Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 2354–2360
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

# QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach

Susheela Sharma [a], Yenamandra S. Prabhakar [b], Prithvi Singh [c], Brij Kishore Sharma [c,*]

[a] *Department of Engineering Chemistry, Sobhasaria Engineering College, Sikar 332 021, India*
[b] *Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow 226 001, India*
[c] *Department of Chemistry, S.K. Government College, Sikar 332 001, India*

Received 18 September 2007; received in revised form 19 December 2007; accepted 10 January 2008
Available online 31 January 2008

**Abstract**

The structure–activity models of the myorelaxant activity of the cromakalim analogues have been investigated with nearly 470 topological descriptors from DRAGON software using Combinatorial Protocol in Multiple Linear Regression (CP-MLR). Among the descriptor classes considered in the study, the binding affinity is correlated with simple functional (FUN), topological (TOPO), atom centered fragments (ACF), empirical (EMP), modified Burden eigenvalues (BCUT), Galvez topological charge indices (GVZ), 2D-autocorrelation (2D-AUTO) and constitutional (CONS) descriptors. The models developed, and the participating descriptors suggest that the substituent groups of 4,6-disubstituted-2,2-dimethylchromans hold scope for further modification in the optimization of activity. The higher path lengths rich in polarizability and lower path length rich in atomic mass in addition to the lower charge indices of the molecule are beneficiary to the activity. The participating descriptors also suggested that certain structural features such as carbon atoms attached to the heteroatom by single or multiple bonding, and lesser or 'no' branching in a molecule are helpful to augment the activity.


*Keywords:* QSAR; Cromakalim analogues; Myorelaxant activity; CP-MLR approach

## 1. Introduction

ATP-sensitive potassium channels ($K_{ATP}$ channels) regulate the flow of potassium ions through the cell membrane. These were identified in a wide range of cell types and are found to link the metabolic state to the electric state of the cell [1–8]. $K_{ATP}$ channels are composed of two different protein subunits in a 4 + 4 stoichiometry [9]. The $K_{ATP}$ channel pore belongs to the inwardly rectifying potassium channel family, which is known as Kir6.x [10]. The second subunit, the sulfonylurea receptor (SUR) subunit, contains the regulatory sites for most drugs [10]. Four variants of SUR, namely SUR1, SUR2A, SUR2B and SUR2C, have been reported [11]. $K_{ATP}$ channels are composed of different subunits according to their tissue localization. For example, SUR1 combined with Kir6.2 forms the pancreatic $K_{ATP}$ channels [12]. The combination of SUR2A and Kir6.2 subunits is found in cardiac and skeletal muscle whereas the smooth muscle $K_{ATP}$ channel is composed of SUR2B and Kir6.1 or Kir6.2 subunits [13]. Such a feature is a key element for the design of new $K_{ATP}$ channel modulators. The main challenge in the development of $K_{ATP}$ modulators as therapeutic agents is to search out compounds with the best selectivity for a single $K_{ATP}$ channel subtype. The pancreatic $K_{ATP}$ channels are well-known to be involved in the insulin-releasing process [14,15] and smooth muscle $K_{ATP}$ channels in the control of muscle tone [16,17], the physiological roles of the different channel subtypes have not yet been thoroughly assessed [8,18].

Several drugs, named as potassium channel openers (PCOs), have been found to activate $K_{ATP}$ channels [19,20], leading to plasma membrane hyperpolarization and reduction in cell excitability. This, in turn, may provoke the relaxation of smooth muscles and/or the inhibition of endocrine releases [21,22]. One such compound cromakalim (Fig. 1) exhibits a marked myorelaxant activity resulting from the smooth muscle $K_{ATP}$

* Corresponding author. Tel.: +91 1572 245200.
*E-mail address:* bksharma_sikar@rediffmail.com (B.K. Sharma).

0223-5234/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.01.020

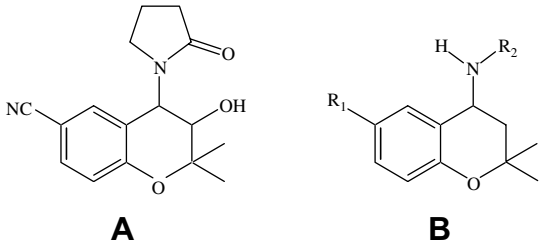


Fig. 1. Structure of cromakalim (A) and its derivatives (B).

channels [23]. However, it is poorly active as an inhibitor of insulin secretion [24,25]. Recently, in an attempt to discover potent and pancreatic β-cell selective PCOs, a series of 4,6-disubstituted-2,2-dimethylchromans, structurally related to cromakalim, were evaluated [26] on insulin secretion from rat pancreatic islets and on the contractile activity of rat aorta rings. This structure–activity relationship (SAR) study was mainly aimed at the incorporation of different substituents at the said moiety (urea, thiourea, carbamate, sulfonylurea and amide). In the present communication, a quantitative SAR (QSAR) study on these analogues (Fig. 1, Table 1) has been conducted to provide the rationale for drug-design and to explore the possible mechanism of their action at molecular level.

