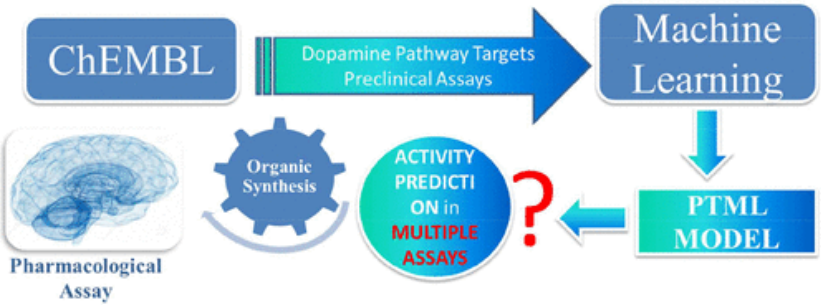
ACS Chemical Neuroscience. 2018; 9(11): 2572-2587

Perturbation theory/machine learning model of ChEMBL data for dopamine targets: docking, synthesis, and assay of new l-prolyl-l-leucyl-glycinamide peptidomimetics

Joana Ferreira da Costa[†], David Silva[†], Olga Caamaño[†], José M. Brea[‡§], Maria Isabel Loza[‡§], Cristian R. Munteanu[∥], Alejandro Pazos[∥⊥], Xerardo García-Mera[†], and Humbert González-Díaz[#g]

† Department of Organic Chemistry, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain
‡ CIMUS, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain
§ Department of Pharmacology, Pharmacy and Pharmaceutical Technology, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain
∥ Instituto de Investigacion Biomedica de A Coruña (INIBIC), Complexo Hospitalario Universitario de A Coruña (CHUAC), A Coruña, 15006, Spain
⊥ Computer Science Department, Faculty of Computer Science, University of A Coruna, 15071 A Coruña, Spain
 # Department of Organic Chemistry II, University of Basque Country UPV/EHU, 48940 Leioa, Spain
g IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

Abstract



Predicting drug–protein interactions (DPIs) for target proteins involved in dopamine pathways is a very important goal in medicinal chemistry. We can tackle this problem using Molecular Docking or Machine Learning (ML) models for one specific protein. Unfortunately, these models fail to account for large and complex big data sets of preclinical assays reported in public databases. This includes multiple conditions of assays, such as different experimental parameters, biological assays, target proteins, cell lines, organism of the target, or organism of assay. On the other hand, perturbation theory (PT) models allow us to predict the properties of a query compound or molecular system in experimental assays with multiple boundary conditions based on a previously known case of reference. In this work, we report the first PTML (PT + ML) study of a large ChEMBL data set of preclinical assays of compounds targeting dopamine pathway proteins. The best PTML model found predicts 50000 cases with accuracy of 70–91% in training and external validation series. We also compared the linear PTML model with alternative PTML models trained with multiple nonlinear methods (artificial neural network (ANN), Random Forest, Deep Learning, *etc.*). Some of the nonlinear methods outperform the linear model but at the cost of a notable increment of the complexity of the model. We illustrated the practical use of the new model with a proof-of-concept theoretical–experimental study. We reported for the first time the organic synthesis, chemical characterization, and

pharmacological assay of a new series of l-prolyl-l-leucyl-glycinamide (PLG) peptidomimetic compounds. In addition, we performed a molecular docking study for some of these compounds with the software Vina AutoDock. The work ends with a PTML model predictive study of the outcomes of the new compounds in a large number of assays. Therefore, this study offers a new computational methodology for predicting the outcome for any compound in new assays. This PTML method focuses on the prediction with a simple linear model of multiple pharmacological parameters (IC_{50}, EC_{50}, K_i, *etc.*) for compounds in assays involving different cell lines used, organisms of the protein target, or organism of assay for proteins in the dopamine pathway.

Keywords:
ChEMBL; Machine learning; Peptide organic synthesis; PLG peptidomimetics

INTRODUCTION

Medicinal Chemistry researchers have at least two important alternatives to study drug–protein interactions (DPIs) for all potential drug targets in one specific pathway. One alternative is the use of experimental techniques, and the other is the use of computational methods. Computational methods, in turn, belong to at least two main classes, machine learning (ML)[1–3] methods or molecular docking techniques implemented on AutoDock Vina software (or others).[4,5] In the case of ML methods, it is possible to calculate numerical parameters to quantify the chemical information on the compounds, peptides, or proteins.[6–22] However, in many cases both classes of methods fail to account for large data sets of preclinical assays with many factors not related to drug or target protein structure. These large sets of data, called big data, are difficult to model not only because of the large quantity of data but also due to the high complexity of it. For instance, the ChEMBL database includes very large sets of experimental outcomes obtained in very different conditions.[23,24] This includes different pharmacological parameters measured (IC_{50}, K_i, K_m, *etc.*), different cell lines used, different organisms of the protein target or organisms of assay, *etc*. In fact, study of the large data sets reported in this database is gaining interest in medicinal chemistry.[25–27] In our opinion, perturbation theory (PT) models allow us to predict the properties of a query compound or molecular system (q) in experimental assays with multiple boundary conditions based on a previous known case of reference (r). Specifically, PTML models combine PT and ML methods to study biomolecular systems.[18,28] In fact, PTML models may have applications in medicinal chemistry, nanotechnology, *etc.* to study large data sets with multiple experimental conditions or other parameters.[29–34] The PTML method uses different types of PT operators to account for perturbations in the structure of the system and the changing conditions of assay, for instance, Moving Average (MA) operators used in Box–Jenkins's ARIMA models for time series analysis.[35] MA operators of structural descriptors have been used to quantify complex data sets in medicinal chemistry and other areas.[36–39]

On the other hand, various biological activities have been assigned to peptidic and peptidomimetic structures. They vary from antiviral activities to neuroprotective properties in different animal models of neurodegenerative processes, such as Huntington's, Parkinson's, and Alzheimer's diseases.[40–42] In this context, the peptide PLG, known as melanocyte-stimulant hormone release-inhibiting factor, possesses a variety of pharmacological activities in the central nervous system.[43] PGL (l-prolyl-l-leucyl-glycinamide) and its analog PAOPA (3(*R*)-(2(*S*)-pyrrolidylcarbonyl)amino-2-oxo-1-pyrrolidineacetamide) modify dopaminergic neurotransmission by acting as allosteric modulators of dopamine D_2 receptor.[44–46] The allosteric modulators are safer alternatives to conventional orthosteric therapeutics; their mechanism involves modulation of endogenous signaling. Studies carried out in cell lines transfected with human dopamine receptor subtypes have shown that PLG and PLG peptidomimetics enhance agonist binding to the D_{2S}, D_{2L}, and D_4 subtypes.[47,48]

In this context, the development of new computational models for the prediction of DPIs for dopamine pathway drug targets is a goal of major importance. In this work, we report for the first time a PTML model able to predict DPIs for a large ChEMBL data set of preclinical compound assays related to dopamine pathway drug targets. The assays include many different pharmacological parameters, target proteins, cell lines, organisms of the target, organisms of assay, *etc.* The data set includes over 50 000 cases of multivariate perturbations in different assay conditions. We illustrated the practical use of the new model with a proof-of-concept theoretical–experimental study. Therefore, for the first time, organic synthesis, chemical characterization, and pharmacological assay of a new series of PLG peptidomimetic compounds are reported. In addition, the molecular docking study for some of these compounds is presented. In addition, we performed a molecular docking study for some of these compounds with the software Vina AutoDock. The work ends with a PTML model predictive study of the outcomes of the new compounds in a large number of assays. Therefore, this study offers a new computational methodology for predicting the outcome for any compound in new assays. This PTML method focuses on prediction with a simple linear model of multiple pharmacological parameters (IC_{50}, EC_{50}, K_i, *etc.*) for compounds in assays involving different cell lines used, organisms of the protein target, or organisms of assay for proteins in the dopamine pathway. In Figure 1, we summarize the workflow used in this work, emphasizing the four main components of the work: (1) PTML model development (general methodology), (2) synthesis of new compounds (experimental example), (3) assay of new compounds (experimental example), and (4) molecular docking (additional studies).

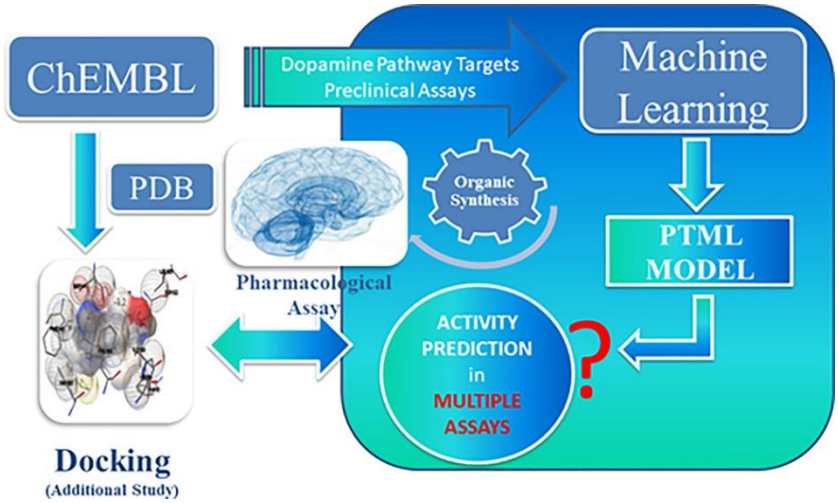
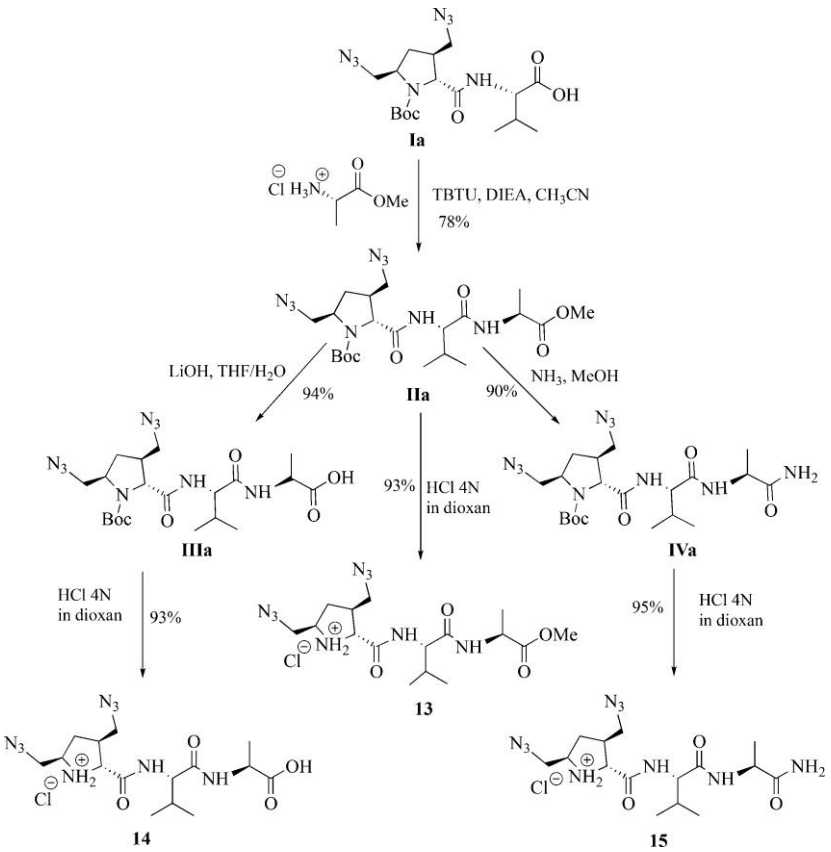


Figure 1. Workflow of the present work.

