

MIANN models in medicinal, physical and organic chemistry

Humberto González-Díaz^{1,2}, Sonia Arrasate², Nuria Sotomayor², Esther Lete², Cristian R. Munteanu³, Alejandro Pazos³, Lina Besada-Porto⁴ and Juan M. Ruso⁴

¹ IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain;

² Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), 48940, Leioa, Bizkaia, Spain;

³ Department of Information and Communication Technologies, Computer Science Faculty, University of A Coruña (UDC), Campus de Elviña, 15071, A Coruña, Spain;

⁴ Soft Matter and Molecular Biophysics Group, Faculty of Physics, University of Santiago de Compostela (USC), 15782, Santiago de Compostela, Spain

Abstract

Reducing costs in terms of time, animal sacrifice, and material resources with computational methods has become a promising goal in Medicinal, Biological, Physical and Organic Chemistry. There are many computational techniques that can be used in this sense. In any case, almost all these methods focus on few fundamental aspects including: type (1) methods to quantify the molecular structure, type (2) methods to link the structure with the biological activity, and others. In particular, MARCH-INSIDE (MI), acronym for Markov Chain Invariants for Networks Simulation and Design, is a well-known method for QSAR analysis useful in step (1). In addition, the bio-inspired Artificial-Intelligence (AI) algorithms called Artificial Neural Networks (ANNs) are among the most powerful type (2) methods. We can combine MI with ANNs in order to seek QSAR models, a strategy which is called herein MIANN (MI & ANN models). One of the first applications of the MIANN strategy was in the development of new QSAR models for drug discovery. MIANN strategy has been expanded to the QSAR study of proteins, protein-drug interactions, and protein-protein interaction networks. In this paper, we review for the first time many interesting aspects of the MIANN strategy including theoretical basis, implementation in web servers, and examples of applications in Medicinal and Biological chemistry. We also report new applications of the MIANN strategy in Medicinal chemistry and the first examples in Physical and Organic Chemistry, as well. In so doing, we developed new MIANN models for several self-assembly physicochemical properties of surfactants and large reaction networks in organic synthesis. In some of the new examples we also present experimental results which were not published up to date.

Keywords: Artificial neural networks; Drug-target networks; Multitarget QSAR; Organic reaction networks; Protein interaction networks; Surfactant QSPR models

1. INTRODUCTION

Reducing costs in terms of time, animal sacrifice, and material resources with computational methods has become a promising goal in Medicinal, Biological, Physical and Organic Chemistry [1]. There are many computational techniques that can be used in this sense, to cite a few: Quantum Mechanics/Molecular Mechanics (QM/MM) [2], Monte Carlo methods (MC) [2, 3], Docking and/or Quantitative Structure-Activity Relationships (QSAR) [4-10]. In any case, almost all these methods focus on few fundamental aspects including: type (1) methods to quantify the molecular structure, type (2) methods to link the structure with the biological activity, and others. Several methods of type (1) use QM and/or Graph theory [11-15], whereas the type (2) methods use Statistical and/or Machine Learning (ML) techniques. DRAGON [16-18], TOPS-MODE [19-22], TOMOCOMD [23, 24], CODESSA [25, 26], and MOE [27] are some of the most used software that implement type (1) methods. Linear Discriminant Analysis (LDA) implemented in STATISCA [28] or ML methods implemented in WEKA [29] are examples of type (2) methods and the software used to carry them out.

In this context, different researchers/journals edited important monographic issues in order to discuss different computational methods. Some of the most recent have been published by *Current Topics in Medicinal Chemistry*. For instance, Bisson has edited a special issue about Computational Chemogenomics in Drug Design and Discovery [30]. Speck-Planche and Cordeiro have guest-edited a special issue about computer-aided techniques for the design of antihepatitis C agents [31]. Prado-Prado and García-Mera have also guest-edited a special issue about computer-aided Drug Design and molecular docking for disorders of the central nervous system and other diseases [32]. González-Díaz has guest-edited two special issues about mt-QSAR and Complex Networks applied to Medicinal Chemistry [33, 34]. In all these issues, and others of the same journal, several reviews and research papers have been published in this area [5, 27, 35-69].

In particular, the bio-inspired Artificial-Intelligence (AI) algorithms called Artificial Neural Networks (ANNs) are among the most powerful type (2) methods. Important applications of ANNs in classifications and feature selection have been published [70-74]. In this paper, we focus on one of these methods called MIANN. Thus, we review for the first time many interesting aspects of the MIANN strategy including theoretical basis, implementation in web servers, and examples of applications in Medicinal and Biological Chemistry. We also report new applications of the MIANN strategy in Medicinal Chemistry and the first examples in Physical and Organic Chemistry, as well. In so doing, we developed new MIANN models for drug-target interactions, several physicochemical properties of surfactants, and large reaction networks in organic synthesis. In some of the new examples we also present experimental results which were not published up to date.

2. THE MIANN APPROACH

MARCH-INSIDE (Markov Chain Invariants for Networks Simulation and Design) is a well-known type (1) method for QSAR analysis. The method has been mentioned in many recent review works published by different groups [65, 75-82]. We can combine MI with different Machine Learning algorithms. In particular, we can combine MI with ANNs in order to seek QSAR models. The name of this strategy is MIANN (MARCH-INSIDE & ANN models).

2.1. Theoretical Basis of MIANN Analysis

2.1.1. Parameters for Drug Structure

The MARCH-INSIDE approach [1, 65, 83] is based on the calculation of the different physicochemical molecular properties as an average of atomic properties (w_j). For instance, it is possible to derive average estimations from molecular descriptors or local indices such as electronegativity values ${}^k\chi(G)$ [84, 85].

$${}^k\chi(G) = \sum_{j \in G} p_k(\chi_j) \cdot \chi_j \quad (1)$$

$$\theta_k(G) = - \sum_{j \in R}^n {}^k p_j(G) \cdot \log[{}^k p_j(G)] \quad (2)$$

$$\pi_k(G) = \sum_{i=j \in R}^n {}^k p_{ij}(G) \quad (3)$$

In this case, $w_j = \chi_j$, the atomic electronegativity. It is also possible to consider isolated atoms ($k = 0$) in a first estimation of the molecular properties ${}^0\chi(G)$, ${}^0\theta(G)$, or ${}^0\pi(G)$. In this case, the probabilities ${}^0p(\chi_j)$ are determined without considering the formation of chemical bonds (simple additive scheme). However, it is possible to consider the gradual effects of the neighboring atoms at different distances in the molecular backbone. In order to reach this goal, the method uses an MCM, which determines the absolute probabilities $p_k(w_j)$ with which the atoms placed at different distances k affect the contribution of the atom j to the molecular property in question.

2.1.2. Parameters for Protein 3D Structures

In this work, the information about the molecular structure of the proteins is codified using the MM method with the ${}^1\Pi$ matrix (the short-term electrostatic interaction matrix). The matrix ${}^1\Pi$ is constructed as a squared matrix ($n \times n$), where n is the number of amino acids (aa) in the protein [86-88]. In previous works we have predicted the protein function based on $\mu_k(R)$ and $\theta_k(R)$ values 3D-Potentials for different types of interactions or molecular fields derived from ${}^1\Pi$. The main types of the used molecular fields are: electrostatic (E), Van der Waals (vdW) and HINT potential [87, 89, 90]. In this paper, we have calculated $\pi_k(R)$ and $\theta_k(R)$ values only for E and HINT potentials. We have omitted the vdW term due to a simple reason, the HINT potential includes a vdW component. The values have been used here as inputs to construct the QSAR model. A detailed explanation on the subject has been previously published. As follows, we give the formula for $\pi_k(R)$, $\theta_k(R)$ and $\xi_k(R)$ and some general explanations:

$$\xi_k(R) = - \sum_{j \in R} {}^k p_j(R) \cdot \xi_0(j) \quad (4)$$

$$\theta_k(R) = - \sum_{j \in R}^n {}^k p_j(R) \cdot \log[{}^k p_j(R)] \quad (5)$$

$$\pi_k(R) = \sum_{i=j \in R}^n {}^k p_{ij}(R) \quad (6)$$

It is remarkable that the spectral moments depend on the probability ${}^k p_{ij}(\mathbf{R})$ with which the effect of the interaction f propagates from amino acid i^{th} to other neighboring amino acids j^{th} and returns to i^{th} after k -steps. On the other hand, both the average electrostatic potential and the entropy measures depend on the absolute probabilities ${}^k p_j(\mathbf{R})$ with which the amino acid j^{th} has an interaction of type f with the rest of amino acids. In any case, both probabilities refer to a first ($k = 1$) direct interaction of type f between amino acids placed at a distance equal to k -times the cut-off distance ($r_{ij} = k \cdot r_{\text{cut-off}}$). The method uses a Markov Chain Model (MCM) to calculate these probabilities, which also depend on the 3D interactions between all pairs of amino acids placed at a distance r_{ij} in r_3 in the protein structure. However, for the sake of simplicity, a truncation or cut-off function α_{ij} is applied in such a way that a short-term interaction takes place in a first approximation only between neighboring aa ($\alpha_{ij} = 1$ if $r_{ij} < r_{\text{cut-off}}$). Otherwise, the interaction is banished ($\alpha_{ij} = 0$). The relationship α_{ij} may be visualized in the form of a protein structure complex network. In this network the nodes are the C_α atoms of the amino acids and the edges connect pairs of amino acids with $\alpha_{ij} = 1$. Euclidean 3D space $\mathbf{r}_3 = (x, y, z)$ coordinates of the C_α atoms of amino acids listed on protein PDB files. For the calculation, all water molecules and metal ions were removed.[1] The MI[1] software performs all these calculations by evaluation of the summation term either for all amino acids or only for some specific groups called regions or orbitals (\mathbf{R}). These regions are often defined in geometric terms and called core, inner, middle or surface region. The protein is virtually divided into the following regions: c corresponds to core, i to inner, m to middle, and s to surface regions, respectively. The diameters of the regions are 0 to 25 for region c , 25 to 50 for region i , 50 to 75 for region m , and 75 to 100 for region s . These values are given in terms of percentage of the longest distance r_{max} with respect to the center of charge. Additionally, we consider the total region (t) that contains all the amino acids in the protein (region diameter 0 to 100% of r_{max}). Consequently, we can calculate different parameters ($\pi_k(\mathbf{R})$, $\xi_k(\mathbf{R})$, and $\theta_k(\mathbf{R})$) for the amino acids contained in a region (c, i, m, s , or t) and placed at a topological distance k within this region (k is the name of the order) [90-94]. In this work, we calculated a total of 90 indices (3 types of indices x 5 types of regions x 6 higher order considered) for each protein.

2.2.2. Pseudo-Folding Parameters for Proteins Sequences

The MARCH-INSIDE approach is based on the calculation of the different molecular parameters of both proteins (Ps) and drugs and/or organic ligands (Ls) of proteins [65, 82]. In this sense, different parameters of protein pseudo-folding in 2D lattice-like spaces have been used before in proteome research [95]. On the other hand, in previous works, we have predicted a protein function based on 3D-potentials for different types of interactions. The main types of potentials used are the averaged values of electrostatic, Van der Waals and HINT potentials [96, 97]. In this paper, we combine for the first time the pseudo-folding electrostatic ξ_k potentials of protein sequences with average electronic parameters of drugs. These values of protein sequences were used as inputs to construct the QSAR model together with the parameters of drug structure. The key of the method is to overcome the higher dimension space bottleneck after previous grouping of the monomers (amino acids) into four groups. These four groups characterize the physicochemical nature of the amino acids as: polar, non-polar, acid, or basic. Classification of monomers (amino acids) as acid or basic prevails over polar/non-polar classification in such a way that the four groups do not overlap each other. Subsequently, each amino acid in the sequence is placed in a Cartesian 2D space starting with the first monomer at the (0, 0) coordinates. The coordinates of the successive amino acids are calculated following simple heuristics, in such a way it can be used for a DNA [98]. Secondly, the method uses the matrix ${}^1\Pi$, which is a squared matrix to characterize the protein sequence pseudo-folding into a 2D lattice-like space (see graphs). Please, note that the number of nodes (n) in the graph may be equal or even smaller than the number of amino acids in the protein sequence. Accordingly, the matrix ${}^1\Pi$ contains the probabilities ${}^1 p_{ij}$ to reach a node n_i moving throughout a walk of length $k = 1$ from another node n_j :

$$p_{ij}(P_n) = \frac{Q_j(P_n)}{\sum_{m=1}^n \alpha_{im} \cdot Q_m(P_n)} \quad (7)$$

Where, Q_j is the charge of the node n_j and a_{ij} equals to 1 if the nodes n_i and n_j are adjacent in the graph and equals to 0 otherwise. The charge of the node Q_j is equal to the sum of the charges of all amino acids projected over this node after protein sequence pseudo-folding. We can list all these Q_j values calculating the elements of a vector of electrostatic potentials $\mathbf{q}_0 \equiv [Q_j/d_j] \equiv [Q_1/d_1, Q_2/d_2, \dots, Q_j/d_j, Q_{j+1}/d_{j+1}, \dots, Q_n/d_n]$ being d_j the Euclidian distance from the node j^{th} and the center of coordinates (0, 0) in the 2D pseudo-folding space. Afterwards, the calculation of the average electrostatic potentials of protein pseudo-folding is carried out directly.

$$\xi_k = \mathbf{p}_0^t \cdot \mathbf{q}_0 = \mathbf{p}_0^t \cdot (\mathbf{1})^k \cdot \mathbf{q}_0 = \sum_{j=1}^n p_k(j) \cdot \xi(j) \quad (8)$$

The average general potentials depend on the initial absolute probabilities $p_0(j)$ and the total potential with which all the amino acids projected over the lattice node j^{th} interact with the rest of amino acids projected over other nodes of the space. These are the probabilities with which the amino acids interact with other amino acids placed at a topological distance d_{ij} equal to k -times the cut-off distance ($d_{ij} = k$). These probabilities are the elements of the vector $\mathbf{p}_0 \equiv [Q_j/d_j] \equiv [Q_1/d_1, Q_2/d_2, \dots, Q_j/d_j, Q_{j+1}/d_{j+1}, \dots, Q_n/d_n]$.

2.2.4. Parameters of PPIs Used in the MIANN Analysis

In principle, we can use as input for the MIANN analysis of PPIs all types of PPI invariants calculated with MI. However, the only MI parameter calculated with MI that has been used for a MIANN analysis of PPIs until now is the entropy $\theta_k(R)$. The used entropy parameters represent the average electrostatic entropy (θ) due to the interactions between all pairs of amino acids allocated inside a specific protein region (R) and placed at a distance k from each other. In order to describe PPIs we need to use $\theta_k(R)$ values of two proteins, $\theta_k(^1R)$ for protein 1 and $\theta_k(^2R)$ for protein 2, in order to generate structural parameters describi>Curr. Top. Med. Chem.</secondary-title></titles><periodical><full-title>Curr. Top. Med. Chem.</full-title></periodical><pages>1883-8</pages><volume>12</volume><number>17</number><dates><year>2012</year></dates><pub-location>Netherlands</pub-location><isbn>1873-4294 (Electronic)1568-0266 (Linking)</isbn><accessioⁿ $\theta_k(R)$: PPI Average Entropy Invariant (ti = a), PPI Entropy Difference Invariant (ti = d), and PPI Entropy Product Invariant (ti = p):

$${}^a\theta_k(R) = {}^a\theta_k(^1R_1, ^2R_1) = \frac{1}{2}[\theta_k(^1R_1) + \theta_k(^2R_1)] \quad (9)$$

$${}^d\theta_k(R) = {}^d\theta_k(^1R_1, ^2R_1) = |[\theta_k(^1R_1) - \theta_k(^2R_1)]| \quad (10)$$

$${}^p\theta_k(R) = {}^p\theta_k(^1R_1, ^2R_1) = \theta_k(^1R_1) \cdot \theta_k(^2R_1) \quad (11)$$

Notably, in order to guarantee that these parameters are invariant to protein labeling as 1 or 2, we have to always use the same ${}^1R = {}^2R = R$ and $k_1 = k_2 = k$ values. In order to calculate the $\theta_k(R)$ values for each protein, the method uses as a source of protein macromolecular descriptors the stochastic matrices ${}^1\mathbb{I}_e$ built up as squared matrices ($n \times n$), where n is the number of amino acids (aa) in the protein. The subscript e points to the electrostatic type of molecular force field. In previous works we have predicted the protein function based on $\theta_k(R)$ values for different types of interactions or molecular fields. The main types of molecular fields used are the following: electrostatic, vdW, and HINT entropies.