Table 1
Observed and modeled binding affinities of cromakalim and its analogues at the rat aorta rings (Fig. 1 for structures)

| Compound number | $R_1$ | $R_2$ | $pED_{50}$ | | | |
|---|---|---|---|---|---|---|
| | | | Obsd.[a] | Calcd. Eq. (1) | Calcd. Eq. (2) | Calcd. Eq. (3) |
| **1** | F | $CONHCH_2CH_3$ | 4.37 | 4.72 | 4.49 | 4.68 |
| **2** | Cl | $CONHCH_2CH_3$ | 5.17 | 4.72 | 4.49 | 4.68 |
| **3** | Br | $CONHCH_2CH_3$ | 5.12 | 4.96 | 4.94 | 4.87 |
| **4** | F | $CONHCH(CH_3)_2$ | 3.96 | 4.25 | 4.01 | 4.05 |
| **5** | Cl | $CONHCH(CH_3)_2$ | 3.84 | 4.28 | 4.39 | 4.16 |
| **6** | Cl | $CONHCH_2C_6H_5$ | 4.82 | 4.82 | 4.90 | 4.93 |
| **7** | F | $CSNHCH_2CH_3$ | 4.75 | 5.14 | 4.71 | 5.11 |
| **8** | Cl | $CSNHCH_2CH_3$ | 5.09 | 5.09 | 4.93 | 5.20 |
| **9** | Br | $CSNHCH_2CH_3$ | 5.43 | 5.17 | 4.96 | 5.20 |
| **10** | F | $CSNHCH(CH_3)_2$ | 4.73 | 4.59 | 4.69 | 4.75 |
| **11** | Cl | $CSNHCH(CH_3)_2$ | 4.96 | 4.58 | 4.91 | 4.82 |
| **12** | Br | $CSNHCH(CH_3)_2$ | 5.06 | 4.71 | 4.85 | 4.80 |
| **13** | F | $CSNHCH_2C_6H_5$ | 5.37 | 5.10 | 4.75 | 4.88 |
| **14** | Cl | $CSNHCH_2C_6H_5$ | 4.70 | 4.97 | 4.98 | 4.94 |
| **15** | Br | $CSNHCH_2C_6H_5$ | 4.61 | 5.04 | 5.04 | 4.94 |
| **16** | F | $COOCH_2CH_3$ | 4.14 | 4.23 | 4.28 | 4.18 |
| **17** | Cl | $COOCH_2CH_3$ | 4.42 | 4.25 | 4.66 | 4.33 |
| **18** | Br | $COOCH_2CH_3$ | 4.59 | 4.55 | 4.76 | 4.49 |
| **19** | Cl | $CONHSO_2C_6H_5$ | 3.68 | 4.10 | 4.25 | 4.14 |
| **20** | Cl | $CONHSO_2C_6H_4$ (4-$CH_3$) | 4.20 | 4.16 | 3.96 | 4.05 |
| **21** | Cl | $CONHSO_2C_6H_4$ (4-Cl) | 4.29 | 3.79 | 4.05 | 3.95 |
| **22** | F | $COCH_3$ | 4.33 | 4.93 | 4.71 | 4.89 |
| **23** | Cl | $COCH_3$ | 5.18 | 4.85 | 5.09 | 4.95 |
| **24** | Br | $COCH_3$ | 5.19 | 5.07 | 5.14 | 4.97 |
| **25**[b] | (±)-Cromakalim | | 6.89 | 6.81 | 6.96 | 6.91 |
| **26**[c] | Br | $CONHCH(CH_3)_2$ | <3.52 | 4.46 | 4.43 | 4.24 |
| **27**[c] | Br | $CONHCH_2C_6H_5$ | <3.52 | 4.94 | 5.01 | 4.92 |
| **28** | F | $CONHCH_2C_6H_5$ | ND[d] | 4.89 | 4.57 | 4.89 |

[a] $pED_{50}$ expressed as negative logarithm on molar basis, represents the myorelaxant activity of KCl-depolarized rat aorta rings; taken from Ref. [26].
[b] Reference compound; cromakalim.
[c] Uncertain activity.
[d] Not determined.

