Innovare Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 6, Issue 5, 2014

Original Article

GROUP- BASED QUANTITATIVE STRUCTURAL ACTIVITY RELATIONSHIP ANALYSIS OF B-CELL LYMPHOMA EXTRA LARGE (BCL-XL) INHIBITORS

NADIA HANIS ABDUL SAMAT[a], ABDUALRAHMAN MOHAMMED ABDUALKADER[b], FARAHIDAH MOHAMED[acd], ABUBAKAR DANJUMA ABDULLAHI[E]

[a]Department of Pharmaceutical Technology, Kulliyyah of Pharmacy, International Islamic University Malaysia (IIUM), 25200 Kuantan, Pahang, Malaysia, [b]Department of Pharmaceutical Chemistry, Kulliyyah of Pharmacy, International Islamic University Malaysia (IIUM), 25200 Kuantan, Pahang, Malaysia, [c]International Institute of Halal Research & Training (INHART), Kulliyyah of Engineering, International Islamic University Malaysia (IIUM), 53100 Kuala Lumpur, Malaysia, [d]IKOP Sdn. Bhd., Pilot Plant Pharmaceutical Manufacturing, Kulliyyah of Pharmacy, International Islamic University Malaysia (IIUM), 25200 Kuantan, Pahang, Malaysia, [e]Department of Basic Medical Sciences, Kulliyyah of Pharmacy, International Islamic University Malaysia (IIUM), 25200 Kuantan, Pahang, Malaysia. Email: bagaruwa@gmail.com

Received: 16 Mar 2014 Revised and Accepted: 16 Apr 2014

ABSTRACT

B-cell Lymphoma Extra Large (Bcl-xL) belongs to B-cell Lymphoma two (Bcl-2) family and owing to its anti-apoptotic role in many cancers, is proven to be an attractive target for anti-cancer therapy. Different classes of potent anti-Bcl-xL small molecules inhibitors have been discovered, and both three-dimensional (3D) and two-dimensional (2D) Quantitative Structural Activity Relationship (QSAR) approaches have been used to study and predict the biological activities of new inhibitors prior to their synthesis.

Objectives: This study was aimed to generate new candidate small inhibitory molecules against Bcl-xL by using G-QSAR analysis of known Bcl-xL inhibitors.

Methods: In the present study, we used group-based QSAR (G-QSAR)—a novel fragment-based method—to develop QSAR models from known Bcl-xL inhibitors. A set of Bcl-xL inhibitors adopted from extant literature was fragmented into three common fragments, and a pool of 214 descriptors was calculated for each one.

Results: Two models were obtained by using different combination of variable selection and model building method; stepwise-multiple linear regression (STP-MLR) and simulated annealing-multiple linear regression (SA-MLR). STP-MLR was found to be the best mode, with $r^2 = 0.80$, $q^2 = 0.70$ and predictive $r^2 = 0.87$.

Conclusion: The G-QSAR results indicate that the generated models are statistically significant and can be used for design and generation of new potent inhibitors.

Keywords: G-QSAR, Bcl-xL inhibitors, 2D descriptors, Simulated annealing, Multiple regression method.

INTRODUCTION

Bcl-2 family of proteins plays a pivotal role in cancer survival by suppressing the pro-apoptotic proteins, increasing the cancer cell growth rate and increasing chemo resistance [1].

Bcl-xL is one of the Bcl-2 family members, found to be over-expressed in many lung cancer cases [2]. It is primarily composed of seven helices α1 to α7, whereby the hydrophobic binding pocket is formed by helices α5 and α6 [3]. Various studies show that targeting Bcl-xL by small molecules results in the arrest of cancer cell growth [4-6]. Various Bcl-xL small molecules inhibitors with potential biological activity against cancers have been designed and synthesized. In addition, computer aided drug design (CADD) approaches, including conventional 2D and 3D QSAR methods, have been performed and extensively utilized to obtain insights on improving the biological activities of Bcl-xL inhibitors [7, 8]. 3D QSAR applies the QSAR theory with consideration of three dimensional structures and protein ligand binding, whereby it depends on the conformers for ligand alignment [9]. However, limitations of 3D QSAR include difficulty in interpreting descriptors, limitation to only congeneric series of compounds and inapplicability of developed models to the design of new ligands [10]. An advanced method in computational drug design, group-based quantity structural activity relationship (G-QSAR) can be used to overcome the shortcomings of 3D QSAR. The method can also be used for non-congeneric molecules, as it does not depend on conformation and alignment of small molecules ;thus, the developed models can be used to design new ligands [10]. In this study, we used G-QSAR to generate models from non-congeneric series of compounds that have been previously investigated in several studies. Here, the G-QSAR models were used to successfully predict the biological activity of Bcl-xL inhibitors and generate highly potent Bcl-xL inhibitors. Two best models (SA-MLR and STP-MLR) were obtained by using different combinations of data and variable selection methodologies as well as a model-building method.

MATERIALS AND METHODS

GQSAR- Data mining and Fragmentation

Bcl-xL inhibitors were sourced from Binding Database website. This is a free database that provides measured binding affinities of interactions between small drug-like molecules and proteins considered to be drug targets [11, 12]. Initially, 321 compounds from various literature sources [13-22] were downloaded in SDF format and subsequently filtered based on their drug-like properties according to Lipinski rules of five [23]. The final number of compounds used for G-QSAR after filtration was 128. However, in order to develop statistically significant models, different compound combinations were used in the training and test sets.

