



Transworld Research Network  
37/661 (2), Fort P.O.  
Trivandrum-695 023  
Kerala, India

Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks, 2010: 1-14 ISBN: 978-81-7895-489-9  
Editors: Humberto González-Díaz and Cristian Robert Munteanu

# 1. Multi-target QSAR of antiviral drugs

Francisco J. Prado-Prado<sup>1</sup> and Humberto González-Díaz<sup>2,\*</sup>

<sup>1</sup>*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela (USC), Santiago de Compostela, 15782, Spain*

<sup>2</sup>*Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Santiago de Compostela (USC), Santiago de Compostela, 15782, Spain*

**Abstract.** Graph theory and have applications at molecular level to describe drug-virus action pairs in antiviral medicinal chemistry research. With graph parameters called Topological Indices (TIs), we can search of Quantitative Structure-Activity Relationship (QSAR) models for prediction and discovery of antimicrobial drugs. In this work, we decided to test the potentialities of one of the classes of TIs. For this test we selected the class of TIs called the node absolute probabilities  $\pi_k(i)$  that can be calculated with the method MARCH-INSIDE based on Markov models. We report a new QSAR model: that can be used to predict drug activity against different viral strains. The model correctly classifies 428 out of 533 cases (80.30%) and 481 out of 596 non-active compounds/virus cases (80.7%). Using this QSAR model we were able to reconstruct a large complex network of observed effective drug-virus pairs with a total Accuracy = 89.8 %. The work opens new directions in the generalization of TIs to develop QSAR/QSPR models for predicting relevant information of systems at different structural levels.

Correspondence/Reprint request: Dr. González-Díaz H, Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Santiago de Compostela (USC), Santiago de Compostela, 15782, Spain  
E-mail: gonzalezdiazh@yahoo.es

## 1. Introduction

Graph theory and Complex Network analysis tools with many applications on the search of rational approaches for antimicrobial drugs discovery. Actually, there are many pathogen microbial species with very different antimicrobial drugs susceptibility. In particular, viral pathogens species are responsible of many human diseases. Examples of diseases caused by viruses include the common cold, which is caused by any one of a variety of related viruses; smallpox; AIDS, which is caused by HIV; and cold sores, which are caused by *herpes simplex*. Other connections are being studied such as the connection of Human Herpes Virus (HHV) in organic neurological diseases such as multiple sclerosis and chronic fatigue syndrome. Recently, it has been shown that cervical cancer is caused at least partly by papillomavirus (which causes papillomas, or warts), representing the first significant evidence in humans for a link between cancer and an infective agent. There is current controversy over whether borna virus, previously thought of primarily as the causative agent of neurological disease in horses, could be responsible for psychiatric illness in humans. The relative ability of viruses to cause disease is described in terms of virulence. This very high number of drug-species combinations may be investigated using networks to group or cluster drugs with similar multi-species activity profile and possibly mechanism of action [1].

In fact, the applications information-mining techniques based on graphs or networks do not limit only to drug discovery. We can use different classes of graph and networks such as: Molecular graphs to describe the structure of the antimicrobial drugs, Artificial Neural Networks (ANN) [2-13] for dataset mining. We can use also Hasse graphs to depict relationships within genetic code [14-20], or Interaction and/or Co-expression networks to represent relationships between proteins, genes, or RNAs [21-35]. Specifically, co-expression networks can be constructed by measuring the expression of pairs or genes in different tissues [36-40]. Similarly, protein networks study experimentally or theoretically established protein-protein interactions [21, 41]. In co-expression networks two RNAs are connected (supposed to be involved in common mechanism of regulation) if the levels of both RNAs for different tissues strongly correlate [42]. We proposed to use the same network approach to study multi-species antimicrobials drug action. The antimicrobial drug plays the role of the RNA molecule and the drug activity against different species activity play the role of RNA level of expression in different tissues. In the co-expression network, we need to measure each RNA tissue profile if we do not have a computational approach to predict it [36, 43].

