

2D-Quantitative Structure Activity Study and the Computational Prediction of Antibacterial Activity for Series of Quinolones Derivatives

Aissaoui Mohamed¹, Hatem Layeb² and Abdelhafid Djerourou³

¹ Synthesis and Biocatalysis Organic Laboratory, Badji Mokhtar University, Annaba, Algeria

² Synthesis and Biocatalysis Organic Laboratory, Badji Mokhtar University, Annaba, Algeria

³ Synthesis and Biocatalysis Organic Laboratory, Badji Mokhtar University, Annaba, Algeria

Abstract: *Quantitative structure–activity relationship (QSAR) models are useful in understanding how chemical structure relates to the biological activity of natural and synthetic chemicals and for design of newer and better therapeutics. In the present study, 15 quinolones derivatives were evaluated as antibacterial inhibitors, expressed by the cytotoxicity of these compounds (MIC). Based on these data, different molecular descriptors were used to solve this problem. A linear QSAR model was developed using Multiple Linear Regression technique, while Genetic Algorithm was adopted for selecting the most appropriate descriptors. The predictive activity of the model was evaluated by means of external validation set and the Y-randomization technique, and its structural chemical domain has been verified by the leverage approach.*

Keywords: *quinolones derivatives; Antibacterial activity; Genetic Algorithm; Multiple Linear Regression.*

1. Introduction

Quinolones are considered as a big family of multi-faceted drugs¹, their chemical synthesis is flexible and can be easily adapted to prepare new congeners with rationally devised structures. This is shown by the description of many thousands of derivatives in the literature. Scientists could accurately describe their QSAR, which is essential for effective drug design. This also gave them the chance to discover new and unprecedented activities, which makes quinolones an endless source of hope and enables further development of new clinically useful drugs. Quinolones are among the most common frameworks present in the bioactive molecules that have dominated the market for more than four decades².

The quinolones act against bacteria by selectively inhibiting the type II topoisomerases DNA gyrase and topoisomerase IV, enzymes that play a critical role in bacterial cell growth and division.³ The lack of structural information about the nature of the interactions between bacteria and quinolones derivatives has thus made it a difficult task to discover good lead compounds. Quantitative structure-activity relationship (QSAR) studies are a powerful method for the design of bioactive compounds and the prediction of activity according to the physical and chemical properties⁴⁻⁷.

Recently a QSAR study on a data set of 10 quinolones derivatives as antibacterial agent was reported by means of quantum chemical descriptors using Multiple Linear Regression in combination with the elimination Stepwise as variable selection algorithm⁸⁻⁹.

The aim of this study is to develop a QSAR model of the antibacterial activity of 15 quinolones derivatives against Gram-positive and Gram-negative organisms, to better understand the structural features of these types of compounds, using 0, 1, 2 and 3D molecular descriptors calculated using the DRAGON software. This study may help us to design new analogues with better biological profile.

2. Materials and methods

2.1. Dataset and biological data

The database consists of 15 quinolones derivatives as antibacterial agent. Their structures and in vitro activity are listed in Fig1 and table 1⁸⁻⁹.

Activities were converted into the corresponding $-\log_1/\text{MIC}$ values, where MIC is the minimal inhibitory concentrations (mmol/mL) or the lowest drug concentrations that prevent visible growth of bacteria.

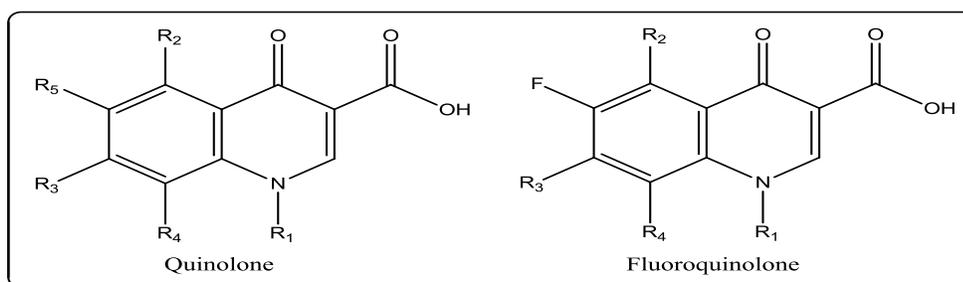


Fig. 1: chemical structure of quinolone and fluoroquinolone derivatives.

