SciForum
Mol2Net

Multi-Viral Targets Entropy QSAR for Antiviral Drugs

Francisco J. Prado-Prado [1,*] and Xerardo García-Mera [2]

1 Biomedical Sciences Department, Health Sciences Division, University of Quintana Roo, 77019 Chetumal, Mexico

2 Department of Organic Chemistry, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

Published: 4 December 2015

Abstract: The antiviral QSAR models today have an important limitation. Only they predict the biological activity of drugs against only one viral species. This is determined due the fact that most of the current reported molecular descriptors encode only information about the molecular structure. As a result, predicting the probability with which a drug is active against different viral species with a single unifying model is a goal of major importance. In this we use the Markov Chain theory to calculate new multi-target entropy to fit a QSAR model that predict by the first time a ms-QSAR model for 900 drugs tested in the literature against 40 viral species and other 207 drugs no tested in the literature using entropy QSAR. We used Linear Discriminant Analysis (LDA) to classify drugs into two classes as active or non-active against the different tested viral species whose data we processed. The model correctly classifies 31 188 out of 31 213 non-active compounds (99.92%) and 432 out of 434 active compounds (99.54%). Overall training predictability was 98.56%. Validation of the model was carried out by means of external predicting series, the model classifying, thus, 15 588 out of 15 606 non-active compounds and 213 out of 217 active compounds. Overall validation predictability was 98.54%. The present work report the first attempts to calculate within a unify framework probabilities of antiviral drugs against different virus species based on entropy analysis.

Keywords: Antiviral drugs; QSAR; Entropy; Mutli-tasking Learning; Markov Chain model; Linear Discriminant Analysis

1. Introduction

Examples of diseases caused by viruses include the common cold (produced by any one of a variety of related viruses), AIDS (caused by HIV) and cold sores (caused by herpes simplex); which produced some of the major health problems in the last 30 years. Other relationships are being studied such as the connection of Human Herpesvirus 6 (HHV6), one of the eight

known members of the human herpes virus family, with organic neurological diseases such as multiple sclerosis and chronic fatigue syndrome. Recently, it has been shown that cervical cancer is caused, at least partially, by papillomavirus, representing the first significant evidence in humans for a link between cancer and an infective agent. The relative ability of viruses to cause disease is described in terms of virulence.

Consequently, there is an increasing interest on the development of rational approaches for discovery of antifungal drugs. In this sense, a very important role may be played by computer-added drug discovery techniques based on Quantitative-Structure-Activity-Relationship (QSAR) models (1). Unfortunately, almost QSAR studies, including those for antiviral activity and others, use limited databases of structurally parent compounds acting against one single fungus species (2). One important step in the evolution of this field was the introduction of QSAR models for heterogeneous series of antimicrobial compounds; see for instance the works of Cronin, de Julián-Ortiz, Galvéz, Gárcía-Domenech, Gosalbez, Marrero-Ponce, Torrens, *et al.* and others (3-15). As a result, researchers may predict very heterogeneous series of compounds but often need to use/develop as many QSAR equations as microbial species are necessary to be predicted. In any case, if you aim to predict activity against different targets you still need to use one different QSAR model for each target.

An interesting alternative, is the prediction of structurally diverse series of antimicrobial compounds (antiviral in this case) against different targets (mechanisms) using complicated non-linear Artificial Neural Networks with multi-class prediction, *e.g.* the work of Vilar *et al.* (16). We can understand strategies developed in this sense as Multi-Objective Optimization (MOOP) techniques; in this case we pretend to optimize the activity of antiviral drugs against many different objectives or targets (viral species). A very useful strategy related to the MOOP problem use Derringer's desirability function desirability function and many QSAR models for different objectives (17). In this sense, it is of major importance the development of unified but simple linear equations explaining the antimicrobial activity, in the present work antiviral activity, of structurally-heterogeneous series of compounds active against as many targets (viral species) as possible. We call this class of QSAR problem the multi-target QSAR (mt-QSAR) (18, 19).