In QSAR studies, the transformation of a chemical structure into numerical descriptors (variables) plays a pivotal role and the meaningful dependent–independent variables' communication results in the evolution of a variety of models with different predictive and diagnostic values. In this way, each model may address different substructural regions and attributes in explaining the chosen phenomenon. This provides scope to understand the predictive and diagnostic aspects of different substructural regions and in averaging them beyond the individual models. For this, it is necessary to characterize the molecules and their structural fragments from different perspectives for the generation of a large number of diverse descriptors. Several computer programs based on graph theory are available to compute characteristic descriptors of the molecules and their structural fragments [27–31]. Moreover, when dealing with a large number of descriptors, for the optimum utilization of information content of the generated datasets, it is necessary to adopt typical protocol(s) to identify the best models as well as information rich descriptors corresponding to the phenomenon under investigation. The Combinatorial Protocol in Multiple Linear Regression (CP-MLR) is an approach, amongst many others, to address the model evolution in high-dimensional QSAR studies. In this milieu, the myorelaxant activity of cromakalim analogues has been analyzed with the molecular descriptors from DRAGON software [30] using a typical descriptor search protocol involving the CP-MLR approach [32–37].

## 2. Materials and methods

### 2.1. Dataset

In this study, 4,6-disubstituted-2,2-dimethylchromans or cromakalim analogues (Table 1) have been considered from the literature report [26] along with their myorelaxant effects, reported in terms of the contractile activity of KCl-depolarized rat aorta rings in the form of logarithm of the inverse of inhibitory concentration ($pED_{50}$, where $ED_{50}$ is the concentration in moles per liter for 50% inhibition against the $K_{ATP}$ channel). The structures of the title compounds have been drawn in ChemDraw [38] using the standard procedure. These structures were ported to DRAGON software [30] for the computation of parameters related to 0D-, 1D-, and 2D-descriptor classes. A total number of 472 descriptors have been obtained corresponding to all the descriptor classes. These descriptor classes along with their definitions and scope in addressing the structural features are given in Table 2. As the total number of descriptors involved in this study is very large, only the names of descriptor classes and the actual descriptor involved in the models have been addressed in the discussion.

### 2.2. CP-MLR

This method involves a 'filter'-based variable selection procedure for the model identification and development in QSAR studies [32–37]. It involves a combinatorial strategy with

Table 2
Descriptor classes used for the analysis of myorelaxant activity of cromakalim analogues and identified categories in modeling the activity

| Descriptor class (acronyms)[a] | Definition and scope | Descriptors' category[b] |
|---|---|---|
| Constitutional (CONST) | Dimensionless or 0D descriptors; independent from molecular connectivity and conformations | II |
| Topological (TOPO) | 2D-descriptor from molecular graphs and independent of conformations | I |
| Molecular walk counts (MWC) | 2D-descriptors representing self-returning walks counts of different lengths | IV |
| Modified Burden eigenvalues (BCUT) | 2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms | II |
| Galvez topological charge indices (GVZ) | 2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix | II |
| 2D-autocorrelations (2DAUTO) | Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the *lag*) | I |
| Functional groups (FUN) | Molecular descriptors based on counting of the chemical functional groups | III |
| Atom centered fragments (ACF) | Molecular descriptors based on counting of 120 atom centered fragments, as defined by Ghose–Crippen | I |
| Empirical (EMP) | 1D-descriptors represent the counts of nonsingle bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule | II |
| Properties (PROP) | 1D-descriptors representing molecular properties of a molecule | IV |

[a] Ref. [30].
[b] Descriptor categories identified at the end of second stage; in this the filter values are as follows: filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.71, and filter-4 as $0.3 \leq Q^2 \leq 1.0$, the number of compounds in each dataset was 25.

appropriately placed 'filters' interfaced with MLR and extracts diverse models having unique combination of descriptors from the dataset. The filters set the thresholds for the descriptors in terms of inter-parameter correlation cutoff limits in subset regressions (filter-1), *t*-values of the regression coefficients (filter-2), internal explanatory power (filter-3; square root of adjusted multiple correlation coefficient of regression equation, *r*-bar), and the external consistency (filter-4; $Q^2$, i.e. cross-validated $R^2$ from the leave-one-out procedure). Throughout this study, for the filters-1, 2 and 4 the thresholds were assigned as 0.3, 2.0, and $0.3 \leq Q^2 \leq 1.0$, respectively. The filter-3 was assigned an initial value of 0.71. In order to collect the descriptors with higher information content, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for the next generation.

### 2.3. Descriptor classification protocol

The three-stage descriptor classification protocol [33] is implemented with the two descriptor combinations (baseline models), as they are the simplest to understand and explain the activity. In the first stage of the classification protocol, the correlations of the activity with two descriptor combinations from the individual descriptor classes (DCs) of the dataset were used to sort the DCs into four categories. They are primary contributors (category I: a DC forms a model with its constituent descriptors), collective contributors (category II: a DC unable to form a model with its constituent descriptors, but forms model(s) in combination with a descriptor from another such DC), secondary contributors (category III: a DC forms a model(s) only in combination with category I) and noncontributors (category IV: a DC unable to form a model(s) in any manner like that of categories I–III). The sorted DCs were collated in the second stage to identify all the 3-descriptor models across the categories. In the last stage, the individual descriptors of all three-descriptor models were pooled to discover the higher models for the activity.

