SciForum
MOL2NET

Multiple Linear Regression Model of Thermolysin Inhibitors as Antihypertensive Pattern.

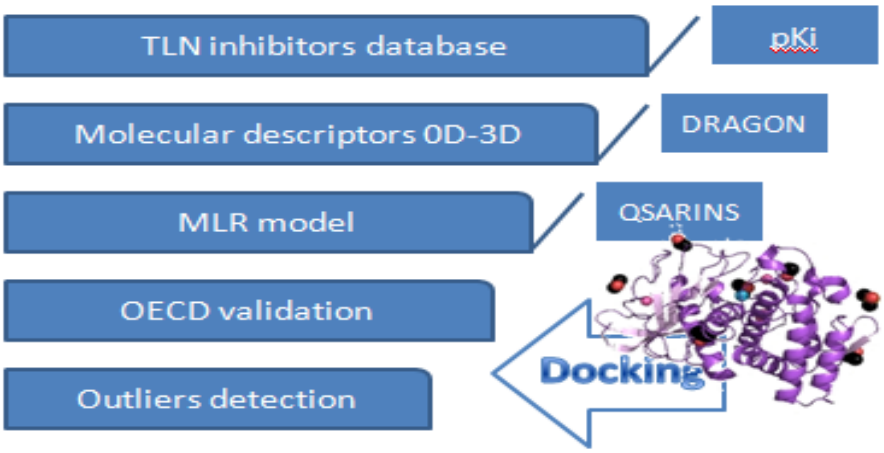
Yudith Cañizares-Carmenate[1], Juan A. Castillo-Garit[1,2,*], Karel Mena-Ulecia[3], Yunier Perera-Sardiña[4], Francisco Torrens[5]

[1] Unit of Computer-Aided Molecular ‘‘Biosilico” Discovery and Bioinformatic Research (CAMD-BIR Unit), Facultad de Química-Farmacia, Universidad Central ‘‘Marta Abreu” de Las Villas, Santa Clara 54830, Villa Clara, Cuba.
[2] Unidad de Toxicología Experimental, Universidad de Ciencias Médicas¨Dr. Serafín Ruiz de Zárate Ruiz¨, Santa Clara 50200, Villa Clara, Cuba.
[3] Centro Interdisciplinario de Neurociencia de Valparaíso, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso, Chile.
[4] Doctorado en Fisicoquímica Molecular, Center of Applied Nanosciences (CENAP), Universidad Andres Bello, Ave. República 275, Santiago, Chile
[5] Institut Universitari de Ciencia Molecular, Universitat de Valencia, Edifici d’Instituts de Paterna, P.O. Box 22085, E-46071 Valencia, Spain
* Corresponding author at: Unidad de Toxicología Experimental, Universidad de Ciencias Médicas¨Dr. Serafín Ruiz de Zárate Ruiz¨, Santa Clara 50200, Villa Clara, Cuba. E-mail addresses: jacgarit@yahoo.es, juancgarit@infomed.sld.cu Tel.: +53 42 273236

Abstract: Thermolysin is a bacterial proteolytic enzyme, considered by many authors as a pharmacological and biological model of other mammalian enzymes, with similar structural characteristics, such as Angiotensin Converting Enzyme and Neutral Endopeptidase. Inhibitors of these enzymes are considered therapeutic targets for common diseases, such as hypertension and heart failure. In this report, a mathematical model of Multiple Linear Regression, for ordinary least squares, and genetic algorithm, for selection of variables, are developed and implemented in QSARINS software, with appropriate parameters for its fitting. The model is extensively validated according to OECD standards, so that its robustness, stability, low correlation of descriptors and good predictive power are proven. In addition, it is found that the model fit is not the product of a random correlation. Two possible outliers are identified in the model application domain but, in a molecular docking study, they show good activity, so we decide to keep both in our database. The obtained model can be used for the virtual screening of compounds, in order to identify new active molecules.

Keywords: *Antihypertensive; Docking; Multiple Linear Regression; QSARINS; Thermolysin.*

Graphical Abstract:



Introduction:

The zinc-metalloproteinases secreted by the gram-positive thermophilic bacterium *Bacillus thermoproteolyticus* is the prototype of the TLN family and has served as a model system to study the inhibition mechanism of other metalloproteinases. Crystallographic data for TLN and various TLN-inhibitor complexes have been used in efforts to model the active site of other TLN-like enzyme [1,2]. These enzymes play a key role in the biosynthesis and metabolism of different bioactive peptides, so that its inhibitors have considerable potential as therapeutic agents [3]. In addition, Thermolysin presents close structural and functional similarities with several mammalian enzymes that are involved in the control of different physiological functions, like neprilysin (NEP) and angiotensin-converting enzyme (ACE), both involved in the control of blood pressure