Table. 1: Minimum inhibitory concentration of quinolones derivatives

| Compd no. | S.aureus ATCC 29213 MIC ($\mu\text{mol/ml}$) | E. coli ATCC 25922 MIC ($\mu\text{mol/ml}$) |
|------------------|---|--|
| Training set No. | | |
| 1 | 0.00196 | 0.251 |
| 2 | 0.00192 | 0.123 |
| 3 | 0.00196 | 0.251 |
| 4 | 0.00192 | 0.0613 |
| 5 | 0.000481 | 0.000481 |
| 6 | 0.25 | 0.000470 |
| 7 | 0.06 | 0.004 |
| 8 | 0.008 | 0.125 |
| 9 | 0.004 | 0.125 |
| 10 | 0.004 | 0.125 |
| Test set No. | | |
| 11 | 0.03 | 0.25 |
| 12 | 0.004 | 0.06 |
| 13 | 0.015 | 0.125 |
| 14 | 0.004 | 0.125 |
| 15 | 0.004 | 0.125 |

2.2. Calculations

2.2.1. Molecular descriptors and geometry optimization

We carried out preliminary MM geometry optimization calculations for each compound of this study, and then using the quantum chemical semi-empirical method AM1¹⁰ included in Gaussian 09 software determined the (x,y,z)-atomic coordinates of the minimal energy conformations for each one. All calculations were carried out at the restricted Hartree-Fock level with no configuration interaction. The molecular structures were optimized with the Polak-Ribiere algorithm until the root mean square gradient was 0.001 K cal mol⁻¹

The resulted geometry was transferred into the Dragon software¹¹ to calculate 4488 descriptors in 20 different classes¹².

2.2.2. Descriptor selection

Once the descriptors had been generated by the descriptor generation routine, variable selection was used to reduce the number of descriptors per compound. Feature (variable) selection is used to choose a subset of descriptors that are best in encoding the activity of interest, since many of the calculated descriptors carry redundant and highly correlated information or very little useful information. Feature selection includes objective methods and subjective methods¹³⁻¹⁵. Objective feature selection uses the independent variables alone to filter out non-useful descriptors without using the dependent variables. This procedure involves:

- All descriptors with same values for all molecules were omitted.
- The input variables in Multiple Linear Regression (MLR) must not be highly correlated. Therefore, one of the two descriptors that has the pairwise correlation coefficient above 0.9 (RN0.9) and has a large correlation coefficient with the other descriptors in each class was eliminated.

The remaining pool of descriptors was then further reduced by subjective feature selection, which searches for an information rich subset of descriptors. Here, the dependent variables, $-\log 1/\text{MIC}$ values, were considered in descriptor selection. The selected subset of descriptors ranging from 3 to 8 descriptors per model was used to map the set of molecular structure to the activity.

Genetic Algorithms (GAs) are stochastic optimization methods that have been inspired by evolutionary principles¹⁶. The distinctive aspect of a GA is that it investigates many solutions simultaneously, each of which explores different regions of parameter space.

The first step of GAs simulation is the creation of N chromosomes (random population), each of which represents a candidate solution of the problem. In the case of feature selection, an appropriate representation of a chromosome can be a numerical string encoding a particular combination of molecular descriptors.

The fitness of each chromosome is evaluated based on a statistical function like the coefficient of determination R^2 , adjusted R, s the standard deviation of the regression, Fisher function a cross validation leave-one-out Q^2_{loo} , which reflect the quality of each candidate solution. According to the Ordinary Least Squared (OLS) method¹⁷ the above statistical parameters are calculated using the observed and calculated responses.

The next step is reproduction, creates new chromosomes from the existing generation. Through the selection operator, better chromosomes can proliferate preferentially. With crossover, each chromosome has an opportunity to exchange information with the others via a mating procedure.

