Scientific paper

QSAR Modelling of [1,2,4]triazino [4,3-a]benzimidazole Acetic Acid Derivatives as Aldose Reductase Inhibitors

Somishetti V. Sambasivarao, Love K. Soni,* Arun K. Gupta and Satish G. Kaskhedikar

Molecular Modelling Study Group, CADD Laboratory, Department of Pharmacy, S.G.S.I.T.S., 23 Park Road, Indore-452003, M. P., India

** Corresponding author: E-mail: lovesoni@hotmail.com*
Tel./ Fax: +91-0731-2546031

Received: 24-10-2007

Abstract

The acetic acid derivatives of [1,2,4]triazino[4,3-*a*]benzimidazole as aldose reductase inhibitors were subjected for QSAR (quantitative structure activity relationship) modeling studies. A total 25 compounds were modelled in MOE. The QSAR model was generated using training set of 17 compounds employing sequential multiple linear regression analysis method. The internal consistency of the training set was confirmed by using leave-one-out (LOO) cross-validation method to ensure the robustness of the model. The model gave conventional and cross-validated r^2 values of 0.920 and 0.723, respectively. The predictive ability of model was further confirmed by a test set of eight compounds, which were not included in the model generation. The predicted activities of the test set were in good agreement with experimentally determined values. The model can be used to improve the activity of [1,2,4]triazino[4,3-*a*]benzimidazole acetic acid derivatives.

Keywords: QSAR, aldose reductase inhibitors, [1,2,4]Triazino[4,3-*a*]benzimidazole acetic acid derivatives, diabetes mellitus

1. Introduction

Diabetes mellitus is a widespread chronic disease, whose current worldwide prevalence of 150 million is predicted to double by year 2025.[1–3] It is always associated with degenerative long-term complications (retinopathies, nephropathies, neuropathies, angiopathies, atherosclerosis and cataracts) that make it one of the leading causes of blindness, renal failure and neuronal pathologies. The increased flux of glucose through the polyol pathway that occurs in hyperglycaemic conditions in tissues possessing insulin-independent glucose transport (nerve, retina, lenses and kidney) is a well-examined factor involved in the onset and progression of such chronic complications.[4–12] Aldose reductase (EC 1.1.1.21, ALR2) is the first enzyme of the polyol pathway and catalyses the NADPH-dependent reduction of glucose to sorbitol. The deprivation of NADPH and NAD+ and the intracellular accumulation of sorbitol result in biochemical imbalances which cause damage in target tissues. ALR2 inhibition thus represents an attractive approach to prevent or control the progression of chronic diabetic complications.[4–12] A variety of ALR2 inhibitors (ARIs) have been reported; however, in clinical studies many of them have exhibited low efficacy or a narrow spectrum of tissue activity, generally because of unfavourable pharmacokinetics, or have proved to produce toxic side-effects.[10,12–15] Currently, epalrestat (Figure 1) is the only ARI inhibitor available on the market.[15] This demands the development of potent inhibitors.

To gain insight into the structural and molecular requirements influencing the aldose reductase inhibitor activity, we herein describe QSAR analysis of [1,2,4]triazino[4,3-*a*]benzimidazole acetic acid derivatives.[16] The relevance of the best QSAR model obtained for the design of novel derivatives should be assessed not only in terms of predictivity, either internal or external, but also in terms of their ability to provide a chemical and structural explana-

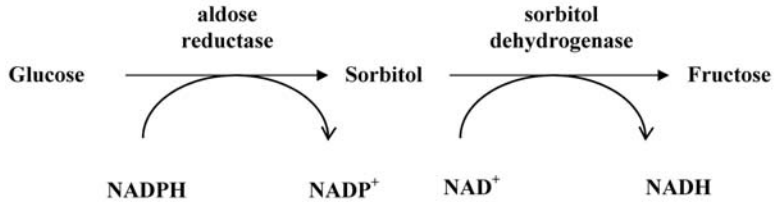


Figure 1: Polyol pathway

tion for their binding interaction. Here we propose a model for the aldose reductase inhibitors and present minimal structural requirements for an aldose reductase inhibitor. These results could serve as a guideline in design of more potent and selective aldose reductase inhibitors.

2. Experimental Methods

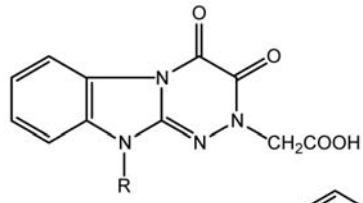
The aldose reductase inhibitory activity data of [1,2,4]triazino[4,3-*a*]benzimidazole acetic acid derivatives, was taken from the reported work of *Da* Settimo *et al* [16] (Table 1).

