Synthesis, Pharmacological and Biological Evaluation of 2-Furoyl-based MIF-1 Peptidomimetics, and the Development of a General-Purpose Model for Allosteric Modulators (ALLOPTML)

Journal:	*ACS Chemical Neuroscience*
Manuscript ID	cn-2020-00687e
Manuscript Type:	Article
Date Submitted by the Author:	22-Oct-2020
Complete List of Authors:	Sampaio-Dias, Ivo; University of Porto Faculty of Sciences, Chemistry and Biochemistry Rodríguez-Borges, José; Universidade do Porto Faculdade de Ciencias, Química e Bioquímica Yáñez-Pérez, Víctor; University of the Basque Country Arrasate, Sonia; Universidad del Pais Vasco, Quimica Organica II Llorente, Javier; University of the Basque Country, Pharmacology Brea, José; Universidade de Santiago de Compostela, Department of Pharmacology Bediaga, Harbil; University of the Basque Country Viña, Dolores; Universidade de Santiago de Compostela, Farmacología Loza, María; Universidade de Santiago de Compostela, Departamento de Farmacologia Caamaño, Olga; Universidade de Santiago de Compostela, Organic Chemistry García-Mera, Xerardo; Universidade de Santiago de Compostela, Dpt. of Organic Chemistry González-Díaz, Humbert; University of the Basque Country, Departament of Organic Chemistry II

Synthesis, Pharmacological and Biological Evaluation of 2-Furoyl-based MIF-1 Peptidomimetics, and the Development of a General-Purpose Model for Allosteric Modulators (ALLOPTML)

Ivo E. Sampaio-Dias[1,*], José E. Rodríguez-Borges[1], Víctor Yáñez-Pérez[2], Sonia Arrasate[3], Javier Llorente[3,4], José M. Brea[5], Harbil Bediaga[2,6], Dolores Viñas[7], María Isabel Loza[5], Olga Caamaño[8], Xerardo García-Mera[8,*], and Humberto González-Díaz[2,9,10,*]

[1] *LAQV/REQUIMTE, Dept. of Chemistry and Biochemistry, University of Porto, 4169-007 Porto, Portugal.*

[2] *Dept. of Organic Chemistry II, University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[3] *Dept. of Pharmacology, University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[4] *Dept. of Pharmacology, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[5] *Innopharma Screening Platform. Biofarma Research group. Centre of Research in Molecular Medicine and Chronic Diseases CIMUS, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[6] *Dept. of Physical Chemistry, University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[7] *Dept. of Pharmacology, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[8] *Dept. of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[9] *Basque Center for Biophysics (CSIC UPV/EHU), University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[10] *IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

Corresponding Authors:

* idias@fc.up.pt (I.E.S.-D.); xerardo.garcia@usc.es (X.G.-M.); humberto.gonzalezdiaz@ehu.es (H.G.-D.).

ABSTRACT.

This work describes the synthesis and pharmacological evaluation of 2-furoyl-based Melanostatin (MIF-1) peptidomimetics as dopamine D_2 modulating agents. Eight novel peptidomimetics were tested for their ability to enhance the maximal effect of tritiated *N*-propylapomorphine ([^{3}H]-NPA) at D_2 receptors (D_2R). In this series, 2-furoyl-L-leucylglycinamide (**6a**) produced a statistically significant increase in the maximal [^{3}H]-NPA response at 10 pM (11 ± 1%), comparable to the effect of MIF-1 (22 ± 9%) at the same concentration. This result supports previous evidence that the replacement of proline residue by heteroaromatic scaffolds are tolerated at the allosteric binding site of MIF-1. Biological assays performed for peptidomimetic **6a** using cortex neurons from 19-day old Wistar-Kyoto rat embryos suggest **6a** displays no neurotoxicity up to 100 μM. Overall, the pharmacological and toxicological profile and the structural simplicity of peptidomimetic **6a** makes this peptidomimetic a potential lead compound for further development and optimization, paving the way for the development of novel modulating agents of D_2R suitable for the treatment of CNS-related diseases. Additionally, the pharmacological data herein reported was used, along with >20000 outcomes of preclinical assays, to seek a general model to assess the potential of a series of compounds as allosteric modulators for a myriad of receptors targets, organisms, cell lines, and biological activity parameters based on Perturbation Theory (PT) ideas and Machine Learning techniques (ML), abbreviated as ALLOPTML. By doing so, ALLOPTML shows specificity Sp = 89.2/89.4% and sensitivity Sn = 71.3/72.2% in training/validation series, respectively. To the best of our knowledge, ALLOPTML is the first general-purpose chemoinformatic tool using a PTML-based model for the multi-output and multi-condition prediction of allosteric compounds, which is expected to save both time and resources during the early drug discovery.

Keywords: Allosteric modulators; Artificial Neural Networks; Big data; ChEMBL; Machine Learning; Melanostatin; Multi-target models; Perturbation Theory.

1. Introduction

Dopamine receptors belong to a complex monoaminergic family of G protein-coupled receptors (GPCRs) represented by five distinct receptors ($D_{1\text{-}5}$ receptors),[1] which are grouped into D_1-like (which mediate excitatory neurotransmission, including D_1 and D_5 isoforms) and D_2-like receptors (associated to inhibitory neurotransmission, comprising D_2, D_3, and D_4 isoforms).[1] These two subgroups differ in their signal transduction, binding profile and physiological effects.[2] Among the $D_{1\text{-}5}$ receptors, D_2-like receptors are considered of utmost importance due to their intrinsic role in the pathophysiology of neurological and psychiatric disorders.[3] Alternative splicing leads to generation of two distinct D_2 receptors (D_2Rs): D_2 short (D_{2S}) and D_2 long (D_{2L}) isoforms, which are associated with presynaptic and postsynaptic populations of D_2R, respectively.[4] The difference between these two splicing isoforms consist of 29 amino acid residues located in the III intracellular loop, responsible for the G protein coupling.[4] The design of ligands with receptor-specificity is considered a herculean task due to the high structural homology shared by D_1-type or D_2-type receptors.[3] For example, while D_2R is the main target to control the positive symptoms of schizophrenia, none of antipsychotics approved are capable to discriminate D_2 from D_3 receptors.[3] In this sense, allosteric regulation is considered a key pharmacological approach to selectively modulate the responsiveness of a specific receptor subtype with negligible or absent interference on the other subtypes. In this sense, allosteric modulators may contribute to the development of new selective and less toxic therapies than those focused on the orthosteric sites.[5] Allosteric modulators bind to a topological region of the receptor (i.e., an allosteric site) different from the binding site used by endogenous ligands (i.e. the orthosteric site). Upon binding, allosteric modulators induce structural modification that can either stabilize the active form of the receptor (positive modulation), inactivate it (negative modulation), or have no influence on the ligand-receptor binding (silent or neutral modulation).[6]

