

3D-QSAR studies on fluroquinolones derivatives as inhibitors for tuberculosis

Atanu Bhattacharjee^{1*}, Baphilinia Jones Myllemngap¹ & Devadasan Velmurugan²

¹Department of Biotechnology and Bioinformatics, North Eastern Hill University, Permanent campus, Shillong-793022, India; ²Centre for Advanced studies in Crystallography and Biophysics, University of Madras, Guindy (Maraimalai) Campus, Chennai-600 025, India; Atanu Bhattacharjee - Email: atanubioinfo@gmail.com; Phone: +919436703339; *Corresponding author

Received April 13, 2012; Accepted April 16, 2012; Published April 30, 2012

Abstract:

A quantitative structure activity relationship (QSAR) study was performed on the fluroquinolones known to have anti-tuberculosis activity. The 3D-QSAR models were generated using stepwise variable selection of the four methods - multiple regression (MR), partial least square regression (PLSR), principal component regression (PCR) and artificial neural networks (kNN-MFA). The statistical result showed a significant correlation coefficient q^2 (90%) for MR model and an external test set of (pred_r²) -1.7535, though the external predictivity showed to improve using kNN-MFA method with pred_r² of -0.4644. Contour maps showed that steric effects dominantly determine the binding affinities. The QSAR models may lead to a better understanding of the structural requirements of anti-tuberculosis compounds and also help in the design of novel molecules.

Keywords: 3D-QSAR, *Mycobacterium tuberculosis*, fluroquinolones, k-nearest neighbor molecular field analysis, Multiple regression, Partial least square regression, Principle component regression.

Background:

Tuberculosis is a common and often a deadly infectious disease caused by *Mycobacterium tuberculosis* H37Rv claiming more lives each year than any other bacterial disease. Despite the availability of effective direct observed treatment or short-course chemotherapy (DOTS) and Bacille Calmette-Guerin (BCG) vaccine, the tubercle bacillus continues to claim more lives than any single effective agent [1]. Due to the extensive period of treatment, patients fail to complete the therapy leading to the emergence of drug resistance-multidrug resistance (MDR) and extensive drug resistance (XDR) in tuberculosis. The widespread emergences of drug resistant strains together with human immunodeficiency virus (HIV) have been witnessed in the increased incidence of tuberculosis in both developing and industrialized countries [1]. It has been estimated that tuberculosis accounts for around 32 per cent deaths in HIV infected individuals [2]. The recent rise in tuberculosis cases and especially the increase in number of drug resistant mycobacteria strains indicate an urgent need to

develop new anti-tuberculosis drugs. The quinolones are a group of a wide range of synthetic antibiotics derived from nalidixic acid. They are characterized by an atom of fluorine and an aryl substituent at position 6 and 7 respectively and are known to exhibit an antibacterial spectrum including *Mycobacterium tuberculosis* [3]. When using fluroquinolones for the treatment of tuberculosis, individual susceptibility, pharmacokinetic and toxicity profiles should be carefully considered [4].

The present study is aimed to understand the structure activity of the available fluroquinolones and to obtain the predictive quantitative structure activity relationship (QSAR) models. A stochastic search method such as multiple regressions (MR), partial least square regression (PLSR), principal component regression (PCR) and an artificial neural network (kNN-MFA) [5-9] was developed to provide an insight to the various interactive fields of different compounds in concord with the in vitro experimental data. Different QSAR approaches have been

developed through the years and the rapid increase in three-dimensional (3D) structure of the biological molecules have led to the development of 3D structural descriptors and associated 3D QSAR methods [10]. The generated models could provide a valuable reference in the design of pharmaceuticals with improved anti-tubercular activity.

Methodology:

3D QSAR study of the molecules was performed using VLife Molecular Design Suite [11] allowing users to choose the probe, grid size, grid interval and other parameters for generation of descriptors.

