Journal of Medicinal Chemistry and Drug Discovery

Available online at http://www.jmcdd.com
October - November, 2013, Vol. 1, No.1, pp 48-58

Research Article

QUANTITATIVE ALIGNMENT INDEPENDENT STRUCTURE ACTIVITY RELATIONSHIP STUDIES FOR A-GLUCOSIDASE INHIBITORS: GUSAR AND 3D-QSAR ANALYSES

J. DESHPANDE

Department of Chemistry, VidyaBharati College, Camp, Amravati, Maharashtra, India.

(**Received**: October 28, 2013; **Accepted**: November 23, 2013)

Abstract

In present study, we have employed alignment independent 3D-descriptors and General Unrestricted Structure Activity Relationships (GUSAR) to explore α-glucosidase inhibitory activity of xanthone derivatives. The analysis reveals that integration of both techniques not only leads to consensus pharmacophore identification but also a better and clear idea about the contribution of individual atom/group in deciding the pharmacological activity. R^2, $R^2_{adj.}$, F and Q^2were the various yardsticks to assess the quality of QSAR and GUSAR models. For the xanthone derivatives, used in present analysis, the consensus GUSAR and conventional 3D-QSAR model have high statistical robustness with R^2 being 0.813 and 0.837 respectively. The present analysis enlightens that the factors which are either missed or neglected by conventional QSAR model(s) can be identified with GUSAR.

Keywords: Xanthone derivatives, GUSAR, 3D-QSAR, α-glucosidase inhibitors.

Introduction

Diabetes has adverse effects on medical and economic conditions of patients. The α-glucosidase, exo-acting carbohydrases, are essential enzymes which are capable of releasing α-D-glucopyranose from the non-reducing ends of a variety of carbohydrates. Literature survey reveals that the inhibition of its catalytic activity causes retardation of glucose absorption and significant reduction in postprandial blood glucose level. Owing to this, efficient α-glucosidase inhibitors are promising chemotherapeutic agents for treatment of hypertension, diabetes, obesity, dyslipidemia, and cardiovascular diseases in patients with metabolic disorder [1, 2]. Myriad number of glucosidic derivatives has been reported to be α-glucosidase inhibitors. However, insufficient structural information about the nature of the interactions between α-glucosidase and the inhibitors has thus made it a hard task to find out good lead compounds [3]. QSAR and GUSAR studies are fruitful methods when little or no information is available about 3D structure of target enzyme. QSAR is used to buildrational mathematical relations (models) between structural

features and the biological activity of the congeneric molecules. In QSAR, the common methodology is to generate linear multi-parametric model(s) which gives idea about role of certain structural features for whole set, but for some molecules, it isplausible that some additional structural features might be more affecting, which are omitted or missed during model generation. In some instances, it is also probable that the structural feature which appears to play crucial (positive or negative) role for "whole set" may have small or negligible impact on biological profile of certain molecules of same data set [4-12]. In these conditions, the use of QSAR model is either erroneous or incorrect. Doable solutions are either to include additional features in QSAR model with more descriptors which may cause "over fitting" or to use QSAR in integration with any other technique. Herein, we report the integration of conventional 3D-QSAR with GUSAR [13-16]. Some of the advantages associated with GUSAR are: (1) For creation of a reliable QSAR model with optimal number and set of descriptors, Self-Consistent Regression (SCR) algorithm is utilized. (2) Quantitative Neighbourhoods of Atoms (QNA) and Multilevel Neighbourhoods of Atoms (MNA) descriptors are used which are better than conventional descriptors to reveal the nature of intermolecular interactions. (3) It predicts the quantitative values of biological activity of chemical compounds on the basis of their "structural formulae" only and there is no need to have information about the 3D structure of ligands and/or target protein(s).

Preparation of the structures and computational softwares:

The 43 molecules were drawn in ChemSketch 12 freeware [17] followed by MM2 energy minimization and biological data addition before further analysis in GUSAR 2010. For better analysis, following settings were used: Y-randomization = 20 iterations, Leave Many Out (LMO) = 20 iterations, No. of leave out = 10%, leverages = 0.99, Similarity = 0.70, kNN RMSE/ Average RMSE = 1, No. of Models = 36. For conventional 3D-QSAR, e-Dragon and Weka 3.7 were used for descriptor generation and model building respectively.