Finally, fitter offspring may appear in the next generation if beneficial mutations take place. As the system gathers more knowledge about the underlying parameter space, the collective search, which may appear to be random at first, begins to gain focus and moves towards more optimal regions.

The reproductive cycle is repeated until a predefined number of generations, a specified convergence criterion or a target fitness score is attained.

The GA simulation conditions were 10 000 generations, number of crossovers were 5000, smoothness factor was 1, mutation probability for adding new term was 50% and 300 model populations. The GA procedure was repeated n-times to confirm that the selected descriptors are the most optimal descriptor set for describing the biological property. The GA was written in the C language and runs on a personal Pentium 3.0 GHz.

All molecular descriptors are used for building QSAR models by Multiple Linear Regression Analysis, as implemented in the Minitab software (version 15.1.0.0.)¹⁸

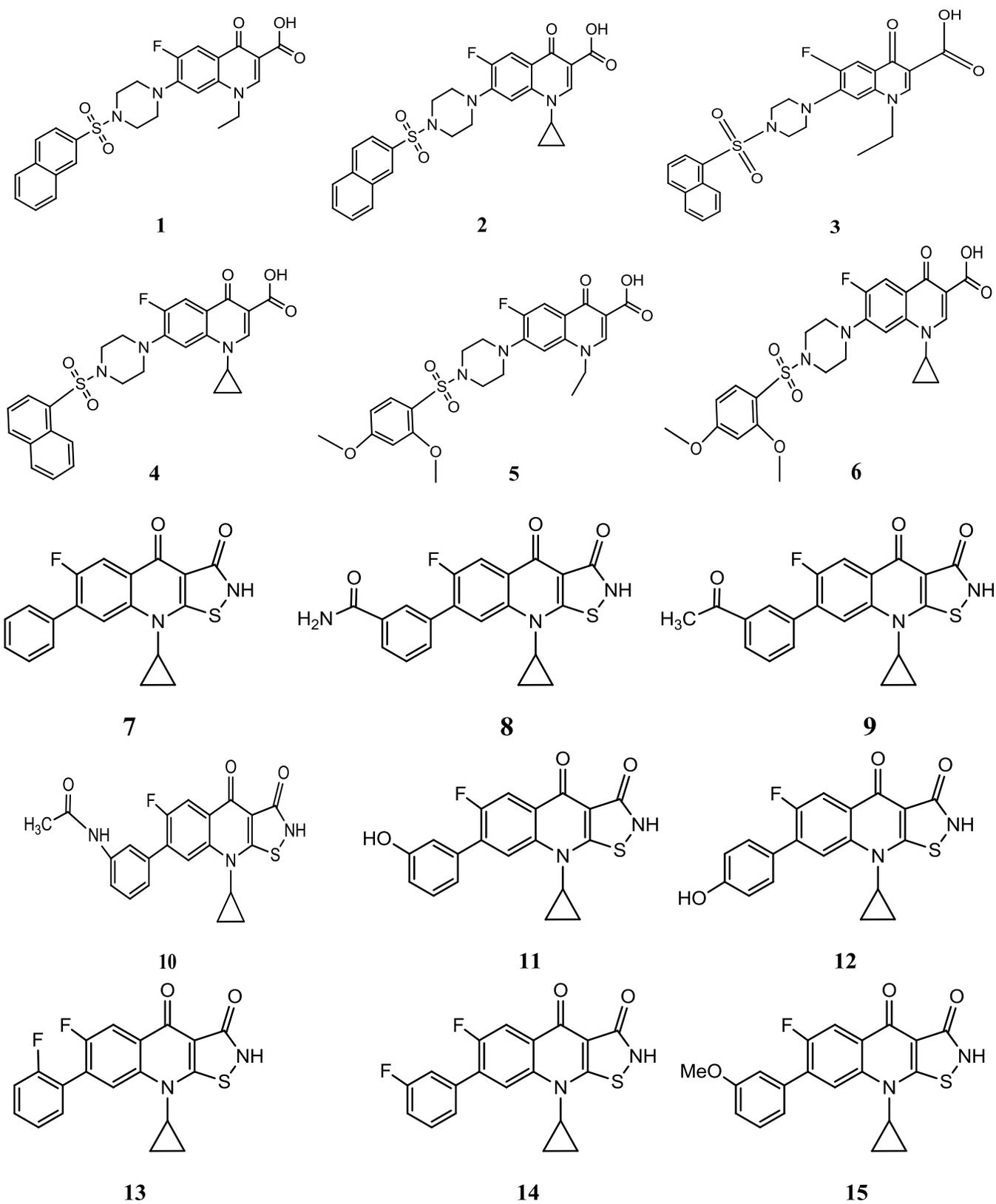


Fig. 2: structure of 15 quinolone derivative

3. Result and Discussion

Table. 2: The experimental and predicted - log1/MIC for Gram-positive and Gram-negative

| Compd no | S. aureus | | E. coli | |
|------------------|-----------|-----------|----------|-----------|
| | Observed | Predicted | Observed | Predicted |
| Training set No. | | | | |
| 1 | 2.71 | 2.76 | 0.60 | 0.68 |
| 2 | 2.72 | 2.83 | 0.91 | 1.05 |
| 3 | 2.71 | 2.76 | 0.60 | 0.87 |
| 4 | 2.72 | 2.83 | 1.21 | 0.87 |
| 5 | 3.32 | 3.17 | 3.32 | 3.38 |
| 6 | 3.33 | 3.24 | 3.33 | 3.52 |
| 7 | 1.22 | 1.32 | 2.34 | 2.6 |
| 8 | 2.10 | 2.19 | 0.90 | 0.88 |
| 9 | 2.34 | 2.31 | 0.90 | 0.81 |
| 10 | 2.34 | 2.40 | 0.90 | 0.77 |
| Test set No. | | | | |
| 11 | 1.52 | 1.53 | 0.6 | 0.65 |