There are near to 2000 chemical molecular descriptors that may be in principle generalized and used to solve the mt-QSAR problem. Many of these indices are known as Topological Indices (TIs) or simply invariants of a molecular graph G. We can rationalize G as a draw composed of vertices (atoms) weighted with physicochemical properties (mass, polarity, electro negativity, or charge) and edges (chemical bonds) (20). In any case, many of these indices have not been extended yet to encode additional information to chemical structure. One alternative to mt-QSAR is the substitution of classic atomic weights by target specific weights. For instance, we introduced and/or reviewed TIs that use atomic weights for the propensity of the atom to interact with different microbial targets (21) or undergoes partition in a biphasic systems or distribution to biological tissues (22-24). The method, called MARCH-INSIDE approach, Markovian Chemicals In Silico Design, calculates TIs using Markov Chain theory. In fact, MARCH-INSIDE define a Markov matrix to derive matrix invariants such as stochastic spectral moments, mean values, absolute probabilities, or entropy measures, for the study of molecular properties. Applications to macromolecules have extended to RNA, proteins, and blood proteome (25-30). In particular, one of the classes of MARCH-INSIDE

descriptors is defined in terms of entropy measures; which have demonstrated flexibility in many bioorganic and medicinal chemistry problems such as: estimation of anticoccidial activity, modelling the interaction between drugs and HIV-packaging-region RNA, and predicting proteins and virus activity (24, 31-33). We give high importance to entropy measures due to it have been largely demonstrate as an excellent function to codify information in molecular systems, see for instance the important works of Graham (34-39). However, have not been studied the proficiency of entropy indices (of MARCH-INSIDE type or not) to solve the mt-QSAR problems in antiviral compounds.

The present study develops the first mt-QSAR model based on entropy indices to predict antiviral activity of drugs against different viral species. The model fits one of the largest datasets used up-to-date in QSAR studies, number of entries 47 000+ cases; which is the result of forming different (antiviral compounds/viral target) pairs.

2. Results and Discussion

One of the main advantages of the present stochastic approach is the possibility of deriving average thermodynamic parameters depending on the probability of the states of the MM. The generalized parameters fit on more clearly physicochemical sense with respect to our previous ones (24, 41, 42). In specific, this work introduces by the first time a linear mt-QSAR equation model useful for prediction and MOOP of the antiviral activity of drugs against different viral target species or objectives. The best model found was:

$$actv = 0.38 \cdot {}^{3}\theta_{s}(s)_{het} - 0.84 \cdot {}^{0}\theta_{s}(s)_{total} - 0.91 \cdot {}^{0}\theta_{s}(s)_{C_{sat}} + 0.89 \cdot {}^{1}\theta_{s}(s)_{C_{sat}} + 2.01 \cdot {}^{0}\theta_{s}(s)_{Csp\&sp2} - 0.32 \cdot {}^{5}\theta_{s}(s)_{h\text{-}het} - 4.71 \quad (8)$$

$$N = 31190 \quad \lambda = 0.38 \quad \chi^2 = 377.43 \quad p < 0.001$$

In the model the coefficient λ is the Wilk's statistics, statistic for the overall discrimination, χ^2 is the Chi-square, and p the error level. In this equation, ${}^{k}\theta_s$ where calculated for the totality (T) of the atoms in the molecule or for specific collections of atoms. These collections are atoms with a common characteristic as for instance are: heteroatom (*Het*), unsaturated Carbon atoms (C_{unst}), saturated Carbon atoms (C_{sat}) and hydrogen bound to heteroatom (*H-Het*. The model correctly classifies 31 188 out of 31 213 non-active compounds (99.92%) and 432 out of 434 active compounds (99.54%). Overall training predictability was 98.56%. Validation of the model was carried out by means of external predicting series, the model classifying, thus, 15 588 out of 15 606 non-active compounds and 213 out of 217 active compounds. Overall validation predictability was 98.54%.

The more interesting fact is that ${}^{k}\theta_s$ have the skill of discerning the active/no-active classification of compounds among a large number of viral species. This property is related to the definition of the ${}^{k}\theta_s$ using species-specific atomic weights (see supplementary material file for method). It allows us to model by the first time a very heterogeneous a diverse data with more than 47 470 cases (one of the largest in QSAR). Another interesting characteristic of the model is that the ${}^{k}\theta_s$ used as molecular descriptors depend both on the molecular structure of the drug and the viral species against which the drug must act. The codification of the molecular structure is basically due to the use of the adjacent factor α_{ij} to encode atom-atom bonding, molecular connectivity. The other aspect that allows encoding molecular structural changes is that the entropy ${}^{k}\theta_s$ are atom-class specific. This property is related to the definition of the ${}^{k}\theta_s$. For example, one change in the molecular structure of, e.g. S by O, necessarily implies a change in the moments of interaction. Moreover, the most interesting fact is that ${}^{k}\mu_s$ are

the molecular descriptors reported for antimicrobial mt-QSAR studies able to distinguish among a large number of viral species. The present work is the first reported mt-QSAR model using entropy ${}^{k}\theta_s$ as a molecular descriptor that allow one predicting antiviral activity of any organic compound against a very large diversity of viral pathogens.

