International Journal of Engineering, Science and Mathematics
Vol.6Issue 4, August 2017,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

2D QSAR model to develop novel DNA Gyrase B inhibitors for the treatment of Mycobacterium tuberculosis

[1]Suchitra Surendran, *[1]Dr V.L. Pushpa, [1]Dr K.B. Manoj

[1]P G and Research Department of chemistry, Sree Narayana College, Kollam, Kerala, India 691001

Abstract

A statistically significant quantitative structure-activity relationship (QSAR) model was built from a diverse dataset of 286 DNA GyrB inhibitors which was considered as an efficient target for Mycobacterium tuberculosis. This model was built to identify the important structural features responsible for DNA Gyrase B ATPase inhibitory activity using a genetic algorithm–multi-linear regression (GA-MLR) methodology. The model obtained was robust, highly predictive and validated thoroughly according to the OECD principles. The statistical quality of the model was verified in terms of coefficient of determination R^2 (0.8713) and cross-validated leave-one-out coefficient Q^2 loo (0.8622). The model can be used for the QSAR based virtual screening of new molecules and hence it is used as a versatile tool to evaluate the drug likeness of hit molecules.

Keywords: DNA Gyrase B, QSAR, GA-MLR

1.Introduction

Even though tuberculosis (TB) is a treatable infectious disease it still remains a threat to public health. Nearly 1.2 million people die from TB every year. One of the major reasons for this is due to the resistance of the bacteria to the existing drugs or the emergence of MDR-TB. This has triggered scientists to search for new drugs [1]. DNA Gyrase is one of the promising enzymes found in all bacteria, but not found in human cells hence the enzyme is an attractive target for effective drug discovery. It comprises of two subunits gyrase A (GyrA) and gyrase B (GyrB). [2]. Fluoroquinolones based drugs have found to inhibit the catalytic activity of GyrA, but the resistance developed due to the mutation in, GyrA domain support DNA GyrB an interesting drug target[3]. Many natural and synthetic lead molecules were designed and biologically tested as Gyrase B inhibitors

and but did not reach the final phase of clinical trials except novobiocin. Though novobiocin was brought to market, it was withdrawn due to its toxicity and permeability issues. Recently lots of efforts have been put in this area and different classes of compounds were identified by targeting the enzyme. However, search of compounds with better activity profile still persists. QSAR is an efficient tool that can be used to determine the important structural features that influence the inhibitory activity and to predict the activity of a new molecular endity prior to its actual synthesis and biological screening. This significantly reduces time, efforts and chemicals involved in the process of drug designing.

The goal of the present work is to develop a statistically significant QSAR model that determine the main structural features responsible for DNA GyrB inhibition and to predict the activity of newly designed compounds. Though many QSAR models were developed earlier, it was limited for some of congeneric series of DNA Gyrase B inhibitors and hence had a narrow chemical space. It was only used to predict the activity of similar molecules. In our work we have taken a large diverse dataset of 286 Gyrase B inhibitors for training our model, hence all the significant structural features that influence Gyrase B activity were captured.

2. Materials and methods

*2.1*Experimental dataset

2D QSAR model was also build on a diverse set of 286 GyrB inhibitors[4–11] belonging to the same bioassay. A diverse set was used to capture all the significant structural features that influence DNA GyrB activity.

2.2Computational Studies

The computational works used for constructing 2D QSAR model were carried out using the softwares PaDEL[12] and QSARINS[13].

2.3. *Molecular descriptor calculation and selection*

Molecular descriptors used for 2D QSAR modelling was calculated using PaDEL 2.21software which is a standalone software , that calculates molecular descriptors and fingerprints using the Chemistry Development Kit (Yap, 2011). The descriptors were generated by retaining the optimized 3D configuration of chemical compounds. Since, all the calculated descriptors (>18,000) do not contain significant information; objective feature selection was employed to reduce the descriptor pool. Nearly constant (>80%), and highly correlated (> 95%) descriptors were eliminated before subjective feature selection

(SFS) using QSARINS-Chem 2.2.1. The reduced set still consists a wide range of theoretical molecular descriptors that takes into account different structural features, viz. constitutional (0D-), mono-dimensional (1D-), bi-dimensional (2D-) and three-dimensional (3D-), capturing and magnifying the diverse aspects of the chemical structures.