Results and discussions:

Theory of GUSAR:

Quantitative Neighbourhoods of Atoms (QNA) and Multilevel Neighbourhoods of Atoms (MNA) descriptors as well as Self-Consistent Regression (SCR) algorithm[13-16] are well implemented and utilized in GUSAR. The calculation of QNA and MNA involves NN for better and accurate results. The basic difference between conventional QSAR and GUSAR lies in the representation of molecule in the space of calculated descriptors. In GUSAR, any molecule is represented as a set of points in two-dimensional (2D) space of QNA descriptors, whereas in conventional QSAR approach, any molecule is represented as a single point in a many-dimensional space of molecular descriptors. In GUSAR, QNA and MNA descriptors are used to build the consensus model; the

calculations of these descriptors are well documented in the literature[13-16]. (4) GUSAR gives output, which is in the form a diagram for individual molecule of the data set, revealing the atoms suitably coloured according to their specific contribution to the activity. (5) GUSAR performs cross validation and Y-randomization and checks the various statistical characteristics to build consensus model during the course of model building. (6) Calculation of descriptors and generation of model is alignment independent. This study may help us to design new analogues with better biological profile.

Experimental/Computational protocol:

Data set: The database consists of 43 recently discovered xanthone derivatives as α-glucosidase inhibitors [3,18]. Their structures and *in vitro* activity are listed in Fig. 1 and Table 1 respectively. Activity values were converted into the corresponding $pIC_{50} = -\log_{10}(IC_{50})$ values, where IC_{50} is the effective concentration of compound required to achieve 50% of inhibition of α-glucosidase.

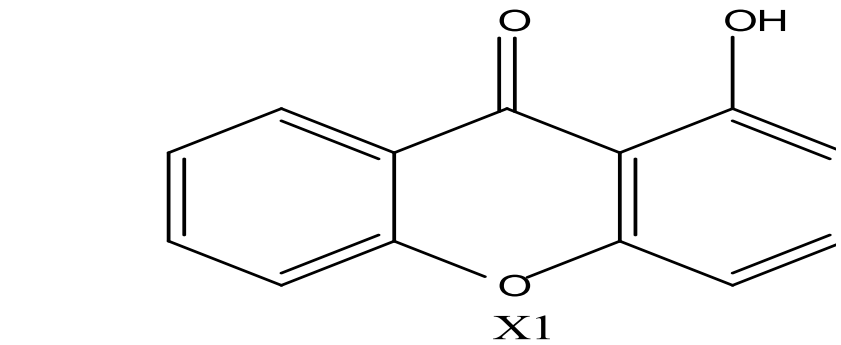
Analysis of GUSAR output:

The output of GUSAR appears in the form of an image in which the atoms are coloured according to their contribution towards biological activity along with various statistical characteristics used to arrive at the consensus model (Figure 2). The obvious limitation of GUSAR is that it neither provides the QSAR model as MLR in interpretable form nor any knowledge about the descriptors that are used to build the consensus model. If QSAR models were produced on the basis of QNA descriptors the involvement of every atom into the predicted value is showed for a studied compound.The contribution is a calculation of activity value for a single atom from the structure of the studied molecule. Explanation of the colours is as following: “Green” means that the impact of the atom approximately corresponds to the predicted activity value for a whole molecule. “Blue” means that the particular atom may decrease the activity. “Red” means that the particular atom may increase the activity. Thus, if one would like to increase the activity, the number of “blue” atoms should be reduced, and the number of “red” atoms should be increased. One can analyze how many fragments have "red" and "blue" colors for finding the most important fragments[14-16].The QNA descriptors based consensus model which was automatically generated and selected among the 36 models by GUSAR 2010was found to be with following statistical characteristics:

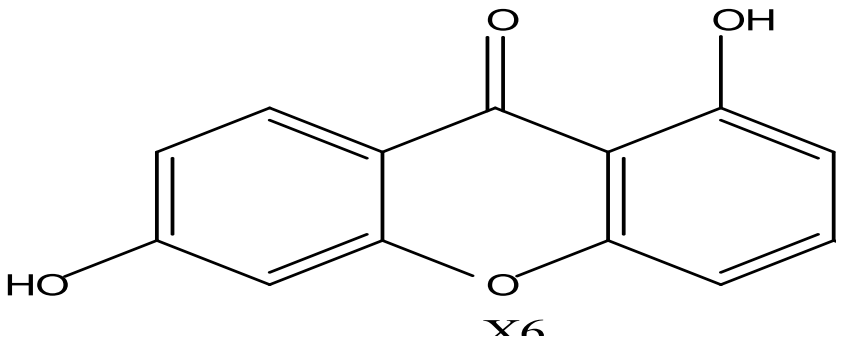
N = 43, R^2 = 0.813, $R^2_{adj.}$ = 0.782, F = 18.207, SD = 0.204, Q^2 = 0.747, V = 6