Positive allosteric modulators (PAMs) operate by lowering the energy barrier necessary to shift the receptor to the active conformation, by stabilizing the active conformation, or both.[7] PAMs do not display any activity or pharmacological effect in the absence of the endo/exogenous agonists, but when combined with an orthosteric agonist, they increase its efficacy, thus minimizing the overall side-effect profile of the agonist.[7]

An example of a PAM with intrinsic D_2-selectivity is the neuropeptide L-prolyl-L-leucylglycinamide,[8-12] also known Melanostatin[8] or as melanocyte stimulating hormone release inhibiting factor 1 (MIF-1, **Figure 1**).[11]

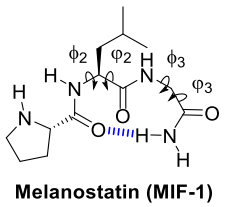


Figure 1. Melanostatin (MIF-1) structure denoting a type-II β-turn as the bioactive conformation assisted by an intramolecular hydrogen bond (dashed blue line).

MIF-1 is formed by exocyclic cleavage of oxytocin hormone and displays several biological activities within the central nervous system (CNS),[13] acting mainly as a neuronal modulating agent in the nigrostriatal pathway.[14] Pharmacologically, MIF-1 interacts with D_2R increasing the effect of dopaminergic agonists on these receptors such as *N*-propylnorapomorphine (NPA), apomorphine, and 2-amino-6,7-dihydroxy-3,4-tetrahydronaphthalene (ADTN).[8] Studies show that MIF-1 potentiates the binding of [^{3}H]-NPA and [^{3}H]-quinpirrole in a dose-dependent manner to D_{2L}, D_{2S}, and D_4 receptor subtypes to give a bell-shaped dose−response curve,[15] without enhancing the binding to D_1 and D_3 receptors.[8, 14] Mechanistically, it is suggested that MIF-1 promotes stabilization of D_2R in a higher affinity state, making this receptor more sensitive to agonists.[14, 16, 17] It is speculated that MIF-1 may have a direct interaction with a domain of the receptor other than its orthosteric site, or that it may display an indirect mechanism associated with the increase of the rate of GTP hydrolysis promoted by G protein.[18] However, this last hypothesis seems less likely as the modulatory effect is not observed in other G protein coupled receptors, such as α_2-adrenergic receptors, or with G_s-coupled receptors, such as D_1R family, showing thus D_2R-specificity.[8]

Regarding its PAM activity and D_2R-selectivity, MIF-1 has been proposed as an attractive drug candidate for further development in the treatment of CNS-related neurodegenerative diseases associated with dopamine receptor sensitivity, such as Parkinson's disease (PD), schizophrenia, tardive dyskinesia, and Gilles de la Tourette syndrome.[19] For example, MIF-1 is widely acknowledge for alleviating the behavioral changes in animal models of PD.[20, 21] Furthermore, this neuropeptide has also the ability to interact with other drugs by modifying their effects[22] (e.g., MIF-1 antagonizes morphine-induced catalepsy as well as dyskinesia induced by antipsychotic drugs).[23-25] X-ray crystallography[26] and structure−activity relationship studies using constrained peptidomimetics[27] suggest that MIF-1 adopts a type-II β-turn formed by intramolecular hydrogen bonding between the carbonyl oxygen atom of the L-proline residue and the *trans* amide proton of the glycinamide residue (depicted as dashed blue lines in **Figure 1**). The presence of the C-terminal carboxamide is considered a key pharmacophore for the modulatory activity of MIF-1 and its active peptidomimetics, however an extended bioactive conformation is also suggested.[16] Modifications in the proline residue of MIF-1 have long been recognized to render bioactive peptidomimetics,[28] indicating that this residue may be suitable for chemical derivatization and the development of novel bioactive PAMs.[10-12]

