	3.227
69	CH_2	$2\text{-}FPhNHCONHCH_2$	Bn	4-FBn	4.180	4.002	4.360
70	CH_2	$tBuNHCONHCH_2$	Bn	Bn	3.310	3.381	3.262
71	CH_2	$tBuNHCONHCH_2$	Bn	2-FBn	3.125	3.210	3.775
72	CH_2	$tBuNHCONHCH_2$	Bn	3,4-diFBn	3.770	3.881	3.838
73	CH_2	$tBuNHCONHCH_2$	Bn	3-F,4-CF_3Bn	4.244	4.920	4.579
74	CH_2	$tBuNHCONHCH_2$	Bn	4-ClBn	5.377	3.880	4.413
75	CH_2	$tBuNHCONHCH_2$	Bn	2,4-diClBn	4.658	3.896	4.648

[a] Compounds **42–59** are from Ref. [11] and compounds **60–75** are from Ref. [20].

as the individual fitness function. An individual in the population is represented by a string of integers, which means the numbering of the columns in the data matrix. In the So and Karplus report [27], the fitness of an individual was determined by a variety of fitness functions which are proportional to the residual error of the training set, the test set, or even the cross-validation set from the neural network simulations. In our approach, we tried the MSE of data fitting for MLR or BRANN predictors, as the case may be, as the individual fitness function. The first step is to create a gene pool (population) of N individuals. Each individual encodes the same number of descriptors; the descriptors are randomly chosen from a common data matrix, and in a way such that (1) no two individuals can have exactly the same set of descriptors and (2) all descriptors in a given individual must be different. The fitness of each individual in this generation is determined by the MSE of the model and scaled using a scaling function. A top scaling fitness function scaled a top fraction of the individuals in a population equally; these individuals have the same probability to be reproduced while the rest are assigned the value 0.

In the next step, a fraction of children of the next generation is produced by crossover (crossover children) and the rest by mutation (mutation children) from the parents. Sexual and asexual reproductions take place so that the new offspring contains characteristics from two or one of its parents. In a sexual reproduction two individuals are selected probabilistically on the basis of their scaled fitness scores and serve as parents. Next, in a crossover, each parent contributes a random selection of half of its descriptor set and a child is constructed by combining these two halves of ''genetic code''. Finally, the rest of the individuals in the new generation are obtained by asexual reproduction when parents selected randomly are subjected to a random mutation in one of their genes; i.e. one descriptor is replaced by another. We also included elitism, which protects the fittest individual in any given generation from crossover or mutation during reproduction. The genetic content of this individual simply moves on to the next generation intact. This selection, crossover, and mutation processes are repeated until all of the N parents in the population are replaced by their children. The fitness score of each member of this new generation is again evaluated and the

Table 5
Experimental and predicted activities of benzimidazol-5-yl amides from the training set

Comp[a]	R_1	R_2	R_3	R_4	$Log(10^6/IC_{50})$		
					Exp.	MLR	BRGNN
84	2-(Me)Bn	iPr	cPr	H	2.495	3.764	2.599
85	2-ClBn	iPr	cPr	H	2.201	2.565	2.615
86	2-(MeO),5-(MeCO)Bn	iPr	cPr	H	3.585	3.605	3.110
87	2-(MeO),5-(NC)Bn	iPr	cPr	H	3.244	3.413	3.377
88	2-(MeO),5-ClBn	iPr	cPr	H	3.796	2.636	3.488
89	2-(MeO),5-(O_2N)Bn	iPr	cPr	H	3.553	2.909	3.140
90	2-(MeO),5-(O_2N)Bn	cPr	tBuNH	H	4.102	3.621	4.296
91	2-(MeO),5-(O_2N)Bn	Et	tBuNH	H	4.377	3.415	3.851
92	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	H	4.745	4.231	4.351
93	2-(MeO),5-(O_2N)Bn	nBu	tBuNH	H	4.638	4.029	4.092
94	2-(MeO),5-(O_2N)Bn	nPent	tBuNH	H	3.678	3.717	3.531
95	2-(MeO),5-(O_2N)Bn	2-(MeO)Et	tBuNH	H	4.244	3.425	4.379
96	2-(MeO),5-(O_2N)Bn	Bn	tBuNH	H	4.553	4.435	4.116
97	2-(MeO),5-(O_2N)Bn	2-PhEt	tBuNH	H	3.444	3.954	3.692
98	2-(MeO),5-(O_2N)Bn	(2-pyridinyl)Me	tBuNH	H	4.367	4.433	4.482
99	2-(MeO),5-(O_2N)Bn	(3-pyridinyl)Me	tBuNH	H	4.481	4.370	4.364
100	2-(MeO),5-(O_2N)Bn	Ph	tBuNH	H	3.456	4.739	3.896
101	2-(MeO),5-(O_2N)Bn	iPr	tBuO	H	3.056	3.977	3.942
102	2-(MeO),5-(O_2N)Bn	iPr	cBu	H	2.824	3.004	3.047
103	2-(MeO),5-(O_2N)Bn	iPr	tBu	H	3.745	2.899	3.442
104	2-(MeO),5-(O_2N)Bn	iPr	$tBuCH_2$	H	3.745	2.974	3.524
105	2-(MeO),5-(O_2N)Bn	iPr	tBuNH	H	4.208	3.850	4.114
106	2-(MeO),5-(O_2N)Bn	iPr	iPrNH	H	3.222	3.744	3.240
107	2-(MeO),5-(O_2N)Bn	iPr	MeNH	H	2.481	3.636	2.853
108	2-(MeO),5-(O_2N)Bn	iPr	cPrNH	H	2.721	3.810	2.698
109	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	(1-pyrrolidinyl)CO	3.215	3.764	3.426
110	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	$(Me)_2NCO$	3.569	3.657	4.013
111	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	MeNHCO	4.387	4.014	4.387
112	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	tBuNHCO	3.387	4.011	3.751
113	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	BnNHCO	3.921	4.143	3.923
114	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	(4-pyridinyl)MeNHCO	4.215	4.318	4.467
115	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	(3-pyridinyl)MeNHCO	4.161	4.225	4.070
116	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	(2-pyridinyl) MeNHCO	4.509	4.382	4.452
117	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	2-(3-pyridinyl)EtNHCO	3.796	4.266	4.342
118	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	3-($(Me)_2N$)nPrNHCO	3.770	3.253	3.494
119	2-(MeO),5-(O_2N)Bn	nPr	tBuNH	3-(Bn,MeN)nPrNHCO	3.721	3.969	3.306

[a] Compounds **84–119** are from Ref. [21].

reproductive cycle is continued until 90% of the generations showed the same target fitness score.

