Available online at www.sciencedirect.com
SCIENCE DIRECT
European Journal of Medicinal Chemistry 41 (2006) 552–557
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://france.elsevier.com/direct/ejmech

Preliminary communication

Quantitative structure activity relationship of benzoxazinone derivatives as neuropeptide Y Y5 receptor antagonists

S. Deswal, N. Roy *

Pharmacocinformatics Division, National Institute of Pharmaceutical Education and Research, Sector 67, Phase X, S.A.S. Nagar, Punjab, India

Received in revised form 30 December 2005; accepted 2 January 2006
Available online 20 March 2006

Abstract

Quantitative structure activity relationship (QSAR) has been established for 30 benzoxazinone derivatives acting as neuropeptide Y Y5 receptor antagonists. The genetic algorithm and multiple linear regression were used to generate the relationship between biological activity and calculated descriptors. Model with good statistical qualities was developed using four descriptors from topological, thermodynamic, spatial and electrotopological class. The validation of the model was done by cross validation, randomization and external test set prediction.

Keywords: QSAR; Neuropeptide Y Y5 receptor antagonists; Obesity; Genetic function approximation

1. Introduction

Obesity has emerged as an epidemic in last several decades especially in the developed countries. In a recent study it was found that more than 1 billion peoples in world are overweight and at least 300 million of them are obese [1]. The cause of its high prevalence in developed countries might be the changing life style of the peoples in these countries i.e. increased consumption of high-energy diet in combination with lack of physical exercise. Although the molecular mechanism underneath is not clearly understood number of pathophysiological alterations associated with obesity are documented. Based on these studies therapeutic interventions are designed and being designed. Some of the agents which are in use or are under investigation for intervention of obesity are central nervous system agents, neuropeptide Y and agouti-related peptide antagonists, gastrointestinal–neural pathway agents like cholecystokinin enhancers and glucagon-like peptide-1 activity enhancers, resting metabolic rate promoters and agents like melanin concentrating hormone antagonists [2]. Obesity it self is not life threatening however it can significantly increase the risk of life threatening diseases like cardiovascular disease, neurological disorders, respiratory disorders, musculoskeletal disorders, endocrine disorders, gastrointestinal disorders, genitourinary disorders and psychological disorders [3,4]. Therefore it is necessary to develop effective and safe antiobesity drugs to reduce the worldwide obesity epidemic.

Neuropeptide Y is a 36 amino acid peptide and belongs to a large family of peptides, which also include peptide YY and pancreatic polypeptide [5]. It is widely distributed in central [6] as well as peripheral nervous system [7] and is found to be associated with several diseases like diabetes mellitus [8], cardiovascular disorders [9], depression [10], anxiety [10], seizures [11], asthma [12], inflammatory diseases [12] and immune system disorders [13]. It has been observed that NPY is a strong feeding stimulant and its administration reduces energy expenditure [14–16]. Though various biological functions of NPY are mediated by five receptor subtypes i.e. Y1, Y2, Y4, Y5 and Y6, the regulation of feeding behavior is mainly mediated via Y1 and Y5 receptor subtypes [17]. In a number of pharmacological studies it has been established that NPY Y5 antagonists are potential antiobese agents [18]. So development of NPY Y5 antagonists can offer effective antiobese drugs.

Quantitative structure activity relationship (QSAR) is a useful method for the design of bioactive compounds and the prediction of activity from the parameters calculated from chemical structure of compound. There are many examples available in literature where QSAR models have been used for screening of compounds from the chemical databases [19–22]. The

* Corresponding author.
E-mail address: nilanjanroy@niper.ac.in (N. Roy).

0223-5234/$ - see front matter © 2006 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.01.012

QSAR models can be developed by linearly correlating the biological activity to the descriptors or the non-linear regression methods such as artificial neural network (ANN) can be used [23]. In present study we carried out a quantitative structure NPY Y5 receptor inhibitory activity relationship of a series of benzoxazinone derivatives using genetic function approximation (GFA) method for variable selection. The objective of study is to develop a model which can be used to screen the compounds for NPY Y5 receptor inhibitory activity from the available chemical databases. The model can be used for virtual screening by applying Lipinski's rule filters for initial screening and then predicting the activity by QSAR model.