2.4. 2D QSAR model building and validation

2D QSAR model for our research was built using the software QSARINS (QSAR-INSubria) developed at the University of Insubria.[13,14] The software uses Multiple Linear Regression (MLR) method to obtain a linear relationship between the inhibitory activity and the physicochemical properties, using ordinary least squares (OLS) algorithm. All subset technique was used to explore all combinations of descriptors in low dimension and Genetic Algorithm (GA) was used for the variable selection.

For the validation of QSAR models, the data set was divided into training and test sets. Internal validation was done using the training set molecules and external validation was based on the test set compounds. Various statistical parameters like R^2, Q^2, RMSE etc were considered for selecting the best model. Higher the value of R^2 corresponds to good fitting of the model. Cross Validation (CV) parameters such as Q^2 Leave One Out (Q^2 loo), Q^2 Leave Many Out (Q^2LMO) and Root Mean Square Error (RMSE) were considered for the confirmation of internal validation [15]. The values of Q^2 loo and Q^2LMO >0.7 indicates the model to have high robustness and internal predictive ability. An acceptable QSAR model also must have a lower RMSE value. Y-randomization is another validation approach employed to check if the model is obtained by chance of correlation. This is associated with parameters like R^2 Yscr and Q^2 Yscr. Lower values of these parameters indicate that models are not obtained by chance of correlation. The external predictive ability of QSAR model is evaluated by parameters like R^2ext, Q^2F1, Q^2F2, Q^2F3,CCCext, RMSEext , and r^2m. A high predictive QSAR model has high values for R^2ext (>0.6), Q^2F1(>0.7), Q^2F2(>0.7), Q^2F3(>0.7),CCCext (0.85) , r^2m (>0.6) and low values of RMSEext[15]. AD is the chemical space defined by molecular descriptors and modelled responses. AD is a plot between leverage values and standard residuals. Leverage values are the diagonal elements of HAT matrix. The QSAR models are generally applied to predict the activity of new classes of compounds falling within the domain of applicability.Compounds having structural dissimilarity, compared to other compounds in the dataset or those having an unexpected biological activity are considered as outliers

[16]. The true external predictive capacity of QSAR model is assessed by applying the QSAR model equation to an external set of 22 Gyr B inhibitor molecules taken from the study of V Jeankumar, S Kotagiri and R Janupallyetal[17], which were not included for the model development.

3.Results and discussion

In this study we have developed a MLR based 2D QSAR model for the identification of important structural features responsible for the inhibition of DNA Gyrase B activity in mycobacterium infections. The subjective feature selection was performed using Genetic Algorithm (GA) method

3.1. Statistical significance of 2D QSAR model

Several models having good correlation with response variable (PIC50) and low intercorrelation between the descriptors was developed upto six significant descriptors. Figure 1 shows the plot of the average value of R^2 and Q^2 vs the number of variables of the model. It is seen that the values for R^2 and Q^2 increased with the increase in the number of significant descriptors.

