



Contents lists available at ScienceDirect

## European Journal of Medicinal Chemistry

journal homepage: <http://www.elsevier.com/locate/ejmech>

## Original article

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*

Francisco Prado-Prado<sup>a,\*</sup>, Xerardo García-Mera<sup>a</sup>, Paula Abeijón<sup>a</sup>, Nerea Alonso<sup>a</sup>, Olga Caamaño<sup>a</sup>, Matilde Yáñez<sup>b</sup>, Teresa Gárate<sup>c</sup>, Mercedes Mezo<sup>d</sup>, Marta González-Warleta<sup>d</sup>, Laura Muiño<sup>e</sup>, Florencio M. Ubeira<sup>e</sup>, Humberto González-Díaz<sup>e</sup>

<sup>a</sup> Department of Organic Chemistry, Faculty of Pharmacy, USC 15782, Spain<sup>b</sup> Department of Pharmacology, USC 15782, Spain<sup>c</sup> Parasitology Service, National Center of Microbiology, Instituto de Salud Carlos III, Majadahonda, Madrid 28220, Spain<sup>d</sup> Centro de Investigaciones Agrarias (INGACAL), Mabegondo, A Coruña 15318, Spain<sup>e</sup> Department of Microbiology and Parasitology, Faculty of Pharmacy, USC 15782, Spain

## ARTICLE INFO

## Article history:

Received 13 October 2010

Received in revised form

10 January 2011

Accepted 13 January 2011

Available online 21 January 2011

## Keywords:

Drug–Protein interaction complex networks

Protein structure networks

Multi-target QSAR

Markov model

Rasagiline derivatives

MAO enzymes

*Fasciola hepatica* proteome

## ABSTRACT

There are many drugs described with very different affinity to a large number of receptors. In this work, we selected Drug-Target pairs (DTPs/nDTPs) of drugs with high affinity/non-affinity for different targets like proteins. Quantitative Structure–Activity Relationships (QSAR) models become a very useful tool in this context to substantially reduce time and resources consuming experiments. Unfortunately, most QSAR models predict activity against only one protein. To solve this problem, we developed here a multi-target QSAR (mt-QSAR) classifier using the MARCH-INSIDE technique to calculate structural parameters of drug and target plus one Artificial Neuronal Network (ANN) to seek the model. The best ANN model found is a Multi-Layer Perceptron (MLP) with profile MLP 32:32–15–1:1. This MLP classifies correctly 623 out of 678 DTPs (Sensitivity = 91.89%) and 2995 out of 3234 nDTPs (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 classifies correctly 313 out of 338 DTPs (Sensitivity = 92.60%) and 1411 out of 1534 nDTP (Specificity = 91.98%) in validation series, corresponding to total Accuracy = 92.09% for validation series (Predictability). This model favorably compares with other LDA and ANN models developed in this work and Machine Learning classifiers published before to address the same problem in different aspects. These mt-QSARs offer also a good opportunity to construct drug–protein Complex Networks (CNs) that can be used to explore large and complex drug–protein receptors databases. Finally, we illustrated two practical uses of this model with two different experiments. In experiment 1, we report prediction, synthesis, characterization, and MAO-A and MAO-B pharmacological assay of 10 rasagiline derivatives promising for anti-Parkinson drug design. In experiment 2, we report sampling, parasite culture, SEC and 1DE 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*; which is promising for anti-parasite drug targets discovery.

© 2011 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

The fast and accurate prediction of interactions between drugs and target proteins is a keystone piece on the combination of

bioinformatics and proteome research toward drug discovery. Therefore, there is a strong incentive to develop new methods capable of detecting these potential drug-target interactions efficiently [1]. In this sense, graphs and complex network theory may play an important role at different stages of modeling process with different degrees of organization of matter [2–9]. In a first stage, we can use molecular graphs to represent and calculate structural parameters for drugs sometimes called Topological Indices (TIs) but

\* Corresponding author. Fax: +34 981 594912.

E-mail address: [fenol1@hotmail.com](mailto:fenol1@hotmail.com) (F. Prado-Prado).

also estimate physicochemical parameters based on a graph method, see our recent reviews [10]. At a higher level we can use graphs to represent structure of the drug-target proteins and calculate characteristic TIs and/or physicochemical parameters of proteins structure or protein interactions networks, see for instance the works of Giuliani [11–17], or our recent reviews [18,19]. Next, we can develop a kind of computer program with network topology called Artificial Neural Networks (ANNs) that after adequate training learn predicts target proteins for a given drug. It means, ANNs are network-like software that may use as inputs TIs and/or physicochemical parameters calculated in the previous steps to predict which network-like molecular structures present or not a desire property, see for instance the works after Caballero and Fernandez et al. [20–23] with applications to both drugs and proteins or works of Zbilut et al. [24,25]. In particular, using the parameters of the drug and the target ANNs may select Drug-Target pairs of drugs with high affinity (DTPs) out of those pairs for drugs with none affinity for different targets (nDTPs). In general, this technique (using or not ANNs) lie within the kind of studies called Quantitative Structure–Activity Relationships (QSAR) models and may become a very useful tool in this context to substantially reduce time and resources consuming experiments. In a last step, the prediction of all possible DTPs/nDTPs in the global set of relationships between protein targets and all drugs form the complex network of drugs and/or targets. For instance, Yildirim and Goh et al. [26] have built a bipartite graph composed of US Food and Drug Administration (US FDA) approved drugs and proteins linked by drug–target binary associations. The resulting network connects most drugs into a highly interlinked giant component, with strong local clustering of drugs of similar types according to Anatomical Therapeutic Chemical classification. Topological analyses of this network quantitatively showed an overabundance of ‘follow-on’ drugs, that is, drugs that target already targeted proteins.

In a previous work, our group have reported a QSAR model base on the MARCH-INSIDE method to predict a large network of DTPs [27]. However, even when this model is useful to predict targets for many proteins it lack of availability for public research due to it was not implemented as an online web server. The problem with many QSAR models is more serious because many of them work only for one target protein or for a limited family of organic compounds. We then develop new statistical methods to predict simultaneously on a large scale unknown DTPs from chemical structure and 3D structure of target proteins. In principle, we can select between more than 1600 different molecular descriptors to solve the former problem [28]. However, not many methods offer one unique software platform to calculate parameters for both drugs and protein structures based on unified theoretic background more easily to rationalize. Our group has introduced elsewhere a Markov Chain Model (MCM) method named **MARKov CHains Invariants for Network SIMulation and DESIGN** (MARCH-INSIDE). The MARCH-INSIDE approach makes use of the same MCM theoretic formulation to calculate the average values of different molecular TIs and physicochemical properties from 2D, 3D, and/or sequence chemical structures including drugs, DNA, RNA, and proteins, see a recent review [29]. MARCH-INSIDE parameters not only offer these advantages but also may used as inputs to train ANNs with the software STATISTICA (e.g.) [30–35].

In this work, we developed a multi-target QSAR (mt-QSAR) classifier using the MARCH-INSIDE technique to calculate structural parameters of drug and target plus one ANN to seek the model. The validation of the model was carried out by means of external predicting series. We also compare this model with other ANN models developed in this work and Machine Learning (ML) classifiers published before to address the same problem. A very good MARCH-INSIDE-QSAR model was obtained, and the subsequent

combined QSAR-CN analysis may become of major importance for the prediction of the activity of new compounds against different targets or the discovery of new targets. In this sense we reported two illustrative experiments that combine both experimental and theoretical studies to show how to use this model in practical situations. In experiment 1, we report by first time mt-QSAR and CN prediction, synthesis, characterization, and MAO-A and MAO-B pharmacological assay of 8 rasagiline derivatives. In experiment 2, we report sampling, parasite culture, sample preparation, 2-DE, MALDI-TOF and -TOF/TOF MS, MASCOT search, MM/MD 3D structure modeling, and QSAR prediction of CN for peptidome of hemoglobin found in parasite *Fasciola hepatica*. In Fig. 1 we depict a flowchart with the main steps given in this work to train and validate the ANN classifier.

## 2. Materials and methods

### 2.1. Computational methods

#### 2.1.1. MARCH-INSIDE technique

*Entropy parameters for drug graphs.* The MARCH-INSIDE approach applied to drugs ( $D$ ) is based on the calculation of the different physicochemical molecular properties as an average of atomic properties for all the molecules or groups of atoms ( $G$ ) [10,19,29]. For instance, it is possible to derive average estimations of information about molecular structure using entropy indices  $D_{\theta_k}(G)$  [36,37].

$$D_{\theta_k}(G) = - \sum_{j \in G}^k p_j(G) \cdot \log [p_j(G)] \quad (1)$$

It is possible to consider isolated atoms ( $k = 0$ ) in the estimation of the molecular properties  $D_{\theta_0}(G)$ . In this case the probabilities  $p_j(\theta_j)$  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 ( $k > 0$ ) in the molecular backbone. In order to reach this goal the method uses an MM, which determines the absolute probabilities  $p_k(\theta_j)$  with which the atoms placed at different distances  $k$  affect the contribution of the atom  $j$  to the molecular property in question. Finally, it is interesting to note that one can sum only the atoms included in a specific group of atoms ( $G$ ) rather than all atoms. In this way we can approach specific classes of average properties such as average entropy for  $sp_3$  carbon atoms ( $C_{sp_3}$ ) or average entropy for heteroatoms (Het). All calculations were performed using the program MARCH-INSIDE [10,19,29], which was developed in-house, see recent reviews for details.

*Entropy parameters of protein residue networks.* In previous works we have predicted protein function based on different protein structural parameters derived from a Markov matrix that account for electrostatic interactions between aminoacid pairs in the 3D structure of the protein. One of the classes of parameters used was called the Shannon Entropy  $T_{\theta_k}(R)$  of the markov matrix. These values are used here as inputs to describe information about the structure of the drug-target proteins ( $T$ ) in order to construct the mt-QSAR models for DTPs. The detailed explanation has been published before [30,38–45] and reviewed in detail more recently [29]. As follows we give the formula for  $T_{\theta_k}(R)$  values and some general explanations:

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

Where,  $p_i(R)$  values are the absolute probabilities with which the effect of the electrostatic interaction propagates from the

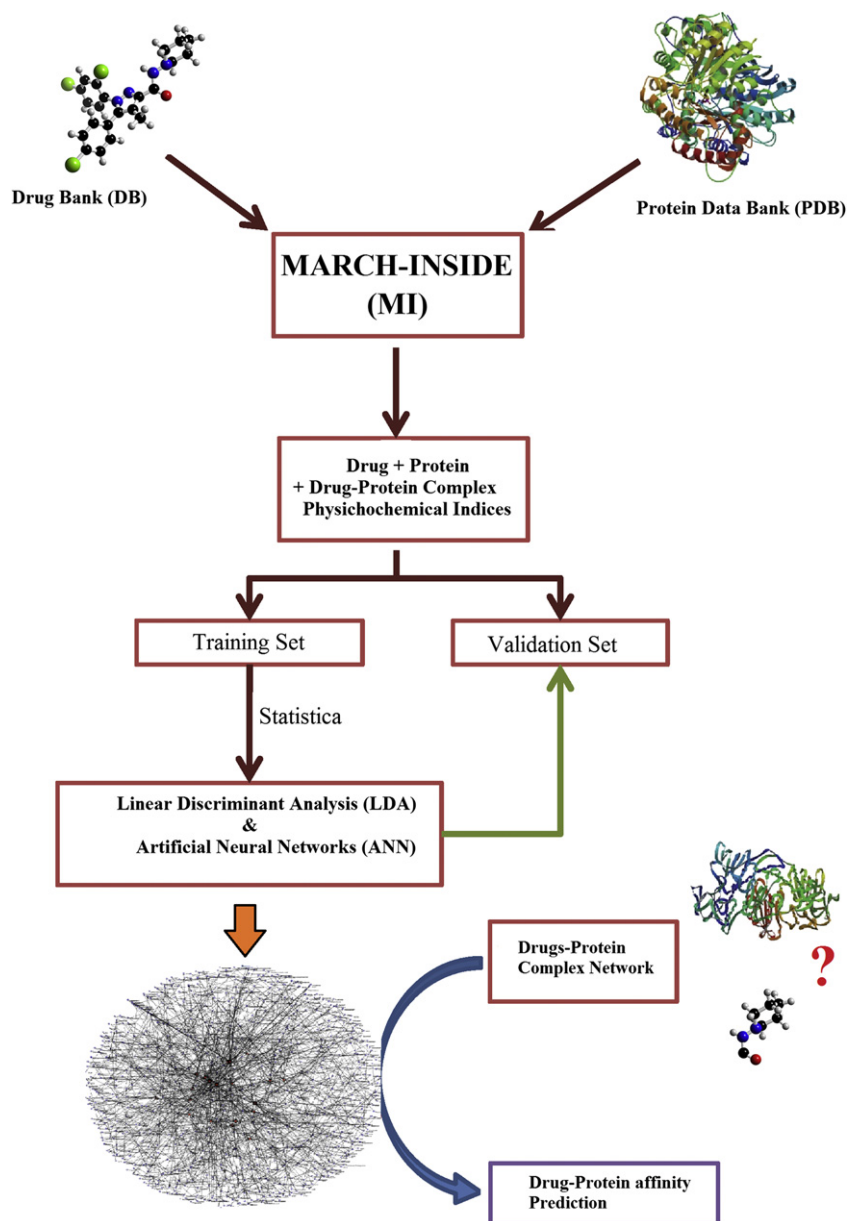


Fig. 1. Flowchart of all steps given in this work to develop the new model.

aminoacid  $i^{\text{th}}$  to other aminoacids  $j^{\text{th}}$  next to it and returns to  $i^{\text{th}}$  after  $k$ -steps. These probabilities refer to: aminoacids considered isolated in the space ( $k = 0$ ), interaction between aminoacids in direct contact ( $k = 1$ ) or spatial ( $k > 1$ ) indirect interactions between aminoacids placed at a distance equal to  $k$ -times the cut-off distance ( $r_{ij} = k \cdot r_{\text{cut-off}}$ ) in the residue network. Euclidean 3D space  $r_3 = (x, y, z)$  coordinates of the  $C_\alpha$  atoms of aminoacids listed in protein PDB files. For calculation, all water molecules and metal ions were removed [19]. All calculations were carried out with our in-house software MARCH-INSIDE 2.0 [19]. For the calculation, the MARCH-INSIDE software always uses the full matrix, never a sub-matrix, but the last summation term may run either for all aminoacids or only for some specific protein regions ( $R$ ) denoted as:  $c$  for core,  $i$  for inner,  $m$  for middle, and  $s$  for surface regions, respectively). Consequently, we can calculate different  $T_{\theta_k}(R)$  for the aminoacids contained in the regions ( $c, i, m, s$ , or  $t$ ) and placed at a topological distance  $k$  each other within this orbit ( $k$  is the order) [32,38,39,46,47]. In this work, we have calculated altogether 5

(types of regions)  $\times$  6(orders considered) = 30  $T_{\theta_k}(R)$  indices for each protein.

**Statistical analysis.** Let be  $D_{\theta_k}(G)$  entropy descriptors molecular that codify information about drug structure and  $T_{\theta_k}(R)$  entropy descriptors that codify information about drug-target proteins; we attempt to develop a simple mt-QSAR model in the form of a linear classifier with the general formula:

$$S(\text{DTP})_{\text{pred}} = \sum_{k=0}^5 a_{G,k} \cdot D_{\theta_k}(G) + \sum_{k=0}^5 b_{R,k} \cdot T_{\theta_k}(R) + c_0 \quad (3)$$

We used Linear Discriminating Analysis (LDA) to fit this discriminant function. The model deals with the classification of a compound set with or without affinity on different receptors. A dummy variable Affinity Class (AC) was used as input to codify the affinity. This variable indicates either high ( $AC = 1$ ) or low ( $AC = 0$ ) affinity of the drug by the receptor.  $S(\text{DTP})_{\text{pred}}$  or DTP affinity predicted score is the output of the model and it is a continuous

dimensionless score that sorts compounds from low to high affinity to the target coinciding DTPs with higher values of  $S(\text{DTP})_{\text{pred}}$  and nDTPs with lowest values. In equation (3),  $b$  represents the coefficients of the classification function, determined by the LDA module of the STATISTICA 6.0 software package [48]. We used Forward Stepwise algorithm for a variable selection. The statistical significance of the LDA model was determined calculating the  $p$ -level ( $p$ ) of error with Chi-square test. We also inspected the Specificity, Sensitivity, and total Accuracy to determine the quality-of-fit to data in training. Cases for training set were selected at random out of the cases in full dataset. The remnant cases were used to validate the model. The validation of the model was corroborated with these external prediction series; these cases were never used to train the model. The ration between training/validation set was 2/1 approximately. This procedure to select training and validation sets is largely known and used to train QSAR models [49–55].

**Dataset.** The dataset was formed by a set of marketed and/or reported drugs/receptors (proteins) pairs where affinity of drugs with the receptors was established taking into consideration the Drug Bank. The dataset was formed to more than 526 drugs with respectively 323 protein receptors, so we were able to collect already 5784 cases (drug/protein receptors) instead of  $526 \times 323$  cases. The dataset were used to perform an LDA model. Overall model classification accuracy was 88.83% (3475/3912 cases) in training, 88.51% (1657/1872) in validation. In addition the dataset was used to develop ANN models to performance the model. Overall ANN model classification accuracy was 92%. To know if the model works well, the model used 8 own compounds rasagiline derivates to predict their protein receptor. The names or codes for all compounds are depicted in Table 1SM and Table 2SM of the supplementary material, due to space constraints, as well as the references consulted to compile the data in this table.

**Drug–Protein (DP) network construction.** In order to achieve the drug and protein affinity with a network approach where one node represents a drug, target, or drug–target pair and the edges express relationships between pairs of drugs and/or targets [26]. We built a bipartite graph composed of Drug Bank and Protein Data Bank-approved drugs and proteins linked by drug–target binary associations. The resulting network connects most drugs into a highly interlinked giant component, with strong local clustering of drugs of similar types. We build two complex networks, a first for the observed data and the second with the data predicted by the model. First, using the Excel software in a column we introduce all the proteins, the drugs used quotation marks in our database. Then in another column lists all the cases, in total 837 vertices. At the beginning of this column puts the total number of vertices, there are currently two columns of the name of drug and protein and their corresponding number of vertices. At the end of the columns are placed bows in the first column put the number of vertices for the drug and in another column corresponding to the protein. The file was saved as a .txt format file. After we had renamed the .txt file as a .net file we read it with the CentiBin software [56,57]. Using CentiBin we can not only represent the network but also highlight all Drugs and Protein (nodes) connected to a specific DRP. To analyze the relationships between drug targets, we measured and calculating closeness parameter, and the centrality. Lastly, the protein and drug centralities were used as input in STATISTICA in order to study the distribution of the network and compare it with other ideal network distributions including normal, exponential.