2. Materials and methods

2.1. Data set

The inhibitory activity of the benzoxazinone derivatives was taken from literature in terms of IC_{50} values [24]. The IC_{50} values were converted to pIC_{50} to get the linear relationship in the equation using following formula

$$pIC_{50} = -\log IC_{50} \quad (1)$$

Total set of 30 compounds was divided in training and test set of 24 and six compounds randomly. In the original article the IC_{50} values were given in nM values. To make the interpretation easy, before conversion to pIC_{50}, IC_{50} values were changed to μM unit, so that the pIC_{50} values become in the positive range only. The structures of compounds used in the study along with observed IC_{50} values are provided in Table 1.

2.2. Molecular modeling

All computational work was performed on Silicon Graphics Fuel Work station. The compounds were built using build model program in the Sybyl 6.9 software package [25]. The energy minimization calculations were performed using AM1 method [26] in MOPAC 6.0. The following specific software options were employed while performing AM1 studies: convergence = normal, optimization = full, state = singlet, net charge = 0 e.u., time limit = 3600 s, keyword = mmok.

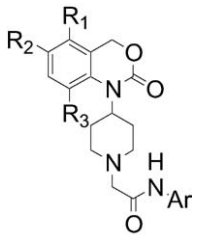
2.3. Descriptor calculation

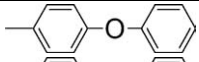
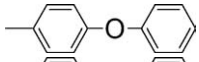
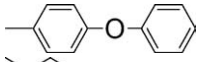
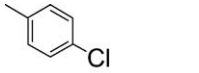
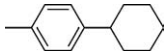
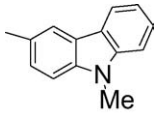
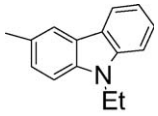
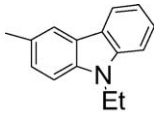
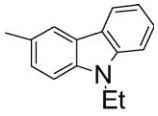
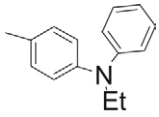
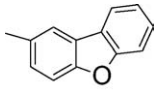
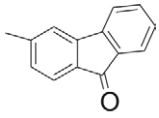
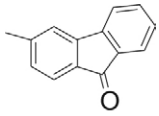
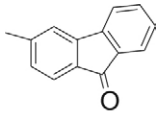
A total of more than 100 descriptors were calculated using Cerius^2 4.10 software package [27]. A brief description of descriptors used which include topological descriptors, spatial descriptors, E-state indices, thermodynamic, electronic and structural descriptors is provided in Table 2.

2.4. Regression analysis

From the total calculated descriptors some of the descriptors e.g. Atype_Br_91, Atype_Br_92, Atype_I_96, Atype_I_97, Atype_I_98, Atype_S_106, Atype_S_107 Atype_P_115, Aty-

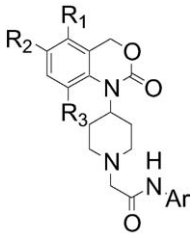
Table 1
Structure and observed IC_{50} of the compounds used for study

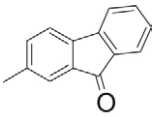
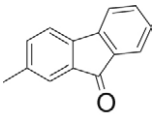
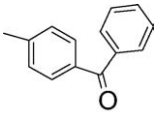
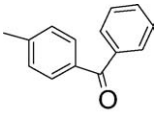
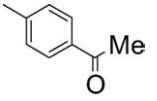
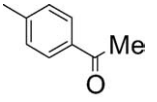
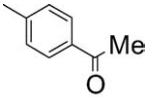
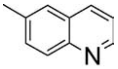
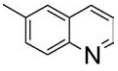
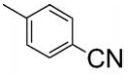
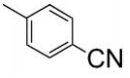
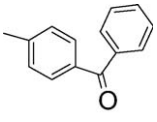
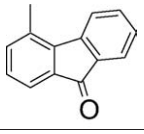