Molecular Design Suite (VLifeMDS 4.2, from VLife Sciences Technologies Pvt. Ltd., India) was used to perform G-QSAR. The 3D compound structures were optimized by conjugate gradient method of Merck Molecular Force Field (MMFF) with 0.01 kcal/mol value of convergence in order to minimize their energy [24]. Based on the structures, the compounds were classified in different scaffolds, including [1,3]dioxolo[4,5-i]phenanthridine; 2,3-dihydro-1-benzofuran-3-one; 3,4,5-trihydrocyclopenta[c]quinolines; benzene-1,2-diol; carbamoylmethyl 2-hydroxy ethanoate; N-(benzenesulfonyl) benzamide; naphthaline-1,6,7-triol and 15-thia-12,17-diazatetracylo[8.7.0^{1,10}.0^{2,7}.0^{12,17}]heptadeca-1(10),2(3),4,6,16-penten-13-one. A common fragmentation pattern was identified

within all selected molecules, which included three fragments for each molecule (as indicated in Fig 1)

i. R1: the terminal region of molecule attached to R2

ii. R2: the part of the molecule that links R1 and R3. It can be single or multiple rings or single rigid chain

iii. R3: The terminal region of molecule attached to R2

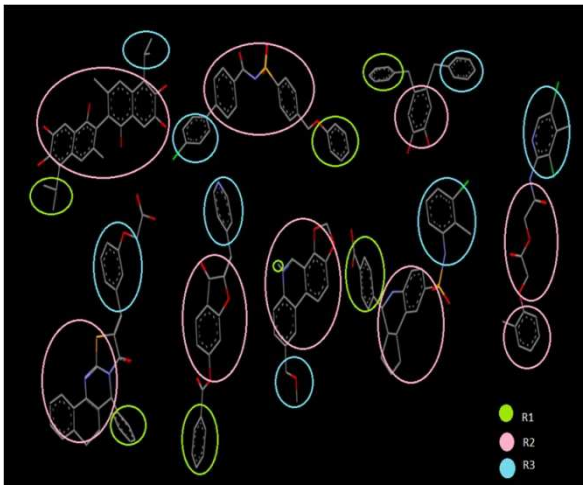

Fig. 1: Fragmentation pattern of the compounds

Molecular Descriptor Calculation

The calculations performed included 462 molecular descriptors, including physicochemical, retention index (chi), atom count, path count, estate numbers, atomic valence connectivity index (chiV), electro-topological, polar surface area, etc. All descriptors with constant values amongst the dataset were deleted, resulting in final 346 unique descriptors that were subsequently subjected to QSAR analysis.

Creation of Training and Test Sets

Sphere exclusion method was used for creation of training and test sets. For both SA-MLR and STP-MLR models, sphere exclusion values were 1.6. To further gain confidence that the training and test set have uniform representation of molecules and examine the reliability of the data selection method, different statistical parameters (unicolumn statistical analysis) of each generated set (maximum value, minimum value, average and standard deviation) were calculated [25].

Variable Selection Method

Robust model generation is contingent on wise selection of variables and model building methods. In this study, simulated annealing (SA) and stepwise-forward (STP) methods that have the ability of selecting multiple variables were used for variable selection, coupled with multiple linear regressions (MLR) for model building.

Simulated Annealing

Simulated annealing (SA) is a virtual process that simulates the real annealing technique, whereby the material is heated and cooled by controlling temperature in order to increase the size of its crystals and reduce their defects which depend on the thermodynamic free energy [26]. It is a global and iterative combinatorial optimization method that does not concur with the first encountered variable configuration. Simulated annealing mimics the physical process in which the system is melted at a very high temperature and then cooled slowly until a steady state is reached [27]. It was applied to select the suitable independent variable combinations from the calculated descriptors pool. In short, system points (descriptors) and cost functions (configuration energy) configuration were optimized [26]. The configuration of the system points was signified by Boltzaman probability factor of distribution, which correlated to the configuration energy (E) and applied temperature (T) [28], as it is known that lowering the temperature leads to the decrease in the system E states. During the simulation, at a certain temperature, a population of subset of descriptors (problem configurations) is generated and the iterative process is continued until an optimal solution, defined by Metropolis algorithms, is found [26]. In this procedure, one descriptor was randomly displaced and energy state difference between two configurations (ΔE= E_new - E_old) was calculated. Thus, when the difference was less or equal to zero (ΔE ≤ 0), the new configuration was accepted and used as starting point for the next iteration. Furthermore, when the energy differences between the two configurations was more or equal to one ΔE ≥ 0, the new descriptor was used for new displacement[29]. In the QSAR context, squared correlation coefficient of the regression (r^2) serves as cost function, while selected descriptors in final equation of the QSAR model represent the system configuration. The aim of QSAR is identifying the descriptor combination with high r^2 value.

Stepwise-Forward

Stepwise-forward variable selection method allows the development of a model through modifying the specifications of the trial model, one independent variable at a time, by adding or removing of the predictor variable using stepping criteria, F. This procedure was repeated until no more variable was left outside the mode created. Stepwise forward starts without predictors (descriptors) in the model. The value of the calculated r^2 is used to evaluate the available descriptors, so that increasing the number of descriptors will lead to an increase in the value of r^2, until it meets the statistical criteria for the inclusion into the model, i.e. the significance level for the increase in the r^2 produced by addition of the descriptor.

The analysis thus terminates when there is no predictor (descriptor) that meets the preset criteria. Second step of this method continues if predictor is being added and involves the re-evaluation of available predictors that are not yet incorporated into the model. Once the predictor that increased r^2 the most is identified, it is added and this process is repeated until no more predictors that satisfy the criteria to be in the entry exist [30].

Model Building

Both models were created by using multiple linear regression method (MLR). MLR estimates the values of the regression coefficients by applying the least squares curve fitting method[31]. Regression equation is given by

$$\mathbf{Y = b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + c}$$

where **Y** is the dependent variable, the $\mathbf{b_i}$ are regression coefficients for corresponding

X (independent variable) and **c** is a regression constant or intercept[31].

Finally, two best models obtained have the combinations of SA-MLR and STP-MLR. Considering the number of molecules and complexity of the developed models, several descriptors were selected from the initial descriptor pool.