### 2.1.2. Theoretical study of hemoglobin peptidome of parasite *F. hepatica*

**MM/MD.** Molecular Mechanics (MM) and Molecular Dynamics (MD) study. For MM study we first introduced the sequence of the 30 peptides in the HyperChem [58]; the optimization of their

geometries was carried out by the Molecular Mechanics Force Field BIO + (CHARMM). In setup we keep the options implemented by default, but allowing a cut-off switching truncation  $r_{\text{in}} = 15 \text{ \AA}$  and  $r_{\text{out}} = 17 \text{ \AA}$ . We refer here to MD in the sense of MD stochastic simulation by the Monte Carlo (MC) method, although some authors understand MD as only the MD deterministic search. The Molecular Dynamics Trajectories (MDTs) or energetic profiles of all the starting structure of peptides were obtained by means of MC method, with the HyperChem package [59,60]. In this sense, the force field AMBER [61] of molecular mechanics was used with distant-dependent dielectric constant (scale factor 1), electrostatic and van der Waals values by default and a cut-off switched function with  $r_{\text{in}} = 15$  and  $r_{\text{out}} = 17 \text{ \AA}$  (see Fig. 2). All the components in the force field were included and the atom type was recalculated by maintaining the current charges. Finally, MD simulation was carried out by use of the Monte Carlo algorithm in the *vacuo* at 300 K and 1000 optimization steps, thus obtaining MDTs with 100 potential energy  $^dE_j$  ( $j = 1, 2, 3, \dots, 100$ ) values for each. We obtained 22 MDTs for 19 peptides. In order to obtain realistic MDTs we monitored an additional parameter in MD algorithms; this is known as the acceptance ratio (ACCR). It appears as ACCR on the list of possible selections in the MC Averages dialog box of HyperChem (see Fig. 2). The ACCR is a running average of the ratio of the number of accepted moves to attempted moves. Varying the step size can have a large effect on the ACCR value. The step size,  $\Delta r$ , is the maximum allowed atomic displacement used in the generation of trial configurations. The default value of  $r$  in HyperChem is  $0.05 \text{ \AA}$  [59]. For most organic molecules, this will result in an ACCR of about 0.5  $\text{\AA}$ , which means that about 50% of all moves are accepted. Increasing the size of the trial displacements may lead to a more complete search of configuration space, but the acceptance ratio will, in general, decrease. Smaller displacements generally lead to higher acceptance ratios but result in more limited sampling. There has been little research to date as regards the optimum value of the acceptance ratio.

## 2.2. Experimental methods

### 2.2.1. Study of rasagiline analogs

**Identification.** Melting points are uncorrected and were determined in a Reichert Kofler Thermopan or in capillary tubes in a Büchi 510 apparatus. Infrared spectra were recorded in a Perkin–Elmer 1640 FTIR spectrophotometer.  $^1\text{H}$  NMR spectra (300 MHz and 500 MHz) and  $^{13}\text{C}$  NMR spectra (75 MHz) were recorded in a Bruker AMX 300 and DRX 500 spectrometer using TMS as internal reference (chemical shifts in  $\delta$  values,  $J$  in Hz). Mass spectra were recorded on a HP5988A spectrometer. FABMS were obtained using MICROMASS AUTOSPEC mass spectrometer. Microanalyses were performed in a Perkin–Elmer 240B elemental analyzer by the Microanalysis Service of the University of Santiago see Table 3SM. Most of reactions were monitored by TLC on pre-coated silica gel plates (Merck 60 F254, 0.25 mm). Synthesized products were purified by flash column chromatography on silica gel (Merck 60, 230–240 mesh) and crystallized if necessary. Solvents were dried by distillation prior use.

**Compound 2.** ( $\pm$ )-3-Amino-3-(thiophen-3-yl)propanoic acid. A solution of thiophen-3-carbaldehyde (2.00 g, 17.83 mmol) in EtOH (5.5 mL) was added ammonium acetate (2.74 g, 35.58 mmol) and malonic acid (1.85 g, 17.78 mmol). The reaction mixture was stirred was refluxed for 7 h. The precipitate formed was filtered and washed with boiling EtOH ( $3 \times 10 \text{ mL}$ ) to give **2** (2.20 g, yield 72%) as a white solid. Mp 223–224 °C. IR (KBr):  $\nu = 2954, 2152, 1532, 1403, 1101, 992, 925, 786 \text{ cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz, TFA- $d$ ):  $\delta = 11.03$  (br s, 3H,  $\text{D}_2\text{O}$  exch.,  $\text{OH} + \text{NH}_2$ ), 6.97–6.91 (m, 2H, 2- $\text{H}_{\text{thiophene}}$  and 5- $\text{H}_{\text{thiophene}}$ ), 6.63–6.61 (m, 1H, 4- $\text{H}_{\text{thiophene}}$ ), 4.58–4.54 (m, 1H, 3-H), 2.92 (dd, 1H,  $J = 18.4, 9.6 \text{ Hz}$ , 2- $\text{HH}$ ), 2.72 (dd, 1H,  $J = 18.4, 4.2 \text{ Hz}$ , 2- $\text{HH}$ ) ppm.  $^{13}\text{C}$



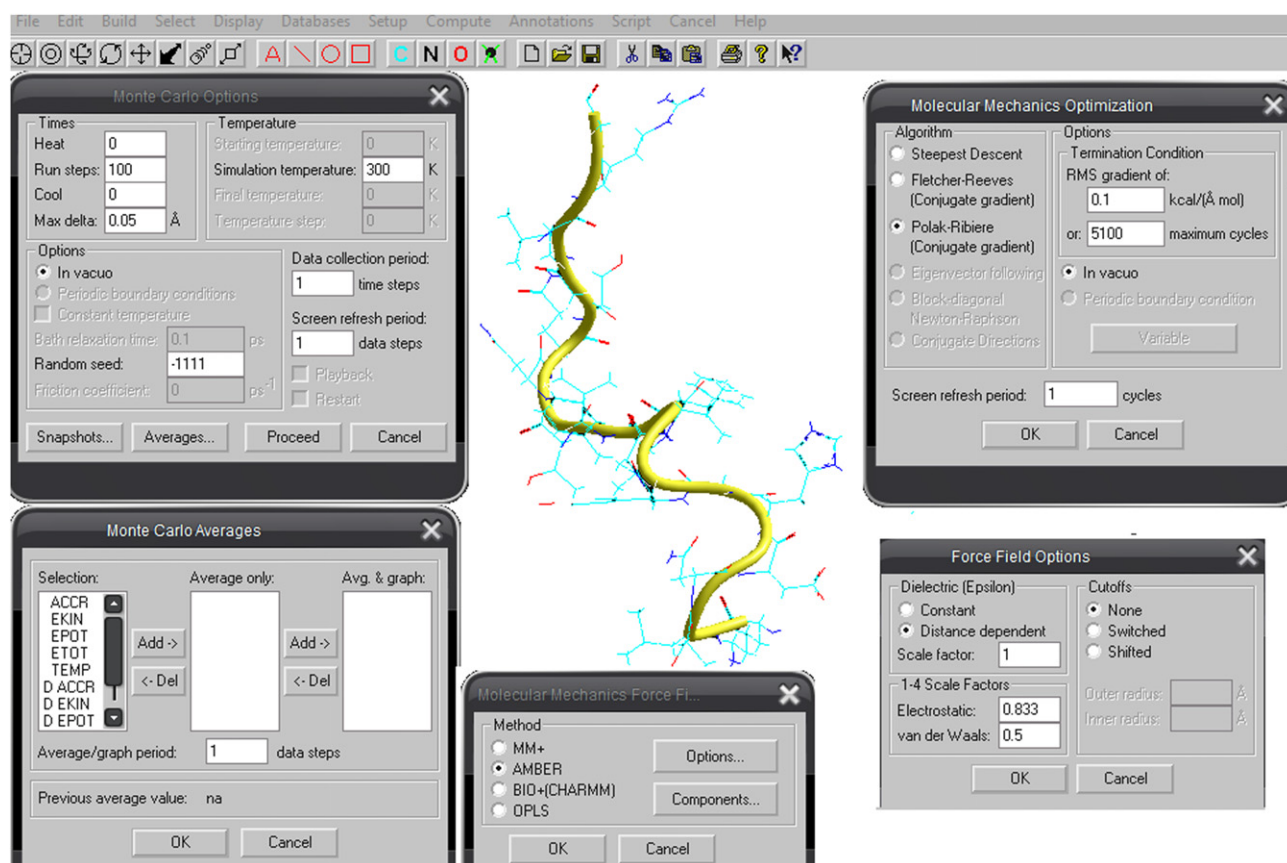


Fig. 2. Snapshot of software HyperChem with MM/MD study for one peptide.

NMR (75 MHz, TFA-*d*):  $\delta$  = 175.01 (CO), 131.92 (C<sub>2</sub><sup>thiophene</sup>), 126.76 (C<sub>4</sub><sup>thiophene</sup>), 123.48 and 122.45 (C<sub>3</sub><sup>thiophene</sup> and C<sub>5</sub><sup>thiophene</sup>), 47.01 (CH), 34.12 (CH<sub>2</sub>) ppm. EIMS: *m/z* (%) = 172 (2) [M + 1]<sup>+</sup>, 171 (17) [M<sup>+</sup>], 112 (100) [M<sup>+</sup> – CH<sub>2</sub>COOH], 111 (3), 110 (17), 85 (54), 84 (4), 83 (3) [M<sup>+</sup> – C<sub>3</sub>H<sub>6</sub>NO<sub>2</sub>], 70 (5), 58 (13). Anal. calcd. for C<sub>7</sub>H<sub>9</sub>NO<sub>2</sub>S (171.21): C 49.10, H 5.30, N 8.18; found C 49.43, H 5.37, N 8.05.

**Compound 3.** (±)-2,2,2-Trifluoro-*N*-(5,6-dihydro-6-oxo-4*H*-cyclopenta[*b*]thiophen-4-yl)acetamide. A solution of **2** (9.50 g, 55.49 mmol) in trifluoroacetic acid (22 mL) was stirred for 30 min at room temperature under argon and then trifluoroacetic anhydride (22 mL) was added. The reaction mixture was refluxed for 5 h and evaporated to dryness and the crude residue (18 g) so obtained was purified on a silica gel column using 1:1 hexane/EtOAc as eluent, to giving pure **3** (10.70 g, 77%) as a white solid. Mp 138–141 °C [Lit. mp 141 °C]. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>):  $\delta$  = 7.97 (d, 1H, *J* = 4.9 Hz, 2-H), 7.17 (d, 1H, *J* = 4.8 Hz, 3-H), 7.00 (br s, 1H, D<sub>2</sub>O exch., NH), 5.59 (t, 1H, *J* = 7.5 Hz, 4-H), 3.50 and 2.83 (AB part of an ABM system, 2H, *J* = 18.6, 7.1, 2.2 Hz, 5-H<sub>2</sub>) ppm. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>):  $\delta$  = 192.78 (C6), 165.31 (CO), 143.05 (C6a), 142.57 (C2), 123.82 (C3), 117.93 (C3a), 114.12 (CF<sub>3</sub>), 48.90 (C5), 46.08 (C4) ppm. EIMS: *m/z* (%) = 250 (6) [M + 1]<sup>+</sup>, 249 (49) [M<sup>+</sup>], 232 (3) [M<sup>+</sup> – OH], 180 (17) [M<sup>+</sup> – CF<sub>3</sub>], 152 (33), 137 (42) [M<sup>+</sup> – NHCOCF<sub>3</sub>], 136 (100) [M<sup>+</sup> – NH<sub>2</sub>COCF<sub>3</sub>], 134 (23), 124 (10), 109 (31), 108 (22), 97 (25), 69 (48). Anal. calcd. for C<sub>9</sub>H<sub>6</sub>F<sub>3</sub>NO<sub>2</sub>S (249.21): C 43.38, H 2.43, N 5.62; found C 43.67, H 2.65, N 5.49.

**Compounds 4a and 4b.** (±)-2,2,2-Trifluoro-*N*-(*cis*-5,6-dihydro-6-hydroxy-4*H*-cyclopenta[*b*]thiophen-4-yl)acetamide and (±)-2,2,2-trifluoro-*N*-(*trans*-5,6-dihydro-6-hydroxy-4*H*-cyclopenta[*b*]thiophen-4-yl)acetamide. NaBH<sub>4</sub> (0.60 g, 15.86 mmol) is added to a solution of compound **3** (2.00 g, 8.02 mmol) in MeOH (8 mL). The resulting mixture is stirred at room temperature for 10 min. The solvent was

removed under reduced pressure and the residue so obtained was partitioned between H<sub>2</sub>O (100 mL) and EtOAc (75 mL). The aqueous phase was extracted with EtOAc (15 × 75 mL), and the pooled organic extracts were dried over Na<sub>2</sub>SO<sub>4</sub>, after which removal of the solvent under reduced pressure afforded 1.72 g of an oily yellow residue that was fractionated chromatographically on a column of silica gel using 6/1 hexane/acetone as eluent. From the first non-void fractions eluted the *cis*-isomer **4a** was isolated (0.65 g, 32%). A second group of fractions afforded a mixture of *trans/cis*-isomers **4a/b** (0.40 g, 20%) and the last group of fractions afforded the *cis*-isomer **4b** (0.62 g, 31%).

**Compound (±)-4a:** mp 135–136 °C. IR (KBr):  $\nu$  = 3272, 3093, 2893, 1697, 1555, 1190, 1041, 995, 939 cm<sup>−1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  = 9.78 (d, 1H, D<sub>2</sub>O exch., *J* = 7.5 Hz, NH), 7.52 (d, 1H, *J* = 4.9 Hz, 2-H), 6.81 (d, 1H, *J* = 4.9 Hz, 3-H), 5.55 (d, 1H, D<sub>2</sub>O exch., *J* = 6.2 Hz, OH), 5.11–5.01 (m, 2H, 4-H + 6-H), 3.15 (dt, 1H, *J* = 13.3, 7.4 Hz, 5-H<sub>H</sub>), 2.14 (dt, 1H, *J* = 13.3, 5.5 Hz, 5-H<sub>HH</sub>) ppm. <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  = 148.42 (CO), 145.45 (C6a), 131.74 (C2), 121.74 (C3), 118.19 (C3a), 114.35 (CF<sub>3</sub>), 69.11 (C6), 48.62 (C4), 46.87 (C5) ppm. EIMS: *m/z* (%) = 251 (3) [M<sup>+</sup>], 233 (100) [M<sup>+</sup> – H<sub>2</sub>O], 139 (23) [M<sup>+</sup> – C<sub>2</sub>HF<sub>3</sub>NO], 138 (17) [M<sup>+</sup> – C<sub>2</sub>H<sub>2</sub>F<sub>3</sub>NO], 137 (31), 136 (49), 121 (20), 110 (43), 97 (10), 69 (31). Anal. calcd. for C<sub>9</sub>H<sub>6</sub>F<sub>3</sub>NO<sub>2</sub>S (251.22): C 43.03, H 3.21, N 5.58; found C 43.37, H 3.43, N 5.65.

**Compound (±)-4b:** mp 118–121 °C. IR (KBr):  $\nu$  = 3299, 2929, 1695, 1550, 1183 cm<sup>−1</sup>. <sup>1</sup>H NMR (300 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  = 9.71 (d, 1H, D<sub>2</sub>O exch., *J* = 7.6 Hz, NH), 7.53 (d, 1H, *J* = 4.9 Hz, 2-H), 6.83 (d, 1H, *J* = 4.9 Hz, 3-H), 5.45 (d, 1H, D<sub>2</sub>O exch., *J* = 6.4 Hz, OH), 5.37–5.30 (m, 1H, 6-H), 5.28–5.24 (m, 1H, 4-H), 2.68–2.52 (m, 2H, 5-H<sub>2</sub>) ppm. <sup>13</sup>C NMR (75 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  = 148.73 (CO), 146.46 (C6a), 132.01 (C2), 121.80 (C3), 118.19 (C3a), 114.36 (CF<sub>3</sub>), 69.76 (C6), 49.82 (C4), 47.28 (C5) ppm. EIMS: *m/z* (%) = 233 (1) [M<sup>+</sup> – H<sub>2</sub>O], 139 (16) [M<sup>+</sup> – C<sub>2</sub>HF<sub>3</sub>NO], 138 (100) [M<sup>+</sup> – C<sub>2</sub>H<sub>2</sub>F<sub>3</sub>NO], 137 (11), 136

(11), 121 (8), 110 (18), 69 (22). Anal. calcd. for  $C_9H_8F_3NO_2S$  (251.02): C 43.03, H 3.21, N 5.58; found C 43.25, H 3.47, N 5.69.

**Compound 5.** (±)-*cis*-4-Amino-5,6-dihydro-4*H*-cyclopenta[*b*]thiophen-6-ol. Compound **4a** (0.30 g, 1.19 mmol) was suspended in a mixture of MeOH/H<sub>2</sub>O (4.5/2.6 mL), and K<sub>2</sub>CO<sub>3</sub> (0.33 g, 2.39 mmol) was added. The mixture was heated to reflux with stirring for 1 h. The reaction mixture was cooled to room temperature and the solvents were evaporated to dryness and the residue so obtained was partitioned between H<sub>2</sub>O (30 mL) and EtOAc (50 mL). The aqueous phase was extracted with EtOAc (4 × 50 mL). The organic phase was separated, dried (Na<sub>2</sub>SO<sub>4</sub>), and evaporated to dryness to give **5a** as a waxy yellow solid (0.18 g, 99%). Mp 101–104 °C. IR and <sup>1</sup>H NMR were coincidences with Lit. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ = 151.25 (C6a), 146.56 (C3a), 131.75 (C2), 121.52 (C3), 70.85 (C6), 51.50 (C5), 51.42 (C4) ppm. EIMS: *m/z* (%) = 155 (2) [M<sup>+</sup>], 138 (22) [M<sup>+</sup> – OH], 137 (15) [M<sup>+</sup> – H<sub>2</sub>O], 136 (36), 112 (100), 111 (16), 110 (46), 109 (13), 85 (19), 69 (11), 58 (10). Anal. calcd. for C<sub>7</sub>H<sub>9</sub>NOS (155.22): C 54.17, H 5.84, N 9.02; found C 54.40, H 5.70, N 9.19.

**Compound 5b.** (±)-*trans*-4-Amino-5,6-dihydro-4*H*-cyclopenta[*b*]thiophen-6-ol. Compound **5b** was obtained, as the sole product, as an orange waxy solid in the same way as **5a**, from **4b** under slightly different reaction conditions (reflux 6 h). The residue obtained was fractionated chromatographically on a column of silica gel using 2/1 CH<sub>2</sub>Cl<sub>2</sub>/MeOH as eluent. From the non-void fractions eluted the *trans*-isomer **5b**, as a orange waxy solid (75%). IR concordat with Lit. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>): δ = 7.44 (d, 1H, *J* = 4.9 Hz, 2-H), 6.91 (d, 1H, *J* = 4.9 Hz, 3-H), 5.16–5.14 (m, 1H, 6-H), 4.33 (t, 1H, *J* = 5.8 Hz, 4-H), 3.98–3.23 (br s, 3H, D<sub>2</sub>O exch., OH + NH<sub>2</sub>), 2.46–2.42 (m, 1H, 5-HH), 2.32–1.97 (m, 1H, 5-HH) ppm. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ = 152.66 (C6a), 146.07 (C3a), 130.93 (C2), 122.12 (C3), 69.68 (C6), 51.65 (C5), 51.54 (C4) ppm. EIMS: *m/z* (%) = 155 (1) [M<sup>+</sup>], 139 (27), 138 (33) [M<sup>+</sup> – OH], 137 (13) [M<sup>+</sup> – H<sub>2</sub>O], 136 (33), 123 (3), 121 (8), 113 (8), 112 (100), 110 (57), 109 (16), 85 (26), 69 (26). HRMS *m/z* calcd. for C<sub>7</sub>H<sub>9</sub>NOS, 155.0405; found, 155.0421.