All the models identified in the last stage have further been put to a randomization test [34,39] by repeated randomization of the activity, without altering the original descriptors' matrix, to discover the chance correlations, if any, associated with them. For this every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted to express the percent chance correlation of the model under scrutiny. The model development procedure has been further verified by creating divergent training and test sets from the compounds of the study.

## 3. Results and discussion

For 4,6-disubstituted-2,2-dimethylchromans (Table 1), several of their descriptors from the 0D- to 2D-descriptor classes have shown significant correlation with their myorelaxant activity. The derived models involving two descriptors from 2DAUTO, TOPO and ACF descriptor classes have been given in Table 3.

Among these the 2DAUTO class descriptors MATS2e (Moran autocorrelation – lag 2/weighted by atomic Sanderson electronegativities) and MATS4m (Moran autocorrelation – lag 4/weighted by atomic masses) correlate negatively to the activity, whereas MATS1p (Moran autocorrelation – lag 1/weighted by atomic polarizabilities) and GATS4e (Geary autocorrelation – lag 4/weighted by atomic Sandersons electronegativities) correlate positively to the activity. Thus suggesting the unfavorable conditions associated with lag 2 and lag 4 weighted by atomic Sandersons electronegativities and atomic masses, respectively. On the other hand, autocorrelation of lag 1 weighted by atomic polarizabilities and lag 4

Table 3
Derived models in two descriptors from category I and the statistical parameters[a]

| Des. class | Constant | Variable-1 | Variable-2 | $r$ | $Q^2$ | $s$ | $F$ |
|---|---|---|---|---|---|---|---|
| 2DAUTO | 3.592 | −6.107MATS2e | 2.436GATS4e | 0.770 | 0.419 | 0.436 | 16.12 |
| | 69.174 | −63.677MATS4m | 14.771MATS1p | 0.754 | 0.354 | 0.449 | 14.53 |
| TOPO | 25.807 | −63.868X2A | −1.642SEige | 0.825 | 0.448 | 0.387 | 23.46 |
| | −74.622 | −0.041DELS | 32.609LP1 | 0.822 | 0.604 | 0.390 | 22.94 |
| ACF | 4.257 | 0.581C-006 | 0.861C-040 | 0.744 | 0.368 | 0.457 | 13.67 |

[a] In all cases, the models were derived for $n = 25$; the statistical parameters $r$, $Q^2$, $s$ and $F$ are discussed in the text.

weighted by atomic Sandersons electronegativities suggests favorable conditions. The TOPO class descriptor LP1 {Lovasz–Pelikan index (leading eigenvalue)} is in favor of activity, whereas, the contribution of X2A (average connectivity index chi-2), SEigv (eigenvalue sum from van der Waals volume weighted distance matrix), and DELS (molecular electrotopological variation) toward activity is negative. The ACF class descriptor, C-006 (CH2RX), is the number of $CH_2$ whose remaining valences are satisfied by alkyl group and heteroatom. C-040 (R–C(=X)–X/R–C#X/X–=C=X) is the number of carbon atoms which are attached to heteroatoms by single/double or triple bonds. These descriptors recommend the certain structural features for better activity.

Even though each individual descriptor class is enriched in information to explain the activity, but their collective influence may lead to the models of highest explained variance. Such models, with increased number of descriptors, were identified by the CP-MLR analysis on successive increment of filter-3. For this, the optimum r-bar value of the preceding model has been used as the new threshold of filter-3 for the next generation. At the end of a search, 52 three-parameter models sharing 35 descriptors among themselves were identified. All these 35 descriptors are listed in Table 4 along with their brief meaning, and total incidence which will serve as a measure of their estimate across these models. The identified descriptors may further have scope in evolving the higher models. The

Table 4
Descriptors identified for modeling the myorelaxant activity of cromakalim analogues along with their average regression coefficients and the total incidence