Model Evaluation and Validation

The developed QSAR models were evaluated and validated internally and externally in order to determine their robustness. The models were evaluated based on n, number of molecules; k, number of variables; and r^2, coefficient determination. Leave-one-out internal validation was carried out, yielding the value of q^2 and for external validation. Moreover, compounds in the test set were predicted by using models generated by the training set and this is represented by the value of pred_r^2. Finally, estimated standard error of the models and validation values were calculated and the best models were selected based on their corresponding r^2 and q^2 values [10, 25].

RESULTS AND DISCUSSION

This G-QSAR study produced two best models based on the statistical significance, namely SA-MLR and STP-MLR models. These models were used to provide insight into the role of molecular properties on the biological activities of Bcl-xL inhibitors.

Statistical Analysis of the Models

As filtration of the compounds using the rule of five eliminated some compounds, 126 compounds were left. The training and test sets selection of compounds for each model varied in terms of dissimilarity value and resulted in different number of sets. Activity prediction and residual between the two values are given in Table 1a and Table 1b for SA-MLR model and Table 2a and Table 2b for STP-MLR model.

In order to obtain statistically significant values for SA-MLR and STP-MLR models, 6 and 2 outliers were deleted respectively, as their values were too far from the norm, which can affect the real statistics of the dataset. These outliers are x/y relation outliers, which were detected by comparing actual (x) and predicted activity (y) based on the two models [32]. Distribution of the molecules in both sets followed good criteria of data selection and this is proven by uni-column statistics for both training and test sets (Table 3).

Table 1: Unicolumn Statistics

	SA-MLR (SE 1.6)		STP-MLR (SE 1.6)	
	Training	**Test**	**Training**	**Test**
Average	1.245	1.862	1.245	1.922
Maximum	3.201	2.706	3.201	2.706
Minimum	0.963	0.983	0.963	0.963
Std. Dev	0.486	0.542	0.486	0.505
Sum	138.205	20.482	138.205	28.834

Firstly, the averages of the respective test sets for the two models were slightly higher than the training set average, indicating the presence of a greater number of active than inactive molecules. Secondly, activity average and standard deviation of both sets were very similar, implying that activity is equally distributed in both training and test sets. Thirdly, the sum of the inhibitory activity of the training set is greater than that of the test set, which suggests that all representative points of training set are well distributed within the entire data (Fig 2a and Fig 2b). A further indication that the test set lies within the training set is the fact that the maximum value in the training set was greater than the maximum in the test set and the minimum value of the test set was greater than the corresponding value in the training set [10].

Two models developed in this study are statistically significant with $\alpha <$ 0.000001 and parameters of each model are presented in Table 4.

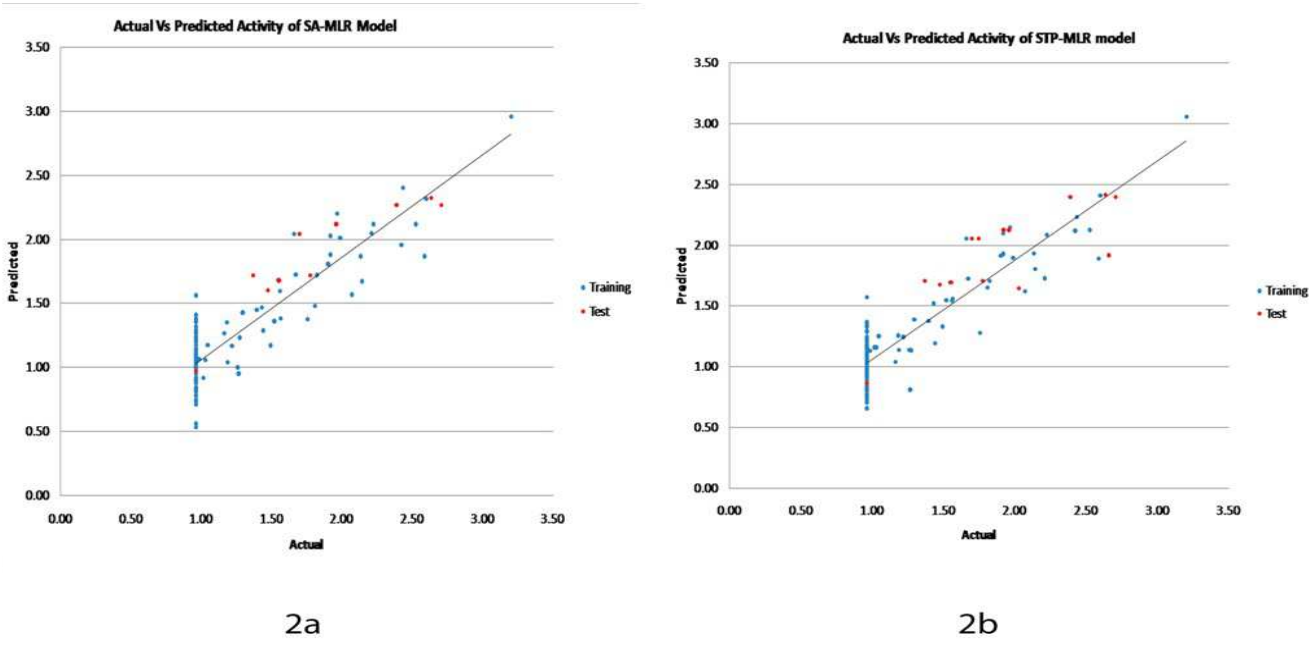

Fig. 2a: Fitness plot for SA-MLR model (actual versus predicted activity), 2b; Fitness plot for STP-MLR model (actual versus predicted activity).

Table 2: Statistical Analysis of Model SA-MLR and STP-MLR

	SA-MLR	STP-MLR
training/test set	111/11	111/15
Degree of freedom	95	93
r2	0.7768	0.79860
q2	0.6766	0.7033
F test	22.0390	21.6904
r2 se	0.2472	0.2373
q2 se	0.2976	0.288
pred_r2	0.9113	0.8708
pred_r2se	0.2514	0.3105

Model SA-MLR

SA-MLR model was built by combination of simulated annealing (for variable section) and multiple linear regression method (for model building).This model can be expressed mathematically using the equation given below.