**Compounds 6a, 7a.** (±)-*cis*-5,6-Dihydro-4-(prop-2-ynylamino)-4*H*-cyclopenta[*b*]thiophen-6-ol (**6a**) and (±)-*cis*-5,6-dihydro-4-(diprop-2-ynylamino)-4*H*-cyclopenta[*b*]thiophen-6-ol (**7a**). To a solution of **5a** (0.64 g, 4.13 mmol) and K<sub>2</sub>CO<sub>3</sub> (0.57 g, 4.13 mmol) in MeCN (13 mL) under argon was added dropwise a solution of propargyl bromide (0.46 mL, 4.13 mmol). The resulting suspension was heated to reflux with stirring for 18 h; whereafter most of volatiles were partitioned between EtOAc (50 mL) and 2 N NaOH (50 mL), the organic phase was extracted with 2 N NaOH (2 × 50 mL). The organic phase was separated, dried (Na<sub>2</sub>SO<sub>4</sub>), and evaporated to dryness. The residue (0.64 g) was flash chromatographed on silica gel, using 3/1 and 1/2 hexane/EtOAc as successive mixtures of eluents. From the first non-void fractions eluted the compound **7a** (0.34 g, 36%) was isolated. A second group of fractions afforded the compound **6a** (0.19 g, 24%).

**Compound (±)-6a:** white solid, mp 59–63 °C. IR (KBr): ν = 3290, 3220, 2922, 2850, 1599, 1494, 1429, 1339, 1183, 1113, 1086, 1060, 1012, 954, 762 cm<sup>−1</sup>. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>): δ = 7.33 (d, 1H, *J* = 4.8 Hz, 2-H), 6.92 (d, 1H, *J* = 5.0 Hz, 3-H), 5.06 (dd, 1H, *J* = 6.7, 2.0 Hz, 6-H), 4.23 (dd, 1H, *J* = 6.7, 2.0 Hz, 4-H), 3.44 (d, 2H, *J* = 2.3 Hz, CH<sub>2</sub>), 3.15 (br. s, 2H, D<sub>2</sub>O exch., OH + NH), 3.05–2.95 (m, 1H, 5-HH), 2.27 (t, 1H, *J* = 2.2 Hz, C≡CH), 2.09 (dt, 1H, *J* = 14.5, 2.0 Hz, 5-HH) ppm. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ = 148.27 (C6a), 147.41 (C3a), 131.19 (C2), 121.62 (C3), 81.33 (C≡CH), 72.20 (C≡CH), 70.60 (C6), 55.38 (C4), 47.62 (C5), 35.89 (CH<sub>2</sub>C≡CH) ppm. EIMS: *m/z* (%) = 193 (3) [M<sup>+</sup>], 192 (16) [M – 1]<sup>+</sup>, 176 (5) [M<sup>+</sup> – OH], 175 (5) [M<sup>+</sup> – H<sub>2</sub>O], 154 (37) [M<sup>+</sup> – CH<sub>2</sub>C≡CH], 140 (11), 139 (100), 138 (20), 137 (19), 136 (19), 135 (4), 122 (16), 121 (13), 111 (18), 110 (16), 109 (12), 85 (11), 77 (11), 69 (10), 65 (10), 57 (12), 55 (12). Anal. calcd. for C<sub>10</sub>H<sub>11</sub>NOS (193.26): C 62.15, H 5.74, N 7.25; found C 62.36, H 5.92, N 7.41.

**Compound (±)-7a:** brown, waxy solid, mp 104–108 °C. IR (KBr): ν = 3285, 3269, 3125, 3077, 2970, 2921, 2844, 2824, 1444, 1333, 1122, 1077, 1033, 904 cm<sup>−1</sup>. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>): δ = 7.34 (d, 1H, *J* = 4.9 Hz, 2-H), 7.00 (d, 1H, *J* = 4.9 Hz, 3-H), 5.09–5.07 (m, 1H, 6-H), 4.19 (dd, 1H, *J* = 6.9, 2.8 Hz, 4-H), 3.63–3.50 (m, 4H, 2 × CH<sub>2</sub>), 2.97–2.87 (m, 1H, 5-HH), 2.82 (br s, 1H, D<sub>2</sub>O exch., OH), 2.45 (dt, 1H, *J* = 14.2, 2.7 Hz, 5-HH), 2.26 (t, 2H, *J* = 2.2 Hz, 2 × C≡CH) ppm. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ = 147.83 (C6a), 146.77 (C3a), 130.95 (C2), 122.88 (C3), 79.47 (2 × C≡CH), 73.50 (2 × C≡CH), 70.27 (C6), 60.86 (C4), 43.90 (C5), 39.85 (2 × CH<sub>2</sub>C≡CH) ppm. EIMS: *m/z* (%) = 231 (1) [M<sup>+</sup>], 230 (4) [M – 1]<sup>+</sup>, 214 (4) [M<sup>+</sup> – OH], 213 (3) [M<sup>+</sup> – H<sub>2</sub>O], 194 (9), 193 (19), 192 (100) [M<sup>+</sup> – CH<sub>2</sub>C≡CH], 188 (11), 174 (19), 173 (11), 139 (47), 138 (9), 137 (14), 136 (13), 135 (7), 122 (17), 121 (13), 111 (21), 110 (17), 85 (16), 84 (12), 71 (16), 57 (17), 55 (10). Anal. calcd. for C<sub>13</sub>H<sub>13</sub>NOS (231.31): C 67.50, H 5.66, N 6.06; found C 67.69, H 5.88, N 6.13.

**Compounds 6b, 7b.** (±)-*trans*-5,6-Dihydro-4-(prop-2-ynylamino)-4*H*-cyclopenta[*b*]thiophen-6-ol (**6b**) and (±)-*trans*-5,6-dihydro-4-(diprop-2-ynylamino)-4*H*-cyclopenta[*b*]thiophen-6-ol (**7b**). Compound **6b** and **7b** were obtained in the same way as **6a** and **7a** from **5b** (0.75 g, 4.84 mmol) but under slightly different reaction conditions [reflux 4.5 h]. The yellow oil obtained (0.8 g) was fractionated chromatographically on a column of silica gel using 3/1 and 1/2 hexane/EtOAc as successive mixtures of eluents. From the first non-void fractions eluted the compound **7b** (0.49 g, 53%) was isolated. A second group of fractions afforded the compound **6b** (0.14 g, 13%).

**Compound (±)-6b:** mp 71–74 °C. IR (KBr): ν = 3268, 3254, 2927, 2853, 1486, 1432, 1336, 1301, 1193, 1115, 1097, 1034 cm<sup>−1</sup>. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>): δ = 7.34 (d, 1H, *J* = 4.9 Hz, 2-H), 6.92 (d, 1H, *J* = 4.9 Hz, 3-H), 5.41–5.38 (m, 1H, 6-H), 4.55–4.51 (m, 1H, 4-H), 3.46 (d, 2H, *J* = 2.3 Hz, CH<sub>2</sub>), 2.61–2.56 (m, 2H, 5-H<sub>2</sub>), 2.26 (t, 1H, *J* = 2.3 Hz, C≡CH), 2.11 (br s, 2H, D<sub>2</sub>O exch., OH + NH) ppm. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ = 149.11 (C6a), 146.44 (C3a), 131.41 (C2), 121.70 (C3), 81.72 (C≡CH), 71.86 (C≡CH), 71.26 (C6), 56.52 (C4), 49.07 (C5), 36.11 (CH<sub>2</sub>) ppm. EIMS: *m/z* (%) = 192 (3) [M – 1]<sup>+</sup>, 176 (6) [M<sup>+</sup> – OH], 175 (2) [M<sup>+</sup> – H<sub>2</sub>O], 174 (10), 155 (9), 154 (100) [M<sup>+</sup> – CH<sub>2</sub>C≡CH], 150 (15), 139 (11), 138 (7), 137 (12), 136 (21), 135 (3), 111 (8), 110 (13), 109 (7), 85 (4), 77 (5), 65 (6). Anal. calcd. for C<sub>10</sub>H<sub>11</sub>NOS (193.26): C 62.15, H 5.74, N 7.25; found C 62.32, H 5.86, N 7.19.

**Compound (±)-7b:** mp 57–59 °C. IR (KBr): ν = 3287, 3203, 2949, 2829, 1440, 1419, 1393, 1365, 1126, 1115, 1031 cm<sup>−1</sup>. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>): δ = 7.37 (d, 1H, *J* = 4.7 Hz, 2-H), 7.00 (d, 1H, *J* = 4.7 Hz, 3-H), 5.49–5.40 (m, 1H, 6-H), 4.64–4.61 (m, 1H, 4-H), 3.55–3.43 (m, 4H, 2 × CH<sub>2</sub>), 2.97 (ddd, 1H, *J* = 14.5, 6.7, 3.7 Hz, 5-HH), 2.45 (ddd, 1H, *J* = 14.5, 7.2, 3.4 Hz, 5-HH), 2.26–2–0.24 (m, 2H, 2 × C≡CH), 1.95 (br s, 1H, D<sub>2</sub>O exch., OH) ppm. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>): δ = 147.47 (C6a), 147.13 (C3a), 131.27 (C2), 122.93 (C3), 79.65 (2 × C≡CH), 73.13 (2 × C≡CH), 71.41 (C6), 62.08 (C4), 44.25 (C5), 39.51 (2 × CH<sub>2</sub>) ppm. EIMS: *m/z* (%) = 232 (1) [M + 1]<sup>+</sup>, 231 (7) [M<sup>+</sup>], 230 (7) [M – 1]<sup>+</sup>, 214 (6) [M<sup>+</sup> – OH], 213 (2) [M<sup>+</sup> – H<sub>2</sub>O], 194 (18), 193 (39), 192 (100) [M<sup>+</sup> – CH<sub>2</sub>C≡CH], 188 (22), 174 (24), 173 (11), 162 (11), 149 (10), 148 (15), 147 (11), 140 (10), 139 (69), 138 (22), 137 (28), 136 (27), 135 (13), 134 (11), 122 (24), 121 (25), 111 (36), 110 (40), 109 (18), 85 (10), 84 (7), 78 (14), 77 (23), 69 (10), 67 (12), 66 (19), 65 (21), 55 (13). Anal. calcd. for C<sub>13</sub>H<sub>13</sub>NOS (231.31): C 67.50, H 5.66, N 6.06; found C 67.77, H 5.81, N 6.19.

**Compound 8a.** (±)-*cis*-4-[(Diprop-2-ynyl)amino]-5,6-dihydro-4*H*-cyclopenta[*b*]thiophen-6-yl acetate. The compound **7a** (50 mg, 0.216 mmol) was stirred with Ac<sub>2</sub>O (0.5 mL) and dry Et<sub>3</sub>N (0.5 mL) for 5.5 h under argon at room temperature, the resulting mixture was concentrated to dryness, and the solid residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (30 mL) and the organic phase was washed successively with saturated NaHCO<sub>3</sub> (3 × 20 mL) and H<sub>2</sub>O (3 × 20 mL), dried (Na<sub>2</sub>SO<sub>4</sub>), and concentrated under reduced pressure. The resulting yellow oil (60 mg) was chromatographed on silica gel with 10/1

hexane/EtOAc as eluent, affording **8a** (51 mg, 86%) as a white solid, mp 59–60 °C. IR (KBr):  $\nu$  = 3281, 3262, 2922, 1721, 1428, 1367, 1305, 1245  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 7.38 (d, 1H,  $J$  = 5.0 Hz, 2-H), 6.97 (d, 1H,  $J$  = 5.0 Hz, 3-H), 5.83 (dd, 1H,  $J$  = 7.6, 2.9 Hz, 6-H), 4.42 (dd, 1H,  $J$  = 7.9, 3.2 Hz, 4-H), 3.60–3.44 (AB system, 2H,  $J$  = 17.0, 2.3 Hz,  $\text{CH}_2$ ), 3.59–3.43 (AB system, 2H,  $J$  = 17.0, 2.3 Hz,  $\text{CH}_2$ ), 3.11–3.01 (m, 1H, 5-HH), 2.57 (dt, 1H,  $J$  = 15.2, 3.2 Hz, 5-HH), 2.24 (t, 2H,  $J$  = 2.3 Hz,  $2 \times \text{C}\equiv\text{CH}$ ), 2.08 (s, 3H,  $\text{CH}_3$ ) ppm.  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 171.22 (CO), 148.58 (C6a), 142.62 (C3a), 132.69 (C2), 122.31 (C3), 80.03 ( $2 \times \text{C}\equiv\text{CH}$ ), 72.98 and 72.51 ( $2 \times \text{C}\equiv\text{CH}$ ), 72.47 (C6), 61.18 (C4), 39.33 (C5), 38.63 ( $2 \times \text{CH}_2$ ), 21.12 ( $\text{C}_7\text{H}_3$ ) ppm. EIMS:  $m/z$  (%) = 273 (2) [ $\text{M}^+$ ], 234 (12) [ $\text{M}^+ - \text{CH}_2\text{C}\equiv\text{CH}$ ], 230 (62), 215 (15), 214 (79), 213 (98), 212 (100), 200 (11), 199 (9), 198 (22), 192 (16), 186 (11), 176 (13), 175 (28), 174 (94), 173 (26), 148 (14), 147 (16), 139 (48), 138 (13), 137 (23), 136 (24), 135 (13), 134 (10), 123 (21), 122 (93), 121 (90), 111 (18), 110 (9), 109 (11), 92 (28), 77 (12), 66 (15), 65 (11). Anal. calcd. for  $\text{C}_{15}\text{H}_{15}\text{NO}_2\text{S}$  (273.35): C 65.91, H 5.53, N 5.12; found C 66.05, H 5.70, N 5.08.

**Compound 8b.** ( $\pm$ )-*trans*-4-[(Diprop-2-ynyl)amino]-5,6-dihydro-4H-cyclopenta[b]thiophen-6-yl acetate. The compound **8b** was obtained in the same way as **8a** from **7b** (53 mg, 0.23 mmol) but under slightly different reaction conditions [room temperature 15 h]. The obtained residue (62 mg) was fractionated chromatographically on a column of silica gel using 10/1 hexane/EtOAc as eluent. From the non-void fractions eluted the compound **8b** (46 mg, 74%) as a yellow oil. IR (film):  $\nu$  = 3286, 1726, 1427, 1371, 1235, 1015  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 7.39 (d, 1H,  $J$  = 5.0 Hz, 2-H), 6.98 (d, 1H,  $J$  = 5.0 Hz, 3-H), 6.09–6.07 (m, 1H, 6-H), 4.68 (ddd,  $J$  = 7.2, 4.7, 1.5 Hz, 1H, 4-H), 3.55–3.42 (m, 4H,  $2 \times \text{CH}_2$ ), 2.97 (ddd, 1H,  $J$  = 14.6, 7.3, 4.4 Hz, 5-HH), 2.60 (ddd, 1H,  $J$  = 14.6, 7.3, 2.9 Hz, 5-HH), 2.25 (t, 2H,  $J$  = 2.3 Hz,  $2 \times \text{C}\equiv\text{CH}$ ), 2.04 (s, 3H,  $\text{CH}_3$ ) ppm.  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 171.21 (CO), 149.09 (C6a), 142.72 (C3a), 132.65 (C2), 122.30 (C3), 79.74 ( $2 \times \text{C}\equiv\text{CH}$ ), 73.95 (C6), 73.11 ( $2 \times \text{C}\equiv\text{CH}$ ), 62.32 (C4), 39.59 (C5), 39.32 ( $2 \times \text{CH}_2$ ), 21.09 ( $\text{CH}_3$ ) ppm. EIMS:  $m/z$  (%) = 272 (1) [ $\text{M} - 1$ ] $^+$ , 234 (5) [ $\text{M}^+ - \text{CH}_2\text{C}\equiv\text{CH}$ ], 230 (34), 215 (10), 214 (49), 213 (100), 212 (58), 199 (6), 198 (15), 192 (10), 186 (7), 176 (8), 175 (19), 174 (63), 173 (18), 148 (11), 147 (12), 139 (31), 138 (9), 137 (18), 136 (18), 135 (10), 134 (8), 123 (13), 122 (42), 121 (66), 111 (12), 110 (6), 109 (9), 92 (10), 77 (9), 66 (14), 65 (10). HRMS  $m/z$  calcd. for  $\text{C}_{15}\text{H}_{15}\text{NO}_2\text{S}$ , 273.0823; found, 273.0841.

**Compound 9a.** ( $\pm$ )-*cis*-4-[(Diprop-2-ynyl)amino]-5,6-dihydro-4H-cyclopenta[b]thiophen-6-yl benzoate. The compound **7a** (50 mg, 0.216 mmol) was stirred with  $\text{BzCl}$  (37.61  $\mu\text{L}$ , 0.324 mmol), dry  $\text{Et}_3\text{N}$  (60.21  $\mu\text{L}$ , 0.432 mmol) and a catalytic amount of DMAP in MeCN (2.16 mL) for 23 h under argon at room temperature. The resulting mixture was concentrated to dryness, and the solid residue was dissolved in EtOAc (30 mL), the solid formed was collected by filtration and the filtrate was concentrated under reduced pressure and the resulting orange oil (130 mg) was chromatographed on silica gel with 20/1 hexane/EtOAc as eluent, affording **9a** (58 mg, 80%) as a yellow oil. IR (film):  $\nu$  = 3290, 2922, 2818, 1707, 1600, 1584, 1265, 1107, 1069, 1024, 1010, 710  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 8.04 (d, 2H,  $J$  = 7.8 Hz, 2- $\text{H}_{\text{benzene}}$  + 6- $\text{H}_{\text{benzene}}$ ), 7.54 (t, 1H,  $J$  = 7.3 Hz, 4- $\text{H}_{\text{benzene}}$ ), 7.44–7.38 (m, 3H, 3- $\text{H}_{\text{benzene}}$  + 5- $\text{H}_{\text{benzene}}$  + 2-H), 6.98 (d, 1H,  $J$  = 4.9 Hz, 3-H), 6.09 (dd,  $J$  = 7.7, 2.1 Hz, 1H, 6-H), 4.50 (dd, 1H,  $J$  = 7.8, 3.1 Hz, 4-H), 3.63–3.49 (AB system, 2H,  $J$  = 17.0, 2.6 Hz,  $\text{CH}_2$ ), 3.62–3.48 (AB system, 2H,  $J$  = 17.0, 2.6 Hz,  $\text{CH}_2$ ), 3.20–3.12 (m, 1H, 5-HH), 2.77 (m, 1H, 5-HH), 2.23 (br s, 2H,  $2 \times \text{C}\equiv\text{CH}$ ) ppm.  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 166.61 (CO), 148.80 (C6a), 142.61 (C3a), 133.10 (C4'), 132.86 (C2), 129.94 (C1'), 129.67 and 128.37 (C2' + C6' and C3' + C5'), 122.28 (C3), 80.07 ( $2 \times \text{C}\equiv\text{CH}$ ), 73.04 (C6), 72.97 ( $2 \times \text{C}\equiv\text{CH}$ ), 61.32 (C4), 39.32 ( $2 \times \text{CH}_2$ ), 38.59 (C5) ppm. EIMS:  $m/z$  (%) = 335 (1) [ $\text{M}^+$ ], 334 (1) [ $\text{M} - 1$ ] $^+$ , 296 (1) [ $\text{M}^+ - \text{CH}_2\text{C}\equiv\text{CH}$ ], 230 (45), 215 (5), 214 (29), 213 (40), 212 (45), 200 (3), 199 (4), 198 (10),

186 (5), 176 (5), 175 (13), 174 (48), 173 (11), 148 (7), 147 (8), 139 (4), 138 (2), 137 (10), 136 (12), 135 (7), 134 (5), 123 (16), 122 (100), 121 (52), 105 (60), 92 (34), 77 (37). HRMS  $m/z$  calcd. for  $\text{C}_{20}\text{H}_{17}\text{NO}_2\text{S}$ , 335.0980; found, 335.0996.

