3D Biostructure Visualisation Using 4D-QSAR Model for Substitute Ureas Binding at the Raf-1 Kinase Receptor Site

Ekrem Mesut SU, Burçin TÜRKMENOĞLU, Yahya GÜZEL

Department of Chemistry, Faculty of Science, Erciyes University, Kayseri, Turkey

Abstract: *To determine F, 4D-QSAR analysis of substitute ureas ligands was applied with Molecular Conformer Electron Topological (MCET), a ligand-based method. The biostructure was found from the 3D substructures of the selected conformers for five different compounds. The electronic properties of atoms presented in the Electron Topological Method (ETM) were used as the 3D structural descriptors (3DSDs) of the molecules. The Genetic Function Approximation (GFA) was used to analyze these descriptors playing a role in activity and to construct a model predicting binding affinity. We have used Atomic Fukui indices as local reactivity parameters for assessing the binding affinity of ligand towards receptor. To gain insights into the nature of nucleophilic and electrophilic of ligand with receptor sites, interaction regions have been defined. Model was validated by applying Leave One Out-Conformer Validation (LOO-CV). Using a pharmacophore structure together with Auxiliary Groups (AG) and Anti-Pharmacophore Shielding groups (APS), 4D QSAR model derived from nonlinear equation was established, then the Q^2 and R^2 possessing fairly values of; 0.71 and 0.60, were calculated respectively.*

Keywords: *Raf-1 Kinase, 4D-QSAR, LOO-CV, ETM, MCET.*

1. INTRODUCTION

Cancer is considered to be one of the major killer diseases worldwide. It is caused by mutations in critical genes that alter normal cell functioning. Kinases are involved in many critical biological signaling pathways essential for the cell cycle regulation. Activation of RAS–RAF- MEK–ERK (kinases) signal transduction pathway initiates a cascade of events that regulates cell growth, proliferation, and differentiation in response to growth factors, cytokines, and hormones [1]. If this pathway is constantly switched on, the cell proliferation can have damaging effects, resulting in cancer [2]. Raf-1 is a validated target for the treatment of cancer [3]. Inhibitors that effectively block the activity of such proteins (Raf-1) could be useful in the treatment of broad-spectrum cancers [4]. Amides [5]

have been demonstrated to inhibit Raf-1 kinase activity. The QSAR application on ureas for the inhibition of Raf-Kinase was determined, and new inhibitors of Raf1-Kinase were also developed using a rational design approach. Understanding the binding of ligands in the active site of a protein is difficult in the absence of a crystal structure. In our continuing effort towards this goal we studied on the chemical structures of ureas using the Molecular Conformer Electron Topological (MCET) method which was developed by us, and was described in our previous publication [6]. This sampling process in turn enables the construction of optimized dynamic spatial 4-Dimensional Quantitative Structure–Activity Relationship (4D-QSAR) models in the form of the 3D biostructure [7]. The 4D-QSAR paradigm has been successfully applied to a variety of chemical classes and biological interaction points [8]. This study was focused on finding a 4D-QSAR model that able to predict binding at the Raf-1 Kinase and even provides clues for mechanism of drug-receptor interaction.

2. MATERIAL METHOD

A series of ureas derivatives (76 compounds) with experimental biological activities was taken from the literature [2]. The sum of set molecules was randomly divided into a training set (n = 49) and a test set (n=27) for confirming 4D-QSAR model. The activity of studied compounds was shown in Table 1.

To perform molecular modeling using SPARTAN '08 (Wavefunction Inc., Irvine, CA, 2000) [9] molecular mechanics, conformer distribution and geometry optimization calculations under the setup menu were selected. For conformer distribution, automated conformational analysis methods were introduced by rotation around each single bond and by the change in the bond angle of atoms. Equilibrium geometry calculations of the conformations were performed using the Density-Functional method with 6-31 G*. After molecular dynamic simulations were completed in water, conformers with less than 2 kcal/mol in relative energy were selected and saved as MolFiles.

The MolFiles database for each conformer was transformed to Electron Topological Matrix (ETM) format using ETM Programmer (ETMP) [10-11]. ETM is the electronic structure and topology with a digital (matrix) form of a molecule. After the conformers for each compound is superimposed onto the template conformer, the best candidate conformer was selected from the conformers' pool. Atomic Fukui indices of the conformers were used as a 3DSDs. For developing 4D-QSAR model, Partial Least-Squares (PLS) regression in combination with GFA was used [12-13]. The molecular activities were computed through the proposed model [14]. Both ETMP (transforming the data to ETM) and MCET (computing the activity), in-house programs, were written in C#.

2.1. Unique Conformer Selection

The molecular structure of each chemical under investigation was represented by many conformers. The total number of all calculated conformers for each molecule was reduced to one hundred conformations. After some conformers, which were more likely to interact with the receptor, were selected similarly from the low energy conformers as those best matching the template conformer, the remaining conformers were excluded and also the duplicate conformers were eliminated.