Dataset for 3D QSAR

Depending on the antibacterial activity and the toxicity to humans, anti-tuberculosis drugs are classified into first-line drugs as the first choice of treatment in new cases of TB and second-line drugs for treatment of patients with *M. tuberculosis* resistant to the first-line drugs and these were considered for QSAR. The chemical structures of the molecules were retrieved from Pubchem compound database at National Center for Biotechnology Information (NCBI) [12]. Activity of the drugs was taken from reported work given as MIC values [3, 13]. The biological activity (IC_{50}) of the molecules were converted to their corresponding pIC_{50} values **Table 1 (see supplementary material)** and used as dependent variables in the QSAR calculations. The molecules were first optimized to their lowest energy state using Merck molecular force field (MMFF) method [14] with RMS gradient of 0.001 and alignment was done based on a template structure. Alignment of the molecules is one of the most important factors in obtaining a reliable model.

Calculation of descriptors and data selection

The aligned molecules were selected for calculation of the descriptors after inserting the biological activity as a dependent variable. The field descriptors were calculated with cutoffs of 10kcal/mol for electrostatic and 30kcal/mol for steric at lattice points of the grid using a methyl probe of charge +1. A grid was generated with grid interval of 2Å in a 3-dimensional lattice around the aligned molecules. The dielectric constant was set to 1.0 and with suitable charge type [11]. Invariable columns not contributing to the QSAR were removed. Biological activity was selected as a dependent variable and the descriptors generated were selected as independent variables. The training and test sets were generated using the sphere exclusion method which allows construction of the training set covering all the descriptor space areas occupied by representative points [15] and the dissimilarity values was set to 15.75. The unicolon statistics of the training and test sets was calculated for the correct selection of the training and test sets.

Variable selection and model building

The aligned molecules were subjected to regression analysis using multiple regressions (MR) which estimates the value of regression coefficient by applying least square curve fitting model, partial least square regression (PLSR), an extension of MR model without imposing any restrictions. Principal component regression (PCR) selects a new set of axis such that for the first set of axis it reflects most of the variations within the data and k-nearest-neighbor (kNN) molecular field analysis (MFA) as building methods. The models were developed using stepwise (forward-backward) variable selection method with

cross-correlation limit set to 0.5 with the term selection criteria as r^2 for MR, PLSR and PCR and q^2 for kNN-MFA. F-test 'in' was set 4.0 and f-test 'out' was set to 3.99 and scaling to autoscaling. KNN-MFA parameter setting was within the range of 2-5 and the distance-based weighted average prediction method was selected for the study.

Discussion:

The fluoroquinolones which show activity against tuberculosis taken for study were optimized to their lowest energy state and aligned with a template. Of the 33 fluoroquinolones considered, 14 quinolones did not align with the template and therefore were not considered for further study. The fluoroquinolones aligned to the template (**Figure 1**) were considered for QSAR studies. The training and test set selected by sphere exclusion method show that the maximum of the test is less than maximum of training and minimum of test is greater than minimum of training set **Table 2 (see supplementary material)** indicating that the test is interpolative and derived within the min-max range of the training set. The mean of the training is higher than the mean of test set showing the presence of relatively less active molecules compared to the active ones. Also, a relatively higher standard deviation in the training set indicates that the training set has a widely distributed activity as compared to the test set. The 3D QSAR study of the fluoroquinolones through MR, PLSR, PCR and kNN-MFA analysis using VLifeMDS resulted in the following statistical model. The statistical model (Model 1) shows a significant activity-descriptors relationship accuracy of 95% ($r^2= 0.9507$) representing part of variation in the observed data, an activity prediction accuracy of 90% ($q^2= 0.9097$) and a decrease in the predictivity for the external test set of about 0.017% ($pred_r^2=-1.7535$).

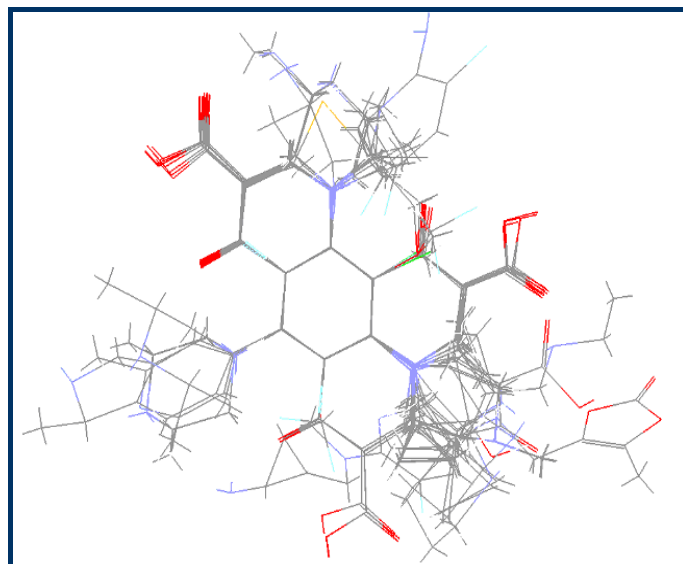


Figure 1: 3D-alignment of the fluoroquinolones based on a template structure

Model 1 (MR)

