Inorganica Chimica Acta 392 (2012) 229–235

Contents lists available at SciVerse ScienceDirect

Inorganica Chimica Acta

journal homepage: www.elsevier.com/locate/ica

Antitumor structure–activity relationship in bis-stannoxane derivatives from pyridine dicarboxylic and benzoic acids

José Antonio Valcarcel [a], Rodrigo Said Razo-Hernández [a], Laura Leticia Valdez-Velázquez [a], Manuel Villanueva García [b], Ángel Andrés Ramos Organillo [a], Oscar F. Vázquez-Vuelvas [a], Miguel A. García Ruiz [c], Zeferino Gómez-Sandoval [a,*]

[a] Facultad de Ciencias Químicas, Universidad de Colima, Km. 9 Carr. Colima-Coquimatlán, C.P. 28400, Coquimatlán, Col., Mexico
[b] Departamento de Ciencias Químicas, División de Ciencias Naturales y Exactas, Universidad de Guanajuato. Noria Alta S/N36050 Guanajuato, Gto., Mexico
[c] Department of Computer Science and Mathematics, Algoma University, 1520 Queen Street East, Sault Ste. Marie, Ontario, Canada

ARTICLE INFO

Article history:
Received 11 January 2012
Received in revised form 14 June 2012
Accepted 15 June 2012
Available online 28 June 2012

Keywords:
Organotin compounds
QSAR
Neural networks
Genetic algorithms
MCF-7
WiDr

ABSTRACT

Organometallic compounds have been proposed as new drugs for cancer treatment due to the large metal mesh with DNA. This study estimated four quantitative structure–activity relationship (QSAR) descriptive models relating antitumor biological activity essays (ID_{50}) in breast (MFC-7) and colon cancer (WiDr) cell lines with stannoxanes compounds derived from 2,6-pyridine-dicarboxylates [$C_5H_3N(COO)_2SnRR'$] (R, R′ = alkyl, aryl) and di-*n*-butyltinbis-benzoates [$(C_6H_5COO)_2SnBu_2$]. A series of thermodynamic, structural and molecular descriptors were calculated from the geometric and electrical optimizations (PM3) of the two organotin series, in order to correlate the biological activity of the tin esters. The use of genetic algorithms and multilinear correlations yielded four mathematical models, each one with four descriptors related to molecular, area/volume, lipophilicity, and molecular dipole polarizability. These descriptors were entered into a back-propagation neural network to obtain theoretical descriptive models with the goal of proposing the development of new organotin molecules with enhanced antitumor activity.

1. Introduction

The coordination and organometallic complexes have an increasing importance in the medicine, particularly in the oncology. The cisplatin in one of the widely used anticancerous agent; however, other platinum complexes Pt(II) are being introduced in the antitumor therapies [1]. The cell resistance to the cisplatin and their analogous is the principal reason in unsuccessful treatments and the clinical relapse. To surpass this problem, not only new Pt(II) and Pt(IV) complexes have been developed, but also complexes with metals other than platinum, with favorable results [2].

The search of new drugs for the treatment of neoplasic illnesses has led to consider compounds based on metals with high biological activity. The main feature of this search focuses in the selectivity, minor side effects and no drug-resistance. In this context, tin(IV) carboxylate complexes have been very attractive during the last two decades [3]. They've shown greater cytotoxicity than many stablished anticancerigenous agents. Furthermore, recent essays in daunomycin-resistant K562/R cell line, indicated that these compounds do not induce acquisition of multidrug resistance [4,5].

Additionally, among these types of molecules, the tin bis-esters have also exhibited *in vitro* activity in cancer cell lines of human breast (MCF-7) and colon (WiDr). The antitumor activities of these molecules, which present a penta and hexacoordinated tin atom, are showed in a higher degree than cisplatin and mitomicin [3–14].

There are reports of biological essays indicating that the coupling target of many organometallic compounds is the DNA, besides some works propose molecular action mechanisms and sites where the coordination metal binds with genetic and mitochondrial material [2]. It is known that the bonding capacity of metals with biomolecules depends on the nature of the electron-donor groups linked to the central metal following the Two-Pole Complementary Principle (TPCP) [15].

Gielen and co-workers [3] have reported studies with organtin compounds in presence of different cell lines to evaluate the biological activity, showing interesting results. This study denotes that the biological activity of these derivatives can be associated with relatively small structural changes, suggesting a structure–activity relationship.

The QSAR models are used with several purposes: (1) to understand the relationship between the structure and the biological activity of known molecules; (2) to evaluate the biological activity of known molecules; (3) to predict the activity of compounds even before synthesizing them; and (4) to identify what kind of physical

* Corresponding author. Fax: +52 31 23 16 1163.
E-mail address: zgomez@ucol.mx (Z. Gómez-Sandoval).

0020-1693/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ica.2012.06.029

and physicochemical properties of the studied molecules have more influence in the biological activity, which is the most important issue in the rationalization or the action mechanism. Therein, a method to obtain quantitative structure–activity relationship models is by means of the implementation of rationalized and systematic statistical techniques [16]. The QSAR methodologies suppose an existing correlation between the biological activity of a molecule and its structure. This relationship can be shown with a mathematical equation with the intention of reproducing, and consequently, to predict a given biological activity of a set of molecules. At that point, the biological activity is seen as a result of contributions of numerous structural factors of the compounds. These factors, known as molecular descriptors, are parameters based on numerical values and represent electronic (quantum mechanical), steric and physicochemical structure of the studied molecules. The main strategy is finding a mathematical equation [17,18] by means of a mathematical-statistical process via least squares [19] applied to the descriptors, with the aim of that the biological activity of a set of compounds fit to a multilinear regression (MLR) model [20].