Pha atoms, common structure for all active compounds under study, are used for the alignment of the conformer to position of a common template. After alignment of different Pha structures, remaining atoms of conformers were superimposed with respect to atoms belonging to a common template. Thus, some consistency was provided over the entire range of orientation at displayed conformers. 3-D similarity score between two compounds depends on the distance tolerance. If distance of corresponding atoms between conformers was less than the tolerance value (0.4 Å), atoms from the current conformer were superimposed onto atoms of template conformer, usually corresponding to a desired target conformation. Therefore, results of MCET are obtained from among the overall orientation, and characterized by a cross-validated correlation coefficient R2 (Q2) for interaction points. The idea underlying 4D-QSAR analysis was related to differences in the Boltzmann average spatial distribution of conformers with respect to the 3D pharmacophores [15]. The resulting activity is averaged over all the selected conformers of the molecule.

2.2. Electron Topological Matrix (ETM)

ETM is the electronic and geometric features obtained from direct quantum-chemical calculations, and information related to both the topological environment and electronic feature of atoms of a molecule. The electronic properties in the molecule is chosen (one at a time) from the atomic charges, Fukui function, HOMO/LUMO coefficients, interaction index (II), valence activities, polarizabilities, etc. These are 3DSDs derived from the 3D structure of the molecule[16]. 3DSDs are sensitive to the orientation and position of superimposed molecules. A detailed description of ETM was reported in our previous publication [10,15,17-19].

2.3. Multiple Structure Alignment Based on Pha

The main problems encountered in constructing a good 4D-QSAR model are related to improper alignment of molecules, uncertainties of Pha structure, greater flexibility of the molecules, and more than one binding mode of the ligands. Only one Pha is responsible for the activity under study, and is approximately the same for the active molecules and not present for the inactive molecules. The Pha and the selected descriptors were identified by a small number of matrix elements (Electron Topological Sub-Matrix: ETSM). The values of the matrix elements of Pha in ETSM for active molecules may vary from one conformer to another within some tolerances. Pha has the same sub-matrix for at least one conformer of the active molecules.. The algorithm in MCET searches for a common Pha structure over a set of molecules. To do this properly, the conformers of each molecule are compared with the template conformer, selected from the lowest energy conformer of the lead molecule as the starting structure. As a result of this comparison, the candidate structures and the key moieties can be obtained from the global minimum structure [17-18]. The conformers which were assumed to have Pha structure were aligned basing on this structure. The coordinate values of atoms in the conformer were fixed into a 3D Cartesian coordinate by the program using the invariant coordinates of the three-ordered atom alignment. MCET is an expert system in which the best common Pha can be selected [20]. Then, the conformers were automatically superimposed with the highest fit value.

2.4. Configuration of the Molecular Descriptors

Unlike traditional QSAR approaches, MCET used the 3DSDs instead of arbitrary descriptors. The descriptors using the atomic coordinates (x, y, z) of a molecule are

therefore called 3DSDs. 3D electronic descriptors are the most general characteristics of a molecule. The selection of descriptors is extremely helpful in correlating interactions from the molecular structure. Theoretically, nothing is better than 3D electronic descriptors and topology given by ETM, which represent molecular ability to interact with other systems [18]. ETM contained all the information about the possible action of the molecule.

The final biostructure consists of the Pha structure, and subsequently AG- and APS-groups, called local electronic reactivity descriptors. These groups were used for modeling biological activity, and presented in the ETSM whose matrix elements are defined from in ETM. Each Pha structure is normally used as the starting point for determining the values of the AG and APS sets, which give the positive and negative contribution to activity, respectively. It was possible to distinguish between the attractive and repulsive effects of the descriptors as the AG and APS groups. They describe the ability of atomic sites to take part in mostly specific interactions, and can also be used to assess the propensity of chemicals to take part in strong or weak interactions.

Fukui function given by Eq. 1 is the derivative of the electron density with respect to the total number of electrons N in the molecule, at a constant external potential on a point r in space around the molecule. It is used as the common descriptors to understand the local electrophilic and nucleophilic reactivities of compound and defined as follows.

$$f(\vec{r}) = \left[\frac{\partial \rho(\vec{r})}{\partial N}\right]_{\nu(r)} \quad (1)$$

Atomic Fukui indices assosiated with the hard and soft acid/base theory were used as indicators of chemical reactivity.

It was state that the regions of a molecule where the Fukui function is large are chemically softer than the regions where the fukui function is small. Atomic Fukui function from conceptual density functional theory can be applied to predict for nucleophilic and electrophilic reactivity of ligand binding sites.

To determine whether a nucleophilic or electrophilic attack is present within the ligand binding site, we examined on atom *condensed Fukui.* The binding affinity of ligand towards receptor is found using the local nucleophilic or electrophilic reactivities of compounds.

2.5. Genetic Function Approximation (GFA)

To select the best model which represents a possible biostructure, GFA combined with PLS was used. To perform data reduction, analysis was done between the observed dependent variable measures and the corresponding set of descriptor values [12-13]. The various atomic positions generated from five different structural molecules were used to form the trial interaction points for GFA model optimization. The subsets of the models were calculated using the Levenberg-Marquardt method in a GFA [13].

