ORIGINAL PAPER

Molecular modelling and QSAR analysis of some structurally diverse N-type calcium channel blockers

Jignesh Mungalpara · Ashish Pandey · Vaibhav Jain ·
C. Gopi Mohan

Abstract A quantitative structure–activity relationship (QSAR) analysis was performed on a data set of 104 molecules showing N-type calcium channel blocking activity. Several types of descriptors, including electro-topological, structural, thermodynamics and ADMET, were used to derive a quantitative relationship between N-type calcium channel blocking activity and structural properties. The genetic algorithm-based genetic function approximation (GFA) method of variable selection was used to generate the 2D-QSAR model. The model was established on a training set of 83 molecules, and validated by a test set of 21 molecules. The model was developed using five information-rich descriptors—Atype_C_24, Atype_N_68, Rotlbonds, S_sssN, and ADME_Solubility—playing an important role in determining N-type calcium channel blocking activity. For the best QSAR model (model 4), the statistics were $r^2=0.798$; $q^2=0.769$; $n=83$ for the training set. This model was further validated using the leave-one-out (LOO) cross-validation approach, Fischer statistics (F), Y-randomisation test, and predictions based on the test data set. The resulting descriptors produced by QSAR model 4 were used to identify physico-chemical features relevant to N-type calcium channel blocking activity.

J. Mungalpara · A. Pandey · V. Jain · C. G. Mohan (✉)
Centre for Pharmacoinformatics,
National Institute of Pharmaceutical Education
and Research (NIPER), Sector 67, S.A.S.,
Nagar 160 062 Punjab, India
e-mail: cmohan@niper.ac.in
e-mail: cgopimohan@yahoo.com

Introduction

Free intracellular Ca^{2+} is an essential element for life and is the most common signal transduction element in cells [1]. An electrochemical gradient exists between extracellular and intracellular calcium concentrations. Calcium enters the cytosol either through plasma membrane voltage-gated calcium channels (VGCCs) or is released from intracellular pools. Under normal physiological conditions, ion channels permit the orderly movement of ions across both plasma and intracellular cell membranes. A number of states, as well as cell death, also occur under pathological conditions in which the disorderly movement of ions through these same channels dominates. The aberrant elevation of intracellular Ca^{2+} levels through altered calcium channel function is related to a variety of serious human pathophysiological conditions, including cardiovascular diseases, muscle disorders, acute or chronic pain, epilepsy, cerebellar ataxia, migraine, mood disorders, and certain types of cancer [2].

Based on their electrophysiological nature, channels are classified into high voltage-activated (HVA) and low voltage-activated (LVA) channels [3, 4]. Of the HVA channels, the L-, N-, P-/Q-, and R-types require a higher depolarisation current for activation. The T-type of LVA channel requires lower depolarisation current to become activated. These channels mediate the influx of Ca^{2+} in response to membrane depolarisation, and play key roles by regulating calcium influx to intracellular medium in response to stimuli [5]. VGCCs differ in function, conductance, activation/inactivation voltage, and sensitivity towards various drugs and toxins.

N-type calcium channels are associated with central and peripheral neurons, being located on pre-synaptic nerve terminals. These channels regulate the calcium flux sub-serving depolarisation-evoked release of neurotransmitter from pre-synaptic endings. At the pre-synaptic nerve terminal, VGCCs open in response to action potentials to allow an influx of calcium ions. The influx is a graded process varying in a linear manner with the frequency of action potentials. These in turn lead to release of various neurotransmitters that diffuse across the synaptic cleft to the postsynaptic membrane and bind to their specific receptors.

It has been suggested that the selective N-type calcium channel blockers (N-type CCBs) could be used as thera-peutic agents for the treatment of pain [6]. Blockage of N-type calcium channels has recently been invoked for treatment of the chronic pain associated with cancer, AIDS and neuropathy. The drug morphine is used for the treatment of chronic pain, but it is associated with serious side effects that limit its long-term use [7]. Omega-conotoxin MVIIA, a 25-amino-acid peptide found in the venom of the piscivorous marine snail (*Conus magus*), is a potent and selective N-type CCBs [8]. SNX-111 is a synthetic version of omega-conotoxin MVIIA (trade name Prialt®), which has recently demonstrated efficacy in animal models of traumatic brain injury, focal cerebral ischaemia, and pain. The United States Food and Drug Administration (FDA) approved SNX-111 (also named ziconotide), an intrathecally administered peptide acting as a novel non-opioid for the treatment of severe chronic pain [6, 9, 10]. The *anti*-hypertensive drug cilnidipine is a 1,4-dihydropyridine-derived long-acting dual CCB, which inhibits both L-type and N-type calcium channels [11].

The group of Yamamoto et al. have recently documented a novel series of 1,4-dihydropyridines and cyproheptadines as potent and selective N-type CCBs [12, 13]. This group also performed a structure–activity relationship (SAR) study on the 2-, 5-, and 6-position of 1,4-dihydropyridine derivatives to identify selective and injectable N-type CCBs with high aqueous solubility [14]. A structure–activity study of L-cysteine-based N-type CCBs carried out by Seko et al. [15] established the selectivity towards this channel over L-type channels. Lars et al. [16] also reported recently the synthesis and SAR of novel 2-arylthiazolidinones as selec-tive N-type CCBs.

In silico techniques have proven their usefulness in pharmaceutical research for the selection/identification and/or design/optimisation of new chemical entities, to trans-form early-stage drug discovery, particularly in terms of time- and cost-savings. Quantitative structure–activity relationship (QSAR) is one of the most important areas in chemometrics, and is a valuable tool that is used exten-sively in drug design and medicinal chemistry [17–19]. Chemical and biological effects are related closely to molecular physico-chemical properties, which can be calculated or predicted from their structure using various kinds of methods [20]. Once a reliable QSAR model is established, we can predict the activities of molecules, and know which structural features play an important role in the biological process. The advances in QSAR studies have widened the scope of rational drug design as well as the search for the mechanisms of drug actions.

Many different methodologies, such as multiple linear regression (MLR), partial least squares (PLS), heuristic method (HM), and different types of artificial neural networks (ANN), can be applied for QSAR development. Recently, genetic function approximation (GFA) has gained great popularity in QSAR research. The GFA method, developed by Rogers and Hopfinger [21], is employed in a statistical analysis to select the relevant descriptors and to generate different QSAR models. Sensitivity analysis of QSAR models is then performed and the best model developed can be used for predicting test set molecules that were not included in the training set molecules. Randomisation tests performed on the model at various intervals of confidence levels ensure its proper validation.