| 12 | 2.34 | 2.37 | 1.22 | 1.35 |
| 13 | 1.82 | 1.9 | 0.9 | 1.2 |
| 14 | 2.34 | 2.5 | 0.9 | 0.86 |
| 15 | 2.34 | 2.1 | 0.9 | 0.82 |

In order to build statistically reliable QSAR model, a large number of compounds and a good algorithm for descriptor selection when exploring a huge descriptor space are needed. A QSAR model is presented for Log 1/MIC of 15 quinolones derivatives as antibacterial agent involving theoretical descriptors, which have been calculated from molecular structure.

The equation that describes the model relative to five variables and using GA-MLR technique is shown with its statistical parameters.

$$-\log 1/\text{MIC} = -0.4727 + 1.17 \text{ MATS7v} + 0.485 \text{ Mor15u} + 0.414 \text{ H5u} + 6.05 \text{ R4e}^+ - 0.480 \text{ nArOR}$$

$$N= 44; \quad R^2=0.857; \quad S=0.197; \quad F=45.53; \quad P < 10^{-4}; \quad Q^2_{\text{cv-100}}=0.815; \quad S_{\text{cv-100}}=0.208$$

Where N is the number of compounds included in the model, s is the standard deviation of the regression, R^2 is the squared correlation coefficient, F is the Fischer ratio, $S_{\text{cv-100}}$ and $Q^2_{\text{cv-100}}$ are the crossvalidation standard deviation and square of the correlation coefficient respectively.

The variables included in the model were designed as follow: the MATS7v represents the Moran autocorrelation -lag 7/weighted by atomic van der Waals volumes and the Mor15u the type of 3D-MoRSE-signal 15/unweighted used that is, H5u for H autocorrelation of lag 5/ unweighted, R4e+ for R maximal autocorrelation of lag 4/weighted by atomic Sanderson electronegativities, and nArOR for number of aromatic ethers (Ar-O-R).

4. Conclusion

The antibacterial activity for 15 quinolone derivatives was modeled with success by Multiple Linear Regression analysis, using Genetic Algorithms as variable selection method.