2.6. Partial Least-Square (PLS) Regression

PLS regression finds a few independent variables in the formulation of an activity model that most efficiently explain variation in both predicted and observed activity [12, 20-21]. 4D-QSAR models were created by applying PLS regression to the properties of atoms located in a set of predicted positions. The accuracy of the models was improved by increasing the number of PLS factors until over fitting started to occur. PLS regression is the most appropriate method to reduce the high number of independent variables produced from the fitted atoms of the superimposed conformers within tolerance values in each position. A set of the reduced independent variables subsequently correlated with the activity to derive the bioactive model [12].

Using PLS the leave-one-out cross-validation (LOO-CV) analysis was performed [20, 22]. Correlation coefficient q^2 [13] calculated from the training set was considered for further analysis. The predictive power of the MCET method was confirmed using a test set excluded during model development. The optimization, alignment and all other steps of these test set molecules were the same as those of the training set molecules and their activities were predicted using the model produced by the training set. The predictive correlation (r^2_{pred}) based on the test set molecules was computed.

2.7. QSAR Model Generation

MCET software was applied for the detection and interpretation of crucial interaction patterns. In the software, firstly, various candidate Pha structures were extracted by comparing the ETMs of all the conformers with the template conformer, and then the descriptor set in the detailed positions for each Pha was automatically created and visualized. A model consist of unique Pha or numerous AG and APS groups, which formed the biostructure. Models were generated from

the Pha, AG and APS, and the best model being the one with the best statistics was defined. A model which was chosen as the biostructure was a 3D description developed by specifying the distance (or bond length) and amount of electronic values such as atomic Fukui indices in the ETM. The biostructure might be generated from the superposition of the active molecules by means of their common features. Given a set of active molecules, the 3D approximate model generation of the biostructure involves three steps: (1) comparing and matching the molecules to identify the key Pha, (2) aligning three-ordered atoms (based on Pha structure) to superimpose the remaining atoms, and (3) analyzing the various positions to define an AG and APS set, the independent variables. The atoms forming the Pha are defined as a basic skeleton. In the presence of the Pha the activity of the molecule may be diminished (partially or completely) by APS which hinders its proper docking with the receptor, or it may be enhanced by AG which provides attraction between L-R. To determine the AG and APS descriptors, we had to examine the structures (conformations) of the superimposed molecules. Their influence could be used to parameterize the receptor points. This parameterization at the receptor was based on AG and APS described by the electronic and geometric values known from the ETM. We suggested a general scheme for quantitative evaluation to estimate their approximate role in the activity. The main idea (somewhat similar to that involved in Pha identification) was to describe each of them by means of structural and electronic parameters and to reveal their role using a minimization procedure, as is usually done in QSAR problems. Then by processing these descriptors for the training set in a comparison with the activities and performing PLS, we obtained the adjustable constants (κ_j) that represent each of the receptor binding parameters in the activity. Taking into account the contribution of AG and APS multiplied by these constants, we obtained a formula for the quantitative prediction of the bioactivity. First, we took into account that their contribution reduced or enhanced L-R binding by an amount E, which reduced (or increased) activity by a factor of exp(-E/kT). We denoted $S_{ni}=E_{ni}/kT$ and introduced the function S in a general way as follows:

$$S_{ni} = \sum_{j=1}^{N} \kappa_{j*} a_{ni}^{(j)} \qquad (2)$$

Where a_{ni}^{j} are the independent variables that describe the jth kind of AG or APS in the ith conformation of nth molecule, N being the number of expected interactions

in spite of the different position of scaffold, and κ_j are variational parameters relative to the interaction point of the receptor. The interaction between L-R was defined by multiplying descriptors of the ligand with the adjustable constant arising from the receptor site as given in Eq. 2. The magnitude of the weighting coefficients within the parameters indicates the relative importance of AG and APS in each position when determining activity. Thus, specific regions in the three-dimensional space, where AG and APS interactions were important, could be identified by superposition, and plotted to derive the pharmacophoric receptor maps. Using this function and taking into account the Boltzmann population of each conformation as a function of its energy and temperature T, we obtained the following general formula of activity: A_0 was a constant (see below), and for the nth molecule, m_n and m_n^{Pha} were the numbers of all selected conformations and of those possessing Pha, respectively.

$$A_n = A_0 \frac{\sum_{i=1}^{m_n^{Pha}} e^{-S_{ni}}\, e^{-E_{ni}/kT}}{\sum_{i=1}^{m_n} e^{-E_{ni}/kT}} \qquad (3)$$

In this formula, the S_{ni} value for conformations that have a Pha contributed to the activity, and these contributions were weighted in accordance with the relative numbers of conformations in the active molecules. These numbers decreased rapidly with the energy increasing of the conformation E_{ni} (at $E_{ni} \sim 2$ kcal/ mol the number of conformations became lower than a 0.02 part of those in the lowest conformation at $E_{n0} = 0$). In the next section we describe how we handled the multiconformation problem. To determine the A_0 constant, we chose a reference molecule (l) from the training set for which the activity was known and calculated A_l after Eq.3.

$$A_\ell = A_0 \frac{\sum_{i=1}^{m_\ell^{Pha}} e^{-S_{\ell i}}\, e^{-E_{\ell i}/kT}}{\sum_{i=1}^{m_\ell} e^{-E_{\ell i}/kT}} \qquad (4)$$