In silico modelling techniques offer an important approach towards predicting potential CCBs during the early stages of drug development [22, 23]. The purpose of the present work was to determine the biological activity of different scaffolds of non-peptidyl derivatives acting as N-type CCBs by creating a robust QSAR model. The developed model should be able to identify and describe important physico-chemical features of the molecules that underpin variations in molecular activity.
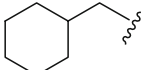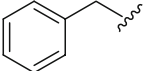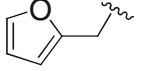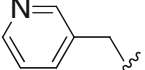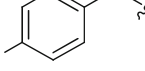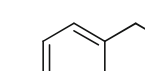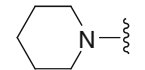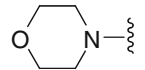
Materials and methods
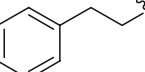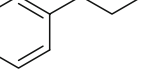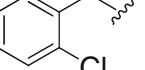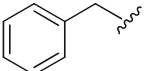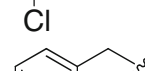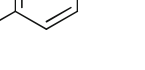
Data set

The data set of 104 molecules used for the 2D-QSAR analyses was collected from the literature [24–29]. The N-type calcium channel blocking potency (IC_{50}) values for the data set molecules vary from 0.04 to 120 μM (a factor of about 12,000), and were measured under the same experimental conditions using a fluorescence-based Ca^{2+} flux assay in IMR32 human neuroblastoma cells [24–29]. The structures and N-type CCB activity (IC_{50}) data of these 104 molecules are presented in Tables 1 and 2.
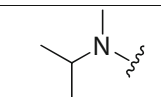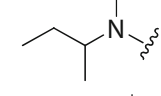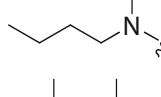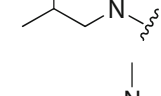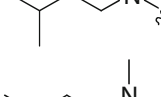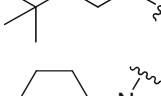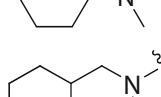
The IC_{50} values in the micromolar (μM) range were converted to the molar (M) range and then to its logarithmic scale (pIC_{50}, M), to reduce the skewness of the data set, which was then used for subsequent QSAR analysis (Eq. 1) as the response variable.

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

Table 1 Structure, and actual and predicted activity of N-type calcium channel blockers (CCBs)

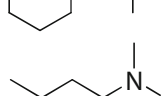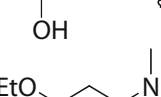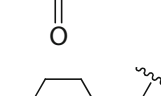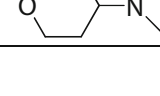| Molecule number	R	Scaffold	[a]IC_{50}(µM)	pIC$_{50}$ Actual	Predicted	Residual
1		A	0.68	6.168	6.326	-0.158
2		A	0.36	6.444	6.300	0.144
3		A	1.60	5.796	5.998	-0.202
4		A	3.60	5.444	5.840	-0.396
5[a]		A	1.70	5.770	6.427	-0.657
6[a]		A	1.90	5.721	6.593	-0.872
7		A	0.56	6.252	6.539	-0.287
8[a]		A	1.00	6.000	6.534	-0.534
9		A	0.23	6.638	6.528	0.110
10		A	0.28	6.553	6.399	0.159
11		A	0.50	6.301	6.304	0.004
12		A	1.00	6.000	6.184	-0.184
13[a]	-NH$_2$	B	3.80	5.420	5.510	-0.091
14	-NHMe	B	2.30	5.638	5.677	-0.039
15	-N-(Me)$_2$	B	2.30	5.638	5.503	0.135
16		B	1.40	5.854	5.836	0.018
17		B	6.40	5.194	5.461	-0.267

Table 1 (continued)

18		B	2.80	5.553	5.822	-0.269
19		B	1.50	5.824	6.058	-0.234
20		B	0.28	6.553	6.130	0.423
21		B	0.38	6.420	6.029	0.391
22[a]		B	0.32	6.495	6.201	0.294
23		B	0.40	6.398	6.274	0.124
24[a]		B	0.20	6.699	6.194	0.505
25		B	0.70	6.155	6.326	-0.171
26		B	2.20	5.658	5.642	0.016
27		B	4.40	5.357	5.807	-0.450
28		B	3.10	5.509	5.543	-0.034

It is essential to assess the predictive power of QSAR models by using a test set of molecules according to the following criteria: (1) the biological activity values of the test set should span the training set several times, but should not exceed the training set by more than 10%; (2) the biological assay methods for both the training set and test set should be the same or comparable; (3) the test set should represent a balanced number of both active and inactive molecules for uniform sampling of the data set. The remaining molecules are taken as the training set for the purpose of QSAR model building.

Molecular modelling

The molecules under study were built using the SYBYL7.1 [30] molecular modelling package installed on a Silicon Graphics Fuel Work station running IRIX 6.5. Since no crystal structure of an N-type calcium channel in complex with any molecule is available, the basic skeleton and conformation of the most active molecule, **53**, having an IC_{50} value of 0.04 μm, was selected, to which Gasteiger-Hückel partial atomic charges [31] were applied and energy minimised by Powell's [32] method using Tripos force field [33] with 0.001 kcal mol^{-1} energy gradient convergence criterion. The remaining molecules were built by making the required substitution on template molecule **53** to which charges were applied and energy minimised as stated previously. These molecules were then used to construct the 2D-QSAR model.

Descriptor calculation and selection

Descriptors were calculated using the Cerius^2 software package [34] which includes: electronic descriptors,

Table 2 Structure, and actual and predicted activity of N-type CCBs
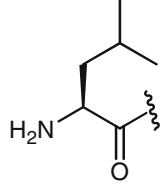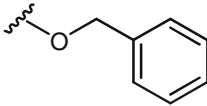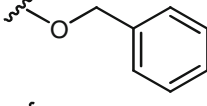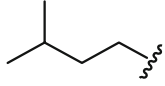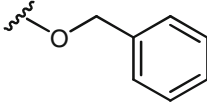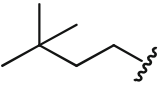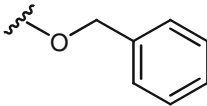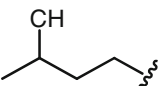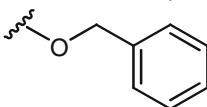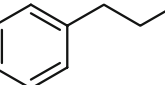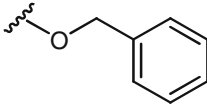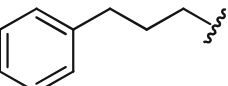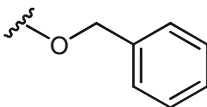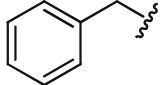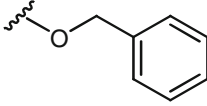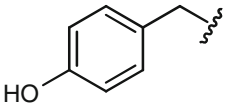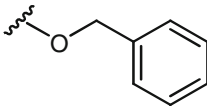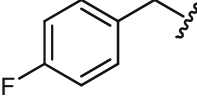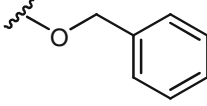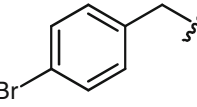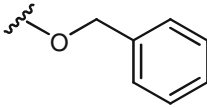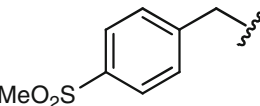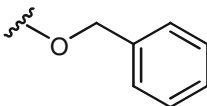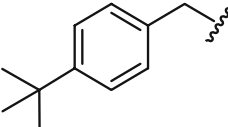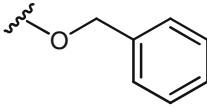
Molecule number	R	X	Scaffold	[a]IC$_{50}$ (μM)	pIC$_{50}$ (M) Actual	pIC$_{50}$ (M) Predicted	Residual
29	*(structure)*	*(structure)*	C	0.67	6.174	5.979	0.196
30	H	*(structure)*	C	4.70	5.328	5.369	-0.041
31	*(structure)*	*(structure)*	C	1.60	5.796	6.067	-0.271
32	*(structure)*	*(structure)*	C	0.60	6.222	6.154	0.068
33	*(structure)*	*(structure)*	C	3.00	5.523	5.370	0.153
34	*(structure)*	*(structure)*	C	0.30	6.523	6.342	0.181
35	*(structure)*	*(structure)*	C	0.50	6.301	6.530	-0.229
36	*(structure)*	*(structure)*	C	0.72	6.143	6.195	-0.052
37	*(structure)*	*(structure)*	C	0.70	6.155	5.959	0.196
38[a]	*(structure)*	*(structure)*	C	0.83	6.081	6.201	-0.120
39	*(structure)*	*(structure)*	C	0.48	6.319	6.403	-0.084
40	*(structure)*	*(structure)*	C	1.80	5.745	5.964	-0.219
41	*(structure)*	*(structure)*	C	0.38	6.420	6.642	-0.222

