XIX SBQT
Simpósio Brasileiro de Química Teórica 2017
12 a 17/Nov, 2017, Águas de Lindóia/SP, Brasil

A new approach for sampling descriptors in 4D-QSAR methodology using computational geometry

Authors: João Vitor Soares Tenório[1], Loïc Pascal Gilles Cerf[1], João Paulo Ataide Martins[2]

(1) Computer Science Department, Federal University of Minas Gerais
(2) Chemistry Department, Federal University of Minas Gerais

Abstract: The quantitative structure-activity relationship (QSAR) is an important research field in theoretical medicinal chemistry, which deals with the prediction of the biological activities of new compounds using mathematical relationships based on structural, physicochemical and conformational properties of previously tested potential agents [1]. A recently developed 4D-QSAR approach named LQTA-QSAR [2], is based on the generation of a conformational ensemble profile, CEP, for each compound followed by the the calculation of 3D descriptors for a set of compounds. This methodology explores jointly the main features of CoMFA [3] and Hopfinger's 4D-QSAR [4] paradigms. LQTA-QSAR makes use of the GROMACS free package to run the molecular dynamics, MD, simulations. The CEPs generated by the MD simulations are aligned and put into a virtual 3D box or grid, which acts like a virtual receptor, and different types of atoms, ions or functional groups, called probes, are used to compute the energy values for the interactions that the selected probe experiences in a respective position of the regular 3D lattice using LQTAgrid program.

One problem observed with the approach described above is that the probe crosses the CEP of the compound and some descriptors presents unrealistic values when the probe falls into or close to an atom of the CEP. Besides, as the 3D lattice plays the role of the receptor, points inside the CEP cannot be associated to a portion of the receptor interacting with the ligand.

This work presents a new approach to generate the 3D region that plays the role of the virtual receptor. Our method takes into account the shape of the CEP using computational geometry algorithms to prevent the probe passes through the points inside the CEP. First, the region which envelop the CEP of each compound in 3D lattice is estimated by generating a convex hull, that is the smallest convex polyhedron containing the atoms of the CEP [6].

After the convex hull generation, the points of the path that the probe must travel are sampled in polar coordinates, to explore the shape of the convex hull efficiently. First, the proposed method requires three parameters: , which is the variation of each angle of the spherical coordinates in each step, Δr which is the radius variation of the coordinate, and N, which is the number of times the radius will be varied in a step

equals to Δr. For each angle step, the initial radius value is defined as the intersection of the angle with the convex hull incremented with the predefined distance equals to 1.5 Å. Then, that point is collected and the remaining N-1 points varying up the radius with step equals to Δr. In this way, the points are sampled respecting a minimum distance of 1.5 Å and the maximum distance of 1,5 Å+(N-1)*Δr from a point to the convex hull. In the present work, these parameters were empirically defined as $\Delta\theta$=3, Δr=1 e N=7.

In order to validate this new approach two data sets of compounds were investigated and QSAR models were built. The first data set (data set 1) is formed by 49 compounds used as target of study in the development of new drugs against prostate cancer. The second data set (data set 2) is formed by 48 compounds with antimalarial activity. After a variable selection with OPS algorithm [5] a PLS model was built for each data set and the results are presented in Table 1.

Table1: Statistical parameters for QSAR models obtained

Data set	Nº of variables	Nº of latent variables	R^2	RMSEC	Q^2	RMSECV
1	10	3	0.58	0.46	0.51	0.50
2	16	8	0.87	0.24	0.73	0.35

These results are very promising, since no models with Q^2 superior to 0.50 were obtained with descriptors generated with the original approach and the descriptors generated with this new approach are more physically meaningful.

In order to obtain better results other regression methods will be tested. Besides, an external validation will be performed.

Key-words: LQTA-QSAR, Molecular dynamics, Convex hull, Computational Geometry

Support: This work has been supported by FAPEMIG and CAPES

References:

[1] M. M. C. Ferreira, J. Braz. Chem. Soc., 13, 742 (2002).
[2] J. P. A. Martins, E. G. Barbosa, K. F. M. Pasqualoto, M. M. C. Ferreira, J. Chem. Inf. Model., 49, 1428 (2009).
[3] R. D. Cramer, D. E. Patterson, J. D. Bunce, J. Am. Chem. Soc., 110, 5959 (1988).
[4] A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, C. Duraiswami, J. Am. Chem. Soc., 119, 10509 (1997).
[5] R. F. Teófilo, J. P. A. Martins, M. M. C. Ferreira, J. Chemom., 23, 32 (2009).
[6] M. Berg, O. Cheong, M. Kreveld "Computational Geometry: Algorithms and Applications." (2008), Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd edition.