PIC = 1.3796 R1-chiV3Cluster - 0.0108 R1-PolarSurfaceAreaIncludingPandS - 0.0914 R1-SaaOE-index - 0.3423 R1-SdssCE-index -0.1383 R1-SsCH3count - 0.2156 R1-SsFcount + 0.1277 R2-chi3 - 0.5757 R2-SaaNHcount + 0.4926 R2-SdScount - 0.0860 R2-SssNHE-index - 0.1539 R3-chi4pathCluster + 0.2167 R3-SaaNHE-index + 0.4595 R3-SaasN(Noxide)count - 0.3001 R3-SdsNcount + 0.3636 R3-SsClE-index + 1.2152

SA-MLR model is considered statistically significant with r^2 value of 0.78, q^2 of 0.68 for internal validation and external validation (pred_r^2) of 0.91. Fifteen descriptors were selected to describe this model, as shown in Fig 3, where 40% (6 descriptors) originated from fragment R1, four were from R2 and the remaining five descriptors from R3.

Model STP-MLR

PIC = 0.3055 R1-6ChainCount - 0.0466 R1-chi0 - 0.0036 R1-PolarSurfaceAreaExcludingPandS - 0.0439 R1-SaaOE-index + 0.1198 R1-SaasCE-index -0.3073 R1-SdssCE-index + 0.1348 R2-chi2 -0.1255 R2-SaaaCcount - 0.6744 R2-SaaNHcount + 0.9290 R2-SdSE-index - 0.1033 R2-SssNHE-index - 0.1678 R2-SulfursCount -0.0787 R3-chi3 +0.8843 R3-SaaNHcount +0.4114 R3-SaasN(Noxide)count - 0.0719 R3-SdsNE-index + 0.3721 R3-SsClE-index + 1.1083

STP-MLR is the second model obtained in this work and was developed through the combination of stepwise forward and multiple linear regressions. Model's equation indicates that there is approximately 80% (r^2: 0.80) of total variance in training set with internal validation (q^2) of 0.70 and external validation (pred_r^2) of 0.73. These statistics confirm that this model is significant and can be used to generate potential bcl-xL inhibitors.

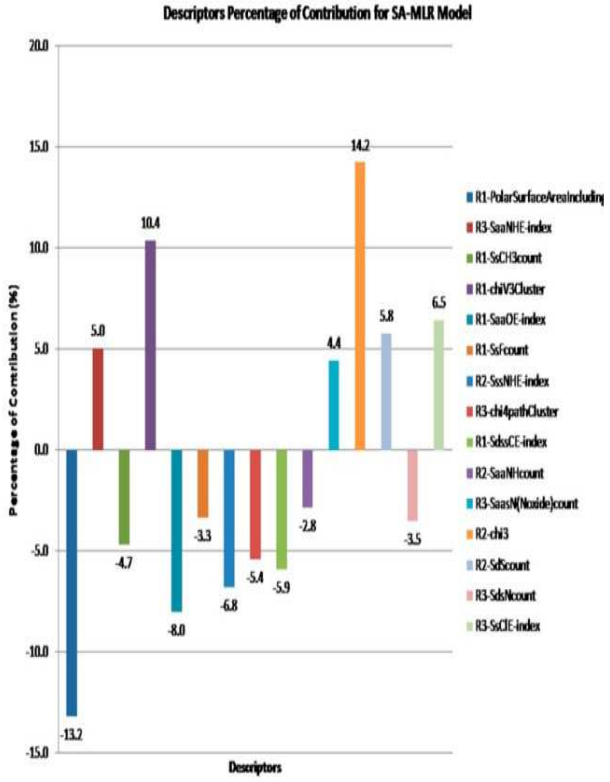

Fig. 3: Contribution plot of SA-MLR

Randomization test (Alpha Rand R^2: 0.0000) shows with > 99.9999999% confidence that the model was generated through QSAR modeling, rather than randomly. The fragments R1 and R2 contributed by 35% each, while fragment R3 contributed the remaining 30% of the activity (Fig 4).

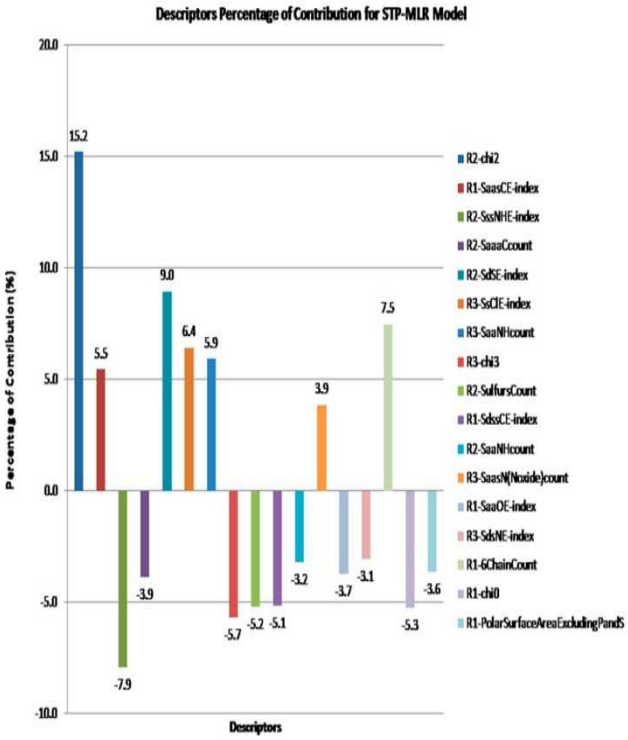

Fig. 4: Contribution plot of STP-MLR

Description of the Models' Descriptors

Electrotopological State Index

Electrotopological state indices (E-state) describe the electronic character and the topological environment of a skeletal atom in the molecule [33]. E-state indices of a certain atom in the molecule provide information on the electronic state of the atom (intrinsic state, I), which depends on π bonds, lone pair electrons and σ bonds that reflect quantitative availability of valence electrons for ligand-target interactions. E-index also explains the perturbation of the intrinsic state of a specific atom (ΔI) by all other atoms in the molecule, taking into consideration the distance between the atoms [34]. E-state indices can be calculated using the equation below [35, 36].