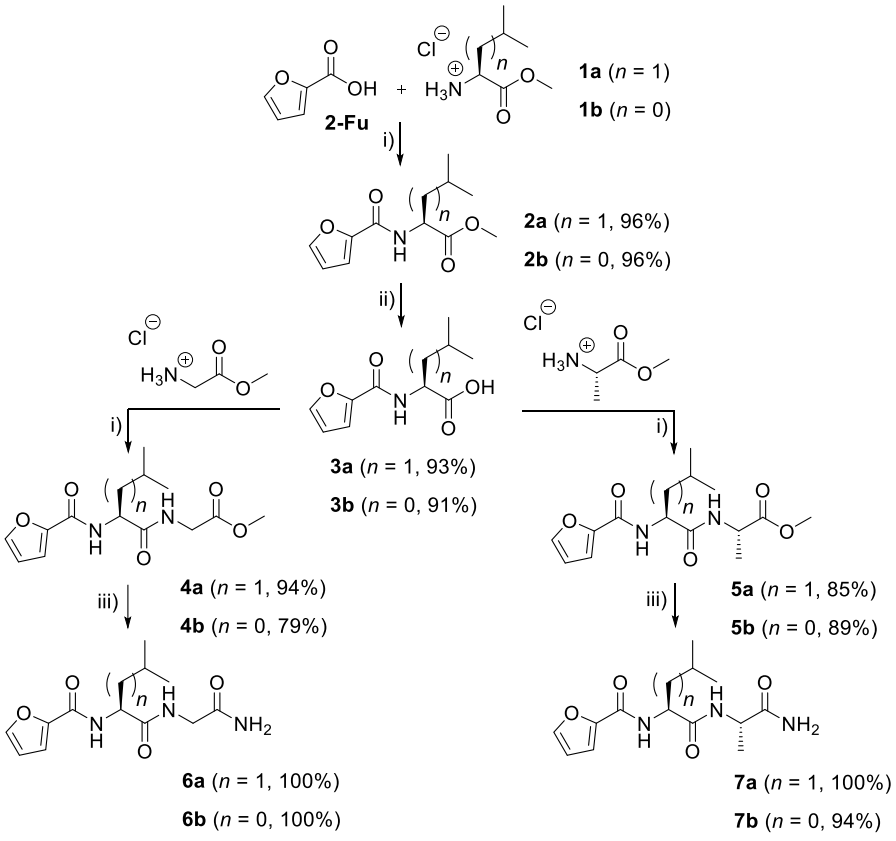
In this work, we aimed to explore the replacement of proline residue in MIF-1 structure by 2-furoic acid (**2-Fu**) for the assembly and pharmacological evaluation of a novel class of MIF-1 peptidomimetics and use the experimental data to construct a general-purpose computational model based on Perturbation-Theory (PT) and Machine Learning (ML) ideas to predict allosteric modulators for a myriad of receptors targets, organisms, cell lines, and biological activity parameters, abbreviated as ALLOPTML.

2. Results and Discussion

Design and chemistry.

Proline has long been stablished as a non-essential amino acid residue for the activity of MIF-1 analogues,[29] thereby offering the possibility for the design of novel MIF-1 peptidomimetics with PAM activity and D_2-selectivity. For example, picolinic acid has been successfully used as proline surrogate for the assembly of MIF-1 analogues with PAM activity and suitable toxicological profile.[11] These findings led us to hypothesize that heterocyclic scaffolds may be tolerated at the putative allosteric binding site of D_2R. In this sense, to gain further insights into the architecture of D_2R allosteric binding site, other heteroaromatic scaffolds should be considered as proline surrogates, allowing to further rationalize the chemical features required for the development of pharmacological active PAMs. Considering the pyrrolidine ring of proline amino acid, five-membered heteroaromatic rings such as furan moieties may represent suitable proline surrogates for the design of MIF-1 peptidomimetics and thus providing useful information about the size and nature of the hydrogen bonding at the D_2R allosteric binding site. To the best of our knowledge, the use and application of furan moieties for the development of MIF-1 derivatives have not yet been reported. In this sense, **2-Fu** was chosen for the assembly of a novel class of MIF-1 peptidomimetics.

For the preparation of 2-furoyl-based MIF-1 peptidomimetics, a diversity-oriented synthesis (DOS) strategy was employed to create a set of structure-related amino acid combinations at the central and C-terminal positions. For this purpose, L-leucine was replaced by L-valine whereas glycine was substituted by L-alanine. The synthetic route for preparation of 2-furoyl peptidomimetics is depicted in **Scheme 1**.



Scheme 1. Synthesis of 2-furoyl-based peptidomimetics **4-7(a,b)**. Reagents and conditions: (i) DIEA, TBTU, CH_2Cl_2; (ii) LiOH, MeOH/H_2O followed by H_2SO_4 1 M; (iii) NH_3 (g), MeOH.

Starting from **2-Fu**, the preparation of amides **2(a,b)** was performed by peptide coupling with L-leucine and L-valine, respectively (**Scheme 1**). For that purpose, TBTU was used as peptide coupling reagent in presence of Hünig's base as tertiary amine.[10, 12, 30] After chromatographic purification, compounds **2(a,b)** were obtained in high yields (96%).

Next, compounds **2(a,b)** were saponified using LiOH followed by acidic work-up with H_2SO_4 1M to obtain the corresponding carboxylic acids **3(a,b)**. Since furan system is sensitive to low pH, at which may result in acid-catalyzed ring opening phenomena,[31] special attention was required during the acidic work-up. Therefore, the pH was carefully adjusted to 4 to prevent the formation of side-products. Both peptide products **3(a,b)** were isolated in good to very good yields (91-93%) with no detection of acyclic side-products (TLC, NMR) under this protocol, as corroborated by the corresponding NMR spectral data (^{1}H and ^{13}C-NMR, **Figures S5-S8** of SI). Carboxylic acids **3(a,b)** were then coupled with either glycine (a) or L-alanine (b) methyl esters using the same peptide coupling protocol, delivering **4(a,b)**/**5(a,b)** in good to very good yields (79-94%).

Finally, pseudopeptides **4(a,b)**/**5(a,b)** were converted into the corresponding C-terminal carboxamides **6(a,b)**/**7(a,b)** upon treatment with gaseous ammonia in methanol, prepared *in situ* by reacting $Ca(OH)_2$ and NH_4Cl at 100 ºC. The reactions proceeded cleanly, albeit it took about 48 h for complete conversion (TLC). After the removal of the volatiles, C-terminal carboxamides **6(a,b)**/**7(a,b)** were obtained in practically quantitative yields (94-100%), without the need of chromatographic purifications. Following this synthetic route 2-furoyl peptidomimetics were successfully prepared in high global yields (69-84%).

Pharmacology.

Eight novel MIF-1 peptidomimetics were tested for their ability to potentiate the maximum binding of the radiolabelled agonist *N*-propylnorapomorphine, [^{3}H]-NPA, to cloned human dopamine D_{2S} receptors and their activity was compared to that of MIF-1 as described in the literature.[8] Peptidomimetics **4-7(a,b)** were tested at eight different concentrations in the range between 1 pM and 10 μM following the protocol previously reported in our research group.[10 12] The whole set of peptidomimetics was evaluated *in silico* to rule out these ligands being promiscuous pan-assay interference compounds (PAINS) and/or aggregators.[32] The experimental results obtained for the binding assays are shown in **Table 1** and **Figure 2**.

