Search and analysis of information on porphyrin-like structures for the design of new compounds

Méndez-Cuesta, Carlos A.[1]; Aguayo-Ortiz, Rodrigo[2]; Chávez-Roa, Aline [1]; Pérez-González, Cuauhtémoc [1]

[1]Departamento de Sistemas Biológicos, Universidad Autónoma Metropolitana, Unidad Xochimilco, Calzada del Hueso 1100, Ciudad de México 04960, México

[2]Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México, Circuito Escolar, Ciudad Universitaria, Coyoacán, Ciudad de México, 04510, Mexico

INTRODUCTION

Porphyrins are organic molecules that have a general structure of aromatic macrocyclics, which consist of four pyrrole units called tetrapyrrolic and are connected through methine bonds. Porphyrins and their derivatives are used as photosensitizers in photodynamic therapy, since they generate cytotoxic singlet oxygen in tumor cells. Furthermore, they are capable of photoinactivation of gram-positive and gram-negative bacteria, to produce reactive oxygen species that are lethal to microbial pathogens. They have been modified in the *meso* position for biomedical applications involving photodiagnosis and cancer therapy. QSAR (Quantitative Structure-Activity Relationship) is a computational method that correlates the structural properties and biological activities of molecules, to obtain a reliable statistical model for the prediction of the activities of new chemical entities.

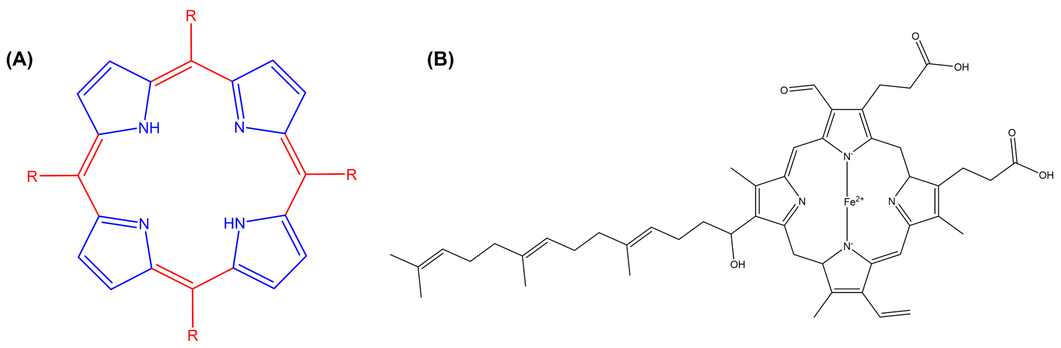


Figure 1. (A) General chemical structure of a porphyrin (B) Hemo group

JUSTIFICATION

According to the National Institute of Statistics and Geography (INEGI), in Mexico, 60,421 deaths were registered due to malignant tumors, the main neoplasms that cause deaths are: lung cancer, breast , colorectal, prostate and stomach, so that early detection and timely treatment of cancer can significantly reduce mortality. Likewise, the emergence of bacteria resistant to antibiotics is a global health problem that is getting worse day by day, so new treatments to eliminate these microorganisms are increasingly urgent to stop these bacterial infections. As the search for new molecules with cytotoxic and antibacterial biological activity is necessary, the use of QSAR models is of great importance for the design of new molecules with the desired biological activity, significantly reducing the number of synthesized candidate molecules to be tested *in vitro* and *in vivo* experiments, which will allow a greater probability of developing a new drug in less time.

METHODOLOGY

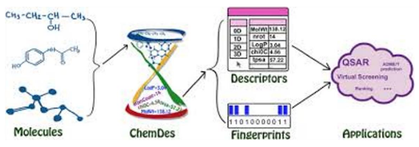
1. Retrieving data from sources

PubMed, CAS SciFinder^n, ELSEVIER, BIDIUAM

2. Activity (IC_{50})

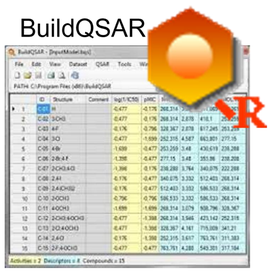
- HCT-116
- HeLa
- MCF-7
- MDA-MB-231
- *Escherichia coli*
- *Staphylococcus aureus*
- *Pseudomonas aeruginosa*