The dual inhibition of neprilysin and the angiotensin receptor may represent an attractive therapeutic approach for a wide range of cardiovascular diseases, including hypertension, diabetes and heart or kidney failure, in which vasoconstriction, volume overload and neurohormonal activation play a role in the pathophysiology [4]. The structural and functional similarities between TLN, NEP and ACE indicate that Thermolysin inhibitors may also inhibit ACE and NEP and be putative antihypertensives [5]. Dual NEP/ACE inhibitors repress simultaneously two key enzymes that participate in cardiovascular function regulation [6]. This type of inhibitors exerted typical actions of ACE inhibitors and/or NEP inhibitors, such as dose-dependent inhibition of angiotensin I–induced hypertension, protection of atrial natriuretic factor, and enhancement of diuresis, natriuresis, and cGMP urinary excretion [7].

The QSAR studies are useful tools for screening chemicals, especially in early stages of the drug discovery process [8,9]. If these are properly developed and rigorously validated [10,11], they become outstanding tools to evaluate only those that are most promising [12,13].

Materials and Methods:

The database was collected from reports of the international literature (can be seen in the full paper [14]). The database of 176 experimental compounds was split into training (133 compounds) and prediction (43 compounds). For the calculation of molecular descriptors (all the families of descriptors 0-3D), we used DRAGON Software [15]. In order to obtain the MLR model, we used the QSARINS (QSAR-Insubria) software [10].The model was validated in accordance with the principles established by the OECD [11].

Results and Discussion:

The best MLR model obtained with its statistical parameters is shown below:

$$\textit{p}\mathbf{Ki} = -23.62 + 0.52\text{x}\mathbf{Mor07u} - 20.77\text{x}\mathbf{Mor12v} - 3.12\text{x}\mathbf{R5v}^{+} + 2.87\text{x}\mathbf{R5p}^{+} + 23.58\text{x}\,\mathbf{B01[N–O]}$$

N= 133	R^2= 0.714	R^2_{adj}= 0.702
s= 1.230	F= 63.248	R^2-R^2_{adj}= 0.011
LOF = 1.688	K_{xx}= 0.231	ΔK= 0.100
	$RMSE_{tr}$= 1.202	MAE_{tr}= 0.975
	RSS_{tr}= 192.024	CCC_{tr}= 0.833

The model presented an R^2 of 0.714, so it manifests a proper fit for modeling Thermolysin inhibition. In addition, an R^2_{adj} of 0.702, which is indicative of the convenience to add a new descriptor to the model and, together with the low value of the LOF parameter of 1.688, we can say that no existing overfitting is in the model, as it presents a good fit with minimum number of descriptors. The correlation among the model´s descriptors is low because the value of K_{xx} is small (0.231); this allows us to assume that the model has very little redundant information in the selected descriptors.

In order to validate our model, we followed the OECD regulatory principles to ensure their validity, checking the model performance, i.e., the fitting, stability in the cross-validation and the ability to predict new compounds [11]. The fitting and stability of the model were evaluated using internal validation procedures; first, we take into account the parameters of the cross-validation Leave-One-Out (LOO).

According to the obtained results, it is possible to affirm that the internal predictions are good since the variance explained in the prediction by LOO (Q^2_{LOO}= 0.6868) has a comparable value with R^2= 0.7135 (see Fig. 1) with a small error in the predictions ($RMSE_{cv}$= 1.2563 and MAE_{cv}= 1.0190).

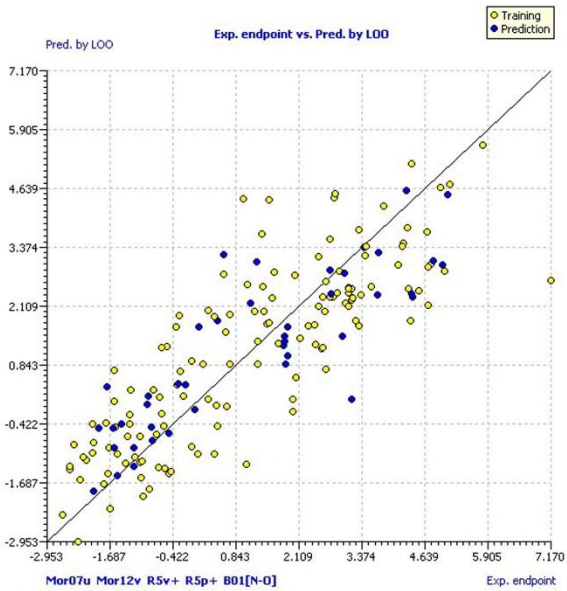


Figure 1: Scatter plot of experimental pKi versus predicted by LOO.