Table 1. Experimental results for MIF-1 derivatives **4-7(a,b)** in the D_2R binding assays.

Cpnd	% [^{3}H]-NPA$_{max}$	AUC	Log (1/C$_{max}$)
4a	11	53.72	10
4b	9	34.35	12
5a	11	27.25	10
5b	6	9.36	12
6a	11	12.98	11
6b	nd	0	nd
7a	3	2.76	12
7b	nd	0	nd
MIF-1	22	44.19	10

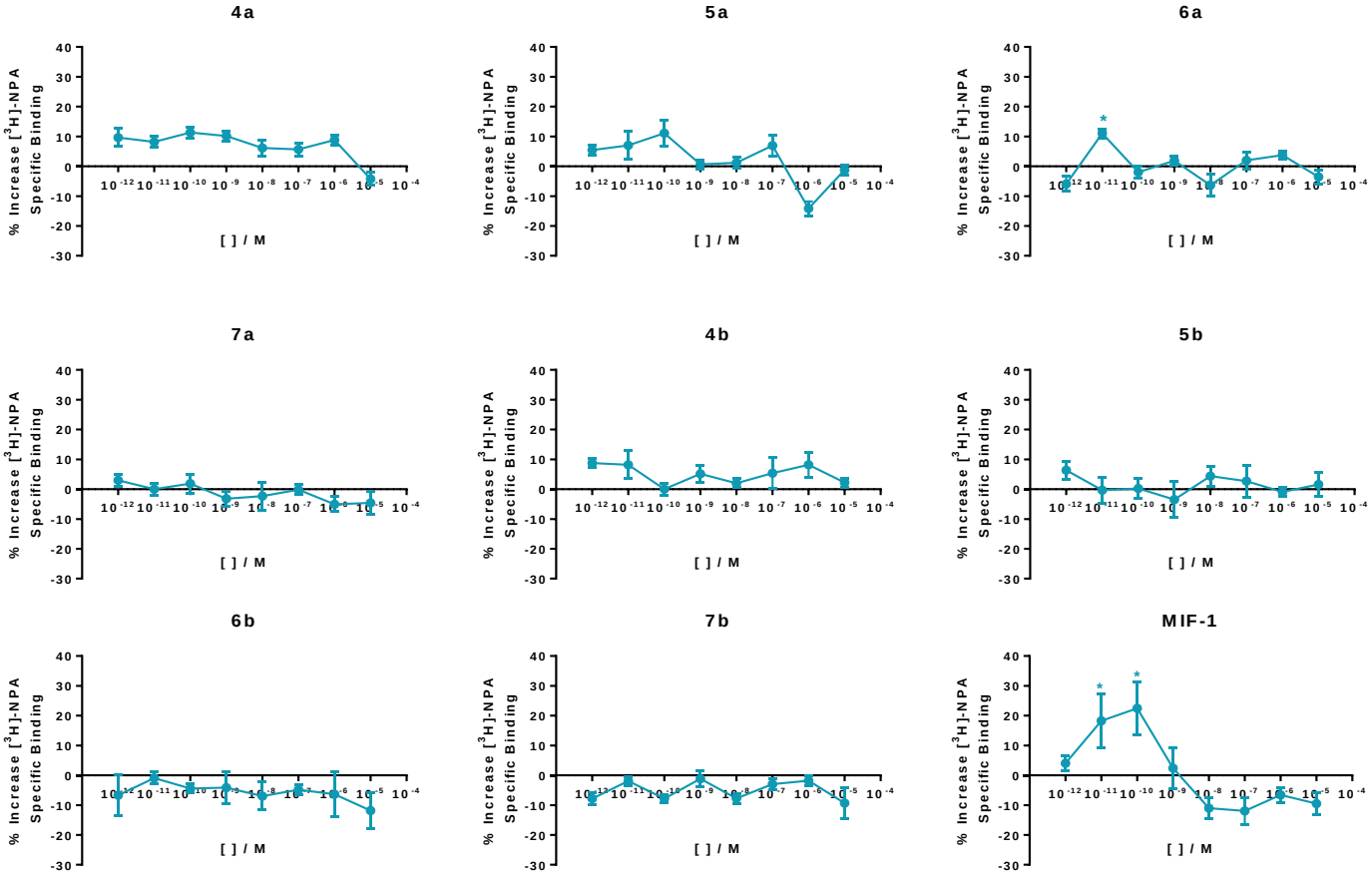


Figure 2. Modulation of the [^{3}H]-NPA binding by 2-furoyl-based MIF-1 peptidomimetics **4-7(a,b)** at eight different concentrations. Points represent the mean ± standard deviation (vertical bars) of three independent experiments carried out with duplicates. *P < 0.05 (ANOVA test; *post-hoc* Dunnet T3 test).

A statistically significant enhancement ($P < 0.05$; ANOVA test; post hoc Dunnet T3 test) of the [^{3}H]-NPA response was observed for peptidomimetic **6a** (**Figure 2**). The maximum effect for **6a** was 11 ± 1% at 10 pM against 22 ± 9% for MIF-1 (positive control) at the same concentration (**Table 1** and **Figure 2**). In respect with the other peptidomimetics, no statistically significant increase in the [^{3}H]-NPA binding was observed, therefore no conclusion can be drawn for these compounds. In this work it was found that MIF-1 displays an increase in the [^{3}H]-NPA binding at lower concentrations than those previously reported by Verma and co-workers.[8] Regarding that allosteric modulators are sensitive to environmental changes elicited by endogenous agonists on GPCRs,[33] this deviation may be attributed to the host cells used for expressing human D_{2S} receptors, which differ from the ones originally reported.[8]