Where N is total number of molecules used, R is correlation coefficient, $R^2_{adj.}$is adjusted R^2, F is value of Fischer’s parameter, SD is standard deviation, Q^2 is the cross-validated R^2 and V is no. of variables used in the model building. The high value of R^2, $R^2_{adj.}$,Q^2, F and low value of SD indicates that the model is statistically very sound

and could be used for future drug designing.The GUSAR analysis was performed for complete data set of 43 molecules. To analyze the contribution of atoms towards binding with receptor, we report the output of GUSAR for four most active and four least active as representative xanthones to get insight into atom-wise contribution towards biological activity(figure 3).



X1



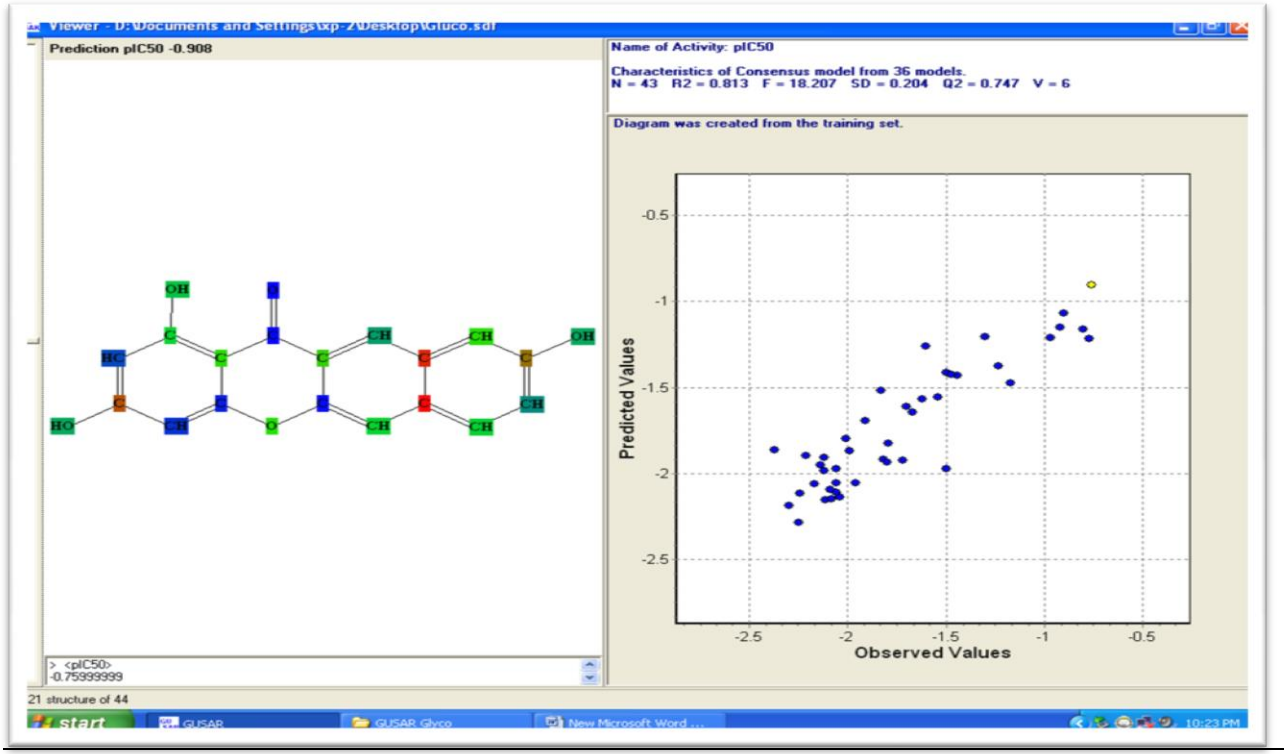
X6

Figure 1.Structures of Xanthone derivatives.