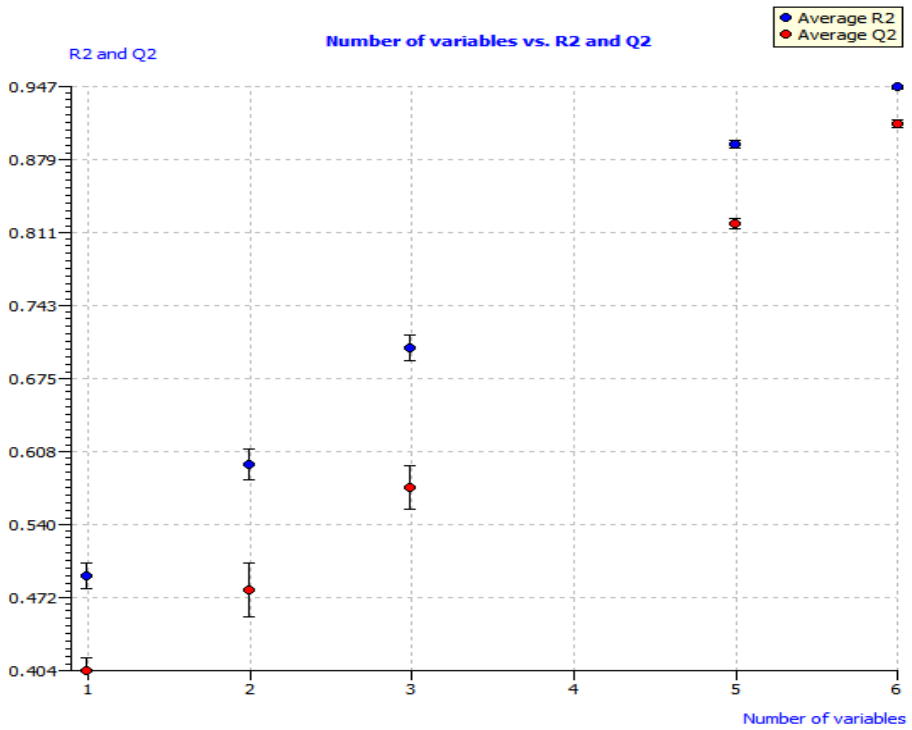

Figure 1. Plot of average of models R^2 and Q^2 vs Number of variables

For selecting the best model, various statistical parameters in terms of fitting, internal validation and external validation criteria were also considered. The details of the parameters are given in table 1

Table 1 Statistical parameters of the developed models

R^2	R^2adj	R^2-R^2adj	LOF	Kxx	Delta K	RMSE tr	MAE tr	RSS tr	CCC tr	s	F
0.8713	0.8677	0.0035	0.2652	0.3911	0.083	0.4876	0.3722	53.7297	0.9312	0.4953	246.9983

Q^2loo	R^2-Q^2loo	RMSE cv	MAE cv	PRESS cv	CCC cv	Q^2 LMO	R^2 Yscr	RMSE AV Yscr	Q^2 Yscr	RMSE ext	MAE ext
0.8622	0.009	0.5044	0.385	57.501	0.9264	0.8596	0.027	1.3404	-0.0373	0.4636	0.339

PRESS ext	R^2ext	Q^2-F1	Q^2-F2	Q^2-F3	CCC ext	r^2m aver.	r^2m delta	k'	k	Clos'	Clos
12.6829	0.8702	0.8695	0.8693	0.8836	0.9321	0.8126	0.0663	0.9951	0.9981	0.0125	0.0009

The OECD regulatory principles were followed to ensure the validity and performance of the model. Higher value of R^2 (>0.6) corresponds to good fitting of the model. The ability to predict the activity of new compounds were checked by external validation. The high value of R^2ext, Q^2-F1, Q^2-F2, Q^2-F3 etc., low value of RMSEext, MAEex and PRESSext suggest the model has a very high predictive ability. Q^2 loo and Q^2 LMO were considered for the internal validation. The very high value of Q^2 loo and Q^2 LMO respectively and their closeness with R^2 suggests the stability of the model. The value of R^2 adj is close to R^2 suggests that the number of descriptors in the model is not too high, thereby, indicating the model is free from over-fitting. This is further supported by the low value of LOF (Lack of fit) for the models. The low Kxx value indicates low inter-correlation among the descriptors used in a model. The condition RMSEtr<RMSEcv is satisfied by the developed model. Y-scrambling method was executed to ensure that the given model is not developed by chance correlation. The low value of R^2Yscr and Q^2Yscr for both the models indicate that the models have not been developed by chance. The applicability of the model was also checked using another external set of molecules. Out of the 22 molecules taken, the model predicted the activity 18 molecules with an error of less than 1 (Table 2).