3. Materials and Methods

3.1. Markov entropy (θ_k) for drug-target k-th step-by-step interaction

One can consider a hypothetical situation in which a drug molecule is free in the space at an arbitrary initial time (t_0). It is then interesting to develop a simple stochastic model for a step-by-step interaction between the atoms of a drug molecule and a molecular receptor in the time of desencadenation of the pharmacological effect. For the sake of simplicity, we are going to consider from now on a general structure less receptor. Understanding as structure-less molecular receptor a model of receptor which chemical structure and position it is not taken into consideration. Specifically, the molecular descriptors used in the present work are called stochastic entropies θ_k, which are entropies describing th connectivity and the distribution of electrons for each atom in the molecule (40). The initial entropy of interaction a j-th atom of the drug with the target ${}^{0}\theta_j(s)$ is considered as a state function so a reversible process of interaction may be came apart on several elemental interactions between the j-th atom and the receptor. The 0 indicates that we refer to the initial interaction, and the argument (s) indicates that this energy depends on the specific viral species. Afterwards, interaction continues and we have to define the interaction probability ${}^{k}\theta_{ij}(s)$ between the j-th atom and the receptor for specific viral specie (s) given that i-th atom has been interacted at previous time t_k. In particular, immediately after of the first interaction ($t_0 = 0$) takes place an interaction ${}^{1}p_{ij}(s)$ at time $t_1 = 1$ and so on. So, one can suppose that, atoms begin its interaction whit the structure-less molecular receptor binding to this receptor in discrete intervals of time t_k. However, there several alternative ways in which such step-by-step binding process may occur (24, 41, 42).

The entropy ${}^{0}\theta_j(s)$ will be considered here as a function of the absolute temperature of the system and the equilibrium local constant of interaction between the j-th atom and the receptor ${}^{0}\gamma_j(s)$ for a give microbial species. Additionally, the energy ${}^{1}\theta_{ij}(s)$ can be defined by analogy as $\gamma_{ij}(s)$ (24, 41, 43):

$${}^{0}\theta_j(s) = -R \cdot T \cdot \log {}^{0}\Gamma_j(s) \quad (1) \qquad {}^{1}\theta_{ij}(s) = -R \cdot T \cdot \log {}^{1}\Gamma_{ij}(s) \quad (1)$$

The present approach to antimicrobial-species-specific-drug-receptor interaction has two main drawbacks. The first is the difficulty on the definition of the constants. In this work, we solve the first question estimating ${}^{0}\gamma_j(s)$ as the rate of occurrence $n_j(s)$ of the j-th atom on active molecules against a given specie with respect to the number of atoms of the j-th class in the molecules tested against the same specie $n_t(s)$. With respect to ${}^{1}\gamma_{ij}(s)$ we must taking into consideration that once the j-th atom have interacted the preferred candidates for the next interaction are such i-th atoms bound to j by a chemical bond. Both constants can be then written down as (24, 41, 43):

$${}^{0}\Gamma_j(s) = \left(\frac{n_j(s)}{n_T(s)} + 1\right) = e^{\frac{{}^{0}\theta_j(s)}{R \cdot T}} \quad (2) \qquad {}^{1}\Gamma_{ij}(s) = \left(\alpha_{ij} \cdot \frac{n_j(s)}{n_T(s)} + 1\right) = e^{\frac{{}^{1}\theta_{ij}(s)}{R \cdot T}} \quad (3)$$