3. Calculation of molecular descriptors



4. Build and analyze QSAR models

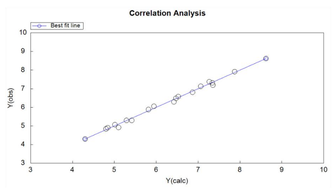
BuildQSAR



5. Model Validation

Internal validation:
-Leave-One-Out
-Q^2
-R^2
-R^2-Q^2
External validation:
-R^{2adj}
-*F*

6. Graphical



RESULTS

Several models were made for each cell line or microorganism, depending on the number of compounds reported with activity (**Figure 2**). Of all these models, the best or the most predictive were chosen and are shown in **Table 1** and **2**. Note that for the cell lines MDA-MB-231, MCF-7 and for the microorganism S. aureus, the external validation was not carried out, because the number of compounds was not sufficient to carry it out, so the analysis was not continued.

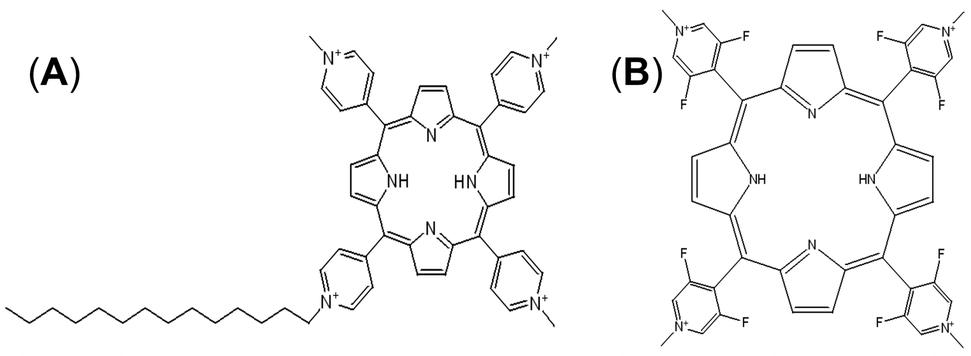


Figure 2. Chemical structure of the molecule with the highest experimental activity (A) *E. coli*; (B) HeLa cell line

Table 1. Summary of the number of molecules to perform the validation

Validation		Cell line or microorganism	No. compounds
Internal validation		MDA-MB-231 (Breast cancer – PR negative)	13
		S. aureus	14
		MCF-7 (Breast cancer – PR positive)	19
	External validation	HCT-116 (Human colon cancer)	27
		HeLa (Cervical cancer)	28
		P. aeruginosa	28
		E. coli	40

Table 2. Summary of the models and their statistical values

Model	Ecuation	R	R^2	R^{2adj}	s	F_QSAR	Q^2	R^2-Q^2	No. descriptors
E. coli	Y1 = + 0.5150 (± 0.2241) X576 - 0.2144 (± 0.0768) X1338 - 1.8925 (± 0.4689) X315 - 0.3938 (± 0.0648) X806 - 0.0270 (± 0.0072) X730 + 0.0026 (± 0.0004) X368 + 5.0329 (± 0.5594)	0.983	0.966	0.957	0.163	101.302	0.951	0.0156	6
P. aeruginosa	Y1 = - 0.2746 (± 0.0973) X732 + 0.0135 (± 0.0042) X464 + 1.0061 (± 0.5444) X311 - 3.7762 (± 1.7083) X1205 - 0.5030 (± 0.1652) X448 + 10.0435 (± 2.4111) X892 - 2.1033 (± 1.8256)	0.978	0.956	0.936	0.131	46.912	0.930	0.0529	6
HeLa	Y1 = - 0.4761 (± 0.2890) X943 + 1.4124 (± 0.5471) X1096 - 0.0158 (± 0.0058) X2 + 5.2739 (± 0.7872)	0.895	0.802	0.765	0.37	21.57	0.651	0.1508	3
HCT-116	Y1 = - 1.0132 (± 0.5596) X249 + 0.2065 (± 0.0594) X1235 - 8.4561 (± 2.5451) X263 - 0.0682 (± 0.0121) X1089 + 0.5229 (± 0.1155) X569 + 8.0716 (± 0.4907)	0.96	0.921	0.898	0.347	39.793	0.856	0.0653	5