2.2.5. ANN Analysis in the Context of the MIANN Strategy

On the one hand, the MIANN strategy involves using as inputs, in order to train the ANNs, the MI parameters of the r^{th} drugs or protein ligands (L_r) in general: ${}^k\chi(L_r)$, ${}^k\theta(L_r)$, or ${}^k\pi(L_r)$. On the other hand, we should use the MI parameters of s^{th} protein sequences or 3D structures: ${}^k\xi(P_s)$, ${}^k\theta(P_s)$, or ${}^k\pi(P_s)$. In the case of a Linear ANN (LNN), the models may have the following general

formulae for the cases of a drug, protein, DPIs, or PPI analysis (examples using only entropies ${}^k\theta(L_r)$ for drugs, ligands, or low-weight molecules and ${}^k\theta(P_s)$ for proteins):

$$S(L_r)_{pred} = \sum_{k=0}^5 \alpha_k \cdot {}^k\theta(L_r) + c_0 \quad (12)$$

$$S(P_s)_{pred} = \sum_{k=0}^5 b_k \cdot {}^k\theta(P_s) + c_0 \quad (13)$$

$$S(DPI_{rs})_{pred} = \sum_{k=0}^5 \alpha_k \cdot {}^k\theta(L_r) + \sum_{k=0}^5 b_k \cdot {}^k\theta(P_s) + c_0 \quad (14)$$

$$S(PPI_{rs})_{pred} = \sum_{k=0}^5 \alpha_k \cdot \frac{1}{2} [\theta_k(P_r) + \theta_k(P_s)] \\ + \sum_{k=0}^5 b_k \cdot |\theta_k(P_r) - \theta_k(P_s)| + \sum_{k=0}^5 c_k \cdot \theta_k(P_r) \cdot \theta_k(P_s) + d_0 \quad (15)$$

The model deals with the classification of a set of compounds, proteins, DPIs, or PPIs, with or without affinity to different protein targets. A dummy input variable Affinity Class (AC) codifies the affinity; AC = 1 for well-known DPIs and AC = 0 otherwise. This variable indicates either high (AC = 1) or low (AC = 0) affinity of the drug to the target protein. The parameter $S(DPI_{rs})_{pred}$ is the output of the model and it is a continuous and dimensionless score that gives higher values for DPIs and lower values for nDPIs. In the model, α_k , b_k , and c_0 represent the coefficients of the MIANN function determined by the LNN technique using the STATISTICA 6.0 software package [99]. We can check Specificity, Sensitivity, total Accuracy, or Area Under ROC curve (AUROC) to determine the quality-of-fit to data in the training and external validation series. DPIs and nDPIs cases in the validation series should not be used to train the model.

2.2.6. Data Set Useful for the MIANN Analysis

In principle, we can withdraw a dataset for the MIANN analysis from different public resources. However, in almost all cases we have obtained the datasets from the Drug Bank (DB, <http://www.drugbank.ca/>) an online source available for public research. From this website, we obtained a list of all drugs approved by the US Food and Drug Administration (US FDA, <http://www.fda.gov/>). We only included DPIs with known affinity of drugs to target with a known 3D structure available in the Protein Data Bank (PDB, <http://www.pdb.org/>) [100]. The data set contains more than >300 drugs with their respective >300 molecular targets. The data also contain negative cases made up of active compounds for known targets but not marketed as active against other targets (nDPIs). Therefore, we were able to collect over >6,000 cases (DPIs/nDPIs). Due to space constraints, the names, DB codes, SMILE codes, target function, and target PDB code for both drugs and/or targets were depicted in the supplementary material file SM2.pdf;.

3. MIANN MODELS IN MEDICINAL CHEMISTRY

3.1. Examples of ANN and MIANN Models in QSAR and mt-QSAR

3.1.1. ANN Model of Drugs Multiplex Toxic Effects

ANN models can be used not only in QSAR but also in mt-QSAR. For instance, in ref. [101], the authors reported the most recent upgrade of the ANN-based strategies for mt-QSAR. In so doing, they used the TOPS-MODE approach to calculate drug molecular descriptors and the STATISTICA software to seek different MIANN models such as Linear Neural Network (LNN), Radial Basis Function (RBF), Probabilistic Neural Networks (PNN) and Multi-Layer Perceptrons (MLP). The best model found was the LNN, which correctly classified 8,258 out of 9,000 (Accuracy = 93.0%) multiplexing assay endpoints of 7,903 drugs (including both training and test series). Each endpoint corresponds to one out of 1,418 assays, 36 molecular or cellular targets, 46 standard type measures, in two possible organisms (human and mouse). Secondly, we have determined experimentally, for the first time, the values of $EC_{50} = 11.41 \mu\text{g/mL}$ and Cytotoxicity = 27.1% for the drug G1 over Balb/C mouse spleen macrophages using flow cytometry. In addition, we have used the LNN model to predict the G1 activity in 1,265 multiplexing assays not measured experimentally (including 152 cytotoxicity assay endpoints). Both experimental and theoretical results point out a low macrophage cytotoxicity of G1. This work breaks new ground for the '*in silico*' multiplexing screening of large libraries of compounds. The same idea behind the application of ANN models in QSAR and mt-QSAR can be applied to seek MIANN models. As follows we review some MIANN models in QSAR and mt-QSAR.

3.1.2. MIANN Model of Drugs with Anti-Cancer Activity

Developing a model for predicting anti-cancer activity of any classes of organic compounds based on molecular structure is a very important goal for a medicinal chemist. Nevertheless, the structural diversity of compounds is so vast that we may need non-linear models such as ANNs instead of linear ones. In a previous work we have used the MIANN strategy to solve this problem [102]. In this work, the MIANN analysis was used to model the anti-cancer activity of organic compounds, which has shown a high average accuracy of 93.79% (training performance) and predictability of 90.88% (validation performance) for the 8:3-MLP topology with different training and predicting series. This MIANN model compares favorably with respect to a previous linear model that showed only 80.49% of accuracy and 79.34% of predictability [103]. (Fig. 1) depicts a graphical abstract of this work. The present MIANN model based on the SmartMLP approach employed shorter training times of only 10h, while previous models have given accuracies of 70-89% only after 25-46 h of training. In order to illustrate the practical use of the model in Bioorganic Medicinal Chemistry, we report the *in silico* prediction, and *in vitro* evaluation of six new synthetic tegafur analogues having IC_{50} values in a broad range between 37.1 and 138 $\mu\text{g/mL}$ for leukemia (L1210/0) and human T-lymphocyte (Molt4/C8, CEM/0) cells. Theoretical predictions coincide very well with experimental results.

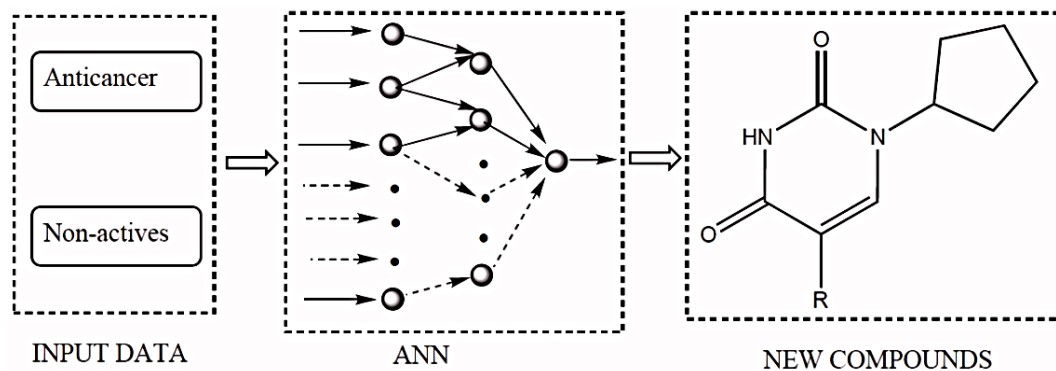


Fig. (1). Graphical abstract of the first work with a MIANN model.

3.1.3. MIANN Approach to *mt*-QSAR of Multi-Target AntiParasite Effects

There are many pathogen parasite species with different susceptibility profile to anti-parasitic drugs. Unfortunately, most QSAR models predict the biological activity of drugs against only one parasite species. Consequently, predicting the probability with which a drug is active against different species with a single unified model is a goal of major importance. In so doing, we used the MIANN strategy to seek the first *mt*-QSAR model for 500 drugs tested in the literature against 16 parasite species and other 207 drugs not tested in the literature using spectral moments. We tested different ANN topologies such as LNN and MLP, RBF and Probabilistic Neural Network (PNN), see (Fig.2). The best MIANN model found had an overall training performance of 87%. The present work reports the first attempts to calculate within unified framework probabilities of anti-parasitic action of drugs against different parasite species based on a spectral moment analysis [104].

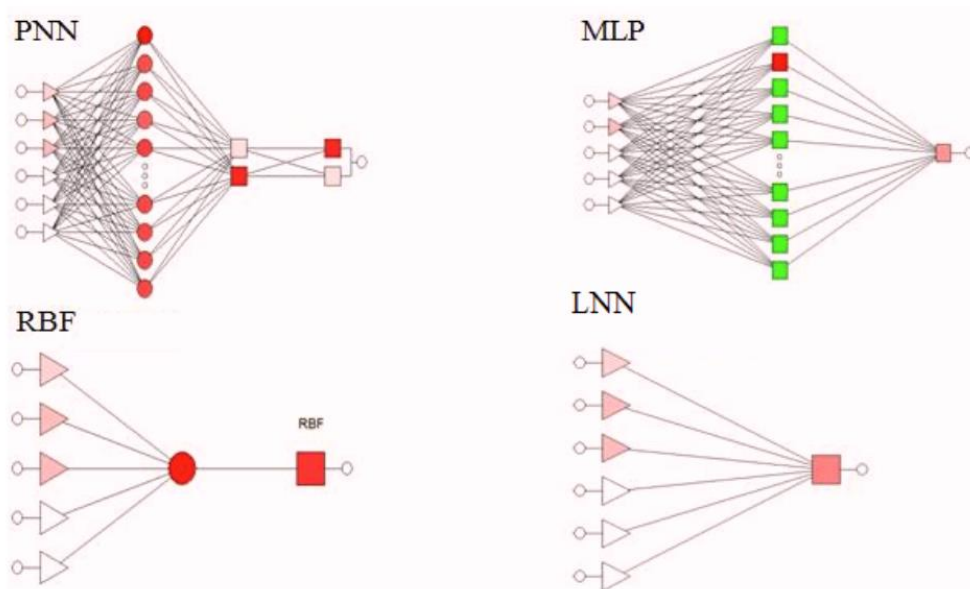


Fig. (2). Different ANN topologies used in MIANN studies.

3.2. Examples of MIANN Models of DPIs

3.2.1. ANN Models of DPIs

Prediction of drug-target proteins interactions (DPIs) is very important for drug discovery. The importance of this topic increases if we take into consideration the high number of experimental assays that are necessary in order to elucidate all possible relationships in DPI complex networks. See for instance, the high complexity of bipartite graph constructed in Ref. [105] in order to represent all DPIs between drugs approved by the US FDA with different target proteins. Hence, the development of new computational methods able to accurately predict DPIs is of major interest [106]. We can discriminate between DPIs of drugs with high affinity and those pairs of drugs with no affinity to different targets (nDPIs) with models that use as input structural parameters of both drugs and protein targets. This type of QSAR-based methods for prediction of DPIs is a particular case of mt-QSAR models [1, 107-109]. To this end, we should take into consideration that they allow the study of multiple target proteins and allow the reconstruction of the respective DPI complex networks. Conversely, almost all QSAR techniques allow the prediction of DPIs for only one target [104, 110-119]. We can use ANN algorithms to seek mt-QSAR models of DPIs. For instance, the researchers combined in Ref. [120] the mt-QSAR ideas with an ANN algorithm to study compounds active against Colorectal Cancer (CRC). CRC is one of the most studied cancers because of its high prevalence and number of deaths. The same authors [121] used ANNs to construct an mt-QSAR model of DPIs for drugs with activity in patients with the Acquired Immunodeficiency Syndrome (AIDS). In both works, the ANN models found classified more than 90% of active and inactive compounds in training and prediction sets. As follows we review some MIANN models for DPIs. (Fig. 3) shows the general workflow used to seek a MIANN model of DPIs.

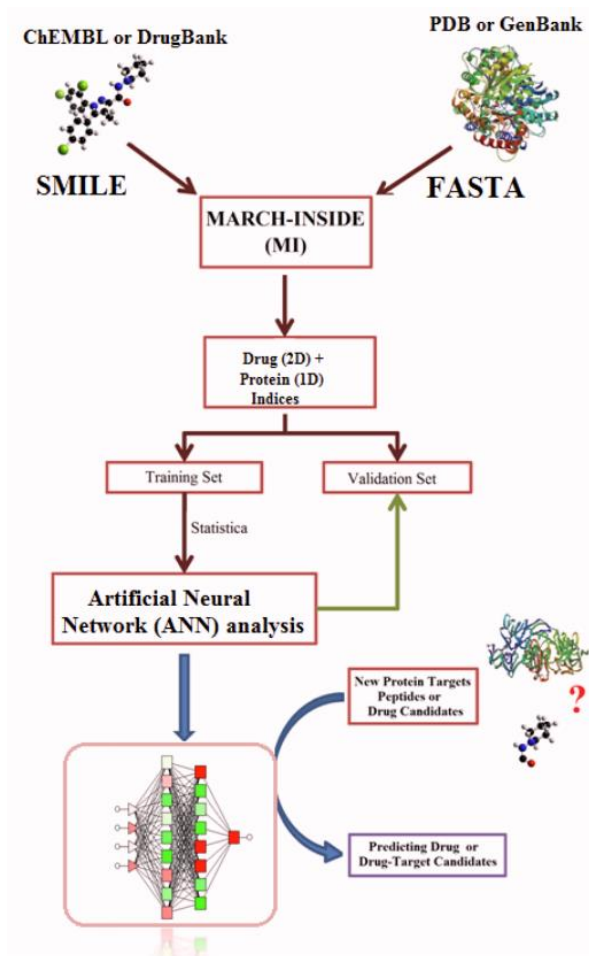


Fig. (3). Workflow of the MIANN strategy for the drug-target interaction network problem.

3.2.2. MI-DRAGON: MIANN Models for DPIs

In a more recent work [122], the scientists combined the MARCH-INSIDE method with the DRAGON software to create 2D MI-DRAGON, a new predictor for DPIs. The MARCH-INSIDE (MI) software was used to calculate 3D structural parameters for targets and the DRAGON software was used to calculate 2D molecular descriptors of drugs. Both classes of parameters were used as input of different ANN algorithms to seek an accurate non-linear mt-QSAR predictor. The best ANN model found is a MLP with profile MLP 21:21-31-1:1. This MLP classifies correctly 303 out of 339 DPIs (Sensitivity = 89.38%) and 480 out of 510 nDPIs (Specificity = 94.12%), corresponding to training Accuracy = 92.23%. The validation of the model was carried out by means of external predicting series with Sensitivity = 92.18% (625/678 DPIs; Specificity = 90.12% (730/780 nDPIs) and Accuracy = 91.06%. 2D MI-DRAGON offers a good opportunity for fast-track calculation of all possible DPIs of one drug enabling to re-construct large drug-target or DPIs complex networks. For instance, we reconstructed the complex network of the US FDA benchmark dataset with 855 nodes (519 drugs + 336 targets). Finally, the practical use of 2D MI-DRAGON was illustrated in one theoretic-experimental study. We reported the prediction, synthesis, and pharmacological assay of 10 different oxisoaporphines with MAO-A inhibitory activity. The most active compound OXO5 presented $IC_{50} = 0.00083 \mu\text{M}$, notably better than the control drug Clorgyline.

3.2.3. NL MIND-BEST: the MIANN Model of DPIs Implemented in a Web Server

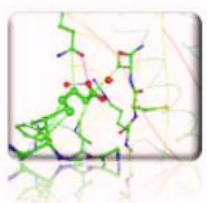
In references [123, 124], the authors developed both linear and MIANN models for DPIs using 2D physicochemical properties of drugs, but 3D physicochemical properties (electrostatic potentials) of the protein. The accuracy of the best linear model was 94.4% (3,859/4,086 cases) for training and 94.9% (1,909/2,012 cases) for the external validation series. In addition, the model was implemented into the Web portal Bio-AIMS. In so doing, we called the new web server MIND-BEST [123], acronym of MARCH-INSIDE Nested Drug-Bank Exploration & Screening Tool. The URL for this server is <http://bio-aims.udc.es/MIND-BEST.php>. This on-line tool is based on PHP/HTML/Python and MARCH-INSIDE routines. This work also illustrates with two examples the practical uses of this server. The experiment 1 includes MIND-BEST prediction, synthesis, characterization, and MAO-A and MAO-B pharmacological assay of eight rasagiline derivatives, promising for anti-Parkinson drug design. Experiment 2 reports the sampling, parasite culture, sample preparation, 2-DE, MALDI-TOF and -TOF/TOF MS, MASCOT search, 3D structure modeling with LOMETS, and MIND-BEST prediction for different peptides found in the proteome of the bird parasite *Trichomonas gallinae*, which is promising for the discovery of antiparasite drug targets.

We also published the MIANN version of the previous web server by the name NL MIND-BEST [124]. In so doing, we solved the DPI problem using the MIANN methodology. In other words, we followed the same flowchart as in MIND-BEST to calculate the parameters of drugs and proteins, but using different non-linear ANNs to seek the model. The best MIANN model found was the MLP 20:20-15-1:1. This MLP correctly classifies 611 out of 678 DPIs (Sensitivity = 90.12%) and 3,083 out of 3,408 nDPIs (Specificity=90.46%), corresponding to training Accuracy = 90.41%. The validation of the model was carried out by means of external predicting series. The model correctly classifies 310 out of 338 DPIs (Sensitivity = 91.72%) and 1,527 out of 1,674 nDPI (Specificity=91.22%) in validation series, corresponding to total Accuracy = 91.30% for validation series (Predictability). This model favorably compares with other ANN models developed in this work and Machine Learning classifiers published before to address the same problem in different aspects. This web server is located at the URL: <http://bio-aims.udc.es/NL-MIND-BEST.php>. (Fig.4) shows the online user interface for the web server NL MIND-BEST. We also illustrated here, with two practical examples, the potential uses of this server. In experiment 1, we report for the first time the Quantum QSAR study, synthesis, characterization, and experimental assay of antiplasmodial and cytotoxic activities of oxoisoaporphine alkaloids derivatives as well as the NL MIND-BEST prediction of potential target proteins. In experiment 2, we report the sampling, parasite culture, sample preparation, 2-DE, MALDI-TOF, and -TOF/TOF MS, MASCOT search, MM/MD 3D structure modeling, and NL MIND-BEST prediction for different peptides found in the proteome of the human parasite *Giardia lamblia*, which is promising for the discovery of anti-parasite drug targets.