| Descriptor[a] | Avg. reg. coeff. (total incidence)[b] | Descriptor[a] | Avg. reg. coeff. (total incidence)[b] | Descriptor[a] | Avg. reg. coeff. (total incidence)[b] |
|---|---|---|---|---|---|
| CONS | | Me | −38.226(1) | Ms | −1.904(2) |
| RBN | −0.246(1)1 | RBF | −16.731(1) | | |
| TOPO | | Rww | −0.099(6) | X2A | −57.842(15) |
| S3K | −0.520(1) | PW3 | 30.321(7) | PW4 | 93.001(3) |
| LP1 | 25.616(3) | SEigv | 0.233(1) | SEige | −1.480(1) |
| BCUT | | BELe1 | 98.722(14) | | |
| GALV | | GGI1 | −0.327(8) | GGI3 | −0.912(1) |
| GGI7 | −2.661(5) | GGI8 | −3.556(12) | JGI4 | 42.439(7) |
| 2D-AUTO | | MATS4m | −54.388(6) | MATS1e | −6.266(1) |
| MATS2e | −3.384(3) | MATS1p | 11.289(10) | MATS4p | 7.621(9) |
| GATS7v | 1.387(1) | GATS3e | 1.235(1) | GATS4e | 2.661(3) |
| GATS7e | −0.823(1) | GATS1p | −1.765(9) | GATS5p | 1.277(1) |
| FUN | | nHDon | 0.346(5) | | |
| ACF | | C-006 | 0.497(1) | C-040 | 0.801(4) |
| O-058 | −0.402(1) | S-108 | 0.510(4) | | |
| EMP | | Hy | 0.524(7) | | |

[a] The descriptors are identified from the three- and four-parameter models emerged from CP-MLR protocol with filter-1 as 0.3; filter-2 as 2.0; filter-3 as 0.82; filter-4 as $0.3 \leq Q^2 \leq 1.0$; a number of compounds in the study are 25; CONST: Me, mean atomic Sanderson electronegativity (scaled on carbon atom); Ms, mean electrotopological state; RBN, number of rotatable bonds; RBF, rotatable bond fraction. TOPO: Rww, reciprocal hyper-detour index; X2A, average connectivity index chi-2; S3K, 3-path Kier alpha-modified shape index; PW3 and PW4, path/walk Randic shape index of order 3 and 4, respectively; LP1, Lovasz–Pelikan index (leading eigenvalue); SEigv and SEIge, eigenvalue sum from van der Waals volume and electronegativity weighted distance matrix, respectively. BCUT: BELe1, the lowest eigenvalue no.1 of Burden matrix/weighted by atomic Sanderson electronegativities. GALV: GGI1, GGI3, GGI7 and GGI8, topological charge indices of order 1, 3, 7 and 8, respectively; JGI4, mean topological charge index of order 4. 2DAUTO: MATS4m, Moran autocorrelation – lag 4/weighted by atomic masses; MATS1e and MATS2e, Moran autocorrelation – lag 1 and 2, respectively/weighted by atomic Sanderson electronegativities; MATS1p and MATS4p, Moran autocorrelation – lag 1 and 4, respectively/weighted by atomic polarizabilities; GATS7v, Geary autocorrelation – lag 7/weighted by atomic van der Waals volumes; GATS3e, GATS4e and GATS7e, Geary autocorrelation – lag 3, 4 and 7, respectively/weighted by atomic Sandersons electronegativities; GATS1p and GATS5p, Geary autocorrelation – lag 1 and 7, respectively/weighted by atomic polarizabilities. FUN: nHDon, number of donor atoms for H-bonds (with N and O). ACF: C-006, CH2RX; C-040, R–C(=X)–X/R–C#X/X–=C=X; O-058, O=; S-108, R=S. EMP: Hy, hydrophilic factor; also see Ref. [30].

[b] The average regression coefficient of the descriptor corresponding to all models and the total number of its incidences; the arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

highest significant models, in three and four descriptors, selected from the study are shown through Eqs. (1)–(3)

$$\mathrm{pED_{50}} = 60.747 - 55.466(12.333)\mathrm{MATS4m} + 12.837(2.649)\mathrm{MATS1p} + 0.598(0.141)\mathrm{C\text{-}040}$$
$$n = 25,\ r = 0.876,\ s = 0.338,\ Q^2 = 0.672,\ F = 23.072 \quad (1)$$

$$\mathrm{pED_{50}} = 24.913 - 65.976(10.109)\mathrm{X2A} - 2.207(0.499)\mathrm{GGI7} + 7.881(1.724)\mathrm{MATS4p}$$
$$n = 25,\ r = 0.872,\ s = 0.343,\ Q^2 = 0.672,\ F = 22.287 \quad (2)$$

$$\mathrm{pED_{50}} = -5.694 + 31.396(5.011)\mathrm{PW3} - 3.111(0.678)\mathrm{GGI8} + 10.219(2.501)\mathrm{MATS1p} + 1.330(0.585)\mathrm{GATS5p}$$
$$n = 25,\ r = 0.899,\ s = 0.314,\ Q^2 = 0.722,\ F = 21.076 \quad (3)$$