$$S_i = I_i + \Delta I_i = I_i + \sum_{j=1}^{A} \frac{I_i - I_j}{(d_{ij} + 1)^k}$$

Three nitrogen-based E-state descriptors, namely sssNHE-index, SaaNHE-index and SdsNE-index, were found to be responsible for activity variation in the two models. SssNHE-index can be defined as the E-state index of –NH group, which is attached to two single bonds. This descriptor contributes negatively to STP-MLR model and SA-MLR with ~-6.8% and ~-7.9% respectively. Therefore, it is not important for the model activity; nonetheless, decreasing the descriptor index at its respective fragments can improve the biological activity of Bcl-xL inhibitors. Fragment R2 in compound 291 (Fig 5) shows this descriptor in which N1 and N4 atoms in piperazine ring are attached to two single bonds. Thus, it is recommended that the E-state index of this atom be reduced for better inhibition. SaaNHE-index is electrotopological index of –NH group binding to two aromatic bonds. This descriptor contributes 5.0% to the activity in SA-MLR model. Thus, the presence of N-

containing aromatic ring (such as indole, pyrrole and indazole) at fragment R3 is important for biological activity. For example, fragment R3 in compound 138 (Fig 5) shows this descriptor, which can be increased by replacing the 1,3-benzodiazole ring with 3H-indazole in order to decrease the distance between two nitrogens (d_{ij}) for improved activity. Lastly, SdsNE-index describes the electrotopological indices of Nitrogen attached to one double bond and one single bond. This descriptor contributes negatively with ~-3.1% to STP-MLR model (fragment R3) and is thus not crucial for the model activity. However, we can also use this negative contributor in order to improve compound's activity. For example, in compound 298 (Fig 5), N1 in pyridazine-3-one at fragment R3 has this descriptor. To improve the activity of this compound, we suggest replacing this pyridazine-3-one with pyrazine-2-one or pyridine-2-one.

Electrotopological indices of carbon atom, such as SaasCE-index and SdssCE-index, are other types of contributing descriptors within this subclass. SaasCE-index signifies the electrotopological indices of carbon attached to two aromatic bonds and one single bond. It contributes to STP-MLR model with ~5.5%. Based on the positive contribution of SaasCE-index, it can be concluded that the activity of compounds can be improved by increasing the value of this descriptor in respective fragments. For instance, activity of compound 165 (Fig 5), which has two instances of SaasCE-index at fragment R1, can be improved by adding more substituent to the benzene ring. This process will increase the value of this descriptor leading to improvement in the activity. In addition, due to its rigid hydrophobic surface, the R1 fragment containing this descriptor can contribute to hydrophobic interaction with the ligand binding pocket. SdssCE-index signifies the electrotopological state indices for number of carbon atom, connected to one double and two single bonds and it contributes negatively to both SA-MLR and STP-MLR models with ~-5.9% and ~-5.1%, respectively. This indicates that sp^2 hybridized carbon (such as alkenes, carbonyl and imine) is not recommended. Fragment R1 in compound 303 (Fig 5) shows three examples of this descriptor, whereby, in order to improve the activity, it is suggested to replace carbonyl group with hydroxyl or methyl group at this fragment. The descriptor SaaOE-index can be defined as the electrotopological state index for number of oxygen atom connected with two aromatic bonds. It contributes negatively to SA-MLR and STP-MLR model, with ~-8% and ~-3.7% respectively. Oxygen atom plays an important role in ligand-target interactions as hydrogen bond acceptor. However, this descriptor contributes negatively which means its presence at fragment R1 is not important for the model activity.

SsClE-index represents the electrotopological indices for number of chlorine attached to a single bond. It is important for model activity, contributing ~6.5% to SA-MR and ~6.4% to STP-MLR model. For example, addition of more chlorine atoms that bind to carbon by single bond in benzene ring of fragment R3 of compound 109 (Fig 5) is suggested as an effective mechanism through which the activity of these models can be improved. Finally, SdSE-index explains the electrotopological indices of sulphur bound to one double bond, which contribute ~9% to STP-MLR model, indicating that the addition of sulphur that can bind to double bond is favorable to the activity. For instance, in fragment R2 of compound 117 (Fig 5), improved activity can be achieved by replacing carbonyl group at C3 of the thiolane ring can with sulphanylidene group.

Electrotopological State Number

In addition to the electrotopological state index, electrotopological state number is another subclass of physicochemical descriptors that gives the total number of certain atom or atom group and their binding properties in the molecule [37]. SaaNHcount contributes negatively in fragment R2 of both models, with ~-2.8% and ~-3.2% for STP-MLR and SA-MLR, respectively. SaaNHcount is the total number of –NH group attached to two aromatic bonds. Pyrazole ring in compound 237 (Fig 6) shows the descriptor, indicating that this compound's activity can be improved by replacing this ring with cyclopenta-1, 3-diene for better inhibition. On the other hand, in model STP-MLR, SaaNHcount descriptor contributes ~5.9% to the activity in fragment R3. Thus, for example, the biological activity of compound 237 can be improved by replacing chlorobenzene with pyrazole ring at STP-MLR model fragment R3.