The quality of each model was proven by the square multiple correlation coefficient (R^2) and the standard deviation (S). The models with R-value above 0.8 were selected and they were tested in cross-validation experiments.

2.4. Bayesian-regularized artificial neural networks

An artificial neural network (ANN) is a layered structure consisting of computing units named neurons and connections between neurons named synapses [30]. Neurons of the input layer are associated with independent variables or input variables; while the output neurons are associated with response variables or output variables. Synapses are oriented connections linking neurons from the input layer to neurons of the hidden layer and neurons from the hidden layer to output neurons. The strength of the synapse from neuron i to neuron j is determined by means of a real value w_{ij}, named weight. Furthermore, each neuron j from the hidden layer, and eventually the output neuron, are associated with a real value b_j named the neuron's bias and with a non-linear function, named the transfer or activation function.

ANNs are data-driven models in the sense that their adjustable parameters are selected in such a way as to minimize some network performance function F. In particular the task concerned with the selection of the optimal values for the adjustable parameters, exploiting the available data set, is called the learning problem. Typically, training aims to reduce the sum of squared errors:

$$F = \mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2 \tag{5}$$

In this equation MSE is the mean of the sum of squares of the network errors, N is the number of compounds, y_i is the predicted biological activity of the compound i and t_i is the experimental biological activity of the compound i.

ANNs have received attention due to their theoretical properties and computational efficiency. The main property of

Table 6
Experimental and predicted activities of non-peptide antagonists for the LHRH receptor from the test set

Comp[a]	X	R_1	R_2	R_3	R_4	$Log(10^6/IC_{50})$		
						Exp.	MLR	BRGNN
5	–	MeO	$PhCH_2N(Me)$	Et	2,6-F	5.398	5.303	5.262
25	–	O_2N	3-MeOPh	–	–	4.699	4.290	4.427
26	–	MeO	iBu	–	–	3.398	4.277	3.866
27	–	MeO	4-MeOPh	–	–	4.523	4.296	4.455
28	–	MeO	3-ClPh	–	–	3.699	4.712	4.312
29	–	EtCONH	3-MeOPh	–	–	6.699	5.159	6.230
30	–	MeONHCONH	H	–	–	7.000	5.921	6.421
40	–	3-thienyl	EtOCO	–	–	5.155	5.051	6.064
41	–	MeONH	EtOCO	–	–	6.155	5.294	6.311
76	S	Et	Bn	3-MeOBn	–	2.319	1.725	2.409
77	S	2-(HO)Et	Bn	4-FBn	–	3.569	3.555	3.349
78	S	$EtOCOCH_2$	Bn	4-FBn	–	3.310	3.573	3.793
79	S	Et	2-(Ph)Et	4-FBn	–	2.000	3.073	2.550
80	CH_2	$EtNHCONHCH_2$	Bn	4-FBn	–	3.161	3.507	3.393
81	CH_2	2-(tBuNHCONH)Et	Bn	4-FBn	–	3.745	4.247	3.822
82	CH_2	$2\text{-}ClPhNHCONHCH_2$	Bn	4-FBn	–	4.081	4.031	5.097
83	CH_2	$tBuNHCONHCH_2$	Bn	4-MeOBn	–	3.469	3.658	3.694
120	–	2-(MeO)Bn	iPr	cPr	H	2.678	2.832	2.844
121	–	2-(MeO),5-(MeOCO)Bn	iPr	cPr	H	3.009	3.189	3.415
122	–	$2\text{-}(MeO),5\text{-}(O_2N)Bn$	(4-pyridinyl)Me	tBuNH	H	4.328	4.551	4.564
123	–	$2\text{-}(MeO),5\text{-}(O_2N)Bn$	iPr	iPr	H	3.602	2.849	3.053
124	–	$2\text{-}(MeO),5\text{-}(O_2N)Bn$	nPr	tBuNH	MeOCO	4.432	3.780	4.579
125	–	$2\text{-}(MeO),5\text{-}(O_2N)Bn$	nPr	tBuNH	HOCO	4.678	4.744	4.639
126	–	$2\text{-}(MeO),5\text{-}(O_2N)Bn$	nPr	tBuNH	2-(4-pyridinyl)EtNHCO	3.886	4.319	5.306
127	–	$2\text{-}(MeO),5\text{-}(O_2N)Bn$	nPr	tBuNH	2-(2-pyridinyl)EtNHCO	4.569	4.222	4.788
128	–	$2\text{-}(MeO),5\text{-}(O_2N)Bn$	nPr	tBuNH	$2\text{-}(H_2N)EtNHCO$	3.678	4.527	3.577

[a] Compound **5** is from Ref. [7], compounds **25–30** are from Ref. [19], compounds **40–41** are from Ref. [18], compounds **76–79** are from Ref. [11], compounds **80–83** are from Ref. [20], compounds **120–128** are from Ref. [21].

ANNs is that they are ''universal approximators'' [31]. This means that ANNs are capable of approximating any mapping between independent variables and response variables when the number of neurons of the hidden layers tends to increase. Unfortunately, it is of little help in estimation problems where increasing in the number of parameters often leads to overfitting: the predictor memorizes the training examples, but it has not learned to generalize to new situations.

For overcoming the deficiencies of neural networks, Bayesian techniques have been successfully applied in the context of both regression and classification problems [32,33]. In Bayesian-regularized ANNs (BRANNs), Bayesian approach yields a posterior distribution of network parameters, conditional on the training data and predictions are expressed in terms of expectations with respect to this posterior distribution.