**Compound 9b.** ( $\pm$ )-*trans*-4-[(Diprop-2-ynyl)amino]-5,6-dihydro-4H-cyclopenta[b]thiophen-6-yl benzoate. The compound **9b** was obtained in the same way as **9a** from **7b** (53 mg, 0.23 mmol) but under slightly different reaction conditions [room temperature 21 h]. The obtained residue (62 mg) was fractionated chromatographically on a column of silica gel using 20/1 hexane/EtOAc as eluent. From the non-void fractions eluted the compound **9b** (72 mg, 99%) as a yellow oil. IR (film):  $\nu$  = 3290, 2923, 2819, 1707, 1600, 1450, 1266, 1107, 1069, 991  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 8.02–7.99 (m, 2H, 2- $\text{H}_{\text{benzene}}$  + 6- $\text{H}_{\text{benzene}}$ ), 7.55–7.50 (m, 1H, 4- $\text{H}_{\text{benzene}}$ ), 7.42–7.38 (m, 3H, 3- $\text{H}_{\text{benzene}}$  + 5- $\text{H}_{\text{benzene}}$  + 2-H), 6.99 (d, 1H,  $J$  = 5.1 Hz, 3-H), 6.32 (d,  $J$  = 6.6 Hz, 1H, 6-H), 4.75 (ddd,  $J$  = 7.4, 4.3, 1.2 Hz, 1H, 4-H), 3.56–3.47 (AB system, 2H,  $J$  = 16.8, 2.3 Hz,  $\text{CH}_2$ ), 3.55–3.46 (AB system, 2H,  $J$  = 16.8, 2.4 Hz,  $\text{CH}_2$ ), 3.09 (ddd, 1H,  $J$  = 14.6, 7.2, 4.3 Hz, 5-HH), 2.76 (ddd, 1H,  $J$  = 14.6, 7.2, 2.7 Hz, 5-HH), 2.25 (t, 2H,  $J$  = 2.3 Hz,  $2 \times \text{C}\equiv\text{CH}$ ) ppm.  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 166.68 (CO), 149.15 (C6a), 142.78 (C3a), 133.06 (C4 $_{\text{benzene}}$ ), 132.75 (C2), 129.95 (C1 $_{\text{benzene}}$ ), 129.65 and 128.30 [C2 $_{\text{benzene}}$  + C3 $_{\text{benzene}}$  + C5 $_{\text{benzene}}$  + C6 $_{\text{benzene}}$ ], 122.29 (C3), 79.72 ( $2 \times \text{C}\equiv\text{CH}$ ), 74.57 (C6), 73.10 ( $2 \times \text{C}\equiv\text{CH}$ ), 62.37 (C4), 39.71 (C5), 39.34 ( $2 \times \text{CH}_2$ ) ppm. EIMS:  $m/z$  (%) = 334 (1) [ $\text{M} - 1$ ] $^+$ , 296 (1) [ $\text{M}^+ - \text{CH}_2\text{C}\equiv\text{CH}$ ], 230 (37), 215 (8), 214 (36), 213 (100), 212 (54), 200 (4), 199 (5), 198 (15), 186 (6), 176 (7), 175 (16), 174 (53), 173 (17), 148 (9), 147 (13), 139 (4), 138 (3), 137 (14), 136 (15), 135 (10), 134 (7), 123 (15), 122 (65), 121 (57), 105 (90), 92 (10), 78 (10), 77 (67). HRMS  $m/z$  calcd. for  $\text{C}_{20}\text{H}_{17}\text{NO}_2\text{S}$ , 335.0980; found, 335.0992.

**MAO Inhibition Assay of rasagiline analogs.** The potential effects of the test drugs on hMAO activity were investigated by measuring their effects on the production of hydrogen peroxide from *p*-tyramine (a common substrate for both hMAO-A and hMAO-B), using the 10-acetyl-3,7-dihydroxyphenoxazine as reagent and microsomal MAO isoforms prepared from insect cells (BTI-TN-5B1-4) infected with recombinant baculovirus containing cDNA inserts for hMAO-A or hMAO-B [36,37]. The production of  $\text{H}_2\text{O}_2$  catalyzed by MAO isoforms can be detected using the previously mentioned reagent, a nonfluorescent, highly sensitive, and stable probe that reacts with  $\text{H}_2\text{O}_2$  in the presence of horseradish peroxidase to produce a fluorescent product, resorufin. In this study, hMAO activity was evaluated using the above method following the general procedure described previously by us. The tested drugs (new compounds and reference inhibitors) inhibited the control enzymatic MAO activities and the inhibition was concentration dependent. The corresponding  $\text{IC}_{50}$  values and MAO-B selectivity ratios [ $\text{IC}_{50}$  (MAO-A)]/[ $\text{IC}_{50}$  (MAO-B)] are shown in Table 5. The assayed compounds themselves do not react directly with the 10-acetyl-3,7-dihydroxyphenoxazine, which indicates that these drugs do not interfere with the measurements. In our experiments and under our experimental conditions, hMAO-A displayed a Michaelis constant ( $\text{SEM}$ ) of 457.17 (38.62  $\mu\text{M}$ ) and a maximum reaction velocity ( $\text{SD}$ ) of 185.67 (12.06 nmol/min/mg protein), whereas hMAO-B showed a  $\text{SEM}$  of 220.33 (32.80  $\mu\text{M}$ ) and a  $\text{SD}$  of 24.32, 1.97 nmol/min/mg protein ( $n = 5$ ).

## 2.2.2. Experimental study *Fasciola* protein fingerprints

**2.2.2.1. Experimental methods. Obtaining the Excretory–Secretory Antigens (ESAs) of *F. hepatica*.** ESAs of *F. hepatica* were obtained as previously described by Mezo et al. [62]. Briefly, liver adult flukes collected from bile ducts of naturally infected cows were washed twice, first in sterile saline solution containing antibiotics (100 IU/mL penicillin and 100  $\mu\text{g}$  streptomycin) and glucose (2 mg/mL) at 38 °C, and then in RPMI 1640 medium, supplemented with 20 mM



HEPES, 0.3 g/L L-glutamine, 2 g/L sodium bicarbonate and antibiotics at 38 °C under 5% CO<sub>2</sub> in air. Flukes were then transferred to a 75 cm<sup>2</sup> tissue culture flask (Iwaki, Sciteck Div. Asahi Techno Glass, Funabashi City, Chiba, Japan) and maintained in culture medium (3 mL/fluke) at 38 °C under 5% CO<sub>2</sub> in air. After 24 h of incubation, the medium containing ESAs was removed, the protease activity inhibited with a protease inhibitor cocktail (SigmaFAST™ Protease Cocktail Tablet, Sigma–Aldrich, Madrid, Spain), and centrifuged at 10,000 g for 20 min at 4 °C. Then, the supernatant was passed through a 0.45 µm pore filter disc, and submitted to a process of washing/concentration with distilled water using Microsep Microconcentrators (3000 molecular weight cut-off; Filtron Technology Corporation, Northborough, Massachusetts) and the concentrate dried using a vacuum concentrator (SpeedVac, Thermo Scientific, Barcelona, Spain) and stored at 4 °C until use.

**Size-Exclusion Chromatography (SEC).** Samples of ESAs of *F. hepatica* were dissolved in PBS at a final protein concentration of 3.5 mg/mL and fractionated by size-exclusion chromatography using an FPLC system (ÄKTA Basic 10, Amersham Biosciences Europe BmbH, Barcelona, Spain) equipped with a column of Superdex 75 HR 10/30 column (Amersham Biosciences) as described previously [62]. Briefly, 0.5 mL samples were eluted at a flow rate of 0.3 mL/min and the protein concentration was monitored at 280 nm. The molecular weight of the eluted proteins was estimated using a mixture of proteins of known molecular weight (Gel Filtration LMW Calibration Kit, Amersham Biosciences). Finally, the protein concentration of each peak obtained was measured by Pirogallol Red Method (Sigma) and stored in aliquots at –80 °C until further analysis.

**Polyacrylamide gel electrophoresis (SDS-PAGE).** The proteins from an aliquot of purified *F. hepatica* ESAs corresponding to peak IV (see Results Section), obtained as described above by exclusion chromatography, were resolved by SDS-PAGE under reducing conditions (2.5 µg of protein per lane) using a mini-vertical gel electrophoresis unit (Mighty Small II, Hoefer Inc., Holliston, MA). One-dimensional electrophoresis (1DE) was performed at 200 V constant current at 20 °C for 1 h using the Laemmli (1970) buffer system with discontinuous slab gels made up of a 5% polyacrylamide stacking gel and a 10–20% linear-gradient polyacrylamide resolving gel, of 1 mm of thickness. 1D-SDS-PAGE molecular weight standards (BioRad) were used to calibrate gels.

**Mass Spectrometry (MS).** Protein bands in the polyacrylamide gels were stained with Imperial Protein Stain (Pierce) overnight and the dye excess removed with distilled water. Then the selected bands (band #2 in this study; see Results Section) were excised using a scalpel, washed twice with water, shrunk 15 min with 100% acetonitrile and dried in a Savant SpeedVac for 30 min. Then, the samples were reduced with 10 mM dithioerythritol in 25 mM ammonium bicarbonate for 30 min at 56 °C and subsequently alkylated with 55 mM iodoacetamide in 25 mM ammonium bicarbonate for 15 min in the dark. Finally, samples were digested with 12.5 ng/µl sequencing grade trypsin (Roche Molecular Biochemicals) in 25 mM ammonium bicarbonate (pH 8.5) overnight at 37 °C. After digestion, the supernatant was collected and 1 µl was spotted onto a MALDI target plate and allowed to air-dry at room temperature. Then, 0.4 µl of a 3 mg/mL of α-cyano-4-hydroxy-transcinnamic acid matrix (Sigma–Aldrich) in 50% acetonitrile were added to the dried peptide digest spots and allowed again to air-dry at room temperature.

MALDI-TOF MS analyses were performed in a 4800 Proteomics Analyzer MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Framingham, MA) at the Genomics and Proteomics Center, Universidad Complutense de Madrid, operated in positive reflector mode, with an accelerating voltage of 20,000 V. All mass spectra were calibrated internally using peptides from the auto digestion of

trypsin. The MALDI-TOF/TOF mass spectrometry analysis produces peptide mass fingerprints. Those observed with a signal-to-noise greater than 10 were collated and represented as a list of mono-isotopic molecular weights. Proteins ambiguously identified by peptide mass fingerprints, were subjected to MS/MS sequencing analyses using the 4800 Proteomics Analyzer (Applied Biosystems, Framingham, MA). The suitable MS spectra precursors for MS/MS sequencing analyses were selected, and fragmentation was carried out using the acquisition method in the 1 kV ion reflector mode, collision induced dissociation (CID) on, and precursor mass window ±5 Da. The plate model & default calibration were optimized for the MS–MS spectra processing. For protein identification, the non-redundant NCBI database or SwissProt was searched using MASCOT 2.1 ([matrixscience.com](http://matrixscience.com)) through the Global Protein Server v3.5 from Applied Biosystems. The Mascot Search Parameters were: i) Taxonomy: metazoa (animals); ii) Database: SwissProt and NCBI nr; iii) Enzyme: Trypsin, allow up 1 missed trypsin cleavage site; iv) Modifications: carbamidomethyl cysteine as fixed modification and oxidized methionine as variable modification; v) Peptide mass tolerance: 50 ppm (PMF) –100 ppm (MS–MS or Combined search); vi) Peptide charge state: +1 MS–MS; vii) Fragments tolerance: 0.3 Da. The parameters for the combined search (Peptide mass fingerprint and MS–MS spectra) were the same described above. In all protein identification, the probability scores were greater than the score fixed by MASCOT as significant with a *p*-value <0.05. Protein identification by *de novo* sequencing from fragmentation spectra of peptides was performed using *de Novo* tool software (Applied Biosystems). Tentative sequences were manually checked and validated, and the homology search for these sequences was obtained using the BLAST tool at (<http://www.ncbi.nlm.nih.gov/BLAST>).

### 3. Results

#### 3.1. DTPs classification models and complex network assembly

**LDA model.** The present is the first mt-QSAR model for the probability of binding organic compounds to very large diversity of receptors based only on the molecular connectivity of the drug and the protein receptor. One application for the present model is predict the protein receptor or active place with a specific drug and vice versa, predict drugs with their proteins. In both cases, receptor susceptibility identification is imperative. Detailed information on the compounds, predicted classification, and probability of affinity on different receptors of the drugs used to seek the model appears in Table 1SM of the supplementary material. Using this model we can predict the different relationships between the drug–protein connectivity same physicochemical property [32]. Common physicochemical properties have been demonstrated to be useful on protein QSAR [63,64]. This work introduces for the first time a single linear mt-QSAR equation model to classify drugs with your respective protein receptor. The best model found was:

$$\begin{aligned} S(\text{DTP})_{\text{pred}} = & -8.51 \cdot {}^D\theta_2(T) + 8.23 \cdot {}^D\theta_4(T) + 2.01 \cdot {}^D\theta_0(X) \\ & + 0.19 \cdot {}^D\theta_0(\text{Het}) + 21.74 \cdot {}^D\theta_2(H - \text{Het}) - 20.16 \cdot {}^D\theta_5(H - \text{Het}) \\ & + 3.78 \cdot {}^T\theta_0(c) + 0.51 \cdot {}^D\theta_0(c) + 0.34 \cdot {}^T\theta_0(i) + 1.25 \cdot {}^T\theta_3(i) \\ & - 0.45 \cdot {}^T\theta_5(i) + 1.97 \cdot {}^T\theta_1(m) - 4.01 \cdot {}^T\theta_2(m) + 3.74 \cdot {}^T\theta_3(m) \\ & - 0.68 \cdot {}^T\theta_5(m) + 0.19 \cdot {}^T\theta_0(s) - 1.05 \cdot {}^T\theta_2(s) - 1.99N \\ = & 5784 \quad \chi^2 = 2241.061 \quad p < 0.001 \end{aligned} \quad (4)$$

In this model the *N* is the number of cases (DTPs) used to train the model and Chi-square ( $\chi^2$ ) is the statistic used to test the



**Table 1**

Detailed list of the symbols and description for all parameters present in the model.

Molecule	Symbol	Atom group	Descriptor name
Drug	$D\theta(T)$	All atoms	Entropy of drug for all atoms at distance $k \leq 2$
Drug	$D\theta(T)$	All atoms	Entropy of drug for all atoms at distance $k \leq 4$
Drug	$D\theta(X)$	Halogens	Entropy of drug for halogens at distance $k = 0$
Drug	$D\theta(Het)$	Heteroatoms	Entropy of drug for heteroatoms at distance $k = 0$
Drug	$D\theta(H-Het)$	Hydrogens bonded to heteroatoms	Entropy of drug for hydrogens-heteroatoms and all the neighbors at distance $k \leq 2$
Drug	$D\theta(H-Het)$	Hydrogens bonded to heteroatoms	Entropy of drug for hydrogens-heteroatoms and all the neighbors at distance $k \leq 5$
Molecule	Symbol	Protein region	Descriptor name
Target protein	$T\theta(c)$	Core	Entropy of all aminoacids in the core of the protein at distance $k = 0$
Target protein	$T\theta(c)$	Core	Entropy of all aminoacids in the core of the protein at distance $k \leq 5$
Target protein	$T\theta(i)$	Inner	Entropy of all aminoacids placed in the inner region at distance $k = 0$
Target protein	$T\theta(i)$	Inner	Entropy of all aminoacids placed in the inner region and all the neighbors at distance $k \leq 3$
Target protein	$T\theta(i)$	Inner	Entropy of all aminoacids placed in the inner region and all the neighbors at distance $k \leq 5$
Target protein	$T\theta(m)$	Middle	Entropy of all aminoacids placed in the middle region and all the neighbors at distance $k \leq 1$
Target protein	$T\theta(m)$	Middle	Entropy of all aminoacids placed in the middle region and all the neighbors at distance $k \leq 2$
Target protein	$T\theta(m)$	Middle	Entropy of all aminoacids placed in the middle region and all the neighbors at distance $k \leq 3$
Target protein	$T\theta(m)$	Middle	Entropy of all aminoacids placed in the middle region and all the neighbors at distance $k \leq 5$
Target protein	$T\theta(s)$	Surface	Entropy of all aminoacids placed in the surface region at distance $k = 0$
Target protein	$T\theta(s)$	Surface	Entropy of all aminoacids placed in the surface region and all the neighbors at distance $k \leq 2$

significance of DTPs/nDTP discrimination with  $p < 0.001$  (level of error). Significant entropy parameters were calculated for the totality ( $T$ ) of the atoms in the molecule or for specific collections of atoms. These collections are atoms with a common characteristic as for instance are: halogens ( $X$ ), heteroatom ( $Het$ ), hydrogen atom bonded to one heteroatom ( $H-Het$ ) or protein region (protein region). In Table 1 we report a detailed list of the symbols and description for all parameters present in the model.

This model, with 18 variables, classifies correctly 597 out of 678 DTPs (Sensitivity of 88.05%) and 2878 out of 3234 nDTP (Drug–Protein Pair for compounds with low connectivity) (Specificity of 89%). Overall training Accuracy was 88.83%. The validation of the model was carried out by means of external predicting series. The model classifies correctly 298 out of 338 DTPs (88.17%) and 1359 out of 1534 nDTPs (88.59%) in validation series. Accuracy for validation series (predictability) was 88.51% (1399 out of 1872 DTPs). These results (Table 2) indicate that we developed an accurate model according to previous reports on the use of LDA in QSAR [65,66].

**ANN models.** The present model shows good results with a relatively small number of parameters (18 parameters) and a linear equation. To show how important is this result, we compared the present model with other models used to address the same problem. We processed our data with different Artificial Neural Networks (ANNs) looking for a better model. Four types of ANNs were used, namely, Probabilistic Neural Network (PNN), Radial Basic Function (RBF), Three Layers Perceptron (MLP-3), and Four Layer Perceptron (MLP-4). The Fig. 3 depicts the networks maps for some of the ANN models tested. In general, at least one ANN of every type tested was statically significant. However, one must note that the profiles of each network indicate that these are highly nonlinear and complicated models.