The C-terminal carboxamide is widely recognized as a key pharmacophore for the activity of MIF-1 and its peptidomimetics,[34,9] which is crucial for these compounds to adopt the postulated type-II β-turn bioactive conformation.[34,9] However, Saitton and co-workers suggested the possibility for an extended bioactive conformation of MIF-1 through the development of bioactive peptidomimetics incorporating a 2,3,4-trisubstituted pyridine scaffold as leucyl mimetic which cannot adopt the type-II β-turn.[16] Like MIF-1, peptidomimetic **6a**, displays the L-leucylglycinamide dipeptide motif, therefore, a type-II β-turn conformation is envisioned for this compound. The simple substitution of proline residue by **2-Fu** proved to effectively mimic the MIF-1 modulatory activity at D_2R allosteric binding site. The activity of **6a** is thus considered of utmost interest since it endorses previous findings that achiral and heteroaromatic scaffolds are tolerated in the putative MIF-1 binding pocket, establishing proline as a key residue for derivatization in the development of MIF-1 analogues with PAM activity.

***In vitro* neurotoxicity profile of 2-furoyl MIF-1 peptidomimetics**.

Low nonspecific cytotoxicity is an important feature of lead compounds in drug development. Therefore, peptidomimetics **4-7(a,b)** were evaluated to assess their neurotoxic profiles using culture of motor cortex neurons from 19 day old Wistar-Kyoto (WKY) rat embryos.[11, 35] Hydrogen peroxide, H_2O_2, was selected as a standard control for neurocytotoxity.[48,72,36] In fact, the majority of the *in vitro* neurotoxicity studies carried out in cortical neurons uses H_2O_2 as an effective inducer of oxidative stress.[37] The use of H_2O_2 creates an imbalance between reactive oxygen species (ROS) levels and scavenging processes.[37] Excessive cell exposure to free radicals subsequently leads to neuronal death. In this assay, cellular viability was estimated based on the ability of cells to reduce MTT after a 24 h incubation period with the test compounds (100 μM). The percentage of cell viability was compared with the dimethyl sulfoxide (DMSO, negative control). All results are expressed as mean ± SEM of five independent experiments (**Figure 3**).

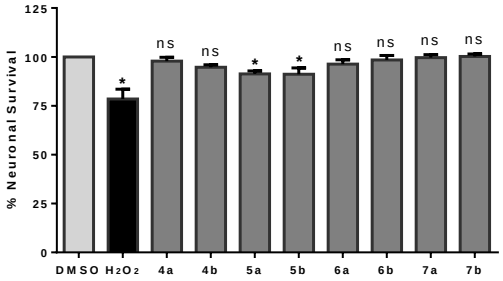


Figure 3. Neurotoxicity assays for peptidomimetics **4−7(a,b)** in cultured neurons from embryonic motor cortices of fetal Wistar-Kyoto rats. Data are the mean ± SEM of n = 5 independent experiments at 100 μM. Negative control: DMSO; Positive control: 100 μM H_2O_2. *Level of statistical significance: $P < 0.05$ *vs.* DMSO condition by *t* test.

As expected, the stress inducer (100 μM H_2O_2) reduced the cell survival to 78.51 ± 4.97% compared to the non-treated cells control ($P < 0.05$). In this assay compounds **5a** and **5b** display a statistically significant decrease in neuronal survival (91.34 ± 1.50% and 91.13 ± 3.20%, respectively, $P < 0.05$), in comparison with DMSO condition, suggesting these peptidomimetics exhibit a marginal neurotoxicity. The cell viability was not affected by the presence of the remaining test compounds when compared to the control (DMSO). These results indicate that 2-furoyl-based MIF-1 analogues display a safe profile up to 100 μM.

PTML-LDA linear models with one-fold reference and moving averages.

The application and combination of computational methods and predictive software tools in medicinal chemistry are considered of utmost interest, which may help to identify, prioritize, and optimize novel chemotypes with desired pharmacological activity profile.[38] Computational models and algorithms capable to make predictions on the activity using different conditions of assay with minimal computational cost are thus highly desired. The development of computer-aided models drives molecular design and decision making in drug discovery, boosting the identification of lead compounds while saving resources and time-consuming assays, being therefore considered an important pillar of modern drug research.[38]

Considering the absence of a robust and general-purpose model for the multi-output and multi-condition prediction of allosteric compounds in medicinal chemistry, we sought to fill this gap. To achieve this goal, we developed and implemented ALLOPTML, abbreviation used to describe a novel predictive model for allosteric modulators based on Perturbation Theory (PT) ideas and Machine Learning (ML) techniques, which may be used for several types of systems and for a hundred different proteins. For this purpose, the pharmacological data of peptidomimetics **4−7(a,b)** disclosed in this work was used to train the model along with the 21439 cases from 8984 different ChEMBL assays of 8957 allosteric compounds for 79 different proteins. For this task 12 different variables ($\mathbf{c}_{0\text{-}11}$) were considered from the ChEMBL database Big Data (BD) sets of compounds covering 38 biological activity parameters ($\mathbf{c}_0$; e.g., efficacy, potency, intrinsic activity, IC_{50}, K_i, K_m); organisms of the protein target ($\mathbf{c}_1$; e.g., human, rat, mouse); organism of assay ($\mathbf{c}_2$; e.g., rat, swine, herpesvirus 5); different cellular lines ($\mathbf{c}_3$; e.g., CHO, HEK293), and so on. The general workflow used in this paper to obtain the PTML model is depicted in **Figure 4**.

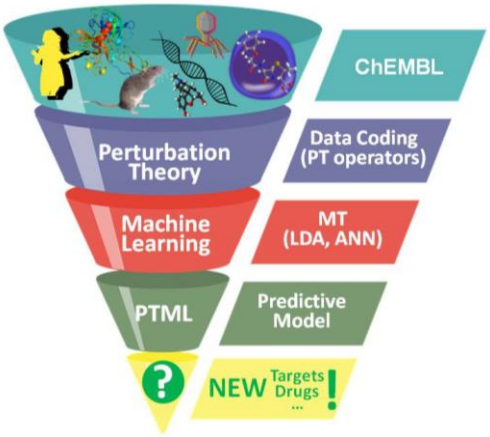


Figure 4. General workflow used to obtain the PTML model.