Please choose drug and protein names to add to the calculation lists (maximum 10)

(A)

Mode 1: Drugs - Standard PDBs



NL-MIND-BEST
Non-Linear MARCH-INSIDE
Nested Drug-Bank
Exploration & Screening Tool
Tool: MARCH-INSIDE (Py)
Data: RCSB PDB

Input drug names:
(Drug name[TAB]SMILE formula)

```

Irinotecan
CCCC1=CC2CN3C(=O)C4=C(C=C3C2=NC2=C1C=C(OC(=O)N1CCCC1)N1CCCC1)C
=C2)C(O)(CC)C(=O)OC4
Simvastatin
CCC(C)
(C)C(=O)OC1CC(C)C=C2C=CC(C)C(CCC3CC(O)CC(=O)O4)C12
  
```

Input a PDB names:
(only PDB ID)

```

1A36
1A8M
  
```

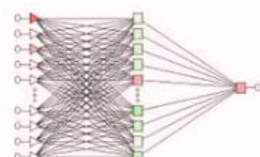
Get Model FDA-approved Drugs*
(use Copy-Paste)

Get Model FDA-approved PDBs*
(use Copy-Paste)

* access on demand

Mode 2: Drugs - LOMETS (PDB) / HyperChem (ENT) files

(B)



MLP 20:20-15-1:1
Test Accuracy = 91.30%
 (Training Accuracy = 90.41%)

Input drug names:
(Drug name[TAB]SMILE formula)

```

Irinotecan
CCCC1=CC2CN3C(=O)C4=C(C=C3C2=NC2=C1C=C(OC(=O)N1CCCC1)N1CCCC1)C
=C2)C(O)(CC)C(=O)OC4
Simvastatin
CCC(C)
(C)C(=O)OC1CC(C)C=C2C=CC(C)C(CCC3CC(O)CC(=O)O4)C12
  
```

Get any Model FDA-approved Drugs
(use Copy-Paste)

or

Use all FDA drugs

Upload & evaluate one PDB from LOMETS or ENT from HyperChem (max. 2MB)

Please select LOMETS PDB / HyperChem ENT file No se ha...archivo

* access on demand

Fig. (4). Online user interface for the web server NL MIND-BEST.

3.2.4. MIANN Model of DPIs Based on Entropy Parameters

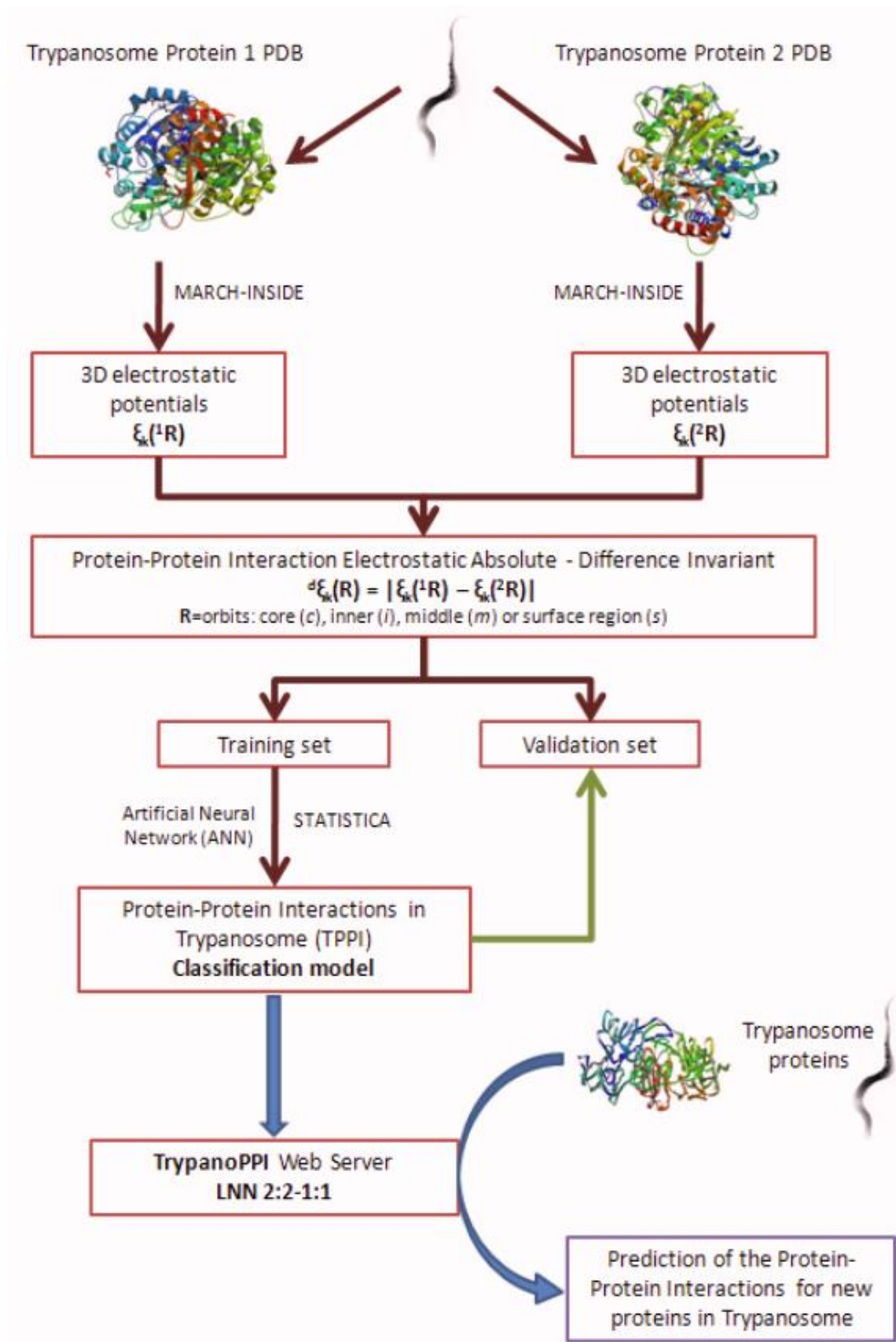
In all these works, we used 2D/3D physicochemical parameters of drugs and proteins calculated with the MI software to apply the MIANN strategy. However, MI calculates not only physicochemical parameters from a more classical point of view, but also different theoretic invariants of molecular structure, which are useful to solve the DPI problem with QSAR models. For instance, Ref. [125] reported an alternative MIANN solution of the DPI problem using entropy parameters of both drug and target structure. The best MIANN model found is the MLP 32:32-15-1:1. This MLP correctly classifies 623 out of 678 DPIs (Sensitivity = 91.89%) and 2,995 out of 3,234 nDPIs (Specificity = 92.61%), corresponding to training Accuracy = 92.48%. The validation of the model was carried out by means of external predicting series. The model correctly classifies 313 out of 338 DPIs (Sensitivity = 92.60%) and 1,411 out of 1,534 nDPIs (Specificity = 91.98%) in validation series, corresponding to total Accuracy = 92.09% for validation series (Predictability). The authors illustrated with two practical examples the use of this model. The first example includes the prediction, synthesis, characterization, and MAO-A and MAO-B pharmacological assay of 10 rasagiline derivatives promising for anti-Parkinson drug design. The second example contains the sampling, parasite culture, SEC and IDE sample preparation,

MALDI-TOF MS and MS/MS analysis, MASCOT search, MM/MD 3D structure modeling, and QSAR prediction for different peptides of hemoglobin found in the proteome of the human parasite *Fasciola hepatica*. This protein is promising for the discovery of anti-parasite drug targets.

4. MIANN MODELS IN BIOLOGICAL CHEMISTRY

4.1. Example of MIANN Model of PPIs

The first MIANN model of PPIs reported up to date was introduced to study PPIs in the parasite Trypanosoma. *Trypanosoma brucei* causes African trypanosomiasis in human (HAT or African sleeping sickness) and Nagana in cattle. The disease threatens over 60 million people and uncounted numbers of cattle in 36 countries of sub-Saharan Africa and has a devastating impact on human health and the economy. On the other hand, *Trypanosoma cruzi* is responsible in South America for Chagas disease, which can cause acute illness and death, especially in young children. In this context, the discovery of novel drug targets in Trypanosome proteome is a major focus for the scientific community. Recently, many researchers have spent important efforts on the study of PPIs in pathogen Trypanosome species concluding that the low sequence identities between some parasite proteins and its human host render these PPIs as highly promising drug targets. To the best of our knowledge, there are no reported general models to predict Unique PPIs in Trypanosome (TPPIs) before the MIANN model published by our group. On the other hand, the 3D structure of an increasing number of Trypanosome proteins is being reported in databases. In this sense, the introduction of a new model to predict TPPIs from the 3D structure of proteins involved in PPI is very important. For this purpose, we introduced new protein-protein complex invariants based on the Markov average electrostatic potential $\xi_k(R_i)$ for amino acids located in different regions (R_i) of i -th protein and placed at a distance k from each other. (Fig. 5) illustrates the workflow used in this previous work [126].



Workflow used in the development of the MIANN server TrypanoPPI.

Fig. (5). Workflow used in the development of the MIANN server TrypanoPPI.

We calculate more than 30 different types of parameters of 7,866 protein pairs (1,023 TPPIs and 6,823 non-TPPIs) from more than 20 organisms, including parasites and human or cattle hosts. We found a very simple linear model that predicts above 90% of TPPIs and non-TPPIs both in training and independent test sub-sets using only two parameters. The parameters were $d_{\xi_k}(s) = |\xi_k(s_1) - \xi_k(s_2)|$ which are the absolute difference between the $\xi_k(s_i)$ values on the surface of the two proteins of the pairs. We also tested non-linear ANN models for comparison purposes but the LNN model gives better results. We implemented this predictor in the web server named TrypanoPPI, freely available to public at <http://bio-aims.udc.es/TrypanoPPI.php> (see user interface in Fig.6). This is the first model that predicts how unique is a protein-protein complex in Trypanosome proteome with respect to other parasites and hosts, providing new opportunities for the discovery of anti-trypanosome drug targets.

Ibero-NBIC Network
RNAS-IMEDIR, TIC
Computer Science Faculty
University of A Coruña
Spain

TrypanoPPI @ Bio-AIMS
Modelling the reality

Home | Links | About

Trypano-PPI
Trypanosome Protein-Protein
Interactions (TPPI)

Tool: MARCH-INSIDE (Python version)
Data: RCSB PDB

PDB-chain lists : Please paste the names of the PDB chains as two lists (maximum 50)
Notes: There is no space between the PDB name and the chain label, no empty new line;
the results will print the combination between the chains from the first list and the chains
from the second one.

1HO2A
1K3TB

1HO2B
1F2CA

Predict

LNN 2:2-1:1
(programed by Humberto González-Díaz)

Test Accuracy = 90.9%
(Training Accuracy = 89.5%)

Fig. (6). Online user interface of the MIANN server TrypanoPPI.

4.2. Example of a MIANN Model for Protein vs. Organism Self/Non-Self Analysis

Infections caused by human parasites (HPs) affect the poorest 500 million people worldwide and unfortunately chemotherapy has become expensive, toxic, and/or less effective due to drug resistance. On the other hand, many 3D structures remain without function annotation in the PDB. We need theoretical models to quickly predict biologically relevant Parasite Self Proteins (PSP), which are expressed differently in a given parasite and are dissimilar to proteins expressed in other parasites and have a high probability to become new vaccines (unique sequence) or drug targets (unique 3D structure). We published a MIANN model for PSPs in eight different HPs (*Ascaris*, *Entamoeba*, *Fasciola*, *Giardia*, *Leishmania*, *Plasmodium*, *Trypanosoma*, and *Toxoplasma*) with 90% accuracy for 15,341 training and validation cases [127]. The input parameters are the spectral moments of the Markov transition matrix for electrostatic interactions associated with the protein residue complex network calculated with the MARCH-INSIDE software. (Fig. 7) depicts the workflow used for the development of the model.

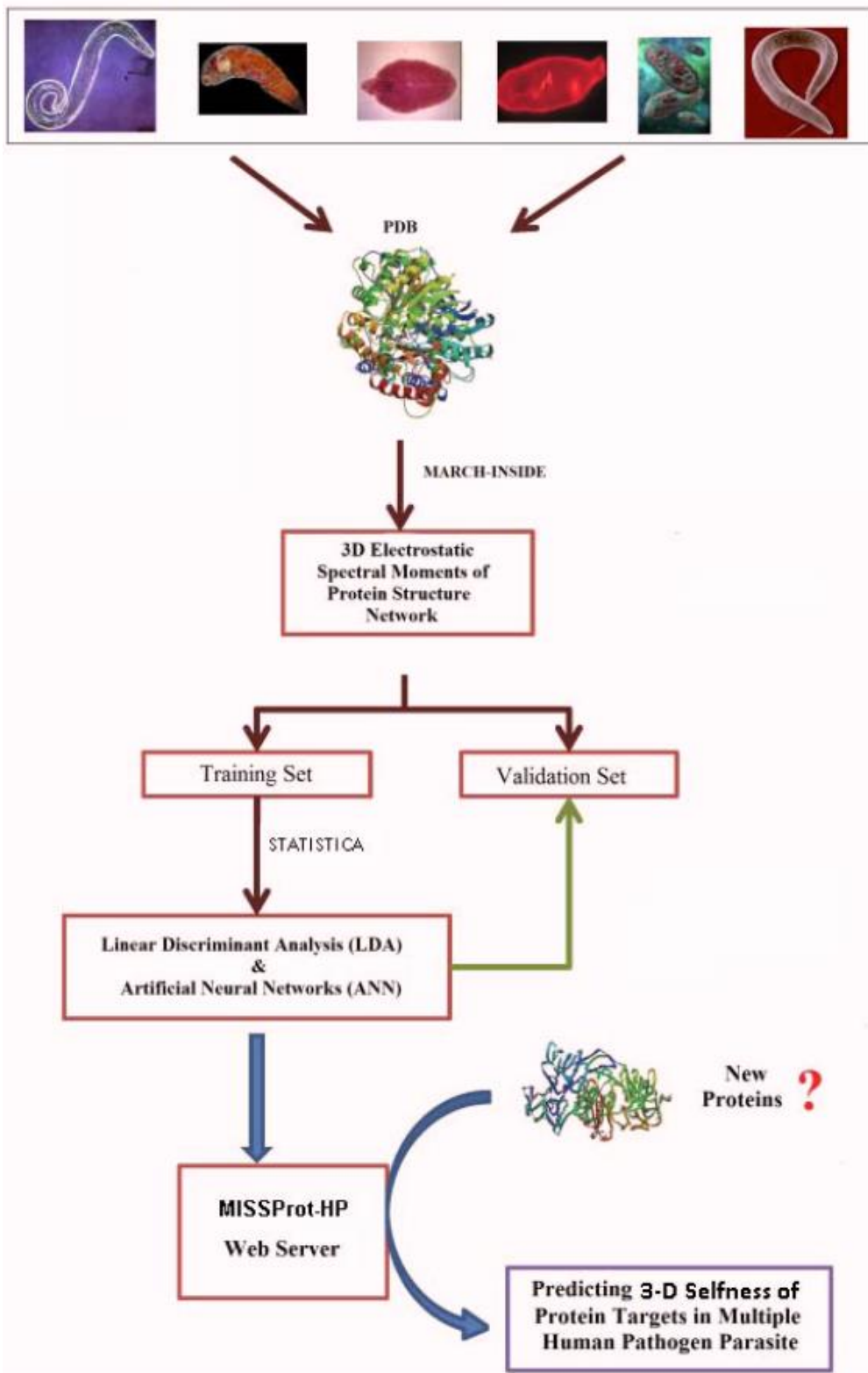


Fig. (7). Workflow used for the development of the MIANN server MISSProt-HP.

We implemented this model in a new web-server called MISS-Prot (MARCH-INSIDE Scores for Self-Proteins). MISS-Prot was programmed using PHP/HTML/Python and MARCH-INSIDE routines and is freely available at: <http://bio-aims.udc.es/MISSProt-HP.php>. This server is easy to use by non-experts in Bioinformatics who can carry out automatic online upload and prediction with 3D structures deposited at PDB (mode 1). We can also study outcomes of Peptide Mass Fingerprinting (PMFs) and MS/MS for query proteins with unknown 3D structures (mode 2). We illustrated the use of MISS-Prot in experimental and/or theoretical studies of peptides from *Fasciola hepatica* cathepsin proteases or present in 10 *Anisakis simplex* allergens (Ani s 1 to Ani s 10). In doing so, we combined electrophoresis (1DE), MALDI-TOF Mass Spectroscopy, and MASCOT to seek sequences, Molecular Mechanics + Molecular Dynamics (MM/MD) to generate 3D structures and MISS-Prot to predict PSP scores. MISS-Prot also allows the prediction of PSP proteins in 16 additional species including parasite hosts, fungi pathogens, disease transmission vectors, and biotechnologically relevant organisms. (Fig. 8) shows the online user interface of the MIANN server.

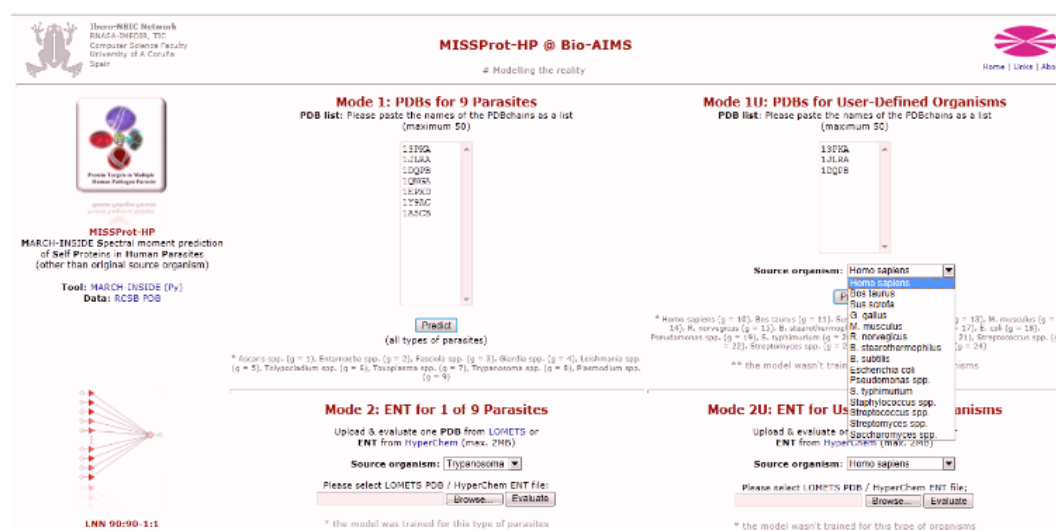


Fig. (8). Online user interface of the MIANN server MISSProt-HP.