Assuming a set of pairs $D = \{x_i, t_i\}$, where $i = 1 - N$ is a label running over the pairs, the data set can be modeled as deviating from this mapping under some additive noise process (v_i):

$$t_i = y_i + v_i \tag{6}$$

If v is modeled as zero-mean Gaussian noise with standard deviation σ_v, then the probability of the data, given the parameters w, is:

$$P(D|w, \beta, M) = \frac{1}{Z_D(\beta)} \exp(-\beta \times \mathrm{MSE}) \tag{7}$$

where M is the particular neural network model used, $\beta = 1/\sigma_v^2$ and the normalization constant is given by $Z_D(\beta) = (\pi/\beta)^{N/2}$. $P(D|w, \beta, M)$ is called the likelihood. The maximum likelihood parameters w_{ML} (the w that minimizes MSE) depend sensitively on the details of the noise in the data.

For completing the interpolation model, a prior probability distribution must be defined which embodies our prior knowledge on the sort of mappings that are ''reasonable'' [34]. Typically, this is quite a broad distribution, reflecting the fact that we only have a vague belief in a range of possible parameter values. Once we have observed the data, Bayes' theorem can be used to update our beliefs, and we obtain the posterior probability density. As a result, the posterior distribution is concentrated on a smaller range of values than the prior distribution. Since a neural network with large weights will usually give rise to a mapping with a large curvature, we favor small values for the network weights. At this point, a prior value is defined that expresses the sort of smoothness that the interpolant is expected to have. The model has the form:

$$P(w|\alpha, M) = \frac{1}{Z_W(\alpha)} \exp(-\alpha \times \mathrm{MSW}) \tag{8}$$

where α represents the inverse variance of the distribution and the normalization constant is given by $Z_W(\alpha) = (\pi/\alpha)^{N/2}$. MSW is the mean of the sum of the squares of the network weights and is commonly referred to as a regularizing function.

Considering the first level of inference, if α and β are known, the posterior probability of the parameters w is:

$$P(w|D,\alpha,\beta,M) = \frac{P(D|w,\beta,M) \times P(w|\alpha,M)}{P(D|\alpha,\beta,M)} \quad (9)$$

where $P(w|D,\alpha,\beta,M)$ is the posterior probability, that is the plausibility of a weight distribution considering the information of the data set in the model used, $P(w|\alpha,M)$ is the prior density, which represents our knowledge of the weights before any data is collected; $P(D|w,\beta,M)$ is the likelihood function, which is the probability of the data occurring, given the weights; and $P(D|\alpha,\beta,M)$ is a normalization factor, which guarantees that the total probability is 1.

Considering that the noise in the training set data is Gaussian and that the prior distribution for the weights is Gaussian, the posterior probability fulfills the relation:

$$P(w|D,\alpha,\beta,M) = \frac{1}{Z_F}\exp(-F) \quad (10)$$

where Z_F depends on objective function parameters. So, under this framework, the minimization of F is identical to finding the (locally) most probable parameters [32].

In short, Bayesian regularization involves modifying the performance function (F) defined in Eq. (5), which is possible when improving generalization by adding an additional term that regularizes the weights by penalizing overly large magnitudes:

$$F = \beta \times \mathrm{MSE} + \alpha \times \mathrm{MSW} \quad (11)$$

The relative size of the objective function parameters α and β dictates the emphasis for getting a smoother network response. MacKay's Bayesian framework automatically adapts the regularization parameters to maximize the evidence of the training data [32].

Bayesian regularization diminishes the inherent complexity of neural networks, being governed by Occam's Razor [24]; in this sense, it produces predictors that are robust and well matched to the data. These properties become BRANNs in accurate predictors for QSAR analysis [35]. The joining with GA feature selection (BRGNN) increases the possibilities of BRANNs for modeling as we indicated in previous works [28,36].

Fully connected, three-layer BRANNs with back-propagation training were implemented in the MATLAB environment [29]. In these nets, the transfer functions of input and output layers were linear, and the hidden layer had neurons with a hyperbolic tangent transfer function. Inputs and targets took the values from independent variables selected by the GA and $\log(10^6/\mathrm{IC}_{50})$ values, respectively; both were normalized prior to network training. BRANN training was carried out according to the Levenberg–Marquardt optimization [37]. The initial value for μ was 0.005 with decrease and increase factors of 0.1 and 10, respectively. The training was stopped when μ became larger than 10^{10}.

2.5. Analysis of the quality of the models

The quality of the fit of the training set of a specific model was measured by its R^2. However, a most important measure is the prediction quality. In this sense, we follow criteria of Golbraikh and Tropsha [38] who formulated that, for a QSAR model to have highly predictive power, it is necessary a high value of correlation coefficient for cross-validation experiments and a high value of correlation coefficient between the predicted and observed activities of compounds from an external test set.

An internal LOO cross-validation process was carried out by estimating R^2 of LOO cross-validation $\left(R^2_{\mathrm{CV}}\right)$. A data point was removed (left-out) from the set, and the model was refitted; the predicted value for that point is then compared to its actual value. This is repeated until each datum has been omitted once; the sum of squares of these deletion residuals can then be used to calculate R^2_{CV}. In addition, the predictive power of the model was also measured by an external validation process that consists in predicting the activity of unknown compounds forming the test set. In this case R^2 of test set fitting is calculated $\left(R^2_{\mathrm{EP}}\right)$.

Ultimately, we carried out randomization experiments. They were performed on the data to ensure that the best linear and non-linear models presented here were the result of a structure-activity relationship and not due to random effects. The dependent variables were randomly scrambled, and models were trained using the above methodologies. A decrease in R^2 was desired to show that scrambled activity values had little correlation to their structures. The results for the scrambled training sets are presented as the maximum values of R^2 found when the training set were scrambled 50 times.

2.6. Kohonen self-organizing neural network

Kohonen SOMs [39] have the special property of effectively creating a spatially organized internal representation of various features of input signals and their abstractions, following an unsupervised and competitive process. In a self-organizing neural network the neurons are arranged in a two-dimensional array to generate a two-dimensional feature map such that similarity in the data is preserved; if two input data vectors are similar, they will be mapped into the same neuron or into neurons close together in the two-dimensional map. Similar features in output vectors will be grouped if adequate variables are selected.

