

Quantitative Structure-Activity Relationship Study of Selective Ligands for the Thyroid Hormone Receptor beta

Huanxiang Liu, Paola Gramatica

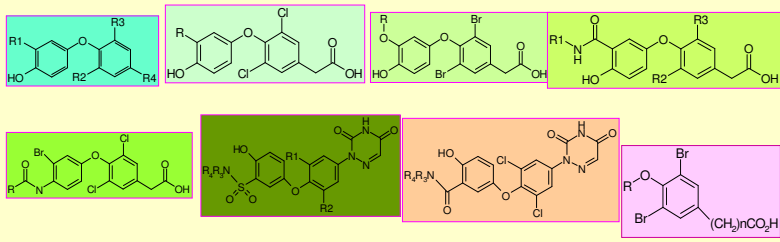
QSAR Research Unit in Environmental Chemistry and Ecotoxicology - INSUBRIA University (Varese - ITALY)

e.mail: paola.gramatica@uninsubria.it; web: <http://www.qsar.it>

Abstract

In this work, an accurate and reliable QSAR model of 85 selective ligands for the thyroid hormone receptor beta 1 (TRB1) was developed, based on theoretical molecular descriptors to predict the binding affinity of compounds with receptor. The structural characteristics of compounds were described wholly by a large amount of molecular structural descriptors calculated by DRAGON. Six most relevant structural descriptors to the studied activity were selected as the inputs of QSAR model by a robust optimization algorithm Genetic Algorithm. The built model was fully accessed by various validation methods, including internal and external validation, Y-randomization test, chemical applicability domain and all the validations indicates that the QSAR model we proposed is robust and satisfactory. Thus, the built QSAR model can be used to fast and accurately predict the binding affinity of compounds (in the defined applicability domain) to TR. At the same time, the model proposed could also identify and provide some insight into what structural features are related to the biological activity of these compounds and provide some instruction for further designing the new selective ligands for TRB1 with high activity.

Data set: The affinity data of 85 ligands to $\beta 1$ isoform of the human thyroid hormone receptor (TRB1), taken from five references in terms of IC50 [1-4] or Ki [5] values and converted to $-\log\text{IC}_{50}$ were used. The involved compounds mainly included the following chemical classes:



Methodology

Molecular Descriptors: The molecular descriptors were calculated by the software DRAGON [6]. The descriptors typology is: OD: Constitutional descriptors; 1D: Empirical, Functional groups, Properties, Atom-centred fragments descriptors; 2D: Autocorrelations, Topological, Molecular walk counts, Galvez topological charge indices, BCUT descriptors; 3D: Geometrical, Randic molecular profiles, WHIM, GETAWAY.

Descriptor Selection and Construction of Models: Genetic Algorithm [7], as a powerful optimization method, was used for variable selection. To make the chemists' job easier, the models were built using the simple Multiple Linear Regression method.

Validation and evaluation of Models: The robustness of the models and their internal predictive ability were evaluated by both Q2 based on leave-one-out cross-validation and bootstrap. The proposed models were also checked for reliability and robustness by permutation testing: new models were recalculated for randomly reordered response (Y scrambling). The external validations were performed both by splitting the original data set into training set and prediction set.

Results and Discussion

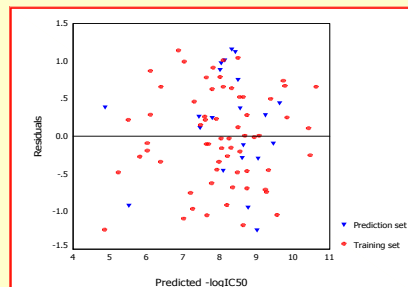
Splitting the data set into training and prediction set by Kohonen Self Organizing Maps (SOM)

We can conclude that the activity of the studied compounds mainly depends on molecular polarity, size, shape and nucleophilic reactivity.

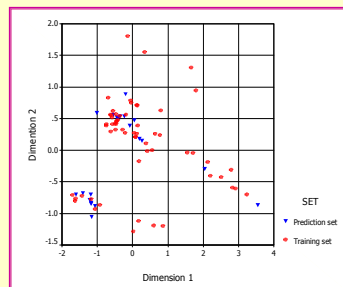
The MLR model between the structural descriptors and the pIC_{50} of the compounds in the training set by SOM splitting.