One network found was MLP and it showed training performance higher than 92%. We compare different types of networks to obtain a better model; Table 2 shows the classification matrix of the different networks. Was taken as the main network (MLP 32:32–15–1:1) because it presents a wider range of variables, presents 32 inputs in the first layer and 32 neurons in second layer, two sets of cases (Training and Validation). Another tested networks found were MLP 39:39–20–21–1:1, RBF 1:1–1–1:1 presents the same type of variables; Linear 64:64–1:1 present many variables and PNN 65:65–3912–2–2:1 has a very low percentage of DTPs leading to possible errors in the model although your accuracy is very well, see Table 2. We depict the ROC-curve for MLP 32:32–15–1:1 to show how reliable was the network model developed, see Fig. 4. Notably, almost the model presented was

under curve higher than 0.5. The model presented an area greater than 0.97. The vitality of this type of procedures developing ANN-QSAR models has been demonstrated before; see, for instance, the works of Fernandez and Caballero [67,68]. The same is true about the other kinds of ANNs tested.

**Complex networks assembly.** Two possible applications for the present model are the bio-molecular screening of drug affinity to different proteins and the construction of multi-protein affinity profile networks for drugs. In order to recall the capacity of the mt-QSAR to predict new CNs we selected a database of recently assayed drugs with their respectively Proteins. With these goals in mind, we constructed first a new observed Drug–Protein (DP) DP-CN, obtaining a CN with 855 vertices and 1016 DP (edges) an average distance equal to 6.66. The same as before, we constructed a new predicted DP-CN obtaining an average distance equal to 5.47 and 1298 DP (edges) see Fig. 6. In this, we illustrate visually both observed DP-CN and predicted DP-CN. The numeric labels of the nodes identify the different inputs (DRPs) used in the analysis.

In Table 3 we show the results of calculating the functions of normal and exponential distribution for proteins and drugs observed and predicted presented in dataset. It is seen in that the protein has a difference ( $d$ ) biggest on normal function to the

**Table 2**

Comparison of LDA and different ANNs classification models.

Model profile	Class	Train			Stat. Par.	Validation		
		DTPs	nDTPs	%		%	DTPs	nDTPs
LDA	DTPs	2878	356	89.0	Sn	88.6	1359	175
17:17–1:1	nDTPs	81	597	88.1	Sp	88.2	40	298
	Total			88.8	Ac	88.5		
LNN	DTPs	606	72	89.4	Sn	88.5	299	39
64:64–1:1	nDTPs	359	2875	88.9	Sp	87.4	193	1341
	Total			89	Ac	87.6		
PNN	DTPs	26	652	3.8	Sn	3.6	12	326
65:65–3912–2–2:1	nDTPs	0	3234	100	Sp	100	0	1534
	Total			83.3	Ac	82.6		
MLP	DTPs	602	76	88.8	Sn	88.8	300	38
39:39–20–21–1:1	nDTPs	354	2880	89.1	Sp	88.9	171	1363
	Total			89	Ac	88.8		
RBF	DTPs	369	309	54.4	Sn	52.4	177	161
1:1–1–1:1	nDTPs	1502	1732	53.6	Sp	52.6	727	807
	Total			53.7	Ac	52.6		
MLP	DTPs	623	55	91.9	Sn	92.6	313	25
MLP 32:32–15–1:1	nDTPs	239	2995	92.6	Sp	92.0	123	1411
	Total			92.5	Ac	92.1		

DTPs: Drug-target pairs for compounds with high affinity; nDTPs: Drug-target pair for compounds with non-affinity.

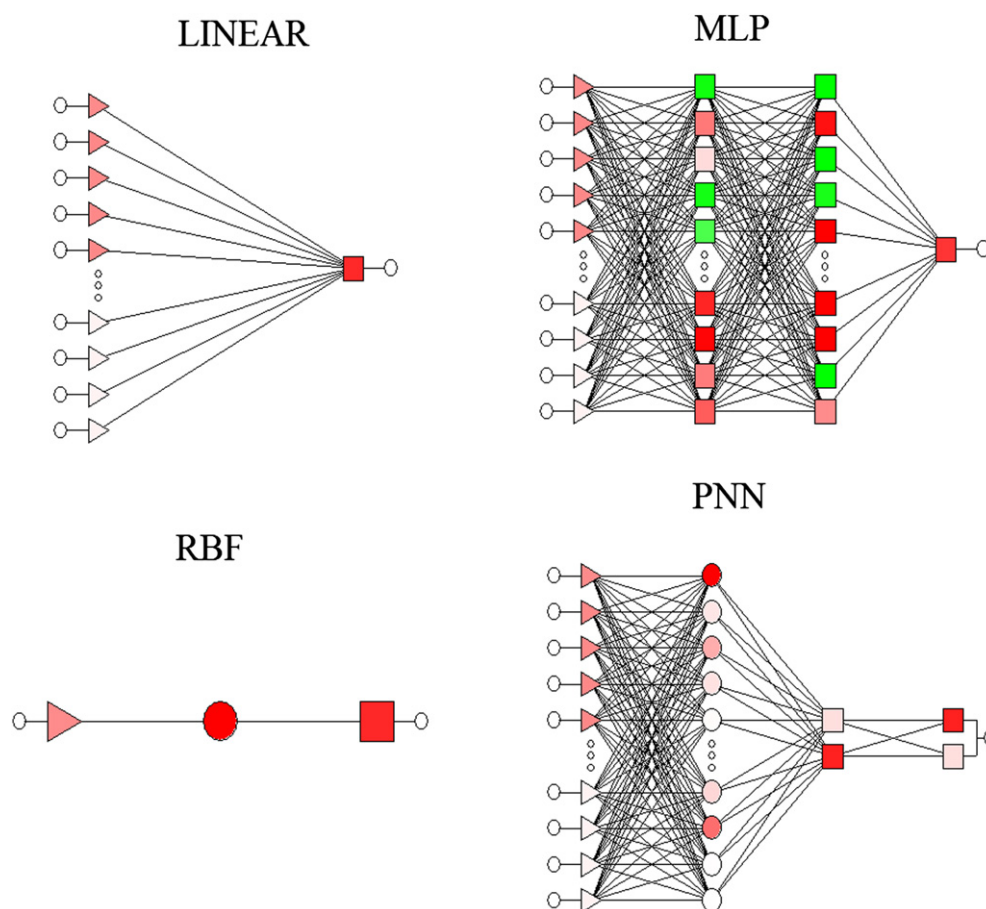


Fig. 3. Topology of some ANN models trained in this work.

exponential function, so that we can conclude that our CN tends to be an exponential distribution. To illustrate these results a node degree distribution for both observed and predicted DPs were performed, see Fig. 7. In this figure we can see that the drugs follow

a normal distribution, while proteins follow an exponential distribution.

In Table 4 we show the results of closeness centrality and the number of node degree for proteins and drugs used in the database.

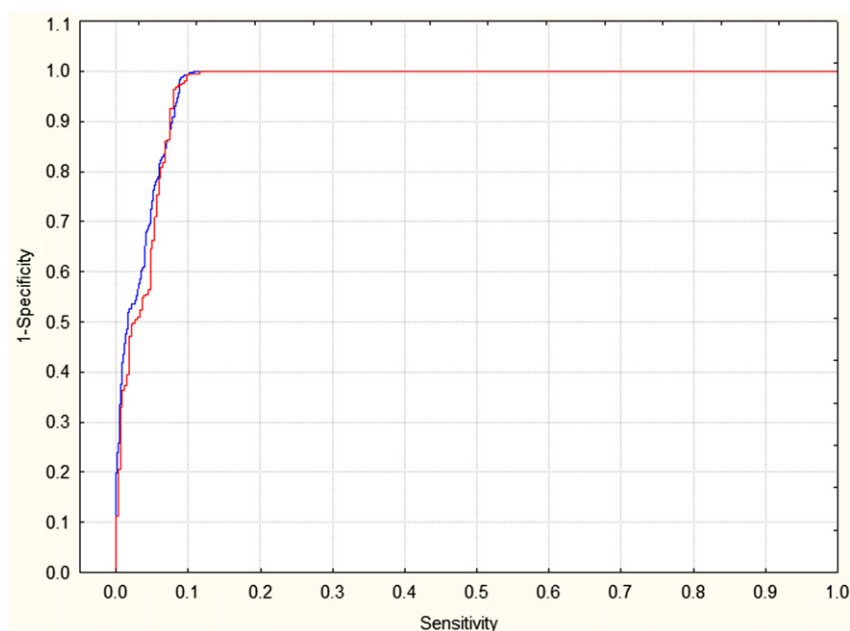
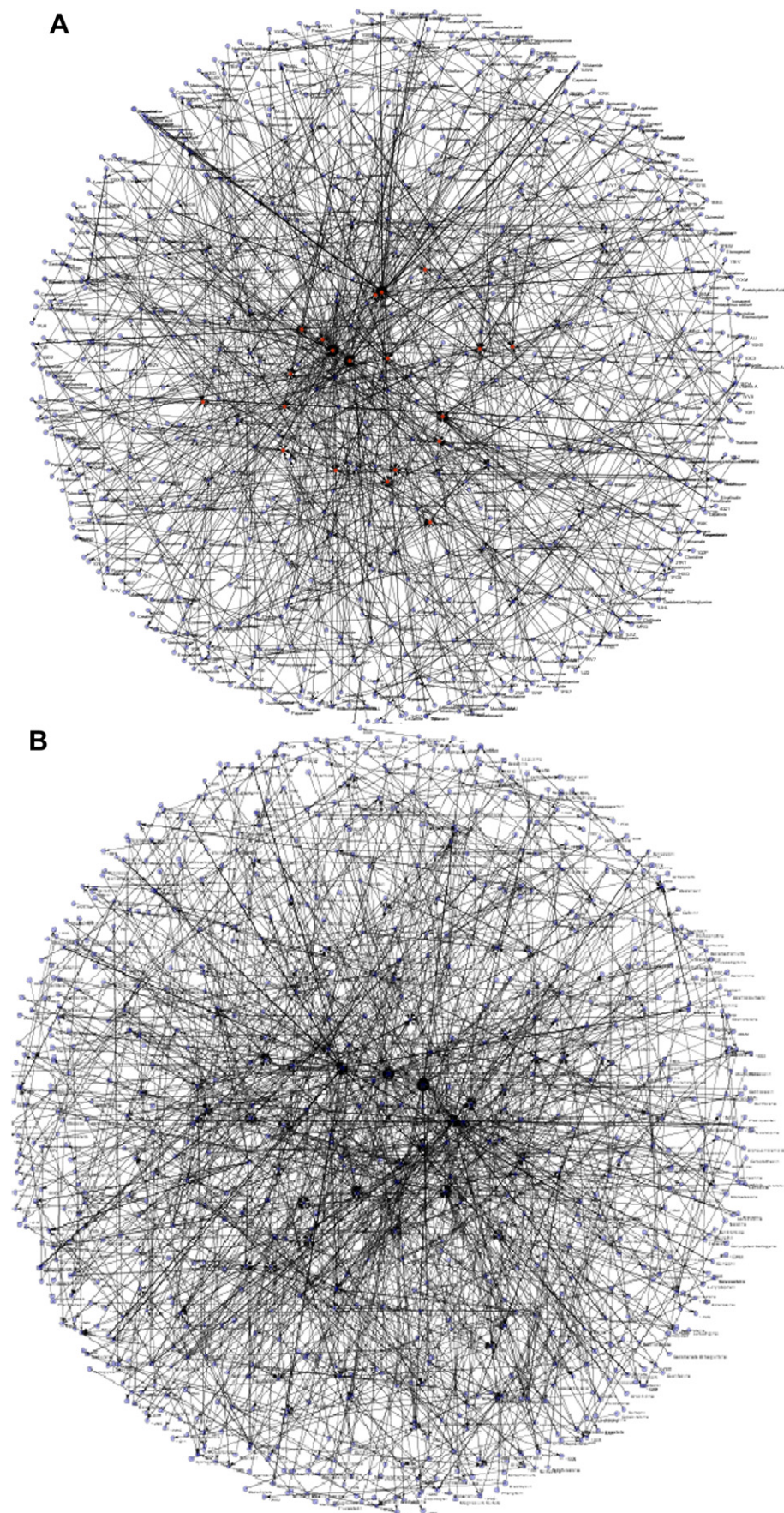


Fig. 4. ROC curve for classifier.







**Fig. 6.** Observed vs. predicted drug-target complex networks.

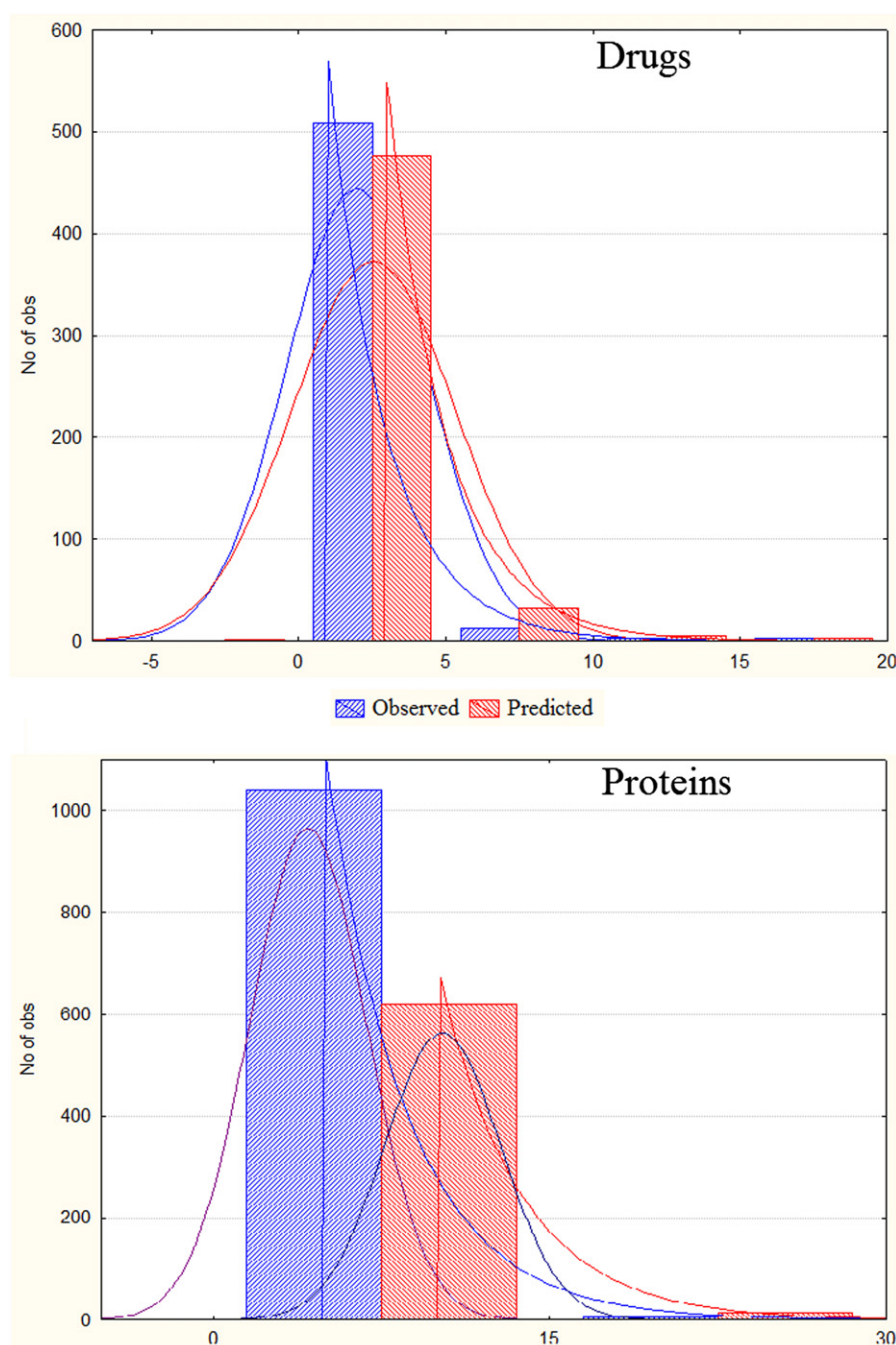


**Table 3**  
Study of network distribution functions.

Sub-network	Network	Distribution	<i>d</i>	<i>p</i>	Fit	Best fit
Proteins	Observed	Normal	0.46	0.01	No	No
		Exponential	0.30	0.01	No	Yes
	Predicted	Normal	0.39	0.01	No	No
		Exponential	0.27	0.01	No	Yes
Drugs	Observed	Normal	0.35	0.01	No	Yes
		Exponential	0.40	0.01	No	No
	Predicted	Normal	0.34	0.01	No	Yes
		Exponential	0.40	0.01	No	No

against illness before descript. On the contrary this protein 1A8M closeness to present a very low node degree with a low closeness, which means that drugs that interact with 1A8M interacts with other proteins and which mean the drugs that interact with 1A8M are not selective, in this case there are a variety of drugs that can interact with protein see [Table 4](#).

In the case of drugs as an example is the Moexipril. This drug is a ACE inhibitor that can be for the treatment of hypertension [72]. Moexipril presents a low closeness equal to 29 with a high node degree 37, this means that the drug is selective for proteins to which is connected to the CN. These results are consistent with the literature [72–74]; Moexipril is a long-acting, non-sulphydryl angiotensine-converting enzyme inhibitor (ACE). On the other



**Fig. 7.** Distribution functions for complex networks sub-sets.

hand Acetazolamide, that is a carbonic anhydrase inhibitor that is used to treat glaucoma, epileptic seizures, benign intracranial hypertension (pseudotumor cerebri), altitude sickness, cystinuria, and dural ectasia [75,76], acetazolamide submit a node degree equal = 5 with a closeness = 22, which indicates that this drug interacts with five proteins these proteins only interact with Acetazolamide; there are specific to this drug.