Learning in a self-organizing feature map occurs for one vector at a time. First, the network identifies the winning neuron, and then, the weights of the winning neuron and the other neurons in its neighborhood are moved closer to the input vector at each learning step. The winning neuron's weights are altered proportional to the learning rate. The weights of neurons in its neighborhood are altered proportional to half the learning rate. The learning rate and the neighborhood distance used to determine which neurons are in the winning neuron's neighborhood are altered during training through two phases: ordering phase that decreases the distances between neurons until the tuning neighborhood distance and tuning phase that tunes the network keeping the ordering learned in the previous phase.

Table 7
Statistic parameters of the linear and non-linear models for the binding affinity (IC_{50}) for the non-peptide antagonists for the human LHRH receptor

Model	Descriptors[a]	Training set			LOO cross-validation		Test set		
		N	R^2	S	R^2_{CV}	S_{CV}	N	R^2_{EP}	S_{EP}
MLR	ATS1v, MATS7m, MATS6p, GATS1e, GATS2e, GATS7e, GATS8e, GATS6p	102	0.746	0.632	0.695	0.693	26	0.733	0.633
BRGNN[b]	ATS1v, MATS2m, MATS8m, MATS3e, MATS4p, GATS1e, GATS7e, GATS6p	102	0.871	0.452	0.752	0.628	26	0.853	0.499

[a] Descriptor definitions: ATS1v: Broto–Moreau autocorrelation coefficient of lag 1 weighted by atomic van der Waals volumes; MATS2m, MATS7m and MATS8m: Moran autocorrelation of lag 2, 7 and 8 weighted by atomic masses; MATS3e: Moran autocorrelation of lag 3 weighted by atomic Sanderson electronegativities; MATS4p and MATS6p: Moran autocorrelation of lag 4 and 6 weighted by atomic polarizabilities; GATS1e, GATS2e, GATS7e and GATS8e: Geary autocorrelation of lag 1, 2, 7 and 8 weighted by atomic Sanderson electronegativities; GATS6p: Geary autocorrelation of lag 6 weighted by atomic polarizabilities.
[b] 8-3-1 architecture.

3. Results and discussion

3.1. Multiple linear regression analysis

In a first approach, MLR analysis combined with a linear GA feature selection was performed on the training set (Tables 1–5). We chose the best model as the most reliable linear relationship between the calculated 2D autocorrelation descriptors and the binding affinity $\log(10^6/IC_{50})$. The best linear model is shown in Eq. (12) (Table 7):

$$\begin{aligned}\log(10^6/IC_{50}) &= 0.078 \times ATS1v + 66.860 \times MATS7m \\ &\quad + 16.266 \times MATS6p + 9.628 \times GATS1e \\ &\quad - 5.299 \times GATS2e + 2.755 \times GATS7e \\ &\quad - 3.231 \times GATS8e + 12.047 \times GATS6p \\ &\quad - 79.635 \\ N = 102, &\quad R^2 = 0.746, \quad S = 0.632, \quad p < 10^{-5}, \\ R^2_{CV} = 0.695, &\quad S_{CV} = 0.693\end{aligned} \tag{12}$$

where N is the number of compounds included in the model, R^2 is the square correlation coefficient, S is the standard deviation of the regression, and p is the significance of the variables in the model. R^2_{CV} and S_{CV} are the correlation coefficient and standard deviation of the LOO cross-validation, respectively.

Eq. (12) shows an eight descriptor-model including one Broto–Moreau's autocorrelation coefficient (ATS1v), two Moran's indices (MATS7m and MATS6p) and five Geary's coefficients (GATS1e, GATS2e, GATS7e, GATS8e and GATS6p). It is noteworthy that there is no significant intercorrelation between these descriptors, as it is seen in Table 8. This model is able to explain about 74% of antagonistic activity variance. Since R^2_{CV} value was about 0.695, the model was considered to be a good predictive one, according to Wold [40].

The predicted binding affinities $\log(10^6/IC_{50})$ for the 102 non-peptide antagonists for the human LHRH receptor, which were used in training process, appears in Tables 1–5. Otherwise, the external validation showed $R^2_{EP} = 0.733$ for the test set ($\log(10^6/IC_{50})$ values for the 26 antagonists used in external predictions appear in Table 6). The plots of predicted versus experimental $\log(10^6/IC_{50})$ values for training and test sets are given in Fig. 2A and B. In spite of the fact that several atomic properties were included in the MLR model, it is notable the predominance of the atomic electronegativities represented by the Geary autocorrelation space.

Finally, randomization studies were performed to verify that the model found by MLR was not due to chance correlations. The maximum R^2-value when the training set was scrambled 50 times and MLR analysis was applied to each randomized data set was $R^2 = 0.168$. From this, we concluded that chance correlation had little or no effect in driving model development.

3.2. Bayesian-regularized artificial neural networks

As biological phenomena are considered non-linear by nature, ANN technique was applied in order to discover the possible existence of non-linear relationships between binding affinity and molecular descriptors that are ignored for the linear

Table 8
Correlation matrix of the descriptors selected by linear GA

	ATS1v	MATS7m	MATS6p	GATS1e	GATS2e	GATS7e	GATS8e	GATS6p
ATS1v	1							
MATS7m	0.222	1						
MATS6p	0.012	0.184	1					
GATS1e	0.026	0.002	0.178	1				
GATS2e	0.000	0.046	0.010	0.355	1			
GATS7e	0.487	0.137	0.001	0.055	0.125	1		
GATS8e	0.168	0.112	0.046	0.024	0.000	0.261	1	
GATS6p	0.010	0.003	0.543	0.090	0.018	0.030	0.080	1