Where, α_{ij} are the elements of the atom adjacency matrix, $n_j(s)$, $n_t(s)$, ${}^{0}\theta_j(s)$, and ${}^{1}\theta_{ij}(s)$ have been defined in the paragraph above, r is the universal gases constant, and t the absolute temperature. The number 1 is added to avoid scale and logarithmic function´s definition problems. The second problem relates to the description of

the interaction process at higher times $t_k > t_1$. Therefore, mm theory enables a simple calculation of the probabilities with which the drug-receptor interaction takes place in the time until the studied effect is achieved. In this work we are going to focus on drugs-microbial structure less target interaction. As depicted in figure 1, this model deals with the calculation of the probabilities ($^{k}p_{ij}$) with which any arbitrary molecular atom j-th bind to the structure less molecular receptor given that other atom i-th has been bound before; along discrete time periods t_k (k = 1, 2, 3, …); (k = 1 in grey), (k = 2 in blue) and (k = 3 in red) throughout the chemical bonding system. The procedure described here considers as states of the mm the atoms of the molecule. The method arranges all the $^{0}\theta_j(s)$ values in a vector **θ (s)** and all the $^{1}\theta_{ij}(s)$ entropies of interaction as a squared table of n x n dimension. After normalization of both the vector and the matrix we can built up the corresponding absolute initial probability vector **φ(s)** and the stochastic matrix $^{1}\mathbf{\Pi(s)}$, which has the elements $^{0}p_j(s)$ and $^{1}p_{ij}(s)$ respectively. The elements $^{0}p_j(s)$ of the above mentioned vector **φ(s)** constitutes the absolute probabilities with which the j-th atom interact with the molecular target or receptor in the species s at the initial time with respect to any atom in the molecule (24, 41, 43):

$$^{0}p_j(s) = \frac{^{0}\theta_j(s)}{\sum_{a=1}^{m} {}^{0}\theta_a(s)} = \frac{-RT \cdot \log\left(\frac{n_j(s)}{n_T(s)} + 1\right)}{\sum_{a=1}^{m} -RT \cdot \log\left(\frac{n_a}{n_T(s)} + 1\right)} = \frac{\log\left(\frac{n_j(s)}{n_T(s)} + 1\right)}{\sum_{a=1}^{m} \log\left(\frac{n_a}{n_T(s)} + 1\right)} \quad (4)$$

Where, m represents all the atoms in the molecule including the j-th, n_a is the rate of occurrence of any atom a including the j-th with value n_j. On the other hand, the matrix is called the 1-step drug-target interaction stochastic matrix. $^{1}\mathbf{\Pi(s)}$ is built too as a squared table of order n, where n represents the number of atoms in the molecule. The elements $^{1}p_{ij}(s)$ of the 1-step drug-target interaction stochastic matrix are the binding probabilities with which a j-th atom bind to a structure less molecular receptor given that other i-th atoms have been interacted before at time $t_1 = 1$ (considering $t_0 = 0$) (18, 24, 41, 43):

$$^{1}p_{ij}(s) = \frac{^{1}\theta_{ij}(s)}{\sum_{a=1}^{n} {}^{1}\theta_{ia}(s)} = \frac{\alpha_{ij} \cdot (-RT) \cdot \log\left(\frac{n_j(s)}{n(s)} + 1\right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot (-RT) \cdot \log\left(\frac{n_a(s)}{n_T(s)} + 1\right)} = \frac{\alpha_{ij} \cdot \log\left(\frac{n_j(s)}{n_T} + 1\right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot \log\left(\frac{n_j(s)}{n_T(s)} + 1\right)} \quad (5)$$

By using, **φ(s),** $^{1}\mathbf{\Pi(s)}$ and chapman-kolgomorov equations one can describe the further evolution of the system.[10-17] summing up all the atomic free energies of interaction $^{0}\theta_j(s)$ pre-multiplied by the absolute probabilities of drug-target interaction $^{a}p_k(j,s)$ one can derive the average changes in entropies $^{k}\theta_s$ of the gradual interaction between the drug and the receptor at a specific time k in a given microbial species (s) (24):

$$^{k}\theta_s = \boldsymbol{\varphi}(\mathbf{s}) \cdot {}^{k}\mathbf{\Pi}(\mathbf{s}) \cdot {}^{0}\theta(\mathbf{s}) = \boldsymbol{\varphi}(\mathbf{s}) \cdot \left[{}^{1}\mathbf{\Pi}(\mathbf{s})\right]^{k} \cdot {}^{0}\theta(\mathbf{s}) = \sum_{j=1}^{n} {}^{k}\theta_j(s) = \sum_{j=1}^{n} {}^{A}p_k(j,s) \cdot {}^{0}\theta_j(s) \quad (6)$$

Such a model is stochastic *per se* (probabilistic step-by-step atom-receptor interaction in time) but also considers molecular connectivity (the step-by-step atom union in space throughout the chemical bonding system).