The biological activity data (IC_{50} in μm) was converted to negative logarithmic mole dose (pIC_{50}) for quantitative structure activity relationship analysis. The molecular modelling study was performed on a P-III processor using MOE[17] and the regression analysis program VALSTAT.[18] The molecular structures of all twenty-three compounds were sketched using the builder module software and energy minimized via steepest descent, conjugate gradient and truncated Newton method in sequence using MMFF94 as force field with energy tolerance value of root mean square gradient 0.001 kcal/mol and maximum number of iteration set to 1000. Conformational search of each energy-minimized structure was performed using stochastic approach. Stochastic conformational search method is similar to RIPS method,[19] which generate new molecular conformation by randomly perturbing the position of each coordinate of each atom in molecule followed by the energy minimization. All conformers generated for each structure were analyzed in conformational geometries panel with great care and the lowest energy conformation of each structure was selected and added to a molecular database to compute various physicochemical properties from three classes: 2D-descriptors based on atoms and connection information of the molecules, i3D-descriptors used three dimensional coordinate information about each molecule, which are invariant to rotations and translations of the conformation, and x3D descriptors which were supported by three dimensional coordinate information, require an absolute frame of reference using QuaSAR module.[20–25]

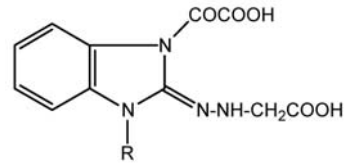
Series was divided into a training set of 17 compounds and a test set of 8 compounds on the basis of structural diversity and cover the complete range of variation in inhibitory activity. The data was transferred to the statisti-

Table 1: Structure and ALR2 inhibition data of acid derivatives.

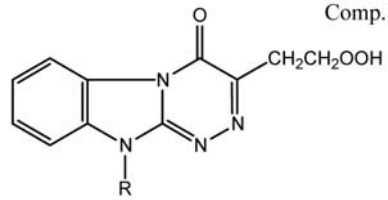
Comp. No.	R	IC_{50} [a]	pIC_{50} [b]
1	CH_3	24.80	4.606
2	$CH_2CH_2CH_3$	37.20	4.429
3	$CH_2C_6H_5$	0.36	6.444
4	$CH_2C_6H_4$–4–CH_3	13.30	4.876
5	$CH_2C_6H_4$–4–OCH_3	42.60	4.371
6	$CH_2C_6H_4$–4–Cl	4.15	5.382
7	$CH_2C_6H_4$–4–F	4.58	5.339
8	$CH_2C_6H_4$–4–CF_3	23.90	4.622
9	$CH_2C_6H_3$–3,4–F_2	4.42	5.355
10	$CH_2C_6H_3$–2–F–4–Br	4.47	5.350
11	CH_2COOH	13.5	4.870
12	CH_3	108.6	3.964
13	$CH_2CH_2CH_3$	46.5	4.333
14	$CH_2C_6H_5$	4.50	5.347
15	$CH_2C_6H_4$–4–CH_3	45.90	4.338
16	$CH_2C_6H_4$–4–OCH_3	44.50	4.352
17	$CH_2C_6H_4$–4–Cl	10.00	5.000
18	$CH_2C_6H_4$–4–F	14.80	4.830
19	$CH_2C_6H_4$–4–CF_3	2.63	5.580
20	$CH_2C_6H_3$–3,4–F_2	9.72	5.012
21	$CH_2C_6H_3$–2–F–4–Br	12.50	4.903
22	CH_2COOH	236.0	3.627
23	H	35.90	4.445
24	CH_3	17.00	4.770
25	$CH_2C_6H_5$	5.44	5.264



Comp. No. 1-11



Comp. No. 12-22



[a] *in vitro* IC_{50} (50% inhibitory concentration in μM).
[b] negative logarithmic of IC_{50} value in mole.

cal program in order to establish a correlation between physicochemical parameters as independent variables and aldose reductase inhibitory activity as dependent variable. The sequential multiple linear regression analysis method was employed. In sequential multiple regression the program searches for all permutations and combinations sequentially for the data set. The best model was selected from the various statistically significant equations on the basis of the observed squared correlation coefficient (r^2), the standard error of the estimate (***SEE***), sequential Fischer test (***F***), the bootstrapping squared correlation coefficient (r^2_{bs}), the bootstrapping standard deviation (S_{bs}), the cross-validated squared correlation coefficient using leave-one-out procedure (q^2), chance statistics (***Chance***), outliers (***Z-score value***), and the predictive squared correlation coefficient of test set (r^2_{pred}).