Disappointingly, QSPR studies are generally focused on the study of limited properties of small molecules. For example, in the case of antimicrobial many QSPR models predict only structurally parent compounds acting against one single microbial species [44-46]. Actually, there are more than 1 600 molecular descriptors that may be in principle generalized and used to solve the QSPR problem in small molecules [47]. In any case, any of these indices have been extended yet to encode information additional to chemical structure [48-50]. In addition, have been reported QSPR-like models based on graph and networks TIs for proteins, DNA, or RNA structures and some authors have extended the applications of TIs to whole blood proteomes, protein-protein interaction networks or even tissues [41, 51, 52]. Anyhow, the applications of TIs in QSPR research are far to cover all the potentialities of TIs and new gateways in molecular, biological, or even social QSPR models are still waiting to be opened. On this line of thinking, our group has introduced a Markov model (MM) method named MARCH-INSIDE: Markovian Chemicals In Silico Design. MARCH-INSIDE generate TIs in the form of matrix invariants such as stochastic entropies, spectral moments, or absolute probabilities for the study of molecular properties [53-57]. Recently the method has been renamed as MARCH-INSIDE 2.0: Markov Chain Invariants for Network Simultation & Design, in order to give a more clear idea of the unexplored potentialities [58].

In this work, we decided to test the potentialities at different structural levels of one of the classes of TIs calculated by MARCH-INSIDE. For this test we selected the class of TIs called the node absolute probabilities  $\pi_k(i)$ . The  $\pi_k(i)$  values represent the absolute probability of reaching node  $i$  after a walk of length  $k$  moving from any node in the network. We calculate the  $\pi_k(i)$  based on a Markov matrix associated to a graph or network. These TIs differ from other MARCH-INSIDE TIs because they are useful only to describe a node or, if we sum several  $\pi_k(i)$  values, we can describe a collection of nodes (atoms, aminoacids, a group of electric plants, a social subgroup...); which form part of larger network systems (molecule, protein, US Electric power system, Society...). It happens because the sum of all  $\pi_k(j)$  values for the whole system is always equal to one for any system a do not give structural information. Consequently, the  $\pi_k(i)$  values may be consider as local node TIs. We commonly have known this class of TIs as networks nodes Centralities. Several node centralities have been defined before and the software CentiBin calculate some of the more used [59]. However, the definition of new Centralities is an active field of research and new centralities have been recently introduced such as sub-graph centrality

introduced by Estrada [60]. Certainly, the  $\pi_k(j)$  values were used in the past by our group [61, 62] but ever at the molecular level only and never for non-molecular problems. In order to both confirm the potentials applications of  $\pi_k(j)$  values at the molecular level and extend these applications beyond traditional frontiers, we are going to develop here a new QSPR models. This QSPR model based on  $\pi_k(j)$  values can be used to predict antiviral drugs activity against multiple virus species.

## 2. Materials and methods

### Multi-target probability centrality for atoms in molecules

By using, Chapman-Kolmogorov equations we can calculate multi-target  ${}^kC_{\pi,s}(j)$  values referred to atoms (nodes) in molecular graphs. As was mentioned above multi-target here means that we obtain different  ${}^kC_{\pi,s}(j)$  values for the same atom in the same molecule when the molecular target (bacteria, virus, parasite, receptor, enzyme, etc.) change. First, we have to calculate the absolute probabilities  ${}^s p_k(j)$  for the interaction in many step of different  $j$ -th atoms with the specific target. Here targets are only different microbial species ( $s$ ). In this sense, we insert the superscript  $s$  in the symbol of the centrality. These values can be determined as the elements of the vectors  ${}^k\pi(s)$ . These vectors are elements of a Markov chain based on the stochastic matrix  ${}^1\Pi$ , which describes probabilities of interaction  ${}^s p_1(i,j)$  of the  $j$ -th atom given that previously other  $i$ -th atom has interacted with the target.