By determining A_0 from this equation and substituting it in Eq.4, we obtained;

$$A_n = A_\ell \frac{\sum_{i=1}^{m_\ell} e^{-E_{\ell i}/kT} \sum_{i=1}^{m_n^{Pha}} e^{-S_{ni}}\, e^{-E_{ni}/kT}}{\sum_{i=1}^{m_n} e^{-E_{ni}/kT} \sum_{i=1}^{m_\ell^{Pha}} e^{-S_{\ell i}}\, e^{-E_{\ell i}/kT}} \qquad (5)$$

Using the experimental data for the activities of the molecules in the training set, we estimated κ_j in Eq.2 by performing PLS. With the constants κ_j determined in this way, we could evaluate the expected activity of any molecular system using Eq.5. In this formula, only the choice of the $a_{ni}^{(j)}$ independent variable remained uncertain. It required some experience and skill. Bersuker at al. have shown how to handle the multiconformation problem successfully by using Electron-Conformational (EC) method [13, 23]. In comparison with the EC method, MCET automatically took into account a set of 3DSDs for all compounds in the training set. This had a significant role for a method that considers the L-R binding, consisting of a large number of positions of the different conformers. By simultaneously solving the rather complex problem of NLME we demonstrated that the parameters for all the compounds in the training set could be defined. To do this we have presented an improved MCET and shown its application to the problem of the binding affinities of ureases. As we have previously shown how to employ MCET algorithm in an earlier paper, we have not repeated it again here [24]. We applied the variable selection method based on GFA to the ureases data and constructed the nonlinear QSAR model in Eq.5. The best QSAR model was selected according to the correlation coefficient ($R^2=0.71$) and the Leave-One-Out Cross Validation (LOO-CV) correlation coefficient ($Q^2=0.60$). The obtained result was accurate and interpretable. Moreover, in order to confirm the predictive ability of the model, a validation test was performed ($R^2_{ext} =0, 60$).

3. RESULTS AND DISCUSSION

The model obtained using the 3D structural descriptors showed good predictive ability. The information about the mechanism of model was based only on the possible descriptors of AG and APS defined according to statistical results. The change of activity calculated using the model was a measure of the magnitude and sign of the interaction points between Ligand-Receptor (L-R). These interactions were either attractive or repulsive depending on whether the charges on both sites were of different or the same type. The sign of the product of the values at L-R sites therefore determines the direction of the force. For example, when we take a point of negative charge of receptor site the local positive charge of ligand acts as an AG group otherwise the negative charge acts an APS. These effects should be considered in quantifying the global reactivity of molecules.

The total set of Raf-1 kinase inhibitors (76 compounds) given were chosen from the literature [2]. In order to design the training and test sets, the complete data set was processed using cluster analysis. The molecules were first divided into two subsets: one training set composed of 49 molecules, and one external test set (marked as *) composed of 27 molecules not included in 4D-QSAR model development but rather employed to analyze predictive performance. The test set comprised of nearly 35 % of the whole set, ensuring that the test set contained representative samples of the training group and included the range of activity values of the training group.

Table 1: *The observed and calculated activities (*Test set compounds)*

Molecule No	Observed (pIC50)	Calculated (pIC50)	Residuals
n01_02	5.959	5.959	0
n02_01	6.639	6.256	0.383
n03_01	6.239	6.707	-0.468
n04_01*	5.199	5.395	-0.196
n05_01	6.699	6.739	-0.04
n06_01	6.099	6.11	-0.011
n07_01	6.059	6.678	-0.619
n08_01	5.069	5.239	-0.17
n09_01	6.919	6.238	0.681
n10_01	5.109	5.341	-0.232
n11_01*	6.149	6.191	-0.042
n12_01	6.799	6.451	0.348
n13_01	6.059	6.485	-0.426
n14_01	5.789	6.087	-0.298
n15_01	6.549	6.437	0.112
n16_01	6.229	5.898	0.331
n17_01*	6.659	6.153	0.506
n18_01	5.459	6.106	-0.647
n19_01*	5.889	6.108	-0.219
n20_01*	5.999	5.884	0.115
n21_01	6.679	6.046	0.633
n22_01	5.439	6.283	-0.844
n23_01	6.299	6.17	0.129
n24_01	6.569	6.056	0.513
n25_01*	6.089	6.697	-0.608
n26_01	6.549	6.033	0.516
n27_01	5.199	5.228	-0.029
n28_01*	5.639	5.208	0.431
n29_01	4.999	5.182	-0.183
n30_01	5.479	6.362	-0.883
n31_01	6.919	5.688	1.231
n32_01*	6.149	6.753	-0.604
n33_01	5.419	5.734	-0.315
n34_01	6.719	6.703	0.016