$\text{Log } p (IC_{50}) = -17.2439 * S_{1379} - 0.0210 * S_{588} + 14.7252 * S_{498} - 0.6401$

$n = 13$, Degree of freedom = 9, $r^2 = 0.9507$, $q^2 = 0.9097$, F test = 57.8086, $r^2 \text{ se} = 0.1606$, $q^2 \text{ se} = 0.2172$, $pred_r^2 = -1.7535$, $pred_r^2 \text{ se} = 0.6996$

The same data was subjected to PLSR method which results in a correlation of 95% ($r^2=0.9498$) and a low prediction accuracy of 89% ($q^2=0.8976$) with a decrease in external predictivity ($\text{pred}_r^2=-1.7493$). S_{1379} and S_{558} contribute negatively and S_{498} contributes positively in both MR and PLSR, thus groups should be altered around the points to reduce the interactions.

Model 2 (PLSR)

Log p (IC50) = $-16.7654 \cdot S_{1379} - 0.02195 \cdot S_{558} + 15.61775 \cdot S_{498}$;
Optimum Components = 2, n = 13, Degree of freedom = 10, $r^2 = 0.9498$, $q^2 = 0.8976$, F test = 94.6739, r^2 se = 0.1536, q^2 se = 0.2194, $\text{pred}_r^2 = -1.7493$, pred_r^2 se = 0.6991

To improve the external predictivity of the model, PCR analysis was performed with the same set of data resulting in a regression coefficient of 97% ($r^2=0.9700$) with an internal predictive accuracy of 86% ($q^2=0.8636$) and a decrease of the external predictivity ($\text{pred}_r^2 = -1.8522$). S_{1379} contributes negatively and S_{560} and S_{618} contribute positively to the activity.

Model 3 (PCR)

Logp (IC50) = $-13.1600 \cdot S_{1379} + 12.7354 \cdot S_{560} + 6.8894 \cdot S_{618} + 0.1260$;
Optimum Components = 2, n = 13, Degree of freedom = 10, $r^2 = 0.9700$, $q^2 = 0.8636$, F test = 161.7857, r^2 se = 0.1188, q^2 se = 0.2533, $\text{pred}_r^2 = -1.8522$, pred_r^2 se = 0.7120

To further improve the external predictivity of the model, kNN-MFA analysis of the fluroquinolone groups was performed and found to be statistically significant in terms of the external predictivity of the test with an internal predictivity of 33% ($q^2=0.3290$) measuring the reliability of the prediction to be reliable and accurate. The pred_r^2 obtained for the test set showed a decreased but an improved result of the external predictivity of $\text{pred}_r^2=-0.4644$. Descriptor S_{242} contributing to the activity negatively indicates that a negative steric potential would favour for the increase activity.

Model 4 (kNN-MFA)

S_{242} (-0.0019 to -0.0016), k Nearest Neighbour= 2, n = 13, Degree of freedom = 11, $q^2 = 0.3290$, q^2 _se = 0.5129, $\text{Pred}_r^2 = -0.4644$, pred_r^2 se = 0.5102.