Table 2 (continued)
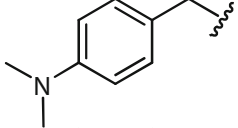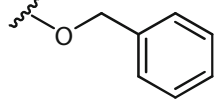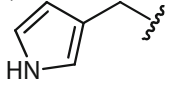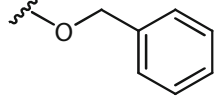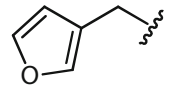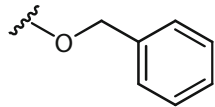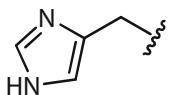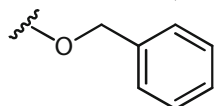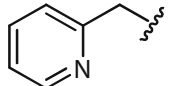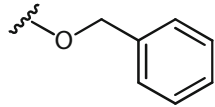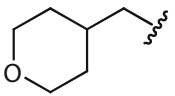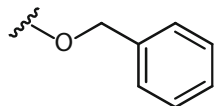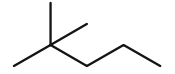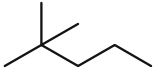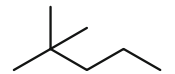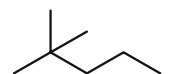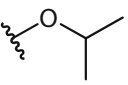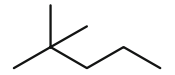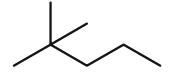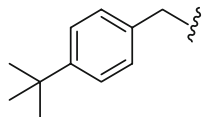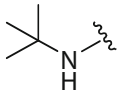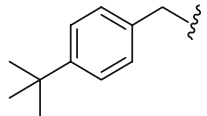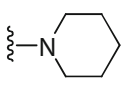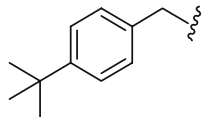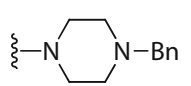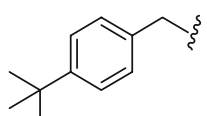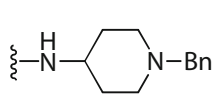
42			C	0.33	6.481	6.355	0.126
43[a]			C	0.58	6.237	5.720	0.517
44[a]			C	3.90	5.409	5.742	-0.333
45			C	3.80	5.420	5.359	0.061
46[a]			C	1.60	5.796	5.779	0.017
47			C	4.80	5.319	5.351	-0.032
48			C	0.90	6.046	6.259	-0.213
49			C	1.30	5.886	5.721	0.165
50			C	2.90	5.538	5.608	-0.070
51		F	C	3.90	5.409	5.267	0.142
52		H	C	7.80	5.108	5.229	-0.121
53[a]			D	0.04	7.398	6.767	0.631
54			D	0.09	7.046	6.728	0.318
55			D	0.50	6.301	6.594	-0.293
56			D	0.27	6.569	6.467	0.102

Table 2 (continued)
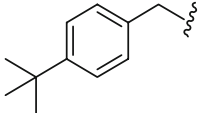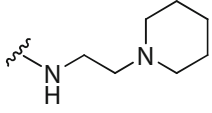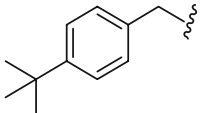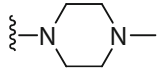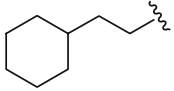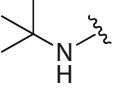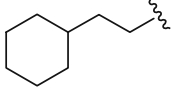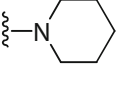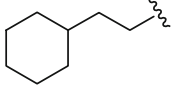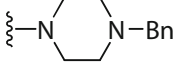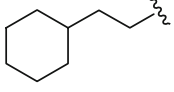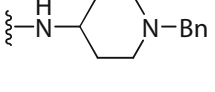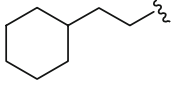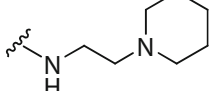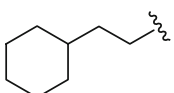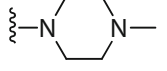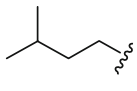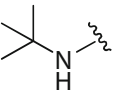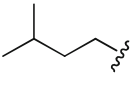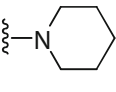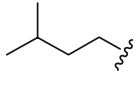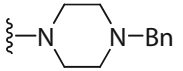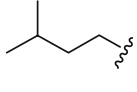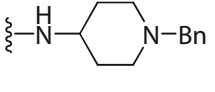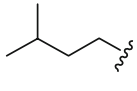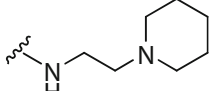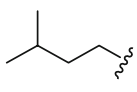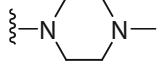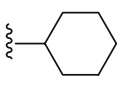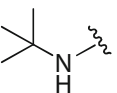
57			D	0.31	6.509	6.384	0.125
58			D	0.44	6.357	6.084	0.237
59			D	0.50	6.301	6.326	-0.025
60			D	0.32	6.495	6.395	0.100
61			D	0.42	6.377	6.367	0.010
62			D	0.31	6.509	6.259	0.250
63			D	1.00	6.000	6.098	-0.098
64[a]			D	1.20	5.921	5.765	0.156
65			D	0.32	6.495	6.201	0.294
66			D	0.66	6.180	6.274	-0.094
67			D	0.59	6.229	6.290	-0.061
68[a]			D	0.88	6.056	6.192	-0.136
69			D	1.10	5.959	5.997	-0.038
70			D	1.20	5.921	5.649	0.273
71[a]			D	0.20	6.699	6.194	0.505