n35_01	5.359	5.54	-0.181
n36_01*	7.339	6.922	0.417
n37_01	6.219	5.987	0.232
n38_01*	5.569	5.531	0.038
n39_01	6.349	6.347	0.002
n40_01*	6.339	6.479	-0.14
n41_01	5.359	5.328	0.031
n42_01	6.139	6.43	-0.291
n43_01	6.289	6.481	-0.192
n44_01*	6.359	6.513	-0.154
n45_01	7.459	7.065	0.394
n46_01	6.279	6.461	-0.182
n47_01	6.349	6.787	-0.438
n48_01*	7.919	7.947	-0.028
n49_01	7.889	6.947	0.942
n50_01*	7.589	7.335	0.254
n51_01	7.549	7.371	0.178
n52_01*	6.519	7.155	-0.636
n53_01	8.219	8.195	0.024
n54_01	7.569	7.328	0.241
n55_01*	6.769	6.12	0.649
n56_01	7.389	7.072	0.317
n57_01	7.059	7.089	-0.03
n58_01	6.619	7.063	-0.444
n59_01	6.919	6.694	0.225
n60_01	6.889	6.886	0.003
n61_01*	6.849	6.904	-0.055
n62_01	5.239	5.368	-0.129
n63_01	6.339	6.306	0.033
n64_01	6.429	6.644	-0.215
n65_01	7.169	6.512	0.657
n66_01	6.889	6.583	0.306
n67_01	7.149	7.04	0.109
n68_01	7.089	6.714	0.375
n69_01*	6.569	7.138	-0.569
n70_01	7.359	7.409	-0.05
n71_01	6.639	6.928	-0.289
n72_01*	6.799	7.002	-0.203
n73_01	6.959	7	-0.041
n74_01*	6.889	7.63	-0.741
n75_01	7.139	7.487	-0.348
N76_01*	6.999	6.43	0.569

We have found that the active regions comprising atoms O2, N1, C14 and C7 as Pha structures in the template conformer. In the Pha hypothesis scoring process, each Pha and its associated biostructure was treated temporarily as a reference in order to assign a score, and the hypotheses were ranked according to the following scores: 1) the alignment score of Pha atoms 2) the superimposed score of oriented atoms around Pha 3) a superposition score of bioactive conformation consisting of Pha, AG and APS. A

biostructure hypothesis was derived to orient the superposition of all individual molecules and to afford a rational and consistent alignment. One of candidate Pha structures, represented the most stabilization of the L-R complex, was depicted as a, b, c and j in Figure 1. Depending on the Pha structure, the nine parameters were obtained using GFA. They could either be attractive (AG) orrepulsive (Aps) depending on the corresponding charges on both ligand and receptor, whichcgiven as d, e, f, g, h, i, k, l and m in Figure 1.
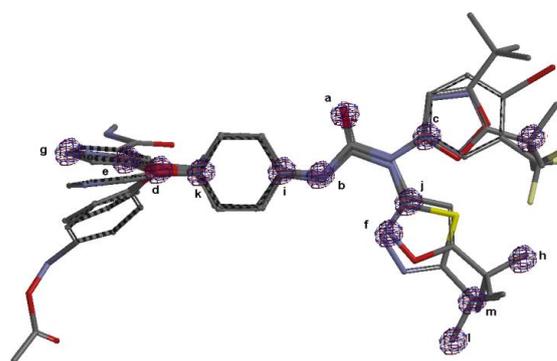
Figure 1: *Common Pha structure for all ligands and AG or APS for some ligands, which are assigned with letters a, b, c etc in Table 2. Pha structure and descriptors*

The molecules for Raf-1 kinase inhibitor were employed in respect to pIC_{50} values to generate regression models utilizing the GFA. According to the GFA, all biostructure hypotheses produced in the previous step were used to build the 4D-QSAR model. This model was constructed from the superposition in the training set, and then its predictive power was validated by the test set molecules. For the interpretation of the results, only the relative magnitude of the interaction points at both receptor and ligand sites and their signs are important, not their absolute values. The degree of increase or decrease in biological activity was strongly dependent on the values of the independent variables and variational parameters used in Eq.2. Biological activity was represented by the total interaction energies between ligand and receptor. The AG or APS in the molecule, assigned to biostructure of the virtual receptor site, were used in the calculation of the interaction energies. Atomic Fukui indices in molecule were multiplied with the magnitude and sign of the adjustable constants, simultaneously obtained using Eq.5. Thus, 4D-QSAR visualization was indicated by Pha, AG and APS [25]. AG or APS on the jth position was marked with letters a-, b-, c-, etc., and the values of κ_j being the adjustable constants were given in Table 2. Positions in Table 2

were defined from most and less active five molecules. After information about the definition of each model was consulted in the statistical analysis, the results were presented in detailed form (as shown in Table 3) for a better understanding of the selected descriptors.