In all above regression equations, $n$ is the number of compounds, $r$ is the correlation coefficient, $Q^2$ is cross-validated index from leave-one-out (LOO) procedure, $s$ is the standard error of the estimate and $F$ is the $F$-ratio between the variances of calculated and observed activities. The values given in the parentheses are the standard errors of the regression coefficients. Above model equations were further subjected to randomization process, where 100 simulations per model were carried out but none of the identified models has shown any chance correlation. These models were then used to calculate the $\mathrm{pED_{50}}$s for all the compounds. The same were found in parity with the observed ones (Table 1). Also, these equations were used to predict the activities of three compounds whose activities were either uncertain (**26**, **27**) or unreported (**28**). Additionally, the above model equations were subjected to external validation. For this, three test sets, each consisting of nearly 25% of the total compounds, have been considered and remaining compounds were included in the corresponding three training sets. Of the three test sets, two were generated in the SYSTAT [40] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the respective descriptors or the activity as the case may be. The selection of the test set from the cluster tree was done in such a way to keep the test compounds at a maximum possible distance from each other. The third test of the compounds corresponds to the random selection procedure. Thus these test sets, each having six compounds, represent different cross-sections of compounds. The predictions of the test sets have been done with the models developed using remaining 19 compounds of the training sets. The residuals of the predictions of three test sets and the corresponding predictive $r^2$-, $s$-, $Q^2$- and $F$-values of the study have been given in Table 5. In order to identify the systematic variation and the type of behavior shown by the congeners, the residuals obtained for the training and test sets derived from the highest significant equation (3) are only given here. The plots given in Fig. 2 show that the

Table 5
Predicted residual activity of different test sets (six compounds each) of the compounds of Table 1 and the statistics of test and training sets

| Compd. no. | Residual[a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Des. clus.[b] | | | Act. clus.[c] | | | Random[d] | | |
| | D1[e] | D2[f] | D3[g] | A1[e] | A2[f] | A3[g] | Ra1[e] | Ra2[f] | Ra3[g] |
| **1** | | | | | | | −0.42 | −0.17 | −0.37 |
| **3** | | | | 0.15 | 0.21 | 0.31 | | | |
| **5** | −0.46 | −0.60 | −0.37 | | | | −0.47 | −0.56 | −0.39 |
| **10** | | | | −0.40 | 0.05 | −0.37 | | | |
| **11** | −0.06 | 0.12 | −0.12 | | | | −0.03 | 0.14 | −0.10 |
| **12** | | | | 0.27 | 0.50 | 0.25 | | | |
| **15** | | | | | | | 0.34 | 0.23 | 0.32 |
| **17** | −0.33 | −0.30 | −0.31 | | | | | | |
| **19** | −0.11 | −0.15 | 0.00 | | | | −0.23 | −0.20 | −0.14 |
| **21** | | | | −0.01 | −0.15 | 0.11 | | | |
| **22** | | | | −0.47 | −0.68 | −0.67 | | | |
| **23** | 0.02 | 0.27 | 0.14 | −0.01 | 0.09 | −0.05 | −0.03 | 0.28 | 0.17 |
| **26** | 0.39 | 0.03 | 0.20 | | | | | | |
| *Test set* | | | | | | | | | |
| $r^2$ | 0.761 | 0.728 | 0.848 | 0.778 | 0.630 | 0.644 | 0.752 | 0.769 | 0.803 |
| *Training set* | | | | | | | | | |
| $r^2$ | 0.764 | 0.767 | 0.796 | 0.760 | 0.785 | 0.828 | 0.767 | 0.760 | 0.803 |
| $s$ | 0.363 | 0.360 | 0.349 | 0.361 | 0.342 | 0.316 | 0.357 | 0.363 | 0.340 |
| $F$ | 16.270 | 16.554 | 13.719 | 15.995 | 18.390 | 17.069 | 16.601 | 15.899 | 14.395 |
| $Q^2$ | 0.537 | 0.645 | 0.622 | 0.616 | 0.700 | 0.703 | 0.632 | 0.636 | 0.659 |

[a] The difference of observed $\mathrm{pED_{50}}$ and calculated $\mathrm{pED_{50}}$; the training models have 19 compounds each.
[b] Test set from the cluster analysis of all descriptor dataset of the compounds.
[c] Test set from the cluster analysis of the activity of the compounds.
[d] Test set from random selection of the compounds.
[e] D, A and Ra represent, respectively, descriptor cluster, activity cluster, and random selections; the training set equations are formed with the descriptors of Eq. (1).
[f] The training set equations are formed with the descriptors of Eq. (2).
[g] The training set equations are formed with the descriptors of Eq. (3).