Table 2: Validation of the model using an external set

External test molecules	IC50 (μM)	pIC50 (M)	Predicted by the regression model	error
4	1.75	5.757	4.695	1.062
5	2.6	5.585	4.452	1.133
6	9.8	5.009	4.712	0.297
7	15.7	4.804	4.527	0.277
8	23.9	4.622	4.425	0.197
9	5.85	5.233	4.668	0.565
10	3.56	5.449	4.583	0.866
11	9.8	5.009	4.574	0.435
12	8.07	5.093	4.354	0.739
13	8.707	5.06	4.579	0.481
14	23.31	4.632	4.438	0.194
15	21.51	4.667	4.324	0.343
16	17.85	4.748	4.535	0.213
17	4.538	5.343	4.491	0.852
18	2.136	5.67	4.703	0.967
19	3.6	5.444	4.495	0.949
20	6.4	5.194	4.718	0.476
21	1.5	5.824	4.626	1.198
22	7.5	5.125	4.601	0.524
23	1.5	5.824	4.379	1.445
24	13.05	4.884	4.584	0.3
25	7.6	5.119	4.517	0.602

Therefore, the developed model satisfies all the recommended threshold values for various statistical parameters suggested by different researchers. Hence, this model is useful to study the structural properties responsible for the Gyrase B inhibitory activity and hence to predict the activities of new compounds.

3.2 .Interpretation of the significant descriptors

A mathematical relationship is established between the biological property and responsible physicochemical properties. The descriptors involved in the model generation along with their regression coefficients and the standard errors are given in the equation shown below.

Log IC50 =3.1108(±0.7038) -0.0540(±.0423) SaasC
+0.7462(±0.3363) maxwHBa
+0.0508(±0.0292) maxHBint2
-0.0229(±0.0203) MDEC-22
+0.0053(±0.0019) Ap
+2.5708(±0.2522) KRFP788

From the equation it is clear that the electrotopological state descriptors SaasC(Sum of atom-type for all methylenes) and MDEC-22 (Molecular distance edge between all secondary carbons) correlates negatively with the Gyrase B inhibitory activity whereas electrotopological state atom type descriptors maxwHBa(Maximum E-States for weak Hydrogen Bond acceptors) and maxHBint2 (Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 2) correlates positively to the inhibitory activity.The 3D descriptor Ap (A total size index / weighted by relative polarizabilities) and Klekota-Roth fingerprint showing the presence of chemical substructures -[NH][CH_2][CH_3] also varies positively

Applicability domain of the model in QSARINS has been provided by the Insubria graph which is based on leverage and model predictions. From the graph of Insubria figure 2 it is clear that out of the 286 molecules, only six molecules come outside the domain, hence these molecules can be considered as potential outliers for the model. Their structures are shown in figure 3. From the analysis of the structures of these molecules, it is found that, three molecules that lie far from the domain limit, compound- 242636822, compound-242636823 and compound-194167335 have an allyl group attached to the urea which significantly reduces the activities of these molecules. If the allyl group was replaced by an ethyl group the activity of the molecule would have increased.