Table 2: *The positions defined from the proper places of the atoms within 5 molecules (n01, n4 n8, n29 and n53).*

Mol. Conf. No	Atom	X-dist.	Y-dist	Z-dist.	K-Value
n01_02	O2	0	0	0	Ka=1.746
n01_02	N1	2.308	0	0	Kb=-1.172
n01_02	C14	0.096	2.862	0	Kc=0.423
n29_01	O3	3.753	-5.419	-0.076	Kd=0.112
n29_01	C17	5.443	-6.995	-0.630	Ke=1.343
n29_01	N1	3.409	2.525	-0.016	Kf=1.667
n08_01	C4	4.475	-8.219	2.124	kg=2.009
n04_01	C18	-2.211	6.812	-0.110	Kh=-2.378
n53_01	C16	-3.572	6.295	0.829	Ki=2.691
n29_01	C7	2.139	2.894	-0.023	Kj=-0.694
n29_01	C10	3.406	-4.063	-0.115	Kk=0.413
n29_01	C4	4.801	6.280	-1.343	Kl=0.833
n53_01	C10	-1.768	4.984	-0.072	Km=-0.239
n04_01	C14	-0.756	4.804	-0.077	Kn=0.136

Table 3: *The ranking values of rm2 in PLS*

Pha structure, and AG- and APS-groups	Contributions to biological activity from Pha and other groups for the rm^2(overall)
Pha (a,b,c and j-)	0.075
Pha, d	0.172
Pha, d,e	0.296
Pha, d,e,f	0.387
Pha, d,e,f,g	0.435
Pha, d,e,f,g,h	0.476
Pha, d,e,f,g,h,i	0.554
Pha, d,e,f,g,h,i,k	0.583
Pha, d,e,f,g,h,i,k,l	0.608
Pha, d,e,f,g,h,i,k,l,m	0.629
Pha, d,e,f,g,h,i,k,l,m,n	0.695

To obtain the PLS statistical parameters, The prediction of generated model was validated using r_m^2(overall) (differences between observed and predicted values of the compounds of the whole set considering both training and test sets). The hypotheses were ranked according to the values of r_m^2(overall). In PLS, the iterations continue until the values of rm2 calculated no longer increases significantly in Table 4. Also, these groups cross validated by Q2 and R2 values were assigned a letter of a, b, c etc. for all the compounds in Table 3. The predicted activities of the training and test set molecules are also listed in Table 4.

Table 4: *Experimental and predicted activities of molecules, relative energy of the conformers and various positions of label a, b, c, etc. which some conformers of molecules include (*Test set compounds)*

Molecule No	Observed (pIC50)	Calculated (pIC50)	Selected Confmr's Energy and Postions of a, b,c
n01_02	5.959	5.959	-3057855.440abcdejkmn
n02_01	6.639	6.256	-3099963.130abcdjkmn
n03_01	6.239	6.707	-3057843.410abcdeijkmn
n05_01	6.699	6.739	-3047858.090abcdhijkmn
n06_01	6.099	6.110	-3151062.360abcfjl
n07_01	6.059	6.678	-3063586.250abcdjkmn
n08_01	5.069	5.239	-3990112.350abcdfgjkl
n09_01	6.919	6.238	-3099950.760abcdgijkmn
n10_01	5.109	5.341	-2954628.350abcdejkmn
n12_01	6.799	6.451	-3203164.430abcdijkmn
n13_01	6.059	6.485	-3096694.040abcdgjkmn
n14_01	5.789	6.087	-3264267.530abcdeijkmn
n15_01	6.549	6.437	-3306376.990abcdgijkmn
n16_01	6.229	5.898	-3005732.830abcdfjkl
n18_01	5.459	6.106	-3151035.640abcdfjkl
n21_01	6.679	6.046	-3255349.040abcdefjkl
n22_01	5.439	6.283	-3594796.350abcdefjkl
n23_01	6.299	6.170	-3358549.470abcdefjkl
n24_01	6.569	6.056	-3255353.740abcdefjkl
n26_01	6.549	6.033	-3099975.480abcdfgjkl
n27_01	5.199	5.228	-3552977.140abcdej
n29_01	4.999	6.182	-3801357.970abcdefjkl
n30_01	5.479	6.362	-3759400.350abcdfjkl
n31_01	6.919	5.688	-3603992.160abcdefjkl
n33_01	5.419	5.734	-3942769.800abcdefjkl
n34_01	6.719	6.703	-3203222.850abcdefjkl
n35_01	5.359	5.540	-3552968.490abcdefjkl
n37_01	6.219	5.987	-4420627.370abcjk
n39_01	6.349	6.347	-4742413.650abcdjkmn
n41_01	5.359	5.328	-3488453.640abcdhijkmn
n42_01	6.139	6.430	-3796283.490abcdjkmn
n43_01	6.289	6.481	-3836419.750abcdjkmn
n45_01	7.459	7.065	-10291866.780abcj
n46_01	6.279	6.461	-3796259.320abcdghijkmn
n47_01	6.349	6.787	-10613648.040abcdghijkmn
n49_01	7.889	6.947	-5185327.550abcdhijkmn
n51_01	7.549	7.371	-5589234.680abcdhijkmn
n53_01	8.219	8.195	-10838028.290abcdhijkmn
n54_01	7.569	7.328	-10941249.990abcjk
n56_01	7.389	7.072	-5486027.220abcdghijkmn
n57_01	7.059	7.089	-5589254.630abcdghijkmn
n58_01	6.619	7.063	-5692479.870abcdghijkmn
n59_01	6.919	6.694	-3646103.950abcdfgjkl
n60_01	6.889	6.886	-3749331.090abcdefjkl
n62_01	5.239	5.368	-3749307.710abcdfgjkl
n63_01	6.339	6.306	-4252731.910abcdefjkl