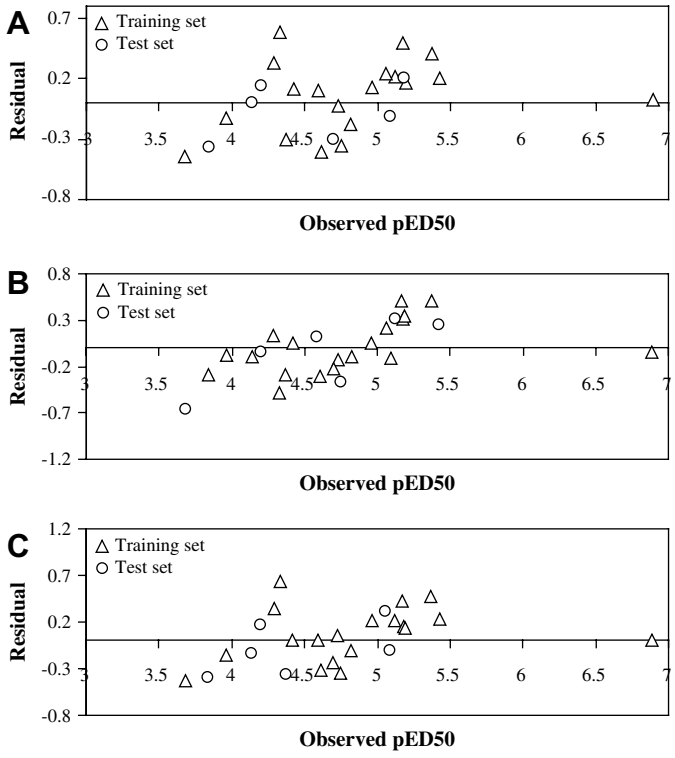


Fig. 2. Plots between observed versus residual pED50's for the cluster of (A) descriptor, (B) activity and (C) random selection.

residuals of the compounds considered in these sets are evenly distributed about the residual axis which, in turn, has validated the derived equation (3) externally.

In equations mentioned above, the 2DAUTO class descriptors, MATS1p and MATS4p, are the Moran autocorrelation of topological structure with path length (*lag*) 2 and 4, respectively, in the graph weighted by atomic polarizabilities. GATS5p is the Geary autocorrelation of *lag* 5 weighted by atomic polarizabilities. The regression coefficients of these descriptors indicate that the higher path lengths rich in polarizability content would be favorable for the improvement of activity. The other participating 2DAUTO descriptor is MATS4m, representing Moran autocorrelation of *lag* 4 in the graph weighted by atomic masses. The negative regression coefficient associated with this descriptor suggests that a lower path length rich in atomic mass would be beneficial. The Galvez topological charge index of order 7 (GGI7) and 8 (GGI8) represents, respectively, the seventh and eighth eigenvalue of the corrected adjacency matrix of a molecule and the lower values of these descriptors are required to improve the myorelaxant activity. The TOPO class descriptor, X2A, is the average connectivity index chi-2. It encodes information about size, branching, cyclization, unsaturation and heteroatom content in a molecule. The more negative or less positive value of this descriptor would enhance the activity. The other TOPO descriptor, PW3 is path/walk Randic shape index of order 3. A higher value of this index would be beneficiary to the activity. The fragment, C-040 [R–C(=X)–X/R–C#X/X–=C=X], a descriptor from ACF class, represents the number of carbon atoms attached to the heteroatom by single or multiple bonding and one valence is satisfied by an alkyl group. This structural feature in molecule would help in augmenting the activity.

## 4. Conclusion

The study has provided detailed structure–activity relationship of the myorelaxant activity of cromakalim analogues (Table 1) in terms of structural requirements. The activity of the compounds is a cumulative influence of different structural features which were identified in terms of individual descriptors. The positive regression coefficient of descriptors MATS1p, MATS4p, GATS5p and MATS4m recommend that the higher path lengths rich in polarizability would be favorable for the improvement of activity whereas, the higher path lengths rich in atomic mass would be unfavorable. The lower charge indices corresponding to the seventh and eighth eigenvalues (GGI7 and GGI8) of the molecule would augment the activity. The regression coefficient of the parameter X2A hints that branching would cause detrimental effect to the activity. The positive correlation of the path/walk Randic shape index of order 3 (PW3) suggests that a higher value of this index is beneficiary to the activity. Also, C-040 suggested that certain structural features are helpful to augment the activity. All the test sets (from the descriptors of the study, the activity, and the random selection procedures) have been predicted well by their corresponding training set equations.

## Acknowledgment

CDRI Communication No. 7453.