4.3. Example of a MIANN Model for Enzyme Class Prediction

4.3.1 EnzClassPred Server

The server EnzClassPred is available in the Bio-AIMS portal (<http://bio-aims.udc.es/EnzClassPred.php>). This MIANN server is a free on-line public tool based on PHP/HTML/Python and MARCH-INSIDE routines. The server is based on the models that we have published before [128]. The input of this server is a PDB id, it can accept maximum 50 entries; for each of them it checks whether the structure has an annotation in the PDB, and if the annotation is present it reports the annotation of the PDB; otherwise it uses the model we have developed and makes a prediction of the structure. (Fig. 9) shows the online user interface of the MIANN server EnzClassPred. When the calculation is launched the user is redirected to the output page; on the top of the output page there is a direct link to the pdb page of each structure, the resume of the model and the EC classification. Below the classification there is a box where it can be specified whether the protein is present in the PDB or not, as already that if the protein is not met, the server activates the model developed earlier thus giving a prediction on the possible classification of the structure. To seek the MIANN model, we have calculated the 3D entropy and

spectral moments of 4,755 proteins (859 enzymes and 3,896 non-enzymes); all these structures were retrieved from the PDB and divided into a positive series of enzymes and a negative series of non-enzymes. The LDA model found is very simple (three variables) and at the same time is able to predict the first EC number with an overall accuracy of 79% for our data set divided into both training and external validation series. In addition, the best non-linear ANN model is notably more complex but has an overall accuracy of >98%.

Ibero-NBIC Network
RNASA-IMEDIR, TIC
Computer Science Faculty
University of A Coruña
Spain

EnzClassPred @ Bio-AIMS
Modelling the reality

Home | Links | About

PDB list : Please paste the names of the PDB as a list (maximum 50)

1XS0
1A0M
1EP9

EnzClassPred
Enzyme Class Prediction

Tool: MARCH-INSIDE (Python version)
Data: RCSB PDB

MLP 4:4-9-8-1:1
(programed by Humberto González-Díaz)

Inputs = 4
Neurons in 1st hidden layer = 9
Neurons in 2nd hidden layer = 8
Output = 1
Training/Members = BP100,CG20,CG4b
Test Accuracy = 98.57%
(Training Accuracy = 98.74%)

Predict
(All enzyme classes)

Fig. (9). Online user interface of the MIANN server EnzClassPred.

5. MIANN MODELS IN PHYSICAL CHEMISTRY

5.1. MIANN QSPR Model of Self-Assembly in Surfactants

5.1.1. Short Introductory Note

Molecular self-assembly is the process through which single molecules arrange themselves spontaneously into different structures [129]. One of the most ubiquitous self-assembly processes in Physical Chemistry is the hierarchical organization of amphiphilic molecules into a huge variety of patterns as micelles, rods or liposomes among others [130, 131]. Such property has recently been employed to design and fabricate for a wide range of biotechnological applications because of their relatively simple structures and easy scale up commercial productions [132]. Some drugs exhibit amphiphile behavior, they tend to self-assembly, usually in a small aggregation number, when dispersed in aqueous solution in a surfactant-like manner [133]. Although drug micelles normally form at concentrations well above the concentration of the drug appearing in body systems, micelles may be present in the pharmaceutical formulation to overcome different challenges including poor bioavailability, stability, side effects or plasma fluctuations of drugs. Imipramine hydrochloride is one of the main tricyclic antidepressant drugs. It is used in clinics as an antidepressant and antipsychotic drug. It possesses a hydrophobic nitrogen-containing heterocycle bound to a short chain carrying a charged amino group [134]. It is widely recognized that the current experimental resources will not be adequate for the vast range of new drug molecules synthesized every day. An especially desirable feature for the pharmaceutical industry is

the ability to know *a priori* and *in situ* the self-assembly properties of these compounds under different conditions.

5.1.2. Materials

Imipramine, 3-(10,11-dihydro-5*H*-dibenzo[*b,f*]azepin-5-yl)-*N,N*-dimethylpropan-1-amine hydrochloride of at least 98.5% purity were purchased from Sigma Chemical Co. and used without further purification. The drug was used as received. All measurements were performed using distilled water with conductivity below $3 \mu\text{S}\cdot\text{cm}^{-1}$ at 298.15 K.

5.1.3. Experimental Methods

Electrical conductivities were measured using a Kyoto Electronics conductometer model CM-117 with a K-121 cell type. The cell constant was determined using KCl solutions. All measurements were taken in a PolyScience Model PS9105 thermostated water bath, at a constant temperature within ± 0.05 K. The determination of the isotherms of conductivity was carried out by continuous dilution of a concentrated sample prepared by weight. The expected duration of the dynamics processes varies from 10^{-8} to 10^{-2} s, i.e., between typical aggregate-solvent exchange time of amphiphilic molecules and typical fusion time of the corresponding aggregates. Thus, the equilibrium is guaranteed a few seconds after dilution [135]. Dynamic light scattering measurements were made at 298.0 ± 0.1 K and at a scattering angle of 90° . Time correlation was analyzed by an ALV-5000 (ALV-GmbH) instrument with vertically polarized incident light of wavelength $\lambda = 488$ nm supplied by a CW diode-pumped Nd:YAG solid-state laser (Coherent, Inc.) operated at 400 mW. Data were analyzed to determine diffusion coefficients using the CONTIN software packages. Hydrodynamic radii were calculated from measuring diffusion coefficients by means of the Stokes-Einstein equation.

5.1.4. Experimental Results

For the studied temperature range, the electrical conductivity concentration dependence shows a monotonic increase with a slight gradual decrease in slope, as expected for self-assembly processes [136, 137]. Critical micellar concentrations (cmc) values were calculated by fitting the experimental isotherms to the non-linear function obtained by direct integration of a Boltzmann-type sigmoid function [138]. The cmc obtained by this method as a function of temperature show the typical U-shaped behavior observed for the critical micelle concentration of surfactant molecules, see (Fig. 10).

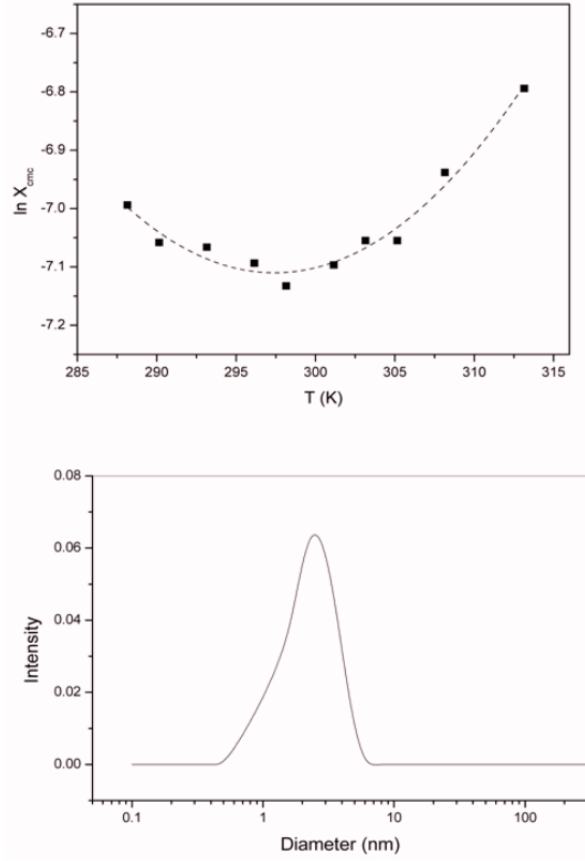


Fig. (10). Top: Fit of the temperature dependence of $\ln x_{cmc}$ using the equation (16); and Botton: Size distribution of a solution of imipramine with a concentration twice the cmc.

For those systems, the cmc decrease with temperature increase like that observed in (Fig. 10) at $T < 297$ K- is typically explained by the net rupture of structured water surrounding the hydrophobic moiety, thus promoting aggregation. The opposite trend at higher temperatures is typically due to the dominant effect coming from the dehydration of the charged head groups accompanied of the subsequent electrostatic repulsion [139]. Thermodynamic parameters corresponding to the aggregation process can be obtained by analyzing the cmc dependence on temperature by means of [140]:

$$\ln x_{cmc} = \ln x_{cmc}^* \left\{ 1 + \beta_1 T^* \left[\frac{1}{2 - \beta} \left(\frac{T^*}{T} - 2 - \frac{\beta_0 - 2}{\beta_1 T^*} \right) + \frac{1}{2 - \beta^*} \left(1 + \frac{\beta_0 - 2}{\beta_1 T^*} \right) \right] \right\} + \frac{\Delta C_{P,m}^0}{(2 - \beta)R} \left(1 - \frac{T^*}{T} - \ln \frac{T}{T^*} \right) + \frac{\alpha}{(2 - \beta)R} \left(\frac{T^{*2} - T^2}{2T} + T^* \ln \frac{T}{T^*} \right) \quad (16)$$

where the degree of ionization dependence on temperature, $\beta = \beta_0 + \beta_1 T$ and the standard change in isobaric molar heat capacity, $\Delta C_{P,m}^0 = \Delta C_{P,m}^0 + \alpha(T - T^*)$, are explicitly considered. $\ln x_{cmc}^*$ and T^* correspond to the minimum value. (Fig. 10) shows the fair fit of equation (16) to experimental data. Results obtained from the fitting were: 297.4 K, -7.11, -363.41 J mol⁻¹ K⁻¹ and -2.67 for T^* , $\ln x_{CAC}^*$, $\Delta C_{P,m}^0$ and α / R respectively. The standard enthalpic and entropic changes due to aggregates formation at the minimum value are given by:

$$\Delta H_m^0 = RT^{*2} \beta_1 \ln x_{CAC}^* \quad [17]$$

$$\Delta S_m^{0*} = \frac{\Delta H_m^{0*}}{T^*} \left(2 + \frac{\beta_0 - 2}{\beta_1 T^*} \right) \quad (18)$$

The dependence of the thermodynamic functions on temperature is obtained from the following expressions:

$$\Delta H_m^0 = \Delta H_m^{0*} + \Delta C_{p,m}^{0*} (T - T^*) + \left(\frac{\alpha}{2} \right) (T - T^*)^2 \quad (19)$$

$$\Delta S_m^0 = \Delta S_m^{0*} + \Delta C_{p,m}^{0*} \ln(T/T^*) + \alpha \{ T - T^* - T^* \ln(T/T^*) \} \quad (20)$$

And the standard Gibbs energy of aggregates formation is given by:

$$\Delta G_m^0 = \Delta H_m^0 - T \Delta S_m^0 \quad (21)$$

Heat capacity, enthalpy, entropy and standard free energy data corresponding to the aggregation process are listed in (Table 1). ΔH_m^0 and ΔS_m^0 are quite sensitive to temperature. ΔH_m^0 values indicate that the imipramine aggregation is increasingly more exothermic at higher temperatures. Negative ΔH_m^0 values suggest the importance of London-dispersion interactions as the major force for aggregation. However, ΔS_m^0 decreased with temperature and remained positive. This aggregation is entropic driven at low temperatures whereas enthalpic contributions become more important at high temperatures. The higher order of water molecules around hydrocarbon chains at lower temperatures could explain this. This fact has been observed for other drugs [141, 142]. In (Table 1) we show the values of critical micelle concentration (cmc in mol kg⁻¹), heat capacity, $\Delta C_{p,m}^0$ (J mol⁻¹ K⁻¹), ΔH_m^0 (J mol⁻¹), ΔS_m^0 (J mol⁻¹ K⁻¹), and standard free energy, ΔG_m^0 (J mol⁻¹) obtained for imipramine at different temperatures, T (K).

Table 1. Values of Different Parameters Obtained for Imipramine at Different Temperatures

| T (K) | cmc | $\Delta C_{p,m}^0$ | ΔH_m^0 | ΔS_m^0 | ΔG_m^0 | β | S(nm ²) | N |
|---------------------------------|----------|--------------------|----------------|----------------|----------------|----------|---------------------|----------|
| 288.2 | 0.051 | 157.71 | 11696 | 51.32 | 26483 | 0.448 | 15.7632 | 9 |
| 290.2 | 0.0478 | 202.09 | 12073 | 50.01 | 26582 | 0.4534 | 15.7632 | 9 |
| 293.2 | 0.0475 | 268.66 | 12559 | 47.48 | 26479 | 0.4615 | 15.7632 | 9 |
| 296.2 | 0.0462 | 335.23 | 13649 | 44.31 | 26771 | 0.4696 | 15.7632 | 9 |
| 298.2 | 0.0444 | 379.61 | 14376 | 41.83 | 26849 | 0.475 | 15.7632 | 9 |
| 301.2 | 0.046 | 446.18 | 15467 | 37.6 | 26790 | 0.4831 | 15.7632 | 9 |
| 303.2 | 0.048 | 490.56 | 16193 | 34.44 | 26632 | 0.4885 | 15.7632 | 9 |
| 305.2 | 0.0481 | 534.94 | 16920 | 31 | 26379 | 0.4939 | 15.7632 | 9 |
| 308.2 | 0.0539 | 601.51 | 18010 | 25.34 | 25819 | 0.502 | 15.7632 | 9 |
| 313.2 | 0.0623 | 712.46 | 19827 | 14.61 | 24402 | 0.5155 | 15.7632 | 9 |
| $S_{ij}(Z_{ij} > 0)$ | | | | | | | | |
| T(K) | cmc | $\Delta C_{p,m}^0$ | ΔH_m^0 | ΔS_m^0 | ΔG_m^0 | β | S(nm ²) | N |
| 288.2 | 0.05311 | -0.0106 | 0.0322 | 0.02195 | 0.03427 | 0.06583 | 0.07995 | 0.05924 |
| 290.2 | 0.0285 | -0.0352 | 0.00759 | -0.0027 | 0.00967 | 0.04123 | 0.05534 | 0.03463 |
| 293.2 | -0.0084 | -0.0721 | -0.0293 | -0.0396 | -0.0272 | 0.00432 | 0.01843 | -0.0023 |
| 296.2 | -0.0453 | -0.109 | -0.0662 | -0.0765 | -0.0642 | -0.0326 | -0.0185 | -0.0392 |
| 298.2 | -0.0699 | -0.1336 | -0.0908 | -0.1011 | -0.0888 | -0.0572 | -0.0431 | -0.0638 |
| 301.2 | -0.1068 | -0.1705 | -0.1277 | -0.138 | -0.1257 | -0.0941 | -0.08 | -0.1007 |
| 303.2 | -0.1314 | -0.1952 | -0.1524 | -0.1626 | -0.1503 | -0.1187 | -0.1046 | -0.1253 |
| 305.2 | -0.1561 | -0.2198 | -0.177 | -0.1872 | -0.1749 | -0.1433 | -0.1292 | -0.1499 |
| 308.2 | -0.193 | -0.2567 | -0.2139 | -0.2241 | -0.2118 | -0.1802 | -0.1661 | -0.1868 |
| 313.2 | -0.2545 | -0.3182 | -0.2754 | -0.2856 | -0.2733 | -0.2418 | -0.2276 | -0.2484 |
| $\langle \chi \rangle$ (anion) | -0.1492 | 0.794355 | 0.08703 | 0.17656 | 0.0172 | -0.24795 | -0.50690 | -0.3305 |
| $\langle \chi \rangle$ (cation) | -0.13596 | 0.266749 | 0.06963 | 0.19678 | 0.09557 | -0.30615 | -0.34548 | -0.08415 |

Dynamic light-scattering measurements were performed to obtain a size distribution of the aggregates. The concentration chosen was twice the cmc, this way we ensured the presence of aggregates. The correlation functions from dynamic light scattering were analyzed by the Contin method. Polydispersity indices generated by this analytical method were less than 0.1, indicative of a reasonable degree of monodispersity of size. (Fig. 2) shows the size distribution obtained. It is a single peak which corresponds to an aggregate with a mean diameter of 2.24 nm. Aggregation number can be estimated from this value. Assuming that the aggregates have spherical shape, the aggregation numbers are obtained from the total surface area of the aggregates and the minimum area per molecule 1.73 nm² (obtained from surface tension measurements [143]). Thus, we obtain a mean value of 8 imipramine molecules per aggregate. This value correlates well with that obtained from static light scattering, 7 [134].