The advance of computing systems has allowed an important progress in the development of programs to obtain reliable and accurate QSAR models; as a consequence, the current software systems have the capacity to calculate a great quantity of descriptors. Additionally, the programs based on artificial intelligence, such as genetic algorithms (GA) and artificial neural networks (ANN), increase the usefulness of these tools in the drug design via QSAR techniques [20,21].

The approximation using genetic algorithms is a statistical method for analyzing complex systems described by numerous variables. This technique allows obtaining important and rejecting non-relevant information through the selection of suitable models from a larger set of models with different variables; in other words, the procedure consists of the evolution of a mathematical models population with a certain target function in order to find the optimal models, in an analogous manner to the genetic evolution [22,23]. Each model is named chromosome, being a binary vector constituted by positions (gens) that corresponds to the diverse descriptors, which will describe a phenotype (biological activity). This analysis carries out by the next procedure: (1) creation of a population: the process begins with the random construction of the models of a population by means of the chromosomes formation via random combinations; (2) interlacing, genes of lower weight from one chromosome are discarded, and the higher weight genes combine with the higher ones of another chromosome. If the new chromosome describes better the phenotype, it is considered as a model, if not, it is discarded until to find the very best chromosome system that adjusts to our phenotype in an evolution process; (3) mutation, a random gene is inserted into a chromosome, with the aim of increasing the success probabilities to represent the phenotype. If the mutation shows an improvement, the chromosome is maintained, if not, it is discarded; (4) condition of finalization, the chromosomes are compared with respect to the phenotype, if the results are not favorable, an iterative process is initiated with the goal of finding optimal results to represent the phenotype [25,26].

The artificial neural networks (ANN) are computing tools based on a learning mechanism of neurons (basic computing unit) in the nervous system of the animals. This computing technique consists of interconnected neurons processing input signals to transform them into output signals [25,26]. In this sense, the back-propagation artificial neural networks (BPANN) consists of a basic architecture formed by input neurons, hidden neurons and output neuron. Each neuron has a random and variable weight factor. The difference between the data (calculated results versus expected results) must be the lowest (close to zero); if not, the neurons weight is adjusted until the convergence criterion is reached; therefore, this technique is known as back-propagation process. Additionally, QSAR models using back-propagation processes are widely employed to obtain models and are very useful for studies of predictive type because these model board problems considering the non-linearity, as well as they predict the biological activity of the studied compounds in a more precise mode and with better values of adjustment in the correlation [24].

This present work has the goal of obtaining QSAR models that describe the relation between the antitumor activity of stannoxanes derived from 2,6-pyridine-dicarboxylate dibuthyl tin [$C_5H_3N(COO)_2SnRR'$] and di-*n*-butyltinbis-benzoate [$(C_6H_5COO)_2SnBu_2$] with several descriptors, implementing improved genetic algorithms and multilinear regression techniques by the application of BPANN. The goal is reproducing and predicting theoretically the biological activity of selected compounds and other not included for obtaining the models. Additionally, the obtained models are analyzed with the aim of finding what kind of molecular properties have more influence in the molecular activity of the compounds being studied, which represents an important issue during the rationalization of action mechanisms and for proposing new drugs.

2. Data set

Two series of stannoxanes derivatives were studied, whose structures and biological activities reported as median infective dose (ID_{50}) with the WiDr and MCF-7 cell lines, have been previously published by Gielen.[3] In the Table 1, the basic structure of compounds 2,6-pyridine-dicarboxylate dibuthyl tin ($C_5H_3N(COO)_2SnR\acute{R}$) is shown, where R1 and R2 are the different substituents. The Table 2 shows the basic structure of the compounds di-*n*-butyltinbis-benzoate [$(C_6H_5COO)_2SnBu_2$], where X, Y and Z represent the ring substituents of the phenyl group.

3. Computational details

1.- The neutral geometry of all stannoxane structures (2,6-pyridine-dicarboxylate dibuthyl tin derivatives constitute the molecules series A and di-*n*-butyl tin bis-benzoate derivatives represent the molecules series (B) were fully optimized without any symmetry restrictions at semiempirical level of theory PM3 [26] and numerical precision with minimal base as implemented by the Spartan® program [27]. The minimum energy structures were verified by calculating the vibrational frequencies at the same level of theory to ensure that the resulting geometries were minima on the potential energy surface.

2.- For the QSAR study, molecular descriptors were calculated from optimized geometries of tin-esters. The selection of the descriptors was initially arbitrary, but the ones with more relevancies were considered, from the chemical point of view, to describe adequately the biological activity of the stannoxanes in the MCF-7andWiDr cell lines reported by Gielen [3]. Therefore, the total of descriptors was 24 and they were calculated with the Spartan® software [27], among the calculated ones were dipole moment, HOMO–LUMO gap, polarizability, ovality, and others.

3.- Other descriptors included in the QSAR study were calculated using the Dragon® computational package [28,29]. This software consists of 1630 descriptors classified in several groups, and three of them were selected: constitutional descriptors (48), of functional groups (154) and of molecular properties (29), giving a total of 231 descriptors.