3D-QSAR MIFs Studies on 3,5-substituted-1,4,2-dioxazoles Derivatives Using Open3DQSAR Tools

M. Abul Kashem Liton¹*, A.C. Bhowmick¹, M. Ashraf Ali²

¹Department of Chemistry, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh
²Bangladesh Chemical Industries Corporation (BCIC), Dhaka, Bangladesh
*Corresponding author: litonchem@hotmail.com

Abstract The MIFs had been applied to a data set of thirty three (33) dioxazole derivatives to generate the 3D-QSAR model at various 3D grid spacing. An excellent cross-validated correlation coefficient $q^2$ (0.903) and conventional correlation coefficient $r^2$ (0.985) were obtained at a 2.0 Å 3D grid spacing, indicating the statistical significance of this class of compounds. The calculated biological activities showed a high degree of agreement with experimental values.

Keywords  MIFs, Open3DQSAR, PLS, LOO, NIPALS

1. Introduction

Amoebiasis [1] is an intestinal infection occurred by the protozoan parasite Entamoeba histolytica (E. histolytica) [2,3]. E. histolytica has the capacity to destroy almost all tissues of the human body, the intestinal mucosa, liver and to lesser extent the brain, skin, cartilage and even bone [4]. Amoebiasis occurs worldwide [5], but is mostly seen in tropical and developing countries, which have bad sanitary and hygienic practices. Acute amoebiasis can present as diarrhoea or dysentery with frequent, small and often bloody stools. Chronic amoebiasis can present with gastrointestinal symptoms plus fatigue, weight loss and occasional fever. Extraintestinal amoebiasis can occur if the parasite spreads to other organs, most commonly the liver where it causes amoebic liver abscess. Amoebic liver abscess presents with fever and right upper quadrant abdominal pain. Amoebic colitis results from ulcerating mucosal lesions caused by the release of parasite-derived hyaluronidases and proteases. Hepatic infection occurs as a consequence of entry of the parasite into the afferent bloodstream. Paromomycin or iodoquinol are the drugs of choice for treatment of asymptomatic infections proven to be caused by E. histolytica [6,7]. For symptomatic intestinal infection, or extraintestinal, infections (e.g., hepatic abscess), the drugs of choice are metronidazole (Flagyl) or tinidazole (Tindamax), immediately followed by treatment with iodoquinol or paromomycin [6,7]. Though medicinal chemistry has vastly developed but there are many difficulties remaining in the determination of E. histolytica for the treatment of amoebiasis. Recent studies tried to improve the treatment of this infection by developing antiamoebic therapy [8], a set of dioxazoles derivatives showed better activity than the reference drug metronidazole; furthermore, they are non toxic to the human kidney epithelial cells. In the other word, QSAR studies were reported to identify important structural features responsible for the antiamoebic activity [9,10]. The most important step is to find the possible compounds with desired biological activity. The quantitative structure-activity relationships (QSAR) are now modern media for drug design. Over the years of development, many methods, algorithms and techniques have been discovered and applied in QSAR studies [11, 12]. Today, QSARs are being applied in many disciplines with much emphasis in drug design. Three-dimensional quantitative structure activity relationships (3D-QSAR) analysis [13] is now the most widely used and improved technique.

2. Materials and Methods

2.1. Experimental Data

3,5-substituted-1,4,2-dioxazoles derivatives and their antiamoebic activities were taken from the literature reported by Bhat et al. [8]. A comparative study of in vitro antiamoebic activity of all synthesized dioxazole derivatives (33 compounds) were performed against HM1:IMSS strain of Entamoeba histolytica. The experimental $IC_{50}$ values of all compounds in μM (micromole) showed very good results compared with $IC_{50}$ values of metronidazole. Most of them are non toxic to the human kidney epithelial cells. The toxicity studies against human kidney epithelial cell line showed that all the compounds were non-toxic with $IC_{50}$ values in the range of 0.41-1.80 μM [8].
Table 1. The structure of dioxazole derivatives (1-33) whose biological activity had been accounted for 3D-QSAR study.