### 3.2. Illustrative experiments

#### 3.2.1. Study of rasagiline analogs (experiment 1)

**Synthesis, characterization, and assay.** The preparation of (±)-*cis*- and *trans*-amino alcohols **5a,5b** and their *N*-propargyl derivatives **6a,6b** and *N,N*-dipropargyl derivatives **7a,7b** has been reported see Fig. 5. In short, the amino alcohols **5a,5b** were prepared from the 2,2,2-trifluoro-*N*-(6-oxo-5,6-dihydro-4*H*-cyclopenta[*b*]thiophen-4-yl)acetamide, **3** that were synthesized by intramolecular cyclization of 3-amino-3-(thiophen-3-yl)propanoic acid, **2** following previously described pathways [77]; thus the β-amino acid **2** were obtained by treatment of thiophen-3-carbaldehyde with ammonium acetate and malonic acid in refluxing ethanol [78]. The treatment of the carboxylic acid **2** with a mixture of boiling trifluoroacetic acid and anhydride gave in one step the oxotri-fluoroacetamide derivative **3**, whose melting point and spectrophotometry infrared has concordat with the literature [77,79]. The reaction of the oxoamide **3** with NaBH<sub>4</sub> in MeOH during 4 days at room temperature (molar ratio **3**/NaBH<sub>4</sub>, 1/5) [80,81] produced, after chromatographic purification, an 86% yield of a mixture of *cis/trans*-**5a/5b** epimers, however attempts of separation of the mixture were unsuccessful, leading only to partial resolution of epimers. Alternatively, the compound **3** were converted to the mixture of *cis/trans*-**4a/4b** epimers (83% yield) by treatment with NaBH<sub>4</sub> in methanol during 10 min at room temperature (molar ratio **3**/NaBH<sub>4</sub>, 1/2) and the subsequent resolution of those by flash column chromatography was efficient (**4a**:32%, **4b**:31%), so the needed separation of the corresponding isomers was performed at this stage of the synthetic route. The trifluoroacetamides hydrolysis of these compounds with K<sub>2</sub>CO<sub>3</sub> in MeOH/H<sub>2</sub>O gave the corresponding amino alcohols **5a** and **5b** with yields of the 99% and 75% respectively. Propargylation of these compounds with propargyl bromide and K<sub>2</sub>CO<sub>3</sub> in acetonitrile [82] afforded the corresponding mixtures *cis* and *trans* of mono- and dipropargyl derivatives **6a/7a** and **6b/7b**. Subsequent

chromatography achieving clean separation of these mixtures: **6a** (24%) and **7a** (36%), and **6b** (13%) and **7b** (753%). Finally the compounds **7a** and **7b** were converted in goods yields, to the corresponding acetyl esters **8a** and **8b**, and benzoyl esters **9a** and **9b**, treatment with acetic anhydride in Et<sub>3</sub>N [83] and with benzoyl chloride in Et<sub>3</sub>N and acetonitrile [84] respectively.

**Prediction of rasagiline analogs vs. MAO proteins.** In this *in silico* experiment we used the 3D structure MAO-A and B proteins with PDB ID 2Z5X and 2VZ2 and the model MLP 32:32–15–1:1 to predict the scores. We also generated the SMILE codes for these compounds and predicted their propensity to form DTPs with MAO-A and MAO-B using the model. In Table 5 we confront the results obtained using this model and the outcomes of the pharmacological assay. The compound Rasagiline a known selective inhibitor for MAO-B was used as control. We consider the observed class for active compounds OC = 1 if compound IC<sub>50</sub> < 50 μM this cut-off is in the similar range than other used in previous works [36,37]. As we can see in this table the only one active compound in the pharmacological assay (OC = 1) was compound **8b** predicted also as active with PC = 1 and high score S(DTP)<sub>pred</sub> = 0.64. Two other compounds (**5b** and **6b**) that are inactive (both as MAO A and MAO B inhibitors) in pharmacological assays (OC = 0) were also predicted as inactive against MAO A (PC = 0) but predicted as active for MAO B (PC = 1). In this case the model fails, however the prediction have a very low activity scores S(DTP)<sub>pred</sub> = 0.15.

**Prediction of rasagiline analogs vs. US FDA drug-target proteins.** An additional use of the model was to predict the activity of the new compounds with respect to all other targets previously approved by US FDA [85]. At the same time we can use this model to predict the selectivity of the new rasagiline derivative **8a** as MAO B inhibitor with respect to all FDA drugs targets and predict possible toxicological effects depending on the other targets predicted for these compounds. This type of experiment is of the major importance due to the cost in terms of animal sacrifice, time, materials and human resources of the experimental assay of all compounds against all these targets, see recent reviews by Duardo-Sanchez et al. [86–89]. In fact, over a decade, the US FDA has been engaged in the applied research, development, and evaluation of computational toxicology methods used to support the safety evaluation of a diverse set of regulated products. The basis for evaluating computational toxicology methods is multi-factorial, including the potential for increased efficiency, reduction in the numbers of animals used, lower costs, and the need to explore emerging

**Table 4**  
Results of node degree and closeness centrality for top-20 proteins and drugs.

Rank	PDB	Function	C <sub>deg</sub>	C <sub>clo</sub>	Drugs	Activity	C <sub>deg</sub>	C <sub>clo</sub>
1	1HOF	α-2A adrenergic receptor	51	20.3	Moexipril	ACE inhibitor	37	29.4
2	1BNA	DNA	49	33.49	Sertraline	SSRI class	18	34.5
3	1HA2	Serum albumin	47	35.29	Levamisole	Anthelmintic	16	27.1
4	1QYX	Estradiol 17-β-dehydrogenase 1	31	27.89	Adenine	Biochemistry metabolism	15	27.6
5	1UWJ	B-Raf proto-oncogene	22	26.64	Digoxin	Atrial fibrillation	15	32.3
6	1EMI	23S rRNA	20	23.67	Atorvastatin	Inhibitor of HMG-coa reductase	11	31.6
7	1CZM	Carbonic anhydrase 1	19	26.77	Pyrazinamide	FAS inhibitor	10	24.9
8	1MLD	Malate dehydrogenase	18	31.47	ADP	AMP/IMP catalysis	8	27.9
9	1P49	Steryl-sulfatase	16	26.98	Mibefradil	Ca <sup>2+</sup> channel blocker	8	27.3
10	1BYW	K <sup>+</sup> voltage-gated H member 2	15	27.62	Alitretinoin	Antineoplastic	7	21.9
11	1T40	Aldose reductase	15	25.2	L-Aspartic acid	Stimulates NMDA receptors	7	23.9
12	1F8U	Acetylcholinesterase	14	23.33	Megestrol	Antineoplastic	7	25.4
13	1N7D	Low-density lipoprotein receptor	14	22.7	Vinblastine	Antineoplastic	7	21.8
14	1SLM	Stromelysin-1	13	21.1	Acitretin	Psoriasis inhibitor	6	29.5
15	1EXX	γ-1 Retinoic acid receptor	12	25.5	Etorphine	Analgesic	6	21.9
16	1E3G	Androgen receptor	11	27.3	Halofantrine	Antimalaria	6	31.2
17	1VJB	Estrogen-related receptor γ	11	23.4	L-Arginine	Vasodilatation	6	20.3
18	1Z8L	Glutamate carboxypeptidase 2	11	24.7	L-Ornithine	Catalyst	6	25.3
19	1A8M	Tumor necrosis factor	10	30.9	Acetazolamide	Carbonic anhydrase inhibitor	5	22.8
20	1OCZ	Cytochrome c oxidase subunit 1	10	26	Adapalene	Anti-acne	5	21.9

**Table 5**

Prediction of rasagiline analogues with the new model.

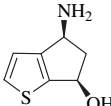
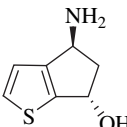
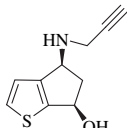
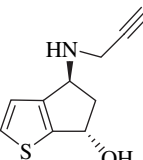
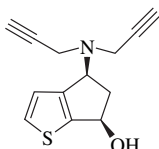
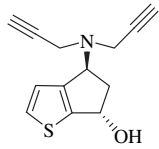
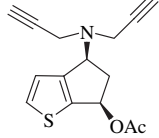
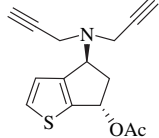
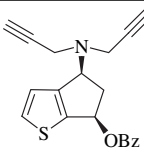
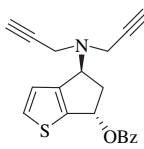
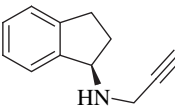
Drug <sup>a</sup>	MAO-B				Compound Structure	MAO-A			
	IC <sub>50</sub> <sup>b</sup>	OC	PC	S(DTP) <sub>pred</sub>		IC <sub>50</sub>	OC	PC	S(DTP) <sub>pred</sub>
<b>5a</b>	>100	0	0	0.70		>100	0	0	0.97
<b>5b</b>	>100	0	1	0.15		>100	0	0	0.88
<b>6a</b>	>100	0	0	0.70		>100	0	0	0.97
<b>6b</b>	>100	0	1	0.15		>100	0	0	0.88
<b>7a</b>	>100	0	0	0.94		>100	0	0	0.99
<b>7b</b>	>100	0	0	0.94		>100	0	0	0.99
<b>8a</b>	>100	0	0	0.92		>100	0	0	0.98
<b>8b</b>	46.25	1	1	0.64		>100	0	0	0.96

Table 5 (continued).

Drug <sup>a</sup>	MAO-B				Compound Structure	MAO-A			
	IC <sub>50</sub> <sup>b</sup>	OC	PC	S(DTP) <sub>pred</sub>		IC <sub>50</sub>	OC	PC	S(DTP) <sub>pred</sub>
<b>9a</b>	>100	0	0	0.94		>100	0	0	0.99
<b>9b</b>	>100	0	0	0.92		>100	0	0	0.98
Rasagiline	0.412 ± 0.04	1	1	0.14		0.0443 ± 0.009	1	0	0.87

<sup>a</sup> Rasagiline was used as positive control.

<sup>b</sup> >100 = compound inactive at 100 μM (highest concentration tested), OC = observed class and PC = Predicted class, OC = 1 if compound IC<sub>50</sub> < 50 μM and PC = 1 if the DTP probability predicted for pair drug-MAO-i enzyme  $p(\text{MAO-i}) > 0.5$  (2Z5X and 2VZ2 are PDB ID of MAO-A and B used to predict *p*-values).

technologies that support the goals of the US FDA's Critical Path Initiative (e.g. to make decision support information available early in the drug review process). The US FDA's efforts have been facilitated by agency-approved data-sharing agreements between government and commercial software developers [90].

For this we used the model to calculate DTPs scores for our compounds (rasagiline derivatives) vs. FDA approved targets. We depict in Table 4SM, all proteins in FDA dataset predicted vs. the 10 rasagiline derivatives. We found that overall the 10 derivatives were predicted negative against almost all proteins in the FDA

database. In Table 6 we depict some results, which show that our rasagiline derivative **8b** is predicted as selective MAO B inhibitor, because it have only interaction with MAO-B, and have not predicted interaction with the proteins used in the database.

Using these results (depicted in Table 4SM), we constructed a DP-CN for rasagiline derivatives and the FDA dataset (see Fig. 8). As a result we obtained a CN with 669 nodes (FDA drugs, proteins, or rasagiline derivatives) and 839 DP (edges, DTPs). As In this network we can see that protein 2BK3 (MAO-B) is predicted to interact with compounds **5b**, **6b** and **8b**, this protein is a known

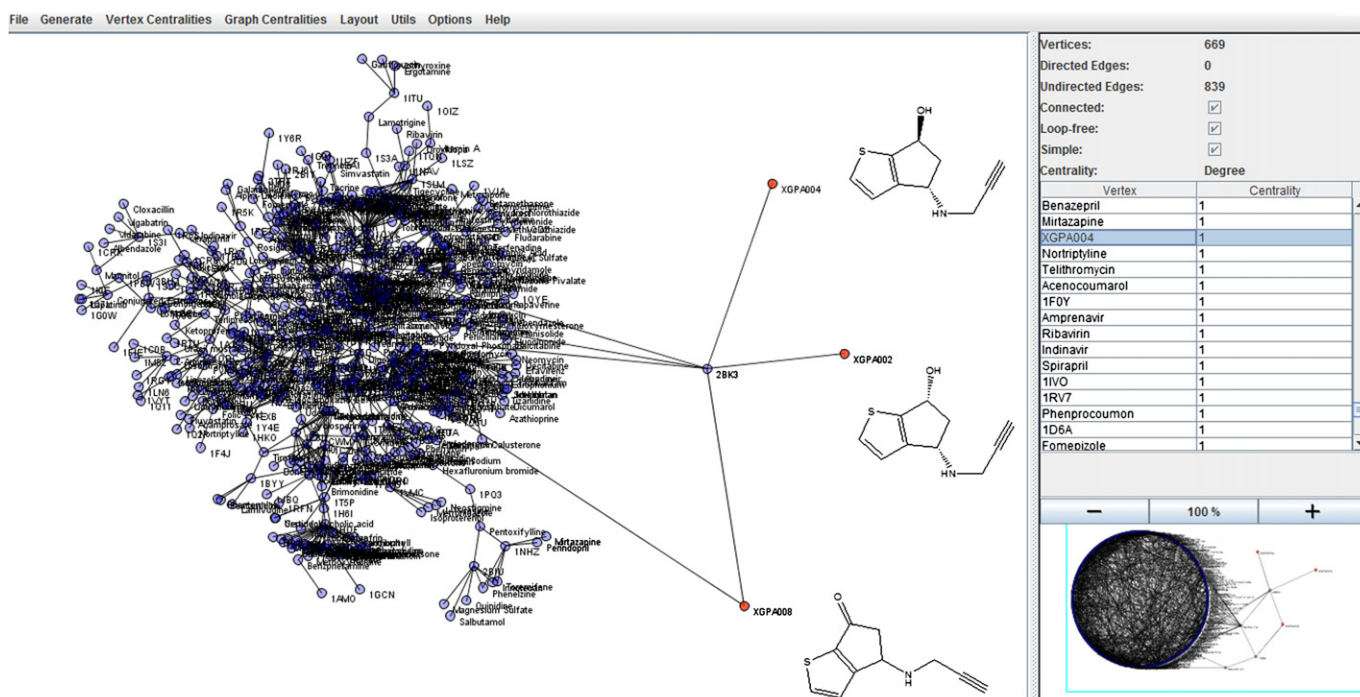
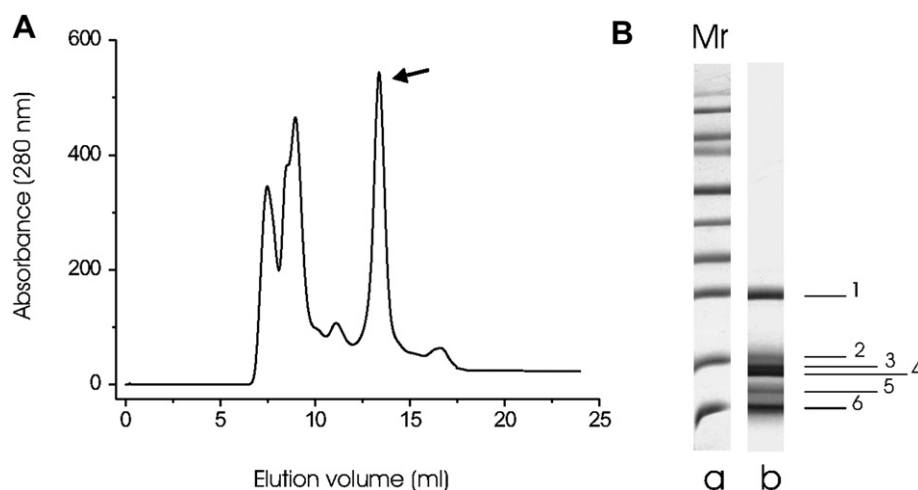


Fig. 8. Drug-target sub-network obtained after coupling the real network for FDA drugs and proteins with new rasagiline derivatives (based on QSAR prediction with MLP model).





**Fig. 9.** ESAs of *F. hepatica*: (A) Size-exclusion chromatogram obtained by FPLC separation of ESAs from *F. hepatica*. The arrow indicates peak number IV of the chromatogram according to Mezo et al. [62]; (B) SDS-PAGE analysis of proteins contained in peak number IV (Lane b). The proteins present in band 2 were analyzed by MS and MS/MS spectrometry. The relative molecular masses of standard markers (lane a) from top to bottom were: 170, 130, 95, 72, 55, 43, 34, 26, 17 and 10 kDa.

rasagiline target [94,95]. These results are satisfactory because they agree with the experimental results presented in this paper where the compound **8b** show MAO-B activity. The use of such complex networks can help us find and predict new drugs–protein interactions, and therefore find new drugs with improved biological activity and fewer side effects, especially in parasite disease.

### 3.2.2. Study of peptidome for *Fasciola* hemoglobin protein (experiment 2)

SEC, 1DE, PMF and MS/MS study of peptides found on *Fasciola* hemoglobin (fHb). In this section we present an example of the practical use of the QSAR model to predict enzyme scores for peptides found in the PMF and MS/MS study of a new query protein. We illustrate an overall view of SEC and 1D electrophoresis study of *F. hepatica* proteome carried out in this work, see Fig. 9. In this figure, we label the bands obtained after 1DE including band number 2.

**Table 6**  
Some illustrative scores obtained in experiments 1 and 2.

Prediction of rasagiline analogues (experiment 1)				Targets <sup>a</sup>
Drugs codes	PDB	PC	Score	
8b	2VZ2	1	0.64	MAO-B MTA phosphorylase
8b	1SD2	1	0.74	
5a	1E18	0	1.00	fHb <sup>b</sup>
5b	1E18	0	0.93	
9a	1A8M	0	1.00	fHb <sup>b</sup>
9b	1A8M	0	1.00	
5a	1AGS	0	1.00	fHb <sup>b</sup>
9b	1BWC	0	1.00	
5a	1BXS	0	1.00	fHb <sup>b</sup>
7b	1D3H	0	1.00	
Prediction of peptides (experiment 2)				fHb <sup>b</sup>
Drugs name	Peptide	PC	Score	
Amphetamine	p1	1	0.82	Yes
Apratinin	p6	1	0.81	Yes
Cisplatin	p16	1	0.78	Yes
Debrisoquin	p14	1	0.57	
Hexachlorophene	p5	1	0.83	
Pentamidine	p22	1	0.58	
Phentermine	p29	1	0.89	
Propofol	p20	1	0.58	
Rasagiline	p5	1	0.55	

<sup>a</sup> We give only the function of positive targets.

<sup>b</sup> fHb = yes means that peptide may be present in hemoglobin of *F. hepatica* (fHb).

The selected band (band number 2 in this study) was excised and the peptides were analyzed by MALDI-TOF MS and MS/MS. Once we obtained the data from MALDI-TOF MS analysis of the query band, the most relevant MS signals were introduced into the MASCOT search engine [91,92]. We obtained 20 hits (template proteins) for this protein with MASCOT scores ( $M_s$ ) higher than 81 ( $p < 0.05$ ), the threshold value for significant match, see Table 7. The maximum scores obtained were 286 and 283, which corresponded to proteins gi|196049684 (16 550 Da) and gi|159461074 (16 681 Da) of *F. hepatica*, both are fHb proteins annotated as: *Fasciola* Chain A hemoglobin (Hb2) and *Fasciola* hemoglobin F2, respectively.