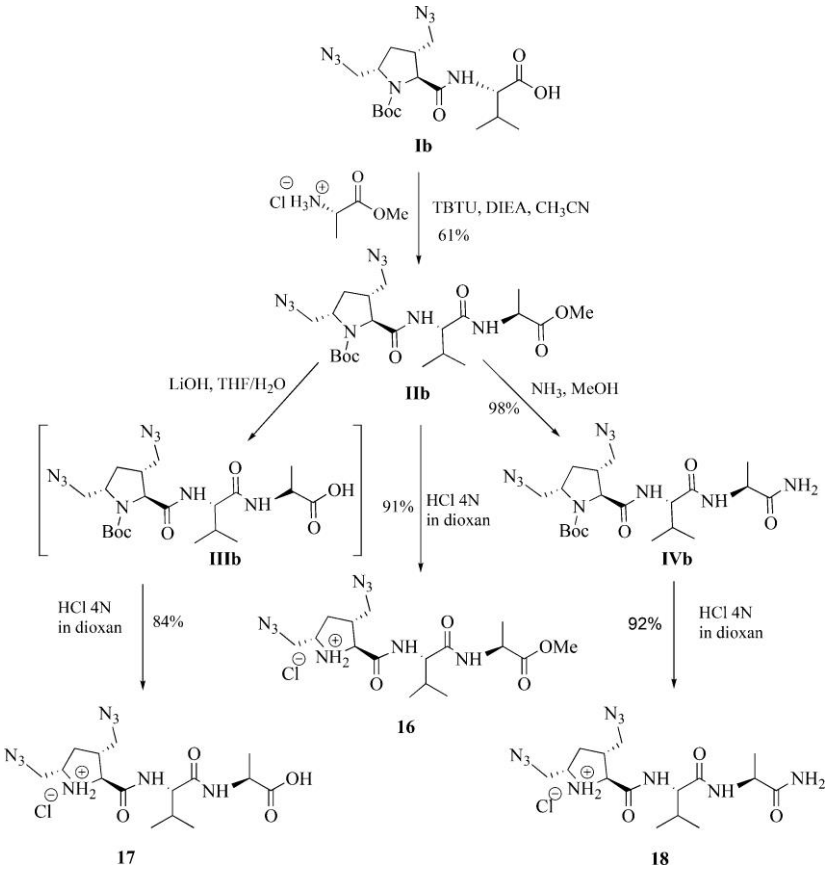
RESULTS AND DISCUSSION

Synthesis of the New PLG Peptidomimetics. Synthesis of compounds **1**–**12** and **19**–**21** was described by us in a previous work.[49] Compounds **13**–**18** were synthesized according to the strategy given in Schemes 1 and 2. As shown in these schemes, they were synthesized from the corresponding dipeptide diastereoisomers **IIa** and **IIb** as described in the earlier work.[49] The coupling of the corresponding carboxylic acids **Ia** and **Ib** with alanine methyl ester hydrochloride provided the corresponding tripeptide esters **IIa** and **IIb**. Compound **IIa** was converted to the corresponding carboxylic acid **IIIa** with LiOH in THF/H_2O. Nevertheless, the hydrolysis of **IIb** with LiOH showed, as in the similar compounds,[49] high lability of the carbamate group during the acidification process with 1 M H_2SO_4; for this reason removal of the Boc protecting group was performed on the dry residue of the reaction mass containing **IIIb**. The transformation of **IIa** and **IIb** into the corresponding primary amides

IVa and **IVb** was achieved by treatment with methanolic ammonia. Finally, removal of the Boc protecting group from **IIa**–**IVa**, **IIb**, and **IVb** by reaction with 4 N HCl in dioxane afforded the corresponding **13**–**18** PLG analogues with excellent yields.



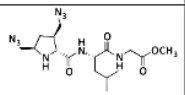
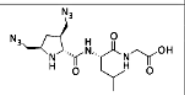
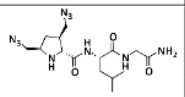
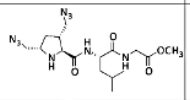
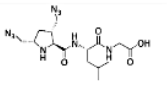
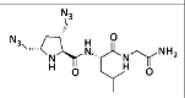
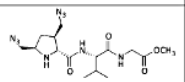
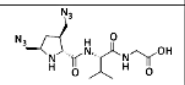
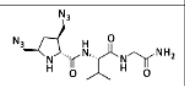
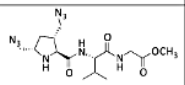
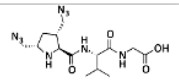
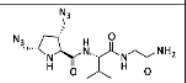
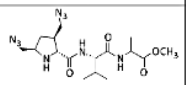
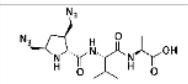
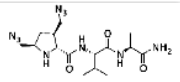
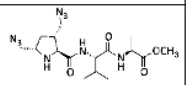
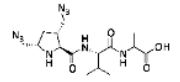
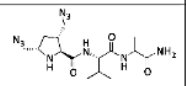
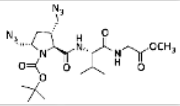
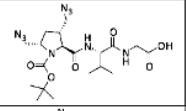
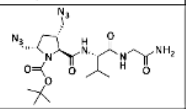
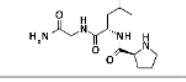
Scheme 1. Synthesis of Compounds **13**–**15**



Scheme 2. Synthesis of Compounds **16**–**18**

Experimental Assay of New PLG Peptidomimetics. In Table 1 and Figure 2, we summarize the experimental results obtained in the pharmacological assay of this series of compounds. The data obtained for PLG analog photoaffinity labeling agents **13**–**21** are shown, and their activity was compared to that of PLG. All the compounds significantly ($P < 0.05$; ANOVA test; post hoc Dunnet T3 test) increased [3H]NPA binding at low concentrations. Compound **13** increased [3H]NPA binding higher than that observed with PLG at the same concentrations. Both compounds **17** and **18** increased [3H]NPA binding: **17** showed 40% increased [3H]NPA binding at the concentration 10^{-7} M, and **18** caused increased [3H]NPA binding higher than 40% in the range of concentrations 10^{-12}–10^{-7} M, showing a better profile than PLG. The effect observed with PLG in our work is similar to that observed in our previous work.[49] However, it is slightly different from that previously reported by Verma *et al.*,[50] showing increased [3H]NPA binding at lower concentrations than those previously reported. This difference can be explained by the different host cell where human D_{2S} receptors were expressed, as it has been previously reported that allosteric modulators are sensitive to environmental changes; which may condition the different active conformations elicited by the endogenous agonists on GPCRs.[51] Most of the compounds showed a bell-shaped curve, which is compatible with previous findings from *in vivo* and clinical experiments using PLG.[50] These data evidence that these compounds are acting as allosteric modulators of dopamine D_{2S} receptors. To view the profiles observed in the series of analogs of the PLG tested, we can conclude that when the change affects the chain of leucine, this happens in both series of compounds. The allosteric modulator effect is lower in the series of compounds **19**–**21**. In contrast, if the leucine is replaced by valine and the amine group of the proline cycle is protected as carbamate, the series of compounds **13**–**18**, where the glycine is replaced by alanine, showed the most similar profiles to PLG in the increase of the union of [3H]NPA to dopamine D_2 receptors.

Table 1. New Experimental Results for Novel PGL Derivatives in Dopamine D_2 Receptor Assay

Cmpd[a]	Structure	%[^{3}H] NPA$_{max}$	AUC	Log (1/C$_{max}$)	Synth. Ref.[b]	Assay Ref.[c]
1		14.0	42.12	9.0	49	49
2		15.0	72.17	10.0	49	49
3		15.0	66.86	11.0	49	49
4		16.0	58.72	10.0	49	49
5		n.a.	n.a.	n.a.	-	-
6		17.0	67.73	10.0	49	49
7		18.0	58.67	11.0	49	49
8		20.0	57.2	11.0	49	49
9		14.0	41.18	10.0	49	49
10		22.0	40.38	12.0	49	49
11		39.0	94.42	11.0	49	49
12		28.0	92.52	12.0	49	49
13		57.0	182.4	11.0	This work	This work
14		21.0	88.45	11.0	This work	This work
15		27.0	64.3	11.0	This work	This work
16		19.0	66.45	5.0	This work	This work
17		47.0	114.1	7.0	This work	This work
18		55.0	266	11.0	This work	This work
19		37.0	139.8	9.0	49	This work
20		32.0	120.7	12.0	49	This work
21		26.0	47.51	10.0	49	This work
PLG		31.0	98.26	11.0	control	control

[a] Compounds studied in this work; compound **5** was not assayed due to failure of the synthetic route. [b] Reference in which the synthesis of the compound was published for the first time. [c] Reference in which the assay of the compound was published for the first time.

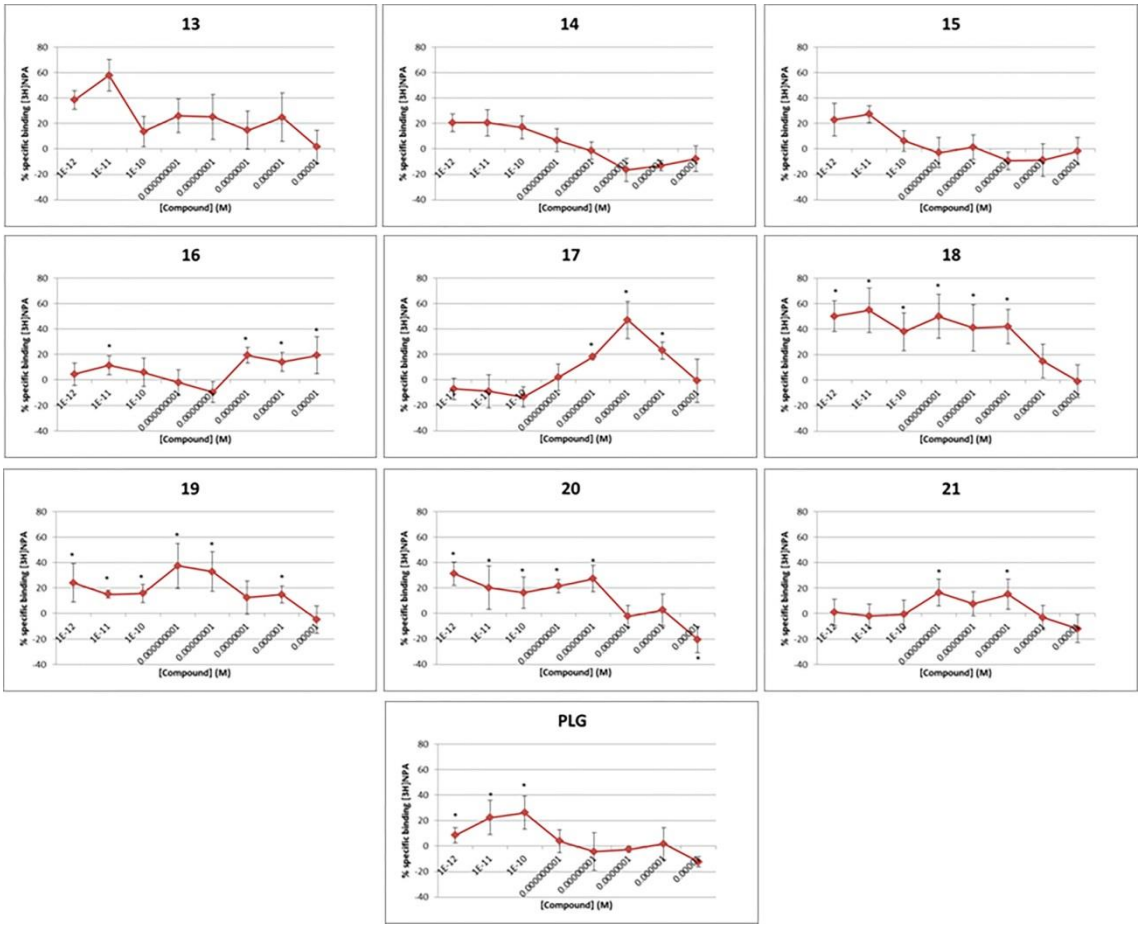


Figure 2. Experimental assay of new PLG peptidomimetics.

Molecular Docking Simulation. After the synthesis and assay of the new series of PLG peptidomimetics, a ligand–protein molecular docking simulation was done for these compounds. Multiple target proteins involved in the dopamine pathway with 3D spatial models reported in the PDB have been selected for this study. The pdb codes of the proteins selected are 1I15, 1OZ5, 1ZHB, 2LYW, 2MHO, 3PBL, 4OAJ, 4ZEL, 5AER, 5I6X, 5I6Z, 5I71, 5I73, 5I74, and 5I75. The affinity energy (AE) values for a total of 20 peptides (compounds) × 15 proteins = 300 DPIs in the dopamine pathway have been calculated. In addition, a two-way joining cluster analysis (TWJCA) of these results was carried out. TWJCA facilitates the visualization and quantitative analysis of the results obtained (see Figure 3). It can be observed that proteins with codes 5AER, 2LYW, 2MHO, and 4OAJ form a cluster with AE > −5 kcal/mol, indicating weak DPIs. In contrast, the other proteins form a cluster with AE = −7 to −9 kcal/mol. This indicates certain coherence in the predicted behavior of this series of compounds in terms of DPI profile.

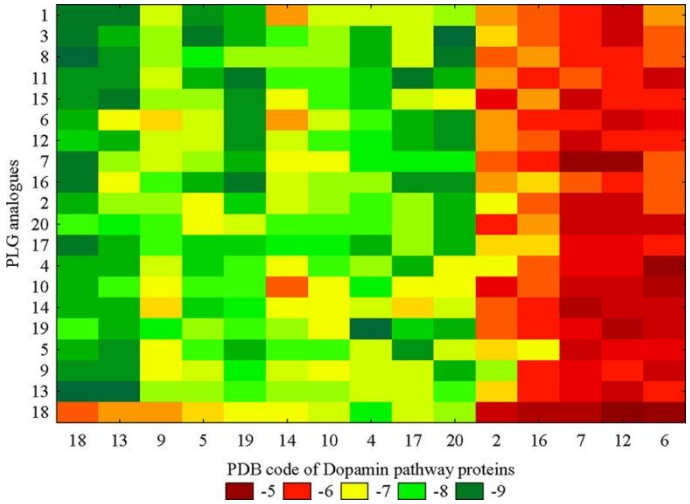


Figure 3. TWJCA of docking results for selected analogues.