Table 2 (continued)
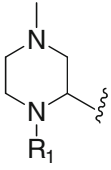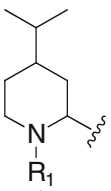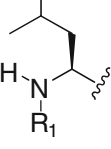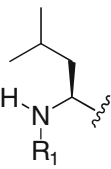
No.	R	R'					
72	cyclohexyl	piperidin-1-yl (–N-piperidine)	D	0.38	6.420	6.265	0.155
73[a]	cyclohexyl	4-benzylpiperazin-1-yl (–N–piperazine–N–Bn)	D	0.76	6.119	6.264	-0.145
74	cyclohexyl	–NH–(1-benzylpiperidin-4-yl) (–N–H–piperidine–N–Bn)	D	1.30	5.886	6.162	-0.276
75	cyclohexyl	–NH–CH₂CH₂–(piperidin-1-yl)	D	0.88	6.056	5.981	0.075
76	cyclohexyl	4-methylpiperazin-1-yl (–N–piperazine–N–)	D	5.60	5.252	5.638	-0.386
77	Me	–NH–C(CH₃)₃ (tert-butylamino)	D	2.80	5.553	5.822	-0.269
78	Me	piperidin-1-yl (–N-piperidine)	D	1.90	5.721	5.584	-0.137
79	Me	4-benzylpiperazin-1-yl (–N–piperazine–N–Bn)	D	1.20	5.921	5.607	0.236
80	Me	–NH–(1-benzylpiperidin-4-yl) (–N–H–piperidine–N–Bn)	D	2.60	5.585	5.607	-0.022
81[a]	Me	–NH–CH₂CH₂–(piperidin-1-yl)	D	4.10	5.387	5.347	0.040
82	Me	4-methylpiperazin-1-yl (–N–piperazine–N–)	D	13.00	4.886	4.968	0.082
83	–NH–CH₂–(pyridin-2-yl)	b	E	1.10	5.959	6.265	-0.306
84	–NH–CH₂–cyclohexyl	b	E	0.29	6.537	6.627	-0.090
85	–NH–CH₂CH₂–C(CH₃)₃	b	E	0.18	6.745	6.603	0.141
86	–NH–CH₂CH₂–CH(OH)CH₃	b	E	1.80	5.745	5.937	-0.192

Table 2 (continued)

87[a]		b	E	120.00	3.921	5.887	-1.966
88[a]		a	E	0.67	6.174	6.876	-0.702
89	H	a	E	1.00	6.000	5.946	0.053
90		b	E	0.19	6.721	6.723	-0.002
91	OH	b	E	1.10	5.959	5.865	0.094
92		NH	F	0.32	6.495	6.201	0.294
93		NH	F	0.37	6.432	6.201	0.231
94		NH	F	0.83	6.081	6.018	0.062
95		NH	F	1.60	5.796	5.478	0.317
96		O	F	1.00	6.000	6.030	-0.030
97[a]		O	F	1.50	5.824	5.974	-0.150
98		O	F	0.32	6.495	6.160	0.334
99		O	F	1.10	5.959	6.142	-0.183
100		O	F	0.78	6.108	6.112	-0.004

Table 2 (continued)

101[a]		O	F	1.00	6.000	5.414	0.585
102		O	F	0.49	6.310	6.392	-0.082
103[a]		NH	F	0.30	6.523	6.676	-0.153
104		O	F	0.40	6.398	6.416	-0.018

Where, R₁= A, B, C, D, E, F and a, b are scaffold from Fig. 1

[a] Indicate molecule taken in the test set

[b] Experimental inhibitory activity of N-type CCBs is from Ref [24–29] and is converted to pIC_{50} using Eq. 1

topological descriptors, information content descriptors, spatial (shadow indices) descriptors, structural descriptors, thermodynamics descriptors, and ADMET descriptors. Before commencing with the development of the QSAR model, the correlation matrix of about 250 descriptors was calculated and highly correlated descriptors, with correlation values above 0.7, were removed. Auto scaling (or unit scale variance) of descriptors was carried out to avoid any bias due to strongly diverging descriptor values. Scaling is also referred to as standardising or normalising descriptors to ensure that they have equal weight (in terms of magnitude) in subsequent analysis [35, 36]. This is a very sensitive procedure because, in the descriptor pool, for most cases we do not know the underlying relationship between the descriptor and the activity and therefore cannot foresee the influence of this scaling process. Furthermore, descriptors with constant values as well as those with poor correlation to the biological property were discarded; some descriptors having zero value were also discarded. Finally, three class of descriptors: electrotopological, thermodynamic and structural, were considered for statistical fitting using the GFA method (Table 3).

Regression analysis

GFA is a genetics-based method of variable selection that combines Holland's genetic algorithm (GA) with Friedman's multivariate adaptive regression splines (MARS) [37]. It works in the following way: first of all a particular number of equations (set at 100 by default in Cerius² software) is generated randomly, then pairs of "parent" equations are chosen randomly from this set of 100 equations and "crossover" operations are performed at random. After some preliminary observations on initial runs, the number of GFA crossovers was set at 5,000 for the present study in order to obtain reasonable convergence. The goodness-of-fit of each progeny equation is then assessed by Friedman's lack of fit (LOF) score, assigned by GFA, which resists over fitting and estimates appropriate number of variables in the equation, and is given by the following formula:

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

Where LSE is the least-squares error, c is the number of basis functions in the model, d is the smoothing parameter, p is the number of descriptors, and m is the number of

Table 3 Equation of quantitative structure–activity relationship (QSAR) models with increasing number of descriptors

Number of descriptors/QSAR model	r^2	QSAR equation
2 (Model 1)	0.584	$pIC_{50} = 6.01117 - 0.26186 * ″ADME\_Solubility″ + 0.265085 * ″Kappa - 1″$
3 (Model 2)	0.669	$pIC_{50} = 6.02994 - 0.201676 * ″Hbondacceptor″ - 0.282349 * ″ADME\_Solubility″ + 0.423022 * ″Rotlbonds″$
4 (Model 3)	0.739	$pIC_{50} = 6.00961 - 0.332063 * ″ADME\_Solubility″ - 0.20311 * ″Atype\_N\_68″ + 0.549302 * ″Kappa - 1″$ $- 0.208374 * ″CHI - V - 2″$
5 (Model 4)	0.798	$pIC_{50} = 6.04274 - 0.325494 * ″ADME\_Solubility″ + 0.118937 * ″Atype\_C\_24″ + 0.369743 * ″Rotlbonds″$ $- 0.266013 * ″Atype\_N\_68″ + 0.125834 * ″S\_sssN″$
6 (Model 5)	0.779	$pIC_{50} = 6.04191 + 0.295619 * ″Rotlbonds″ - 0.19678 * ″Jurs - DPSA - 2″ - 0.282395 * ″SC - 0″$ $- 0.269756 * ″ADME\_Solubility″ - 0.229987 * ″Atype\_N\_68″ + 0.580605 * ″MR″$

observations in the training set. The smoothing parameter, which controls the scoring bias between equations of different sizes, was set at a 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 statistical parameters such as regression coefficient, adjusted regression coefficient, regression coefficient cross validation and F-test values.