5.1.5. The MIANN Models of Multiple Self-Assembly Properties of Surfactants

In this work we have tried to obtain for the first time a linear model able to predict the probability with which a compound (i-th) presents an experimental value of the property (j-th) higher than the average value of this property for this data set. The best model found using LDA was the following:

$$S_i(Z_{ij} > 0) = 3,91375 + 0,011557 \cdot e^{C_{salt} (mol \cdot L^{-1})} - 0,012303 \cdot T(K)_j + 0,244894 \cdot {}^5\chi(anion)_i - 0,047320 \cdot {}^5\chi(cation)_i + 1,919078 \cdot {}^5\langle\chi\rangle(anions)_j + 3,436077 \cdot {}^5\langle\chi\rangle(cations)_j \quad n = 1085 \quad \chi^2 = 221,2429 \quad p(\chi^2) < 0.01 \quad (22)$$

In this equation, $S_{ij}(Z_{ij} > 0)$ is a real-valued score (output of the linear model) that can be used to discriminate between surfactants with high probability $p(Z_{ij} > 0)$ of showing a high $Z_{ij} > 0$ and others with high probability $p(Z_{ij} \leq 0)$ of presenting a $Z_{ij} \leq 0$. Considering that Z_{ij} is a standardization coefficient used to scale the values of the different properties to a single adimensional scale, we calculated this parameter as follows: $Z_{ij} = (Y_{ij} - \langle Y \rangle_j) / SD_j$. Where, Y_{ij} is the value of the j-th property of the i-th surfactant; $\langle Y \rangle_j$ is the average value of its property in the data set and SD_j is the standard deviation. These parameters allowed us to classify the surfactants as “active” $\Rightarrow C = 1 \Rightarrow Z_{ij} > 0 \Rightarrow Y_{ij} > \langle Y \rangle_j$ or “non-active” $\Rightarrow C = -1 \Rightarrow Z_{ij} \leq 0 \Rightarrow Y_{ij} \leq \langle Y \rangle_j$, that is, surfactants with a higher or lower-than-the-average value of the observed property j-th. The independent terms of the equation are: C_{salt} is the concentration of the salt in the solvent used; ${}^5\chi$ is the average electronegativity of the surfactant for the anion or cation part; and $\langle \chi \rangle$ are the average value of χ for those surfactants with $Z_{ij} > 0$. These terms were calculated using the MARCH-INSIDE software. In addition, n is the number of cases used to train the model and χ^2 is Chi-square statistics with a given p-level = $p(\chi^2)$. (Table 1) shows the overall results obtained for this model in training and validation series. We used this linear model to predict the behavior of imipramine in other experimental conditions, different from those used in the previous experiments. In the bottom part of (Table 2) we found the values of $S_{ij}(Z_{ij} > 0)$ predicted with this model for imipramine NaCl at 0.1 mol·Kg⁻¹.

Table 2. Results of the LDA vs. MIANN Analysis for Multiple Properties of Surfactants

| MIANN model | Train | | | | Statistic Parameter | Validation | | |
|--------------------------|-----------------|-------|---------|------|---------------------|------------|---------|-------|
| | C _{ij} | Total | Correct | (%) | | (%) | Correct | Total |
| LDA | 1 | 403 | 293 | 72.7 | Sensitivity | 73.7 | 98 | 133 |
| | 0 | 682 | 492 | 72.1 | Specificity | 72.8 | 166 | 228 |
| PNN 6:6-1085-2-2:1 | 1 | 403 | 56 | 13.9 | Sensitivity | 15.0 | 20 | 133 |
| | 0 | 682 | 681 | 99.9 | Specificity | 100.0 | 228 | 228 |
| MLP 6:6-8-1:1 (MLP1) | 1 | 403 | 340 | 84.4 | Sensitivity | 85.7 | 114 | 133 |
| | 0 | 682 | 527 | 77.3 | Specificity | 77.2 | 176 | 228 |
| MLP 6:6-10-1:1 (MLP2) | 1 | 403 | 347 | 86.1 | Sensitivity | 83.5 | 111 | 133 |
| | 0 | 682 | 529 | 77.6 | Specificity | 76.3 | 174 | 228 |
| RBF 6:6-1-1:1 | 1 | 403 | 272 | 67.5 | Sensitivity | 69.2 | 92 | 133 |
| | 0 | 682 | 410 | 60.1 | Specificity | 62.3 | 142 | 228 |
| LNN 6:6-1:1 | 1 | 403 | 297 | 73.7 | Sensitivity | 75.2 | 100 | 133 |
| | 0 | 682 | 452 | 66.3 | Specificity | 67.1 | 153 | 228 |

Table 3. Results of the MIANN Analysis One Enantioselective Reactions Network

| MIANN Model profile | Diff-eeR(%) ≥ 100 | Training | | | Statistic Parameter | CV | | |
|----------------------|-----------------------|----------|---------|------|---------------------|------|---------|-------|
| | | Total | Correct | % | | % | Correct | Total |
| MLP1 10:10-8-1:1 | yes | 2683 | 2498 | 93.1 | Sensitivity | 92.6 | 1260 | 1361 |
| | no | 14721 | 13617 | 92.5 | Specificity | 93.1 | 6831 | 7341 |
| MLP1 10:10-11-1:1 | yes | 2683 | 2544 | 94.8 | Sensitivity | 93.8 | 1277 | 1361 |
| | no | 14721 | 13800 | 93.7 | Specificity | 94.2 | 6918 | 7341 |
| MLP1 10:10-12-1:1 | yes | 2683 | 2546 | 94.9 | Sensitivity | 93.7 | 1275 | 1361 |
| | no | 14721 | 13810 | 93.8 | Specificity | 94.3 | 6921 | 7341 |
| MLP2 10:10-7-6-1:1 | yes | 2683 | 2513 | 93.7 | Sensitivity | 93.5 | 1272 | 1361 |
| | no | 14721 | 13695 | 93.0 | Specificity | 93.6 | 6874 | 7341 |
| MLP2 10:10-10-11-1:1 | yes | 2683 | 2508 | 93.5 | Sensitivity | 93.2 | 1268 | 1361 |
| | no | 14721 | 13645 | 92.7 | Specificity | 93.3 | 6850 | 7341 |
| RBF 10:10-1-1:1 | yes | 2683 | 1796 | 66.9 | Sensitivity | 66.1 | 899 | 1361 |
| | no | 14721 | 9846 | 66.9 | Specificity | 67.4 | 4949 | 7341 |
| MLP2 10:10-9-12-1:1 | yes | 2683 | 2531 | 94.3 | Sensitivity | 93.5 | 1272 | 1361 |
| | no | 14721 | 13762 | 93.5 | Specificity | 93.9 | 6891 | 7341 |
| LNN 10:10-1:1 | yes | 2683 | 1879 | 70.0 | Sensitivity | 71.4 | 972 | 1361 |
| | no | 14721 | 10461 | 71.1 | Specificity | 69.8 | 5121 | 7341 |

Next, we carried out an ANN analysis of the data set used before in order to obtain MIANN models. As usual, we carried out AI experiments with different ANN topologies. (Table 4) shows the most interesting results found. As we can note, some MIANN models improve the results obtained by LDA. For instance, the MIANN model MLP1 with profile MLP 6:6-8-1:1 improves LDA in terms of Sensitivity both in Training and external Validation series. Sensitivity of LDA was $\approx 72\%$ whereas this MIANN model presented a Sensitivity $> 84\%$ in both series. It means that MIANN is able to improve LDA by 10 percent points. However, this is considered at cost of complicating the linear model, transforming it into a non-linear ANN with $H_1 = 8$ (hidden neurons). (Table 2) depicts the overall results found with different MIANN models as well. Last, (Fig. 12) depicts the topology of the MLP1 model and the AUROC values of some of the MIANN models trained here. Notably, both the MLP1 and MLP2 have an AUROC = 0.89. However, it is well-known that MLP2 needs a second hidden layer of neurons, 10 more neurons in the present case (profile MLP 6:6-10-1:1), which complicates the model to gain nothing in terms of Sensitivity or Specificity with respect to MLP1, see (Table 2).

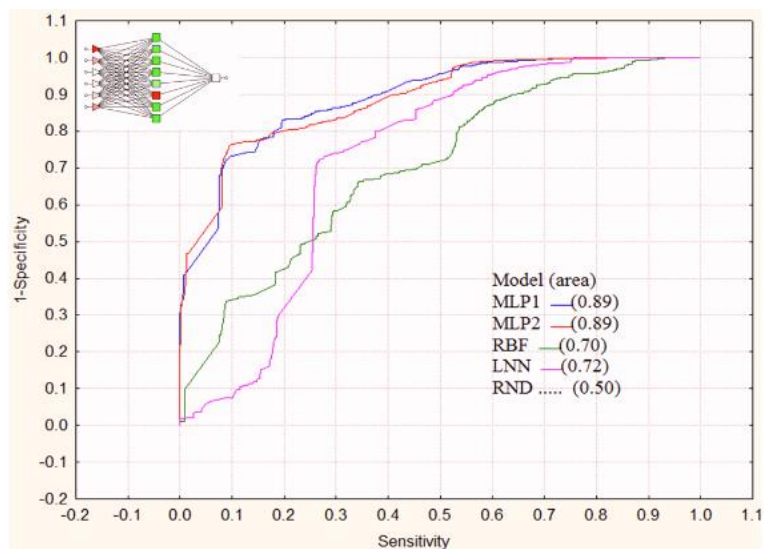


Fig. (11). Topology of the ANN MLP1 and AUROC plots of MIANN models for surfactants.

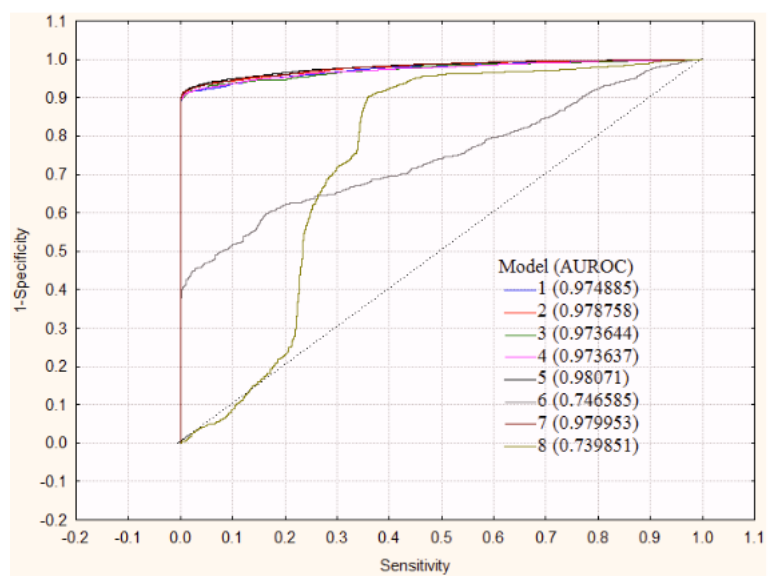


Fig. (12). AUROC analysis of the MIANN models obtained for the reaction network.

6. MIANN MODELS IN ORGANIC CHEMISTRY

6.1. MIANN Approach to QSRR Models of Reaction Networks

6.1.1. Short Introductory Note to QSRR Models

The asymmetric 1,2-addition of organometallic reagents to imines is a powerful tool to form carbon-carbon bonds. Thus, it is possible to introduce a new stereogenic center in organic molecules [144-153], providing ready access to enantiomerically enriched amines with a stereogenic center at the α -position, an important structural feature in many biologically active compounds. These optically active amines are also important compounds because of their broad range of applications such as chiral auxiliaries, resolving agents and building blocks for the synthesis of natural and unnatural compounds, and their pharmacological properties [154-158]. In this kind of reactions many variables, substrates, organo-lithium reagents, chiral ligands, products and variables of reaction condition are involved. Therefore, there is a huge field of possible reactions to investigate.

Quantitative Structure-Reactivity Relationships (QSRR) studies, based on molecular descriptors of chemical structure, may play an important role in the prediction of biological activity or specific property of a reaction. For example, the authors of ref. [159] describe QSRR for kinetic chain-transfer constants for 90 agents on styrene polymerization at 60° C in which three- and five-parameter correlations were obtained with R^2 of 0.725 and 0.818, respectively. Other scientists propose substructural fragments as a simple and safe way to encode molecular structures in a matrix containing the occurrence of fragments of a given type [160]. Satoh and coworkers investigated a dataset of 131 reactions focusing on the changes of electronic features on the oxygen atoms at the reaction sites by principal component analysis and self-organizing neural networks analyses [161]. On the other hand, Long and Niu developed Quantitative Structure-Property Relationship (QSPR) for rate constants (k) of alkyl naphthalene reactions with chlorine, hydroxyl and nitrate radicals using partial least squares (PLS) regression [162]. Different scientists investigated the prediction of oxidoreductase-catalyzed reactions based on atomic properties of metabolites [163]. As above-mentioned, QSRR models may be used to predict effect of changes in reaction variables over enantioselectivity but we also need tools to describe the huge amount of information generated. This sort of problem may be investigated using reaction networks to regroup reactions with inverse results in which the enantiomeric excess and configuration are changed from R to S .

In a previous work [164] we have constructed a MLR-QSRR equation to investigate which variables influence more strongly the change on enantioselectivity. The most important variables were the differences between the initial and final reaction for: product partition coefficient (ΔP_p), chiral ligands hardness (ΔH_l), solvent dipolar moment (ΔD_s), reaction time (Δt_r), reaction temperature (ΔT_r), addition temperature (ΔT_a), average enantiomeric excess for reactions using the same procedure (ΔA_e), substrate molar refractivity (ΔM_i), and steric constant (ΔS_o) and hardness of organolithium compounds (ΔP_o), respectively. Using these variables, the best model found was:

$$\begin{aligned} \Delta ee(R)\% &= -6.60 + 5.80 \cdot \Delta P_p - 4.63 \cdot \Delta H_l - 23.08 \cdot \Delta D_s + 44.18 \cdot \Delta t_r - 1.23 \cdot \Delta T_r \\ &\quad - 0.18 \Delta T_a + 0.24 \Delta A_e + 1.90 \cdot \Delta S_o - 8.22 \cdot \Delta P_o - 0.24 \cdot \Delta M_i \quad n \\ &= 17404 \quad R^2 = 0.803 \quad R^2_{adjusted} = 0.803 \quad F = 7120.7 \quad p < 0.00001 \end{aligned} \quad (22)$$

where, n is the number of cases (reaction pairs) used to train the model, R^2 and $R^2_{adjusted}$ are the training and adjusted square regression coefficients, F is Fisher ratio, and p the level of error. All these reactions have been previously reported in the literature [165-177]. This model, with ten variables, predicts correctly 80.3% of variance of the data set with a standard error of 29.35%. Notably, the adjusted values of R^2 and $R^2_{adjusted}$ are equal, which indicates that the model is not over-fitted due to incorporating an elevated number of parameters.

In order to achieve the enantiomeric excess and configuration of the product with a network approach where one node represents a reaction and the edges show reaction pairs with high propensity to *R/S* chirality inversion, we carried out the following steps. First, we calculated the observed and QSRR-predicted average-scores that numerically characterize the propensity of one reaction to yield *R/S* chirality inversion. These scores were labelled as Obs. Avg.Δee(*R*)% and Pred.Avg.Δee(*R*)%:

$$Obs. Avg. \Delta ee(R)\%_v = \frac{1}{228} \sum_{w=1}^{w=228} \Delta ee(R)\%_{obs.}(v, w) = \frac{1}{228} \sum_{w=1}^{w=228} (ee(R)\%_{obs}(v) - ee(R)\%_{obs}(w)) \quad (23)$$

$$Pr ed. Avg. \Delta ee(R)\%_v = \frac{1}{228} \sum_{w=1}^{w=228} \Delta ee(R)\%_{pred.}(v, w) = \frac{1}{228} \sum_{w=1}^{w=228} (ee(R)\%_{pred}(v) - ee(R)\%_{pred}(w)) \quad (24)$$

Where, Obs.Avg.Δee(*R*)%_v is the difference between observed *R* enantiomeric excess for reaction *v* minus observed *R* enantiomeric excess for reaction *w* and Pred.Avg. Δee(*R*)% is the difference between predicted *R* enantiomeric excess for reaction *v* minus observed *R* enantiomeric excess for reaction *w*. Then, we used these scores as inputs in a Microsoft-Excel sheet to calculate the elements of the Boolean or Adjacency matrix (*A*) associated to the reaction network as follows:

$$A = \begin{cases} \text{if} & sign(ee(R)\%_{obs}(v)) = sign(ee(R)\%_{obs}(w)) & \text{then } \alpha_{vw} = 0 \\ \text{else} & & \\ \text{if} & [Obs Avg \Delta ee(R)\%_v - Obs Avg \Delta ee(R)\%_v] \leq cut - off & \text{then } \alpha_{vw} = 0 \\ \text{else} & & \alpha_{vw} = 1 \end{cases} \quad (25)$$

In order to validate the model we used it to predict 26,106 reactions pairs never used to train the model (validation series). In this series the results were: $R^2 = 79.98\%$, $F = 1043 \cdot 10^2$ and $p < 0.00001$. The model explains correctly 80.0% of variance of the data set with a standard error of 29.79% in the validation series. These results indicate that we developed an accurate model according to previous reports on the use of MLR in QSRR [178-180]. In this previous work, we constructed the observed reaction using the observed values, considering the experimental data. Next, we predicted the reaction network with the QSRR model and last we compared both networks. In order to compare them, we used the sensitivity, specificity and accuracy by means of a Chi-Square test. The value obtained for the $p < 0.00001$ was with Chi-square = 293.364, demonstrating that both networks are very similar.

6.1.2. MIANN-QSRR Model of Reaction Networks for 1,2-Addition of Organometallic Reagents to Imines

We propose herein, for the first time, a MIANN-QSRR model able to predict the difference in enantiomeric excess for *R*-product between two pairs of reactions (Δee(*R*)%) in the reaction network, which reaches similar/dissimilar enantioselectivity after the modification of reaction variables. These QSRR models combine the MI and ANN software to predict the configuration of the new stereogenic center formed in the synthesis of amines, taking into consideration similar reaction pairs in which the enantiomeric excess increases or reduces.