3.2. Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR using the MARCH-INSIDE methodology, as defined previously, with the general formula:

$$Actv = a_0 \cdot {}^{0}\theta_s + a_1 \cdot {}^{1}\theta_s + a_2 \cdot {}^{2}\theta_s + a_3 \cdot {}^{3}\theta_s \ldots\ldots + a_k \cdot {}^{k}\theta_s + b_0 \quad (7)$$

Here, $^{k}\theta_s$ act as the microbial species specific molecule-target interaction descriptors. The calculation of these indices has been explained in supplementary material by space reasons. We selected Linear Discriminant Analysis (LDA) to fit the classification functions. The model deals with the classification of a set of compounds as active or not against different microbial species(43). A dummy variable (Actv) was used to codify the antimicrobial activity. This variable

indicates either the presence (Actv = 1) or absence (Actv = –1) of antimicrobial activity of the drug against the specific species. In equation (1), a_k represents the coefficients of the classification function and b_0 the independent term, determined by the least square method as implemented in the LDA module of the STATISTICA 6.0 software package(44). Forward stepwise was fixed as the strategy for variable selection(43). The quality of LDA models was determined by examining Wilk's U statistic, Fisher ratio (F), and the p-level (p). We also inspected the percentage of good classification and the ratios between the cases and variables in the equation and variables to be explored in order to avoid over-fitting or chance correlation. Validation of the model was corroborated by re-substitution of cases in four predicting series (43, 44).

3.3. Data set

The data set was formed by a set of marketed and/or very recently reported antiviral drugs which low reported MIC_{50} < 10 μM against different virus. The data set was conformed to more of 1100 different drugs experimentally tested against some species of a list of 40 virus. Not all drugs were tested in the literature against all listed species so we were able to collect 47 470 cases (drug/species pairs) instead of 1100 x 40 cases.

Entropy based mt-QSAR equation is able to predict the biological activity of antiviral drugs in more general situations than the traditional QSAR models; which the major limitation is predict the biological activity of drugs against only one viral species. The present model with a very large data set improves significantly the previous QSAR models and may help to perform MOOP of drug activity against different viral species. This mt-QSAR methodology improves models using entropy as a molecular descriptor that allow predicting antiviral activity of any organic compound against a very large diversity of viral pathogens.

4. Conclusions

References and Notes

1. Prado-Prado J, Martinez de la Vega O, Uriarte E, Ubeira FM, Chou K-C, González-Díaz H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug–drug complex networks. Bioorg Med Chem. 2008;doi:10.1016/j.bmc.2008.11.075.

2. Fratev F, Benfenati E. 3D-QSAR and molecular mechanics study for the differences in the azole activity against yeastlike and filamentous fungi and their relation to P450DM inhibition. 1. 3-substituted-4(3H)-quinazolinones. Journal of chemical information and modeling. 2005 May-Jun;45(3):634-44.

3. Cronin MT, Aptula AO, Dearden JC, Duffy JC, Netzeva TI, Patel H, et al. Structure-based classification of antibacterial activity. J Chem Inf Comput Sci. 2002 Jul-Aug;42(4):869-78.

4. Marrero-Ponce Y, Castillo-Garit JA, Olazabal E, Serrano HS, Morales A, Castanedo N, et al. Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic. Bioorg Med Chem. 2005 Feb 15;13(4):1005-20.

5. Marrero-Ponce Y, Medina-Marrero R, Torrens F, Martinez Y, Romero-Zaldivar V, Castro EA. Atom, atom-type, and total nonstochastic and stochastic quadratic fingerprints: a promising approach for modeling of antibacterial activity. Bioorg Med Chem. 2005 Apr 15;13(8):2881-99.