QSAR model validation

The QSAR model developed was evaluated rigorously using Y-randomisation test [38] and test set predictions. Y-randomisation test confirms whether the model is obtained by chance correlation, and is a true structure–activity relationship to validate the adequacy of the training set molecules. The steps followed in the randomisation test are (1) repeatedly scrambling the activity data in the training set molecules, (2) using the randomised data to generate QSAR equations, and (3) comparing the resulting scores with the score of the original QSAR equation generated with non-randomised data. If the activity prediction of the random model is comparable to that of the original equation, the set of observations is not sufficient to support the model. The randomisation test was performed at different confidence intervals (90%, 95%, 98% and 99%). More randomisation tests are run for higher confidence levels. For a 90% confidence level, there are 9 trials run, 19 trials for 95%, 49 trials for 98% and 99 trials for 99%. The correlation coefficient (r) value of the original model was much higher than any of the trials using permuted data, hence showing that the model developed is statistically significant and robust. The inter-correlation of the descriptors used in the final model was checked and the descriptors were found to be reasonably orthogonal. The predictive properties of the developed QSAR model were tested more rigorously by predicting the N-type channel blocking potency of 21 test set molecules.

Results and discussion

A 2D-QSAR analysis was performed to explore the structure–activity relationship of different scaffolds of non-peptidyl derivatives acting as N-type CCBs. These channel blockers possess a variety of scaffolds (**A**, **B**, **C**, **D**, **E**, and **F**), di-substituted with different R and X groups, as presented in Fig. 1 and Tables 1 and 2. The data set of 104 molecules was divided into a training set of 83 molecules and a test set of 21 molecules taking into account structural diversity and activity range. The training set molecules was then used to generate the 2D-QSAR models, while the test set molecules were selected for model validation.

A large pool of descriptors are available to build initial QSAR models but, for a good QSAR model, a minimal set of information-rich descriptors that can shed light on the structural features important for the biological activity of the training set is required. Including too many descriptors in the model, even if they contain relevant information, can result in over fitting of the model, with the resultant loss of ability of the model to generalise unseen molecules. Selection of descriptors is based on their correlation to biological activity and capability of producing MLR models with moderate correlation coefficients. A statistically meaningful QSAR model was constructed by checking the variation of different statistical parameters against the number of descriptors.

Initially, we were unable to develop a robust QSAR model for this data set. The QSAR statistics obtained were not significant, showing very low squared correlation coefficient and cross-validated correlation coefficient. We then adopted an auto-scaling procedure for the descriptors. This methodology led to a marked improvement in the QSAR statistics, and the model developed was more robust, as explained in detail below. Scaling of the descriptors has been carried out successfully by different research groups for a variety of QSAR case studies in order to improve their predictive power [39, 40].

Fig. 1 Different scaffolds of N-type calcium channel blockers (CCBs)
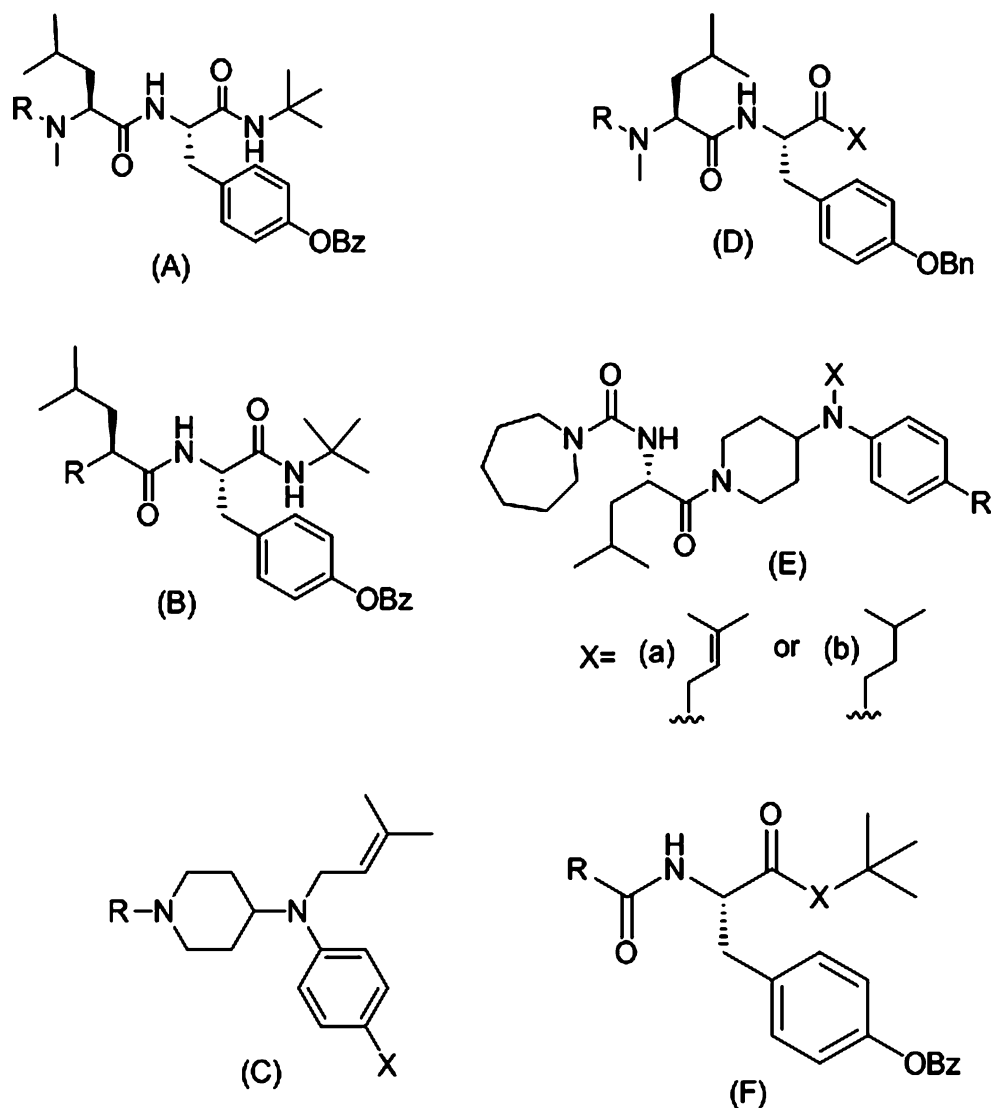
Sensitivity analysis of the 2D-QSAR model

Sensitivity of the QSAR models was assessed mainly using two statistical parameters: the cross-validated correlation coefficient (q^2), and Fischer ratio (F). The correlation coefficient (r^2) can be increased easily by the number of terms in the QSAR equation, so we took q^2 as the limiting factor for controlling the number of descriptors to be used in the model. In order to assess the sensitivity, five different QSAR models were created by increasing the number of descriptors stepwise from two to six. The QSAR equations along with the number of descriptors are presented in Table 3. Different statistical parameters used to validate these QSAR models are shown in Table S1 in the electronic supplementary material (ESM). The variation of r^2 and q^2 values with the number of descriptors is presented in ESM Fig. S1. The r^2 and q^2 values increase until the number of descriptors in the QSAR equation reached five. When the number of descriptors in the QSAR

equation was six (model 5), there was a decrease in q^2 value indicating the model sensitivity, as presented in Fig. S1 and Table S1. Thus, the number of descriptors was restricted to five and the QSAR equation obtained for model 4 is given below.

$$
\begin{aligned}
\text{pIC}_{50} &= 6.04274 - 0.325494 * ''\text{ADME\_Solubility}'' \\
&\quad + 0.118937 * ''\text{Atype\_C\_24}'' + 0.369743 * ''\text{Rotlbonds}'' \\
&\quad - 0.266013 * ''\text{Atype\_N\_68}'' + 0.125834 * ''\text{S\_sssN}'' \\
N &= 83; \text{LOF} = 0.048; r^2 = 0.798; r^2_{adj} = 0.785; F-\text{test} = 62.27; \\
&\quad \text{LSE} = 0.038; r = 0.893; q^2 = 0.769
\end{aligned}
\tag{3}
$$

Where N is the number of molecules in the training set, LOF is lack of fit score, r^2 is the squared correlation coefficient, r^2_{adj} is the square of adjusted correlation coefficient, F-test is a variance-related statistic that 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 assumed to be good if the F-test is above a threshold value, LSE is least square error,

r is correlation coefficient, and q^2 is the square of the correlation coefficient of the cross-validation.