The specificity for one target is given using target specific weights in the definition of the elements of the matrix  ${}^1\Pi$ . The theoretic foundations of the method have been given in previous works, so we do not detail it here but refer the reader to these works [63, 64]. After that, the entropy centrality is very ease to calculate applying the Shannon's formula to each element  ${}^s p_k(j)$  of the vectors  ${}^k\pi(s)$  and obtain the entropy centrality measures  ${}^kC_{\pi,s}(j)$ . As in the example 1 we can sum the  ${}^kC_{\pi,s}(j)$  values for specific atom sets (AS), or the same groups of nodes, to create local molecular descriptors for the drug-target interaction. Herein the AS used were: halogens (X), insaturated carbons ( $C_{ins}$ ), saturated carbons ( $C_{sat}$ ), heteroatom (Het), and hydrogen atoms bound to heteroatom (H-Het). The corresponding symbols of the local entropy centrality for these AS are:  ${}^kC_{\pi,s}(X)$ ,  ${}^kC_{\pi,s}(C_{ins})$ ,  ${}^kC_{\pi,s}(C_{sat})$ ,  ${}^kC_{\pi,s}(Het)$ ,  ${}^kC_{\pi,s}(H-Het)$  and  ${}^kC_{\pi,s}(T)$ . In this study, we calculated the first six classes of entropy centrality ( $k = 0$  to  $5$ ) for the 5 AS in total  $6 \cdot 5 = 30$  molecular local centralities for each drug [64]. The theoretic foundations of the method have

been given in previous works, so we do not detail it here but refer the reader to these works [64, 65]:

$${}^k\pi_s = {}^0\pi(s) \cdot [{}^1\Pi(s)]^k = [\pi_0(1,s), \pi_0(2,s), \pi_0(3,s), \dots, \pi_0(n,s)] \cdot \begin{bmatrix} {}^1\pi_{11}(s) & {}^1\pi_{12}(s) & \dots & \dots & {}^1\pi_{1n}(s) \\ {}^1\pi_{21}(s) & {}^1\pi_{22}(s) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ {}^1\pi_{n1}(s) & \dots & \dots & \dots & {}^1\pi_{nn}(s) \end{bmatrix}^k \quad (1)$$

The  ${}^A\pi_k(j,s)$  can be summed for specific sets of atoms (AS) to create local molecular descriptors for the drug-target interaction. Herein, the AS used were the following: halogens (X), unsaturated carbons ( $C_{ins}$ ), saturated carbons ( $C_{sat}$ ), heteroatoms (Het), and hydrogens bound to heteroatoms (H-Het). The corresponding symbols of the local absolute probabilities for these AS are:  ${}^A\pi_k(X,s)$ ,  ${}^A\pi_k(C_{ins},s)$ ,  ${}^A\pi_k(C_{sat},s)$ ,  ${}^A\pi_k(Het,s)$ ,  ${}^A\pi_k(H-Het,s)$ . In this study, we calculated the first six classes of probabilities ( $k = 0$  to 5) for the 5 AS in total  $6 \cdot 5 = 30$  molecular descriptors [64].

## Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSPR using the MARCH-INSIDE methodology, as defined previously, with the general formulae [62]:

$$Actv = {}^c b_0 \cdot {}^A\pi_0(C,s) + {}^c b_1 \cdot {}^A\pi_1(C,s) + {}^c b_2 \cdot {}^A\pi_2(C,s) + {}^c b_3 \cdot {}^A\pi_3(C,s) + \dots + {}^c b_k \cdot {}^A\pi_4(C,s) + b \quad (2)$$

Here, the absolute probabilities  ${}^A\pi_k(C,s)$  play the role of molecule-target interaction descriptors for specific microbial species. We selected Linear Discriminant Analysis (LDA) [66] to fit the classification functions. The model deals with the classification of a set of compounds as active or not against different microbial species. 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 microbe species in question. In equation (8),  $b_k$  represents the coefficients of the classification function, determined by the LDA module of the STATISTICA 6.0 software package [67] using forward stepwise strategy for variable selection. 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. Validation of the model was corroborated with external prediction series.