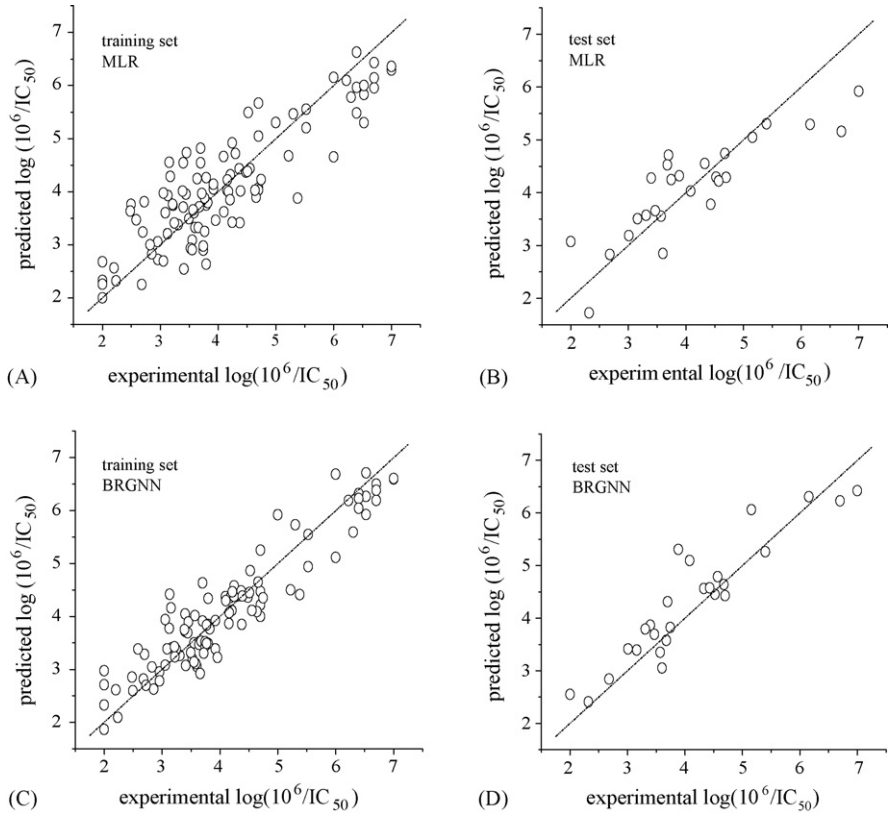

Fig. 2. Predicted vs. experimental activities for the data set. (A) MLR training set, (B) MLR test set, (C) BRGNN training set and (D) BRGNN test set.

model. We extracted eight descriptors by means of the non-linear GA for constructing the BRGNN model. Since we used a reduced architecture (8-2-1) in order to offer stability to the feature selection process, we explored BRANN architectures testing different quantities of neurons in the hidden layer. Results were quite stable because the Bayesian regularization avoids overfitting, however the inclusion of three hidden neurons gave the better statistics according to LOO cross-validation experiments.

BRGNN approach yielded an optimum variable subset that was similar in comparison with the descriptor subset of the linear model (Table 7). BRGNN analysis includes one Broto–Moreau's autocorrelation coefficient (ATS1v), four Moran's indices (MATS2m, MATS8m, MATS3e and MATS4p) and three Geary's coefficients (GATS1e, GATS7e and GATS6p). As it was obtained using MLR approach, there is no significant intercorrelation between the selected descriptors (Table 9).

The predictions of $\log(10^6/IC_{50})$ values for the 102 non-peptide antagonists using the BRGNN model appears in Tables 1–5; otherwise, external predictions appear in Table 6. The plots of predicted versus experimental $\log(10^6/IC_{50})$ values for training and test sets are given in Fig. 2C and D. An improvement in the reliability on the modeling of the LHRH antagonistic activity was achieved by the BRGNN procedure (Table 7). The non-linear model overcame the MLR one by fitting the training set with a higher R^2 of 0.871 in comparison with 0.746 the linear model. When comparing this predictor with the MLR model developed in this work, we found that, beside the improvement on the fitting of training set, the most remarkable result is the increment in predictive power. In fact, BRGNN model exhibited higher predictive power measured by internal and external validation processes. BRGNN exhibits a higher R^2_{CV} of LOO cross-validation $\left(R^2_{CV} = 0.752\right)$ in comparison to MLR model $\left(R^2_{CV} = 0.695\right)$. In addition, the increment in the quality

Table 9
Correlation matrix of the descriptors selected by non-linear GA

	ATS1v	MATS2m	MATS8m	MATS3e	MATS4p	GATS1e	GATS7e	GATS6p
ATS1v	1							
MATS2m	0.074	1						
MATS8m	0.060	0.053	1					
MATS3e	0.314	0.000	0.048	1				
MATS4p	0.031	0.007	0.009	0.314	1			
GATS1e	0.026	0.022	0.336	0.002	0.001	1		
GATS7e	0.487	0.001	0.048	0.537	0.337	0.055	1	
GATS6p	0.010	0.209	0.302	0.002	0.089	0.090	0.030	1

of the prediction of the external test set is showy, from $R^2_{EP} = 0.733$ for the MLR model to $R^2_{EP} = 0.853$ for the BRGNN one (Table 7). In fine, BRGNNs overcame linear models by increasing data fitting and the predictors exhibit a remarkable improvement in predictive power. These results prove the greater adjusting of non-linear analysis to the present paradigm.

Like in the linear approach, the scrambled models were performed to verify that the model found by BRGNN approach was not due to chance correlations. The maximum R^2-value when the training set was scrambled 50 times and BRGNN analysis was applied to each randomized data set was $R^2 = 0.150$. The much lower value of R^2 for the scrambled data confirms that the correlation found previously was not due to chance correlations.

3.3. Kohonen self-organizing neural network analysis

In order to achieve data differentiation using the eight descriptors in the best BRGNN predictor, a SOM with 13 × 13 neurons was mapped with these descriptors as input vectors and both training and test sets were included in training process. Neurons were initially located at a gridtop topology map. The ordering phase was realized in 1000 steps with learning rate = 0.9 until tuning neighborhood distance (1.0) was achieved. The tuning phase learning rate was 0.02. Training was performed for a period of 2000 epochs in an unsupervised manner.

Fig. 3 depicts the Kohonen SOM for the 128 human LHRH antagonists. It is clearly visible that compounds were adequately distributed across the entire map: 77 of a total of 169 neurons were occupied. As it is observed, compounds with common structural features are placed in well-defined zones. Moreover, analysis of the map also reveals that compounds that share similar biological activities are grouped into neighboring areas. The most active compounds (thieno[2,3-*b*]pyridine-4-ones, thieno[2,3-*d*]pyrimidine-2,4-diones and imidazo[1,2-*a*]pyrimidin-5-ones) are grouped in the upper-left zone, meanwhile the lower-right zone is mainly occupied by benzimidazol-5-yl amides with $\log(10^6/IC_{50})$ around 4.0 and 5.0. Otherwise, the less active benzimidazol-5-yl amides are located in lower-left zone and the other less-active compounds (benzimidazole-5-sulfonamides) are in the upper-right zone. The accurate discrimination among more and less active compounds demonstrates the sensitivity of the BRGNN approach and preferred descriptor combination.