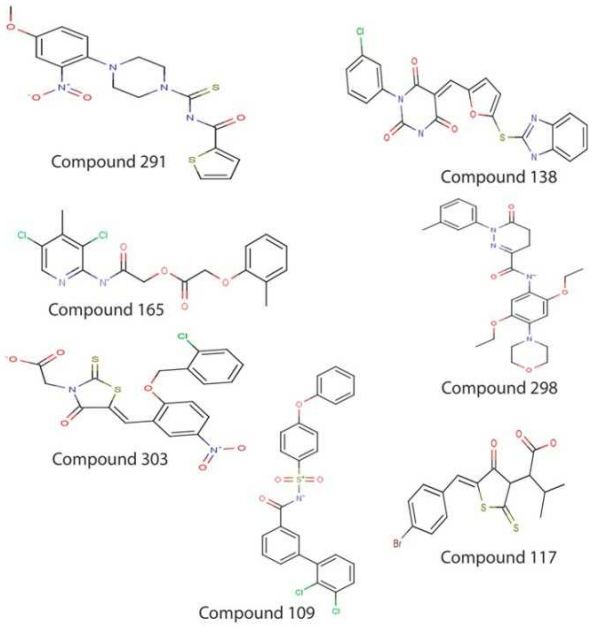

Fig. 5: Examples of compounds representing electrotopological index descriptors

SaasN(Noxide)count explains the electrotopological state indices for the number of nitro-oxide group connected via two aromatic and one single bond. This descriptor originates from fragment R3 and contributes to biological activity in SA-MLR and STP-MLR by ~4.4% and 3.9% respectively. It could be found in an aromatic ring with nitrogen attached to oxygen via single bond and become N-oxide ring, such as pyridine N-oxide. Furthermore, by adding more structure, such as pyridine N-oxide to fragment R3, biological activity can be improved in both models.

Fragment R3 of model SA-MLR has SdsNcount descriptor that explains the total number of nitrogen atom binding to one single bond and one double bond. The descriptor contributes to the activity by ~-3.5% and is thus not favorable for model activity. Hence, its removal will improve the biological activity.

SdScount from fragment R2 of SA-MLR signifies the total number of sulphur atom connected with one double bond and it contributes ~5.8% to the activity. Sulphur can be potential hydrogen bond acceptor because of its electronegativity. Therefore, owing to the presence of sulphur in this fragment, hydrogen bond might be formed between the compound and the active site of Bcl-xL. For instance, in fragment R2 of compound 117 (Fig 6), in order to improve activity, carbonyl group at C3 of the thiolane ring can be replaced with sulphanylidene group.

SsCH3count signifies the total number of –CH3 group attached to one single bond and contributes by has ~-4.7% in R1 of SA-MLR. Its negative contribution indicates that branching is not favorable for the activity. Another descriptor found in R1 of SA-MLR is SsFcount describes the total number of fluorine atoms connected via a single bond. It contributes ~-3.3% to activity, indicating that the presence of fluorine is detrimental to the activity. SaaaCcount descriptor was found in fragment R2 of STP-MLR model and it signifies the total number of carbon binding via three aromatic bonds. This descriptor is found in fused aromatic chains, such as indole, indazole and purine, where the carbon between these fused rings is attached to three aromatic bonds. This descriptor contributes ~-3.9% to activity and is represented in fragment R2 of compound 106 (Fig 6). Replacement of naphthalen-1-ol with phenol can be suggested in this case, as this might improve the activity of the compound.

Chi (Retention Index)

Chi0, chi2 and chi3 descriptors represent the retention index by certain order derived directly from gradient retention times. This descriptor is referring to Kovats retention index (*I*), which is a characteristic of a gas chromatographed compound on a given column at a definite temperature [37]. Chi0, chi2 and chi3 signify retention index by zero, second and third order, respectively. Chi0 from fragment R1 contributes negatively to STP-MLR model (~-5.3%), while Chi2 from fragment R2 of the same model contributes by ~15.2% and is thus very important to the model activity.

Finally, Chi3 from fragment R2 provides significant and positive contribution (~14.2%) to model SA-MLR. The retention time indicates the hydrophobicity of the compound, whereby the higher the value of *I*, the high the hydrophobicity of the compound [37]. Therefore, fragment R2 of both models might interact with hydrophobic pocket of Bcl-xL protein [38].

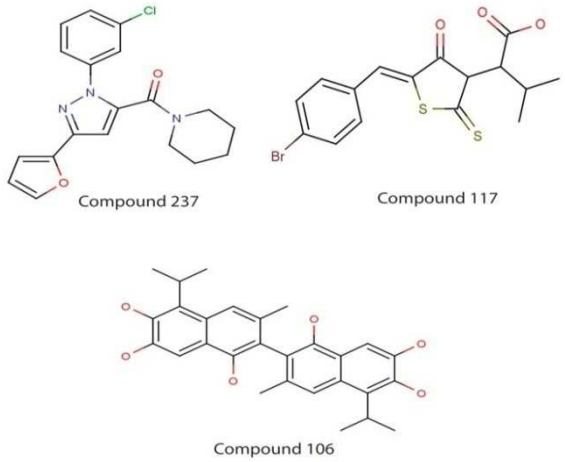

Fig. 6: Examples of compounds representing electrotopological state number descriptor

Polar Surface Area

Polar surface area is one of the molecular surface area descriptors that are very important in understanding the structure and chemical properties of a molecule, affecting its ability to bind ligands, solvents and other biological molecules [37]. Two descriptors on this molecular surface are PolarSurfaceAreaIncludingPandS and PolarSurfaceAreaExcludingPandS, which describe the part of molecular surface associated or excluded with sulphur and phosphorous, including their bonded hydrogen atoms affecting the hydrogen bonding ability of the molecule. Both descriptors contribute negatively to the activity, whereby PolarSurfaceAreaIncludingPandS from fragment R1 of SA-MLR contributes by ~-13.2% and PolarSurfaceAreaExcludingPandS from fragment R1 of STP-MLR provides contribution of ~-3.6%. Therefore, in order to improve compound's activity, removal of polar structure on fragment R1 is highly recommended. For instance, in compound 303 (Fig 7), carbonyl groups at fragment R1 can be replaced with hydroxyl group for better activity.