## 2. Results and discussion

### QSAR model for antiviral molecules

One of the main advantages of the present approach is that the generalized parameters  ${}^kC_{\pi,s}(j)$  fit on more large and complex databases than the previous ones. This work introduces by the first time a single linear mt-QSAR equation model in order to predict the antibacterial activity of drugs against different species. The data set used here was established by a set of marketed and/or very recently reported antiviral drugs with low reported  $MIC_{50} < 10 \mu M$  against different virus strains. The data set have different drugs experimentally tested against some species of a list of more than 40 viruses. We do not found in the literature the experimental values for all compounds against all listed virus species so we were able to collect 950 cases (drug/virus pairs). The names or codes and activity for all compounds as well as the references used to collect it have been saved into in a supplementary material file, available upon author request.

$$\begin{aligned}
 S(DVP) = & -16.47 \cdot {}^1C_{\pi}(C_{C_{sp} \& Sp_2}) + 17.34 \cdot {}^2C_{\pi}(C_{C_{sp} \& Sp_2}) - 7.05 \cdot {}^0C_{\pi}(C_{C_{sp_3}}) \\
 & - 3.06 \cdot {}^0C_{\pi}(H - Het) - 9.69 \cdot {}^0C_{\pi}(C_{C_{sp} \& Sp_2}) - 0.91 \cdot {}^5C_{\pi}(H - Het) + 0.18 \\
 U = & 0.48 \quad F = 94.767 \quad p < 0.001
 \end{aligned}
 \tag{3}$$

$S(DVP)$ , the output of the model, is a real value variable (not probability) that scores Drug-Virus activity specificity. In this equation,  ${}^kC_{\pi,s}(j)$  where summed for the totality (T) of the atoms in the molecule or for specific atom sets (AS) as we referred above. These collections are atoms with a common characteristic as for instance are: saturated Carbon atoms ( $C_{sat}$ ), hydrogen atoms linked to the hetero-atoms ( $H-Het$ ). The model correctly classifies 428 out of 533 cases (80.30%) and 481 out of 596 non-active compounds/virus cases (80.7%). Overall training Accuracy was 80.5%.

### Construction of a drug-virus network

Next, we used the outputs of the mt-QSAR as inputs to constructs the first CN for antiviral drugs and species based on  ${}^kC_{\pi,s}(j)$  values. In previous works, we constructed by the first time mt-QSAR models accounting for pairs of anti-parasite [68, 69] antifungal [61, 64] or antiviral drugs [70] with similar/dissimilar multi-species activity profile and represented it as large

networks. In this work, we have to manage with a very high number of possible Drug-Virus Pairs (DVPs). These DVPs may be investigated using CNs to regroup or cluster drugs with similar multi-bacterial affinity profile. In DVP-CN, the DVPs are nodes interconnected by the edges if they have similar drug-bacteria activity. We need to measure the activity of the drug on different bacterial if we cannot predict it. We propose to construct here, by the first time, a DVP-CN taking into consideration only the number of DVPs predicted by the mt-QSAR model based on  ${}^kC_{\pi}(j)$  values. In order to construct this CN we have given the following steps:

1. First, we calculated two types of activity Z-scores (drug score and bacteria score) for both experimental and QSAR-predicted values:

$$z_{obs}(d) = \frac{\log MIC_i}{\log MIC_{max}} \quad (4)$$

$$z_{pred}(d) = p(-)_i \quad (5)$$

where  $d$  is the score affinity, either observed score ( $s_{obs}$ ) or predicted score ( $s_{pred}$ ).  $s_{obs}$  were calculated on the experimental data ( $IC_{50}$ ). We calculated the  $s_{pred}$  of each one of the 950 compounds with all the studied viral strains here by substituting the molecular descriptors into the QSAR equation using the Microsoft Excel application [71]. Mean is the average either of  $s_{obs}$  or  $s_{pred}$  for the DBP. We calculate the distance matrix between all DBP using a Euclidean distance:

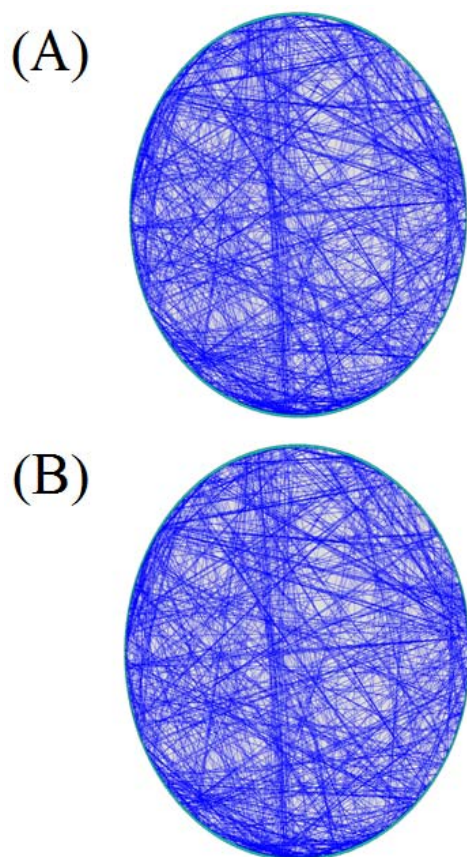
$${}^{obs}d_{ij} = \frac{1}{\log MIC_{max}} \cdot |\log MIC_i - \log MIC_j| \quad (6)$$

$${}^{pred}d_{ij} = |p(-)_i - p(-)_j| \quad (7)$$

2. Using Microsoft Excel [71] again, we transformed the DBPs distance matrices derived into Boolean matrices. The elements of this type of matrix are equal to 1 if two DBPs have a Euclidean distance  $d_{ij} <$  a cut-off value. We explore the threshold values in a range from  $\log IC_{s_{obs}}$  until  $IC_{s_{pred}}$  trying to obtain average DBP node degree equal to 1 and minimizing the number of disconnected DBPs [61].

3. The Boolean matrix was saved as a .txt format file. After, renamed the .txt file as a .mat file we read it with the software CentiBin [59] and renamed the .mat file as a .net file and read it with Pajek software. Using Pajek, we can not only represent the network but also highlight all DBP (nodes) connected to a specific DBP and calculating connectivity parameters [72].

4. Last, we compared the observed and predicted DBP-DBP networks pair-to-pair calculating the total Accuracy values. The mt-QSAR predicted correctly 751 334 drug-virus similar/dissimilar pairs out of 836 310 pairs. Thus, the mt-QSAR based on  ${}^kC_{\pi}(j)$  values predicts the real CN (based on  $MIC_{50}$  values) with an Accuracy of 89.8% (number of similarity/dissimilarity relationships present in both CNs). The cut-off values that maximize the similarity between the two CN were used in order to decided if to nodes are connected or not:  $p(\text{real}) = (\log MIC_{50} / \log MIC_{\max}) = 0.03$  for the observed CN and  $p(\text{predicted}) = 0.017$ . In **Figure 1** we illustrate both, the CN observed and the CN predicted with the mt-QSAR model.



**Figure 1.** Drug-Virus Complex Networks: (A) observed network and (B) predicted network.

**Table 1.** Results of the Leave-Specie-Out validation for the mt-QSAR model.

Virus	# Cases	LSO (%)	Accuracy (%)
Cowpox Virus	24	91.7	87.4
Epstein-Barr Virus	16	87.5	87.4
Hepatitis B Virus	36	88.9	87.3
Hepatitis C Virus	114	89.5	86.5
Herpes Simplex Virus 1	55	96.4	87.0
HIV-1	409	98.5	87.3
Human Cytomegalovirus	124	99.2	87.0
Human Papillomavirus	37	81.1	87.4
Influenza Virus	38	89.5	87.2
Vaccinia Virus	51	94.1	86.9
Varicella-Zoster Virus	55	100	87.0

## Acknowledgements

Prado-Prado, F. acknowledges financial support from Xunta the Galicia and European Social Fund (ESF) for a one-year post-doctoral position (Research Project IN89A 2008/75-0). González-Díaz, H. acknowledges financial support of Program Isidro Parga Pondal and one-year post-doctoral position (Research Project IN89A 2008/117-0), both funded by Xunta the Galicia and ESF.