Table 1: Experimental, predicted pIC_{50} (GUSAR model) and pIC_{50} (QSAR model)

Compound	pIC_{50} (Exp.)	pIC_{50} (Calc.) GUSAR model	Residual (Exp.-Pred. GUSAR)	pIC_{50} (Calc.) QSAR model	Residual (Exp.-Pred. QSAR)
X1	-2.25	-2.2969	0.04690	-2.1648	-0.0852
X10	-2.14	-1.9232	-0.21680	-2.0157	-0.1243
X11	-1.67	-1.634	-0.03600	-1.5887	-0.0813
X12	-1.7	-1.5769	-0.12310	-1.5156	-0.1844
X13	-2.24	-2.1009	-0.13910	-2.2326	-0.0074
X14	-2.04	-2.1491	0.10910	-2.1330	0.0930
X15	-2.11	-2.1563	0.04630	-2.1492	0.0392
X16	-2.08	-2.1517	0.07170	-2.1067	0.0267
X17	-2.06	-2.1166	0.05660	-2.1347	0.0747
X18	-2.09	-2.0906	0.00060	-2.1211	0.0311
X19	-2.06	-2.0551	-0.00490	-2.0990	0.0390

X20	-1.99	-1.8371	-0.15290	-2.1296	0.1396
X21	-1.82	-1.9318	0.11180	-1.7610	-0.0590
X22	-1.72	-1.9474	0.22740	-1.7066	-0.0134
X23	-2.06	-1.9598	-0.10020	-2.1466	0.0866
X24	-1.79	-1.8398	0.04980	-1.9910	0.2010
X25	-1.8	-1.9498	0.14980	-1.8826	0.0826
X26	-2.12	-1.9624	-0.15760	-2.1135	-0.0065
X27	-0.97	-1.2345	0.26450	-1.2784	0.3084
X28	-0.76	-0.92998	0.16998	-0.5640	-0.1960
X29	-0.9	-1.0914	0.19140	-0.7724	-0.1276
X3	-1.96	-2.0709	0.11090	-2.0027	0.0427
X30	-1.5	-1.4028	-0.09720	-1.5902	0.0902
X31	-1.3	-1.191	-0.10900	-1.2184	-0.0816
X32	-1.44	-1.4217	-0.01830	-1.3535	-0.0865
X33	-1.6	-1.2123	-0.38770	-1.3363	-0.2637
X34	-1.54	-1.5649	0.02490	-1.3926	-0.1474
X35	-2.37	-1.7651	-0.60490	-2.0317	-0.3383
X36	-2.01	-1.7577	-0.25230	-1.7468	-0.2632
X37	-2.17	-2.0528	-0.11720	-2.0122	-0.1578
X38	-2.3	-2.1719	-0.12810	-2.0233	-0.2767
X39	-0.77	-1.2801	0.51010	-1.1308	0.3608
X4	-2.12	-1.8899	-0.23010	-2.0305	-0.0895
X40	-0.8	-1.2147	0.41470	-1.0029	0.2029
X41	-0.92	-1.1851	0.26510	-1.0058	0.0858
X42	-1.47	-1.4164	-0.05360	-1.6140	0.1440
X43	-1.83	-1.482	-0.34800	-1.5408	-0.2892
X5	-1.91	-1.6665	-0.24350	-1.7783	-0.1317
X6	-1.62	-1.5605	-0.05950	-1.7508	0.1308
X7	-1.17	-1.5091	0.33910	-1.4105	0.2405
X8	-1.23	-1.4243	0.19430	-1.4730	0.2430
X9	-1.5	-2.0335	0.53350	-2.0276	0.5276