n64_01	6.429	6.644	-4149512.980abcdfjkl
n65_01	7.169	6.512	-4191613.530abcdfgjkl
n66_01	6.889	6.583	-5246402.280abcdehijkmn
n67_01	7.149	7.040	-6102105.610abcdjkmn
n68_01	7.089	6.714	-6102105.610abchijkmn
n70_01	7.359	7.409	-5791924.960abcdhijkmn
n71_01	6.639	6.928	-6101550.620abcdgjkmn
n73_01	6.959	7.000	-5650289.700abcdehijkmn
n76_01	7.139	7.487	-6144210.720abcdghijkmn
n04_01*	5.199	6.395	-3142085.810abcdghjkn
n11_01*	6.149	6.191	-3161057.350abcdeijkmn
n17_01*	6.659	6.153	-3947953.460abcdfjkl
n19_01*	5.889	6.108	-4264545.230abcdefjkl
n20_01*	5.999	5.884	-3161103.940abcdefjkl
n25_01*	6.089	6.697	-3458646.810abcdfjkl
n28_01*	5.639	5.208	-3203198.000abcdefjkl
n32_01*	6.149	6.753	-3358555.430abcdfjkl
n36_01*	7.339	6.922	-4742386.500abcdghijkmn
n38_01*	5.569	6.531	-4158192.560abcdjkmn
n40_01*	6.339	6.479	-3535737.050abcdgjkmn
n44_01*	6.359	6.513	-3733224.030abcdjkmn
n48_01*	7.919	6.947	-5288543.960abcdghijkmn
n50_01*	7.589	7.335	-5391768.560abcdghijkmn
n52_01*	6.519	7.155	-5391715.170abcdghijkmn
n55_01*	6.769	6.120	-10941195.790abcdghijkmn
n61_01*	6.849	6.904	-3852551.090abcdfgjkl
n69_01*	6.569	7.138	-6007851.300abcdehijkmn
n72_01*	6.799	7.002	-6101578.640abcjk
n74_01*	6.889	7.630	-6092630.840abcdhijkmn
n80_01*	6.999	6.430	-4782755.160abcdgjkmn

The regression line for the observed and MCET predicted activity was shown in Figure 2. MCET has been shown to be both useful and reliable for the construction of quantitative models with a 3D biostructure, especially for sets of flexible ligand analogue
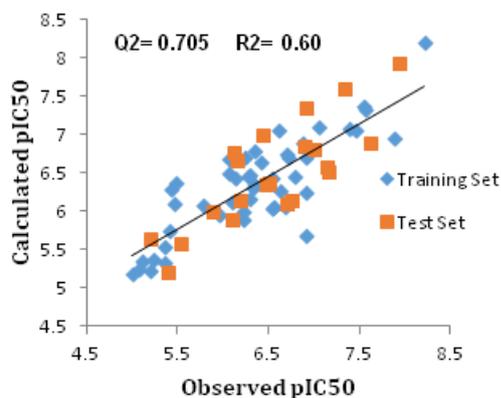
Figure 2: *Fitness graph between observed and predicted binding affinity for training and test set molecules.*

The best QSAR model generated with the following statistical parameters values: q 2 = 0.705; r 2 = 0.60; which were statistically more significant than the values reported in ref. 2 [Q2COMFA = 0.53; Q2COMSIA = 0.44].

4. CONCLUSION

MCET developed by us, described in ref. [6, 24], outlined atomic Fukui indices as the 3DSDs to be coded in QSAR descriptors. As the main part of the present study, we constructed a model consisting of Pha, AG and APS groups, the biostructure found from the 3D sub-structures of five different compounds. A meaningful 4D-QSAR model was derived using Atomic Fukui indices for the series of ureas derivatives. Furthermore, the 4D-QSAR model also explained how atomic Fukui indices in the different positions influenced the Raf-1.The robustness of the constructed model used in this study has good stability and great predictive power, as assessed by the internal and external validations. The selected 3DSDs in the 4D-QSAR model can illustrate the contributing electronic and steric properties that affect the activity of ureas derivatives. By interpreting 3DSDs in the regression model, we conclude that a Pha structure consisting of four atoms was essential for activity, and ten AG or APS groups surrounding its structure contribute to the activity. The present study provided useful guidelines for developing ureas derivatives as potent active molecules in ligand-based drug design approaches. Finally, the accuracy and predictability of the proposed model were selected by best statistical results of $rm^2_{(Overall)}$, Q^2 or R^2. As a result compared to COMFA and COMSIA methods, better results were obtained with MCET method [9].

REFERENCES

[1] Robinson, M. J.; Cobb, M. H. Curr. Opin. Cell Biol. 1997, 9, 180–186.

[2] R. Thaimattam et al. / Bioorg. Med. Chem. 12 (2004) 6415–6425

[3] Lyons, J. F.; Wilhelm, S. M.; Hibner, B.; Bollag, G.Endocr.-Relat. Cancer 2001, 8, 219–225.

[4] Morrison, D. K. Nature 2004, 428, 813–815.

[5] Hall-Jackson, C. A.; Eyers, P. A.; Cohen, P.; Goedert, M.; Boyle, F. T.; Hewitt, N.; Plant, H.; Hedge, P. Chem. Biol.1999, 6, 559–568.