3. Results and Discussion

In order to explore the physico-chemical requirements for aldose reductase inhibition in terms of molecular characteristics, the series was subjected to QSAR analysis. Training set was used to explore the conformational and geometrical related physicochemical properties that could help to understand the probable binding site of the drug with the enzyme. A correlation was established between physicochemical parameters and ALR2 inhibitory activity using sequential multiple linear regression technique. Several statistically significant equations were obtained, from which the following equation was chosen as model.

pIC_{50} = 0.386 (±0.079) *a_aro* –0.574 (±0.135) *diameter* –4.407 (±0.968) brotR+8.083 n = 17, r = 0.920, r^2 = 0.846, r^2_{adj} = 0.811, *SEE* = 0.216, *QF* = 4.260, *PE* = 0.0248, F = 23.868, FIT = 2.754, AIC = 0.075

Model has a better correlation coefficient (r = 0.920), which account for 92.0% of the variance in the activity. The multi-variant model shows that the dependent variable can be predicted from a linear combination of the independent variables. The P value is less than 0.001 for each physicochemical parameter involved in model generation. The data showed overall internal statistical significance level better than 99.9% as it exceeded the tabulated $F_{(3,13\ \alpha\ 0.001)}$ = 11.9.

A high correlation coefficient alone is not enough to select the equation as a model and hence various statistical approaches were employed to confirm the robustness and the practical applicability of the equations. The model was further tested for outlier by the Z-score method and no compound was found to be an outlier (Table 2) which suggested that the model is able to explain the structurally diverse analogues of the series and is helpful in designing more potent compounds using physicochemical parameters. In randomized biological activity test model shows that chance correlation was less than 0.1%.

The internal consistency of the training set was confirmed by using leave-one-out (LOO) cross-validation method to ensure the robustness of the model. A q^2 value (in the biological activity data of leave one compound) of 0.3 corresponds to a confidence limit greater than 95%, which minimizes the risk of finding significant explanatory equation for the biological activity just by mere chance. The cross-validated squared correlation co-efficient (q^2 = 0.723), predictive residual sum of square (S_{PRESS} = 0.290) and standard error of prediction (S_{DEP} = 0.253) suggested a good internal consistency as well as predictive ability of the biological activity with low S_{DEP} (Figure 2 and Table 2). Expressions that have significant internal consistency may not be applicable for the analogs, which were never used in generation of correlation. Therefore, the predictive power of model was further confirmed by a test set of eight compounds.

The robustness and wide applicability of the model was further explained by significant r^2_{pred} value (0.348) of the test set data (Figure 3 and Table 3). In general the model fulfills the statistical validation criteria to a signifi-

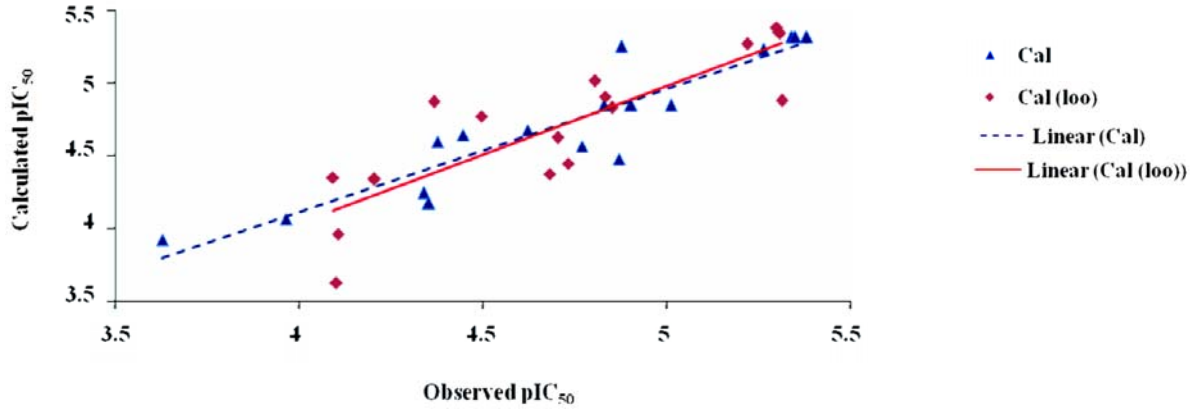


Figure 2: The scattered plot between observed pIC_{50} and calculated pIC_{50} values.