6. Marrero-Ponce Y, Meneses-Marcel A, Castillo-Garit JA, Machado-Tugores Y, Escario JA, Barrio AG, et al. Predicting antitrichomonal activity: a computational screening using atom-based bilinear indices and experimental proofs. Bioorg Med Chem. 2006 Oct 1;14(19):6502-24.
7. Montero-Torres A, Vega MC, Marrero-Ponce Y, Rolon M, Gomez-Barrio A, Escario JA, et al. A novel non-stochastic quadratic fingerprints-based approach for the 'in silico' discovery of new antitrypanosomal compounds. Bioorg Med Chem. 2005 Nov 15;13(22):6264-75.
8. Meneses-Marcel A, Marrero-Ponce Y, Machado-Tugores Y, Montero-Torres A, Pereira DM, Escario JA, et al. A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: outcomes of in silico studies supported by experimental results. Bioorg Med Chem Lett. 2005 Sep 1;15(17):3838-43.
9. Vega MC, Montero-Torres A, Marrero-Ponce Y, Rolon M, Gomez-Barrio A, Escario JA, et al. New ligand-based approach for the discovery of antitrypanosomal compounds. Bioorg Med Chem Lett. 2006 Apr 1;16(7):1898-904.
10. Marrero-Ponce Y, Meneses-Marcel A, Rivera-Borroto OM, Garcia-Domenech R, De Julian-Ortiz JV, Montero A, et al. Bond-based linear indices in QSAR: computational discovery of novel anti-trichomonal compounds. J Comput Aided Mol Des. 2008 Aug;22(8):523-40.
11. Garcia-Domenech R, Galvez J, de Julian-Ortiz JV, Pogliani L. Some new trends in chemical graph theory. Chem Rev. 2008 Mar;108(3):1127-69.
12. Marrero-Ponce Y, Khan MT, Casanola-Martin GM, Ather A, Sultankhodzhaev MN, Garcia-Domenech R, et al. Bond-based 2D TOMOCOMD-CARDD approach for drug discovery: aiding decision-making in 'in silico' selection of new lead tyrosinase inhibitors. J Comput Aided Mol Des. 2007 Apr;21(4):167-88.
13. Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, Garcia-Domenech R, Munoz C, Guna R, et al. Search of chemical scaffolds for novel antituberculosis agents. J Biomol Screen. 2005 Apr;10(3):206-14.
14. Garcia-Garcia A, Galvez J, de Julian-Ortiz JV, Garcia-Domenech R, Munoz C, Guna R, et al. New agents active against Mycobacterium avium complex selected by molecular topology: a virtual screening method. J Antimicrob Chemother. 2004 Jan;53(1):65-73.
15. Meneses-Marcel A, Rivera-Borroto OM, Marrero-Ponce Y, Montero A, Machado Tugores Y, Escario JA, et al. New antitrichomonal drug-like chemicals selected by bond (edge)-based TOMOCOMD-CARDD descriptors. J Biomol Screen. 2008 Sep;13(8):785-94.
16. Vilar S, Santana L, Uriarte E. Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. J Med Chem. 2006;49(3):1118-24.
17. Cruz-Monteagudo M, Borges F, Cordeiro MN, Cagide Fajin JL, Morell C, Ruiz RM, et al. Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries. J Comb Chem. 2008 Nov-Dec;10(6):897-913.
18. González-Díaz H, Prado-Prado FJ, Santana L, Uriarte E. Unify QSAR approach to antimicrobials. Part 1: Predicting antifungal activity against different species. Bioorg Med Chem. 2006 Jun 5;14 5973–80.