3.4. 2D autocorrelation space and LHRH antagonistic activity

Despite the obvious stints of information from 2D representations of molecules, 2D autocorrelation descriptors have found successful applications in the performance of QSAR modeling [41,42]. Although more clear information can be obtained from ''traditional'' descriptors, such as log *P* or pK_a, several authors pointed out problems with the definition of a chemistry space using them [43]. In general, ''traditional'' descriptors often are highly correlated and cannot distinguish the details of important substructural differences. In the other side, 3D descriptors are highly dependent on conformations of molecules; consequently an adequate three-dimensional application requires the identification of active conformations (the receptor-bound conformation) among multiple molecular shapes of the compounds.

The 2D autocorrelation descriptors represent the topological structure of the compounds, but are more complex in nature when compared to the classical topological descriptors. The

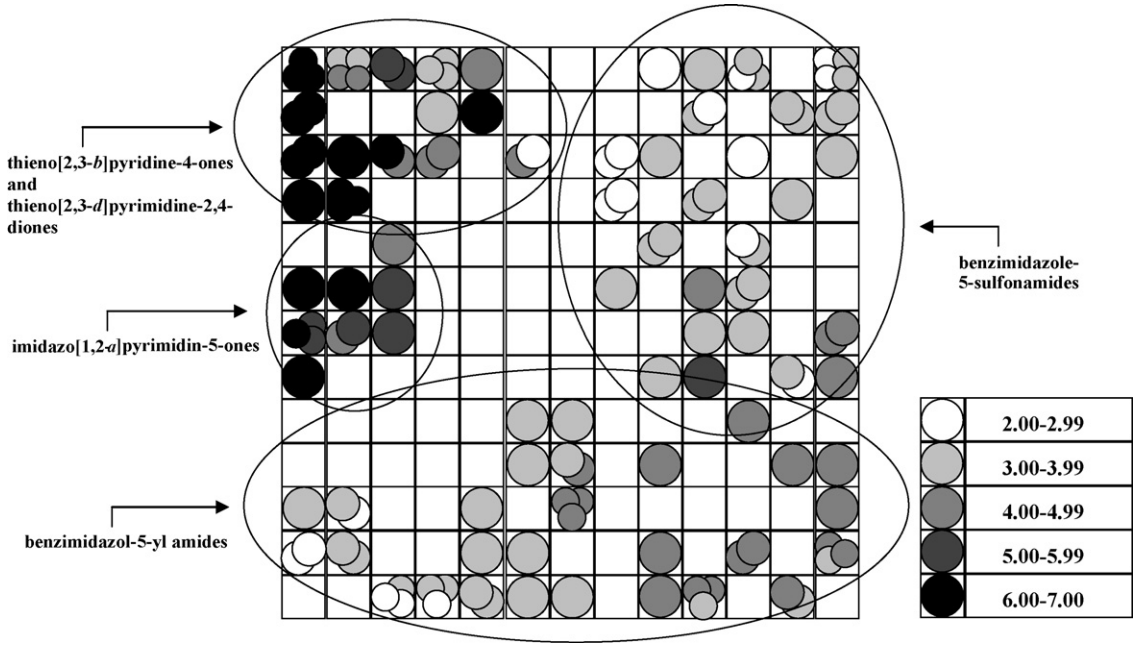

Fig. 3. SOM for the data set using descriptors selected by non-linear genetic algorithm. Squares and circles denote neurons and compounds, respectively. Legend at right decodes the ranges of LHRH antagonistic activities ($\log(10^6/IC_{50})$). Zones in the map grouping compounds with common structural features are indicated.

computation of these descriptors involves the summation of different autocorrelation functions corresponding to the different fragment lengths and leads to different autocorrelation vectors corresponding to the lengths of the structural fragments. Bearing in mind this aspect, the interpretation of 2D autocorrelation descriptors is uneasy.

Basically, the pool of 2D autocorrelation descriptors defines a wide 2D space. In behalf of a greater applicability, physicochemical properties are inserted as weighting components. As a result, these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. For a sound application, we found that the presence of very few structural keys yields performance better than the complete set. In point of fact, the BRGNN selected descriptor combination, which includes atomic electronegativities, masses and polarizabilities as the most relevant structural keys, illustrates that a certain distribution of these properties is necessarily required for typifying the LHRH antagonistic activity.

The structural design of non-peptide LHRH antagonists has faced up to the mimicking of a type II β-turn involving residues 5–8 (Tyr-Gly-Leu-Arg) of LHRH that has been considered to be the dominant structure for its binding to the receptor [44]. In this sense, a bicyclic heterocycle ''scaffold'' bearing important functional groups for receptor binding provides a non-peptide ligand of the LHRH receptor. The ligand-binding pocket affinity of the receptor is mainly governed by hydrophobic interactions. According to SAR studies, the introduction of different functional groups into a common bicyclic scaffold leads to discrepancies with regard to the value of activity. Our BRGNN model encodes this effect by means of a complex relation between the topological structure and deformability and shape of the antagonists.

Beyond previous analysis, both linear and non-linear models here obtained showed that the electronic distribution through the heterocyclic ''scaffold'' has a great influence on the ligand-receptor interaction for the studied compounds. These facts well agree with recent reports in which the LHRH antagonistic activity is highly influenced by the electronic distribution across the heterocyclic frame [11,20,21]. In addition, it has been confirmed that Asp302 from human LHRH receptor recognizes positively charged residues (Arg8 in LHRH). Therefore, a sufficiently positively charged group is another requirement for high-affinity binding to the LHRH receptor.

Indubitably, the identification of such possible structural features that contribute to the biological activity is desirable in order to orientate the synthesis of novel derivatives by minimizing the groups that contribute negatively to the activity and maximizing those with positive contribution. However, it is straightforward to realize that the biological activity of a molecule is the result of the contribution of all its fragments and is not necessarily determined by the influence of only one specific fragment. The computation of the 2D autocorrelation descriptors implicates the integration of the structural fragments. We proposed an explanation to the 2D autocorrelation space; however, a further attempt to uncoil the content of these descriptors would involve traversing backward from a higher state to a lower one.