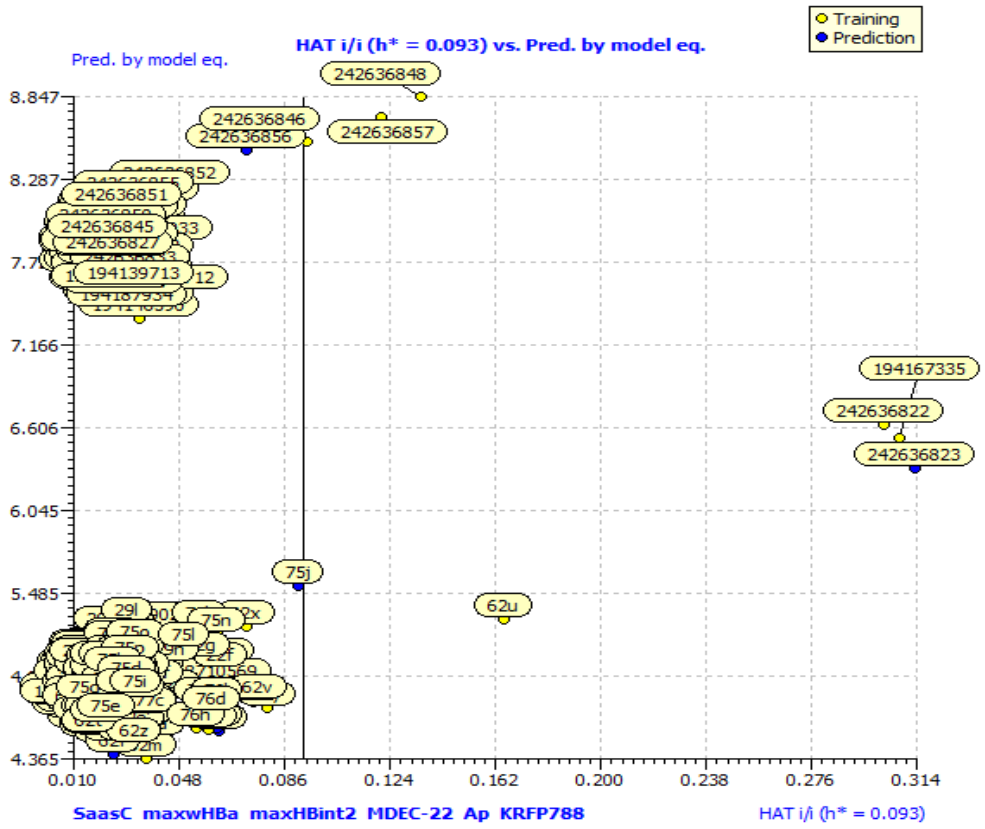

Figure 2 :Insubria graph used to define the applicability domain.