| Comp. ID | R₁ | R₂ | Comp. ID | R₁ | R₂ | Comp. ID | R₁ | R₂ |
|----------|----|----|----------|----|----|----------|----|----|
| 1        | H  | Cl | 12       | H  | Cl | 23       | H  | Cl |
| 2        | H  | Cl | 13       | H  | Cl | 24       | H  | Cl |
| 3        | H  | CH₃| 14       | H  | CH₃| 25       | H  | CH₃|
| 4        | H  | C₂H| 15       | H  | C₂H| 26       | H  | C₂H₅|
| 5        | CH₃| Cl | 16       | CH₃| Cl | 27       | CH₃| Cl |
| 6        | H₂  | C-CH₃| 17 | H₂  | C-CH₃| 28 | H₂  | C-CH₃|
| 7        | CH₃| B   | 18       | CH₃| Br | 29       | CH₃| Br |
| 8        | H₂  | C-CH₃| 19 | H₂  | C-CH₃| 30 | H₂  | C-CH₃|
| 9        | CH₃|     | 20       | CH₃|     | 31       | CH₃|     |
| 10       | H   | Cl | 21       | H   | Cl | 32       | H   | Cl |
| 11       | H   | Cl | 22       | H   | Cl | 33       | H   | Cl |

*Structures were taken from Ref [8].*
2.2. Data Set

3D-QSAR modeling was carried out through MIFs analysis by dividing the dataset of 33 molecules into training set [14] of 27 molecules and test set of 6 molecules in a random manner (the testing set is marked by *). According to research methodology, all experimental $IC_{50}$ values (μM) were converted to negative logarithm of $IC_{50}$, i.e., $pIC_{50}$ and used as dependent variable in 3D-QSAR study. The structures of these compounds are given in Table 1 and their biological activities are incorporated in Table 4.

3. Computational Details

3.1. Structure Generation

At first all the 3D structures were generated by Gauss View 03 and minimization was performed with the MOPAC2012 [15] software using semi-empirical method (AM1). All geometric variables were finally optimized for each compound using Gaussian03W [16] version 6.0 program at the level of B3LYP/6-31G(d,p) theory and the low energy conformers were ensured with all real frequency by the frequency calculations and this lowest energy conformations were used in the open3DQSAR tools [17] for MIFs analysis.

3.2. Alignment

The molecules were superimposed using the atom-based alignment by the open3DALIGN tools [18, 19] given in Figure 2. From the data set, the compound, 10 shown in Figure 1 was selected as the template to construct other compounds because of its high biological activity and representative chemical structure, and the alignment was completed by open3DALIGN workstation [18, 19]. Except for some special notes, default values were chosen. Then their geometries were optimized by the RMS gradient [20] criterion method on MMFF94s force-field implemented in TINKER by using command option of open3DALIGN. The energy convergence criterion in alignment is 0.01 kcal/mol.

3.3. Generating MIFs

The Molecular Interaction Fields (MIFs) are the interaction energies between a probe atom (or a molecule) and a set of aligned molecules, which are used to establish the three-dimensional quantitative structure activity relationship (3D-QSAR) equations. To generate the MIFs, a probe atom is systematically moved from one point to another for each aligned molecule within a defined 3D grid [21]. At each grid point, the interaction energy is calculated between the probe and the target molecule. In this study, the 33 aligned molecules were placed in various 3D cubic lattice spacing. The steric (van der Waals) and electrostatic (Coulombic) interaction energies were calculated for each molecule at each grid point using an alkyl carbon probe (default) with automatically assigned charges using OpenBabel utilities. Energies lower than -40.0 kcal/mol and greater than 40.0 kcal/mol were cutoff because a few high values in the dataset may severely bias the model.

3.4. MIFs Study

The MIFs analysis described here were performed on an open3DQSAR tools [17] from Sourceforge.net using Partial Least Square (PLS) [22] technique through the NIPALS algorithm methodology [23]. Steric and electrostatic fields are included in the analysis and various grid spacing was applied. All of the 33 compounds for the data set were superimposed (see Figure 2) onto a template using an atom by atom least-square fit, and one of the most active compounds (10) was used as the reference molecule. After alignment the molecules were put to the various 3D grid [21] spacing. The steric and electrostatic fields were then calculated using a ‘CR’ Alkyl Carbon atom with default charge and the cutoff energy was set between -40 to 40 kcal/mol. Regression analysis of the resulting field matrix was performed by Partial Least Squares (PLS) [22] technique. To obtain the 3D-QSAR models, PLS analysis was performed using steric and electrostatic field alone and also combination. Cross-validation [24] in PLS was carried out using the leave-one-out method (LOO) [25, 26] to check the predictive ability of the models and to determine the optimal number of components to be used in the final 3D-QSAR models. The optimum number of PLS components for the final, non-validated analysis was chosen based on the smallest standard error of estimation (SEE) values from the cross-validated analysis.
4. Result and Discussion

To develop an effective 3D-QSAR model some parameters such as the cross-validated correlation coefficient ($q^2$), non-cross-validated correlation coefficient ($r^2$), standard error estimate (SEE) and F-statistic values have been taken under consideration. The LOO cross-validation was carried out first for 3D-QSAR model. Then the number of components identified in the LOO cross-validation process was used in the final non-cross-validated PLS run. The optimal number of components was determined by selecting highest $q^2$ value.