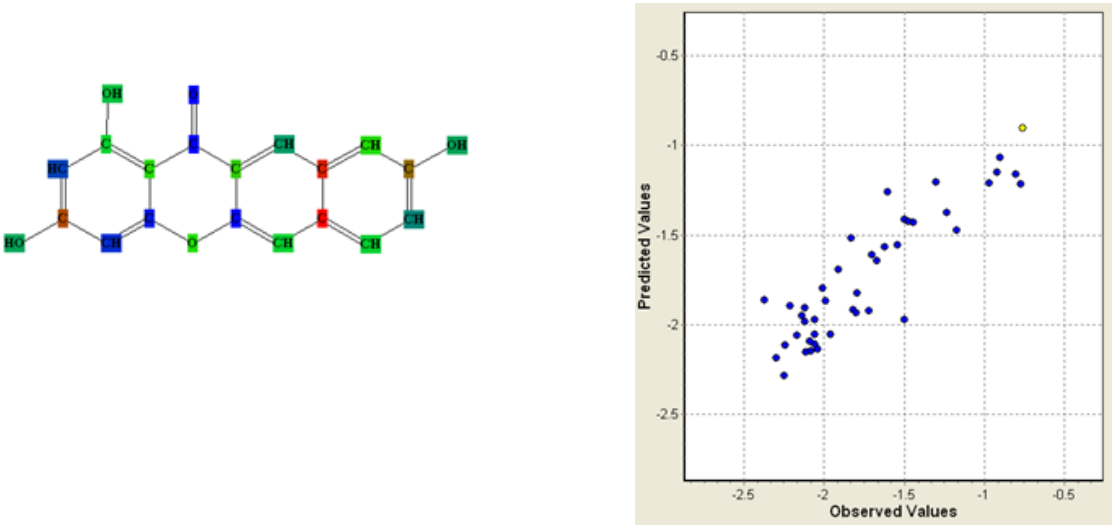
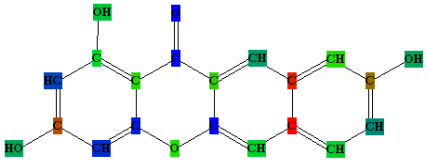
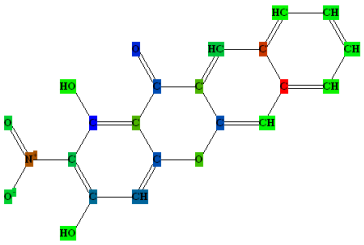


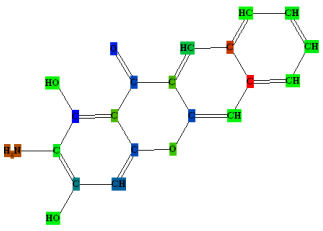
Figure 2: Output of GUSAR (On left side: most active molecule **28**with atoms coloured according to their contribution in deciding the biological activity and on right side: various statistical characteristics and a graph between observed & predicted values)



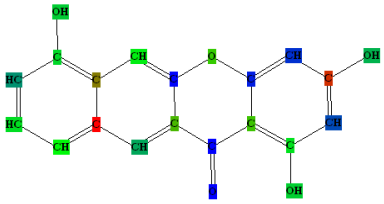
X28



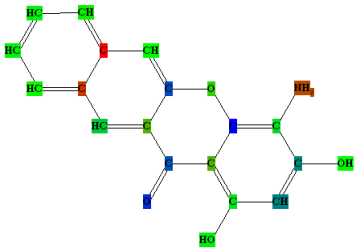
X39



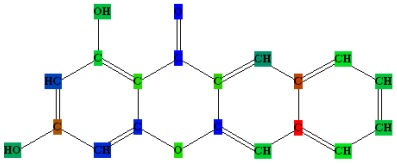
X40



X29

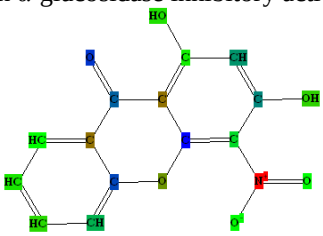


X41

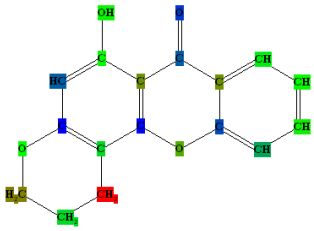


X27

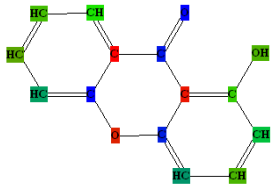
High α-glucosidase inhibitory activityin descending order X28>X39>X40>X29>X41>X27