4. Conclusions

We applied QSAR methodology for modeling the specific LHRH antagonistic activity of 128 non-peptidic heterocyclic compounds using 2D autocorrelation descriptors. For selecting the pertinent and informative variables we carried out linear and non-linear GA searches. Satisfactory quantitative models have been obtained using MLR and BRGNN approaches, achieving also good results in leave-one-out cross validations, external predictions and random tests. The better results achieved by BRGNN approach suggest that structure-LHRH antagonistic activity relationship is a complex phenomenon that can be appropriately described by non-linear analysis and the chemical characteristics it captures.

The present work demonstrates the wide applicability of BRGNN approach like a robust and sensitive method to quantify structure-activity relationships. In general, ANNs are more appropriate than other methods for particular problems at hand. Their general ability to model non-linear effects makes them useful tools to solve relationships when traditional linear models fail. Moreover, Bayesian-regularization is a way to avoid the well-known shortcomings of neural networks; in this sense the combination of BRANNs with GA (BRGNN) lead to a powerful method for the scientific community interested in QSAR developing.

References

[1] (a) A.V. Schally, A. Arimura, A.J. Kastin, H. Matsuo, Y. Baba, T.W. Redding, R.M.G. Nair, L. Debeljuk, W.F. White, Gonadotropin-releasing hormone: one polypeptide regulates secretion of luteinizing and follicle-stimulating hormones, Science 173 (1971) 1036–1038;
(b) H. Matsuo, Y. Baba, R.M.G. Nair, A. Arimura, A.V. Schally, Structure of the porcine LH- and FSH-releasing hormone. I. The proposed amino acid sequence, Biochem. Biophys. Res. Commun. 43 (1971) 1334–1339.

[2] (a) M. Filicori, C. Flamigni, GnRH Agonists and antagonists current clinical status, Drugs 35 (1988) 63–82;
(b) A.J. Friedman, The biochemistry, physiology, and pharmacology of gonadotropin releasing hormone (GnRH) and GnRH analogs, in: R.L. Barbieri, A.J. Friedmann (Eds.), Gonadotropin Releasing Hormone Analogs Applications in Gynecology, Elsevier, New York, 1990, pp. 1–15.

[3] J.J. Nestor Jr., R. Tahilramani, T.L. Ho, G.I. McRae, B.H. Vickery, Luteinizing hormone-releasing hormone antagonists containing very hydrophobic amino acids, J. Med. Chem. 27 (1984) 1170–1174.

[4] F. Haviv, T.D. Fitzpatrick, C.J. Nichols, R.E. Swenson, N.A. Mort, E.N. Bush, G. Diaz, A.T. Nguyen, M.R. Holst, V.A. Cybulski, J.A. Leal, G. Bammert, N.S. Rhutasel, P.W. Dodge, E.S. Johnson, J.B. Cannon, J. Knittle, J. Greer, The effect of NMeTyr6 substitution in luteinizing hormone-releasing hormone antagonists, J. Med. Chem. 36 (1993) 928–933.

[5] A. Janecka, T. Janecki, S. Shan, C. Bowers, K. Folkerst, Novel, potent luteinizing hormone-releasing hormone antagonists with improved solubility in water, J. Med. Chem. 37 (1994) 2238–2241.

[6] M.J. Karten, An overview of GnRH antagonist development: two decades of progress, in: W.F. Crowley, Jr., P.M. Conn (Eds.), Modes of Action of GnRH and GnRH Analogs, Elsevier, New York, 1992, pp. 277–297.

[7] N. Cho, M. Harada, T. Imaeda, T. Imada, H. Matsumoto, Y. Hayase, S. Sasaki, S. Furuya, N. Suzuki, S. Okubo, K. Ogi, S. Endo, H. Onda, M. Fujino, Discovery of a novel, potent, and orally active non-peptide