6.1.3. MI Methods Used

The MIANN model used here is based on the calculation of the different physicochemical molecular properties (λ_m) for substrates, organolithium reagents, chiral ligands and products (λ_s , λ_o , λ_i , λ_p), respectively. These λ_m are calculated as an average of atomic properties (λ_j). For instance, it is possible to derive average estimations of refractivities (MR_s , MR_o , MR_i , MR_p), partition coefficients (P_s , P_o , P_i , P_p), and hardness (η_s , η_o , η_i , η_p) that we are going to use in this work, as seen in the equation below [85]:

$$\lambda_m = \frac{1}{6} \sum_{k=0}^5 {}^k\lambda = \frac{1}{6} \sum_{k=0}^5 \sum_j p_k(\lambda_j) \cdot \lambda_j \quad (26)$$

It is possible to consider isolated atoms ($k = 0$) in the estimation of the molecular properties ${}^0\eta$, ${}^0\chi$, ${}^0\text{MR}$, ${}^0\alpha$, ${}^0\text{P}$. In this case the probabilities ${}^0p(\lambda_j)$ are determined without considering the formation of chemical bonds (simple additive scheme). However, it is also possible to consider the gradual effects of the neighboring atoms at different distances in the molecular backbone. In order to reach this goal, the method uses an MM, which determines the absolute probabilities ${}^k p(\lambda_j)$ with which the atoms placed at different distances k affect the contribution of the atom j to the molecular property in question.

$${}^k\lambda = [{}^0p(\lambda_1) \ {}^0p(\lambda_2) \ \dots \ {}^0p(\lambda_n)] \begin{bmatrix} {}^1p_{1,2} & {}^1p_{1,2} & {}^1p_{1,3} & \cdot & {}^1p_{1,n} \\ {}^1p_{2,1} & {}^1p_{2,2} & {}^1p_{2,3} & \cdot & {}^1p_{2,n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ {}^1p_{n,1} & \cdot & \cdot & \cdot & {}^1p_{n,n} \end{bmatrix}^k \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \lambda_n \end{bmatrix} = \sum_{j=1}^n {}^k p(\lambda_j) \cdot \lambda_j \quad (27)$$

Where, from left to right, the first term is ${}^k\lambda$, which is the average molecular property considering the effects of all the atoms placed at distance k over every atomic property λ_j . The vector on the left side of the equation contains the probabilities ${}^0p(\lambda_j)$ for every atom in the molecule, without considering chemical bonds. The matrix in the center of the equation is the so-called stochastic matrix. The values of this matrix (${}^1p_{ij}$) are the probabilities with which every atom affects the parameters of the atom bonded to it. Both kinds of probabilities ${}^0p(\lambda_j)$ and ${}^1p_{ij}$ are easily calculated from atomic parameters (λ_j) and the chemical bonding information:

$${}^0p_{ij} = \frac{\lambda_j}{\sum_{k=1}^n \lambda_k} \quad (28)$$

$${}^1p_{ij} = \frac{\delta_{ij} \cdot \lambda_j}{\sum_{k=1}^n \delta_{ik} \cdot \lambda_k} \quad (29)$$

The only difference is that in the probabilities ${}^0p(\lambda_j)$ we consider isolated atoms by carrying out the sum in the denominator over all n atoms in the molecule. On the other hand, for ${}^1p_{ij}$ chemical bonding is taken into consideration by means of the factor δ_{ij} . This factor has the value 1 if atoms i and j are chemically bonded and it is 0 otherwise. All calculations were performed using the MARCH-INSIDE program version 3.0 [181], which can be obtained for free academic use, upon request, from the corresponding author of the present work.

6.1.4. ANN Analysis

Given the λ_m of the above-mentioned molecular parameters and λ_{orv} of the other reaction variables such as (T(a), T(r), t(r)), we can calculate the differences $\Delta\lambda = \lambda_{orv}(r2) - \lambda_{orv}(r1)$ for any reaction pairs. Using these $\Delta\lambda_m$ and $\Delta\lambda_{orv}$ values as input we performed different ANN analyses to fit the QSRR model, including the LNN equation with the form:

$$\Delta ee(R)\%_{pred} = \sum_{s,l,o,p} b_m \cdot \Delta\lambda_m + \sum_{orv} b_{orv} \cdot \Delta\lambda_{orv} + b_0 \quad (30)$$

The parameter $\Delta\%ee(R)_{pred}$ (the prediction of the difference in enantiomeric excess for R -product between two pairs of reactions) is the output of the model. In equation (6), b represents the coefficients of the variables in the model, determined with the ANN module of the STATISTICA 6.0 software package [99]. The data set included a series of reported organolithium addition to imines in presence of chiral ligands reactions.

6.1.5. Results of the MIANN Model

We studied the previous data set with different MIANN approaches: LNN, MLP1, MLP2, and RBF. (Table 3) shows the best results found. The LNN model gives good classification results with Sensitivity (Sn) and Specificity (Sp) higher than 70%, except for Sp in Cross-Validation (CV) series. These results of the MIANN analysis confirm the findings of our previous work. There is a strong linear relationship between the MI parameters studied and the inversion of chirality in the present reaction network, taking into consideration the change of specific reaction parameters as well. In any case, the use of non-linear MIANN strategy notably improved the results found with the linear MIANN model obtained by LNN. More specifically, the model 1 is a MIANN model with topology MLP1 that shows very high (>90%) values of Sn and Sp both in training and cv series, using only one hidden layer with 8 hidden neurons. (Fig. 12) shows the AUROC values of the MIANN model developed for the reaction network.

CONCLUSIONS

The MIANN models may be used to model very different data and reduce experimental costs in different areas of the molecular sciences, including Physical, Organic, and Medicinal Chemistry. In almost all cases the non-linear MIANN models improve the results obtained with a linear method.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

The authors acknowledge Ministry of Science and Innovation (Ministerio de Ciencia e Innovación -MICIN), (CTQ2009-07733), and the University of the Basque Country (UPV/EHU), (UFI11/22, and GIU 0946), for their financial support. We also thank financial support from Xunta de Galicia (10PXIB206258PR). Munteanu CR acknowledges the funding support for the “Isidro Parga Pondal” program from Xunta de Galicia (Spain) and the European Social Fund (ESF). This work is partially supported by the “IberoAmerican Network of the Nano-Bio-Info-Cogno Convergent Technologies”, Ibero-NBIC Network (209RT-0366) funded by CYTED (Spain).

REFERENCES

- [1] González-Díaz, H.; González-Díaz, Y.; Santana, L.; Ubeira, F.M.; Uriarte, E. Proteomics, networks and connectivity indices. *Proteomics*, 2008, 8, 750-778.
- [2] Jorgensen, W.L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.*, 2005, 26(16), 1689-1700.
- [3] Perez-Riverol, Y.; Vera, R.; Mazola, Y.; Musacchio, A. A parallel systematic-monte carlo algorithm for exploring conformational space. *Curr. Top. Med. Chem.*, 2012, 12(16), 1790-1796.
- [4] Audie, J.; Swanson, J. Recent work in the development and application of protein-peptide docking. *Future Med. Chem.*, 2012, 4(12), 1619-1644.
- [5] Speck-Planche, A.; Kleandrova, V.V. QSAR and molecular docking techniques for the discovery of potent monoamine oxidase B inhibitors: computer-aided generation of new rasagiline bioisosteres. *Curr. Top. Med. Chem.*, 2012, 12(16), 1734-1747.
- [6] Biesiada, J.; Porollo, A.; Velayutham, P.; Kouril, M.; Meller, J. Survey of public domain software for docking simulations and virtual screening. *Hum. Genomics*, 2011, 5(5), 497-505.

- [7] Bottegoni, G. Protein-ligand docking. *Front. Biosci.*, 2011, 16, 2289-2306.
- [8] Cavasotto, C.N. Homology models in docking and high-throughput docking. *Curr. Top. Med. Chem.*, 2011, 11(12), 1528-1534.
- [9] Garcia, I.; Fall, Y.; Gomez, G. QSAR, Docking, and CoMFA Studies of GSK3 Inhibitors. *Curr. Pharm. Des.*, 2010, 16(24), 2666-2675.
- [10] Kim, K.H.; Gaisina, I.; Gallier, F.; Holzle, D.; Blond, S.Y.; Mesecar, A.; Kozikowski, A.P. Use of molecular modeling, docking, and 3D-QSAR studies for the determination of the binding mode of benzofuran-3-yl-(indol-3-yl)maleimides as GSK-3 inhibitors. *J. Mol. Model.*, 2009, 15, 1463-1479.
- [11] Estrada, E.; Delgado, E.J.; Alderete, J.B.; Jaña, G.A. Quantum-connectivity descriptors in modeling solubility of environmentally important organic compounds. *J. Comput. Chem.*, 2004, 25(14), 1787-1796.
- [12] Besalu, E.; Girones, X.; Amat, L.; Carbo-Dorca, R. Molecular quantum similarity and the fundamentals of QSAR. *Acc. Chem. Res.*, 2002, 35(5), 289-295.
- [13] Rincon, D.A.; Cordeiro, M.N.; Mosquera, R.A. On the electronic structure of cocaine and its metabolites. *J. Phys. Chem. A*, 2009, 113(50), 13937-13942.
- [14] Mandado, M.; Gonzalez-Moa, M.J.; Mosquera, R.A. Chemical graph theory and n-center electron delocalization indices: a study on polycyclic aromatic hydrocarbons. *J. Comput. Chem.*, 2007, 28(10), 1625-1633.
- [15] Gonzalez-Diaz, H.; Gonzalez-Diaz, Y.; Santana, L.; Ubeira, F.M.; Uriarte, E. Proteomics, networks and connectivity indices. *Proteomics*, 2008, 8(4), 750-778.
- [16] Helguera, A.M.; Combes, R.D.; Gonzalez, M.P.; Cordeiro, M.N. Applications of 2D descriptors in drug design: a DRAGON tale. *Curr. Top. Med. Chem.*, 2008, 8(18), 1628-1655.
- [17] Casanola-Martin, G.M.; Marrero-Ponce, Y.; Khan, M.T.; Ather, A.; Khan, K.M.; Torrens, F.; Rotondo, R. Dragon method for finding novel tyrosinase inhibitors: Biosilico identification and experimental in vitro assays. *Eur. J. Med. Chem.*, 2007, 42(11-12), 1370-1381.
- [18] Tetko, I.V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V.A.; Radchenko, E.V.; Zefirov, N.S.; Makarenko, A.S.; Tanchuk, V.Y.; Prokopenko, V.V. Virtual computational chemistry laboratory--design and description. *J. Comput. Aided Mol. Des.*, 2005, 19(6), 453-463.
- [19] Marzaro, G.; Chilin, A.; Guiotto, A.; Uriarte, E.; Brun, P.; Castagliuolo, I.; Tonus, F.; Gonzalez-Diaz, H. Using the TOPS-MODE approach to fit multi-target QSAR models for tyrosine kinases inhibitors. *Eur. J. Med. Chem.*, 2011, 46(6), 2185-2192.
- [20] Vilar, S.; Estrada, E.; Uriarte, E.; Santana, L.; Gutierrez, Y. In silico studies toward the discovery of new anti-HIV nucleoside compounds through the use of TOPS-MODE and 2D/3D connectivity indices. 2. Purine derivatives. *J. Chem. Inf. Model.*, 2005, 45(2), 502-514.
- [21] Estrada, E.; Quincoces, J.A.; Patlewicz, G. Creating molecular diversity from antioxidants in Brazilian propolis. Combination of TOPS-MODE QSAR and virtual structure generation. *Mol. Divers.*, 2004, 8(1), 21-33.
- [22] Estrada, E.; Gonzalez, H. What are the limits of applicability for graph theoretic descriptors in QSPR/QSAR? Modeling dipole moments of aromatic compounds with TOPS-MODE descriptors. *J. Chem. Inf. Comput. Sci.*, 2003, 43(1), 75-84.
- [23] Marrero-Ponce Y, C.-G.J., Olazabal E, Serrano HS, Morales A, Castañedo N, Ibarra-Velarde F, Huesca-Guillen A, Jorge E, del Valle A, Torrens F, Castro EA. TOMOCOMD-CARDD, a novel approach for computer-aided 'rational' drug design: I. Theoretical and experimental assessment of a promising method for computational screening and in silico design of new anthelmintic compounds. *J. Comput. Aided Mol. Des.*, 2004, 18(10), 615-634.
- [24] Marrero-Ponce, Y.; Medina-Marrero, R.; Castro, A.E.; Ramos de Armas, R.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F. Protein Quadratic Indices of the "Macromolecular Pseudograph's -Carbon Atom Adjacency Matrix". 1. Prediction of Arc Repressor Alanine-mutant's Stability. *Molecules*, 2004, 9, 1124-1147.
- [25] Katritzky, A.R.; Oliferenko, A.; Lomaka, A.; Karelson, M. Six-membered cyclic ureas as HIV-1 protease inhibitors: a QSAR study based on CODESSA PRO approach. Quantitative structure-activity relationships. *Bioorg. Med. Chem. Lett.*, 2002, 12(23), 3453-3457.
- [26] Katritzky, A.R.; Perumal, S.; Petrukhin, R.; Kleinpeter, E. Codessa-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities. *J. Chem. Inf. Comput. Sci.*, 2001, 41(3), 569-574.
- [27] Vilar, S.; Cozza, G.; Moro, S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr. Top. Med. Chem.*, 2008, 8(18), 1555-1572.

- [28] Hill, T.; Lewicki, P. STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining. StatSoft: Tulsa, 2006 [29]
- [29] Frank, E.; Hall, M.; Trigg, L.; Holmes, G.; Witten, I.H. Data mining in bioinformatics using Weka. *Bioinformatics*, 2004, 20(15), 2479-2481.
- [30] Bisson, W.H. Drug repurposing in chemical genomics: can we learn from the past to improve the future? *Curr. Top. Med. Chem.*, 2012, 12(17), 1883-1888.
- [31] Speck-Planche, A.; Cordeiro, M.N. Computer-aided drug design methodologies toward the design of anti-hepatitis C agents. *Curr. Top. Med. Chem.*, 2012, 12(8), 802-813.
- [32] Prado-Prado, F.; Garcia-Mera, X. QSAR models for computer-aided drug design and molecular docking for disorders of the central nervous system and other diseases. *Curr. Top. Med. Chem.*, 2012, 12(16), 1731-1733.
- [33] Gonzalez-Diaz, H. Quantitative studies on Structure-Activity and Structure-Property Relationships (QSAR/QSPR). *Curr. Top. Med. Chem.*, 2008, 8(18), 1554.
- [34] Gonzalez-Diaz, H. Editorial: QSAR/QSPR Models as Enabling Technologies for Drug & Targets Discovery in: Medicinal Chemistry, Microbiology-Parasitology, Neurosciences, Bioinformatics, Proteomics and Other Biomedical Sciences. *Curr. Top. Med. Chem.*, 2012, 12(8), 799-801.
- [35] Alderson, R.G.; De Ferrari, L.; Mavridis, L.; McDonagh, J.L.; Mitchell, J.B.; Nath, N. Enzyme informatics. *Curr. Top. Med. Chem.*, 2012, 12(17), 1911-1923.
- [36] Bisson, W.H. Editorial: computational chemogenomics in drug design and discovery. *Curr. Top. Med. Chem.*, 2012, 12(17), 1867-1868.
- [37] Castillo-Garit, J.A.; Abad, C.; Rodriguez-Borges, J.E.; Marrero-Ponce, Y.; Torrens, F. A review of QSAR studies to discover new drug-like compounds actives against leishmaniasis and trypanosomiasis. *Curr. Top. Med. Chem.*, 2012, 12(8), 852-865.
- [38] Cedeno, W.; Alex, S.; Jaeger, E.P.; Agrafiotis, D.K.; Lobanov, V.S. An integrated data management framework for drug discovery-from data capturing to decision support. *Curr. Top. Med. Chem.*, 2012, 12(11), 1237-1242.
- [39] Chatterjee, A.K.; Yeung, B.K. Back to the future: lessons learned in modern target-based and whole-cell lead optimization of antimalarials. *Curr. Top. Med. Chem.*, 2012, 12(5), 473-483.
- [40] Chen, J.; Wang, Y.; Guo, D.; Shen, B. A systems biology perspective on rational design of peptide vaccine against virus infections. *Curr. Top. Med. Chem.*, 2012, 12(12), 1310-1319.
- [41] Cordero, F.; Beccuti, M.; Donatelli, S.; Calogero, R.A. Large disclosing the nature of computational tools for the analysis of next generation sequencing data. *Curr. Top. Med. Chem.*, 2012, 12(12), 1320-1330.
- [42] Dave, K.; Lahiry, A. Conotoxins: review and docking studies to determine potentials of conotoxin as an anticancer drug molecule. *Curr. Top. Med. Chem.*, 2012, 12(8), 845-851.
- [43] Dave, K.; Panchal, H. Review on chemogenomics approach: interpreting antagonist activity of secreted frizzled-related protein 1 in glaucoma disease with in-silico docking. *Curr. Top. Med. Chem.*, 2012, 12(16), 1834-1842.
- [44] Faivre, C.; Barbolosi, D.; Iliadis, A. A new model for determining the MTD during phase-I trials in pediatric oncology. *Curr. Top. Med. Chem.*, 2012, 12 (15), 1660-1664.
- [45] Garcia, I.; Fall, Y.; Gomez, G. Review of synthesis, biological assay, and QSAR studies of HMGR inhibitors. *Curr. Top. Med. Chem.*, 2012, 12(8), 895-919.
- [46] Jayadeepa, R.M.; Niveditha, M.S. Computational approaches to screen candidate ligands with anti- Parkinson's activity using R programming. *Curr. Top. Med. Chem.*, 2012, 12(16), 1807-1814.
- [47] Khan, M.T.; Mischiati, C.; Ather, A.; Ohyama, T.; Dedachi, K.; Borgatti, M.; Kurita, N.; Gambari, R. Structure-based analysis of the molecular recognitions between HIV-1 TAR-RNA and transcription factor nuclear factor-kappaB (NFkB). *Curr. Top. Med. Chem.*, 2012, 12 (8), 814-827.
- [48] Kobe, B.; Boden, M. Computational modelling of linear motif-mediated protein interactions. *Curr. Top. Med. Chem.*, 2012, 12(14), 1553-1561.
- [49] Kramer, C.; Lewis, R. QSARs, Data and Error in the Modern Age of Drug Discovery. *Curr. Top. Med. Chem.*, 2012, 12(17), 1896-1902.
- [50] Kufareva, I.; Chen, Y.C.; Ilatovskiy, A.V.; Abagyan, R. Compound Activity Prediction Using Models of Binding Pockets or Ligand Properties in 3D. *Curr. Top. Med. Chem.*, 2012, 12(17), 1869-1882.
- [51] Lin, J.H. Target prediction of small molecules with information of key molecular interactions. *Curr. Top. Med. Chem.*, 2012, 12(17), 1903-1910.
- [52] Luan, F.; Borges, F.; Cordeiro, M.N. Recent advances on A(3) adenosine receptor antagonists by QSAR tools. *Curr. Top. Med. Chem.*, 2012, 12(8), 878-894.
- [53] Mortier, J.; Rakers, C.; Frederick, R.; Wolber, G. Computational tools for in silico fragment-based drug design. *Curr. Top. Med. Chem.*, 2012, 12(17), 1935-1943.