Figure 2 presents scatter plots of the actual (experimental) pIC_{50} values for the training set (blue squares) and test set molecules (red squares) versus the calculated pIC_{50} values predicted by QSAR model 4.

Validation of 2D-QSAR model 4

Cross-validation of the QSAR model 4 by the leave one out (LOO) procedure showed a statistically significant correlation coefficient of $q^2=0.776$. Furthermore, the random-isation test, performed for 99 random trials (Table S2) showed a large difference of about 0.7 in the r values for non-random and random trials, with the observed deviation of random values being only 0.08, suggesting that the selected training set was adequate to support the developed QSAR model. Also, none of the random r values were greater than the non-random r values. Furthermore, the correlation matrix of the descriptors used in developing the QSAR equation of model 4 is significant (presented in Table S3).
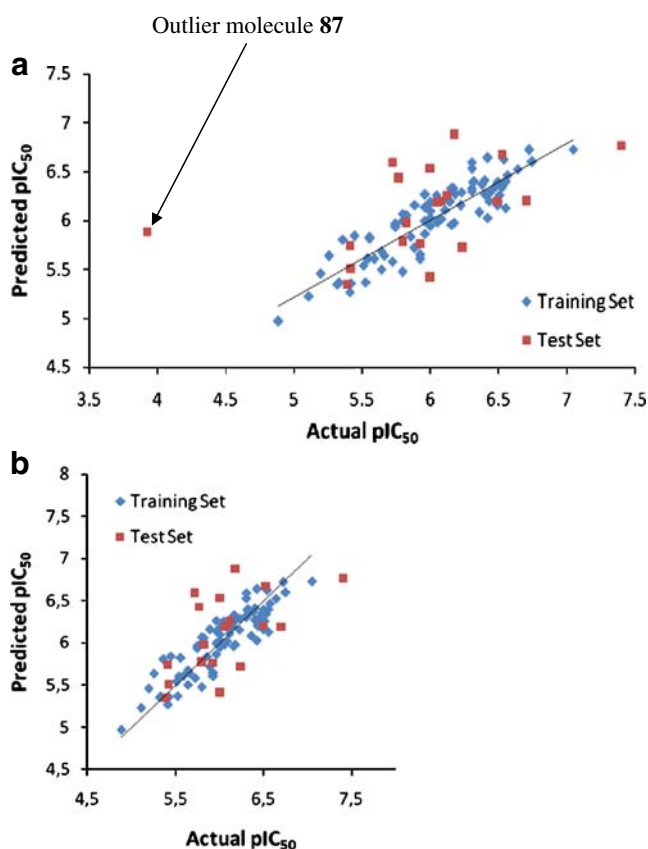
Fig. 2 Scatter plot of actual vs predicted activity of the training set (*blue squares*) and test set (*red squares*) molecules of quantitative structure–activity relationship (QSAR) model 4, either including (**a**) or excluding (**b**) molecule 87

In the test set prediction, the residual values for 20 of the 21 molecules were found to be within the acceptable range of 1. A scatter plot of the experimental and predicted pIC_{50} for the test set molecules is presented in Fig. 2. Overall, the developed 2D-QSAR model 4 is robust and was found to be statistically significant.

Importance of molecular descriptors

Five important molecular descriptors appeared in QSAR model 4: Rotlbonds, Atype_C_24, S_sssN, Atype_N_68, and ADME_Solubility. Among these, Rotlbonds, Atype_C_24, and S_sssN descriptors correlated positively with activity, while ADME_Solubility and Atype_N_68 descriptors correlated negatively with N-type CCB activity.

The type, definition and meaning of these five descriptors are described below along with their importance in understanding N-type CCB activity.

(1) Rotlbonds is a structural descriptor and indicates number of rotational bonds in a molecule. The molecular activity and the number of rotatable bonds in the studied molecules can be broadly correlated, and this descriptor showed positive correlation with N-type CCB activity in QSAR model 4. The 'Opera' rule needs to be addressed on the basis of the number of rotatable bonds, which in turn controls the structural flexibility of the molecule required for attaining bioactive conformation for drug-like molecules [41].

(2) Atype_C_24 is one of the atom type AlogP descriptors that appeared in QSAR model 4. In this descriptor, the C atom in the molecule is linked in R- -CR- -R manner, where R represents any group linked through carbon and "- -" represents aromatic bonds as in benzene or delocalised bonds like the N–O bond of a nitro group [42, 43]. Atype descriptors are thermodynamic descriptors defining the presence of that type of atom in the molecule. Various atom type AlogP descriptors can be used to calculate the logP of molecules. In this strategy, halogens and hydrogen are classified by the hybridisation and oxidation state of the carbon to which they are bonded; carbon atoms are classified by their hybridisation state and the chemical nature of their neighbouring atoms. In other words, the aromaticity associated with the C atom as part of the aromatic ring is favourable for CCB activity. Thus the positive slope of the Atype_C_24 descriptor in QSAR model 4 revealed that N-type CCB activity increases with an increase in hydrophobicity associated with this carbon.

(3) S_sssN is a thermodynamic descriptor and is defined as the summation of the electrotopological indices for all the N atoms present in a molecule that is connected

by three single bonds except hydrogen. Zhihua et al. [44] showed that electrotopological descriptors have both excellent structural selectivity and good activity estimation for peptide analogues. The numerical values of electrotopological descriptors computed for each atom in a molecule encode information about both the topological environment of that atom and electronic interactions due to all other atoms in the molecule. The topological relationship is based on the graph distance to each atom. The electronic aspect is based on the intrinsic state and perturbation due to the intrinsic state differences between atoms in the molecule [44, 45]. The developed QSAR model 4 revealed a positive correlation with the descriptor S_sssN, suggesting that the nature of the electronic environment in this molecule is necessary for potent N-type CCB activity.

The other two descriptors Atype_N_68 and ADME_Solubility selected by QSAR model 4 exhibited a negative correlation with the N-type CCB activity, as discussed in detail below.

(4) Atype_N_68 descriptor is of atom-type and is defined as N present in an Al_3N context in the molecule. Al represents aliphatic groups and the N atom linked with the three aliphatic groups in the molecule (tertiary nature). The negative slope of this descriptor in QSAR model 4 revealed that N-type CCB activity decreases with an increase in hydrophobicity associated with this nitrogen.