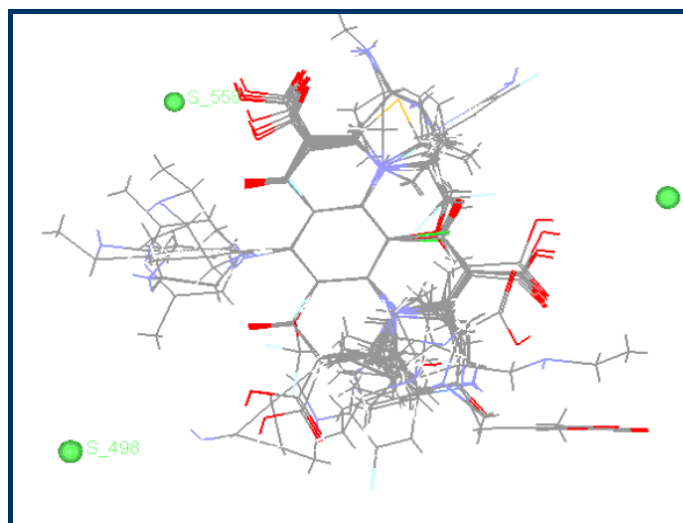


Figure 2: Contribution plot of interaction to the MR model

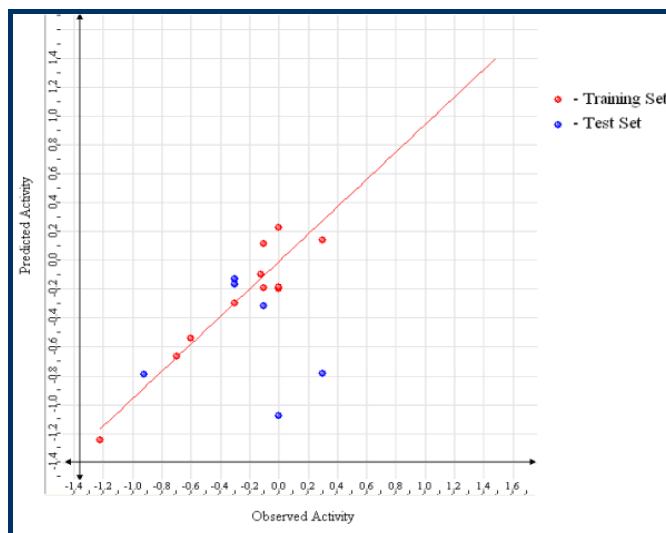


Figure 3: Fitness plot of observed vs predicted activities of the MR model

The PCR contour plots (Figure 2) showed the relative position and range of the corresponding important steric field in the model. The experimental and predicted activity of the model is shown as the fitness plot (Figure 3) for PCR. The nearness of the predicted to the observed activity is reported in Table 3 (see supplementary material). The models obtained for QSAR showed that the steric interactions play an important role in determining the biological activity of the models (Figure 4). The negative range in the steric descriptors indicates that the negative steric potential is favourable for the increase in the activity and a less bulky substituent group is favored in the region.

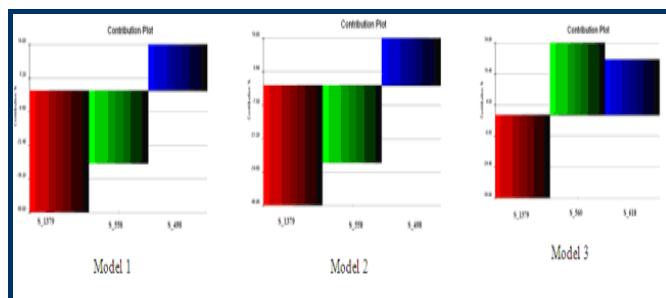


Figure 4: Contribution chart of statistically significant models

Conclusion:

The present study was aimed at deriving the predictive 3D QSAR models capable of revealing the structural requirement for anti-tuberculosis inhibitors. Comparison of the different statistical parameters of the four models suggests model 1 for further consideration having a better internal predictivity of $q^2=0.9097$ and an external test of $\text{pred}_r^2=-1.7535$. Model 4 (kNN-MFA) though has a bad internal predictivity of $q^2 = 0.3290$, the external predictivity ($\text{pred}_r^2=-0.4644$) of the test is better as compared to the other models. The models show that steric effects dominantly determine the binding affinities indicating both a positive and a negative contribution suggesting a more bulky group is favored in the region where there is positive contribution and a less bulky group is favourable in the region with negative contribution. Modification of these regions will lead to better and improved compounds for the treatment of

tuberculosis. The MR and kNN-MFA models developed show potential predictive ability as determined by testing the external test set. The QSAR models may lead to a better understanding of the structural requirements of anti-tuberculosis compounds and also help in designing of novel molecules.