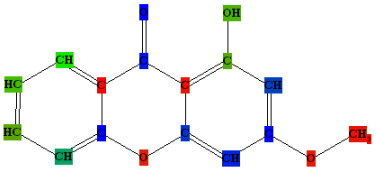
X35



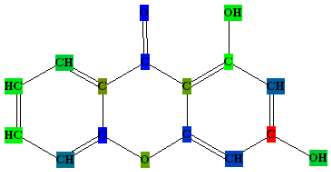
X38



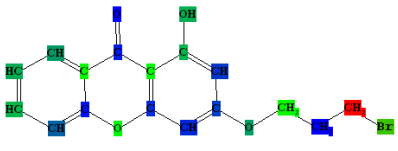
X1



X13



X2



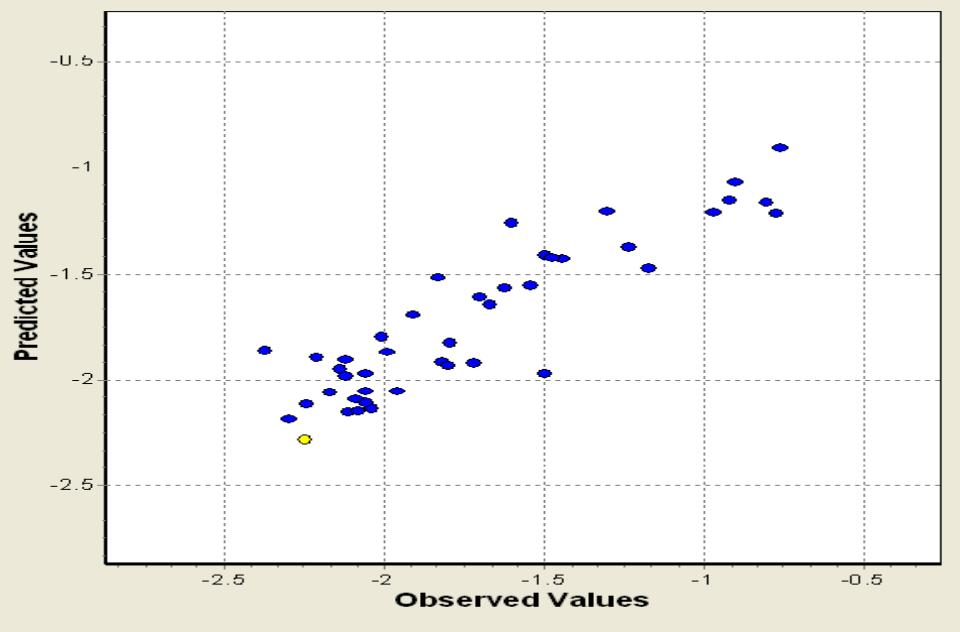
X37

Low α-glucosidase inhibitory activityin descending order X35>X38>X1>X13>X2>X37

Figure 3: Contribution of each atom/fragment towards biological activity. (Red-Positive, Blue-Negative and Green-No Effect).

Figure4 is a graph between observed and predicted values for biologic activity which reveals that there is good relation between experimental and predicted values.

Figure 4. Graph between observed and predicted values for biologic activity.



Conventional 3D-QSAR analysis

For conventional 3D-QSAR, e-Dragon was used to calculate 1D, 2D and 3D descriptors. The QSAR model was built using Genetic Algorithm (GA) implemented in Weka 3.7. GA has the advantage that it can select optimum number and set of descriptors for model generation. The best QSAR model based on five descriptors is as follows:

pIC_{50} = - 3.481 + 0.172 D.B. + 0.496 Nπ + 0.220 a_don + 0.851 MATS7v + 0.167 Hs

N = 43, s = 0.198, R^2 = 0.837, R^2_{adj} = 0.815, PRESS = 2.112, R^2_{pred} = 0.764

Where N is number of compounds in data set, R is the correlation coefficient, R^2 is the coefficient of determination, $R^2_{adj}$is adjusted coefficient of determination.

We might have developed a four parametric equation inadvertently. Therefore, in order to prove that the model is not a chance-comer, we have calculated R^2_{pred} and R^2_{adj} also. The rationale for using R^2_{adj} is that it varies with number of descriptors used and its value reduces with rise in the number of redundant descriptors. The high value of R, R^2 and R^2_{pred}indicates that model has excellent statistical significance. Moreover, the value of R^2_{adj}which is considered as better parameter to judge the predictive power compared to R^2, is close to the value of R^2 thereby validating the high predictive power of model [6-9].The QSAR model reveals that activity is directly related to D.B. (number of double bonds in the molecule), Nπ, a_don(number of donor atoms), MATS7v (a 2D autocorrelation descriptor), and Hs(number of hydrogen atoms). All these descriptors have positive coefficient which means the values of these descriptors should be increased to enhance the activity. From QSAR analysis, it is clear that D.B. (number of double bonds in the molecule) affects biological activity, but which double bonds have higher influence is uncertain from the conventional QSAR model. The success of GUSAR lies in giving answers to this type of specific questions. On the same basis, GUSAR tells about the type and position of π bonds, number of donor atoms, number of hydrogen atoms which affects the biological activity. Thus, a combine use of GUSAR and QSAR gives better, interpretable and clear idea of specific fragments/atoms that influences the biological activity.