The statistical results of MIFs studies are summarized in Table 2. The predicted antiamoebic activities for the dioxazole derivatives versus their experimental activities are listed in Table 4 and the correlation between the predicted activities and the experimental activities is depicted in Figure 3.

The MIFs based PLS calculation resulted in several models and among them, the final model selection is an important issue. To obtain the 3D-QSAR models, PLS analysis was performed using each of the steric and electrostatic MIFs alone and also in combination varying on 3D grid spacing. Seven types of model were produced on varying the grid spacing (see Table 2). All the models comparatively showed good statistical results except the model 5 (Table 2). Model-3, which uses both steric and electrostatic fields on 2.0Å grid spacing, was thus chosen as the working MIFs model, whose validity and predictability were assessed by the $r^2$ value of 0.985 and $q^2$ value of 0.903 with 5 components, $F$ is 293.46, and a standard error of estimation (SEE) of 0.035. The steric and electrostatic contributions were 79.89% and 20.11% respectively. A graphical inspection of the experimental with calculated ($pIC_{50}$) values immediately indicated that the overall fit of the molecule was satisfactory for all the models proposed in Table 2. The results of the analyses are shown in Table 2, from which it is obvious that the best model is 3; that is to say, the combined fields (steric and electrostatic) at a 2.0Å gave the best statistical results.

The external predictive ability of the MIFs model is extremely important in terms of the applicability of the MIFs model. Therefore, it was decided to use the $r^2_{pred}$ as a criterion for final selection of the one best model. As reflected by the Table 3, all the models showed comparatively good $r^2_{pred}$ values except model 5 and model 3 has the highest $r^2_{pred}$ value of 0.974 and hence the lowest standard error of prediction (SEP) value of 0.048 for the test set (six compounds, indicated by *, Table 4). Therefore, model 3 was selected as the best MIFs model.

### Table 2. The statistical analysis of training set through MIFs studies.

| Model No | 3D-Grid spacing | Compounds | Components | $q^2$ | $r^2$ | SEE | $F$ | Steric contribution (%) | Electrostatic contribution (%) |
|----------|------------------|-----------|------------|-------|-------|-----|----|--------------------------|-------------------------------|
| 1        | 1.0 Å            | 27        | 5          | 0.862 | 0.978 | 0.042 | 188.96 | 75.91 | 24.09 |
| 2        | 1.5 Å            | 27        | 5          | 0.862 | 0.980 | 0.039 | 222.63 | 76.25 | 23.75 |
| 3        | 2.0 Å            | 27        | 5          | 0.903 | 0.985 | 0.035 | 293.46 | 79.89 | 20.11 |
| 4        | 2.0 Å            | 27        | 5          | 0.877 | 0.978 | 0.042 | 226.39 | 100  | 0    |
| 5        | 2.0 Å            | 27        | 5          | 0.272 | 0.763 | 0.137 | 21.55  | 0    | 100  |
| 6        | 2.5 Å            | 27        | 5          | 0.859 | 0.975 | 0.044 | 173.23 | 78.86 | 21.14 |
| 7        | 3.0 Å            | 27        | 5          | 0.855 | 0.975 | 0.044 | 172.12 | 74.55 | 25.45 |

### Table 3. The statistical analysis of test set through MIFs studies.

| Model no | 3D-Grid spacing | $r^2_{pred}$ | SEP | Steric contribution (%) | Electrostatic contribution (%) |
|----------|------------------|--------------|-----|--------------------------|-------------------------------|
| 1        | 1.0 Å            | 0.954        | 0.064 | 75.91 | 24.09 |
| 2        | 1.5 Å            | 0.962        | 0.058 | 76.25 | 23.75 |
| 3        | 2.0 Å            | 0.974        | 0.048 | 79.89 | 20.11 |
| 4        | 2.0 Å            | 0.965        | 0.056 | 100  | 0    |
| 5        | 2.0 Å            | 0.021        | 0.295 | 0    | 100  |
| 6        | 2.5 Å            | 0.948        | 0.067 | 78.86 | 21.14 |
| 7        | 3.0 Å            | 0.947        | 0.068 | 74.55 | 25.45 |