(5) ADME_Solubility belongs to the ADME set of descriptors, and is defined as the base 10 logarithm of the molar solubility of each molecule in water, predicted by linear regression methodology [46]. The presence of this descriptor in QSAR model 4 was correlated with a negative contribution towards N-type CCB activity.

All five above-mentioned descriptors provided insight into the physico-chemical requirements necessary for designing potent N-type CCBs.

The N-type CCB dataset has six different structurally diverse scaffolds, as presented in Fig. 1. Scaffold A has 12 molecules (**1–12**), with activity ranging from 0.23 to 3.6 μM. Scaffold B has 16 molecules (**13–28**), with activity ranging from 0.20 to 6.4 μM. Table 1 shows the structure and activity of these two different scaffold-containing molecules. Other series of molecules from the dataset include scaffold C, with 24 molecules (**29–52**), and activity ranging from 0.30 to 7.8 μM; scaffold D with 30 molecules (**53–82**), and activity ranging from 0.04 to 13.0 μM; scaffold E with 9 molecules (**83–91**), and activity ranging from 0.18 to 120.0 μM; and scaffold F with 13 molecules (**92–104**), and activity ranging from 0.30 to 1.60 μM, respectively. The

structure and activity of these latter four different scaffold-containing molecules is presented in Table 2.

The most potent molecules (**53** and **54**) are present in the scaffold D group. The contribution of the electronic environment of the descriptor S_sssN in these two molecules has similar chemical features, as reflected in its R- and X- substitutions. The presence of this descriptor is very important for the potent activity of these molecules, which was further confirmed by the positive slope in QSAR model 4. Further, the functional groups attached at the R region in molecules **53** and **54** have aromatic features, and do not completely follow the nature of the chemical environment of the Atype_N_68 descriptor. In other words, molecule **53** has aliphatic amine groups, which are secondary in nature and do not strictly follow the Al_3N manner. In support of this observation, this descriptor also showed a negative slope in QSAR model 4.

The scaffold D group also has the least potent molecule (**82**) with an IC_{50} value of 13 μM, which is 1,000-fold less potent than molecule **53**. The R-Me group substitution in molecule **82** is strictly in accordance with the chemical nature of the Atype_N_68 descriptor. This molecule also has two other substitutions with the same chemical features in the X-region, favouring lower activity (Table 2). The presence of an Atype_N_68 descriptor type chemical feature might be the key point determining the low activity of molecule **82**.

We also analysed the contribution of different descriptor values in QSAR model 4 for Atype_N_68 and S_sssN descriptors, with respect to the most potent molecules **53** and **54** as well as that of the least potent molecule **82**. These values were used mainly for the development of QSAR model 4. For the most potent molecules **53** and **54**, the Atype_N_68 descriptor has a descriptor value of 0.079. For the least potent molecule **82**, this value is 1.61. The negative correlation of the Atype_N_68 descriptor in building QSAR model 4 revealed that, for the most potent molecules, its contribution (or descriptor value) should be less, and for the least potent molecules its contribution should be more. Thus, the high descriptor value of the Atype_N_68 descriptor for molecule **82** suggests its contribution in making the molecule less active, viz., this descriptor is negatively correlated with QSAR model 4. On the other hand, the most potent molecules (**53** and **54**) have lower descriptor values as mentioned above, suggesting that the contribution of Atype_N_68 descriptor is less, which in turn favour greater molecular activity. The S_sssN descriptor has descriptor values 0.972 and 0.770 for the most potent molecules **53** and **54**. For the least potent molecule **82**, this value is 1.15, which is less than that observed for the Atype_N_68 descriptor as mentioned above. This analysis shows clearly that the positive (and negative)

contribution of the value of the descriptors S_sssN (and Atype_N_68) in generating QSAR model 4 is in agreement with each molecule's N-type CCB activity. Thus, the above analysis clearly demonstrates that the influence of Atype_N_68 and S_sssN descriptors is very important in distinguishing between potent and less potent N-type CCBs.

The main outlier of QSAR model 4 is molecule **87**, which fits neither in the training set nor the test set. This molecule belongs to the scaffold E group of the dataset. Scaffold E group molecules (**83–91**) have medium-range potency (0.18–1.80 μM), with the exception of molecule **87**, with a potency of 120 μM. Most of these medium-range potency molecules have aliphatic substitutions at the R- and X- regions, which are thus less electronegative in nature. Molecule **87**, in which the nitrogen of the imidazole ring exhibits double bond connectivity, is more polar than other molecules in this series. This chemical feature is present only in molecule **87** of this E series, and is not supported by the descriptor S_sssN. Further, the S_sssN descriptor correlated positively with the activity as shown by the QSAR equation of model 4. All the above observations clearly support that conclusion that molecule **87** is inactive. Hence, the influence of these descriptors in building QSAR model 4 will be essential to the prediction of the activity of diverse classes of N-type CCBs.

Conclusions

Using 104 N-type CCBs, a robust QSAR model was developed using GFA methodology. The best QSAR model (4) was created using five important molecular descriptors: Atype_C_24, Rotlbonds, and S_sssN descriptors have positive correlation, whereas ADME_Solubility and Atype_N_68 descriptors have negative correlation with N-type CCB activity.

QSAR model 4 was validated by cross-validation, randomisation, and test set prediction techniques. Model 4 is statistically and chemically sound and explains more than 95% of the variance in the experimental activity with good predictive power, as evidenced from the predicted activity of the test set molecules. The structural and electrotopological index descriptors were found to play a major role in determining N-type CCB activity. The rotational bond and atom-type descriptors highlight the importance of spatial aspects in designing these CCBs. The influence of electrotopological descriptor S_sssN was promising, and showed that the tertiary nitrogen with its linear aliphatic amine contributions must be taken into account when designing new inhibitors against this channel. Analysis of atom-wise contributions to hydrophobicity will probably help to

appropriately take into account those atom types that are essential for determining CCB activity.

2D-QSAR modelling can help in determining which parts of a molecule can be modified to increase affinity and efficacy, providing valuable guidance in the drug discovery process. These models are useful because they rationalise a large number of experimental observations, and allow saving of both time and money in the drug design process, and represent a step forward in the *in silico* identification of potent N-type CCBs.