So far, finding allosteric binding sites and lead compounds has been largely serendipitous, accomplished by medicinal chemists through preclinical assays performing high-throughput screening with a high number of combinations of experimental conditions ($\mathbf{c}_{a,j}$). Besides, the existence allo-network drugs, i.e., allosteric compounds that do not act directly over the pharmacological target but over a second protein which in turn interacts directly or indirectly with the target protein makes this task even more challenging.[39-41] In addition to experimental techniques, computational techniques can be used to predict new drugs against different targets.[42] In this sense, many researchers have developed Cheminformatics models for the discovery of new compounds. Nevertheless, almost all these models are designed for homologous series of compounds, one target, and/or a single cell line. Other models, instead, use heterogeneous series of compounds, although are unable to incorporate information about the target, cell line, organism of assay, among other features.[43-58] This drawback can be partially faced by ML techniques through which diverse molecular descriptors, encoding the chemical structure of different drugs, can be calculated.[59-75] ML techniques have been already carried out to predict allosteric pockets on proteins.[76] Unfortunately, this method fail to account for large BD sets of preclinical assays, such as the ChEMBL database, difficult to study due to the high complexity of the data in addition to the huge volume of highly heterogeneous preclinical assays.[77, 78]

To tackle this problem we have associated PT and ML modeling.[71, 79]. In fact, PTML (PT + ML) models have been used in many fields of knowledge such as Medicinal Chemistry, Proteomics or Nanotechnology for modeling large data sets with BD features.[35, 80-92] Probably due to the potential benefits of allosteric drugs, a renewed interest in allostery has led to the development of a number of computational approaches to understanding allostery.[93] Nevertheless, to the best of our knowledge, a general-purpose PTML models for multiple preclinical assays of allosteric compounds have not been reported yet.

PTML-LDA linear models with one-fold reference and moving averages.

The domain of the PTML model is the array $\mathbf{c}_j$ for all sub-sets (assays) with different standard experimental conditions. PTML model starts with the expected value of a given activity and adds the effect of different perturbations in the system. Therefore, the model has two types of input variables: the reference or expected-value function $f(v_{ij})_{ref}$ and the PT operators $\Delta D_k(\mathbf{c}_j)$. The input variable $f(v_{ij})_{ref}$ represents the expected value of a particular biological activity for one compound in a given biological activity c_0. The other PT operators are moving averages (MA) calculated for one sub-set of multiple conditions at time. The PT operators measure the deviation of D_{ki} from the expected value of $\langle D_k(\mathbf{c}_j)\rangle$ (average value) of this descriptor for different sets of drugs c_j.[94-96] Hence, we can calculate the PT operators as follow $\Delta D_k(c_j) = D_{ki} - \langle D_k(c_j)\rangle$. They depend on the value of the molecular descriptors D_{ki} of type k used to quantify the structure of the i^{-th} molecule. In this work, the specific molecular descriptors used are D_1 = ALOGP (n-Octanol/Water Partition Coefficient) and D_2 = PSA (Polar Surface Area), taken directly from ChEMBL data set.

Table 2. Principal PT operators used as input (c_0-c_8)

MA Type [a]	$\mathbf{c}_{j,l}$	Condition	Symbol	Operator Formula	Operator Information
MA	c_0	Activity types	$f(v_{ij})_{expt}$	$n(f(v_{ij})_{obs}=1)/n_j$	Expected value of probability $p(f(v_{ij})=1)_{expt}$ for the activity v_{ij} of type c_0
	c_0	Activity type	$\Delta D_k(c_0)$	$D_k - \langle D_k(c_j)\rangle$	$\Delta D_k(\mathbf{c}_{j,l})$ operators account for changes (Δ) on the chemical structure of the compound, quantified by the molecular descriptor D_k, with respect to the expected value of the descriptor $\langle D_k(c_{j,l})\rangle$ for molecules assayed under the same single condition of assay c_j.
	c_1	Orgs.	$\Delta D_1(c_1)$		
	c_2	Assay Orgs.	$\Delta D_1(c_2)$		
	c_3	Cell line	$\Delta D_1(c_3)$		
	c_4	Target	$\Delta D_1(c_4)$		
MMA	$\mathbf{c}_j$	Multiple Condition Array	$\Delta D_k(\mathbf{c}_j)$	$D_k - \langle D_k(\mathbf{c}_j)\rangle$	$\Delta D_k(\mathbf{c}_{j,l})$ account for changes (Δ) on the chemical structure of the compound, quantified by the molecular descriptor D_k, with respect to the expected value of the descriptor $\langle D_k(\mathbf{c}_j)\rangle$ for a molecules assayed under the same sub-set of multiple conditions of assay at the same time represented by selected cells of the matrix of conditions $\mathbf{c}_j$.

[a]MA = Moving Average operators *vs*. MMA = Multiple Moving Average operators

Using linear discriminant analysis (LDA)[97] linear classification models can be developed. See details about the form of a PTML linear model-equation on methods section. The equation of the best model found using this kind of PT operators is the following:

$$f(\mathrm{v_{i,j}})_{calc} = -3.62490 + 8{,}82060 \cdot f(\mathrm{v_{i,j}}/c0)_{ref} - 0.02087 \cdot \Delta\mathrm{D_2(c_1)} - 0.21686 \cdot \Delta\mathrm{D_1(c_3)} + 0.02817 \cdot \Delta\mathrm{D_2(c_3)} + 0.33521 \cdot \Delta\mathrm{D_1(c_4)} + 0.01765 \cdot \Delta\mathrm{D_2(c_4)} - 0.01966 \cdot \Delta\mathrm{D_2(c_8)} \quad (1)$$

$$\mathrm{n} = 16080 \quad \chi^2 = 3927.97 \quad \mathrm{p} < 0.05$$

The statistical parameters of the model are: the number of cases used to train the model, n; the Chi-square test, $\chi 2$; and the p-level, p. In **Table 2** it is shown a more detailed explanation about all the input variables analyzed to seek the PTML-LDA linear model. The output of the model $f(v_{ij})_{calc}$ is a scoring function of the value v_{ij} of biological activity of the i^{-th} drug in the conditions of assay c_j. In the particular case of an LDA model $f(v_{ij})_{calc}$ is not a probability and therefore it is not in the range 0-1. However, for a given value of $f(v_{ij})_{calc}$, the LDA algorithm, which uses the Mahalanobis's distance metric,[97] can calculate the respective values of posterior probabilities $p(f(v_{ij}) = 1)_{pred}$.