We provide in Table 8, detailed information on the results of the MS and MASCOT search engine for band number 2. This table includes the 9 most interesting peptides matching with the fHb sequence. We found an excellent match between the mass of the peptide detected and the mass of the template peptide recorded in MASCOT database with known sequence for all these peptides. After that, we decided to investigate the structure–function relationships for all sequences of the 9 fHb peptides found on the PMF

**Table 7**  
Top-20 template proteins in *F. hepatica* found by MASCOT search.

Protein	Accession	Score	Function
1	gi 196049684	16,550	286 Chain A, hemoglobin (Hb2) <i>Fasciola hepatica</i>
2	gi 159461074	16,681	283 Hemoglobin F2 <i>Fasciola hepatica</i>
3	gi 47116941	14,671	79 Fatty acid-binding protein type 3
4	gi 209964147	20742	69 D,D-heptose 1,7-bisphosphate phosphatase 1
5	gi 73539911	69,948	68 Thiamine biosynthesis protein thic
6	gi 53713210	34,431	64 Putative transcription regulator
7	gi 162149431	56,150	64 Hypothetical protein GDI3669
8	gi 156844632	52,531	62 Hypothetical protein Kpol-1058p57
9	gi 51245092	32,033	61 Cysteine synthase A
10	gi 188590974	69,946	60 Thiamin biosynthesis protein
11	gi 164428423	274,657	60 Hypothetical protein NCU00551
12	gi 58268372	88,612	60 Myosin heavy chain
13	gi 114800067	52,676	60 Hypothetical protein HNE_2630
14	gi 37522203	19,619	59 Hypothetical protein gll2634
15	gi 46203356	41,788	59 COG0673 dehydrogenases
16	gi 3913328	59,940	59 Cytochrome P450 52A10
17	gi 146278964	46,771	59 Peptidase M16 domain-containing protein
18	gi 6679022	38,881	59 Neutrophil cytosolic factor 4
19	gi 189184533	29,079	58 Hypothetical protein OTT_1626
20	gi 148284486	29,137	58 Competence protein F

**Table 8**MASCOT study of hemoglobin peptidome in *Fasciola hepatica*.

Pept	Sequences	fHb <sup>a</sup>	Observed	Mr(expt)	Mr(calc)	Delta
p1	kaasnpvsleeri	Yes	1172.59	1171.58	1171.58	−0.002
p2	kaasnpvsleerivqgakd	Yes	1768.80	1767.79	1767.95	−0.1577
p3	karpvtkdqftgaapifkf	Yes	1960.03	1959.02	1959.09	−0.0745
p4	kavnnyhkv		944.58	943.58	943.49	0.0883
p5	kcpentthvvre		1212.56	1211.56	1211.57	−0.0154
p6	kdnvgqsegiry	Yes	1074.56	1073.55	1073.51	0.0414
p7	kdqftgaapifkf	Yes	1307.68	1306.68	1306.69	−0.0168
p8	kdsdskisqvqkc		1234.56	1233.56	1233.62	−0.0631
p9	kffqglkkq	Yes	980.60	979.60	979.59	0.0119
p10	kflhvmqiaaakm	Yes	1504.70	1503.70	1503.79	−0.0967
p11	kiaahaadlakg		980.60	979.60	979.55	0.0523
p12	kiahfscmcgpkf		1307.65	1306.64	1306.56	0.0778
p13	klegsenmdavlqkl		1504.70	1503.70	1503.72	−0.0264
p14	klitsskpeiftlegnkm		1877.92	1876.91	1877.01	−0.0996
p15	klldhgyfvfvtnqsgvarg		2236.24	2235.23	2235.14	0.0853
p16	klqgltkdnvgqsegiry	Yes	1714.81	1713.80	1713.90	−0.099
p17	kmiatvtvgdvka	Yes	1133.61	1132.60	1132.62	−0.0113
p18	kmiatvtvgdvkavnnyhkv		1960.03	1959.02	1959.02	−0.0051
p19	ktlfaahpeyisyfsl		1773.79	1772.78	1772.88	−0.0987
p20	ktvisftfgeefkeetadgrt		2264.07	2263.06	2263.06	−0.0009
p21	kwclahhke		951.55	950.54	950.45	0.0886
p22	rcrkpepgmldldcdw		1859.90	1858.90	1858.89	0.0125
p23	rdawrgaafldrd		1277.61	1276.61	1276.63	−0.0246
p24	kwclahhke		1114.61	1113.61	1113.58	0.0278
p25	rdlqaaeaagirg		1114.61	1113.605	1113.57	0.0278
p26	rgaafldrdgvlndhgyvhrr		2225.17	2224.16	2224.11	0.0456
p27	rghlftggdlsefvqgallr		1960.03	1959.02	1959.02	−0.0017
p28	rklitsskpeiftlegnkm		2006.10	2005.10	2005.11	−0.0135
p29	rkpepgmldldcdwvpvdr		2196.99	2195.99	2196.08	−0.0965
p30	rredvewiqgaitavkl		1714.81	1713.80	1713.90	−0.103

<sup>a</sup> fHb = yes means that peptide may be present in hemoglobin of *F. hepatica* (fHb).

of the new protein and other peptides reported by MASCOT as well (30 peptides in total).

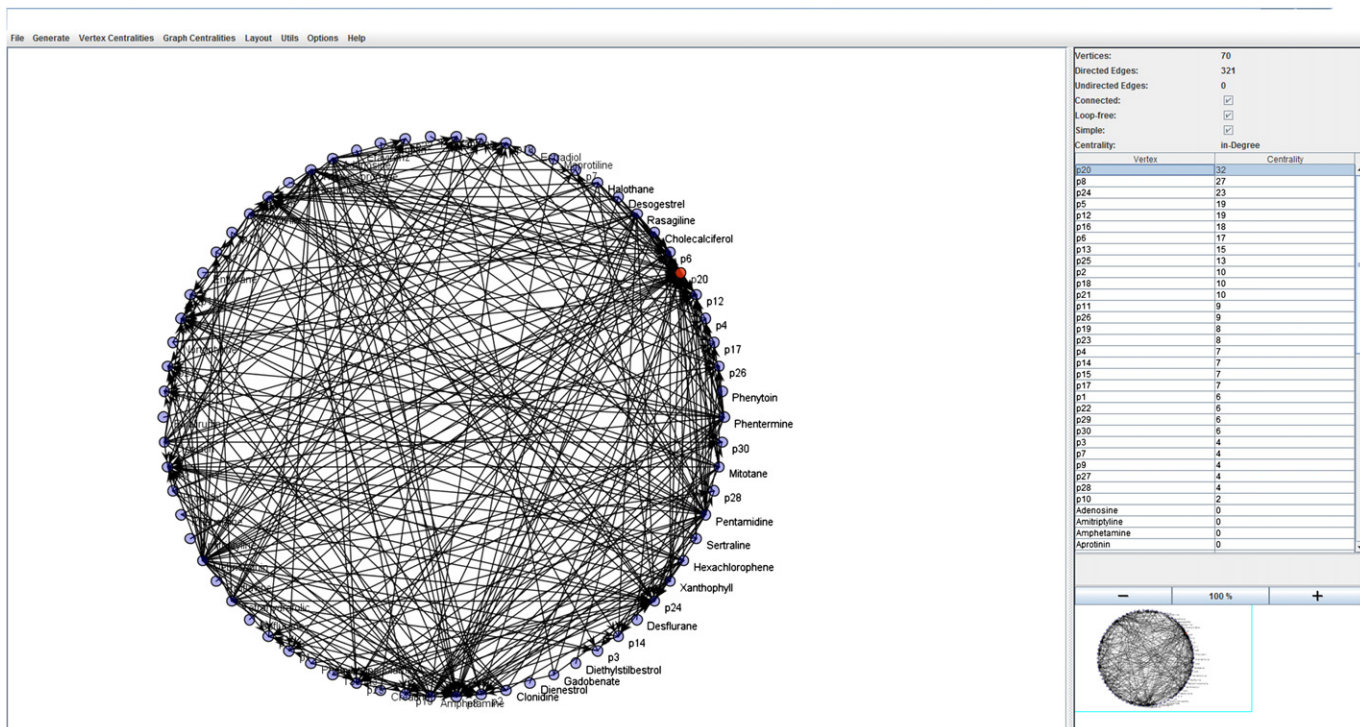
**MM/MC study of peptides found on new protein PMF.** Our main interest on the study of the peptides in the PMF of the new unknown proteins is to find which of them make a positive

**Table 9**MM/MD study of hemoglobin peptidome in *Fasciola hepatica*.

Input			MM/MD		
Pept.	sequences	fHb <sup>a</sup>	ACCR	EPOT	DEPOT
p1	kaasnpvsleeri	Yes	0.62	−224.29	26.91
p2	kaasnpvsleerivqgakd	Yes	0.60	−237.99	35.00
p3	karpvtkdqftgaapifkf	Yes	0.47	288.65	61.95
p4	kavnnyhkv		0.47	24.38	34.27
p5	kcpentthvvre		0.67	18.15	24.15
p6	kdnvgqsegiry	Yes	0.65	−26.93	21.92
p7	kdqftgaapifkf	Yes	0.63	55.90	27.43
p8	kdsdskisqvqkc		0.61	−152.86	36.88
p9	kffqglkkq	Yes	0.63	56.88	34.29
p10	kflhvmqiaaakm	Yes	0.62	−118.13	25.75
p11	kiaahaadlakg		0.65	−34.73	15.87
p12	kiahfscmcgpkf		0.61	−139.91	23.68
p13	klegsenmdavlqkl		0.62	−130.95	25.73
p14	klitsskpeiftlegnkm		0.64	−5.93	33.40
p15	klldhgyfvfvtnqsgvarg		0.64	−69.69	30.24
p16	klqgltkdnvgqsegiry	Yes	0.60	−203.86	26.48
p17	kmiatvtvgdvka	Yes	0.63	49.97	19.80
p18	kmiatvtvgdvkavnnyhkv		0.64	86.49	24.13
p19	ktlfaahpeyisyfsl		0.66	−66.69	25.78
p20	ktvisftfgeefkeetadgrt		0.64	19.60	35.22
p21	kwclahhke		0.62	−145.51	38.69
p22	rcrkpepgmldldcdw		0.65	−55.35	19.04
p23	rdawrgaafldrd		0.63	−38.58	35.68
p24	kwclahhke		0.64	−133.69	37.46
p25	rdlqaaeaagirg		0.62	−209.57	33.44
p26	rgaafldrdgvlndhgyvhrr		0.63	76.47	18.94
p27	rghlftggdlsefvqgallr		0.60	−208.79	28.09
p28	rklitsskpeiftlegnkm		0.64	−168.13	34.40
p29	rkpepgmldldcdwvpvdr		0.59	−362.27	25.37
p30	rredvewiqgaitavkl		0.60	−253.95	22.92

<sup>a</sup> fHb = yes means that peptide may be present in hemoglobin of *F. hepatica* (fHb).

contribution to the ligand interaction. This may allow us to select peptides for drug design and/or obtain information for drug-target discovery. We therefore have to calculate the  $\theta_k$  for all peptides and substitute these values in the QSAR model to predict ligand

**Fig. 10.** Drug-target sub-network for FDA drugs vs. 30 peptides of different proteins including 9 peptides of fHb.

interaction score for one ligand (levulinic acid). For this, we first need the 3D structures of the peptides in order to calculate the  $\theta_k$  values. For this study we used the same 30 peptides found by PMF of the new protein (9 peptides from fHb and the rest from other proteins). Unfortunately, we only have the sequences of the peptides but not the 3D structures. We therefore first obtained the optimal 3D folded structures by use of an MM geometry optimization for the 30 peptides (see Fig. 2). We complemented the MM by MC search in order to explore alternative geometrical structures for the peptides. We summarized the results of MC simulation of these peptides in Table 9. In this table we reported the initial energy ( $E_0$ ) based on the starting structure constructed with standard parameters for  $\alpha$ -helices (bond distances, angles, and dihedral angles) set as default on the sequence editor of Hyperchem [59,60]. We also reported the ( $E_1$ ) obtained after optimization of the structure with AMBER force field obtained by MC method using 1000 steps for 30 peptides. Lastly, we report the ACCR values for the MDT of the 30 peptides in Table 9. In the MD study most researchers tend to try for an average ACCR value around 0.5; smaller values may be appropriate when longer runs are acceptable and more extensive sampling is necessary. In the present study all the ACCR values were between 0.47 and 0.66 because MC simulation has been realized by 1000 steps; in consequence, we can accept the MD results as valid [59,60].

**Assemble of drug-target binding site network.** We use our best model to predict whether the FDA drugs used in the database have interaction or affinity with the 30 peptides, but we put special interest on the 9 peptides from fHb. The reason for this interest in fHb peptides is due to the work of Dewilde et al. [93]; which described for the first time fHb, a potential immunogen, in the search for an effective vaccine. Here mt-QSAR models in combination with complex networks may be helpful to predict possible drugs that interact with fHb effectively with fewer adverse reactions. In this sense, we assembled drug-target binding site network with the predicted probabilities of binding of organic compounds to 30 peptides of different proteins found with MASCOT (including 9 peptides of fHb). In so doing, we used the MARCH-INSIDE 2.0 software to calculate the  $\theta_k$  values for the 30 peptides. We substituted these values in the model and predicted the probability with the 30 peptides interact with FDA drugs. In bottom of Table 6 (experiment 2), we show some results used to assemble the drug-target binding site network. We constructed a new observed Drug-Protein DP-CN, obtaining a CN with 70 vertices (peptides or drugs) and 321 DP (edges, peptide-drug pairs) see Fig. 10.

#### 4. Conclusions

Combining entropy parameters, calculated with MARCH-INSIDE approach to codify information of drug and target structure, and ANNs is possible to seek one mt-QSAR classifier to predict the probability of drugs to bind more than 500 different drug-target proteins approved by FDA of USA with Accuracy >90%. This ANN classifier is useful to carry out Data Mining of PDB in order to discovery new drug targets for any drug. We can also predict the interactions with specific drugs of peptides or protein 3D structures generated by MM/MD modeling.

#### Acknowledgments

The authors appreciate the technical assistance and advice of Lola Gutiérrez, from the Genomics and Proteomics Center, Universidad Complutense de Madrid. H. González-Díaz thanks sponsorships for a research position at the University of Santiago de Compostela from the *Isidro Parga Pondal* Program, Xunta de Galicia. F. Prado-Prado thanks sponsorships for research position at the

University of Santiago de Compostela from Angeles Alvariño, Xunta de Galicia. N. Alonso thanks sponsorships for research position at the University of Santiago de Compostela from *FPU program*, Xunta de Galicia. The present study was partially supported by grants FAU2006-00021-C03-00 and AGL2010-22290-C03-01 (Ministerio de Ciencia e Innovación, Spain), 07CSA008203PR (Xunta de Galicia, Spain) and by the European Fund for Regional Development (FEDER). The authors also would like to dedicate the present work to Prof. F. Orallo (*in memoriam*) by his kind attention, support, and friendship.

#### Appendix. Supplementary material

Supplementary data related to this article can be found online at doi:10.1016/j.ejmech.2011.01.023.

#### References

- [1] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug-target interaction networks from the integration of chemical and genomic spaces, *Bioinformatics* 24 (2008) i232–240.
- [2] A. Giuliani, Collective motions and specific effectors: a statistical mechanics perspective on biological regulation, *BMC Genomics* 11 (Suppl. 1) (2010) S2.
- [3] P.K. Dhar, A. Giuliani, Laws of biology: why so few? *Syst. Biol. Synth.* 4 (2010) 7–13.
- [4] S. Bornholdt, H.G. Schuster, *Handbook of Graphs and Complex Networks: From the Genome to the Internet*. WILEY-VCH GmbH & CO. KGa., Weinheim, 2003.
- [5] E. Estrada, Virtual identification of essential proteins within the protein interaction network of yeast, *Proteomics* 6 (2006) 35–40.
- [6] E. Estrada, Protein bipartivity and essentiality in the yeast protein-protein interaction network, *J. Proteome Res.* 5 (2006) 2177–2184.
- [7] A. Réka, A.-L. Barabasi, Statistical mechanics of complex networks, *Rev. Mod. Phys.* 74 (2002) 47–97.
- [8] A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.* 5 (2004) 101–113.
- [9] A.L. Barabasi, Network medicine—from obesity to the “diseasome”, *N. Engl. J. Med.* 357 (2007) 404–407.
- [10] H. González-Díaz, S. Vilar, L. Santana, E. Uriarte, Medicinal chemistry and bioinformatics – current trends in drugs discovery with networks topological indices, *Curr. Top. Med. Chem.* 7 (2007) 1025–1039.
- [11] A. Giuliani, L. Di Paola, R. Setola, Proteins as networks: a mesoscopic approach using haemoglobin molecule as case study, *Curr. Proteomics* 6 (2009) 235–245.
- [12] A. Krishnan, J.P. Zbilut, M. Tomita, A. Giuliani, Proteins as networks: usefulness of graph theory in protein science, *Curr. Protein Pept. Sci.* 9 (2008) 28–38.
- [13] A. Krishnan, A. Giuliani, J.P. Zbilut, M. Tomita, Implications from a network-based topological analysis of ubiquitin unfolding simulations, *PLoS ONE* 3 (2008) e2149.
- [14] M.C. Palumbo, A. Colosimo, A. Giuliani, L. Farina, Essentiality is an emergent property of metabolic network wiring, *FEBS Lett.* 581 (2007) 2485–2489.
- [15] A. Krishnan, A. Giuliani, J.P. Zbilut, M. Tomita, Network scaling invariants help to elucidate basic topological principles of proteins, *J. Proteome Res.* 6 (2007) 3924–3934.
- [16] A. Krishnan, A. Giuliani, M. Tomita, Indeterminacy of reverse engineering of gene regulatory networks: the curse of gene elasticity, *PLoS ONE* 2 (2007) e562.
- [17] K. Tun, P.K. Dhar, M.C. Palumbo, A. Giuliani, Metabolic pathways variability and sequence/networks comparisons, *BMC Bioinf.* 7 (2006) 24.
- [18] L.G. Pérez-Montoto, F. Prado-Prado, F.M. Ubeira, H. González-Díaz, Study of parasitic infections, cancer, and other diseases with mass-spectrometry and quantitative proteome-disease relationships, *Curr. Proteomics* 6 (2009) 246–261.
- [19] H. González-Díaz, Y. González-Díaz, L. Santana, F.M. Ubeira, E. Uriarte, Proteomics, networks and connectivity indices, *Proteomics* 8 (2008) 750–778.
- [20] J. Caballero, M. Fernandez, 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.* 8 (2008) 1580–1605.
- [21] L. Fernández, J. Caballero, J.I. Abreu, M. Fernández, Aminoacid sequence autocorrelation vectors and Bayesian-regularized genetic neural networks for modeling protein conformational stability: gene v protein mutants, *Proteins* 67 (2007) 834–852.
- [22] M. Fernández, J. Caballero, L. Fernández, J.I. Abreu, M. Garriga, Protein radial distribution function (P-RDF) and Bayesian-regularized genetic neural networks for modeling protein conformational stability: chymotrypsin inhibitor 2 mutants, *J. Mol. Graph. Model.* 26 (2007) 748–759.