Table 2: Calculated and predicted pIC_{50} values (by LOO method) of training set with residual and Z-score value of training set.

Comp No.	Observed pIC_{50}	Calculated pIC_{50}	Residual	Z-value	Predicted pIC_{50} (LOO)	Residual (LOO)
4	4.876	5.251	–0.375	–1.926	5.314	–0.438
5	4.376	4.596	–0.220	–1.128	4.682	–0.305
6	5.382	5.313	0.069	0.352	5.300	0.082
7	5.339	5.313	0.026	0.132	5.308	0.031
8	4.622	4.677	–0.055	–0.285	4.703	–0.081
10	5.350	5.313	0.037	0.186	5.306	0.044
11	4.870	4.477	0.393	2.018	4.366	0.504
12	3.964	4.066	–0.102	–0.524	4.105	–0.141
15	4.338	4.242	0.096	0.494	4.204	0.134
16	4.352	4.173	0.179	0.919	4.091	0.261
18	4.830	4.847	–0.017	–0.087	4.851	–0.021
20	5.012	4.847	0.165	0.851	4.805	0.207
21	4.903	4.847	0.056	0.290	4.832	0.071
22	3.627	3.923	–0.296	–1.518	4.101	–0.474
23	4.445	4.645	–0.200	–1.028	4.731	–0.286
24	4.770	4.565	0.205	1.051	4.495	0.275
25	5.264	5.225	0.039	0.203	5.219	0.045

cant extent to be a useful theoretical base for proposing more active compounds.

In the model a_aro contributed positively while diameter and brotR contributed negatively to the observed variance in the activity. a_aro is representative of a number of aromatic atoms and might be responsible for π–π interaction with the receptor and aromatic stacking with aromatic amino acids. brotR is fraction of rotatable bonds and depends on molecular flexibility, as measured by fraction of rotatable bonds. This descriptor affects rigidity or flexibility of molecules and result in the partitioning behavior of solutes. The molecules are in their vast majority hydrophilic, with a fraction of rotatable bonds commonly between 0.1 and 0.4, and are also crucial for hydrophilic interaction with receptor. Diameter is the largest value in the distance matrix and it might be crucial for accommodation of the molecule in the receptor pocket.

Table 3: Observed and predicted pIC_{50} values of test set with residual of test set.

Comp No.	Observed pIC_{50}	Predicted pIC_{50}	Residual
1	4.606	4.682	–0.076
2	4.429	4.537	–0.108
3	6.444	5.888	0.556
9	5.355	5.313	0.042
13	4.333	4.021	0.312
14	5.347	5.421	–0.074
17	5.000	4.847	0.153
19	5.580	4.242	1.338

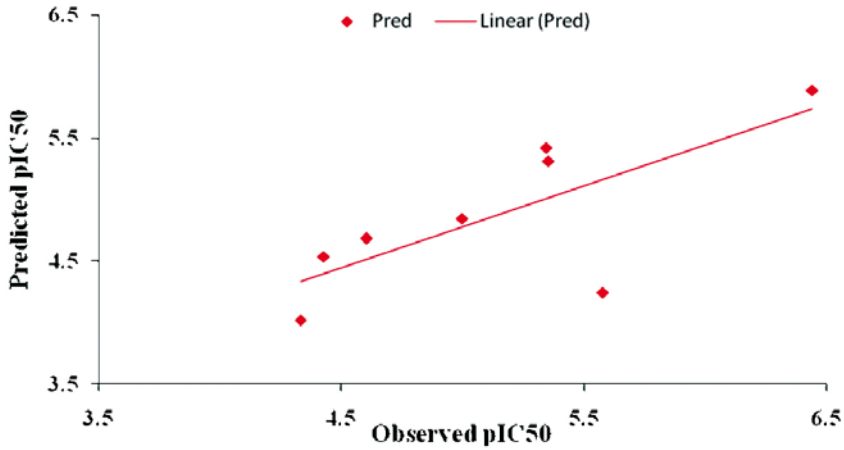


Figure 3: The scattered plot between observed pIC_{50} and predicted pIC_{50} values with residual presentation for test set.

On the basis of above made discussion we can assert that the model could be explored further to design potent [1,2,4]triazino[4,3-*a*]benzimidazole acetic acid derivatives.