The proposed model has good stability and predictively, The chemical applicability domain of the studied model served as a valuable tool to filter out dissimilar and outlier compounds. The most important selected descriptors have a clear mechanistic meaning: they related to both Van Der Waals volumes and electronegativity of atoms and groups.

5. Acknowledgements

This work was supported financially by The General Directorate for Scientific Research and Technological Development (DG-RSDT), Algerian Ministry of Scientific Research, Applied Organic Laboratory (FNR 2000).

6. References

- [1] A. Ahmed, M. Daneshtalab. Nonclassical Biological Activities of Quinolone Derivatives. *J Pharm Pharmaceut Sci.* 2011; 1:52 – 72
- [2] Bambeke, F. V.; Michot, J. M.; Eldere, J. V.; Tulkens, P. M. *Clin. Microbiol. Infect.* 2005, 11, 256
A. Cichocki and R. Unbehaven, *Neural Networks for Optimization and Signal Processing*, 1st ed. Chichester, U.K.: Wiley, 1993, ch. 2, pp. 45-47. (**Book style**)
- [3] Andriole, V. T. *Clin. Infect. Dis.* 2005, 41, S113
<http://dx.doi.org/10.1086/428051>
- [4] M. Karelson (Ed.), *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, 2000.
- [5] M. Fernandez, J. Caballero, A.M. Helguera, E.A. Castro, M.P. Gonzalez, *Bioorg. Med. Chem.* 2005; 13:3269–3277.
<http://dx.doi.org/10.1016/j.bmc.2005.02.038>
- [6] S. Liane, M.P. Gonzalez, Y. Fall, G. Gomez, *Eur. J. Med. Chem.* 2007;42:64–70.
<http://dx.doi.org/10.1016/j.ejmech.2006.08.005>
- [7] A. Tropsha, W. Zheng, *Curr. Pharm. Des.* 2001; 7:125–133.
<http://dx.doi.org/10.2174/1381612013397834>
- [8] Alaa A.-M. Abdel-Aziz.; Yousif A. Asiri.; Mohamed H.M. Al-Agamy. *European Journal of Medicinal Chemistry* 2011;46:5487-5497
- [9] 9 Jason A. Wiles.; Qiuping Wang.; Edlaine Lucien.; Akihiro Hashimoto.; Yongsheng Song.; Jijun Cheng.; Christopher W. Marlor.; Yangsi Ou.; Steven D. Podos.; Jane A. Thanassi.; Christy L. Thoma.; Milind Deshpande.; Michael J. Pucci.; Barton J. Bradbury. *Bioorganic & Medicinal Chemistry Letters.* 2006; 16:1272–1276
- [10] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, *J. Am. Chem. Soc.* 1985; 107:3902–3909.
<http://dx.doi.org/10.1021/ja00307a051>
- [11] R. Todeschini, Milano Chemometrics and QSPR Group, <http://www.disat.unimib.it/vhtml>.
- [12] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors (Methods and Principles in Medicinal Chemistry)*, Wiley-VCH, Weinheim, 2000.
<http://dx.doi.org/10.1002/9783527613106>
- [13] B.E. Mattioni, P.C. Jurs, *J. Mol. Graph. Model.* 2003 ;21:391–419.
[http://dx.doi.org/10.1016/S1093-3263\(02\)00187-0](http://dx.doi.org/10.1016/S1093-3263(02)00187-0)
- [14] R. Guha, J.R. Serra, P.C. Jurs, *J. Mol. Graph. Model.* 2004;23:1–14.
<http://dx.doi.org/10.1016/j.jmglm.2004.03.003>
- [15] H. Linnan, P.C. Jurs, *J. Mol. Graph. Model.* 2005;23:503–523
<http://dx.doi.org/10.1016/j.jmglm.2005.03.003>
- [16] D.E. Clark, *Evolutionary Algorithms in Molecular Design*, Wiley-VCH, Weinheim, 2000
<http://dx.doi.org/10.1002/9783527613168>
- [17] S. Weisberg, *Applied Linear Regression*, Third edition John Wiley and Sons, New York, 2005.
- [18] Minitab 15.1.0.0. Minitab Inc (www.minitab.com).