- [23] M. Fernández, F. Caballero, L. Fernández, J.I. Abreu, G. Acosta, Classification of conformational stability of protein mutants from 3D pseudo-folding graph representation of protein sequences using support vector machines, *Proteins* 70 (2008) 167–175.
- [24] J.P. Zbilut, A. Giuliani, A. Colosimo, J.C. Mitchell, M. Colafranceschi, N. Marwan, C.L. Webber Jr., V.N. Uversky, Charge and hydrophobicity patterning along the sequence predicts the folding mechanism and aggregation of proteins: a computational approach, *J. Proteome Res.* 3 (2004) 1243–1253.
- [25] J.P. Zbilut, A. Colosimo, F. Conti, M. Colafranceschi, C. Manetti, M. Valerio, C.L. Webber Jr., A. Giuliani, Protein aggregation/folding: the role of deterministic singularities of sequence hydrophobicity as determined by nonlinear signal analysis of acylphosphatase and abeta(1–40), *Biophys. J.* 85 (2003) 3544–3557.
- [26] M.A. Yildirim, K.I. Goh, M.E. Cusick, A.L. Barabasi, M. Vidal, Drug-target network, *Nat. Biotechnol.* 25 (2007) 1119–1126.
- [27] D. Viña, E. Uriarte, F. Orallo, H. González-Díaz, Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors, *Mol. Pharmacol.* 6 (2009) 825–835.
- [28] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*. Wiley-VCH, 2002.
- [29] H. Gonzalez-Díaz, F. Prado-Prado, F.M. Ubeira, Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach, *Curr. Top. Med. Chem.* 8 (2008) 1676–1690.
- [30] G. Agüero-Chapin, J. Varona-Santos, G.A. de la Riva, A. Antunes, T. Gonzalez-Villa, E. Uriarte, H. González-Díaz, Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from *coffea arabica* and prediction of a new sequence, *J. Proteome Res.* 8 (2009) 2122–2128.
- [31] R. Concu, M.A. Dea-Ayuela, L.G. Perez-Montoto, F. Bolas-Fernandez, F.J. Prado-Prado, G. Podda, E. Uriarte, F.M. Ubeira, H. Gonzalez-Díaz, Prediction of enzyme classes from 3D structure: a general model and examples of experimental-theoretic scoring of peptide mass fingerprints of *Leishmania* proteins, *J. Proteome Res.* 8 (2009) 4372–4382.
- [32] H. Gonzalez-Díaz, L. Saiz-Urra, R. Molina, L. Santana, E. Uriarte, A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions, *J. Proteome Res.* 6 (2007) 904–908.
- [33] C.R. Munteanu, J.M. Vazquez, J. Dorado, A.P. Sierra, A. Sanchez-Gonzalez, F.J. Prado-Prado, H. Gonzalez-Díaz, Complex network spectral moments for ATCUN motif DNA cleavage: first predictive study on proteins of human pathogen parasites, *J. Proteome Res.* 8 (2009) 5219–5228.
- [34] Y. Rodriguez-Soca, C.R. Munteanu, F.J. Prado-Prado, J. Dorado, A. Pazos Sierra, H. Gonzalez-Díaz, Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions, *J. Proteome Res.* (2009). doi:10.1021/pr900827b.
- [35] R. Concu, M.A. Dea-Ayuela, L.G. Perez-Montoto, F.J. Prado-Prado, E. Uriarte, F. Bolas-Fernandez, G. Podda, A. Pazos, C.R. Munteanu, F.M. Ubeira, H. Gonzalez-Díaz, 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in *Leishmania* parasites, *Biochim. Biophys. Acta* 1794 (2009) 1784–1794.
- [36] L. Santana, H. Gonzalez-Díaz, E. Quezada, E. Uriarte, M. Yanez, D. Vina, F. Orallo, Quantitative structure-activity relationship and complex network approach to monoamine oxidase A and B inhibitors, *J. Med. Chem.* 51 (2008) 6740–6751.
- [37] L. Santana, E. Uriarte, H. González-Díaz, G. Zagotto, R. Soto-Otero, E. Mendez-Alvarez, A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins, *J. Med. Chem.* 49 (2006) 1149–1156.
- [38] H. González-Díaz, Y. Pérez-Castillo, G. Podda, E. Uriarte, Computational chemistry comparison of stable/nonstable protein mutants classification models based on 3D and topological indices, *J. Comput. Chem.* 28 (2007) 1990–1995.
- [39] H. Gonzalez-Díaz, L. Saiz-Urra, R. Molina, Y. Gonzalez-Díaz, A. Sanchez-Gonzalez, Computational chemistry approach to protein kinase recognition using 3D stochastic van der Waals spectral moments, *J. Comput. Chem.* 28 (2007) 1042–1048.
- [40] G. Agüero-Chapin, A. Antunes, F.M. Ubeira, K.C. Chou, H. Gonzalez-Díaz, Comparative study of topological indices of macro/supramolecular RNA complex networks, *J. Chem. Inf. Model.* 48 (2008) 2265–2277.
- [41] M. Cruz-Monteagudo, C.R. Munteanu, F. Borges, M.N.D.S. Cordeiro, E. Uriarte, K.-C. Chou, H. González-Díaz, Stochastic molecular descriptors for polymers. 4. Study of complex mixtures with topological indices of mass spectra spiral and star networks: the blood proteome case, *Polymer* 49 (2008) 5575–5587.
- [42] M.A. Dea-Ayuela, Y. Perez-Castillo, A. Meneses-Marcel, F.M. Ubeira, F. Bolas-Fernandez, K.C. Chou, H. Gonzalez-Díaz, HP-Lattice QSAR for dynein proteins: experimental proteomics (2D-electrophoresis, mass spectrometry) and theoretic study of a *Leishmania infantum* sequence, *Bioorg. Med. Chem.* 16 (2008) 7770–7776.
- [43] G. Agüero-Chapin, H. Gonzalez-Díaz, G. de la Riva, E. Rodriguez, A. Sanchez-Rodriguez, G. Podda, R.I. Vazquez-Padron, MMM-QSAR recognition of ribonucleases without alignment: comparison with an HMM model and isolation from *Schizosaccharomyces pombe*, prediction, and experimental assay of a new sequence, *J. Chem. Inf. Model.* 48 (2008) 434–448.
- [44] G. Ferino, H. Gonzalez-Díaz, G. Delogu, G. Podda, E. Uriarte, Using spectral moments of spiral networks based on PSA/mass spectra outcomes to derive quantitative proteome–disease relationships (QPDRs) and predicting prostate cancer, *Biochem. Biophys. Res. Commun.* 372 (2008) 320–325.
- [45] H. Gonzalez-Díaz, M.A. Dea-Ayuela, L.G. Perez-Montoto, F.J. Prado-Prado, G. Agüero-Chapin, F. Bolas-Fernandez, R.I. Vazquez-Padron, F.M. Ubeira, QSAR for RNases and theoretic-experimental study of molecular diversity on peptide mass fingerprints of a new *Leishmania infantum* protein, *Mol. Divers.* (2009).
- [46] H. Gonzalez-Díaz, R. Molina, E. Uriarte, Recognition of stable protein mutants with 3D stochastic average electrostatic potentials, *FEBS Lett.* 579 (2005) 4297–4301.
- [47] R. Concu, G. Podda, E. Uriarte, H. Gonzalez-Díaz, 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.* 30 (2009) 1510–1520.
- [48] StatSoft, Inc., STATISTICA (Data Analysis Software System), Version 6.0. Statsoft Inc., 2002. [www.statsoft.com](http://www.statsoft.com).
- [49] G.M. Casanola-Martin, Y. Marrero-Ponce, M.T. Khan, S.B. Khan, F. Torrens, F. Perez-Jimenez, A. Rescigno, C. Abad, Bond-based 2D quadratic fingerprints in QSAR studies: virtual and in vitro tyrosinase inhibitory activity elucidation, *Chem. Biol. Drug Des.* 76 (2010) 538–545.
- [50] J.A. Castillo-Garit, M.C. Vega, M. Rolon, Y. Marrero-Ponce, V.V. Kouznetsov, D.F. Torres, A. Gomez-Barrio, A.A. Bello, A. Montero, F. Torrens, F. Perez-Gimenez, Computational discovery of novel trypanosomicidal drug-like chemicals by using bond-based non-stochastic and stochastic quadratic maps and linear discriminant analysis, *Eur. J. Pharm. Sci.* 39 (2010) 30–36.
- [51] R. Gozalbes, F. Barbosa, E. Nicolai, D. Horvath, N. Froloff, Development and validation of a pharmacophore-based QSAR model for the prediction of CNS activity, *ChemMedChem* 4 (2009) 204–209.
- [52] Y. Marrero-Ponce, A. Meneses-Marcel, O.M. Rivera-Borroto, R. Garcia-Domech, J.V. De Julian-Ortiz, A. Montero, J.A. Escario, A.G. Barrio, D.M. Pereira, J.J. Nogal, R. Grau, F. Torrens, C. Vogel, V.J. Aran, Bond-based linear indices in QSAR: computational discovery of novel anti-trichomonal compounds, *J. Comput. Aided Mol. Des.* 22 (2008) 523–540.
- [53] S.J. Patankar, P.C. Jurs, Classification of inhibitors of protein tyrosine phosphatase 1B using molecular structure based descriptors, *J. Chem. Inf. Comput. Sci.* 43 (2003) 885–899.
- [54] M. Murcia-Soler, F. Perez-Gimenez, F.J. Garcia-March, M.T. Salabert-Salvador, W. Diaz-Villanueva, P. Medina-Casamayor, Discrimination and selection of new potential antibacterial compounds using simple topological descriptors, *J. Mol. Graph. Model.* 21 (2003) 375–390.
- [55] R.A. Cercos-del-Pozo, F. Perez-Gimenez, M.T. Salabert-Salvador, F.J. Garcia-March, Discrimination and molecular design of new theoretical hypolipaeic agents using the molecular connectivity functions, *J. Chem. Inf. Comput. Sci.* 40 (2000) 178–184.
- [56] B.H. Junker, D. Koschützki, F. Schreiber, Exploration of biological network centralities with CentiBiN, *BMC Bioinf.* 7 (2006) 219.
- [57] D. Koschützki, pp. CentiBiN Version 1.4.2, centralities in biological networks<sup>®</sup> 2004–2006 Dirk Koschützki research group network analysis. IPK Gate-rsleben, Germany, 2006.
- [58] Hypercube Inc., Hyperchem Software. Release 7.5 for Windows, Molecular Modeling System. Hypercube Inc., Gainesville, FL, USA, 2002.
- [59] M. Froimowitz, HyperChem: a software package for computational chemistry and molecular modeling, *BioTechniques* 14 (1993) 1010–1013.
- [60] I. Hypercube, Hyperchem Inc. Gainesville, FL, USA, 2002.
- [61] Y. Liu, D.L. Beveridge, Exploratory studies of ab initio protein structure prediction: multiple copy simulated annealing, AMBER energy functions, and a generalized born/solvent accessibility solvation model, *Proteins* 46 (2002) 128–146.
- [62] M. Mezo, M. Gonzalez-Warleta, F.M. Ubeira, Optimized serodiagnosis of sheep fascioliasis by Fast-D protein liquid chromatography fractionation of *Fasciola hepatica* excretory–secretory antigens, *J. Parasitol.* 89 (2003) 843–849.
- [63] O. Ivanciuc, N. Oezguen, V.S. Mathura, C.H. Schein, Y. Xu, W. Braun, Using property based sequence motifs and 3D modeling to determine structure and functional regions of proteins, *Curr. Med. Chem.* 11 (2004) 583–593.
- [64] C.H. Schein, O. Ivanciuc, W. Braun, Common physical–chemical properties correlate with similar structure of the IgE epitopes of peanut allergens, *J. Agric. Food Chem.* 53 (2005) 8752–8759.
- [65] Y.M. Alvarez-Ginarte, Y. Marrero-Ponce, J.A. Ruiz-Garcia, L.A. Montero-Cabrera, J.M. Vega, P. Noheda Marin, R. Crespo-Otero, F.T. Zaragoza, R. Garcia-Domech, Applying pattern recognition methods plus quantum and physico-chemical molecular descriptors to analyze the anabolic activity of structurally diverse steroids, *J. Comput. Chem.* (2007).
- [66] A.H. Morales, J.E. Rodríguez-Borges, X. García-Mera, F. Fernández, M.N. Dias-Sueiro-Cordeiro, Probing the anticancer activity of nucleoside analogs: a QSAR model approach using an internally consistent training set, *J. Med. Chem.* 50 (2007) 1537–1545.
- [67] M. Fernandez, J. Caballero, A. Tundidor-Camba, Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl)amino]acetamide derivatives as matrix metalloproteinase inhibitors, *Bioorg. Med. Chem.* 14 (2006) 4137–4150.
- [68] J. Caballero, M. Fernandez, Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks, *J. Mol. Model.* 12 (2006) 168–181.
- [69] D.A. Chung, E.R. Zuiderweg, C.B. Fowler, O.S. Soyler, H.I. Mosberg, R.R. Neubig, NMR structure of the second intracellular loop of the alpha 2A adrenergic



- receptor: evidence for a novel cytoplasmic helix, *Biochemistry* 41 (2002) 3596–3604.
- [70] M. Katragadda, M.W. Maciejewski, P.L. Yeagle, Structural studies of the putative helix 8 in the human beta(2) adrenergic receptor: an NMR study, *Biochim. Biophys. Acta* 1663 (2004) 74–81.
- [71] K. Khafizov, G. Lattanzi, P. Carloni, G protein inactive and active forms investigated by simulation methods, *Proteins* 75 (2009) 919–930.
- [72] H. Kalasz, G. Petroianu, K. Tekes, I. Klebovich, K. Ludanyi, Z. Gulyas, Metabolism of moexipril to moexiprilat: determination of in vitro metabolism using HPLC-ES-MS, *Med. Chem.* 3 (2007) 101–106.
- [73] G.S. Chrysant, P.K. Nguyen, Moexipril and left ventricular hypertrophy, *Vasc. Health Risk Manage.* 3 (2007) 23–30.
- [74] F. Sayegh, J. Topouchian, M. Hlawaty, M. Olzewska, R. Asmar, Regression of left ventricular hypertrophy with moexipril, an angiotensin-converting enzyme inhibitor, in hypertensive patients, *Am. J. Therapeut.* 12 (2005) 3–8.
- [75] A. Innocenti, R.A. Hall, C. Schlicker, A. Scozzafava, C. Steegborn, F.A. Muhlschlegel, C.T. Supuran, Carbonic anhydrase inhibitors. Inhibition and homology modeling studies of the fungal beta-carbonic anhydrase from *Candida albicans* with sulfonamides, *Bioorg. Med. Chem.* 17 (2009) 4503–4509.
- [76] C. Temperini, A. Cecchi, A. Scozzafava, C.T. Supuran, Carbonic anhydrase inhibitors. Comparison of chlorthalidone, indapamide, trichloromethiazide, and furosemide X-ray crystal structures in adducts with isozyme II, when several water molecules make the difference, *Bioorg. Med. Chem.* 17 (2009) 1214–1221.
- [77] P. Dallemagne, S. Rault, M. Cugnon de Sévricourt, K.M. Hassan, M. Robba, *Tetrahedron Lett.* 23 (1986) 2607–2610.
- [78] S. Rault, P. Dallemagne, M. Robba, *Bull. Soc. Chim. Fr.* 6 (1987) 1079–1083.
- [79] M.A. Quermonne, P. Dallemagne, J. Louchahi-Raould, J.C. Pilo, S. Rault, M. Robba, *Eur. J. Med. Chem.* 27 (1992) 961–965.
- [80] P. Dallemagne, S. Rault, J.C. Pilo, M.P. Foloppe, M. Robba, *Tetrahedron Lett.* 32 (1991) 6327–6328.
- [81] P. Dallemagne, S. Rault, M. Gordaliza, M. Robba, *Heterocycles* 26 (1987) 3233–3237.
- [82] J. Guillon, G. Hebert, P. Dallemagne, J.M. Leger, C. Vidaillac, C. The, V. Lisowski, S. Rault, J. Demotes-Mainard, C. Jarry, Synthesis and initial results for MAO-B inhibition by new N-propargyl-3-pyrrol-1-ylindanamine derivatives, analogs of rasagiline, *J. Enzyme Inhib. Med. Chem.* 18 (2003) 147–153.
- [83] P. Abejón, J.M. Blanco, F. Fernández, M.D. García, O. Caamaño, Synthesis of two precursors of heterocarbocyclic nucleoside analogues, *Eur. J. Med. Chem.* (2003) 759–764.
- [84] M. Sakeda, S. Ichikawa, A. Matsuda, S. Shuto, The first radical method for the introduction of an ethynyl group using a silicon tether and its application to the synthesis of 2'-deoxy-2'-C-ethynynucleosides, *J. Org. Chem.* 68 (2003) 3465–3475.
- [85] N.T. Nguyen, D.M. Cook, L.A. Bero, The decision-making process of US food and drug administration advisory committees on switches from prescription to over-the-counter status: a comparative case study, *Clin. Ther.* 28 (2006) 1231–1243.
- [86] A. Duardo-Sanchez, G. Patlewicz, A. Lopez-Diaz, Current topics on software use in medicinal chemistry: intellectual property, taxes, and regulatory issues, *Curr. Top. Med. Chem.* 8 (2008) 1666–1675.
- [87] H. González-Díaz, F. Prado-Prado, L.G. Pérez-Montoto, A. Duardo-Sánchez, A. López-Díaz, QSAR models for proteins of parasitic organisms, plants and human guests: theory, applications, legal protection, taxes, and regulatory issues, *Curr. Proteomics* 6 (2009) 214–227.
- [88] H. González-Díaz, A. Duardo-Sanchez, F.M. Ubeira, F. Prado-Prado, L.G. Pérez-Montoto, R. Concu, G. Podda, B. Shen, Review of MARCH-INSIDE & complex networks prediction of drugs: aDMET, anti-parasite activity, metabolizing enzymes and cardiotoxicity proteome Biomarkers, *Curr. Drug Metab.* 11 (2010) 379–406.
- [89] H. González-Díaz, F. Romaris, A. Duardo-Sanchez, L.G. Perez-Montoto, F. Prado-Prado, G. Patlewicz, F.M. Ubeira, Predicting drugs and proteins in parasite infections with topological indices of complex networks: theoretical backgrounds, applications, and legal issues, *Curr. Pharm. Des.* 16 (2010) 2737–2764.
- [90] C. Yang, L.G. Valerio Jr., K.B. Arvidson, Computational toxicology approaches at the US food and drug administration, *Altern. Lab. Anim.* 37 (2009) 523–531.
- [91] Z. Lei, A.M. Elmer, B.S. Watson, R.A. Dixon, P.J. Mendes, L.W. Sumner, A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume *Medicago truncatula*, *Mol. Cell Proteomics* 4 (2005) 1812–1825.
- [92] M.C. Giddings, A.A. Shah, R. Gesteland, B. Moore, Genome-based peptide fingerprint scanning, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 20–25.
- [93] S. Dewilde, A.I. Ioaniteanu, L. Kiger, K. Gilany, M.C. Marden, S. Van Doorslaer, J. Vercruyssen, A. Pesce, M. Nardini, M. Bolognesi, L. Moens, The hemoglobins of the trematodes *Fasciola hepatica* and *Paramphistomum epiclitum*: a molecular biological, physico-chemical, kinetic, and vaccination study, *Protein Sci.* 17 (2008) 1653–1662.
- [94] F. Hubalek, C. Binda, A. Khalil, M. Li, A. Mattevi, N. Castagnoli, D.E. Edmondson, Demonstration of isoleucine 199 as a structural determinant for the selective inhibition of human monoamine oxidase B by specific reversible inhibitors, *J. Biol. Chem.* 280 (2005) 15761–15766.
- [95] C. Binda, F. Hubalek, M. Li, N. Castagnoli, D.E. Edmondson, A. Mattevi, Structure of the human mitochondrial monoamine oxidase B: new chemical implications for neuroprotectant drug design, *Neurology* 67 (2006) S5–S7.