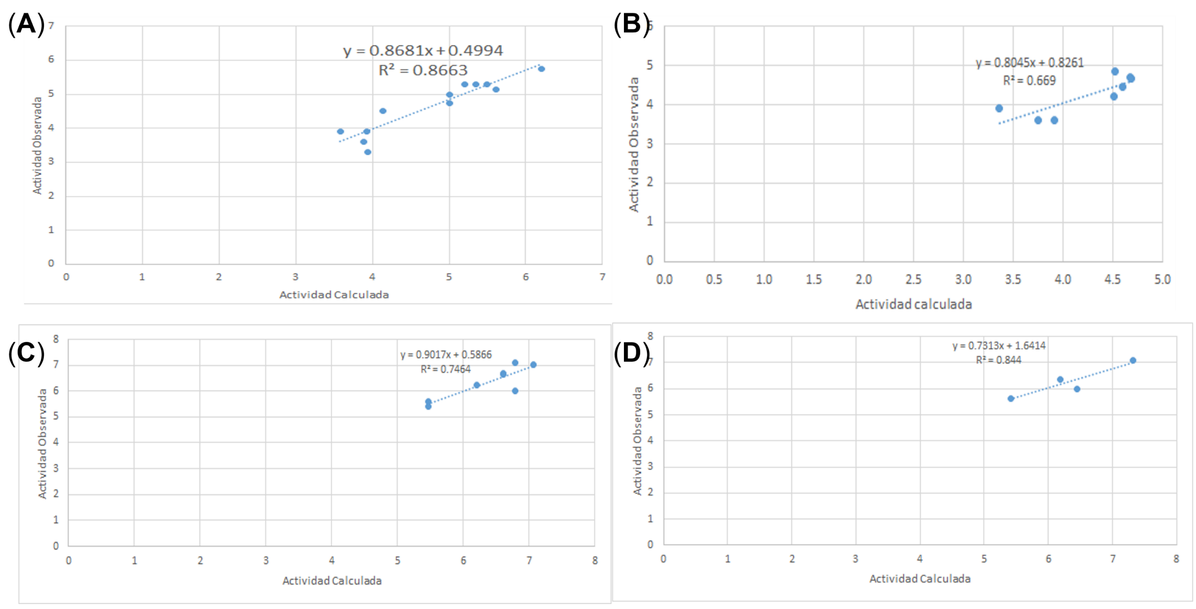


Figure 3. Correlation analysis of the selected model with the test group (A) *E. coli*; (B) *P. aeruginosa;* (C) HeLa; (D) HCT-116.

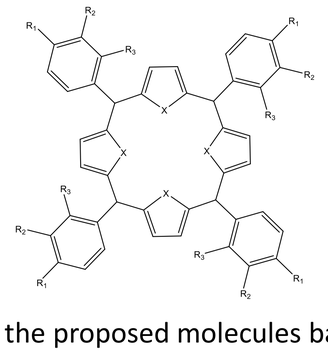


Figure 4. General structure of the proposed molecules based on the models obtained. X=N, S, O; R_1= -OH, -OR, -NH_2, -NO_2, -R; R_2= -OH, -OR, -NH_2, -NO_2, -R; R_3= -OH, -OR, -NH_2, -NO_2, -R. R groups can be combined in different ways.

CONCLUSIONS

1. The QSAR models of HeLa, HCT-116, *E. coli* and *P. aeruginosa*, meet the minimum criteria of statistical quality and with predictive capacity to design new derivatives of tetraphenylporphyrin.
2. New derivatives were designed whose predictions with the QSAR models propose that they will have greater biological activity than the compounds reported.
3. A more exhaustive search of the data on anticancer and antibacterial activity is recommended, in order to carry out studies with a greater number of molecules of each cell line and improve the models designed.

REFERENCES

1. Ajesh, T. (2012). Journal of Medicinal Chemistry, 55, 5110-5120.
2. Banfi, S. (2006). Journal of Photochemistry and Photobiology B: Biology, 85, 28-38.
3. Ensuncho, A. (2017). Información Tecnológica, 28(2), 191-202.
4. Kwon, S. (2019). BMC Bioinformatics, 20(521), 1-12