## References

1. Prado-Prado FJ, de la Vega OM, Uriarte E, Ubeira FM, Chou KC, 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.* 2009;17:569–75.
2. Katritzky AR, Dobchev DA, Fara DC, Karelson M. QSAR studies on 1-phenylbenzimidazoles as inhibitors of the platelet-derived growth factor. *Bioorg Med Chem.* 2005 Dec 15;13(24):6598-608.
3. Basak SC, Grunwald GD, Gute BD, Balasubramanian K, Opitz D. Use of statistical and neural net approaches in predicting toxicity of chemicals. *J Chem Inf Comput Sci.* 2000 Jul-Aug;40(4):885-90.
4. Baskin, II, Ait AO, Halberstam NM, Palyulin VA, Zefirov NS. An approach to the interpretation of backpropagation neural network models in QSAR studies. *SAR QSAR Environ Res.* 2002 Mar;13(1):35-41.

5. Benigni R, Giuliani A. Quantitative structure-activity relationship (QSAR) studies in genetic toxicology: mathematical models and the "biological activity" term of the relationship. *Mutat Res.* 1994 Apr 15;306(2):181-6.
6. Fernandez M, Caballero J. Bayesian-regularized genetic neural networks applied to the modeling of non-peptide antagonists for the human luteinizing hormone-releasing hormone receptor. *J Mol Graph Model.* 2006 Feb 28.
7. Fernandez M, Caballero J, Tundidor-Camba A. Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors. *Bioorg Med Chem.* 2006 Jun 15;14(12):4137-50.
8. Fernandez M, Tundidor-Camba A, Caballero J. Modeling of cyclin-dependent kinase inhibition by 1H-pyrazolo[3,4-d]pyrimidine derivatives using artificial neural network ensembles. *Journal of chemical information and modeling.* 2005 Nov-Dec;45(6):1884-95.
9. Fernandez M, Caballero J, Helguera AM, Castro EA, Gonzalez MP. Quantitative structure-activity relationship to predict differential inhibition of aldose reductase by flavonoid compounds. *Bioorg Med Chem.* 2005 May 2;13(9):3269-77.
10. Caballero J, Garriga M, Fernandez M. Genetic neural network modeling of the selective inhibition of the intermediate-conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channel by some triarylmethanes using topological charge indexes descriptors. *J Comput Aided Mol Des.* 2005 Nov;19(11):771-89.
11. Caballero J, Fernandez M. Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks. *J Mol Model* 2006 Jan;12(2):168-81.
12. Caballero J, Fernandez M. Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks. *J Mol Model (Online).* 2005 Oct 21:1-14.
13. Caballero J, Fernandez L, Abreu JI, Fernandez M. Amino Acid Sequence Autocorrelation vectors and ensembles of Bayesian-Regularized Genetic Neural Networks for prediction of conformational stability of human lysozyme mutants. *Journal of chemical information and modeling.* 2006 May-Jun;46(3):1255-68.
14. Sanchez R, Grau R. A genetic code Boolean structure. II. The genetic information system as a Boolean information system. *Bulletin of mathematical biology.* 2005 Sep;67(5):1017-29.
15. Sanchez R, Grau R. A novel algebraic structure of the genetic code over the Galois field of four DNA bases. *Acta biotheoretica.* 2006;54(1):27-42.
16. Sanchez R, Grau R, Morgado E. A novel Lie algebra of the genetic code over the Galois field of four DNA bases. *Mathematical biosciences.* 2006 Jul;202(1):156-74.
17. Sanchez R, Morgado E, Grau R. A genetic code Boolean structure. I. The meaning of Boolean deductions. *Bulletin of mathematical biology.* 2005 Jan;67(1):1-14.
18. Sanchez R, Morgado E, Grau R. Gene algebra from a genetic code algebraic structure. *Journal of mathematical biology.* 2005 Oct;51(4):431-57.
19. Bashford JD, Jarvis PD. The genetic code as a periodic table: algebraic aspects. *Bio Systems.* 2000 Aug-Sep;57(3):147-61.