In Table 2, we report the average value (Avg), standard deviation (SD), and minimum value (Min) of the AE (kcal/mol) for all 300 DPIs. The protein with the lowest average affinity energy, AE = −8.6 kcal/mol, for this series of peptides is the dopamine D_2 receptor (PDB 1I15). This protein also has one of the more negative AE = −9.4 kcal/mol (min energy) for peptide **14** (PVA-l-Val-l-Ala). Interestingly, compound **18**, the compound that increased [3H]NPA binding to dopamine receptors also has a relatively stronger DPI predicted (AE = −9.2 kcal/mol) with dopamine D_2 receptor (PDB 1I15). The compound **18** also has a relevant (AE = −8.0 kcal/mol) but weaker predicted interaction with dopamine D_3 receptor (human), see Table 2. In Figure 4, we depict the details of the DPI for the compound **18** and protein model 1I15. The more relevant amino acids in the binding pocket seem to be Trp115, Ser122, Ser141, Phe82, Leu125, Trp182, Cys90, Phe186, Ser118, Val87, His189, Phe145, and Ile190

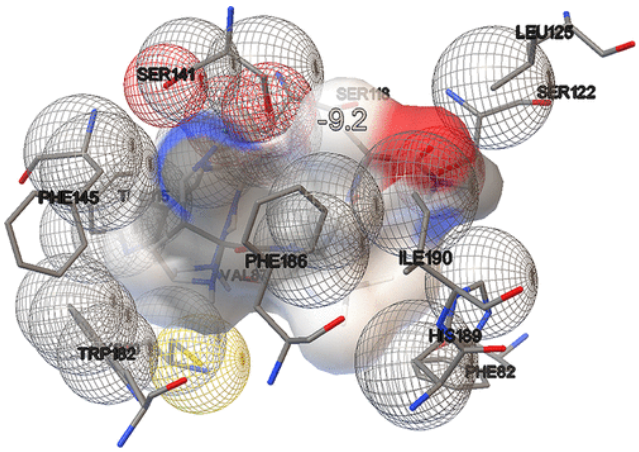
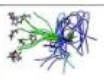
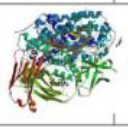


Figure 4. Docking interactions between peptide **18** (ligand) and PDB 1I15 (target).

Table 2. Results of the Docking Study of Dopamine Pathway Proteins *versus* PLG Peptidomimetics

Protein PDB	Protein Name	Protein model	AE(kcal/mol) [a]					AE_{min}
			18	**17**	Avg	SD	Min	Cmpd
1I15	Dopamine D2 Receptor		-9.2	-9.1	-8.6	0.7	-9.4	**14**
1ZHB	Dopamine β-monooxigenase (rat)		-8.0	-7.3	-7.3	0.4	-8	**18**
1OZ5	D1 dopamine receptor	-	-6.9	-6.5	-6.5	0.5	-7.5	**10**
2LYW	Neurotensin 1 receptor		-5.9	-6.3	-5.8	0.3	-6.3	**12**
2MHO	5HTDC receptor (peptide)		-5.8	-6	-5.7	0.3	-6.2	**12**
3PBL	Dopamine D3 receptor (human)		-8.0	-7.6	-7.5	0.3	-8	**18**
4OAJ	5HT2A receptor		-6.1	-6.4	-5.9	0.5	-6.5	**1**
4ZEL	Dopamine β-hydroxylase		-7.8	-7.8	-7.4	0.4	-8.1	**20**
5AER	Dopamine D2 receptor		-6.8	-6.9	-6.3	0.4	-7.0	**6**
5I75	Na-dependent serotonin transporter		-8.4	-9.2	-8.3	0.60	-9.2	**17**

[a] AE (kcal/mol) values for compounds **18** and **17**, and average (Avg), standard deviation (SD), and minimum (Min) value of all the 300 DPIs studied.

PTML-Linear Discriminant Analysis (LDA) Model. The previous experimental and computational molecular docking studies are very useful to know details of DPIs for new compounds. However, the experimental studies have and important cost in terms of sacrifice of laboratory animals, material resources, and time. The molecular docking calculations have an important cost in computational resources too. Therefore, the experimental or molecular docking study of large sets of experimental conditions may be a goal hard to accomplish for many research groups. Previous publications showed that the PTML analysis of big data sets may be a useful tool. PTML is especially useful if it is a study with multiple pharmacological parameters for a series of compounds in different target proteins, organisms of assay, cell lines, *etc*. In the current work, the first PTML model for the DPIs of the dopamine pathway was developed. A large data set of preclinical assays was downloaded from the ChEMBL database. The

data set is large and complex including different target proteins, cellular lines, organisms, *etc*. PT operators of molecular descriptors have been used as inputs to train a PTML classifier. In Table 3, we depict the results for the best PTML linear model obtained (PTML-LDA model): PTML model obtained with the LDA algorithms. In this table, the overall results for training and validation subsets are presented. In training series, the model presented high values of specificity = Sp (%) = 72.8, sensitivity = Sn (%) = 72.4, and overall accuracy = Ac (%) = 72.7. The model was stable in external validation series with values of Sp (%) = 72.7, Sn (%) = 71.4, and Ac (%) = 72.6. The equation of this model is presented in eq 1.

$$\begin{aligned} f(\varepsilon_{ij})_{\text{new}} = & -10.780386430140 - 0.000000000020 f(\varepsilon_{ij})_{\text{expt}} + 0.440071875560 \Delta\mathcal{D}_1(c_0) \\ & + 0.465335484664 \Delta\mathcal{D}_1(c_1) - 0.541834505781 \Delta\mathcal{D}_1(c_2) \\ & - 0.127705300409 \Delta\mathcal{D}_1(c_3) - 0.114637007349 \Delta\mathcal{D}_1(c_4) \\ & - 0.095637330548 \Delta\mathcal{D}_2(c_5) - 0.054733584740 \Delta\mathcal{D}_2(c_6) \\ & - 0.056732915285 \Delta\mathcal{D}_2(c_7) \quad n = 41082 \quad \chi^2 = 5564.0 \quad p < 0.05 \end{aligned}$$

Table 3. PTML Analysis Results (LDA Method) for DPIs in Dopamine Pathway

			predicted sets		
observed sets[a]	statistical parameter[b]	predicted statistics	n_j	$L_{ij} = 0$	$L_{ij} = 1$
Training Series					
$L_{ij} = 0$	Sp	0.701	37080	26005	11075
$L_{ij} = 1$	Sn	0.839	4001	644	3357
total	Ac	0.715			
External Validation Series					
$L_{ij} = 0$	Sp	0.702	12364	8675	3689
$L_{ij} = 1$	Sn	0.833	1329	222	1107
total	Ac	0.714			

[a] The classification classes are (1) drugs with biological effect over the receptor stronger than the average ($L_{ij} = 1$ when $\varepsilon_{ij} > \text{avg}(\varepsilon_{ij})_j$) and (2) $L_{ij} = 0$ otherwise.
[b] Sn = sensitivity, Sp = specificity, and Ac = accuracy

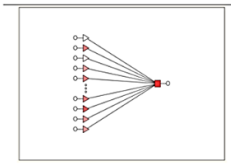
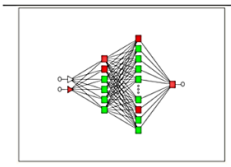
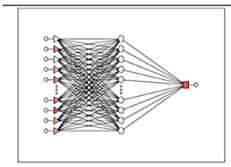
PTML model starts with the expected value of activity and adds the effect of different perturbations in the system. Consequently, the model has two types of input variables: the expected-value function, $f(\varepsilon_{ij})_{\text{expt}}$, and the PT operators, $\Delta D_1(\mathbf{c}_j)$. The input variable $f(\varepsilon_{ij})_{\text{expt}}$ represents the expected value of biological activity for one compound in experimental conditions $\mathbf{c}_j$. We investigated two types of PT operators, $\Delta D_1(\mathbf{c}_j)$ and $\Delta D_2(\mathbf{c}_j)$. The type 1 operators, $\Delta D_1(\mathbf{c}_j) = \text{ALOGP}_i - \langle\text{ALOGP}(\mathbf{c}_j)\rangle$, account for the effects on biological activity of perturbations of the drug hydrophobicity. We measured these perturbations as a deviation ΔALOGP in the value of ALOGP_i of the *i*th drug with respect to the average value, $\langle\text{ALOGP}(\mathbf{c}_j)\rangle$, for all drugs tested under condition $\mathbf{c}_j$. The operators selected by the forward-stepwise strategy of variable selection are sensitive to perturbations of the organism of assay, cell line, *etc.* On the other hand, the type 2 operators, $\Delta D_2(\mathbf{c}_j)$, account for the effect on biological activity of perturbations of the polar surface area (PSA) of the drug. These perturbations are calculated as a deviation, $\Delta\text{PSA}(\mathbf{c}_j)$, in the value of PSA_i of the *i*th drug with respect to the average value, $\langle\text{PSA}(\mathbf{c}_j)\rangle$, for all drugs tested under condition $\mathbf{c}_j$. Table 4presents more detailed explanation about all input variables. After obtaining this model, the next step was to illustrate its use by using a practical example. Therefore, the synthesis, biological assay, and computational study of a new series of PLG peptidomimetic compounds have been presented (see next section).

Table 4. Input Variables of the PTML Model

condition (c_j)	operator symbol	source variable	operator type formula	operator information
activity type (c_0)	$f(\varepsilon_{ij})$	$p(\Delta\varepsilon_{ij}>0)_{ref}$	$n(\Delta\varepsilon_{ij}>0)/n_j$	expected value of probability $p(L_{ij} = 1)$ for a given type of activity (ε_{ij})
activity type (c_0)	$\Delta D_1(c_0)$	ALOGP	$ALOGP_i - \langle ALOGP(c_j)\rangle$	quantifies changes or perturbations in the hydrophobicity of the compound due to changes in the chemical structure of the compound ($ALOGP_i$) and changes in the conditions of assay ($\langle ALOGP(c_j)\rangle$)
target (c_1)	$\Delta D_1(c_1)$			
organism of target (c_2)	$\Delta D_1(c_2)$			
organism of assay (c_3)	$\Delta D_1(c_3)$			
cell line (c_4)	$\Delta D_1(c_4)$			
assay type (c_5)	$\Delta D_2(c_5)$	PSA	$PSA_i - \langle PSA(c_j)\rangle$	quantifies change or perturbations in the hydrophobicity of the compound due to changes in the chemical structure of the compound (PSA_i) and changes in the conditions of assay ($\langle PSA(c_j)\rangle$)
confidence (c_6)	$\Delta D_2(c_6)$			
curation (c_7)	$\Delta D_2(c_7)$			

PTML-Artificial Neural Network (ANN) Nonlinear Models. In addition to the linear models, different nonlinear ML methods have been used. ANN algorithms have been tested, in a first instance, to find a better PTML model with higher accuracy (Ac) and sensitivity (Sn); see results in Table 5. The simplest PTML-ANN model is the linear neural network (LNN). The LNN model is similar to the LDA in performance but with 19 input variables. The multilayer perceptron (MLP) model has also similar performance and is even more complicated with a hidden layer of neurons. It is straightforward to realize that both the PTML-LNN and the PTML-MLP models do not outperformed the PTML-LDA model. However, the nonlinear PTML-radial basis function (RBF) model has values of Ac and Sn in the range 80–83% in training and external validation series. These values, in the case of Ac, are 10% higher than the values obtained with the PTML-LDA model. The PTML-RBF model has also good area under the receiver operator characteristic (AUROC) curve values in the range 0.89–0.88 in training and validation series. This demonstrates that the classification results are higher than the values expected for a random model (AUROC = 0.5).(52) In any case, the PTML-RBF model with >700 neurons in the hidden layer is also notably more complicated than the PTML-LDA model.

Table 5. Results for the ANN-PTML Nonlinear Models[a]

PTML-ANN modelprofile	Set L_{ij}	Training series 1	Training series 0	Training series (%)	Stat. Param.	Validation series (%)	Validation series 1	Validation series 0
 LNN 19:19-1:1	1	3071	8488	76.7	Sn	77.1	1025	2865
	0	931	28592	77.1	Acc	76.8	304	9499
 MLP 2:2-6-12-1:1	1	2975	9664	74.3	Sn	73.7	980	3275
	0	1027	27416	73.9	Acc	73.5	349	9089
 RBF 19:19-786-1:1	1	3327	6643	83.1	Sn	80.2	1066	2322
	0	675	30437	82.1	Acc	81.2	263	10042

[a] Stat. Param. = statistical parameters, Sn = sensitivity, and Acc = accuracy.

PTML Nonlinear Models. In addition, we used the software Weka(53) in order to test other nonlinear ML methods with the same features from the linear model. The Ac values for test subset are superior to the ANN results: 0.866 with Bayes Nets,(54) 0.915 with Random Tree,(55) 0.927 with Random Forest, and 0.908 with Decision Table.(56) All the AUROC values are greater than 0.78. The last tested classifier was a Deep Learning model (a hidden layer with 1000) using Keras/python. The test accuracy obtained was Ac = 0.909 (training Ac = 0.908). This is a superior value compared with ANN and Bayes Net, but not if we compare it with Random Tree, Random Forest, and Decision Trees. If all the features are included, the accuracy is improved, Ac = 0.913 (test) and 0.917 (training). These values are greater than all the Weka ML classifier results. However, taking into consideration the complexity factor, the understanding of the model, and the differences between the Ac values, the linear model was the choice for the predictions in the next sections.