After calculating $p(f(v_{ij}) = 1)_{pred}$ it can be easily calculated the Boolean function $f(v_{ij})_{pred} = 1$ when $p(f(v_{ij}) = 1)_{pred} > 0.5$ or $f(v_{ij})_{pred} = 0$ otherwise. The values of $f(v_{ij})_{pred} = 1$ or 0 must be compared with the respective observed values $f(v_{ij})_{obs} = 1$ or 0 to calculate the specificity (Sp, (%)), sensitivity (Sn, (%)), and overall accuracy (Ac, (%)) of the model. Hence, when $f(v_{ij})_{pred} = f(v_{ij})_{obs}$ the case is correctly classified.[97] The present one-condition moving averages model resulted in values of Sp = 79.7 %, Sn = 68.6 %, and Ac = 77.8 % in training series and in very similar values in external validation series. It is worth mentioning that the data points (compound-assay pair) used in validation series have not been used to train the model.

We used the forward-stepwise strategy[97] of variable selection to detect the more important perturbations on different conditions c_j related to pharmacologic property, like assay organism, cell line, among others. See details about these variables on **Table 2**. In this particular model the ML-selected operators measure perturbations on the values of ALOGP (D_1) and PSA (D_2) compared to other sub-sets c_j of drugs where c_j is: c_1, organisms; c_3, cell line; c_4, ChEMBL targets and c_8, binding domains. The parameters ALOGP and PSA are extensively used in Medicinal Chemistry. While ALOGP is related to the lipophilicity, PSA strongly reflects the hydrogen bonding capacity and polarity of a given drug.[98] Accordingly, these parameters mirror the ability of drugs to pass through biological membranes or interact with protein hydrophobic pockets. For example, for any drug designed to act on receptors in the central nervous system it is assumed that a PSA less than 90 $Å^2$ is required to be capable of penetrating the blood–brain barrier.[99, 100]

This model can be used to estimate the activity of a putative new allosteric modulator in different conditions of assay. To begin with, the expected probability of activity $p(f(v_{ij})_{obs} = 1)_{ref}$ must be substituted on the equation, see **Table 3**.

Table 3. One-condition averages, cutoff, desirability $d(c_0)$, among others, for selected biological parameters.

Condition c_0[a]	Input parameters used to specify $\mathbf{c}_{0,l}$[b]						
Activity	$<D_1(\mathbf{c}_0)>$	$<D_2(\mathbf{c}_0)>$	$n_j(c_0)$	$n_j(f(v_{ij}) = 1)_{obs}$	$p(f(v_{ij}) = 1)_{ref}$	cutoff	$d(\mathbf{c}_0)$
Potency (nM)	2.93	83.57	8173	128	0.016	100.0	-1
EC_{50} (nM)	3.51	69.15	4177	649	0.155	100.0	-1
IC_{50} (nM)	3.43	71.60	2256	432	0.191	100.0	-1
Activity (%)	3.63	73.74	1500	593	0.395	48.9	1
$\%_{max}$ (%)	3.50	82.28	1123	487	0.434	70.9	1
E_{max} (%)	3.66	64.44	941	215	0.228	128.3	1
Inhibition (%)	5.03	84.22	777	424	0.546	40.8	1
Efficacy (%)	4.20	80.19	144	85	0.590	55.5	1
Allosteric Enhancer (%)	4.04	75.37	132	47	0.356	22.2	1
K_i (nM)	4.14	130.73	58	24	0.414	100.0	-1

[a]Condition $c_j = c_0$ = subscript j refers to the type of activity parameter measured; subscript l are the levels of the j^{th} condition (specific pharmacological parameters).

Note that these values change for different activities, like potency (nM), inhibition (%), IC_{50} (nM), activity (%), E_{max} (%), *etc*. Therefore, this model can predict different kinds of activity parameters (c_0) for any compounds. For that, the values of ALOGP and PSA for the new compound (taken from ChEMBL or calculated with software) must be fitted in the model.

The expected values of probability, $p(f(v_{ij})_{obs} = 1)_{expt}$, must be calculated by the formula: $p(f(v_{ij})_{obs} = 1)_{ref} = n(f(v_{ij}) = 1)_{obs}/n_j$. Where, $n(f(v_{ij}) = 1)_{obs}$ is the number of compounds with a desired level of activity for the condition c_j and n_j is the number of compounds assayed for the same condition c_j. The desired level of activity of a compound is $f(v_{ij})_{obs}$ and when that particular c_0 is desirable $f(v_{ij})_{obs}$ is defined to be 1, or else it is set as 0. However, depending on the desirability ($d(c_0)$) of a given activity there are two ways to set $f(v_{ij})_{obs} = 1$. For a property desirability $d(c_0) = 1$ and when the value of activity is greater than its cutoff ($v_{ij} >$ cutoff) the desired level of activity is $f(v_{ij})_{obs} = 1$. A compound also has a desired level of activity $f(v_{ij})_{obs} = 1$ when the value of activity is smaller than its cutoff ($v_{ij} <$ cutoff) for not desirable properties, i.e., $d(c_0) = -1$. Otherwise, the compound is not considered to have a desired level of activity so $f(v_{ij})_{obs}$ is set as 0. Therefore, it is essential to realize that for properties of the compound that we want to maximize, such as activity (%), the property desirability $d(c_0)$ is set as 1, otherwise $d(c_0)$ is set as -1 (e.g., EC_{50} (nM)). The cutoff for any particular activity (v_{ij}) was set as the average (cutoff = $<v_{ij}>$) of all the values of that activity unless it is described with units in nM (e.g., IC_{50}), in which case the cutoff was set as 100 nM.