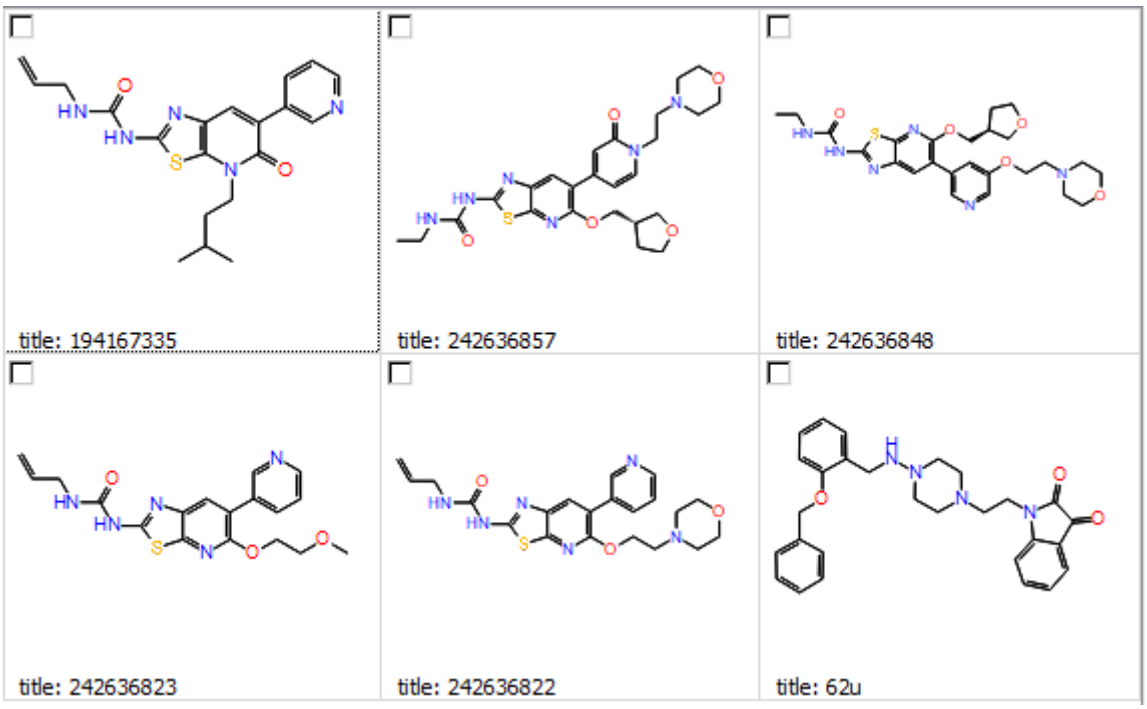

Figure 3: Structures of the potential outliers

4.CONCLUSION

A statistically significant 2D QSAR model was built on a diverse set of DNA Gyrase B inhibitors. A highly predictive model with R^2 = 0.8713 and Q^2loo = 0.8622 was obtained. Since the model was built using a diverse dataset, all important structural features that influence Gyrase B activity was captured. From the mathematical linear equation obtained, we can conclude that the presence of methylene group will decrease the biological activity whereas, the presence of hydrogen bond acceptors at suitable positions would increase the biological activities. The model developed can be used for the virtual screening of large database and also to predict the activities of newly designed molecules.

Bibliography

[1] K.J. Seung, S. Keshavjee, M.L. Rich, Drug-Resistant Tuberculosis, (2015) 1–20.

[2] J.J. Champoux, DNA T OPOISOMERASES : Structure , FunctChampoux, J. J. (2001). DNA T OPOISOMERASES : Structure , Function , 369–413.ion , Annu. Rev. Biochem. 70 (2001) 369–413.

[3] S. Matrat, N. Veziris, C. Mayer, V. Jarlier, C. Truffot-Pernot, J. Camuset, E. Bouvet, E. Cambau, A. Aubry, Functional analysis of DNA gyrase mutant enzymes carrying mutations at position 88 in the A subunit found in clinical strains of Mycobacterium tuberculosis resistant to fluoroquinolones, Antimicrob. Agents Chemother. 50 (2006) 4170–4173. https://doi.org/10.1128/AAC.00944-06.

[4] K. Indrasena, K. Srihari, J. Renuka, K. Shruthi, A. Chuppala, V. Ullas, J. Padma, K. Sudhakar, P. Yogeeswari, D. Sriram, Bioorganic & Medicinal Chemistry An efficient synthesis and biological screening of benzofuran and benzo [d] isothiazole derivatives for Mycobacterium tuberculosis DNA GyrB inhibition, Bioorg. Med. Chem. 22 (2014) 6552–6563. https://doi.org/10.1016/j.bmc.2014.10.016.

[5] S.R. Ghorpade, M. Kale, A. Raichurkar, S.H. P, D. Mckinney, M. Mramaiah, K. Koushik, L.K. Jena, V. Shinde, S. Rudrapatna, V. Humnabadkar, P. Madhavapeddi, H. Basavarajappa, A. Ghosh, R. Murthy, S. Guptha, S. Sharma, P. Vachaspati, K.N. Kumar, J. Giridhar, J. Reddy, V. Panduga, S. Ganguly, S. Gaonkar, N.K.C. N, D.

Ogg, J.A. Tucker, P.A. Boriack-, S.M. De Sousa, V. Sambandamurthy, U. Kranthi, S. Barde, Thiazolopyridine Ureas as Novel Antitubercular Agents Acting Through Inhibition of DNA Gyrase B, (2013). https://doi.org/10.1021/jm401268f.

[6] V.U. Jeankumar, J. Renuka, V.K. Pulla, V. Soni, J.P. Sridevi, P. Suryadevara, M. Shravan, R. Medishetti, P. Kulkarni, P. Yogeeswari, D. Sriram, Development of novel N-linked aminopiperidine-based mycobacterial DNA gyrase B inhibitors: Scaffold hopping from known antibacterial leads, Int. J. Antimicrob. Agents. (2014) 4–13. https://doi.org/10.1016/j.ijantimicag.2013.12.006.