PTML Study of New PLG Peptidomimetics. After obtaining the best model, its use was demonstrated with a practical example. In the previous section, the synthesis and biological assay of a new series of PLG peptidomimetic compounds have been presented. In this section, the use of the PTML model to predict the min–max scale probabilities, $*p(\Delta\varepsilon_{ij} > 0)$ of having a biological activity ε_{ij} higher than the expected value $\langle\varepsilon_{ij}\rangle$ will be illustrated. This is equivalent to the probability of having $\Delta\varepsilon_{ij} = (\varepsilon_{ij} - \langle\varepsilon_{ij}\rangle) > 0$. In order to achieve this goal, the parameters used to specify both the structure of the compounds and the conditions $\mathbf{c}_j$ for different biological assays were set in the PT operators of the model. In order to specify the molecular structure of the peptides, the molecular descriptors D_1 = ALOGP and D_2 = PSA were used for each peptide. In addition, the expected values of the parameters for different conditions need to be used to specify the conditions of assay (see Table 6).

Table 6. Expected Values $\langle D_k \rangle$ of Molecular Descriptors for Assay Conditions (Selected Examples)

c_0 = Activity	$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j	c_1 = target protein	$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j
K_i (nM)	4.12	47.07	28030	dopamine D_2 receptor	4.16	54.74	11887
IC_{50} (nM)	3.79	44.31	10740	dopamine D_3 receptor	4.14	54.48	6844
ratio	4.27	40.10	3034	dopamine transporter	3.95	35.64	5753
inhibition (%)	3.67	46.45	2877	dopamine D_4 receptor	3.83	43.39	3487
EC_{50} (nM)	3.44	57.44	1595	dopamine D_1 receptor	3.68	57.18	2995
potency (nM)	2.97	67.04	1267	5HT&D transporters	4.55	32.09	1141
activity (%)	3.84	55.83	1127	NE&D transporters	3.99	36.17	551
selectivity ratio	4.01	42.21	577	dopamine β-hydroxylase	2.12	54.21	280
c_3 = assay organism		$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j	c_2 = protein organism	$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j
Mus musculus		3.63	61.58	232	*Homo sapiens*	4.02	50.08	33240
Homo sapiens		4.03	51.11	21699	*Rattusnorvegicus*	3.91	42.33	19265
Rattusnorvegicus		3.94	43.62	7342	*Bostaurus*	2.55	53.42	1600
Cricetulusgriseus		3.40	25.89	82	*Mus musculus*	3.94	47.27	613
Carassiusauratus		3.71	75.86	87	*Susscrofa*	4.58	51.66	352
Canis lupus familiariz		3.38	46.67	50	*Carassiusauratus*	3.71	75.86	87
Bostaurus		3.26	41.80	429	*Canis lupus familiariz*	2.03	81.58	75
Susscrofa		4.39	48.18	443	*Chlorocebusaethiops*	3.99	33.62	42
c_4 = cell line		$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j	c_5 = assay type	$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j
CHO		4.17	50.68	6846	binding (B)	3.99	46.41	46964
HEK293		3.96	43.95	4835	functional (F)	3.62	53.60	8336
CHO-K1		4.05	41.52	1079				
Sf9		4.81	46.21	227				
C6		4.70	49.92	122				
COS-7		4.78	18.38	29				
LLC-PK1		4.08	45.54	15				
SCH		4.01	34.60	11				
HeLa		4.25	77.75	7				
SK-N-MC		4.33	26.90	4				
c_6 = target conf. score		$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j	c_7 = data curation	$\langle D_1(c_j)\rangle$	$\langle D_2(c_j)\rangle$	n_j
single protein		3.98	50.34	28648	expert	3.97	43.75	24129
multiple proteins		4.14	42.14	2539	autocuration	3.90	50.97	28000
homologous protein		3.84	45.17	20645	intermediate	4.01	45.59	3173

For instance, with the values $\langle D_1(c_0)\rangle$ = 3.67 and $\langle D_2(c_0)\rangle$ = 47.07, we specify that we want to calculate the $*p(\Delta\varepsilon_{ij} > 0)$ with which a compound shows a c_0 = inhibition (%) of c_1 = dopamine D_2 receptor higher than the expected. In addition, it is necessary to select other sets of values from the table to specify different conditions $\mathbf{c}_j$ such as organism of assay, organism of the protein, cell line, type of assay, *etc*. Please, note that this table is only a summary of some selected examples of conditions that can be predicted with the model. In the Supporting Information, the full table with all the possible conditions to be predicted with this model is presented. PTML model was used to predict the min–max scale probabilities, $*p(\Delta\varepsilon_{ij} > 0)$, for compound **18**. Thus, the current study predicted $*p(\Delta\varepsilon_{ij} > 0)$ for over 150 different assays(different conditions $\mathbf{c}_j$ = c_0, c_1, c_2, *etc*.) of compound **18**. Table 7 shows this type of results for selected examples involving different c_0 = activity parameter (units) such as efficacy (%),

inhibition (%), *etc*. The examples also involve different c_1 = target proteins including dopamine (D), norepinephrine (NE), and serotonin (5HTA) receptors (R) and transporters (T). Notably, the model allows us to make the inference for these values of $*p(\Delta\varepsilon_{ij} > 0)$ in different organisms including *H. sapiens* (humans), *R. norvegicus* (rat), and *M. musculus* (mouse). Notably, the values activity (%), efficacy (%), and inhibition (%) of compound **18** for receptors (D_1R to D_5R) are predicted to be higher than the expected value for these assays in humans with $*p(\Delta\varepsilon_{ij} > 0) > 0.5$. In addition, K_i (nM) values are predicted to have values notably lower than expected for the same receptors in humans with $*p(\Delta\varepsilon_{ij} > 0) < 0.25$. These results also support the selection of compound **18** as a potential lead compound for further discovery and optimization of PLG peptidomimetics derivatives.

Table 7. Prediction of Activity Profiles of Compound **18** for Dopamine Pathway Proteins[a] (Selected Examples)

Activity (Units)	Protein name	H. sapiens	R. norvegicus	M. musculus	B. taurus
Activity (%)	D1R	0.56	0.58	0.61	0.00
Activity (%)	D2R	0.59	0.00	0.00	0.00
Activity (%)	D3R	0.60	0.58	0.00	0.00
Activity (%)	D4R	0.57	0.00	0.00	0.00
Activity (%)	D5R	0.58	0.57	0.55	0.00
Activity (%)	DT	0.59	0.00	0.00	0.00
Agonist (%)	D3R	0.69	0.00	0.00	0.00
Agonist (%)	D4R	0.69	0.00	0.00	0.00
Efficacy (%)	D1R	0.63	0.00	0.00	0.00
Efficacy (%)	D2R	0.71	0.00	0.00	0.00
Efficacy (%)	D3R	0.63	0.63	0.00	0.00
Efficacy (%)	D4R	0.58	0.61	0.00	0.00
Inhibition (%)	D1R	0.58	0.59	0.00	0.00
Inhibition (%)	D2R	0.58	0.60	0.00	0.00
Inhibition (%)	D3R	0.58	0.55	0.59	0.51
Inhibition (%)	D4R	0.61	0.00	0.00	0.00
Inhibition (%)	DT	0.60	0.51	0.00	0.00
Intrinsic activity	D2R	0.68	0.62	0.00	0.00
Intrinsic activity (%)	D1R	0.62	0.60	0.00	0.00
Intrinsic activity (%)	D2R	0.63	0.58	0.00	0.00
Intrinsic activity (%)	D3R	0.65	0.00	0.00	0.00
Intrinsic activity (%)	D4R	0.77	0.00	0.00	0.00
Intrinsic activity (nM)	D1R	0.76	0.00	0.00	0.00
Intrinsic activity (nM)	D2R	0.00	0.63	0.00	0.00
Intrinsic activity (nM)	D4R	0.00	0.75	0.00	0.00
Ki (nM)	D/NER	0.57	0.00	0.00	0.00
Ki (nM)	D1R	0.17	0.13	0.00	0.00
Ki (nM)	D2R	0.19	0.17	0.17	0.18
Ki (nM)	D3R	0.19	0.17	0.00	0.00
Ki (nM)	D4R	0.19	0.12	0.00	0.00
Ki (nM)	D5R	0.11	0.00	0.00	0.00
Ki (nM)	DT	0.19	0.19	0.21	0.03
Selectivity index	D2R	0.47	0.00	0.00	0.00
Selectivity ratio	D/5HTAR	0.64	0.00	0.00	0.00
Selectivity ratio	NE/DT	0.50	0.45	0.00	0.00

[a] Protein names: dopamine (D), norepinephrine (NE), and serotonin (5HTA) receptors (R) and transporters (T).

CONCLUSIONS

We demonstrated that PTML big data analysis method is able to account for large data sets of preclinical assays reported in ChEMBL for compounds targeting dopamine pathway proteins. The principal relevance of this finding is not only in the large size of the data set but in the high diversity of types of assays, target proteins, cell lines, organisms of target, organisms of assays, *etc*. present in the data set studied. We included a comparative study to nonlinear ANN algorithms. ANN algorithms do not improve the results of the linear PTML model. We also presented for the first time a proof-of-concept example of the practical uses of this model. In so doing, six novel analogs of PLG modified in the l-

proline and l-leucine scaffolds have been synthesized and chemically characterized. They have been tested as mimetics of the neuropeptide PLG for their ability to potentiate the binding of the dopamine receptor agonist [3H]NPA to cloned human dopamine D_{2S} receptors. All the compounds show significant activity in enhancing the binding of the dopamine D_2 receptor agonist [3H]NPA. We used the PTML model to predict the results of multiple preclinical assays for the new compounds. In addition, a molecular docking study for some of these compounds was also reported as a complementary study.

METHODS

General Chemistry Procedures. All chemicals were of reagent grade, were obtained from Aldrich Chemical Co., and were used without further purification. All air sensitive reactions were carried out under argon. Flash chromatography was performed on flash silica gel SDS type 60 A.C.C., 35–70 μm, and analytical TLC was carried out on precoated silica gel plates (Merck 60 F_{254}, 0.2 mm). Melting points were measured on a Reichert Kofler Thermopan apparatus and are uncorrected. Na-D line polarimetry was carried out at 25 °C on a PerkinElmer 241 polarimeter. Infrared spectra were recorded on a JASCO FT/IR-4100 spectrophotometer. ^{1}H and ^{13}C NMR spectra were recorded on a Varian Mercury 300 or a Varian Inova 400 spectrometers at 300 or 400 and 75 or 100 MHz, respectively, using TMS as an internal standard (chemical shifts in δ values, *J* in hertz). Mass spectra were recorded on Micromass Autospec or Bruker Microtof spectrometers. Microanalyses were performed on a Thermo Finnigan Flash 1112.

Methyl d-[(3*R*,5*S*)-3,5-Bis(azidomethyl)-1-(*tert*-butoxycarbonyl)]prolyl-l-valyl-l-alanilate (IIa). The synthesis of this compound has been reported before.(49) To a solution of the dipeptide acid **Ia** (310.0 mg, 0.73 mmol) in CH_3CN anhydrous (18.6 mL) at 0 °C under argon was added *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium tetrafluoroborate (TBTU, 375.02 mg, 1.17 mmol). The resulting suspension was stirred for 30 min at 0 °C. l-Alanine methyl ester hydrochloride (122.27 mg, 0.88 mmol) in CH_3CN anhydrous (18.6 mL) and *N*,*N*-diisopropylethylamine (DIEA, 0.51 mL, 2.92 mmol) were added. The reaction mixture was stirred at 0 °C for 1 h and at room until complete consumption of the starting material (TLC; 21 h).The solvent was removed *in vacuo*, the residue was dissolved in EtOAc (100 mL), and this solution was washed with saturated aq. $NaHCO_3$ (2 × 50 mL) and brine (2 × 50 mL). The organic layer was dried over Na_2SO_4 and evaporated to dryness. The solid crude product was purified by flash column chromatography with hexane/EtOAc 1:2 as eluent to afford by concentration **IIa** (290 mg, 78%), white solid; mp = 161–163 °C. $[\alpha]^{25}_D$ −59.6 (*c* 0.97, MeOH). IR (solid) ν: 3283, 2977, 2933, 2101, 1739, 1693, 1639, 1539, 1450, 1379, 1365, 1267, 1217, 1164, 1137, 1056 cm^{-1}. ^{1}H NMR ($CDCl_3$, rotamers present) δ: 7.18, 6.34, and 6.15 (3 br s, 2H D_2O exchange, 2NH), 4.52–4.03 (4 m, 3H), 3.98–3.91 (m,1H), 3.72, 3.70, and 3.66 (3 s, 3H, CO_2CH_3), 3.49–3.46 (m, 2H), 3.41 (dd, *J* = 12.1, 2.3 Hz, 1H), 2.51–2.30 (2 m, 3H), 2.06–2.01, 1.86–1.79 and 1.74–1.61 (3 m, 2H), 1.43 and 1.34 (2 s, 9 H, 3 CH_3), 1.38 (d, *J* = 7.0 Hz, 3H, CH_{3Ala}), 0.94 and 0.88 (2 d, *J* = 7.1 Hz, 2 CH_{3Val}). ^{13}C NMR ($CDCl_3$, rotamers present) δ: 172.95 (C), 171.95 and 171.48 (C), 170.77 and 170.30 (C), 154.19 (C), 81.79 and 80.94 (C), 65.90 and 65.61 (CH), 58.44, 57.97, and 57.58 (CH), 54.30(CH_2), 53.76 and 53.52 (CH_2), 52.58 and 52.26 (CH_3), 48.39 and 48.12 (CH), 42.56 (CH), 41.16 (CH), 32.08, 30.95, and 29.71 (CH_2), 29.13 (CH), 28.29 (3 CH_3), 19.37 and 19.06 (CH_3), 18.27 and 18.12 (CH_3), 17.34 and 16.99 (CH_3). ESI-TOF HRMS, *m*/*z*: $[M]^+$, ($C_{21}H_{36}N_9O_6$) 510.2801, requires 510.2783.