Compounds	R1	R2	R3	Ar	IC_{50} (nM)
1	H	H	H		20
2	H	Me	H		104
3	H	Cl	H		60.7
4	H	H	H		300
5	H	Cl	H		112.4
6	H	H	H		9.6
7	Cl	H	H		54.6
8	H	Cl	H		100
9	H	H	Me		50
10	H	H	OMe		765.1
11	H	H	OMe		55.7
12	H	H	H		23.3
13	H	H	Me		50
14	H	Cl	H		25

(continued)

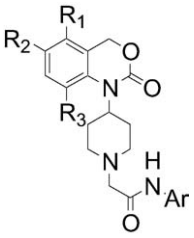
Table 1 *(continued)*

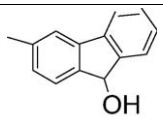
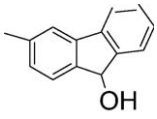
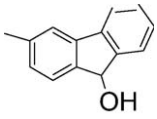


Compounds	R1	R2	R3	Ar	IC_{50} (nM)
15	H	Me	H		86.7
16	H	Cl	H		69
17	H	H	Me		39.6
18	H	Cl	H		11.2
19	H	H	H		17.3
20	H	Cl	H		21
21	H	Me	H		26
22	H	F	H		7.7
23	H	Cl	H		148
24	H	H	H		132
25	H	Cl	H		199
26	H	H	H		7.6
27	H	H	H		138.2

(continued)

Table 1 *(continued)*



Compounds	R1	R2	R3	Ar	IC_{50} (nM)
28	H	Cl	H		8
29	H	H	H		8.7
30	H	H	Me		30

pe_P_116, S_ddC, S_sNH3, S_sNH2 and S_dNH were rejected because they contain a value of 0 for all the compounds. The reason for the value of 0 for all the compounds for these descriptors was that there is no atom corresponding to these descriptors in any of the compounds. Further, the inter-correlation of descriptors was taken in to account and highly correlated descriptors were grouped together manually by analyzing the correlation matrix. Only one descriptor was then taken for further study from each group of highly correlated descriptors. Only remaining descriptors were considered for model development by GFA method. The GFA method works in the following way: first of all few equations (set at 100 by default in the Cerius[2] software) are generated randomly. Then pairs of "parent" equations are chosen for "crossover" operations from this set of 100 equations randomly. The number of crossing over was set by default at 5000. The goodness of each progeny

Table 2
Description of the parameters used in the study

Type	Descriptors
Electronic	Sum of partial charges, sum of formal charges, dipole moment, HOMO, LUMO
Spatial	Radius of gyration, shadow indices, area, density, PMI, Vm
Structural	Molecular weight, number of rotatable bonds, number of hydrogen-bond acceptors, number of hydrogen-bond donors
Thermodynamic	log of the partition coefficient, log of the partition coefficient atom-type value, desolvation free energy of water, desolvation free energy of octanol, heat of formation, molar refractivity
Topological	Wiener index, Zagreb index, Hosoya index, Kier and Hall molecular connectivity index, Balaban indices
E-state indices	Electrotopological-state indices

Table 3
Statistical assessment of equations with varying number of descriptors

Number of descriptors	Equation	LOF	r^2	r^2_{adj}	F-test	LSE	r	q^2	BSr^2
1	pIC_{50} = 3.612 – 0.579*S_sssN	0.269	0.229	0.194	6.540	0.226	0.479	–0.605	0.228
2	pIC_{50} = 2.645 + 0.781*Atype_C_8 – 0.549*S_sssN	0.258	0.388	0.329	6.650	0.179	0.623	–0.455	
3	pIC_{50} = 3.5076 + 0.840*Atype_C_8 – 0.013*Shadow_XZ -0.513*S_sssN	0.276	0.469	0.389	5.880	0.156	0.685	0.370	0.469
4	pIC_{50} = 1.389 – 0.640*S_sssN + 0.303*CHI-2 – 0.034*Shadow_XZ + 0.638*Atype_C_8	0.185	0.720	0.661	12.194	0.082	0.848	0.616	0.720
5	pIC_{50} = 1.442 – 0.591*S_sssN – 0.033*Shadow_XZ – 0.003*Vm + 0.626*Atype_C_8 + 0.366*CHI-2	0.239	0.722	0.644	9.335	0.081	0.850	0.513	0.724