Conclusions

From the present analyses, it is clear that combined use of GUSAR and QSAR is highly useful in finding the specific fragments/atoms which governs the biologic profile of the molecule. The structural features which control the activity but either neglected or missed by the conventional QSAR model(s) can be traced using GUSAR. In present work, GUSAR was able to indicate the specific double bonds, π bonds, number of donor atoms, number of hydrogen atoms which governs the α-glucosidase inhibitory activity of xanthone derivatives.

Acknowledgement

We are thankful to Weka and e-Dragon development team, ACD labs for providing ACD ChemSketch 12 freeware and GUSAR development team for providing GUSAR especially to Dr. Vladimir Poroikov, Dr.Alexey Lagunin and Dr.AlexeyZakharov for helpful discussions. Special thanks to Prof. Vijay Masand for allowing to use GUSAR and for helpful discussions throughout the work.

References:

[1]. Hussan S.S., Ph.D. Thesis, University of Lübeck, 2005.

[2]. Modi P., *Curr. Drug Dis. Tech.*, 2007, **4**, 39-47.

[3]. Saqib U., Siddiqi M.I., *Int. J. Integrat.Bio.*, 2009, **5 (1)**, 13-19.

[4]. Jawarkar R. D., Masand V. H., Patil K. N., Mahajan D. T. , Youssoufi M. H.,Hadda T. B., Kumbhare S. L., *Der Pharm. Chem.*,2010, **2(6)**, 302-310.

[5]. Jawarkar R. D., Mahajan D. T. , Masand V. H., Hadda T. B., Kurhade G. H., *Der Pharm.Lett.*, 2010, **2(6)**, 350-357.

[6]. Manikrao A. M., Khatale P. N., Jawarkar R. D., Vyas J. V., Mahajan D. T., Masand V. H., Hadda T. B., *J. Comput. Method. Mol. Design*,2011, **1 (2)**, 9-13.

[7]. Masand V. H., Patil K. N., Mahajan D. T., Jawarkar R. D., Nazerruddin G. M., *Der Pharm. Chem.*, 2010, **2(4)**, 298-308

[8]. Masand V. H., Patil K. N., Nazerruddin G. M., Jawarkar R. D., Bajaj S. O., *Org. Chem.: An Ind. J.*, 2010, **6(1)**, 30-38.

[9]. Christoph H., *Molecu. Diversity*, 2006, **10**, 147–158.

[10]. Kumar R., Kulkarni A., Jayaraman V. K.,Kulkarni B. D., *Inter. Electro. J. Mol. Des.*, 2004, **3**, 118–133.

[11]. Myint K. Z., Xie X. Q., *Int. J. Mol. Sci.*, 2010, **11**, 3846-3866.

[12]. Cedeño W., Agrafiotis D. K., J. Comp. Aid. Mol. Des., 2003, **17**, 255–263.

[13]. Poroikov V. V., Filimonov D. A., Borodina Y.V., Gloriozova T. A., Sitnikov V. B., Sadovnikov S.V., Sosnov A. V., *Pharma. Chem. J.*, 2004, **38(4)**, 188-190.

[14]. Filimonov D. A., Zakharov A.V., Lagunin A.A., Poroikov V.V., *SAR and QSAR Environ. Res.*, 2009, **20 (7-8)**, 679-709.

[15]. Filimonov D. A., Poroikov V. V., Borodina Y. V., Gloriozova T. A., *J. Chem. Inf. Comput. Sci.*, 1999**,39**, 666-670.

[16]. FilimonovD. A., Akimov D.V., Poroikov V. V., *Pharm. Chem. J.*, 2004, **38 (1)**, 21-24.s

[17]. www.acdlabs.com

[18]. Masand V. H., Mahajan D. T., Patil K. N., Chinchkhede K. D., Jawarkar R. D., Ben Hadda T., Alafeefy A. A, Shibi I. G, (2012) k-NN, quantum mechanical and field similarity based analysis of xanthone derivatives as α-glucosidase inhibitors, *Med Chem Res*, 2012, pp. 1-12, DOI:10.1007/s00044-012-9995-z.