d-[(3*S*,5*R*)-3,5-Bis(azidomethyl)-1-(tert-butoxycarbonyl)]prolyl-l-leucyl-l-alanine (IIIa). To a stirred solution of **IIa** (100 mg, 0.196 mmol) in THF (1 mL) at 0 °C was added dropwise a solution of LiOH (58.68 mg, 2.45 mmol) in THF/H_2O (1/1) (4 mL), and the reaction mixture was stirred for 0.5 h, warmed to room temperature, and stirred for a further stirred 52 h. The THF was removed, the residue was dissolved in H_2O (10 mL), and this solution was adjusted to pH 4 with 1 M H_2SO_4. The solution was concentrated to give a white solid, which was triturated with warm Et_2O/EtOH 1:3 (40 mL) to afford, after concentration, a white solid (91 mg, 94%); mp = 120–123 °C. $[\alpha]^{25}_D$–34.2 (*c* 0.81, MeOH). IR (solid) ν: 3301, 2972, 2933, 2099, 1651, 1610, 1524, 1454, 1393, 1367, 1256, 1141, 774 cm^{-1}. ^{1}H NMR ($CDCl_3$, rotamers present) δ: [11.04–11.02 and 10.80–10.78 (2 m), 7.13 (d, *J* = 7 Hz) and 7.02 (d, *J* = 7.4 Hz), 3H, D_2O exchange, CO_2H and 2 NH], 4.64–4.60 (m, 1H), 4.42–4.29 (m, 2H), 4.04–3.99 (m, 1H), 3.77–3.31 (m, 4H), 2.93–2.83 (m, 1H), 2.38–2.37 (m, 2H), 2.0–1.91 (m, 1H), 1.73–1.70 (m 1H), 1.42–

1.20 (m, 12H, 4 CH_3, 3 $CH_{3tert\text{-}butyl}$ and CH_{3Ala}), 0.94–0.85 (m, 6H, 2 CH_{3Val}). ^{13}C NMR ($CDCl_3$, rotamers present) δ: 177.76 and 172.67 (C), 172.78 (C), 171.89 (C), 155.36 (C), 81.91 and 81.85 (C), 64.10 (CH), 59.58 and 58.42 (CH), 54.11 (CH_2), 53.60 (CH_2), 50.61 (CH), 40.93 (CH), 31.80 (CH_2), 29.65 (CH), 28.17 (3 CH_3), 27.98 (CH), 21.78 (CH_3), 19.84 (CH_3), 18.87 (CH_3). ESI-TOF HRMS, *m*/*z*: $[M]^+$, ($C_{20}H_{34}N_9O_6$) 496.2631, requires 496.2627.

d-[(3*S*,5*R*)-3,5-Bis(azidomethyl)-1-(*tert*-butoxycarbonyl)]prolyl-l-valyl-l-alanylamide (**IVa**). Methyl ester **IIIa** (100 mg, 0.20 mmol) was treated with a concentrated solution of ammonia in MeOH (10 mL) at room temperature. The solution was stirred for 27 h, and the solvent and excess ammonia were removed *in vacuo*. The solid residue was successively dissolved in CH_2Cl_2 (3 × 10 mL) and then evaporated to dryness to afford **IVa** (90 mg, 90%) as a white solid; mp = 212–214 °C. $[\alpha]^{25}_D$– 38.2 (*c*0.82, MeOH). IR (solid) ν: 3262, 3202, 2973, 2099, 1678, 1636, 1534, 1451, 1365, 1255, 1164, 1146 cm^{-1}. ^1H NMR ($CDCl_3$, rotamers present) δ: [7.45 (d, *J* = 8.8 Hz, 1H, D_2O exchange), 6.71 (br s, 1H, D_2O exchange), 6.06 (d, *J* = 6.2 Hz, 1H, D_2O exchange), 5.12 (br s, 1H, D_2O exchange), 2 NH + NH_2], 4.61–4.53 (m, 1H), 4.18 (dd, *J* = 6.6, 3.6 Hz, 1H), 4.13–4.08 (m, 1H), 3.94 (d, *J* = 7.4 Hz, 1H), 3.86 (dd, *J* = 12.3, 4.8 Hz, 1H), 3.55–3.43 (m, 2H), 3.36 (dd, *J* = 12.3, 1.6 Hz, 1H), 2.63–2.50 (2 m, 2H), 2.34 (dt, *J* = 13.3, 8.5 Hz, 1H), 1.74–1.68 (m, 1H), 1.55 (s, 3H, CH_{3Ala}), 1.47, 1.45, and 1.44 (apparent 4s, 9H, 3CH_3), 1.04, 0.98, and 0.94 (3d, *J* = 7.0 Hz, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 175.75 (C), 173.09 and 172.48 (C), 172.04 (C), 155.16 (C), 80.65 and 80.41 (C), 65.67 and 65.49 (CH), 59.98, 59.50, 58.68, and 58.13 (CH), 58.68 and 58.13 (CH), 54.49 and 53.95 (CH_2), 53.02 (CH_2), 49.41 (CH), 43.04 (CH), 41.79 (CH), 32.33, 31.09, and 30.04 (CH_2), 31.85 and 31.45 (CH), 28.51, 28.56, and 20.40 (3 CH_3), 19.94 and 19.83(CH_3), 19.55 and 19.37 (CH_3), 18.86 and 18.77(CH_3). ESI-TOF HRMS, *m*/*z*: $[M]^+$, ($C_{20}H_{35}N_{10}O_5$) 495.2805, requires 495.2786.

Methyl d-[(3*R*,5*S*)-3,5-Bis(azidomethyl)]prolyl-l-valyl-l-alanilate Hydrochloride (**13**). Compound **IIa** (50.7 mg, 0.1 mmol) was dissolved in 4 N HCl in dioxane (2 mL), and the reaction mixture was stirred at room temperature for 6.5 h. The reaction mass was concentrated, and the residue was azeotroped with CH_2Cl_2 (5 × 5 mL) to afford **13** (41 mg, 93%), as a low melting point beige solid. $[\alpha]^{25}_D$ −45.3 (*c* 0.96, MeOH). IR (solid) ν: 2963, 2927, 2101, 1741, 1650, 1541, 1455, 1276, 1160, 1053 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 9.53–9.52, 9.40–9.38, 9.29–9.27 and 7.41 (3 br s, 4H, D_2O exchange, 2 NH, $^+NH_2$), 4.95–4.72 and 4.55–4.53 (2 m, 3H), 4.01–3.86 (m, 2H), 3.72–3.59 (m, 3H), 3.55 (s, 3H, CH_3), 2.83–2.81, 2.44–2.39 and 2.22–2.20 (3 m, 3H), 1.69–1.61 (m, 1H), 1.52, 1.36, and 1.29 (3d, *J* = 7.0, 7.1, and 7.0 Hz, 3H, CH_{3Ala}), 1.10–1.02 (m, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 173.40 (C), 171.33 and 171.19 (C), 171.10 (C), 66.84 (CH_2), 62.84 (CH), 59.0, 58.73 and 58.63 (2 CH), 53.86 and 52.88 (CH_2), 51.66 (CH_3), 48.40 (CH), 44.14 (CH), 43.52 (CH), 33.64 (CH_2), 31.51 (CH), 19.44 and 19.32 (CH_3), 18.32 and 18.21 (CH_3), 17.80 and 17.06 (CH_3). ESI-TOF HRMS, *m*/*z*: $[M]^+$, ($C_{16}H_{28}N_9O4$) 410.2263, requires 410.2259.

d-[(3*S*,5*R*)-3,5-Bis(azidomethyl)]prolyl-l-valyl-l-alanine Hydrochloride (**14**). The same procedure as used to make compound **13** was used for **14** from **IIIa** (50.3 mg, 0.10 mmol) to give **14** (43 mg, 93%) as a low melting point beige solid. $[\alpha]^{25}_D$ −18.2 (*c* 0.95, MeOH). IR (solid) ν: 3246, 2964, 2931, 2104, 1649, 1556, 1455, 1407, 1262, 1162 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 9.60–9.54 and 7.10 (m and br s, 5H, D_2O exchange, OH, 2 NH, $^+NH_2$), 4.95–4.87 and 4.74–4.72 (2 m, 3H), 4.04–3.89, 3.81–3.74 and 3.69–3.59 (3 m, 6H), 2.86–2.85, 2.53–2.48, 2.25–2.24 (3 m, 3H), 1.72–1.69 and 1.53–1.46 (2 m, 3H, CH_{3Ala}), 1.25–1.03 (m, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 175.89 (C), 171.88 (C), 171.25 (C), 63.00 (CH), 61.56 (CH), 59.50 and 59.52 (CH), 54.22 and 53.07 (CH_2), 49.29 (CH), 44.50 (CH), 33.94 and 29.98 (CH_2), 31.80 (CH), 19.76 (CH_3), 18.79 (CH_3), 18.02 (CH_3). ESI-TOF HRMS, *m*/*z*: $[M]^+$, ($C_{15}H_{26}N_9O_4$) 396.2111, requires 396.2102.

d-[(3*S*,5*R*)-3,5-Bis(azidomethyl)]prolyl-l-valyl-l-alanylamide Hydrochloride (**15**). The same procedure as used to make compound **13** was used for **15** from **IVa** (50 mg, 0.1 mmol) to give **15** (42 mg, 95%) as a low melting point beige solid. $[\alpha]^{25}_D$ −36.2 (*c* 0.96, MeOH). IR (solid) ν: 3243, 2963, 2928, 2102, 1653, 1544, 1453, 1373, 1268, 1121 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 9.69 and 9.52 (2 d, *J* = 6.8 and 8.8 Hz, respectively, 2H, 2NH), 8.71 and 8.10 (2 br s, 2H, D_2O exchange, NH_2), 7.15 (br s, 2H, D_2O exchange, $^+NH_2$), 5.22–5.20 (m, 1H), 4.97–4.94 (m, 1H), 4.20–4.15 (m, 1H), 4.01–3.82 (m,

3H), 3.70–3.64 (m, 2H), 2.92–2.89 (m, 1H), 2.43–2.36 (m, 2H), 1.85–1.83 (m, 1H), 1.51 (d, J = 5.9 Hz, 3H, CH_{3Ala}), 1.11 (dd, J = 22.4, 6.8 Hz, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 175.75 (C), 171.53 (C), 170.84 (C), 62.52 and 61.57 (CH), 61.59 (CH_2), 60.11 and 59.45 CH), 54.07 and 52.42 (CH_2), 49.47 (CH), 44.70 and 43.87 (CH), 33.85 and 29.98 (CH_2), 32.10 (CH), 19.73 (CH_3), 19.13 (CH_3), 18.70 (CH_3). ESI-TOF HRMS, m/z: $[M]^+$, ($C_{15}H_{27}N_{10}O_3$) 395.2269, requires 395.2262.