A stronger technique included in the QSARINS is Leaving-Many-Out (LMO), which was developed leaving out the 30% of the dataset to study the behavior of our model. According to this, the model is considered stable because the R^2= 0.7135 and Q^2_{LMO}= 0.6810 values are comparable, and calculation in each iteration of LMO and their averages are comparable to the values of R^2 and Q^2_{LOO} of the model.

We demonstrates that the model was not the result of a casual correlation, using the Y-scrambling procedure, placing the answers at random, so that there was no correlation with descriptors and, as a consequence, the model performance decayed dramatically.

The predictive ability of the model was tested using a series of external predictions (external validation). Using this procedure, we checked the ability of the model to predict new compounds and their statistical parameters showed values equivalent to the model R^2_{ext} = 0.723, $RMSE_{ext}$ = 1.182, MAE_{ext} = 0.91, $PRESS_{ext}$ = 60.091, CCC_{ext} = 0.817.

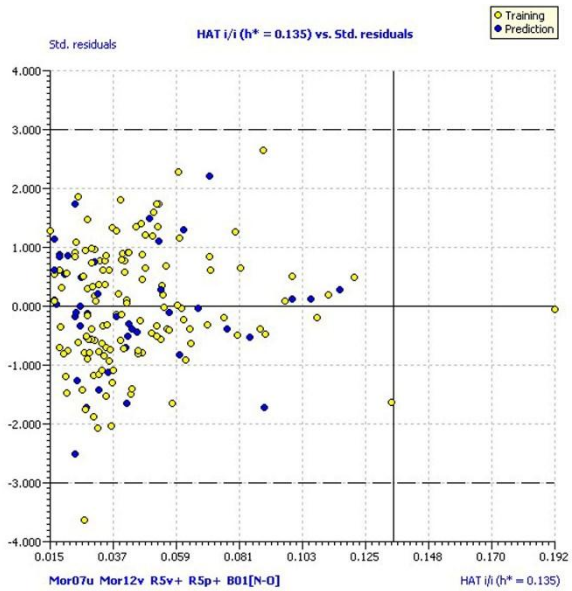


Figure 2: Williams plot. Hat diagonal values versus standardized residuals.

In this study, we used the approach of leverage (*h*) and standardized residuals described in the technical literature [16]. Fig. 2 shows the William graph of the model, where yellow circles represent the compounds of the training and blue circles represent the prediction set. As shown in this figure, most of the compounds are within the applicability domain of the model. There is only one compound (compound 34) with leverage value greater than the critical leverage (*h**= 0.192), but showing residual within the

limits, and one compound (compound 43), having residual off- limits (3.62) though this is within critical leverage. Both cases must be taken into account as potential outlier compounds.
In this sense, we made experiments of molecular docking of these 'outlier' compounds and a group of compounds with similar structural characteristics, since the model provided three-dimensional information of molecules, and it would be interesting to observe their behavior in the active site of Thermolysin. Based on these results, we decided to keep both compounds in the database.

Conclusions: In the present study, a QSAR-RLM model was developed using molecular descriptors of Dragon software, which adequately predicted the inhibitory activity of the enzyme Thermolysin. This model fulfilled all regulatory principles established by the OECD; the robustness of the model was tested through exercised internal validation (LOO, LMO and Y-scrambling), and its predictability was determined with an external prediction set (external validation). Two possible outliers were identified in the model application domain but, in a molecular docking study, they showed good activity, so we decided to keep both in our database.

Acknowledgements:
The authors thank to Prof. Paola Gramatica of the University of Insubria for kindly provide us a free version of QSARINS software to develop this work.

Conflicts of Interest: The authors declare no conflict of interest

Notes:
For the full content of the results presented here, see:
Cañizares-Carmenate, Y., Mena-Ulecia, K., Perera-Sardiña, Y., Torrens, F., Castillo-Garit, J. A. **2016**. **An approach to identify new antihypertensive agents using Thermolysin as model: In silico study based on QSARINS and docking.** *Arab. J. Chem.* (which can be downloaded free of cost using the following link: http://dx.doi.org/10.1016/j.arabjc.2016.10.003)

References:

[1] Bohacek, R.; De Lombaert, S.; McMartin, C.; Priestle, J.; Gru1tter, M. Three-Dimensional Models of ACE and NEP Inhibitors and Their Use in the Design of Potent Dual ACE/NEP Inhibitors. *J. Am. Chem. Soc.* **1996**, *118*, 8231-8249.
[2] Tiraboschi, G. A three-dimensional construction of the active site (region 507-749) of human neutral endopeptidase (EC.3.4.24.11). *Protein Eng* **1999**, *12*, 141-149.
[3] Holland, D. R.; Barclay, P. L.; Danilewicz, J. C.; Matthews, B. W.; James, K. Inhibition of Thermolysin and Neutral Endopeptidase 24.11 by a Novel Glutaramide Derivative: X-ray Structure Determination of the Thermolysin-Inhibitor Complex. *Biochemistry ,,* **1994**, *33*, 51-56.
[4] Manzur, F.; Villarreal, T.; Moneriz, C. Dual inhibition of neprilysin and angiotensin II receptor: promising new strategy in the treatment of cardiovascular disease. . *Rev Colomb Cardiol* **2013**, *20*, 386-393.
[5] Khan, M. T. H.; Fuskevåg, O.; Sylte, I. Discovery of Potent Thermolysin Inhibitors Using Structure Based Virtual Screening and Binding Assays. *J. Med. Chem.* **2009**, *52*, 48-61.
[6] Ruilope, L. M.; Dukat, A.; Böhm, M.; Lacourcière, Y.; Gong, J.; Lefkowitz, M. P. Blood pressure reduction with LCZ696, a novel dual-acting inhibitor of the angiotensin II receptor and neprilysin: a randomised, double-blind, placebo-controlled, active comparator study. *The Lancet* **2012**, *375* 1255-1266.
[7] Laurent, S.; Boutouyrie, P.; Azizi, M.; Marie, C.; Gros, C.; Schwartz, J. C.; Lecomte, J. M.; Bralet, J. Antihypertensive Effects of Fasidotril, a Dual Inhibitor of Neprilysin and Angiotensin-Converting Enzyme, in Rats and Humans. *Hypertension* **2000**, *35*, 1148-1153.

[8] Castillo-Garit, J. A.; Abad, C.; Rodriguez-Borges, J. E.; Marrero-Ponce, Y.; Torrens, F. A Review of QSAR studies to Discover New Drug-like Compounds Actives Against Leishmaniasis and Trypanosomiasis. *Curr. Top. Med. Chem.* **2012**, *12*, 852-865.
[9] Castillo-Garit, J. A.; Marrero-Ponce, Y.; Barigye, S. J.; Medina-Marrero, R.; Bernal, M. G.; Vega, J. M. G. d. l.; Torrens, F.; Arán, V. J.; Pérez-Giménez, F.; García-Domenech, R.; Acevedo-Barrios, R. In silico Antibacterial Activity Modeling Based on the TOMOCOMD-CARDD Approach. *J. Braz. Chem. Soc.* **2015**, *26*, 1218-1226.
[10] Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34*, 2121-2132.
[11] Worth, A. P.; Bassan, A.; De Bruijn, J.; Gallego-Saliner, A.; Netzeva, T.; Patlewicz, G.; Pavan, M.; Tsakovska, I.; Eisenreich, S. The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods. *SAR and QSAR in Environmental Research* **2007**, *18*, 111–125.
[12] Castillo-Garit, J. A.; Del Toro-Cortés, O.; Kouznetsov, V. V.; Puentes, C. O.; Romero Bohórquez, A. R.; Vega, M. C.; Rolón, M.; Escario, J. A.; Gómez-Barrio, A.; Marrero-Ponce, Y.; Torrens, F.; Abad, C. Identification In Silico and In Vitro of Novel Trypanosomicidal Drug-Like Compounds. *Chem. Biol. Drug Des.* **2012**, *80*, 38-45.
[13] Castillo-Garit, J. A.; del Toro-Cortés, O.; Vega, M. C.; Rolón, M.; Rojas de Arias, A.; Casañola-Martin, G. M.; Escario, J. A.; Gómez-Barrio, A.; Marrero-Ponce, Y.; Torrens, F.; Abad, C. Bond-based bilinear indices for computational discovery of novel trypanosomicidal drug-like compounds through virtual screening. *Eur. J. Med. Chem.* **2015**, *96*, 238-244.
[14] Cañizares-Carmenate, Y.; Mena-Ulecia, K.; Perera-Sardiña, Y.; Torrens, F.; Castillo-Garit, J. A. An approach to identify new antihypertensive agents using Thermolysin as model: In silico study based on QSARINS and docking. *Arab. J. Chem.* **2016**, *article in press*, http://dx.doi.org/10.1016/j.arabjc.2016.1010.1003.
[15] Talete, s. r. l.: DRAGON for Windows (Software for Molecular Descriptor Calculations). 5.5 ed., 2007.
[16] Atkinson, A. C.: *Plots, Transformations, and Regression. An Introduction to Graphical Methods of Diagnostic Regression Analysis*; Clarendon Press, 1985.