Table 4
Correlation matrix of the descriptors used in equation

	S_sssN	Shadow_XZ	Atype_C_8	CHI-2	Activity
S_sssN	1.000				
Shadow_XZ	0.097	1.000			
Atype_C_8	–0.063	0.099	1.000		
CHI-2	0.202	0.675	0.201	1.000	
Activity	–0.479	–0.286	0.428	0.155	1.000

equation is assessed by Friedman's lack of fit (LOF) score, which is described by following formula

$$\mathrm{LOF} = \mathrm{LSE}/\{1 - (c + dp)/m\}^2 \quad (2)$$

Where LSE is the least-squares error, c is the number of basis functions in the model, d is smoothing parameter, p is the number of descriptors and m is the number of observations in the training set [28]. The smoothing parameter that controls the scoring bias between equations of different sizes was set at default value of 1.0 and the new term was added with a probability of 50%. Only the linear equation terms were used for model building. The best equation out of the 100 equations was taken based on the statistical parameters such as regression coefficient, adjusted regression coefficient, regression coefficient cross validation and F-test values.

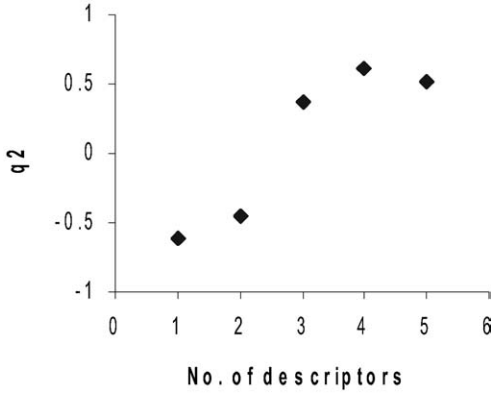


Fig. 1. The cross-validated r^2 (q^2) as a function of number of descriptors. The number of descriptors was varied from 1 to 6 and corresponding value of cross validated r^2 (q^2) was predicted to find the optimum number of descriptors necessary for QSAR development.

3. Results and discussion

As a rule of thumb, data set should be approximately five times more than the number of descriptors used in the model [29]. Thus descriptor reduction was done as described in Section 2. The results of the best QSAR model using one to five descriptors are given in Table 3. As the r^2 value can be easily increased by increasing the number of descriptors in the model, so cross validated correlation coefficient (q^2) was used as a parameter to select the optimum number of descriptors. The variation in cross validation correlation coefficient (q^2) as a function of number of descriptors is shown in Fig. 1. The best model in term of q^2 was obtained with four descriptors and is given as:

$$\mathrm{pIC50} = 1.389 - 0.640 * \mathrm{S_sssN} + 0.303 * \mathrm{CHI-2} - 0.034 * \mathrm{Shadow_XZ} + 0.638 * \mathrm{Atype_C8} \quad (3)$$

$N = 24$; LOF = 0.185; $r^2 = 0.720$, $r^2_{adj} = 0.661$; F-test = 12.194; LSE = 0.082; $r = 0.848$; $q^2 = 0.616$; $BSr^2 = 0.720$, $r^2_{pred} = 0.706$.

Where N is number of compounds in training set, LOF is Lack of Fit score, r^2 is squared correlation coefficient, $r^2_{adj.}$ is square of adjusted correlation coefficient, F-test is a variance-related static which compares two models differing by one or more variables to see if the more complex model is more reliable than the less complex one, the model is supposed to be good if the F-test is above a threshold value, LSE is least-square error, r is correlation coefficient, q^2 is the square of the correlation coefficient of the cross validation, BSr^2 is Bootstrapping correlation coefficient, r^2_{pred} is the predicted correlation coefficient calculated from the predicted activity of the test set compounds.