- [54] Ortore, G.; Tuccinardi, T.; Martinelli, A. Computational studies on translocator protein (TSPO) and its ligands. *Curr. Top. Med. Chem.*, 2012, 12(4), 352-359.
- [55] Popelier, P. New insights in atom-atom interactions for future drug design. *Curr. Top. Med. Chem.*, 2012, 12(17), 1924-1934.
- [56] Prado-Prado, F.; Garcia-Mera, X.; Escobar, M.; Alonso, N.; Caamano, O.; Yanez, M.; Gonzalez-Diaz, H. 3D MI-DRAGON: New Model for the Reconstruction of US FDA Drug-Target Network and Theoretical-Experimental Studies of Inhibitors of Rasagiline Derivatives for AChE. *Curr. Top. Med. Chem.*, 2012, 12(16), 1843-1865.
- [57] Riera-Fernandez, I.; Martin-Romalde, R.; Prado-Prado, F.J.; Escobar, M.; Munteanu, C.R.; Concu, R.; Duardo-Sanchez, A.; Gonzalez-Diaz, H. From QSAR models of Drugs to Complex Networks: State-of-Art Review and Introduction of New Markov-Spectral Moments Indices. *Current Topics in Medicinal Chemistry*, 2012, 12(8), 927-960.
- [58] Saladino, G.; Gervasio, F.L. New insights in protein kinase conformational dynamics. *Curr. Top. Med. Chem.*, 2012, 12(17), 1889-1895.
- [59] Sharma, N.; Ethiraj, K.R.; Yadav, M.; Nayarisseri, S.A.; Chaurasiya, M.; Vankudavath, R.N.; Rao, K.R. Identification of LOGP Values and Electronegativities As Structural Insights to Model Inhibitory Activity of HIV-1 Capsid Inhibitors - A SVM and MLR Aided QSAR Studies. *Curr. Top. Med. Chem.*, 2012, 12(16), 1763-1774.
- [60] Van Calenbergh, S.; Pochet, S.; Munier-Lehmann, H. Drug design and identification of potent leads against mycobacterium tuberculosis thymidine monophosphate kinase. *Curr. Top. Med. Chem.*, 2012, 12(7), 694-705.
- [61] Zhang, T.; Zhao, M.; Pang, Y.; Zhang, W.; Angela Liu, L.; Wei, D.Q. Recent progress on bioinformatics, functional genomics, and metabolomics research of cytochrome P450 and its impact on drug discovery. *Curr. Top. Med. Chem.*, 2012, 12(12), 1346-1355.
- [62] Gonzalez-Diaz, H.; Romaris, F.; Duardo-Sanchez, A.; Perez-Montoto, L.G.; Prado-Prado, F.; Patlewicz, G.; Ubeira, F.M. Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues. *Curr. Pharm. Des.*, 2010, 16(24), 2737-2764.
- [63] Caballero, J.; Fernandez, M. Artificial neural networks from MATLAB in medicinal chemistry. Bayesian-regularized genetic neural networks (BRGNN): application to the prediction of the antagonistic activity against human platelet thrombin receptor (PAR-1). *Curr. Top. Med. Chem.*, 2008, 8(18), 1580-1605.
- [64] Duardo-Sanchez, A.; Patlewicz, G.; Lopez-Diaz, A. Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues. *Curr. Top. Med. Chem.*, 2008, 8(18), 1666-1675.
- [65] Gonzalez-Diaz, H.; Prado-Prado, F.; Ubeira, F.M. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr. Top. Med. Chem.*, 2008, 8(18), 1676-1690.
- [66] Ivanciuc, O. Weka machine learning for predicting the phospholipidosis inducing potential. *Curr. Top. Med. Chem.*, 2008, 8(18), 1691-1709.
- [67] Wang, J.F.; Wei, D.Q.; Chou, K.C. Drug candidates from traditional chinese medicines. *Curr. Top. Med. Chem.*, 2008, 8(18), 1656-1665.
- [68] Gonzalez-Diaz, 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), 1015-1029.
- [69] Rodriguez-Soca, Y.; Munteanu, C.R.; Dorado, J.; Rabuñal, J.; Pazos, A.; González-Díaz, H. Plasmod-PPI: a web-server predicting complex biopolymer targets in Plasmodium with entropy measures of protein-protein interactions. *Polymer*, 2010, 51(1), 264-273.
- [70] Rivero, D.; Dorado, J.; Rabuñal, J.R.; Pazos, A. Modifying genetic programming for artificial neural network development for data mining. *Soft Comput.*, 2009, 13(3), 291-305.
- [71] Rivero, D.; Dorado, J.; Rabuñal, J.R.; Pazos, A.; Pereira, J. Artificial neural network development by means of genetic programming with graph codification. *Trans. Eng., Comput. Technol.*, 2006, 16, 209-214.
- [72] Rivero, D.; Dorado, J.; Rabuñal, J.; Pazos, A. Using genetic programming for artificial neural network development and simplification. *Proceedings of the 5th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics*, 2006, 65-71.
- [73] Gestal, M.; Gómez-Carracedo, M.P.; Andrade, J.M.; Dorado, J.; Fernández, E.; Prada, D.; Pazos, A. Selection of variables by genetic algorithms to classify apple beverages by artificial neural networks. *Appl. Artif. Intell.*, 2005, 19(2), 181-198.
- [74] Dorado, J.; Rabuñal, J.R.; Rivero, D.; Santos, A.; Pazos, A. In: Automatic recurrent ANN rule extraction with genetic programming, *Proceedings of the Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference; IEEE, 2002; pp 1552-1557.*

- [75] Bhattacharjee, B.; Jayadeepa, R.M.; Banerjee, S.; Joshi, J.; Middha, S.K.; Mole, J.P.; Samuel, J. Review of Complex Network and Gene Ontology in pharmacology approaches: Mapping natural compounds on potential drug target Colon Cancer network. *Curr. Bioinf.*, 2011, 6(1), 44-52.
- [76] Dave, K.; Banerjee, A. Bioinformatics analysis of functional relations between CNPs regions. *Curr. Bioinf.*, 2011, 6(1), 122-128.
- [77] Duardo-Sanchez, A.; Patlewicz, G.; González-Díaz, H. A Review of Network Topological Indices from Chem-Bioinformatics to Legal Sciences and back. *Curr. Bioinf.*, 2011, 6(11), 53-70.
- [78] García, I.; Fall, Y.; Gómez, G. Trends in Bioinformatics and Chemoinformatics of Vitamin D analogues and their protein targets. *Curr. Bioinf.*, 2011, 6(1), 16-24.
- [79] Ivanciuc, T.; Ivanciuc, O.; Klein, D.J. Network-QSAR with Reaction Poset Quantitative Superstructure-Activity Relationships (QSSAR) for PCB Chromatographic Properties. *Curr. Bioinf.*, 2011, 6(1), 25-34.
- [80] Prado-Prado, F.; Escobar-Cubiella, M.; García-Mera, X. Review of Bioinformatics and QSAR studies of -secretase inhibitors. *Curr. Bioinf.*, 2011, 6(1), 3-15.
- [81] Wan, S.B.; Hu, L.L.; Niu, S.; Wang, K.; Cai, Y.D.; Lu, W.C.; Chou, K.C. Identification of multiple subcellular locations for proteins in budding yeast. *Curr. Bioinf.*, 2011, 6(1), 71-80.
- [82] Gonzalez-Diaz, H.; Duardo-Sanchez, A.; Ubeira, F.M.; Prado-Prado, F.; Perez-Montoto, L.G.; Concu, R.; Podda, G.; Shen, B. Review of MARCH-INSIDE & complex networks prediction of drugs: ADMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome biomarkers. *Curr. Drug Metab.*, 2010, 11(4), 379-406.
- [83] 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), 1025-1039.
- [84] Santana, L.; Gonzalez-Diaz, H.; Quezada, E.; Uriarte, E.; Yanez, M.; Vina, D.; Orallo, F. Quantitative structure-activity relationship and complex network approach to monoamine oxidase a and B inhibitors. *J. Med. Chem.*, 2008, 51(21), 6740-6751.
- [85] Santana, L.; Uriarte, E.; González-Díaz, H.; Zagotto, G.; Soto-Otero, R.; Mendez-Alvarez, E. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J. Med. Chem.*, 2006, 49(3), 1149-1156.
- [86] González-Díaz, H.; Pérez-Bello, A.; Uriarte, E. Stochastic molecular descriptors for polymers. 3. Markov electrostatic moments as polymer 2D-folding descriptors: RNA–QSAR for mycobacterial promoters. *Polymer*, 2005, 46 6461–6473.
- [87] Saiz-Urra, L.; González-Díaz, H.; Uriarte, E. Proteins Markovian 3D-QSAR with spherically-truncated average electrostatic potentials. *Bioorg. Med. Chem.*, 2005, 13(11), 3641-3647.
- [88] González-Díaz, H.; Uriarte, E.; Ramos de Armas, R. Predicting stability of Arc repressor mutants with protein stochastic moments. *Bioorg. Med. Chem.*, 2005, 13(2), 323-331.
- [89] Concu, R.; Podda, G.; Uriarte, E.; Gonzalez-Diaz, H. Computational chemistry study of 3D-structure-function relationships for enzymes based on Markov models for protein electrostatic, HINT, and van der Waals potentials. *J. Comput. Chem.*, 2008, 30(9), 1510-1520.
- [90] González-Díaz, 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. *J. Proteome Res.*, 2007, 6(2), 904-908.
- [91] Gonzalez-Diaz, H.; Saiz-Urra, L.; Molina, R.; Gonzalez-Diaz, Y.; Sanchez-Gonzalez, A. Computational chemistry approach to protein kinase recognition using 3D stochastic van der Waals spectral moments. *J. Comput. Chem.*, 2007, 28(6), 1042-1048.
- [92] Gonzalez-Diaz, H.; Molina, R.; Uriarte, E. Recognition of stable protein mutants with 3D stochastic average electrostatic potentials. *FEBS Lett.*, 2005, 579(20), 4297-4301.
- [93] Concu, R.; Podda, G.; Uriarte, E.; Gonzalez-Diaz, H. Computational chemistry study of 3D-structure-function relationships for enzymes based on Markov models for protein electrostatic, HINT, and van der Waals potentials. *J. Comput. Chem.*, 2009, 30, 1510-1520.
- [94] González-Díaz, H.; Pérez-Castillo, Y.; Podda, G.; Uriarte, E. Computational Chemistry Comparison of Stable/Nonstable Protein Mutants Classification Models Based on 3D and Topological Indices. *J. Comput. Chem.*, 2007, 28, 1990-1995.
- [95] Agüero-Chapin, G.; Varona-Santos, J.; de la Riva, G.A.; Antunes, A.; Gonzalez-Villa, T.; Uriarte, E.; Gonzalez-Diaz, H. Alignment-Free Prediction of Polygalacturonases with Pseudofolding Topological Indices: Experimental Isolation from *Coffea arabica* and Prediction of a New Sequence. *J. Proteome Res.*, 2009, 8(4), 2122-2128.
- [96] Humberto González-Díaz, F.P.-P.a.F.M.U. Predicting Antimicrobial Drugs and Targets with the MARCH INSIDE Approach. *Curr. Top. Med. Chem.*, 2008, 8(18), 1676-1690.

- [97] Concu, R.; Podda, G.; Uriarte, E.; Gonzalez-Diaz, H. Computational chemistry study of 3D-structure-function relationships for enzymes based on Markov models for protein electrostatic, HINT, and van der Waals potentials. *J. Comput. Chem.*, 2009, 30(9), 1510-1520.
- [98] Agüero-Chapin, G.; Gonzalez-Diaz, H.; Molina, R.; VaronaSantos, J.; Uriarte, E.; Gonzalez-Diaz, Y. Novel 2D maps and coupling numbers for protein sequences. The first QSAR study of polygalacturonases; isolation and prediction of a novel sequence from *Psidium guajava* L. *FEBS Lett.*, 2006, 580 723-730.
- [99] StatSoft.Inc. STATISTICA, (data analysis software system), version 6.0, www.statsoft.com, 6.0; 2002.
- [100] Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 1977, 112(3), 535-542.
- [101] Tenorio-Borroto, E.; Penuelas Rivas, C.G.; Vasquez Chagoyan, J.C.; Castanedo, N.; Prado-Prado, F.J.; Garcia-Mera, X.; Gonzalez-Diaz, H. ANN multiplexing model of drugs effect on macrophages; heoretical and flow cytometry study on the cytotoxicity of the antimicrobial drug G1 in spleen. *Bioorg. Med. Chem.*, 2012.
- [102] Gonzalez-Diaz, H.; Bonet, I.; Teran, C.; De Clercq, E.; Bello, R.; Garcia, M.M.; Santana, L.; Uriarte, E. ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds. *Eur. J. Med. Chem.*, 2007, 42 (5), 580-585.
- [103] González-Díaz, H.; Gia, O.; Uriarte, E.; Hernandez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J.A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L.A.; Cabrera, M.A. Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design I: discovery of anticancer compounds. *J. Mol. Model.*, 2003, 9(6), 395-407.
- [104] Prado-Prado, F.J.; Garcia-Mera, X.; Gonzalez-Diaz, H. Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorg. Med. Chem.*, 2010, 18(6), 2225-2231.
- [105] Yildirim, M.A.; Goh, K.I.; Cusick, M.E.; Barabasi, A.L.; Vidal, M. Drug-target network. *Nat. Biotechnol.*, 2007, 25(10), 1119-1126.
- [106] Yamanishi, Y.; Araki, M.; Gutteridge, A.; Honda, W.; Kanehisa, M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 2008, 24(13), i232-240.
- [107] Nunez, M.B.; Maguna, F.P.; Okulik, N.B.; Castro, E.A. QSAR modeling of the MAO inhibitory activity of xanthenes derivatives. *Bioorg. Med. Chem. Lett.*, 2004, 14(22), 5611-5617.
- [108] Terada, M.; Inaba, M.; Yano, Y.; Hasuma, T.; Nishizawa, Y.; Morii, H.; Otani, S. Growth-inhibitory effect of a high glucose concentration on osteoblast-like cells. *Bone*, 1998, 22 (1), 17-23.
- [109] Freund, J.A.; Poschel, T. *Lect. Notes Phys*; Springer-Verlag: Berlin, Germany, 2000.
- [110] Speck-Planche, A.; Kleandrova, V.V. In silico design of multi-target inhibitors for C-C chemokine receptors using substructural descriptors. *Mol. Divers.*, 2012, 16(1), 183-191.
- [111] Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N. In silico discovery and virtual screening of multi-target inhibitors for proteins in *Mycobacterium tuberculosis*. *Comb. Chem. High Throughput Screen.*, 2012, 15 (8), 666-673.
- [112] Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N. Chemoinformatics in Multi-Target Drug Discovery for Anti-Cancer Therapy: In Silico Design Of Potent And Versatile Anti-Brain Tumor Agents. *Anticancer Agents Med. Chem.*, 2011, 12(6), 678-685.
- [113] Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N. Multi-target drug discovery in anti-cancer therapy: Fragment-based approach toward the design of potent and versatile anti-prostate cancer agents. *Bioorg. Med. Chem.*, 2011, 19(21), 6239-6244.
- [114] Prado-Prado, F.J.; Garcia, I.; Garcia-Mera, X.; Gonzalez-Diaz, H. Entropy multi-target QSAR model for prediction of antiviral drug complex networks. *Chemom. Intell. Lab. Syst.*, 2011, 107(2), 227-233.
- [115] Prado-Prado, F.J.; Ubeira, F.M.; Borges, F.; Gonzalez-Diaz, H. Unified QSAR & Network-Based Computational Chemistry Approach to Antimicrobials. II. Multiple Distance and Triadic Census Analysis of Antiparasitic Drugs Complex Networks. *J. Comput. Chem.*, 2010, 31(1), 164-173.
- [116] Prado-Prado, F.J.; Borges, F.; Perez-Montoto, L.G.; Gonzalez-Diaz, H. Multi-target spectral moment: QSAR for antifungal drugs vs. different fungi species. *Eur. J. Med. Chem.*, 2009, 44(10), 4051-4056.