Å means Angstrom (1Å =10⁻⁸ cm)
Table 4. Experimental and calculated pIC$_{50}$ values as well as residues for dioxazole derivatives (1-33)*.

| Comp ID | pIC$_{50}$ (Exp.) | pIC$_{50}$ (Cal.) | Residue | Comp ID | pIC$_{50}$ (Exp.) | pIC$_{50}$ (Cal.) | Residue |
|---------|-------------------|-------------------|---------|---------|-------------------|-------------------|---------|
| 1       | -0.092            | -0.017            | -0.075  | 18      | 0.433            | 0.445            | -0.012  |
| 2*      | -0.292            | -0.196            | -0.096  | 19      | 0.367            | 0.364            | 0.003   |
| 3       | 0.494             | 0.451             | 0.043   | 20      | 0.238            | 0.265            | -0.027  |
| 4       | 0.486             | 0.440             | 0.046   | 21      | -0.208           | -0.264           | 0.056   |
| 5       | 0.461             | 0.440             | 0.021   | 22*     | -0.041           | -0.044           | 0.003   |
| 6       | 0.400             | 0.385             | 0.015   | 23      | 0.053            | -0.008           | 0.061   |
| 7       | 0.364             | 0.373             | -0.009  | 24      | -0.276           | -0.234           | -0.042  |
| 8*      | 0.408             | 0.366             | 0.042   | 25      | 0.433            | 0.426            | 0.007   |
| 9       | 0.210             | 0.178             | 0.032   | 26      | 0.373            | 0.415            | -0.042  |
| 10      | -0.387            | -0.348            | -0.039  | 27*     | 0.433            | 0.423            | 0.010   |
| 11      | -0.143            | -0.129            | -0.014  | 28      | 0.389            | 0.399            | -0.010  |
| 12      | 0.083             | 0.053             | 0.030   | 29      | 0.417            | 0.382            | 0.035   |
| 13      | -0.125            | -0.172            | 0.047   | 30      | 0.403            | 0.399            | 0.004   |
| 14      | 0.452             | 0.494             | -0.042  | 31      | 0.199            | 0.210            | -0.011  |
| 15*     | 0.444             | 0.483             | -0.039  | 32      | -0.319           | -0.319           | 0.000   |
| 16      | 0.468             | 0.491             | -0.023  | 33*     | -0.066           | -0.100           | 0.034   |
| 17      | 0.441             | 0.462             | -0.021  |         |                  |                  |         |

*a The experimental biological activities were taken from Ref [8]

Indicates the test set

Figure 3. The graph of experimental versus calculated pIC$_{50}$ activities of dioxazole derivatives

5. Conclusions

We have carried out MIFs studies for thirty three (33) dioxazole derivatives against experimental biological activities. The atom based alignment with varying the 3D grid spacing method was used to provide the model for MIFs analysis. Our present studies have established that the model derived through MIFs studies is quite reliable and significant. We have investigated that the PLS analysis at 2.0Å 3D grid spacing by Open3DQSAR tools has presented an excellent statistical results in terms of $q^2$ and $r^2$ values for dioxazole analogues and showed a high degree of agreement with the experimental antiamoebic activities. So it is to say that the dioxazole analogues have a good antiamoebic activity. In addition dioxazole may also find roles in the treatment of gastrointestinal infections caused by the protozoan parasite *E. histolytica*.

Acknowledgement

The authors would like to thank the scientist for the development of Open3DSAR tools and also grateful to free for the public users.

REFERENCES
[1] J. Harries. Amoebiasis, a review, J. R. Soc. Med., vol.75, no.3, 191-197, 1982.

[2] WHO, “Amoebiasis, Report of a WHO Expert Committee.” WHO Technical Report Series, 421, 1-52, 1969.

[3] WHO, “WHO/PAHO/UNESCO report, A consultation with expert on amoebiasis. Mexico City, Mexico 28-29 January 1997.” Epidemiological Bulletin, vol.18, no.1, 13-14, 1997.

[4] K.V.G.K. Tilak. Lecture on Amoebiasis, 2nd February, APICON2013 Conference, Coimbatore, Tamil Nadu, India, 2013.

[5] L. Samuel, J. Stanley, L. Sharon, Reed. Initial Microbes and microbial toxin: paradigms for microbialmucosal. Interactions VI. Entamoeba histolytica: parasite-host interactions, Am. J. Phy. Gastrointest Liver Physiol., vol.280, no.6, 1049-1053, 2001.