Acknowledgement:

The authors are thankful to the Department of Biotechnology, Govt. of India for providing the necessary funds for the study

References:

- [1] Cole ST *et al.* *Nature* 1998 **393**: 537 [PMID: 9634230]
- [2] Kochi A, *Geneva: World Health Organization* 1996
- [3] Bryskier A & Lowther J, *Expert Opin Investig Drugs*. 2002 **11**: 233 [PMID: 11829714]
- [4] Berning SE, *Drugs*. 2001 **61**: 9 [PMID: 11217874]
- [5] Berk RA *Berk, R.A., Ed.; SAGE Publications Ltd: London* 2003 125.
- [6] Wold *et al.* *Kubinyi, H. Ed.; ESCOM Science Publishers: Leiden* 1993 523
- [7] Dunteman GH, *Dunteman, G.H., Ed.; SAGE Publications Ltd: London* 1989 65.
- [8] Ajmani S *et al.* *J Chem Inf Model*. 2006 **46**: 24 [PMID: 16426036]
- [9] Zheng W & Tropsha A, *J Chem Inf Comput Sci*. 2000 **40**: 185 [PMID: 10661566]
- [10] Noolvi MN & Patel HM, *J Basic Clin Pharm*. 2010 **1**: 153.
- [11] www.vlifesciences.com
- [12] <http://www.ncbi.nlm.nih.gov/pccompound>
- [13] Rastogi N *et al.* *Curr Microbiol*. 1996 **33**: 167 [PMID: 8672093]
- [14] Halgren TA, *J Comp Chem*. 1996 **17**: 490
- [15] Golbraikh A & Tropsha A, *J Comput Aided Molm Des*. 2002 **16**: 357 [PMID: 12489684]

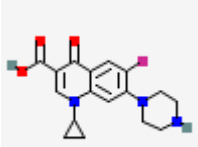
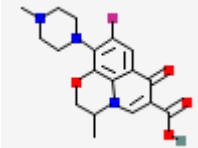
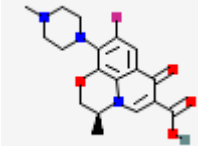
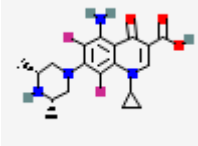
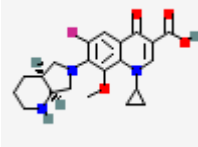
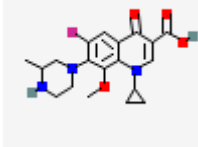
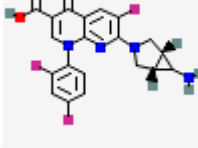
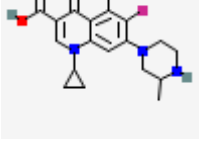
Edited by P Kanguane

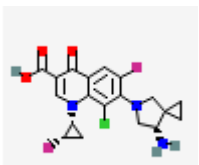
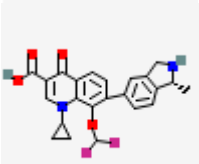
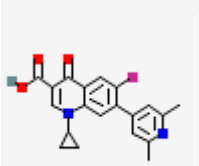
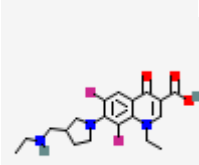
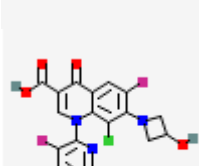
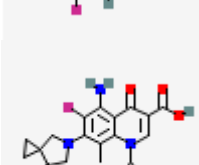
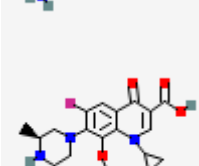
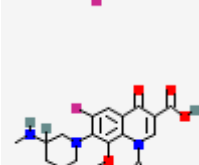
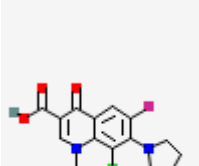
Citation: Bhattacharjee *et al.* *Bioinformation* 8(8): 381-387 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

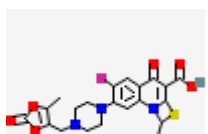
Supplementary material:

Table 1: Structure, biological activity data and calculated log values of the compounds considered for QSAR