4. Acknowledgements

Authors are grateful to the Director, Shri G. S. Institute of Technology and Science, Indore, for providing facilities for this work and Tata Elxsi, India for providing the MOE software. The author AKG is grateful to CSIR, New Delhi for providing senior research fellowship.

5. References

1. J. Diamond, *Nature* **2003**, *423*, 599–602.
2. P. Zimmet, *Diabetes Metab.* **2003**, *29*, 6S9–6S18
3. H. King, R. E. Aubert, W. H. Herman, *Diabetes Care* **1998**, *21*,1414–1431.
4. M. Brownlee, *Nature* **2001**, *414*, 813–820.
5. S. Suzen, E. Buyukbingol, *Curr. Med. Chem.* **2003**, *10*, 1329–1352.
6. D. Porte Jr, M. W. Schwartz, *Science* **1996**, *272*, 699–700.
7. P. F. Kador, J. H. Kinoshita, N. E. Sharpless, *J. Med. Chem.* **1985**, *28*, 841–849.
8. P. F. Kador, *Med. Res. Rev*. **1988**, *8*, 325–352.
9. C. Yabe-Nishimura, *Pharmacol. Rev.* **1998**, *20*, 21–33.
10. L. Costantino, G. Rastelli, G. Cignarella, P. Vianello, D. Barlocco, *Exp. Opin. Ther. Patents*, **1997**, *7*, 843–858.
11. B. H. Wolffenbuttel, T. W. van Haeften, *Drugs* **1995**, *50*, 263–288.
12. S. Miyamoto, *Chem-Bio. Informatics J.* **2002**, *2*, 74–85.
13. E. R. Larson, C. A. Lipinski, R. Sarges, *Med. Res. Rev.* **1988**, *8*, 159–186.
14. L. Costantino, G. Rastelli, P. Vianello, G. Cignarella, D. Barlocco, *Med. Res. Rev.* **1999**, *19*, 3–23.
15. L. Costantino, G. Rastelli, M. C. Gamberoni, D. Barlocco, *Exp. Opin. Ther. Patents* **2000**, *10*, 1245–1262.
16. F. Da Settimo, G. Primofiore, A. Da Settimo, C. La Motta, S. Taliani, F. Simorini, E. Novellino, G. Greco, A. Lavecchia, E. Boldrini, *J. Med. Chem.* **2001**, *44,* 4359–4369
17. MOE User's Manual, Tata Elxsi Ltd., Bangalore.
18. A. K. Gupta, B. M. Arockia, S. G. Kaskhedikar, *Indian J. Pharm. Sci.* **2004**, *66*, 396–402.
19. D. M. Ferguson, D. J. Raber, *J. Am. Chem. Soc.*, **1989**, *111*, 4371–4378.
20. L. H. Hall, L. B. Kier: in D. Boyd, K. Lipkowitz (Ed.): Reviews of Computational Chemistry, VCH Publishers, Inc., 1991, pp 367–422.
21. L. H. Hall and L. B. Kier, *Eur. J. Med. Chem. – Chimica Therapeutica.* **1977**, *12*, 307–312.
22. S. A. Wildman, G. M. Crippen, *J. Chem. Inf. Comput. Sci. 1999*, *39*, 868–837.
23. A. Lin, QuaSAR-descriptor. J. Chem. Compute. Group
24. P. A. Labute, *J. Mol. Graph. Model.* **2000**, *18*, 464–477.
25. N. Baurin, J. -C. Mozziconzcci, E. Arnoult, P. Chavette, C. Marot, L. Morin-Allory, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 276–285.

Povzetek

V prispevku je podana modelna QSAR študija derivatov [1,2,4]triazino[4,3-*a*]benzimidazola z ocetno kislino, kot potencialnih inhibitorjev aldolne reduktaze. Modeliranih je bilo 25 spojin z uporabo MOE metode. QSAR model je bil generiran na setu 17 spojin z uporabo zaporedne mnogokratne linearne regresijske analize. Doslednost izbranega seta in njegova trdnost je bila potrjena z uporabo leave-one-out (LOO) križno-validacijske metode. Model daje konvencionalno r^2 vrednost 0.920 in križno-validacijsko r^2 vrednost 0.723. Zmožnost predvidevanja generiranega modela je bila potrjena s testom na setu osmih spojin, ki niso bile vključene pri generiranju modela. Sposobnost predvidevanja modela na testnem setu se je dobro ujemala z eksperimentalno določenimi vrednostmi. Generiran model bi lahko uporabili za izboljšanje aktivnosti derivatov [1,2,4]triazino[4,3-*a*]benzimidazola z ocetno kislino.