antagonist of the human luteinizing hormone-releasing hormone (LHRH) receptor, J. Med. Chem. 41 (1998) 4190–4195.
[8] R.J. De Vita, D.D. Hollings, M.T. Goulet, M.J. Wyvratt, M.H. Fisher, J.-L. Lo, Y.T. Yang, K. Cheng, R.G. Smith, Identification and initial structure-activity relationships of a novel non-peptide quinolone GnRH receptor antagonist, Bioorg. Med. Chem. Lett. 9 (1999) 2615–2620.
[9] L. Chu, J.E. Hutchins, A.E. Weber, J.-L. Lo, Y.-T. Yang, K. Cheng, R.G. Smith, M.H. Fisher, M.J. Wyvratt, M.T. Goulet, Initial structure-activity relationship of a novel class of non-peptidyl GnRH receptor antagonists: 2-arylindoles, Bioorg. Med. Chem. Lett. 11 (2001) 509–513.
[10] Y.-F. Zhu, R.S. Struthers, P.J. Connors Jr., Y. Gao, T.D. Gross, J. Saunders, K. Wilcoxen, G.J. Reinhart, N. Ling, C. Chen, Initial structure-activity relationship studies of a novel series of Pyrrolo[1,2-*a*]pyrimid-7-ones as GnRH receptor antagonists, Bioorg. Med. Chem. Lett. 12 (2002) 399–402.
[11] K. Hashimoto, M. Tatsuta, M. Kataoka, K. Yasoshima, Y. Shogase, M. Shimazaki, T. Yura, Y. Li, N. Yamamoto, J.B. Gupta, K. Urbahns, Benzimidazole derivatives as novel non-peptide luteinizing hormone-releasing hormone (LHRH) antagonists. Part I. Benzimidazole-5-sulfonamides, Bioorg. Med. Chem. Lett. 15 (2005) 799–803.
[12] C. Hansch, A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, ACS Professional Reference Book, American Chemical Society, Washington DC, 1995.
[13] H. Bauknecht, A. Zell, H. Bayer, P. Levi, M. Wagener, J. Sadowski, J. Gasteiger, Locating biologically active compounds in medium-sized heterogeneous datasets by topological autocorrelation vectors: dopamine and benzodiazepine agonists, J. Chem. Inf. Comp. Sci. 36 (1996) 1205–1213.
[14] G. Moreau, P. Broto, Autocorrelation of a topological structure: a new molecular descriptor, Nouv. J. Chim. 4 (1980) 359–360.
[15] P.A.P. Moran, Notes on continuous stochastic processes, Biometrika 37 (1950) 17–23.
[16] R.F. Geary, The contiguity ratio and statistical mapping, The Incorporated Statistician 5 (1954) 115–145.
[17] Dragon, version 3.0, Milano Chemometrics and QSAR research group, Milan, 2003.
[18] S. Sasaki, T. Imaeda, Y. Hayase, Y. Shimizu, S. Kasai, N. Cho, M. Harada, N. Suzuki, S. Furuya, M. Fujino, A new class of potent non-peptide luteinizing hormone-releasing hormone (LHRH) antagonists: design and synthesis of 2-phenylimidazo[1,2-*a*]pyrimidin-5-ones, Bioorg. Med. Chem. Lett. 12 (2002) 2073–2077.
[19] S. Sasaki, N. Cho, Y. Nara, M. Harada, S. Endo, N. Suzuki, S. Furuya, M. Fujino, Discovery of a thieno[2,3-*d*]pyrimidine-2,4-dione bearing a *p*-methoxyureidophenyl moiety at the 6-position: a highly potent and orally bioavailable non-peptide antagonist for the human luteinizing hormone-releasing hormone receptor, J. Med. Chem. 46 (2003) 113–124.
[20] Y. Li, M. Kataoka, M. Tatsuta, K. Yasoshima, T. Yura, K. Urbahns, A. Kiba, N. Yamamoto, J.B. Gupta, K. Hashimoto, Benzimidazole derivatives as novel non-peptide luteinizing hormone-releasing hormone (LHRH) antagonists. Part II. Benzimidazole-5-sulfonamides, Bioorg. Med. Chem. Lett. 15 (2005) 805–807.
[21] M. Tatsuta, M. Kataoka, K. Yasoshima, S. Sakakibara, Y. Shogase, M. Shimazaki, T. Yura, Y. Li, N. Yamamoto, J. Gupta, K. Urbahns, Benzimidazoles as non-peptide luteinizing hormone-releasing hormone (LHRH) antagonists. Part III. Discovery of 1-(1H-benzimidazol-5-yl)-3-*tert*-butylurea derivatives, Bioorg. Med. Chem. Lett. 15 (2005) 2265–2269.
[22] J.J.P. Stewart, Optimization of parameters for semi-empirical methods I-method, J. Comp. Chem. 10 (1989) 210–222.
[23] MOPAC, version 6.0, Frank J. Seiler Research Laboratory, US Air Force academy, Colorado Springs, CO, 1993.
[24] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Comp. Sci. 44 (2004) 1–12.
[25] H. Holland, Adaption in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, MI, 1975.
[26] H.M. Cartwright, Applications of Artificial Intelligence in Chemistry, Oxford University Press, Oxford, 1993.
[27] S.S. So, M. Karplus, Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks, J. Med. Chem. 39 (1996) 1521–1530.
[28] J. Caballero, M. Fernández, Linear and non-linear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks, J. Mol. Model. 12 (2006) 168–181.
[29] MATLAB, version 7.0, The MathWorks Inc., 2004. Web: http://www.mathworks.com.
[30] J. Zupan, J. Gasteiger, Neural networks: a new method for solving chemical problems or just a passing fase? Anal. Chim. Acta 248 (1991) 1–30.
[31] (a) H. White, Learning in artificial neural networks: a statistical perspective, Neural Comput. 1 (1989) 425–464;
(b) K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366.
[32] (a) D.J.C. Mackay, Bayesian interpolation, Neural Comput. 4 (1992) 415–447;
(b) D.J.C. Mackay, A practical Bayesian framework for backprop networks, Neural Comput. 4 (1992) 448–472.
[33] R.M. Neal, Bayesian learning for neural networks, Lecture Notes in Statistics, vol. 118, Springer-Verlag, New York, 1996.
[34] J. Lampinen, A. Vehtari, Bayesian approach for neural networks—review and case studies, Neural Networks 14 (2001) 7–24.
[35] (a) F.R. Burden, D.A. Winkler, Robust QSAR models using Bayesian regularized neural networks, J. Med. Chem. 42 (1999) 3183–3187;
(b) D.A. Winkler, F.R. Burden, Bayesian neural nets for modeling in drug discovery, Biosilico 2 (2004) 104–111.
[36] (a) M.P. González, J. Caballero, A. Tundidor-Camba, A.M. Helguera, M. Fernández, Modeling of farnesyltransferase inhibition by some thiol and non-thiol peptidomimetic inhibitors using genetic neural networks and RDF approaches, Bioorg. Med. Chem. 14 (2006) 200–213;
(b) M. Fernández, J. Caballero, Modeling of activity of cyclic urea HIV-1 protease inhibitors using regularized-artificial neural networks, Bioorg. Med. Chem. 14 (2006) 280–294.
[37] F.D. Foresee, M.T. Hagan, Gauss–Newton approximation to Bayesian learning, in: Proceedings of the 1997 International Joint Conference on Neural Networks, 1997, pp. 1930–1935.
[38] A. Golbraikh, A. Tropsha, Beware of q^2! J. Mol. Graph. Model. 20 (2002) 269–276.
[39] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybern. 43 (1982) 59–69.
[40] S. Wold, Validation of QSARs, Quant. Struct. Act. Relat. 10 (1991) 191–193.
[41] M. Fernández, A. Tundidor-Camba, J. Caballero, 2D autocorrelation modeling of the activity of trihalobenzocycloheptapyridine analogues as farnesyl protein transferase inhibitors, Mol. Simulat. 31 (2005) 575–584.
[42] M.K. Gupta, R. Sagar, A.K. Shaw, Y.S. Prabhakar, CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols—topological descriptors in modeling the activity, Bioorg. Med. Chem. 13 (2005) 343–351.
[43] R.S. Pearlman, K.M. Smith, Metric validation and the receptor-relevant subspace concept, J. Chem. Inf. Comp. Sci. 39 (1999) 28–35.
[44] M.J. Karten, J.E. Rivier, Gonadotropin-releasing hormone analog design. Structure-function studies toward the development of agonists and antagonists: rationale and perspective, Endocr. Rev. 7 (1986) 44–66.