Methyl l-[(3*R*,5*S*)-3,5-Bis(azidomethyl)-1-(tert-butoxycarbonyl)]prolyl-l-valyl-l-alanilate (IIb). The same procedure as used to make compound **IIa** was used for **IIb** from **Ib** (110 mg, 0.26 mmol) to give **IIb** (132 mg, 61%) as a white solid, mp = 104–106 °C. $[\alpha]^{25}_D$ −11.9 (c 92, MeOH). IR (solid) ν: 3296, 2977, 2099, 1745, 1698, 1644, 1542, 1383, 1366, 1256, 1210, 1164, 1137, 1055 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 7.39, 6.50, 6.39, and 6.30 (4 br s, 2H D_2O exchange, 2 NH), 4.55–4.50 and 4.43–4.39 (2 m, 1H), 4.21–4.19 and 4.08–3.96 (2 m, 3H), 3.71, 3.70, and 3.69 (3 s, 3H, CO_2CH_3), 3.52–3.39 (m, 3H), 2.56–2.19, 2.13–2.09, 1.90–1.83 and 1.72–1.63 (4 m, 5H), 1.47–1.32 (m, 3H, CH_{3Ala}), 1.42, 1.41, 1.39, 1.38, and 1.36 (5s, 9H, 3 CH_3), 0.95–0.86 (m, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 173.08 (C), 172.06 and 171.59 (C), 170.89 and 170.40 (C), 154.22 (C), 81.60 and 81.07 (C), 65.85, 65.48, and 65.30 (CH), 58.52, 57.83, 57.69, and 57.33 (CH), 54.31 and 53.99 (CH_2), 53.76 and 53.33 (CH_2), 52.5 and 52.33 (CH_3), 48.41 and 48.09 (CH), 42.22 (CH), 41.28 and 40.74 (CH), 32.19, 30.76, and 29.63 (CH_2), 31.22 (CH), 28.84 and 28.27 (3 CH_3), 19.43, 19.11, 18.07, 17.70, 17.31, and 16.54 (3CH_3). ESI-TOF HRMS, m/z: $[M]^+$, ($C_{21}H_{36}N_9O_6$) 510.2800, requires 510.2783.

l-[(3*S*,5*R*)-3,5-Bis(azidomethyl)-1-(*tert*-butoxycarbonyl)]prolyl-l-valyl-l-alanylamide (IVb). The same procedure as used to make compound **IVa** was used for **IVb** from **IIb** (60 mg, 0.12 mmol) to give **IVb** (58 mg, 97%) as white solid, mp = 148–152 °C. $[\alpha]^{25}_D$ −5.8 (c 0.96, MeOH). IR ν: 3326, 3210, 2979, 2098, 2100, 1682, 1648, 1541, 1456, 1340 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 9.56–9.53, 9.38–9.32, 9.21–9.12, 9.05–9.02, 8.35–8.24, 8.22–8.20, 8.10–8.03 and 7.90–7.88 (8 m, 4H D_2O exchange, 2 NH + NH_2), 5.11–5.06 (m, 1H), 4.71–4.66, 4.61–4.55, 4.41–4.36, 4.32–4.30 and 4.16–4.01 (5 m, 3H), 3.76–3.73 and 3.61–3.51 (2 m, 3H), 2.75–2.64 (m, 1H), 2.57–2.34 (m, 2H), 1.83–1.77 (m, 1H), 1.63 and 1.61 (2s, 9H, 3 CH_3), 1.55–1.51 (m, 3H, CH_{3Ala}), 1.33–1.28 (m, 1H), 1.17 and 1.14–1.11 (d, J = 6.9 Hz and m, 6H, 2 CH_{3Val}). ESI-TOF HRMS, m/z: $[M + Na]^+$, ($C_{20}H_{34}N_{10}NaO_5$) 517.2620, requires 517.2606.

Methyl l-[(3*R*,5*S*)-3,5-Bis(azidomethyl)]prolyl-l-valyl-l-alanilate Hydrochloride (16). The same procedure as used to make compound **13** was used for **16** from **IIb** (50 mg, 0.10 mmol) to give **16** (40 mg, 91%) as a low melting point beige solid. $[\alpha]^{25}_D$ −17.0(c 0.81, MeOH). IR ν: 3052, 2961, 2927, 2103, 1737, 1648, 1545, 1453, 1267, 1211, 1161, 1055 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 9.64–9.58, 9.50–9.38, 9.37–9.35 and 9.20–9.19 (4 m, 2H, D_2O exchange, 2 NH), 7.40 (br s, 2H, D_2O exchange, $^+NH_2$), 4.96–4.84 (2 m, 1 H), 4.75–4.63 (m, 1H), 4.02–3.89 (m, 3H), 3.81–3.58 (m, 3H), 3.61 and 3.54 (2s, 3H, CH_3), 2.91–2.87 (m, 1H), 2.51–2.49 and 2.38–2.26 (2 m, 2H), 1.73–1.68 (m, 1H), 1.63–1.57 (m, 1H), 1.47 and 1.42 (2 d, J = 7.2 and 7.1 Hz, respectively, 3H, CH_{3Ala}), 1.16–0.94 (m, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 173.83 (C), 171.84 (C), 170.77 (C), 63.07 and 62.97 (CH), 61.59 (CH_2), 59.73, 59.52, 59.31, and 59.24 (CH), 53.96, 53.62, and 52.97 (CH_2), 52.05 and 51.96 (CH_3), 48.93 and 48.74 (CH), 44.75, 44.26, and 43.87 (CH), 34.02, and 29.99 (CH_2), 33.65 (CH), 31.75 and 31.71 (CH), 19.73 and 19.59 (CH_3), 18.63 and 18.47 (CH_3), 17.58 and 17.45 (CH_3). ESI-TOF HRMS, m/z: $[M]^+$, ($C_{16}H_{28}N_9O_4$) 410.2263, requires 410.2259.

l-[(3*S*,5*R*)-3,5-Bis(azidomethyl)]prolyl-l-valyl-l-alanine Hydrochloride (17). To a stirred solution of **IIb** (60 mg, 0.12 mmol) in THF (1 mL) at 0 °C was added dropwise a solution of LiOH (36 mg, 1.5 mmol) in THF/H_2O (1/1) (3 mL), and the resultant mixture was stirred for 0.5 h at this temperature and then 18 h at room temperature. The THF was removed, the residue was dissolved in H_2O (7.5 mL), and the resulting aqueous solution was adjusted to pH 4 by the addition of 1 M H_2SO_4. White solid obtained after removal of H_2O to dryness was extracted with a mixture of Et_2O/EtOH (1:3) (2 × 15 mL) to afford, after elimination of solvent to dryness, a white solid (0.39 mg) of mp = 173–177 °C, which without further purification was dissolved in 4 N HCl in dioxane (2 mL), and the reaction mixture was stirred at room temperature for 6.5 h. The elimination of solvents to vacuum through coevaporation with CH_2Cl_2 (5 × 5 mL) led to **17** (42.5 mg, 84%) as a solid beige of low melting point. $[\alpha]^{25}_D$ +7.2 (c 0.79, MeOH). IR ν:

3382, 2104, 1655, 1543, 1456, 1270, 1049 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 9.72–9.70, 9.68–9.64, 9.47–9.45 and 9.34–9.31 (4 m, 2H D_2O exchange, 2 NH), 6.61 (br s, 3H D_2O exchange, $^+NH_2$ and OH), 4.98–4.87 and 4.70–4.63 (2 m, 2H), 3.95–3.93, 3.88–3.85 and 3.73–3.57 (3 m, 8H), 2.86–2.85, 2.52–2.35 and 2.21–2.17 (4m, 2H), 1.66–1.56 and 1.53–1.52 (2 m, 3H, CH_{3Ala}), 1.20–1.03, 0.98, and 0.92 (m and 2 d, J = 6.6 Hz, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 175.84 (C), 171.87 (C), 170.91 (C), 62.94 (CH), 61.57 (CH_2), 59.69 and 59.55 (CH), 53.81 and 53.07 (CH_2), 49.19 (CH), 44.41 (CH), 43.90 (CH), 34.05 and 29.97 (CH_2), 31.68 (CH), 190.80 and 19.70 (CH_3), 18.79 (CH_3), 18.13 (CH_3). ESI-TOF HRMS, m/z: $[M]^+$, ($C_{15}H_{26}N_9O_4$) 396.2096, requires 396.2102.

l-[(3*S*,5*R*)-3,5-Bis(azidomethyl)]prolyl-l-valyl-l-alanylamide Hydrochloride (**18**). The same procedure as used to make compound **13** was used for **18** from **IVb** (41 mg, 0.08 mmol) to give **18** (33 mg, 92%) as a low melting beige solid. $[\alpha]^{25}_D$ +17.8 (*c* 0.81, MeOH). IR ν: 3279, 2964, 2928, 2102, 1672, 1648, 1635, 1542, 1454, 1373, 1265, 1121 cm^{-1}. ^1H NMR (pyridine-d_5, rotamers present) δ: 9.47, 9,35, 8.40, and 8.02 (2 d, J = 8.8 and 7.5 Hz, respectively, and 2 br s, 4H, D_2O exchange, 2 NH and NH_2), 6.95 (br s, 2H, $^+NH_2$), 5.07, 4.89–4.85, 4.59–4.57 (t, J = 7.1 Hz and 2 m, 3H), 3.94–3.83 (m, 3H), 3.72–3.51 (m, 3H), 2.86–2.79 (m, 1H), 2.44–2.20 (3 m, 2H), 1.66–1.63 and 1.55–1.51 (2 m, 3H, CH_{3Ala}), 1.22–1.17, 1.03, and 0.99 (m and 2 dd, J = 14.8, 6.6 Hz and 25, 6.5 Hz, respectively, 6H, 2 CH_{3Val}). ^{13}C NMR (pyridine-d_5, rotamers present) δ: 175.54 (C), 171.58 (C), 171.07 (C), 62.97 (CH), 61.60 (CH_2), 59.73 and 59.38 (CH), 54.06, 53.77, and 53.04 (CH_2), 49.43 (CH), 44.73 (CH), 43.88 (CH), 34.04 and 30.0 (CH_2), 31.81 (CH), 19.71 (CH_3), 18.95 and 18.82 (CH_3), 18.72 and 18.61 (CH_3). ESI-TOF HRMS, m/z: $[M]^+$, ($C_{15}H_{27}N_{10}O_3$) 395.2251, requires 395.2262.

Pharmacological Assay of New PLG Peptidomimetics. Eleven novel PLG mimetics based on a 3,5-disubstituted proline scaffold were tested for their ability to potentiate the binding of the dopamine receptor agonist *N*-propylnorapomorphine ([3H]NPA) to cloned human dopamine D_{2S} receptors, as described by Verma *et al.*(50) These compounds were tested for their ability to increase [3H]NPA binding at eight different concentrations in the range between 1 and 10 mM. The data obtained for PLG analog photoaffinity labeling agents **13**–**21** are shown, and their activity was compared to that of PLG. All of the compounds significantly ($P < 0.05$, ANOVA test, post hoc Dunnet T3 test) increased [3H]NPA binding at low concentrations. Compound **13** increased [3H]NPA binding more than that observed with PLG at the same concentrations. Both compounds **17** and **18** increased [3H]NPA binding: **17** showed 40% increased [3H]NPA binding at the concentration 10^{-7} M, and **18** caused increased [3H]NPA binding higher than 40% in the range of concentrations 10^{-12}–10^{-7} M, showing a better profile than PLG. The effect observed with PLG in our work is similar to that observed in our previous work(49) and slightly different from that previously reported by Verma *et al.*,(50) showing increased [3H]NPA binding at lower concentrations than those previously reported. This difference can be explained by the different host cell where human D_{2S} receptors were expressed as it has been previously reported that allosteric modulators are sensitive to environmental changes, which may condition the different active conformations elicited by the endogenous agonists on GPCRs.(57) Most of the compounds showed a bell-shaped curve, which is compatible with previous findings from *in vivo* and clinical experiments using PLG.(50) These data evidence that these compounds are acting as allosteric modulators of dopamine D_{2S} receptors. To view the profiles observed in the series of analogs of PLG tested, we can conclude that the change affects the chain of leucine. This happen in both series of compounds; however, the allosteric modulator effect is lower in the series of compounds **19**–**21**, counterparts in which the leucine is replaced by valine and the amine group of the proline cycle is protected as carbamate, than in the series of compounds **13**–**18**, where the glycine is replaced by alanine, which showed most similar profiles to PLG in the increase of the union of [3H]NPA to dopamine D_2 receptors.

Docking Simulation. Molecular docking techniques were used as a complementary tool useful to measure the DPI variable L_{ij}. A transformation similar to previous publications has been applied: $L_{ij} = 1$ when the deviation of AE_{ij} from the expected value is $\Delta AE_{ij} > 0$. This deviation was calculated as $\Delta AE_{ij} = AE_{ij} - \langle AE_{ij} \rangle$. The expected value of the parameter is equal to the average value $\langle AE_{ij} \rangle$ for all the calculated peptides. The strength of the interactions was quantified by the affinity energy AE_{ij} (kcal/mol) of ligands with protein targets using the free software AutoDock Vina.(58) The entire processing was done in the BioCAI cluster at University of A Coruna (Spain). The docking flow has several steps that

include the ligand and protein processing, molecule format conversion, and geometry optimization before the docking calculations. Thus, the ligands are presented as a list of SMILES formulas, and they are converted into PDB with optimized 3D structure using the free tool Babel.(59) The protein targets are filtered for only the first PDB model, and the nonprotein part is eliminated (water molecules, other ligands, etc.). The PDB of ligands and proteins are converted into PDBQT format using AutoDock tools scripts (prepare_ligand4.py and prepare_receptor4.py).(60) The protein target is considered rigid in all docking calculations, and the interaction search considers the entire surface of the targets. The docking flow is based on python and bash scripts, including the reading of the final results. The cutoff for stable interactions is considered AE < −7.0 kcal/mol.(61) The results are based on the first docking conformer of the ligands with reference root-mean-square deviation of atomic positions RMSD = 0.(62) Authors will release the atomic coordinates and experimental data upon article publication.