The four descriptors selected by GFA to develop the model, belong to four different descriptor classes. S_sssN is an elec-

Table 5
Observed and predicted activity of training set compounds

Compounds	Observed pIC_{50}	Predicted pIC_{50}	Residual	Compounds	Observed pIC_{50}	Predicted pIC_{50}	Residual
1	1.699	1.461	0.238	16	1.161	1.406	–0.245
2	0.983	0.787	0.196	17	1.402	1.998	–0.596
4	0.523	0698	–0.175	19	1.762	1.490	0.272
5	0.949	0.762	0.187	20	1.678	1.547	0.131
6	2.018	1.384	0.634	22	2.114	1.626	0.488
8	1.000	1.187	–0.187	23	0.830	1.410	–0.580
9	1.301	1.110	0.191	24	0.879	1.061	–0.182
10	0.116	0.115	0.001	25	0.701	0.835	–0.134
11	1.254	1.278	–0.024	26	2.119	1.940	0.179
13	1.301	1.291	0.010	27	0.859	1.153	–0.294
14	1.602	1.458	0.144	28	2.097	2.103	–0.006
15	1.062	1.315	–0.253	29	2.06	2.054	0.006

Table 6
Observed and predicted activity of test set compounds

Compounds	Observed pIC_{50}	Predicted pIC_{50}	Residual
3	0.217	0.758	–0.541
7	1.263	1.316	–0.053
12	1.633	1.416	0.217
18	1.951	2.002	–0.051
21	1.585	1.522	0.063
30	1.523	1.959	–0.436

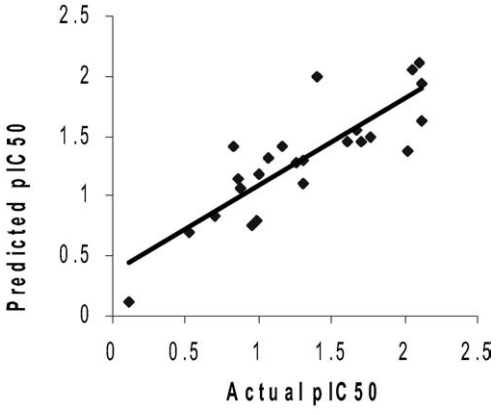


Fig. 2. Correlation between observed and predicted activity of the training set. The experimental activity of compounds in training set was plotted against the activity predicted by the model.

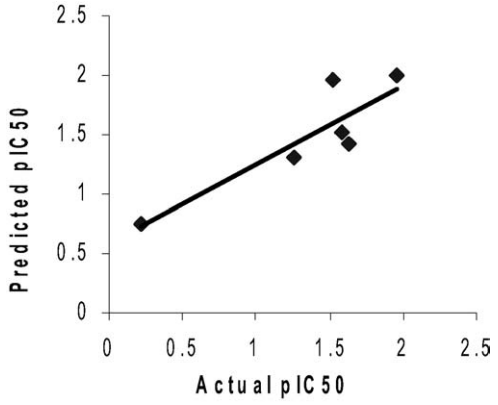


Fig. 3. Correlation between observed and predicted activity of the test set. The experimental activity of compounds in test set was plotted against the activity predicted by the model.

trotopological state index. The E-state indices encode not only the information about the topological environment of an atom, but also the electronic interactions from other atoms in the molecule. Thus, E-state is able to provide useful information on structure features that mostly relate to the property to be modeled [30]. Symbol S_sssN indicates the electrotopological environment of a nitrogen atom which is connected by three single bonds. In case, when there are more than one such nitrogen atoms in molecule, the descriptor indicates the summation for various such atoms. Thus it is clear that the nitrogen atoms and the electronic environment surrounding them are critical for inhibitory activity. CHI-2 is Kier and Hall molecular connectivity index of order 2. The connectivity indices belong to topological class of descriptors and are single valued parameters that can be calculated from the 2D graph representation of molecules. They characterize structures according to their size, degree of branching and overall shape [31]. Shadow_XZ is spatial descriptor calculated by projecting the molecular surface on XZ plane. These descriptors help to characterize the shape of molecules. Atype_C_8 is atom type AlogP descriptor. Various atom type AlogP descriptors are used to calculate the logP of molecule that characterize the lipophilicity of molecule.