19. González-Díaz H, Prado-Prado F. Unified QSAR and Network-Based Computational Chemistry Approach to Antimicrobials, Part 1: Multispecies Activity Models for Antifungals. J Comput Chem. 2008;29:656-7.
20. Todeschini R, Consonni V. Handbook of Molecular Descriptors. Mannhold R, Kubinyi H, Timmerman H, editors: Wiley-VCH; 2002.
21. Gonzalez-Diaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. Curr Top Med Chem. 2008;8(18):1676-90.
22. González-Díaz H, Cabrera-Pérez MA, Agüero-Chapín G, Cruz-Monteagudo M, Castañedo-Cancio N, del Río MA, et al. Multi-target QSPR assemble of a Complex Network for the distribution of chemicals to biphasic systems and biological tissues. Chemometrics Intellig Lab Syst. 2008;94:160-5.
23. Cruz-Monteagudo M, González-Díaz H, Agüero-Chapin G, Santana L, Borges F, Domínguez RE, et al. Computational Chemistry Development of a Unified Free Energy Markov Model for the Distribution of 1300 Chemicals to 38 Different Environmental or Biological Systems. J Comput Chem. 2007; 28:1909-22.
24. González-Díaz H, Aguero G, Cabrera MA, Molina R, Santana L, Uriarte E, et al. Unified Markov thermodynamics based on stochastic forms to classify drugs considering molecular structure, partition system, and biological species: distribution of the antimicrobial G1 on rat tissues. Bioorg Med Chem Lett. 2005 Feb 1;15(3):551-7.
25. González-Díaz H, Uriarte E. Proteins QSAR with Markov average electrostatic potentials. Bioorg Med Chem Lett. 2005 Nov 15;15(22):5088-94.
26. Saiz-Urra L, González-Díaz H, Uriarte E. Proteins Markovian 3D-QSAR with spherically-truncated average electrostatic potentials. Bioorg Med Chem. 2005 Jun 1;13(11):3641-7.
27. Ferino G, Delogu G, Podda G, Uriarte E, González-Díaz H. Quantitative Proteome-Disease Relationships (QPDRs) in Clinical Chemistry: Prediction of Prostate Cancer with Spectral Moments of PSA/MS Star Networks. In: Mitchem BHaS, Ch.L., editor. Clinical Chemistry Research (ISBN: 978-1-60692-517-1). NY: Nova Science Publisher; 2009.
28. Concu R, Podda G, Uriarte E, González-Díaz H. A New Computational Chemistry & Complex Networks approach to Structure-Function and Similarity Relationships in Protein Enzymes. In: Collett CTaR, C.D., editor. Handbook of Computational Chemistry Research: Nova Science Publishers 2009.
29. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. Proteomics. 2008;8:750-78.
30. González-Díaz H, Vilar S, Santana L, Uriarte E. Medicinal Chemistry and Bioinformatics – Current Trends in Drugs Discovery with Networks Topological Indices. Curr Top Med Chem. 2007;7(10):1025-39.
31. Gonzalez-Diaz H, Saiz-Urra L, Molina R, Santana L, Uriarte E. A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. Journal of proteome research. 2007 Feb;6(2):904-8.
32. González-Díaz H, Marrero Y, Hernandez I, Bastida I, Tenorio E, Nasco O, et al. 3D-MEDNEs: an alternative "in silico" technique for chemical research in toxicology. 1. prediction of chemically induced agranulocytosis. Chem Res Toxicol. 2003 Oct;16(10):1318-27.

33. González-Díaz H, Molina R, Uriarte E. Markov entropy backbone electrostatic descriptors for predicting proteins biological activity. Bioorg Med Chem Lett. 2004 Sep 20;14(18):4691-5.
34. Graham DJ. Information Content in Organic Molecules: Brownian Processing at Low Levels. Journal of chemical information and modeling. 2007;47(2):376-89.
35. Graham DJ, Schacht D. Base Information Content in Organic Molecular Formulae. J Chem Inf Comput Sci. 2000;40:942.
36. Graham DJ. Information Content in Organic Molecules: Structure Considerations Based on Integer Statistics. J Chem Inf Comput Sci. 2002;42:215.
37. Graham DJ, Malarkey C, Schulmerich MV. Information Content in Organic Molecules: Quantification and Statistical Structure via Brownian Processing. . J Chem Inf Comput Sci. 2004;44(1601).
38. Graham DJ, Schulmerich MV. Information Content in Organic Molecules: Reaction Pathway Analysis via Brownian Processing. J Chem Inf Comput Sci. 2004;44(1612).
39. Graham DJ. Information Content and Organic Molecules: Aggregation States and Solvent Effects. Journal of chemical information and modeling. 2005;45(1223).
40. Gonzalez-Diaz H, Tenorio E, Castanedo N, Santana L, Uriarte E. 3D QSAR Markov model for drug-induced eosinophilia--theoretical prediction and preliminary experimental assay of the antimicrobial drug G1. Bioorg Med Chem. 2005 Mar 1;13(5):1523-30.
41. González-Díaz H, Cruz-Monteagudo M, Molina R, Tenorio E, Uriarte E. Predicting multiple drugs side effects with a general drug-target interaction thermodynamic Markov model. Bioorg Med Chem. 2005 Feb 15;13(4):1119-29.
42. Cruz-Monteagudo M, González-Díaz H. Unified drug-target interaction thermodynamic Markov model using stochastic entropies to predict multiple drugs side effects. Eur J Med Chem. 2005 Oct;40(10):1030-41.
43. Van Waterbeemd H. Discriminant Analysis for Activity Prediction. In: Van Waterbeemd H, editor. Chemometric methods in molecular design. New York: Wiley-VCH; 1995. p. 265-82.
44. StatSoft.Inc. STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft, Inc. 6.0 ed2002. p. STATISTICA (data analysis software system), version 6.0, www.statsoft.com.Statsoft.