ChEMBL Data Preprocessing. In this work, the values of the experimental parameter ε_{ij} that characterizes the effect of the *i*th molecule (m_i) over the *j*th target have been downloaded from ChEMBL. This parameter depends on a series of experimental conditions $\mathbf{c}_j = (c_0, c_1, c_2, ..., c_n)$. The conditions could be, for example, c_0 = type of biological activity parameter ε_{ij} measured (IC_{50}, K_i, K_m, *etc*.), c_1 = *j*th target protein, and c_2 = organism of assay (*H. sapiens*, *R. norvegicus*, *etc*.). Unfortunately, the values ε_{ij} reported in ChEMBL are not exact numbers in many cases: there are several examples such as $K_i > 1000$ or $K_i < 1000$ instead of $K_i = 1000$ (the exact value). On the other hand, the experimental values differ for IC_{50}, K_i, K_m, *etc*. Therefore, it is better to develop a classification model than a regression model for the DPI. Consequently, the values ε_{ij} downloaded from ChEMBL have been preprocessed in order to transform them into the discrete variable L_{ij}. The transformation used the following rule: $L_{ij} = 1$ when the deviation of ε_{ij} from the expected value is $\Delta\varepsilon_{ij} > 0$ (experimental perturbation with respect to the expected value). This deviation was calculated as $\Delta\varepsilon_{ij} = \varepsilon_{ij} - \langle\varepsilon_{ij}\rangle$.The expected value of the parameter is equal to the average value $\langle\varepsilon_{ij}\rangle$ for all the drugs with the same *j*th experimental parameter. In this context, the DPI variable L_{ij} is a discrete variable. The value $L_{ij} = 1$ is indicative of a DPI stronger than the expected, and $L_{ij} = 0$ may be indicative of a DPI weak or lower than expected.

PTML Model of ChEMBL Data Set. González-Díaz *et al.* introduced a general-purpose PTML modeling technique useful to quantify the effect of perturbations in complex biomolecular systems.(63,64) Using a PTML model, it is possible to predict the values of the scoring function $f(\varepsilon_{ij})_{new}$ for the *i*th compound in the *j*th preclinical assay. The PTML model uses as input the expected value of biological activity, $f(\varepsilon_{ij})_{expt}$ for one compound assayed in the conditions $\mathbf{c}_j$ and the values of the PT operators $\Delta D_k(m_i, \mathbf{c}_j)$. The expected value $f(\varepsilon_{ij})_{expt} = \langle\varepsilon_{ij}\rangle$ is the average value of the biological activity parameter ε_{ij} for all cases in ChEMBL data set with the same c_0 = activity parameter ε_{ij}(units). These PT operators $\Delta D_k(m_i, \mathbf{c}_j) = D_k(m_j) - \langle D_k(\mathbf{c}_j)\rangle$ are intended to account for the changes (perturbations) in the system with respect to the expected values. Specifically, perturbations on the value of the molecular descriptors of the drug, $D_k(m_j)$, with respect to the expected value, $\langle D_k(\mathbf{c}_j)\rangle$, for a drug measured under the conditions of the experiment $\mathbf{c}_j$. These PT operators resemble the Box–Jenkins MA operators.(18,28) We used both linear and nonlinear ML algorithms called linear discriminant analysis (LDA) and artificial neural network (ANN) to seek alternative linear and nonlinear models.(52) The compact and developed forms of a PTML linear model are given as follows.

$$f(\varepsilon_{ij})_{\text{new}} = a_0 \cdot f(\varepsilon_{ij})_{\text{expt}} + \sum_{k=1}^{k_{max}} \sum_{j=0}^{j_{max}} a_{jk} \cdot \Delta D_k(m_i, c_j) + e_0 \tag{2}$$

$$f(\varepsilon_{ij})_{\text{new}} = a_0 \cdot f(\varepsilon_{ij})_{\text{expt}} + \sum_{k=1}^{k_{max}} \sum_{j=0}^{j_{max}} a_{jk} \cdot (D_k(m_i)_{\text{new}} - \langle D_k(c_j)\rangle_{\text{ref}}) + e_0 \tag{3}$$

In addition, several ML methods from Weka(53) have been used: Bayes Nets(54), RandomTree(55), Random Forest, and Decision Table.(56) The Weka parameters are the default ones. The last classifier tested was a Deep Learning model using the following parameters: a single hidden layer with 1000 neurons with ReLU activation functions, batch normalization, Adam optimizer, accuracy metrics, binary cross entropy loose function, and 5000 epochs for training using 2048 batch size.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschemneuro.8b00083.

NMR and IR characterization of new compounds (PDF)
Details of the results of the PTML model (XLSX)
Expected values for all the molecular descriptors used in the calculation of the moving averages (PDF)

AUTHOR INFORMATION

Author Contributions

J.F.C., D.S., O.C., and X.G.M. performed the organic synthesis and characterization of new peptidomiemtic compounds. J.M.B. and M.I.L. carried out the pharmacological assay of peptidomimetic compounds. A.P and C.R.M. carried out the molecular docking and nonlinear ML studies. H.G.D. preprocessed the ChEMBL data set and developed the linear PTML model. All authors contributed to the writing of their respective sections in the paper.

Funding

Grants from Ministry of Economy and Competitiveness, MINECO, Spain (FEDER CTQ2016-74881-P), Basque government (IT1045-16), Xunta de Galicia (GPC2014/003, CN 2012/069, ED431D 2017/16, ED431D 2017/23, and GRC2014/049), and European Regional Development Fund (ERDF) are gratefully acknowledged for their financial support. In addition, the authors would like to acknowledge support from the Galician Network for R&D of Medicines (REGID, the General Directorate of Culture, Education and University Management of Xunta de Galicia and the Galician Network for Colorectal Cancer Research (REGICC) (Ref. ED431D 2017/23).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Grants from Ministry of Economy and Competitiveness, MINECO, Spain (FEDER CTQ2016-74881-P), Basque government (IT1045-16), and Xunta de Galicia (GPC2014/003, CN 2012/069, and GRC2014/049) are gratefully acknowledged for their financial support. In addition, the authors would like to acknowledge support from the Galician Network for R&D of Medicines (REGID, the General Directorate of Culture, Education and University Management of Xunta de Galicia and the Galician Network for Colorectal Cancer Research (REGICC) (Ref. ED431D 2017/23). This work is also supported by “Collaborative Project in Genomic Data Integration (CICLOGEN)” PI17/01826 funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013–2016 and the European Regional Development Funds (FEDER). This project was also supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia ED431D 2017/16 and "Drug Discovery Galician Network" Ref. ED431G/01 and the “Galician Network for

Colorectal Cancer Research" (Ref. ED431D 2017/23) and finally by the Spanish Ministry of Economy and Competitiveness for its support with the funding of the unique installation BIOCAI (UNLC08-1E-002, UNLC13-13-3503) and the European Regional Development Funds (FEDER) by the European Union.

REFERENCES.

(1) Vilar, S., Santana, L., and Uriarte, E. (2006) Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. *J. Med. Chem. 49*, 1118– 1124, DOI: 10.1021/jm050932j

(2) Santana, L., Uriarte, E., Gonzalez-Diaz, H., Zagotto, G., Soto-Otero, R., and Mendez-Alvarez, E. (2006) A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. *J. Med. Chem. 49*, 1149– 56, DOI: 10.1021/jm0509849

(3) Santana, L., Gonzalez-Diaz, H., Quezada, E., Uriarte, E., Yanez, M., Vina, D., and Orallo, F. (2008) Quantitative structure-activity relationship and complex network approach to monoamine oxidase A and B inhibitors. *J. Med. Chem. 51*, 6740– 51, DOI: 10.1021/jm800656v

(4) Li, C., Xu, L., Wolan, D. W., Wilson, I. A., and Olson, A. J. (2004) Virtual screening of human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase against the NCI diversity set by use of AutoDock to identify novel nonfolate inhibitors. *J. Med. Chem. 47*, 6681– 90, DOI: 10.1021/jm049504o

(5) Ongarora, B. G., Fontenot, K. R., Hu, X., Sehgal, I., Satyanarayana-Jois, S. D., and Vicente, M. G. (2012) Phthalocyanine-peptide conjugates for epidermal growth factor receptor targeting. *J. Med. Chem. 55*, 3725– 38, DOI: 10.1021/jm201544y

(6) Han, L., Cui, J., Lin, H., Ji, Z., Cao, Z., Li, Y., and Chen, Y. (2006) Recent progresses in the application of machine learning approach for predicting protein functional class independent of sequence similarity. *Proteomics 6*, 4023– 37, DOI: 10.1002/pmic.200500938

(7) Chou, K. C. and Shen, H. B. (2010) Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One 5*, e11335, DOI: 10.1371/journal.pone.0011335

(8) Xiao, X., Wu, Z. C., and Chou, K. C. (2011) A multi-label classifier for predicting the subcellular localization of gram-negative bacterial proteins with both single and multiple sites. *PLoS One 6* (6), e20592, DOI: 10.1371/journal.pone.0020592

(9) Shen, H. B. and Chou, K. C. (2009) QuatIdent: a web server for identifying protein quaternary structural attribute by fusing functional domain and sequential evolution information. *J. Proteome Res. 8*, 1577– 84, DOI: 10.1021/pr800957q

(10) Chou, K. C. and Shen, H. B. (2006) Large-scale predictions of gram-negative bacterial protein subcellular locations. *J. Proteome Res. 5*, 3420– 8, DOI: 10.1021/pr060404b

(11) Rodriguez-Soca, Y., Munteanu, C. R., Dorado, J., Pazos, A., Prado-Prado, F. J., and Gonzalez-Diaz, H. (2010) Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. *J. Proteome Res. 9*, 1182– 90, DOI: 10.1021/pr900827b

(12) Munteanu, C. R., Vazquez, J. M., Dorado, J., Sierra, A. P., Sanchez-Gonzalez, A., Prado-Prado, F. J., and Gonzalez-Diaz, H. (2009) Complex network spectral moments for ATCUN motif DNA cleavage: first predictive study on proteins of human pathogen parasites. *J. Proteome Res. 8*, 5219– 28, DOI: 10.1021/pr900556g

(13) Gonzalez-Diaz, H., Saiz-Urra, L., Molina, R., Santana, L., and Uriarte, E. (2007) A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. *J. Proteome Res. 6*, 904– 8, DOI: 10.1021/pr060493s

(14) Gonzalez-Diaz, H., Prado-Prado, F., Garcia-Mera, X., Alonso, N., Abeijon, P., Caamano, O., Yanez, M., Munteanu, C. R., Pazos, A., Dea-Ayuela, M. A., Gomez-Munoz, M. T., Garijo, M. M., Sansano, J., and Ubeira, F. M. (2011) MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from Trichomonas gallinae. *J. Proteome Res. 10*, 1698– 718, DOI: 10.1021/pr101009e

(15) Concu, R., Dea-Ayuela, M. A., Perez-Montoto, L. G., Bolas-Fernandez, F., Prado-Prado, F. J., Podda, G., Uriarte, E., Ubeira, F. M., and Gonzalez-Diaz, H. (2009) 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*, 4372– 82, DOI: 10.1021/pr9003163

(16) Agüero-Chapin, G., Varona-Santos, J., de la Riva, G. A., Antunes, A., González-Villa, T., Uriarte, E., and Gonzalez-Diaz, H. (2009) Alignment-free prediction of polygalacturonases with

pseudofolding topological indices: experimental isolation from Coffea arabica and prediction of a new sequence. *J. Proteome Res. 8*, 2122– 8, DOI: 10.1021/pr800867y

(17) Martínez-Arzate, S. G., Tenorio-Borroto, E., Barbabosa Pliego, A., Diaz-Albiter, H., Vazquez-Chagoyan, J. C., and Gonzalez-Diaz, H. (2017) PTML Model for Proteome Mining of B-cell Epitopes and Theoretic-Experimental Study of Bm86 Protein Sequences from Colima Mexico. *J. Proteome Res. 16*, 4093, DOI: 10.1021/acs.jproteome.7b00477

(18) González-Díaz, H., Molina-Ruiz, R., and Hernandez, I. (2007) MARCH-INSIDE version 3.0 (MARkov CHains INvariants for SImulation & DEsign); Windows supported version under request to the main author, contact e-mail: gonzalezdiazh@yahoo.es.