The inter correlation of the descriptors used was checked and is provided in the form of correlation matrix (Table 4). The developed model was used to predict the activity of test set compounds and the predicted activity for training and test set compounds are given in Tables 5 and 6, respectively. The correlation of predicted activity to the observed activity is shown in Figs. 2 and 3 for training and test set, respectively. Further validation of the developed model was done by randomization test. The test was done by repeatedly permuting the activity values of the data set and using the permuted values to generate QSAR models and then comparing the resulting scores with the score of the original QSAR model generated from non-randomized activity values. If the original QSAR model is statistically significant, its score should be significantly better than those from permuted data [32]. The randomization test was performed at 90%, 95%, 98% and 99% confidence interval. The higher the confidence level, the more randomization tests are run. For a 90% confidence level, nine trials are run, 19 trials at 95%, 49 trials at 98% and 99 trials at 99%. The r value of the original model was much higher than

Table 7
Results of randomization test performed to check the validation of model

Confidence level	90%	95%	98%	99%
Total trials	9	19	48	99
r from non-random	0.848	0.848	0.848	0.848
Random r's > non-random	0	0	0	0
Random r's < non-random	9	19	48	99
Mean value of r from random trial	0.326	0.391	0.406	0.407
Standard deviation of random trials	0.112	0.113	0.128	0.132
Standard deviation from non-random r to mean	4.657	4.064	3.464	3.334

any of the trials using permuted data. Hence, the model is statistically significant and robust. The results of randomization test at various confidence levels are shown in Table 7.

So it is clear that the model satisfy all the validation criteria i.e. leave one out cross validation, randomization test and external set prediction. Good results are obtained in each of the validation technique. From Table 3, it is also clear that electrotopological state indices S_sssN is most important descriptor for the neuropeptide Y Y5 receptor inhibitory activity as it is present in all the equations. So it can give an idea about the activity of the compounds. In each of the equation, the coefficient associated with S_sssN is negative, which shows that equations are valid and with increase in the value of S_sssN, there is a decrease in the activity. Since E-state indices are the measure of the availability of the π and/or lone pair electrons on the atoms, it indicates that nitrogen might be playing important role in the interaction of the compounds with the receptor. Nitrogen can be assumed to be involved in some charge-transfer phenomenon with the receptor where the stability of the charge-transfer complex formed will depend upon the value of S_sssN, that is, on the availability of π or the lone pair electrons on the nitrogen. Further, the descriptors Shadow_XZ and CHI-2 show that the overall shape, size and branching in the molecule is critical for activity and should be considered during drug design.

4. Conclusion

The QSAR model of neuropeptide Y Y5 receptor inhibitory activity have been developed based on topological, E-state, spatial and thermodynamic descriptors to estimate and predict relative antagonistic activity of 30 benzoxazinone derivatives. The predictive ability of model was demonstrated by using leave one out cross validation technique, randomization test as well as external test set prediction. The results presented above show that these descriptors can be used to describe the structure activity relationship of neuropeptide Y Y5 antagonists and its performance based on statistical parameters is satisfying.