[7] J. Renuka, K.I. Reddy, K. Srihari, V.U. Jeankumar, M. Shravan, J.P. Sridevi, P. Yogeeswari, K.S. Babu, D. Sriram, Design , synthesis , biological evaluation of substituted benzofurans as DNA gyrase B inhibitors of Mycobacterium tuberculosis, Bioorg. Med. Chem. (2014). https://doi.org/10.1016/j.bmc.2014.06.041.

[8] R.R. Kale, M.G. Kale, D. Waterson, A. Raichurkar, S.P. Hameed, M.R. Manjunatha, B.K.K. Reddy, K. Malolanarasimhan, V. Shinde, K. Koushik, L. Kumar, S. Menasinakai, V. Humnabadkar, P. Madhavapeddi, H. Basavarajappa, S. Sharma, R. Nandishaiah, M. Kumar, S. Ganguly, V. Ahuja, S. Gaonkar, C.N.N. Kumar, D. Ogg, P.A. Boriack-sjodin, V.K. Sambandamurthy, S.M. De Sousa, S.R. Ghorpade, Bioorganic & Medicinal Chemistry Letters Thiazolopyridone ureas as DNA gyrase B inhibitors : Optimization of antitubercular activity and efficacy, Bioorg. Med. Chem. Lett. 24 (2014) 870–879. https://doi.org/10.1016/j.bmcl.2013.12.080.

[9] S. Saxena, G. Samala, J. Renuka, J.P. Sridevi, P. Yogeeswari, D. Sriram, Development of 2-amino-5-phenylthiophene-3-carboxamide derivatives as novel inhibitors of Mycobacterium tuberculosis DNA GyrB domain, Elsevier Ltd, 2015. https://doi.org/10.1016/j.bmc.2015.02.032.

[10] B. Medapi, P. Suryadevara, J. Renuka, J.P. Sridevi, P. Yogeeswari, D. Sriram, 4-Aminoquinoline derivatives as novel Mycobacterium tuberculosis GyrB inhibitors: Structural optimization, synthesis and biological evaluation, Eur. J. Med. Chem. 103 (2015) 1–16. https://doi.org/10.1016/j.ejmech.2015.06.032.

[11] B. Medapi, J. Renuka, S. Saxena, J. Padma, R. Medishetti, P. Kulkarni, P.

Yogeeswari, D. Sriram, Bioorganic & Medicinal Chemistry Design and synthesis of novel quinoline – aminopiperidine hybrid analogues as Mycobacterium tuberculosis DNA gyraseB inhibitors, Bioorg. Med. Chem. 23 (2015) 2062–2078. https://doi.org/10.1016/j.bmc.2015.03.004.

[12] A. Allouche, Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares, J. Comput. Chem. 32 (2012) 174–182. https://doi.org/10.1002/jcc.

[13] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: A new software for the development, analysis, and validation of QSAR MLR models, J. Comput. Chem. 34 (2013) 2121–2132. https://doi.org/10.1002/jcc.23361.

[14] P. Gramatica, S. Cassani, N. Chirico, QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS, J. Comput. Chem. 35 (2014) 1036–1044. https://doi.org/10.1002/jcc.23576.

[15] P. Gramatica, Principles of QSAR models validation: Internal and external, QSAR Comb. Sci. 26 (2007) 694–701. https://doi.org/10.1002/qsar.200610151.

[16] R.P. Verma, C. Hansch, An approach toward the problem of outliers in QSAR, Bioorganic Med. Chem. 13 (2005) 4597–4621. https://doi.org/10.1016/j.bmc.2005.05.002.

[17] V.U. Jeankumar, S. Kotagiri, R. Janupally, P. Suryadevara, J.P. Sridevi, R. Medishetti, P. Kulkarni, P. Yogeeswari, D. Sriram, Exploring the Gyrase ATPase domain for tailoring newer anti-tubercular drugs : Hit to lead optimization of a novel class of thiazole inhibitors, Bioorg. Med. Chem. (2014). https://doi.org/10.1016/j.bmc.2014.12.001.