[6] W. A. Petri, U. Singh. Enteric Amebiasis. In: Tropical Infectious Diseases: Principles, pathogens, and practice, 2nd edition. R. Guerrant, D. H. Walker, P. F. Weller (eds.), Elsevier, Philadelphia, 967-983, 2006.

[7] J. I. Ravdin, W.M. Stauffer. Entamoeba histolytica (amoebiasis). In: Mandell GL, Bennett JE, Dolin R, editors. Mandell, Bennet, & Dolin: Principles and Practice of Infectious Diseases. 6th ed. Philadelphia: Churchill Livingstone, 3097–3111, 2005.

[8] A. R. Bhat, F. Athar, A. Azam. New derivatives of 3,5-substituted-1,4,2-dioxazoles: Synthesis and activity against Entamoeba histolytica, Eur. J. Med. Chem., vol.44, 926-936, 2009.

[9] S. Mbarki, K. Dguigu, M. El Hallouai. Construction of 3D-QSAR models to predict antiamoebic activities of pyrazoline and dioxazoles derivatives, J. Mater. Environ. Sci., vol.2, no.1, 61-70, 2011.

[10] N. Adhikari, M. K. Maiti, J. Tarun. Exploring structural requirements of 1-N-substituted thiocarbamoyl-3-phenyl-2-pyrazolines as antiamoebic agents using comparative QSAR modeling, Bioorganic& medicinal chemistry letters, vol.20, 4021-4026, 2010.

[11] I. He, and P. C. Jurs. Assessing the reliability of a QSAR model's predictions, Journal of Molecular Graphics and Modelling, vol.23, 503-523, 2005.

[12] D. V. Eldred, C. L. Weikel, P. C. Jurs, and K. L.E. Kaiser. Prediction of fathead minnow acute toxicity of organic compounds from molecular structure, Chem. Res. Toxicol., vol.12, 670-678, 1999.

[13] H. Kubinyi, 3D-QSAR in Drug Design. Theory, Methods and Applications, Leiden: (ed.) ESCOM, 1993.

[14] Z. C. Shi et al. Modulation of peripheral serotonin levels by novel tryptophan hydroxylase inhibitors for the potential treatment of functional gastrointestinal disorders, J. Med. Chem., vol.51, 3684-3687, 2008.

[15] J. P. James, Stewart. MOPAC2012, Stewart Computational Chemistry, Version 12.347W, 2012.

[16] M. J. Frisch et al. Gaussian 03W, Version 6.0, Gaussian, Inc., Pittsburgh PA., 2003.

[17] P. Tosco, T. Balle. An open-source software aimed at highthroughput chemometric analysis of molecular interaction fields. Journal of Molecular Modeling, vol.17: 201-208, 2011.

[18] N. J. Richmond, P. Willett, R. D. Clark. Alignment of three-dimensional molecules using an image recognition algorithm, J. Mol. Graph. Model.vol.23, 199-209, 2004.

[19] J. Taminau, G. Thijs, H. De Winter. Pharao: Pharmacophore alignment and optimization, J. Mol. Graph. Model. vol.27, 161-169, 2008.

[20] M. Petitjean. Interactive Maximal Common 3D Substructure Searching with the Combined SDM/RMS Algorithm, Computers Chem., vol.22, 463-465, 1998.

[21] R. D. Cramer III, M. Milne. Abstract of Papers of the Am. Chem. Soc. M. 44, Computer Chemistry Section, 1979.

[22] L. Stahle, S. Wold. Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study, J. Chemom., vol.11, 185-196, 1987.

[23] S. Wold, M. Jöjöström, L. Eriksson. PLS-regression: a basic tool of chemometrics, Chemometrics Intell. Lab. Syst., vol.58, 109-130, 2001.

[24] R. D. Cramer, J. D. Bunce, D. E. Patterson. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies, Quant. Struct. Act. Relatsh., vol.7, 18-25,1998.

[25] V. Centner, D. L. Massart, O. E. de Noord, S. de Jong, vandeginste, C. Stiera. Elimination of uninformatives variables for multivariate calibration, Anal. Chem., vol.68, 3851-3858, 1996.

[26] R. Grohmann, T. J. Schindler. Toward robust QSRR models: Synergistic utilization of robust regression and variable elimination, J. Comput. Chem., vol.29, 847-860, 2008.