Variable	The meaning of variables	Reg.coeff.	Em.coeff	Std.Coef.
Intercept	Constant	9.996	1.424	0
GATS4c	Geary autocorrelation - lag 1 weighted by atomic Sanderson electronegativities	-7.683	1.621	-0.262
FR10q	Eigenvalue 8 from edge adjacency matrix weighted+edge degrees	-2.906	0.388	-1.399
FR10p	Eigenvalue 7 from edge adjacency matrix weighted-dipole moments	4.411	0.526	1.665
CH06	topological charge index of order 6	1.994	0.653	0.316
MOE6	R maximal autocorrelation of lag 6 index weighted by atomic van der Waals volume	60.550	10.996	0.339
H04	the number of the H atom attached to alpha C atom	-0.383	0.080	-0.287

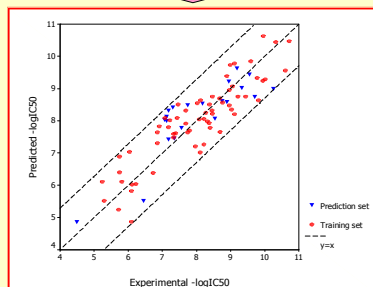
Model parameters: $n=64$, $R^2=0.836$, $R^2_{cv}=0.819$, $Q^2_{cv}=0.780$, $RMSE=0.350$, $K_{cv}=0.439$, $K_{cv}=0.466$, $\sigma=0.582$



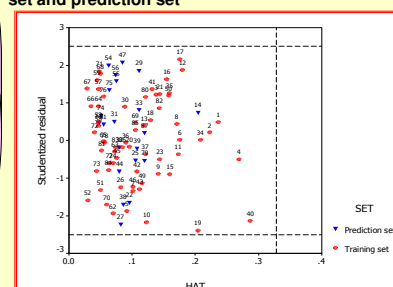
Residuals vs. Predicted $-\log\text{IC}_{50}$ values for training set and prediction set



MDS map for the training set and prediction set split by SOM neural network



Predicted $-\log\text{IC}_{50}$ values vs. experimental values for training set and prediction set (Two side lines express the confidence interval of 95%)



Williams plot

From the above results, both for the training set and external prediction set, it can be seen the model we suggest matches the high quality parameters not only with good fitting power, but mainly with high capability of assessing external data.

In order to ensure that the results were not conditioned by the data distribution in descriptor space, the built model was also validated by the training and prediction set selected from random method and activity sampling.

Splitting method	Training set		RMSE	Prediction set		RMSE
	R^2	Q^2		R^2_{pred}	Q^2_{cv}	
SOM	0.836	0.793	0.550	0.730	0.711	0.702
Randomness	0.813	0.764	0.600	0.808	0.529	0.554

The satisfactory results proved that the built model was not conditioned by the data distribution in structural space, as well as its robustness and reliability.

Conclusion

The built QSAR model was accessed comprehensively (internal and external validation) and all the validations indicates that it is robust and satisfactory, and that the selected descriptors can account for the structural features responsible for the binding affinity of compounds to TR, as well as GA is an effective method to select descriptors.

By interpreting the molecular descriptors in the regression model, we can conclude that the activity of the studied compounds mainly depends on molecular polarity, size, shape and nucleophilicity.

The QSAR model developed in this study can provide a useful tool to predict the activity of the new compounds and also to design new compounds with high activity.

References This work is now published in: *Bioorg. Med. Chem.*, 2007, 15 (15), 5251-5261.

- Ye, L. et al. *J. Med. Chem.* 2003, 46, 1580-1588.
- Dow, R.L. et al. *Bioorg. Med. Chem. Letter.* 2003, 13, 379.
- Hangeland, J.J. et al. *Bioorg. Med. Chem. Letter.* 2004, 14, 3549.
- Li, Y.-L. et al. *Bioorg. Med. Chem. Letter.* 2006, 16, 884-886.
- Collazo, A. M. G. et al. *Bioorg. Med. Chem. Letter.* 2006, 16, 1240-1244.
- Todeschini, R. et al. DRAGON-Software for the calculation of molecular descriptors. Version 5.3 for Windows, Talet srl, Milan, Italy, 2005.
- Rogers, D. and Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* 1994, 34, 854-866.