- [117] Prado-Prado, F.J.; Gonzalez-Diaz, H.; Martinez de la Vega, O.; Ubeira, F.M.; Chou, K.-C. 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. *Bioorganic & Medicinal Chemistry*, 2008, 16(11), 5871-5880.
- [118] Prado-Prado, F.J.; Gonzalez-Diaz, 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(2), 897-902.
- [119] Gonzalez-Diaz, H.; Prado-Prado, F.J.; Santana, L.; Uriarte, E. Unify QSAR approach to antimicrobials. Part I: Predicting antifungal activity against different species. *Bioorg. Med. Chem.*, 2006, 14(17), 5973-5980.
- [120] Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N. Rational drug design for anti-cancer chemotherapy: multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. *Bioorg. Med. Chem.*, 2012, 20(15), 4848-4855.
- [121] Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N. A ligand-based approach for the in silico discovery of multi-target inhibitors for proteins associated with HIV infection. *Mol. BioSyst.*, 2012, 8(8), 2188-2196.
- [122] Prado-Prado, F.; Garcia-Mera, X.; Escobar, M.; Sobarzo-Sanchez, E.; Yanez, M.; Riera-Fernandez, P.; Gonzalez-Diaz, H. 2D MI-DRAGON: a new predictor for protein-ligands interactions and theoretic-experimental studies of US FDA drug-target network, oxoisoaporphine inhibitors for MAO-A and human parasite proteins. *Eur. J. Med. Chem.*, 2011, 46(12), 5838-5851.
- [123] Gonzalez-Diaz, H.; Prado-Prado, F.; Garcia-Mera, X.; Alonso, N.; Abeijon, P.; Caamano, O.; Yanez, M.; Munteanu, C.R.; Pazos, A.; Dea-Ayuela, M.A.; Gomez-Munoz, M.T.; Garijo, M.M.; Sansano, J.; Ubeira, F.M. MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from *Trichomonas gallinae*. *J. Proteome Res.*, 2011, 10(4), 1698-1718.
- [124] Gonzalez-Diaz, H.; Prado-Prado, F.; Sobarzo-Sanchez, E.; Haddad, M.; Maurel Chevalley, S.; Valentin, A.; Quetin-Leclercq, J.; DeaAyuela, M.A.; Teresa Gomez-Munos, M.; Munteanu, C.R.; Jose Torres-Labandeira, J.; Garcia-Mera, X.; Tapia, R.A.; Ubeira, F.M. NL MIND-BEST: A web server for ligands and proteins discovery-Theoretic-experimental study of proteins of *Giardia lamblia* and new compounds active against *Plasmodium falciparum*. *J. Theor. Biol.*, 2011, 276(1), 229-249.
- [125] Prado-Prado, F.; Garcia-Mera, X.; Abeijon, P.; Alonso, N.; Caamano, O.; Yanez, M.; Garate, T.; Mezo, M.; Gonzalez-Warleta, M.; Muino, L.; Ubeira, F.M.; Gonzalez-Diaz, H. Using entropy of drug and protein graphs to predict FDA drug-target network: Theoretic-experimental study of MAO inhibitors and hemoglobin peptides from *Fasciola hepatica*. *Eur. J. Med. Chem.*, 2011, 46(4), 1074-1094.
- [126] Rodriguez-Soca, Y.; Munteanu, C.R.; Dorado, J.; Pazos, A.; Prado-Prado, F.J.; Gonzalez-Diaz, H. Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. *J. Proteome Res.*, 2010, 9(2), 1182-1190.
- [127] Gonzalez-Diaz, H.; Muino, L.; Anadon, A.M.; Romaris, F.; Prado-Prado, F.J.; Munteanu, C.R.; Dorado, J.; Sierra, A.P.; Mezo, M.; Gonzalez-Warleta, M.; Garate, T.; Ubeira, F.M. MISS-Prot: web server for self/non-self discrimination of protein residue networks in parasites; theory and experiments in *Fasciola* peptides and *Anisakis* allergens. *Mol. BioSyst.*, 2011, 7(6), 1938-1955.
- [128] Concu, R.; Dea-Ayuela, M.A.; Perez-Montoto, L.G.; Prado-Prado, F.J.; Uriarte, E.; Bolas-Fernandez, F.; Podda, G.; Pazos, A.; Munteanu, C.R.; Ubeira, F.M.; Gonzalez-Diaz, H. 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in *Leishmania* parasites. *Biochim. Biophys. Acta*, 2009, 1794(12), 1784-1794.
- [129] Whitesides, G.M.; Grzybowski, B. Self-Assembly at All Scales. *Science*, 2002, 295(5564), 2418-2421.
- [130] Ruiz, R.; Tedeschl, L.O.; Marini, J.C.; Fox, D.G.; Pell, A.N.; Jarvis, G.; Russell, J.B. The effect of a ruminal nitrogen (N) deficiency in dairy cows: evaluation of the cornell net carbohydrate and protein system ruminal N deficiency adjustment. *J. Dairy Sci.*, 2002, 85(11), 2986-2999.
- [131] Hassan, N.; Ruso, J.M.; Piñeiro, A.n. Hydrogenated/Fluorinated Catanionic Surfactants as Potential Templates for Nanostructure Design. *Langmuir*, 2011, 27(16), 9719-9728.

- [132] Nagai, A.; Nagai, Y.; Qu, H.; Zhang, S. Dynamic Behaviors of Lipid-Like Self-Assembling Peptide A6D and A6K Nanotubes. *J. Nanosci. Nanotechnol.*, 2007, 7(7), 2246-2252.
- [133] Blanco, E.; Verdes, P.V.; Ruso, J.M.; Prieto, G.; Sarmiento, F. Interactions in binary mixed systems involving betablockers with different lipophilicity as a function of temperature and mixed ratios. *Colloids Surf. A Physicochem. Eng. Asp.*, 2009, 334(1-3), 116-123.
- [134] Soriano, A.; Marco, F.; Martínez, J.A.; Pisos, E.; Almela, M.; Dimova, V.P.; Alamo, D.; Ortega, M.; Lopez, J.; Mensa, J. Influence of vancomycin minimum inhibitory concentration on the treatment of methicillin-resistant *Staphylococcus aureus* bacteremia. *Clin. Infect. Dis.*, 2008, 46(2), 193-200.
- [135] Bakszt, R.; Wernimont, A.; Allali-Hassani, A.; Mok, M.W.; Hills, T.; Hui, R.; Pizarro, J.C. The crystal structure of *Toxoplasma gondii* pyruvate kinase 1. *PLoS ONE*, 2010, 5(9), e12736.
- [136] Yoshida, N.; Matsuoka, K.; Moroi, Y. Micelle Formation of n -Decyltrimethylammonium Perfluorocarboxylates. *J. Colloid Interface Sci.*, 1997, 187(2), 388-395.
- [137] López Fontán, J.L.; Costa, J.; Ruso, J.M.; Prieto, G.; Sarmiento, F. Electrical Conductivities and Critical Micelle Concentrations (Determined by the Local Polynomial Regression Method) of Imipramine and Clomipramine Hydrochlorides from (283 to 313) K. *J. Chem. Eng. Data*, 2004, 49(4), 1008-1012.
- [138] González-Pérez, A.; Ruso, J.M.; Romero, M.J.; Blanco, E.; Prieto, G.; Sarmiento, F. Application of thermodynamic models to study micellar properties of sodium perfluoroalkyl carboxylates in aqueous solutions. *Chem. Phys.*, 2005, 313(1-3), 245-259.
- [139] Blanco, E.; González-Pérez, A.; Ruso, J.M.; Pedrido, R.; Prieto, G.; Sarmiento, F. A comparative study of the physicochemical properties of perfluorinated and hydrogenated amphiphiles. *J. Colloid Interface Sci.*, 2005, 288(1), 247-260.
- [140] Rodríguez, J.R.; González-Pérez, A.; Del Castillo, J.L.; Czapkiewicz, J. Thermodynamics of Micellization of Alkyldimethylbenzylammonium Chlorides in Aqueous Solutions. *J. Colloid Interface Sci.*, 2002, 250(2), 438-443.
- [141] Taboada, P.; Attwood, D.; García, M.; Jones, M.N.; Ruso, J.M.; Mosquera, V.; Sarmiento, F. Thermodynamics of Association of Structurally Related Amphiphilic Penicillins. *J. Colloid Interface Sci.*, 2000, 221(2), 242-245.
- [142] Ruso, J.M.; Attwood, D.; Rey, C.; Taboada, P.; Mosquera, V.; Sarmiento, F. Light Scattering and NMR Studies of the Self-Association of the Amphiphilic Molecule Propranolol Hydrochloride in Aqueous Electrolyte Solutions. *J. Phys. Chem. B*, 1999, 103(34), 7092-7096.
- [143] Taboada, P.; Ruso, J.M.; Garcia, M.; Mosquera, V. Surface properties of some amphiphilic antidepressant drugs. *Colloids Surf. A Physicochem. Eng. Asp.*, 2001, 179(1), 125-128.
- [144] Klein, J. In: *The Chemistry of Double-Bonded Functional Groups: Supplement A*; Patai, S., Ed.; Wiley: Chichester, 1989; Vol. 2, Part 1, Chapter 10.
- [145] Volkmann, R.A. In: *Comprehensive Organic Synthesis, Additions to C-X pi-Bonds, Part 1*; Schreiber, S.L., Ed.; Pergamon Press: Oxford, 1991; Vol. 1.
- [146] Kleinman, E.F.V., R. A. In: *Comprehensive Organic Synthesis, Additions to C-X pi-Bonds, Part 2*; Heathcock, C.H., Ed.; Pergamon Press: Oxford, 1991; Vol. 2.
- [147] Berrisford, D.J. Catalytic asymmetric C-C bond formation: new enolato- and organolithium chemistry. *Angew. Chem., Int. Ed. Engl.*, 1995, 34, 178-180.
- [148] Risch, N.A., M. In: *Methods of Organic Chemistry. Stereoselective Synthesis [Houben-Weyl]*; Helmchen, G.H., R. W.; Mulzer, J.; Schaumann, E., Ed.; Thieme: Stuttgart, 1996; Vol. Workbench Edition E21, Vol. 3.
- [149] North, M. Amines and amides. *Contemp. Org. Synth.*, 1996, 3, 323-343.
- [150] Denmark, S.E.; Nicaise, O.J.-C.C. Ligand-Mediated Addition of Organometallic Reagents to Azomethine Functions. *Chem. Commun.*, 1996, 9, 999-1004.
- [151] Enders, D.; Reinhold, U. Asymmetric Synthesis of Amines by Nucleophilic 1,2-Addition of Organometallic Reagents to the CN-Double Bond. *Tetrahedron: Asymmetry*, 1997, 8, 1895-1946.
- [152] Bloch, R. Addition of organometallic reagents to C=N bonds: reactivity and selectivity. *Chem. Rev.*, 1998, 98, 1404-1438.
- [153] Denmark, S.E.N., O. J.-C. In: *Comprehensive Asymmetric Catalysis*; Jacobsen, E.N.P., A.; Yamamoto, H., Ed.; Springer-Verlag: Berlin, 1999.
- [154] Seyden-Penne, J. *Chiral Auxiliaries and Ligands in Asymmetric Synthesis*; Wiley: New York, 1995.
- [155] Jacques, J.C., A.; Wilen, S. H. *Enantiomers, Racemates, and Resolution*; Wiley: New York, 1981.
- [156] Eliel, E.L.W., S. H.; Mander, L. N. *Stereochemistry of Organic Compounds*; Wiley: New York, 1994.

- [157] Moser, H.; Ryhs, G.; Santer, H. Effect of atropisomerism and chiral center on the biological activity of metolachlor, *Z. Naturforsch.*, 1982, 37B, 451-462.
- [158] Ariens, E.J. QSAR methods and stereoselectivity in action. *Prog. Clin. Biol. Res.*, 1989, 291, 3-6.
- [159] Ignatz-Hoover, F.; Petrukhin, R.; Karelson, M.; Katritzky, A.R. QSRR correlation of free-radical polymerization chain-transfer constants for styrene. *J. Chem. Inf. Comput. Sci.*, 2001, 41(2), 295-299.
- [160] Varnek, A.F., D.; Hoonakker, F.; Solov'ev, V. P. Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures. *J. Comput. Aided Mol. Des.*, 2005, 19, 693-703.
- [161] Satoh, H.; Sacher, O.; Nakata, T.; Chen, L.; Gasteiger, J.; Funatsu, K. Classification of organic reactions: similarity of reactions based on changes in the electric features of oxygen atoms at the reaction sites. *J. Chem. Inf. Comput. Sci.*, 1998,38, 210-219.
- [162] Long, X.N., J. Estimation of gas-phase reaction rate constants of alkyl naphthalenes with chlorine, hydroxyl and nitrate radicals. *Chemosphere*, 2007, 67, 2028-2034.
- [163] Mu, F.U., P. J.; Unkefer, C. J.; Hlavacek, W. S. Prediction of oxidoreductase-catalyzed reactions based on atomic properties of metabolites. *Bioinformatics*, 2006, 22(24), 3082-3088.
- [164] Arrasate, S.; Sotomayor, N.; Lete, E.; González-Díaz, H. In: *Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks*; González-Díaz, H.; Munteanu, C.R., Eds.; Transworld Research Network: Kerala, 2010; pp. 53-68.
- [165] Munteanu, C.R.; Dorado, J.; Pazos-Sierra, A.; Prado-Prado, F.; Perez-Montoto, L.G.; Vilar, S.; Ubeira, F.M.; Sanchez-Gonzalez, A.; Cruz-Monteagudo, M.; Arrasate, S.; Sotomayor, N.; Lete, E.; Duardo-Sanchez, A.; Diaz-Lopez, A.; Patlewicz, G.; Gonzalez-Díaz, H. In: *Towards an information theory of Complex Networks: Statistical Methods and Applications*; Dehmer, M.; Emmert-Streib, F.; Mehler, A. Eds.; Birkhäuser/Springer Publisher: Germany, 2009.
- [166] Denmark, S.E.S., C. M. Effect of Ligand Structure in the Bisoxazoline Mediated Asymmetric Addition of Methyllithium to Imines. *J. Org. Chem.*, 2000, 65, 5875-5878.
- [167] Hasegawa, M.T., D.; Tomioka, K. Facile Asymmetric Synthesis of α -Amino Acids Employing Chiral Ligand-Mediated Asymmetric Addition Reactions of Phenyllithium with Imines. *Tetrahedron*, 2000, 56, 10153-10158.
- [168] Taniyama, D.H., M.; Tomioka, K. A facile asymmetric synthesis of 1-substituted tetrahydroisoquinoline based on a chiral ligand-mediated addition of organolithium to imine. *Tetrahedron: Asymmetry*, 1999, 10, 221-223.
- [169] Inoue, I.S., M.; Koga, K.; Kanai, M.; Tomioka, K. Enantioselective Reaction of An Imine with Methyllithium Catalyzed by A Chiral Ligand. *Tetrahedron: Asymmetry*, 1995, 6, 2527-2533.
- [170] Denmark, S.E.N., N.; Nicaise, O. J.-C. Asymmetric Addition of Organolithium Reagents to Imines. *J. Am. Chem. Soc.*, 1994, 116, 8797-8798.
- [171] Freedman, D.S.; Otvos, J.D.; Jeyarajah, E.J.; Barboriak, J.J.; Anderson, A.J.; Walker, J.A. Relation of lipoprotein subclasses as measured by proton nuclear magnetic resonance spectroscopy to coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.*, 1998, 18(7), 1046-1053.
- [172] Inoue, I.S., M.; Koga, K.; Tomioka, K. Asymmetric 1,2-Addition of Organolithium to Aldimines Catalyzed by Chiral Ligand. *Tetrahedron*, 1994, 50, 4429-4438.
- [173] Leopoldo, M.; Berardi, F.; Colabufo, N.A.; Contino, M.; Lacivita, E.; Perrone, R.; Tortorella, V. Studies on 1-arylpiperazine derivatives with affinity for rat 5-HT₇ and 5-HT_{1A} receptors. *J. Pharm. Pharmacol.*, 2004, 56(2), 247-255.
- [174] Cabello, N.K., J.-C.; Gille, S.; Alexakis, A.; Bernardinelli, G.; Pinchard, L.; Caille, J.-C. Simple 1,2-Diamine Ligands for Asymmetric Addition of Aryllithium Reagents to Imines. *Eur. J. Org. Chem.*, 2005, 4835-4842.
- [175] Kizirian, J.-C.C., N.; Pinchard, L.; Caille, J.-C.; Alexakis, A.; Enantioselective addition of methyllithium to aromatic imines catalyzed by C₂ symmetric tertiary diamines. *Tetrahedron*, 2005, 61, 8939-8946.
- [176] Jaffe, J.D.; Mani, D.R.; Leptos, K.C.; Church, G.M.; Gillette, M.A.; Carr, S.A. PEPPEr, a platform for experimental proteomic pattern recognition. *Mol. Cell. Proteomics*, 2006, 5(10), 1927-1941.
- [177] Cabello, N.K., J.-C.; Alexakis, A.; Enantioselective addition of aryllithium reagents to aromatic imines mediated by 1,2-diamine ligands. *Tetrahedron Lett.*, 2004, 45, 4639-4642.
- [178] Cheng, Z.; Ren, J.; Li, Y.; Chang, W.; Chen, Z. Study on the multiple mechanisms underlying the reaction between hydroxyl radical and phenolic compounds by qualitative structure and activity relationship. *Bioorg. Med. Chem.*, 2002, 10(12), 4067-4073.

- [179] Pompe, M.V., M.; Randic, M.; Balaban, A. T. Using variable and fixed topological indices for the prediction of reaction rate constants of volatile unsaturated hydrocarbons with OH radicals. *Molecules*, 2004, 9, 1160-1176.
- [180] Ren, Y.L., H.; Yao, X.; Liu, M. Prediction of ozone tropospheric degradation rate constants by projection pursuit regression. *Anal. Chim. Acta*, 2007, 589, 150-158.
- [181] González-Díaz, H.; Molina-Ruiz, R.; Hernandez, I. MARCH-INSIDE v3.0 (MARKov CHains INvariants for SIMulation & DESign); Windows supported version under request to the main author contact email: gonzalezdiazh@yahoo.es. 2007.