Cluster, Path Cluster and Chain Path Count

Cluster, Path Cluster and Chain Path Count are molecular graph descriptors. These descriptors depict the connectivity of the atoms within the molecules regardless of bond types, angles, torsion, geometry, etc. [37]. From both models, three descriptors of molecular graph can be identified, namely ChiV3cluster, 6ChainCount and Chi4pathCluster. ChiV3Cluster signifies the valence molecular connectivity index (MCI) of third order cluster. In this work, MCI was calculated from vertex degree δ of the atoms in H depleted molecular graph and is thus proposed to measure the molecular branching. ChiV3Cluster is the only positive contributor from fragment R1 in model SA-MLR, with ~10.4% contribution. Therefore, it is highly recommended to include this descriptor into its respective fragment, as it has the potential for improvement in biological activity of Bcl-xL inhibitors. 6ChainCount, which signifies total number of six-member rings in a compound, contributes to the activity by 7.5%, indicating that addition of any six-member rings, such as benzene or pyridine, to fragment R1 will improve the biological activity. Chi4pathCluster from fragment R3 signifies molecular connectivity index of fourth order path cluster.

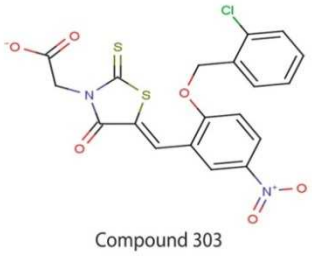

Fig. 7: Compound 303 represents polar surface area descriptor

Path cluster is a description that refers to the molecular graph with the vertex degree > 2 in the subgraph, while fourth order indicates the path number of length 4[37]. This descriptor contributes negatively to the activity of SA-MLR model (~-5.4%), thus its presence is not important. Compound 252 (Fig 8) has high value of Chi4pathCluster at fragment R3. Therefore, in order to improve its activity, it is recommended to replace naphthalene-2-ol with phenol ring, thus decreasing the number of possibilities of having fourth order path structure.

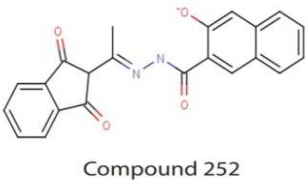

Fig. 8: Compound 252 represents Chi4pathCluster descriptor

Element Count

SulphurCount signifies number of sulphur atoms in the fragment. Increasing the number of sulphur atom in the same fragment does not improve the activity, as this descriptor contributes by ~-5.2%. For instance, replacement of thiophene ring with pyrrole or furan at fragment R2 of compound 228 (Fig 9) will help improve the activity of this compound.

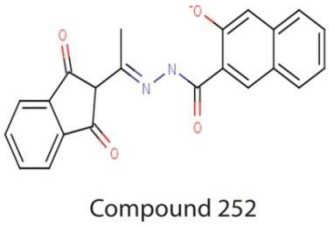

Fig. 9: Compound 228 represents SulphurCount descriptor

CONCLUSION

The aim of this work was to obtain clues for designing new molecules using the novel G-QSAR method. The statistically

significant GQSAR models developed as a part of this study can be used for the prediction of the activity of an external set of compounds. The SA-MLR and STP-MLR models described here will be further used to generate compounds with improved biological activity against Bcl-xL protein with potential therapeutic uses in the treatment of various cancers that depend on Bcl-xL for survival.

ACKNOWLEDGEMENT

Authors would like to express their gratitude to the International Islamic University Malaysia for funding this research under the research matching grant RMGS 10-003-0013.