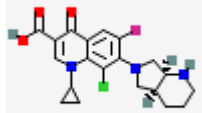
Sl. No.	Compounds	Structure	IC ₅₀ (ug/ml)	Log IC ₅₀	References
1.	CID_2764_Ciprofloxacin [†]		0.5	-0.301	12
2.	CID_4583_Ofloxacin		0.75	-0.124	12
3.	CID_149096_Levofloxacin [†]		0.5	-0.301	13
4.	CID_60464_Sparfloxacin		0.2	-0.698	13
5.	CID_152946_Moxifloxacin		0.5	-0.301	13
6.	CID_5379_Gatifloxacin [†]		0.12	-0.920	13
7.	CID_62959_Trovafloxacin		32	1.505	13
8.	CID_72474_Grepafloxacin.		1.0	0	13

9.	CID_461399_Satifloxacin [†]		0.1	0	13
10.	CID_124093_Garenoxacin		1.0	0	13
11.	CID_121960_WIN57273		1.0	0	13
12.	CID_121833_CI-934		2.0	0.301	13
13.	CID_487101_WQ-3034		0.78	-0.107	13
14.	CID_477670_Olamufloxacin [†]		0.78	-0.107	13
15.	CID_189912_Cadrofloxacin		0.25	-0.602	13
16.	CID_65958_Balofloxacin [†]		2.0	0.301	13
17.	CID_60063_Clinafloxacin		0.5	-0.301	13

18. CID_65947_Prulifloxacin 0.78 -0.107 13



19. CID_00119375_BAYy3118 0.06 -1.221 13



T= Test set

Table 2: Uni-column Statistics for training and test sets

Set	Average	Max	Min	StdDev	Sum
Training	-0.1273	1.5050	-1.2210	0.6262	-1.6550
Test	-0.2213	0.3010	-0.9200	0.4088	-1.3280

Table 3: Experimental and predicted activity of the compounds

Sl. No.	Compounds	Exp Log IC ₅₀	Pred_MR	Pred_PLSR	Pred_PCR	Pred_kNN-MFA
1.	CID_2764_Ciprofloxacin ^T	-0.301	-0.130536	-0.118652	-0.118652	0.091470
2.	CID_4583_Ofloxacin	-0.124	-0.096736	-0.068184	-0.068184	-0.080296
3.	CID_149096_Levofloxacin ^T	-0.301	-0.164698	-0.141946	-0.141946	-0.225776
4.	CID_60464_Sparfloxacin	-0.698	-0.662853	-0.693905	-0.693905	-0.741774
5.	CID_152946_Moxifloxacin	-0.301	-0.166971	-0.155746	-0.155746	-0.396863
6.	CID_5379_Gatifloxacin ^T	-0.920	-0.788032	-0.790338	-0.790338	-0.697438
7.	CID_62959_Trovafloxacin	1.505	1.441020	1.420574	1.420574	1.475813
8.	CID_72474_Grepafloxacin	0	-0.194400	-0.181942	-0.181942	-0.003357
9.	CID_461399_Satifloxacin ^T	0	-1.076203	-1.071753	-1.071753	-0.804463
10.	CID_124093_Garenoxacin	0	-0.182312	-0.179561	-0.179561	-0.096659
11.	CID_121960_WIN57273	0	0.226791	0.208722	0.208722	0.158183
12.	CID_121833_CI-934	0.301	0.140702	0.148581	0.148581	0.072382
13.	CID_487101_WQ-3034	-0.107	0.113440	0.136374	0.136374	0.079528
14.	CID_477670_Olamufloxacin ^T	-0.107	-0.312990	-0.287095	-0.287095	-0.362230
15.	CID_189912_Cadrofloxacin	-0.602	-0.540839	-0.550700	-0.550700	-0.584534
16.	CID_65958_Balofloxacin ^T	0.301	-0.785967	-0.788517	-0.788517	-0.969202
17.	CID_60063_Clinafloxacin	-0.301	-0.294585	-0.280946	-0.280946	-0.231961
18.	CID_65947_Prulifloxacin	-0.107	-0.189217	-0.201832	-0.201832	-0.110201
19.	CID_00119375_BAYy3118	-1.221	-1.249021	-1.256416	-1.256416	-1.195300

T= Test set; Pred_MR, Pred_PLSR, Pred_PCR, Pred_kNN-MFA: Predicted activity of each model