## References

[1] A. Noma, Nature 305 (1983) 147–148.
[2] D.L. Cook, C.N. Hales, Nature 311 (1984) 271–273.
[3] H. Bernardi, A.M. Fosset, M. Lazdunski, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 9816–9820.
[4] N.B. Standen, J.M. Quayle, N.W. Davies, J.E. Brayden, Y. Huang, M.T. Nelson, Science 245 (1989) 177–180.
[5] B. Allard, M. Lazdunski, Eur. J. Pharmacol. 236 (1993) 419–426.
[6] J.M. Quayle, M.T. Nelson, N.B. Standen, Physiol. Rev. 77 (1997) 1165–1232.
[7] J. Bryan, L. Aguilar-Bryan, Curr. Opin. Cell Biol. 9 (1997) 553–559.
[8] S. Seino, T. Miki, Prog. Biophys. Mol. Biol. 81 (2003) 133–176.
[9] A.P. Babenko, J.A. Aguilar-Bryan, Annu. Rev. Physiol. 60 (1998) 667–687.
[10] N. D'hahan, H. Jacquet, C. Moreau, P. Catty, M. Vivaudou, Mol. Pharmacol. 56 (1999) 308–315.
[11] S. Seino, Annu. Rev. Physiol. 61 (1999) 337–362.
[12] N. Inagaki, T. Gonio, J.P. Clement, Science 270 (1995) 1166–1170.
[13] A. Hambrock, C. Löfter-Walz, U. Delabar, Y. Horio, Y. Kurachi, U. Quast, Mol. Pharmacol. 55 (1999) 832–840.
[14] O.H. Petersen, M.J. Dunne, Pflueger's Arch. 414 (1989) S115–S120.
[15] P. Lebrun, Rev. Fr. Endocrinol. Clin. Nutr. Metab. 34 (1993) 241–254.
[16] H.A. Kolb, Rev. Physiol. Biochem. Pharmacol. 15 (1990) 51–79.
[17] J.E. Brayden, Clin. Exp. Pharmacol. Physiol. 29 (2002) 312–316.
[18] W.A. Cotzee, Cardiovasc. Drugs Ther. 6 (1992) 201–208.
[19] R. Mannhold, Med. Res. Rev. 24 (2004) 213–266.

[20] M.J. Coghlan, W.A. Carroll, M. Gopalakrishnan, J. Med. Chem. 44 (2001) 1627–1653.
[21] P. Lebrun, V. Devreux, M. Hermann, A. Herchuelz, J. Pharmacol. Exp. Ther. 250 (1989) 1011–1018.
[22] U. Quast, Trends Pharmacol. Sci. 14 (1993) 332–337.
[23] G. Edwards, A.-H. Weston, Annu. Rev. Pharmacol. Toxicol. 33 (1993) 597–637.
[24] P. Petit, D. Hillaire-Buys, A.-K. Mir, M.-M. Loubatires-Mariani, Fundam. Clin. Pharmacol. 255 (1992) 948–954.
[25] P. Lebrun, M.-H. Antoine, V. Devreux, M. Hermann, J. Pharmacol. Exp. Ther. 255 (1990) 948–954.
[26] S. Sebille, P. de Tullio, B. Becker, M.-H. Antoine, S. Boverie, B. Pirotte, P. Lebrun, J. Med. Chem. 48 (2005) 614–621.
[27] S.C. Basak, D.K. Harriss, V.R. Magnuson, POLLY, University of Minnesota, Duluth, MN, 1988.
[28] Molconn-Z, Ver. 2.07, eduSoft Lc, A Virginia Corporation, Ashland, VA 23005, USA. Available from: www.edusoft-lc.com.
[29] (a) A.R. Katritzky, V. Lobnov, M. Karelson, CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis), University of Florida, Gainesville, FL, 1994;
(b) A.R. Katritzky, S. Perumal, R. Petrukhin, E. Kleinpeter, J. Chem. Inf. Comput. Sci. 41 (2001) 569–574.
[30] DRAGON software version 3.0-2003. By Todeschini R, Consonni V, Mauri A, Pavan M. Milano, Italy. Available from: http//disat.unimib.it/chm/Dragon.htm.
[31] M.P. Gonzalez, A.M. Helguera, J. Comput.-Aided Mol. Des. 17 (2003) 665–672.
[32] Y.S. Prabhakar, QSAR Comb. Sci. 22 (2003) 583–595.
[33] M.K. Gupta, Y.S.J. Prabhakar, Chem. Inf. Model 46 (2006) 93–102.
[34] Y.S. Prabhakar, V.R. Solomon, R.K. Rawal, M.K. Gupta, S.B. Katti, QSAR Comb. Sci. 23 (2004) 234–244.
[35] Y.S. Prabhakar Available from: Internet Electron J. Mol. Des. 3 (2004) 150–162 http://www.biochempress.com.
[36] M.K. Gupta, R. Sagar, A.K. Shaw, Y.S. Prabhakar, Bioorg. Med. Chem. 13 (2005) 343–351.
[37] M. Saquib, M.K. Gupta, R. Sagar, Y.S. Prabhakar, A.K. Shaw, R. Kumar, P.R. Maulik, A.N. Gaikwad, S. Sinha, A.K. Srivastava, V. Chaturvedi, R. Srivastava, B.S. Srivastava, J. Med. Chem. 50 (2007) 2942–2950.
[38] Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA.
[39] S.S. So, M. Karplus, J. Med. Chem. 40 (1997) 4347–4359.
[40] SYSTAT, Version 7.0; SPSS Inc., 444 North Michigan Avenue, Chicago, IL 60611.