REFERENCES

1. Reed, J. Double identity for proteins of the Bcl-2 family. Nature. 1997; 387,: 773-6.
2. Hanahan, D., Weinberg, R.A., Francisco, S. The Hallmarks of Cancer Review University of California at San Francisco. 2000; 100: 57-70.
3. Muchmore, S.W., Sattler, M., Liang, H., Meadows, R.P., Harlan, J.E., Yoon, H.S., et al. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. Nature. 1996; 381: 335-41.
4. Emi, M., Kim, R., Tanabe, K., Uchida, Y., Toge, T. Targeted therapy against Bcl-2-related proteins in breast cancer cells. Breast Cancer Research. 2005; 7: R940 - R52.
5. Kang, M.H., Reynolds, C.P. Bcl-2 inhibitors: targeting mitochondrial apoptotic pathways in cancer therapy. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009; 15: 1126-32.
6. Hussain, S., Plückthun, A., Allen, T.M., Zangemeister-Wittke, U. Chemosensitization of carcinoma cells using epithelial cell adhesion molecule–targeted liposomal antisense against bcl-2/bcl-xL. Molecular Cancer Therapeutics. 2006; 5: 3170-80.
7. Almerico, A.M., Totone, M., Lauria, A. 3D-QSAR pharmacophore modeling and in silico screening of new Bcl-xl inhibitors. European Journal of Medicinal Chemistry. 2010; 45: 4774-82.
8. Srivastava, a.K., Srivastava, A., Jaiswal, M., Nath, A. QSAR studies on anti-apoptotic Bcl-2 protein inhibitors. Journal of Saudi Chemical Society. 2009; 13: 259-62.
9. Kubinyi, H. 3D Qsar in Drug Design: Volume 1: Theory Methods and Applications, Springer, 1993.
10. Ajmani, S., Agrawal, A., Kulkarni, S.A. A comprehensive structure–activity analysis of protein kinase B-alpha (Akt1) inhibitors. Journal of Molecular Graphics and Modelling. 2010; 28: 683-94.
11. Liu, T., Lin, Y., Wen, X., Jorissen, R.N., Gilson, M.K. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic acids research. 2007; 35: D198-D201.
12. Chen, X., Lin, Y., Liu, M., Gilson, M.K. The Binding Database: data management and interface design. Bioinformatics. 2002; 18: 130-9.
13. Bruncko, M., Oost, T.K., Belli, B.A., Ding, H., Joseph, M.K., Kunzer, A., et al. Studies Leading to Potent, Dual Inhibitors of Bcl-2 and Bcl-xL. Journal of Medicinal Chemistry. 2007; 50: 641-62.
14. Kitada, S., Leone, M., Sareth, S., Zhai, D., Reed, J.C., Pellecchia, M. Discovery, Characterization, and Structure–Activity Relationships Studies of Proapoptotic Polyphenols Targeting B-Cell Lymphocyte/Leukemia-2 Proteins. Journal of Medicinal Chemistry. 2003; 46: 4259-64.
15. Leverrier, A.l., Martin, M.-T.r.s., Servy, C., Ouazzani, J., Retailleau, P., Awang, K., et al. Rearranged Diterpenoids from the Biotransformation of ent-Trachyloban-18-oic Acid by Rhizopus arrhizus. Journal of Natural Products. 2010; 73: 1121-5.
16. Siskou, I.C., Rekka, E.A., Kourounakis, A.P., Chrysselis, M.C., Tsiakitzis, K., Kourounakis, P.N. Design and study of some novel ibuprofen derivatives with potential nootropic and neuroprotective properties. Bioorganic & Medicinal Chemistry. 2007; 15: 951-61.
17. Sleebs, B.E., Czabotar, P.E., Fairbrother, W.J., Fairlie, W.D., Flygare, J.A., Huang, D.C.S., et al. Quinazoline Sulfonamides as Dual Binders of the Proteins B-Cell Lymphoma 2 and B-Cell Lymphoma Extra Long with Potent Proapoptotic Cell-Based Activity. Journal of Medicinal Chemistry. 2011; 54: 1914-26.
18. Wang, G., Nikolovska-Coleska, Z., Yang, C.-Y., Wang, R., Tang, G., Guo, J., et al. Structure-Based Design of Potent Small-Molecule Inhibitors of Anti-Apoptotic Bcl-2 Proteins. Journal of Medicinal Chemistry. 2006; 49: 6139-42.
19. Wei, J., Kitada, S., Stebbins, J.L., Placzek, W., Zhai, D., Wu, B., et al. Synthesis and Biological Evaluation of Apogossypolone Derivatives as Pan-active Inhibitors of Antiapoptotic B-Cell Lymphoma/Leukemia-2 (Bcl-2) Family Proteins. Journal of Medicinal Chemistry. 2010; 53: 8000-11.
20. Wei, J., Stebbins, J.L., Kitada, S., Dash, R., Placzek, W., Rega, M.F., et al. BI-97C1, an Optically Pure Apogossypol Derivative as Pan-Active Inhibitor of Antiapoptotic B-Cell Lymphoma/Leukemia-2 (Bcl-2) Family Proteins. Journal of Medicinal Chemistry. 2010; 53: 4166-76.
21. Wendt, M.D., Shen, W., Kunzer, A., McClellan, W.J., Bruncko, M., Oost, T.K., et al. Discovery and Structure–Activity Relationship of Antagonists of B-Cell Lymphoma 2 Family Proteins with Chemopotentiation Activity in Vitro and in Vivo. Journal of Medicinal Chemistry. 2006; 49: 1165-81.
22. Wei, J., Kitada, S., Rega, M.F., Stebbins, J.L., Zhai, D., Cellitti, J., et al. Apogossypol Derivatives as Pan-Active Inhibitors of Antiapoptotic B-Cell Lymphoma/Leukemia-2 (Bcl-2) Family Proteins. Journal of Medicinal Chemistry. 2009; 52: 4511-23.
23. Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 2001; 46: 3-26.
24. Alzate-Morales, J.H., Vergara-Jaque, A., Caballero, J. Computational Study on the Interaction of N1 Substituted Pyrazole Derivatives with B-Raf Kinase: An Unusual Water Wire Hydrogen-Bond Network and Novel Interactions at the Entrance of the Active Site. Journal of Chemical Information and Modeling. 2010; 50: 1101-12.
25. Ajmani, S., Jadhav, K., Kulkarni, S.A. Group-Based QSAR (G-QSAR): Mitigating Interpretation Challenges in QSAR. QSAR & Combinatorial Science. 2009; 28: 36-51.
26. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P. Optimization by Simulated Annealing Science 1983; 220: 671-80.
27. Klein, R.W., Dubes, R.C. Experiments in projection and clustering by simulated annealing. Pattern Recognition. 1989; 22: 213-20.
28. Ingber, L. Simulated annealing: Practice versus theory. Mathematical and computer modelling. 1993; 18: 29-57.
29. Dueck, G., Scheuer, T. Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing. Journal of computational physics. 1990; 90: 161-75.
30. Armstrong, N.A., James, K.C. Pharmaceutical experimental design and interpretation, Informa Healthcare, 1996.
31. Aiken, L.S., West, S.G., Pitts, S.C. Multiple linear regression. Handbook of psychology. 2003.
32. Furusjö, E., Svenson, A., Rahmberg, M., Andersson, M. The importance of outlier detection and training set selection for reliable environmental QSAR predictions. Chemosphere. 2006; 63: 99-108.
33. Huuskonen, J. QSAR modeling with the electrotopological state indices: predicting the toxicity of organic chemicals. Chemosphere. 2003; 50: 949-53.
34. Kier, L.B., Hall, L.H. The E-state as an extended free valence. Journal of chemical information and computer sciences. 1997; 37: 548-52.
35. Hall, L., Kier, L. The E-state as the basis for molecular structure space definition and structure similarity. Journal of chemical information and computer sciences. 2000; 40: 784-91.
36. Hall, L.H., Mohney, B., Kier, L.B. The electrotopological state: structure information at the atomic level for molecular graphs. Journal of Chemical Information and Modeling. 1991; 31: 76-82.
37. Todeschini, R., Consonni, V. Handbook of Molecular Descriptors, WILEY-VCH Verlag, 2000.
38. Zhang, Y.-H., Bhunia, A., Wan, K.F., Lee, M.C., Chan, S.-L., Yu, V.C.-K., et al. Chelerythrine and Sanguinarine Dock at Distinct Sites on Bcl-XL that are Not the Classic BH3 Binding Cleft. Journal of Molecular Biology. 2006; 364: 536-49.