References

[1] S. Nammi, S. Koka, K.M. Chinnala, K.M. Boini, Nutr. J. 3 (2004) 3.
[2] H.E. Bays, Obes. Res. 12 (2004) 1197–1211.
[3] S. Fisekovic, Bosn. J. Basic Med. Sci. 5 (2005) 30–34.
[4] B.M. Maxwell, MedGenMed 7 (2005) 77.
[5] D.R. Gehlert, P.A. Hipskind, Expert Opin. Investig. Drugs 6 (1997) 1827–1838.
[6] Y.S. Allen, T.E. Adrian, J.M. Allen, K. Tatemoto, T.J. Crow, S.R. Bloom, J.M. Polak, Science 221 (1983) 877–879.
[7] J.M. Lundberg, A. Franco-Cereceda, J.S. Lacroix, J. Pernow, N. Ann, Y Acad Sci. 611 (1990) 166–174.
[8] S. Dryden, H. Frankish, Q. Wang, G. Williams, Eur. J. Clin. Invest. 24 (1994) 293–308.
[9] J. Pons, E.W. Lee, L. Li, J. Kitlinska, Curr. Opin. Investig. Drugs 5 (2004) 957–962.
[10] M. Heilig, Neuropeptides 38 (2004) 213–224.
[11] S.C. Baraban, Neuropeptides 38 (2004) 261–265.
[12] D.A. Groneberg, G. Folkerts, C. Peiser, K.F. Chung, A. Fischer, Pulm. Pharmacol. Ther. 17 (2004) 173–180.
[13] S. Bedoui, S. Miyake, R.H. Straub, S. von Horsten, T. Yamamura, Trends Immunol. 25 (2004) 508–512.
[14] C. Cabrele, M. Langer, R. Bader, H.A. Wieland, H.N. Doods, O. Zerbe, A.G. Beck-Sickinger, J. Biol. Chem. 275 (2000) 36043–36048.
[15] B.G. Stanley, S.F. Leibowitz, Proc. Natl. Acad. Sci. USA 82 (1985) 3940–3943.
[16] D.C. Jimerson, B.E. Wolfe, CNS Spectr. 9 (2004) 516–522.
[17] A. Lecklin, I. Lundell, L. Paananen, J.E. Wikberg, P.T. Mannisto, D. Larhammar, J. Br. Pharmacol. 135 (2002) 2029–2037.
[18] L. Juanenea, S. Galiano, O. Erviti, A. Moreno, S. Perez, I. Aldana, A. Monge, Bioorg. Med. Chem. 12 (2004) 4717–4723.
[19] T.J. Hou, J.M. Wang, N. Liao, X.J. Xu, J. Chem. Inf. Comput. Sci. 39 (1999) 775–781.
[20] S. Oloff, R.B. Mailman, A. Tropsha, J. Med. Chem. 48 (2005) 7322–7332.
[21] J.L. Medina-Franco, A. Golbraikh, S. Oloff, R. Castillo, A. Tropsha, J. Comput. Aided Mol. Des. 19 (2005) 229–242.
[22] M. Shen, C. Beguin, A. Golbraikh, J.P. Stables, H. Kohn, A. Tropsha, J. Med. Chem. 47 (2004) 2356–2364.
[23] X.C. Fu, G.P. Wang, W.Q. Liang, Q.S. Yu, Pharmazie 59 (2004) 126–130.
[24] A. Torrens, J. Mas, A. Port, J.A. Castrillo, O. Sanfeliu, X. Guitart, A. Dordal, G. Romero, M.A. Fisas, E. Sanchez, E. Hernandez, P. Perez, R. Perez, H. Buschmann, J. Med. Chem. 48 (2005) 2080–2092.
[25] Sybyl 6.9 Tripos Associates Inc., 1699 S Hanley Rd., St. Louis, MO 631444, USA.
[26] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.P. Stewart, J. Ame. Chem. Soc. 107 (1985) 3902–3909.
[27] Cerius2 version 4.10, Accelrys Inc., 6985 Scranton Road, San Diego, CA, USA.
[28] D. Rogers, A.J. Hopfinger, J. Chem. Info. Comput. Sci. 34 (1994) 854–866.
[29] J.G. Topliss, R.P. Edwards, J. Med. Chem. 22 (1979) 1238–1244.
[30] K. Rose, L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 42 (2002) 651–666.
[31] J. Liu, L. Yang, Y. Li, D. Pan, A.J. Hopfinger, J. Comput. Aided Mol. Des. 19 (2005) 567–583.
[32] K. Roy, J.T. Leonard, J Chem. Inf. Model. 45 (2005) 1352–1368.