[6] Yilmaz H, Güzel Y, Önal Z, Altiparmak G, Özhan Kocakaya Ş. 4D-Qsar Study Of P56(Lck) Protein Tyrosine Kinase Inhibitory Activity Of Flavonoid Derivatives Using MCET Method. *Journal of*

Bulletin of the Korean Chemical Society (2012) 32: 4352-4360.

[7] Hopfinger, A. J.; Reaka, A.; Venkatarangan, P.; Duca, J. S.; Wang, S. Construction of a Virtual High Throughput Screen by 4D-QSAR Analysis: Application to a Combinatorial Library of Glucose Inhibitors of Glycogen Phosphorylase b. J. Chem. Inf. Comput. Sci. 1999, 39, 1151–1160.

[8] Senese, C. L.; Duca, J.; Pan, D.; Hopfinger, A. J.; Tseng, Y. J. 4D fingerprints, universal QSAR and QSPR descriptors. J. Chem. Inf. Comput. Sci. 2004, 44, 1526–1539.

[9] Luiz C. S. Pinheiro, Júlio C. Borges, Maurício S. dos Santos, Vitor F. Ferreira, Alice M. R. Bernardino, Rodrigo Tonioni, Plínio C. Sathler, Helena C. Castro, Dilvani O. Santos, Samara B. Nascimento , Saulo C. Bourguignon, Uiaran O. Magalhães, Lúcio Cabral, and Carlos R. Rodrigues. Searching for new antileishmanial lead drug candidates: Synthesis, biological and theoretical evaluations of promising thieno[2,3-b]pyridine derivatives journal of Microbiology and Antimicrobials Vol. 4(1), pp. 32-39, 2012

[10] Bersuker IB, Dimoglo AS. The Electron-Topological Approach to the QSAR Problem, in book: "Reviews in Computational Chemistry". Lipkowitz KB, and Boyd DB, eds., VCH, New-York, 1991, pp.423-460.

[11] Bersuker IB, Dimoglo AS, Yu M, Gorbachov P, Vlad F, Pesaro M. Origin of musk fragrance activity: The electron-topologic approach. *New J. Chem.* (1991) 15: 307-320.

[12] Glen, W. G.; D, W. J.; Scott, D. R. Principal components analysis and partial least squares. Tetrahedron Comput. Methods 1989, 2, 349– 354.

[13] Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quant. Struct.-Act. Relat., v.13, n.3, p.285-294, 1994.

[14] Karelson M, Lobanov VS, Katritzky, AR. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem Rev.* (1996) 96: 1027-1044.

[15] Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. J. Am. Chem. Soc. 1997, 119, 10509–10524.

[16] Prasad VB. Modeling and Informatics in Drug Design. *Precilinical Development Handbook.* (2008) p:1-47.

[17] Guzel Y, Saripinar E, Yildirim I. Electron-topological (ET) investigation of structure-antagonist activity of a series of dibenzo(a,d)cycloalkenimines. *J. Mol. Struct. (Theochem)* (1997) 418: 83-91.

[18] Guzel Y, Ozturk E. *Stu*dy of the binding affinity for corticosteroid-binding globulin (CBG) using the electron topological method (ETM) as three-dimensional quantitative structure–activity relationship (3D QSAR). *Bioorg. Med. Chem.* (2003), 11: 4383-4388.

[19] Dimoglo AS, Beda AA, Shvets NM, Gorvachov MY, Kheifits LA, Aulchenko IS. Investigation of the relationship between sandalwood odor and chemical structure: Electron-topological approach. *New. J. Chem.* (1995) 19: 149-154.

[20] Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. J. Comput. Aided Mol. Des. 2006, 20, 647.

[21] Bersuker, I. B.; Bahceci, S.; Boggs, E. J.; Pearlman, R. S. An Electron- Conformational Method of Identification of Pharmacophore and Anti-Pharmacophore Shielding: Application to Rice Blast Activity. J. Comput.-Aided Mol. Des. 1999, 13, 419-434.

[22] Bersuker, I. B.; Bahceci, S.; Boggs, E. J. The Electron-Conformational Method of Identification of Pharmacophore and Anti-Pharmacophore Shielding. In Pharmacophore Perception, DeVelopment, and Use in Drug Design; Guner, O. F., Ed.; International University Line: La Jolla, 2000; pp 457-474.

[23] Isaac B. Bersuker QSAR without arbitrary descriptors: the electron-conformational method J Comput Aided Mol Des (2008) 22:423–430

[24] Yilmaz H, Güzel Y, Boz, M, Türkmenoğlu B. Pharmacophore And Functional Group Identification Of 4,4 '-Dihydroxydiphenylmethane As Bisphenol-A (Bsa) Derivative. *Tropical Journal of Pharmaceutical Research.* (2014) 13: 117-129.

[25] Bersuker, I. B.; Bahceci, S.; Boggs, E. J.; Pearlman, R. S. A Novel Electron-Conformational Approach to Molecular Modeling for QSAR by Identification of Pharmacophore and Anti-Pharmacophore Shielding. SAR QSAR EnViron. Res. 1999, 10, 157-173.