20. Beland P, Allen TF. The origin and evolution of the genetic code. *Journal of theoretical biology*. 1994 Oct 21;170(4):359-65.
21. Zhang Z, Grigorov MG. Similarity networks of protein binding sites. *Proteins*. 2006 Feb 1;62(2):470-8.
22. Voy BH, Scharff JA, Perkins AD, Saxton AM, Borate B, Chesler EJ, et al. Extracting gene networks for low-dose radiation using graph theoretical algorithms. *PLoS Comput Biol*. 2006 Jul 21;2(7):e89.
23. Tanaka T, Ikeo K, Gojobori T. Evolution of metabolic networks by gain and loss of enzymatic reaction in eukaryotes. *Gene*. 2006 Jan 3;365:88-94.
24. Sun S, Zhao Y, Jiao Y, Yin Y, Cai L, Zhang Y, et al. Faster and more accurate global protein function assignment from protein interaction networks using the MFGO algorithm. *FEBS Lett*. 2006 Mar 20;580(7):1891-6.
25. Gelfand MS. Evolution of transcriptional regulatory networks in microbial genomes. *Curr Opin Struct Biol*. 2006 Jun;16(3):420-9.
26. Barabasi AL. Sociology. Network theory--the emergence of the creative enterprise. *Science*. 2005 Apr 29;308(5722):639-41.
27. Barabasi AL, Freeh VW, Jeong H, Brockman JB. Parasitic computing. *Nature*. 2001 Aug 30;412(6850):894-7.
28. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature reviews*. 2004 Feb;5(2):101-13.
29. de Menezes MA, Barabasi AL. Fluctuations in network dynamics. *Physical review letters*. 2004 Jan 16;92(2):028701.
30. Dezsó Z, Barabasi AL. Halting viruses in scale-free networks. *Physical review*. 2002 May;65(5 Pt 2):055103.
31. Dezsó Z, Oltvai ZN, Barabasi AL. Bioinformatics analysis of experimentally determined protein complexes in the yeast *Saccharomyces cerevisiae*. *Genome research*. 2003 Nov;13(11):2450-4.
32. Dobrin R, Beg QK, Barabasi AL, Oltvai ZN. Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. *BMC bioinformatics*. 2004 Jan 30;5:10.
33. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001 May 3;411(6833):41-2.
34. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. *Nature*. 2000 Oct 5;407(6804):651-4.
35. Oliveira JG, Barabasi AL. Human dynamics: Darwin and Einstein correspondence patterns. *Nature*. 2005 Oct 27;437(7063):1251.
36. Yu X, Lin J, Masuda T, Esumi N, Zack DJ, Qian J. Genome-wide prediction and characterization of interactions between transcription factors in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2006;34(3):917-27.
37. Carter SL, Brechbuhler CM, Griffin M, Bond AT. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics*. 2004 Sep 22;20(14):2242-50.
38. Carlson MR, Zhang B, Fang Z, Mischel PS, Horvath S, Nelson SF. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics*. 2006;7:40.

39. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*. 2005;4:Article17.
40. Reverter A, Barris W, McWilliam S, Byrne KA, Wang YH, Tan SH, et al. Validation of alternative methods of data normalization in gene co-expression studies. *Bioinformatics*. 2005 Apr 1;21(7):1112-20.
41. Estrada E. Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics*. 2006 Jan;6(1):35-40.
42. Yu X, Lin J, Zack DJ, Qian J. Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Res*. 2006;34(17):4925-36.
43. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*. 2006;7 Suppl 1:S7.
44. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*: Wiley-VCH 2002.
45. Otzen T, Wempe EG, Kunz B, Bartels R, Lehwerk-Yvetot G, Hansel W, et al. Folate-synthesizing enzyme system as target for development of inhibitors and inhibitor combinations against *Candida albicans*-synthesis and biological activity of new 2,4-diaminopyrimidines and 4'-substituted 4-aminodiphenyl sulfones. *J Med Chem*. 2004 Jan 1;47(1):240-53.
46. 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.
47. Kubinyi H. Quantitative structure-activity relationships (QSAR) and molecular modelling in cancer research. *J Cancer Res Clin Oncol*. 1990;116(6):529-37.
48. 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.
49. 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.
50. Marrero-Ponce Y, Montero-Torres A, Zaldivar CR, Veitia MI, Perez MM, Sanchez RN. Non-stochastic and stochastic linear indices of the 'molecular pseudograph's atom adjacency matrix': application to 'in silico' studies for the rational discovery of new antimalarial compounds. *Bioorg Med Chem*. 2005 Feb 15;13(4):1293-304.
51. González-Díaz H, González-Díaz Y, Santana L, Ubeira FM, Uriarte E. Proteomics, networks and connectivity indices. *Proteomics*. 2008;8:750-78.
52. Manke T, Demetrius L, Vingron M. Lethality and entropy of protein interaction networks. *Genome Inform Ser*. 2005;16(1):159-63.