References

1. Carafoli E, Santella L, Branca D, Brini M (2001) Generation, control, and processing of cellular calcium signals. Crit Rev Biochem Mol Biol 36:107–260
2. Snutch TP, Larry RS (2009) Voltage-gated calcium channels. Encyclopedia of neuroscience. Academic, Oxford, pp 427–441
3. Catterall WA (2000) Structure and regulation of voltage-gated Ca^{2+} channels. Annu Rev Cell Dev Biol 16:521–555
4. Ertel EA, Campbell KP, Harpold MM et al (2000) Nomenclature of voltage-gated calcium channels. Neuron 25:533–535
5. Catterall WA, Perez-Reyes E, Snutch TP, Striessnig J (2005) International union of pharmacology. XLVIII. Nomenclature and structure-function relationships of voltage-gated calcium channels. Pharmacol Rev 57:411–425
6. McGivern JG (2006) Targeting N-type and T-type calcium channels for the treatment of pain. Drug Discov Today 11:245–253
7. McQuay H (1999) Opioids in pain management. Lancet 353:2229–2232
8. Olivera BM, Cruz LJ, de Santos V (1987) Neuronal calcium channel antagonists. Discrimination between calcium channel subtypes using ω-conotoxin from *Conus magus* venom. Biochemistry 26:2086–2090
9. Brose WG, Gutlove DP, Luther RR, Bowersox SS, McGuire D (1997) Use of intrathecal SNX-111, a novel, N-type, voltage-sensitive, calcium channel blocker, in the management of intractable brachial plexus avulsion pain. Clin J Pain 13:256–259
10. Atanassoff PG, Hartmannsgruber MWB, Thrasher J et al (2000) Ziconotide, a new n-type calcium channel blocker, administered intrathecally for acute postoperative pain. Reg Anesth Pain Med 25:274–278
11. Takahara A, Fujita S, Moki K, Ono Y, Koganei H, Iwayama S, Yamamoto H (2003) Neuronal Ca2+ channel blocking action of an antihypertensive drug, cilnidipine, IMR-32 human neuroblastoma cells. Hypertens Res 26:743–747
12. Yamamoto T, Niwa S, Ohno S et al (2006) Structure-activity relationship study of 1, 4- dihydropyridine derivatives blocking N-type calcium channels. Bioorg Med Chem Lett 16:798–802
13. Yamamoto T, Niwa S, Iwayama S et al (2006) Discovery, structure-activity relationship study, and oral analgesic efficacy of cyproheptadine derivatives possessing N-type calcium channel inhibitory activity. Bioorg Med Chem 14:5333–5339
14. Yamamoto T, Niwa S, Ohno S et al (2008) The structure-activity relationship study on 2-, 5-, and 6-position of the water soluble 1,

4-dihydropyridine derivatives blocking N-type calcium channels. Bioorg Med Chem Lett 18:4813–4816

15. Seko T, Kato M, Kohno H et al (2002) Structure-activity study of L-cysteine-based N-type calcium channel blockers: optimization of N- and C-terminal substituents. Bioorg Med Chem Lett 12:915–918

16. Knutsen LJS, Hobbs CJ, Earnshaw CG et al (2007) Synthesis and SAR of novel 2-arylthiazolidinones as selective analgesic N-type calcium channel blockers. Bioorg Med Chem Lett 17:662–667

17. Manly CJ, Louise-May S, Hammer JD (2001) The impact of informatics and computational chemistry on synthesis and screening. Drug Discov Today 6:1101–1110

18. Hansch C, Leo A (1995) Exploring QSAR: fundamentals and applications in chemistry and biology. American Chemical Society, Washington

19. Kubinyi H (1993) QSAR: Hansch analysis and related approaches. VCH, New York

20. Abraham DJ (2003) Burger's medicinal chemistry and drug discovery, volume 1: drug discovery. Wiley, New York

21. Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure-property relationships. J Chem Inf Comput Sci 34:854–866

22. Gupta SP, Veerman A, Bagaria P (2004) Quantitative structure–activity relationship studies on some series of calcium channel blockers. Mol Divers 8:357–363

23. Mohan CG, Gandhi T (2008) Therapeutic potential of voltage gated calcium channels. Mini Rev Med Chem 8:1285–1290

24. Hu LY, Ryder TR, Nikam SS, Millerman E, Szoke BG, Rafferty MF (1999) Synthesis and biological evaluation of substituted 4-(OBz) phenylalanine derivatives as novel N-type calcium channel blockers. Bioorg Med Chem Lett 9:1121–1126

25. Hu LY, Ryder TR, Rafferty MF et al (1999) N, N-Dialkyl-dipeptidylamines as novel N-type calcium channel blockers. Bioorg Med Chem 9:907–912

26. Hu LY, Ryder TR, Rafferty MF et al (2000) The discovery of [1-(4-dimethylamino-benzyl)-piperidin-4-yl]-[4-(3, 3-dimethylbutyl)-phenyl]-(3-methyl-but-2-enyl)-amine, an N-type Ca^{2+} channel blocker with oral activity for analgesia. Bioorg Med Chem 8:1203–1212

27. Hu LY, Ryder RT, Rafferty MF et al (1999) Structure–activity relationship of N-methyl-N-aralkyl-peptidylamines as novel N-type calcium channel blockers. Bioorg Med Chem Lett 9:2151–2156

28. Ryder RT, Hu LY, Rafferty MF et al (1999) Structure–activity relationship at the proximal phenyl group in a series of non-peptidyl N-type calcium channel antagonists. Bioorg Med Chem Lett 9:2453–2458

29. Ryder RT, Hu LY, Rafferty MF, Millerman E, Szoke BG, Tarczy-Hornoch K (1999) Multiple parallel synthesis of N, N-dialkyldipeptidylamines as N-type calcium channel blockers. Bioorg Med Chem Lett 9:1813–1818

30. Tripos (2008) Molecular Modeling Software, Version 7.1. Tripos Associates, St. Louis, MO

31. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36:3219–3228

32. Fletcher R, Powell MJD (1963) A rapidly convergent descent method for minimization. Comput J 6:163–168

33. Clark M, Cramer RD, Opdenbosch NV (1989) Validation of the general purpose Tripos 5. 2 force field. J Comput Chem 10:982–1012

34. Cerius2, Version 4.10 (2005) San Diego, CA

35. Leach AR, Gillet VJ (2003) An introduction to chemoinformatics. Springer, New York

36. Jackson JE (1991) A user's guide to principal components. Wiley, New York

37. Friedman JH (1991) Multivariate adaptive regression splines. Ann Stat 19:1–67

38. Rucker C, Rucker G, Meringer M (2007) y-randomization and its variants in QSPR/QSAR. J Chem Inf Model 47:2345–2357

39. Wang XS, Tang H, Golbraikh A, Tropsha A (2008) Combinatorial QSAR modeling of specificity and subtype selectivity of ligands binding to serotonin receptors 5HT1E and 5HT1F. J Chem Inf Model 48:997–1013

40. Brett AT, Lori BP, Reynolds CH (2002) Chemical information based scaling of molecular descriptors: a universal chemical scale for library design and analysis. J Chem Inf Comput Sci 42:879–884

41. Oprea TI, Allu TK, Fara DC, Rad RF, Ostopovici L, Bologa CG (2007) Lead-like, drug-like or "Pub-like": how different are they? J Comput Aided Mol Des 21:113–119

42. Ghose AK, Crippen GM (1987) Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. J Chem Inf Comp Sci 27:21–35

43. Viswanadhan VN, Ghose AK, Reyankar GR, Robins RK (1989) Atomic physico-chemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. J Chem Inf Comp Sci 29:163–172

44. Zhihua L, Yuzhang W, Xuejun Q, Yuegang Z, Bing N, Ying W (2002) Use of a novel electrotopological descriptor for the prediction of biological activity of peptide analogues. Int J Pept Res Ther 9:273–281

45. Hall LH, Kier LB (1995) Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. J Chem Inf Comp Sci 35:1039–1045

46. Cheng A, Merz K Jr (2003) Prediction of aqueous solubility of a diverse set of compounds using quantitative structure–property relationships. J Med Chem 46:3572–3580