(19) Fernández, L., Caballero, J., Abreu, J. I., and Fernández, M. (2007) Amino Acid Sequence Autocorrelation Vectors and Bayesian-Regularized Genetic Neural Networks for Modeling Protein Conformational Stability: Gene V Protein Mutants. *Proteins: Struct., Funct., Genet. 67*, 834– 852, DOI: 10.1002/prot.21349

(20) Caballero, J., Fernandez, L., Abreu, J. I., and Fernandez, M. (2006) Amino Acid Sequence Autocorrelation vectors and ensembles of Bayesian-Regularized Genetic Neural Networks for prediction of conformational stability of human lysozyme mutants. *J. Chem. Inf. Model. 46*, 1255– 68, DOI: 10.1021/ci050507z

(21) Perez-Riverol, Y., Audain, E., Millan, A., Ramos, Y., Sanchez, A., Vizcaino, J. A., Wang, R., Muller, M., Machado, Y. J., Betancourt, L. H., Gonzalez, L. J., Padron, G., and Besada, V. (2012) Isoelectric point optimization using peptide descriptors and support vector machines. *J. Proteomics 75*, 2269– 74, DOI: 10.1016/j.jprot.2012.01.029

(22) Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrian-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magarinos, M. P., Overington, J. P., Papadatos, G., Smit, I., and Leach, A. R. (2017) The ChEMBL database in 2017. *Nucleic Acids Res. 45*, D945– D954, DOI: 10.1093/nar/gkw1074

(23) Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overington, J. P. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res. 40*, D1100– 7, DOI: 10.1093/nar/gkr777

(24) Svensson, F., Bender, A., and Bailey, D. (2018) Fragment-Based Drug Discovery of Phosphodiesterase Inhibitors. *J. Med. Chem. 61*, 1415, DOI: 10.1021/acs.jmedchem.7b00404

(25) Stumpfe, D., Dimova, D., and Bajorath, J. (2016) Computational Method for the Systematic Identification of Analog Series and Key Compounds Representing Series and Their Biological Activity Profiles. *J. Med. Chem. 59*, 7667– 76, DOI: 10.1021/acs.jmedchem.6b00906

(26) Jansen, C., Kooistra, A. J., Kanev, G. K., Leurs, R., de Esch, I. J., and de Graaf, C. (2016) PDEStrIAn: A Phosphodiesterase Structure and Ligand Interaction Annotated Database As a Tool for Structure-Based Drug Design. *J. Med. Chem. 59*, 7029– 65, DOI: 10.1021/acs.jmedchem.5b01813

(27) Gonzalez-Diaz, H., Arrasate, S., Gomez-SanJuan, A., Sotomayor, N., Lete, E., Besada-Porto, L., and Ruso, J. M. (2013) General Theory for Multiple Input-Output Perturbations in Complex Molecular Systems. 1. Linear QSPR Electronegativity Models in Physical, Organic, and Medicinal Chemistry. *Curr. Top. Med. Chem. 13*, 1713– 1741, DOI: 10.2174/1568026611313140011

(28) Blazquez-Barbadillo, C., Aranzamendi, E., Coya, E., Lete, E., Sotomayor, N., and Gonzalez-Diaz, H. (2016) Perturbation theory model of reactivity and enantioselectivity of palladium-catalyzed Heck-Heck cascade reactions. *RSC Adv. 6*, 38602– 38610, DOI: 10.1039/C6RA08751E

(29) Casanola-Martin, G. M., Le-Thi-Thu, H., Perez-Gimenez, F., Marrero-Ponce, Y., Merino-Sanjuan, M., Abad, C., and Gonzalez-Diaz, H. (2016) Multi-output Model with Box-Jenkins Operators of Quadratic Indices for Prediction of Malaria and Cancer Inhibitors Targeting Ubiquitin-Proteasome Pathway (UPP) Proteins. *Curr. Protein Pept. Sci. 17*, 220– 227, DOI: 10.2174/1389203717999160226173500

(30) Romero-Duran, F. J., Alonso, N., Yanez, M., Caamano, O., Garcia-Mera, X., and Gonzalez-Diaz, H. (2016) Brain-inspired cheminformatics of drug-target brain interactome, synthesis, and assay of TVP1022 derivatives. *Neuropharmacology 103*, 270– 278, DOI: 10.1016/j.neuropharm.2015.12.019

(31) Kleandrova, V. V., Luan, F., Gonzalez-Diaz, H., Ruso, J. M., Speck-Planche, A., and Cordeiro, M. N. D. S. (2014) Computational Tool for Risk Assessment of Nanomaterials: Novel QSTR-Perturbation Model for Simultaneous Prediction of Ecotoxicity and Cytotoxicity of Uncoated and Coated Nanoparticles under Multiple Experimental Conditions. *Environ. Sci. Technol. 48*, 14686– 14694, DOI: 10.1021/es503861x

(32) Luan, F., Kleandrova, V. V., Gonzalez-Diaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., and Cordeiro, M. N. (2014) Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale 6*, 10623– 30, DOI: 10.1039/C4NR01285B
(33) Alonso, N., Caamano, O., Romero-Duran, F. J., Luan, F., D. S. Cordeiro, M. N., Yanez, M., Gonzalez-Diaz, H., and Garcia-Mera, X. (2013) Model for High-Throughput Screening of Multitarget Drugs in Chemical Neurosciences: Synthesis, Assay, and Theoretic Study of Rasagiline Carbamates. *ACS Chem. Neurosci. 4*, 1393– 1403, DOI: 10.1021/cn400111n
(34) Box, G. E. P., and Jenkins, G. M. (1970) *Time series analysis*, p 553, Holden-Day.
(35) Speck-Planche, A. and Dias Soeiro Cordeiro, M. N. (2017) Speeding up Early Drug Discovery in Antiviral Research: A Fragment-Based in Silico Approach for the Design of Virtual Anti-Hepatitis C Leads. *ACS Comb. Sci. 19*, 501– 512, DOI: 10.1021/acscombsci.7b00039
(36) Kleandrova, V. V., Ruso, J. M., Speck-Planche, A., and Dias Soeiro Cordeiro, M. N. (2016) Enabling the Discovery and Virtual Screening of Potent and Safe Antimicrobial Peptides. Simultaneous Prediction of Antibacterial Activity and Cytotoxicity. *ACS Comb. Sci. 18*, 490– 8, DOI: 10.1021/acscombsci.6b00063
(37) Speck-Planche, A. and Cordeiro, M. N. (2015) Computer-aided discovery in antimicrobial research: In silico model for virtual screening of potent and safe anti-pseudomonas agents. *Comb. Chem. High Throughput Screening 18*, 305– 14, DOI: 10.2174/1386207318666150305144249
(38) Speck-Planche, A. and Cordeiro, M. N. (2014) Simultaneous virtual prediction of anti-Escherichia coli activities and ADMET profiles: A chemoinformatic complementary approach for high-throughput screening. *ACS Comb. Sci. 16*, 78– 84, DOI: 10.1021/co400115s
(39) Sizonenko, S. V., Sirimanne, E. S., Williams, C. E., and Gluckman, P. D. (2001) Neuroprotective effects of the N-terminal tripeptide of IGF-1, glycine-proline-glutamate, in the immature rat brain after hypoxic-ischemic injury. *Brain Res. 922*, 42– 50, DOI: 10.1016/S0006-8993(01)03148-1
(40) Guan, J., Krishnamurthi, R., Waldvogel, H. J., Faull, R. L., Clark, R., and Gluckman, P. (2000) N-terminal tripeptide of IGF-1 (GPE) prevents the loss of TH positive neurons after 6-OHDA induced nigral lesion in rats. *Brain Res. 859*, 286– 92, DOI: 10.1016/S0006-8993(00)01988-0
(41) Alexi, T., Hughes, P. E., Van Roon-Mom, W. M. C., Faull, R. L. M., Williams, C. E., Clark, R. G., and Gluckman, P. D. (1999) The IGF-I amino-terminal tripeptide glycine-prolineglutamate (GPE) is neuroprotective to striatum in the quinolinic acid lesion animal model of Huntington's disease. *Exp. Neurol. 159*, 84– 97, DOI: 10.1006/exnr.1999.7168
(42) Mishra, R. K., Chiu, S., Chiu, P., and Mishra, C. P. (1983) Pharmacology of L-prolyl- L-leucylglycinamide (PLG): a review. *Methods Find. Exp. Clin. Pharmacol. 5*, 203– 233
(43) Bhargava, H. N. (1984) Effects of prolyl-leucyl-glycinamide and cyclo(leucyl-glycine) on the supersensitivity of dopamine receptors in brain induced by chronic administration of haloperidol to rats. *Neuropharmacology 23*, 439– 44, DOI: 10.1016/0028-3908(84)90252-1
(44) Srivastava, L. K., Bajwa, S. B., Johnson, R. L., and Mishra, R. K. (1988) Interaction of L-prolyl-L-leucyl glycinamide with dopamine D2 receptor: evidence for modulation of agonist affinity states in bovine striatal membranes. *J. Neurochem. 50*, 960– 8, DOI: 10.1111/j.1471-4159.1988.tb03005.x
(45) Mishra, R. K., Srivastava, L. K., and Johnson, R. L. (1990) Modulation of high-affinity CNS dopamine D2 receptor by L-pro-L-leu-glycinamide (PLG) analogue 3(R)-(N-L-prolylamino)-2-oxo-1-pyrrolidineacetamide. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry 14*, 821– 7, DOI: 10.1016/0278-5846(90)90054-K
(46) Beyaert, M. G. R., Daya, R. P., Dyck, B. A., Johnson, R. L., and Mishra, R. K. (2013) PAOPA, a potent dopamine D2 receptor allosteric modulator, prevents and reverses behavioural and biochemical abnormalities in an amphetamine-sensitized preclinical animal model of schizophrenia. *Eur. Neuropsychopharmacol. 23*, 253– 262, DOI: 10.1016/j.euroneuro.2012.04.010
(47) Basu, D., Tian, Y., Bhandari, J., Jiang, J. R., Hui, P., Johnson, R. L., and Mishra, R. K. (2013) Effects of the dopamine D2 allosteric modulator, PAOPA, on the expression of GRK2, arrestin-3, ERK1/2, and on receptor internalization. *PLoS One 8*, e70736, DOI: 10.1371/journal.pone.0070736
(48) Ferreira da Costa, J., Caamano, O., Fernandez, F., Garcia-Mera, X., Sampaio-Dias, I. E., Brea, J. M., and Cadavid, M. I. (2013) Synthesis and allosteric modulation of the dopamine receptor by peptide analogs of L-prolyl-L-leucyl-glycinamide (PLG) modified in the L-proline or L-proline and L-leucine scaffolds. *Eur. J. Med. Chem. 69*, 146– 58, DOI: 10.1016/j.ejmech.2013.08.001
(49) Verma, V., Mann, A., Costain, W., Pontoriero, G., Castellano, J. M., Skoblenick, K., Gupta, S. K., Pristupa, Z., Niznik, H. B., Johnson, R. L., Nair, V. D., and Mishra, R. K. (2005) Modulation of agonist binding to human dopamine receptor subtypes by L-prolyl-L-leucyl-glycinamide and a peptidomimetic analog. *J. Pharmacol. Exp. Ther. 315*, 1228– 36, DOI: 10.1124/jpet.105.091256
(50) Keov, P., Sexton, P. M., and Christopoulos, A. (2011) Allosteric modulation of G protein-coupled receptors: a pharmacological perspective. *Neuropharmacology 60*, 24– 35, DOI: 10.1016/j.neuropharm.2010.07.010

(51) Hill, T., and Lewicki, P. (2006) *STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*, Vol. 1, p 813. StatSoft, Tulsa, OK.
(52) Smith, T. C., and Frank, E. (2016) Introducing Machine Learning Concepts with WEKA, *Statistical Genomics: Methods and Protocols*, pp 353– 378, Springer, New York, NY.
(53) John, G. H., and Langley, P. (1995) *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp 338– 345, Morgan Kaufmann, San Mateo.
(54) Fineberg, H. V. (1980) Decision trees: construction, uses, and limits. *Bull. Cancer 67*, 395– 404
(55) Kohavi, R. (1995) *Proceedings of 8th European Conference on Machine Learning*, pp 174– 189, Springer-Verlag, Heraclion, Greece.
(56) Christopoulos, A. and Kenakin, T. (2002) G protein-coupled receptor allosterism and complexing. *Pharmacol. Rev. 54*, 323– 74, DOI: 10.1124/pr.54.2.323
(57) Trott, O. and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. 31*, 455– 61, DOI: 10.1002/jcc.21334
(58) O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: An open chemical toolbox. *J. Cheminf. 3*, 33, DOI: 10.1186/1758-2946-3-33
(59) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem. 30*, 2785– 91, DOI: 10.1002/jcc.21256
(60) Chang, M. W., Lindstrom, W., Olson, A. J., and Belew, R. K. (2007) Analysis of HIV wild-type and mutant structures via in silico docking against diverse ligand libraries. *J. Chem. Inf. Model. 47*, 1258– 62, DOI: 10.1021/ci700044s
(61) Coutsias, E. A., Seok, C., and Dill, K. A. (2004) Using quaternions to calculate RMSD. *J. Comput. Chem. 25*, 1849– 57, DOI: 10.1002/jcc.20110
(62) Gonzalez-Diaz, H., Perez-Montoto, L. G., and Ubeira, F. M. (2014) Model for vaccine design by prediction of B-epitopes of IEDB given perturbations in peptide sequence, in vivo process, experimental techniques, and source or host organisms. *J. Immunol. Res. 2014*, 768515, DOI: 10.1155/2014/768515
(63) Gonzalez-Diaz, H., Arrasate, S., Gomez-SanJuan, A., Sotomayor, N., Lete, E., Besada-Porto, L., and Ruso, J. M. (2013) General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry. *Curr. Top. Med. Chem. 13*, 1713– 41, DOI: 10.2174/1568026611313140011