53. González-Díaz H, Olazabal E, Castanedo N, Sanchez IH, Morales A, Serrano HS, et al. Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer aided molecular design II: experimental and theoretical assessment of a novel method for virtual screening of fasciolicides. *J Mol Model (Online)*. 2002 Aug;8(8):237-45.
54. González-Díaz H, Gia O, Uriarte E, Hernadez I, Ramos R, Chaviano M, et al. Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design I: discovery of anticancer compounds. *J Mol Model*. 2003 Dec;9(6):395-407.
55. 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., ed. *Clinical Chemistry Research (ISBN: 978-1-60692-517-1)*. NY: Nova Science Publisher 2009.
56. 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., ed. *Handbook of Computational Chemistry Research*: Nova Science Publishers 2009.
57. González-Díaz H, Prado-Prado F, Ubeira FM. Predicting Antimicrobial Drugs and Targets with the MARCH-INSIDE approach. *Cuerrent Topics in Medicinal Chemistry*. 2008;8(18):1676-90.
58. Cruz-Montegudo M, González-Díaz H, Borges F, Dominguez ER, Cordeiro MN. 3D-MEDNEs: An Alternative "in Silico" Technique for Chemical Research in Toxicology. 2. Quantitative Proteome-Toxicity Relationships (QPTR) based on Mass Spectrum Spiral Entropy. *Chem Res Toxicol*. 2008(21):619-32.
59. Junker BH, Koschuetzki D, Schreiber F. Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics*. 2006 Apr 21;7(1):219.
60. Estrada E, Rodriguez-Velazquez JA. Subgraph centrality in complex networks. *Phys Rev E*. 2005 May;71(5 Pt 2):056103.
61. 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.
62. González-Díaz H, Sanchez IH, Uriarte E, Santana L. Symmetry considerations in Markovian chemicals 'in silico' design (MARCH-INSIDE) I: central chirality codification, classification of ACE inhibitors and prediction of sigma-receptor antagonist activities. *Comput Biol Chem*. 2003 Jul;27(3):217-27.
63. Prado-Prado F, González-Díaz H, Santana L, Uriarte E. Unified QSAR approach to antimicrobials. Part 2: Predicting activity against more than 90 different species in order to halt antibacterial resistance. *Bioorg Med Chem*. 2007;15 897-902.
64. 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.
65. Prado-Prado F, González-Díaz H, Santana L, Uriarte E. Unified QSAR approach to antimicrobials. Part 2: Predicting activity against more than 90 different species in order to halt antibacterial resistance. *Bioorg Med Chem*. 2007;15:897-902.

66. Van Waterbeemd H. Chemometric methods in molecular design. New York: Wiley-VCH 1995.
67. StatSoft.Inc. STATISTICA (data analysis software system), version 6.0, [www.statsoft.com](http://www.statsoft.com). Statsoft, Inc. 6.0 ed 2002.
68. Prado-Prado FJ, González-Díaz H, Martínez de la Vega O, Ubeira FM, Chou KC. Unified QSAR approach to antimicrobials. Part 3: First multi-tasking QSAR model for Input-Coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. *Bioorg Med Chem*. 2008;16:5871–80.
69. González-Díaz 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.
70. Prado-Prado J, Martínez 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.
71. Microsoft.Corp. Microsoft Excel 2002:Microsoft Excel
72. Batagelj V, Mrvar A. Pajek 1.15. 2006.
73. Devah P. Mark of a Criminal Record. *Am J Soc*. 2003(108):937-75.