In order to predict a new compound, the expected values of the molecular descriptors $D_k(c_j)$ (ALOGP and PSA) for different conditions ($<D_k(c_j)>$) must be included in the model. **Table 4** shows selected values of the averages $<D_1(c_j)>$. Note that these values change for different conditions, so the model gives a different result for one compound if a condition is changed. For instance, $<D_1(c_j)> = 3.44$ for *H. sapiens* and $<D_1(c_j)> = 3.51$ for *M. musculus*. Consequently, the model is able to predict a different activity in human and mouse for the same drug. The full list of the values of one-condition moving averages appears in the supplementary material file SI00.xlsx.

Table 4. One-condition averages and number of cases for selected conditions of assay.

Condition c_1	Parameters used to specify c_1			Condition c_1	Parameters used to specify c_1		
Org. of target	$<D_1(c_1)>$	$<D_2(c_1)>$	$n_j(c_1)$	Org. of target	$<D_1(c_1)>$	$<D_2(c_1)>$	$n_j(c_1)$
H. sapiens	3.44	76.40	15564	*Glycine max*	6.40	63.31	4
R. norvegicus	3.55	75.90	3075	*H. herpesvirus 5*	4.15	47.45	18
Mus musculus	3.51	66.39	107	*S. scrofa*	3.82	75.43	34
Condition c_2	Parameters used to specify c_2			Condition c_2	Parameters used to specify c_2		
Assay organism	$<D_1(c_2)>$	$<D_2(c_2)>$	$n_j(c_2)$	Assay organism	$<D_1(c_2)>$	$<D_2(c_2)>$	$n_j(c_2)_j$
H. sapiens	3.43	76.99	15046	*Glycine max*	6.40	63.31	4
R. norvegicus	3.35	76.68	3550	*H. herpesvirus 5*	4.15	47.45	36
Mus musculus	3.29	70.12	130	*S. scrofa*	3.82	75.43	12
Condition c_3[a]	Parameters used to specify c_3			Condition c_3[a]	Parameters used to specify c_3		
Cell line	$<D_1(c_3)>$	$<D_2(c_3)>$	$n_j(c_3)$	Cell line	$<D_1(c_3)>$	$<D_2(c_3)>$	$n_j(c_3)$
CHO	4.23	79.13	1924	CHO-K1	3.06	88.96	277
HEK293	3.54	57.83	1715	PAM	3.32	76.43	12
Condition c_4[b]	Parameters used to specify c_4			Condition c_4[b]	Parameters used to specify c_4		
Target	$<D_1(c_4)>$	$<D_2(c_4)>$	$n_j(c_4)$	Target	$<D_1(c_4)>$	$<D_2(c_4)>$	$n_j(c_4)$
$GABA_A$	2.76	83.41	2650	A_1R	5.31	83.99	1598
Caspase-1	2.92	85.12	2870	$mGluR_5$	3.04	66.42	1322
Caspase-7	2.94	85.58	2283	M_1R	3.03	68.68	841

[a]Species full name: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* and *Sus scrofa*. [b] A_1R, Adenosine A_1 receptor; $GABA_A$, γ-Aminobutyric acid type A receptor; $mGluR_5$, Metabotropic glutamate receptor 5; M_1R, Muscarinic acetylcholine receptor M_1.

PTML-LDA model with m-fold moving averages.

The PT operators used in the previous model are based on single-condition averages (the simplest case). There is another possibility of calculating PT operators based on multi-condition averages (MMA) [101]. It means that the average calculation runs over all cases with the same set of conditions $\mathbf{c}_j$. Recall that, in this context, $\mathbf{c}_j$ (with **c** in boldface) refers to a vector of multiple conditions $\mathbf{c}_j = (c_0, \ldots, c_j)$. A complex case would be the use of all conditions together or all the conditions from the previous model. However, it is possible to calculate the

averages using different combinations of conditions. For instance, some conditions suggested by the previous approximation (c_1, c_3 and c_4) and some other conditions essential to be in the model (c_0, biological activity; and c_2, assay organism) can be forced into a new model. Therefore, considering that the simpler the model the better it is, the MMA equation ended up as follows:

$$f(\mathrm{v_{ij}})_{calc} = -5.19375867457001000 + 11.75070469825080000 \cdot f(\mathrm{v_{ij}}/c_0)_{ref} + 0.12834055618459200 \cdot \Delta\mathrm{D}_1(\mathrm{c_0, c_1, c_2, c_3, c_4}) + 0.00671023382639323 \cdot \Delta\mathrm{D}_2(\mathrm{c_0, c_1, c_2, c_3, c_4}) \quad (2)$$

$$\mathrm{n} = 16080 \quad \chi^2 = 8556.26 \quad \mathrm{p} < 0.05$$

In training series, the model presented high values of specificity, Sp = 89.2%; sensitivity, Sn = 71.3%; and accuracy, Ac = 86.1%. Likewise, the MMA model was also reliable in external validation series: Sp = 89.4%, Sn = 72.2%, and Ac = 86.4%. These values are in the range considered as useful for classification models with application in Medicinal Chemistry.[102] In **Table 5** the results obtained with this model are compared with the ones from the 1-fold PTML model with simple MAs (one MA for each experimental condition and descriptor).

As can be noted the PTML model with m-fold MMAs (one MMA include all experimental conditions) has higher values of Sp, Sn, and Ac. Specifically, Sp increased approximately 10% with respect to the MA model, Sn surpasses the line of 70%, and Ac was over 85% both in training and validation series. This is an important improvement taking into consideration that the number of variables was reduced from 7 variables in the first model to only 3 variables in the second one, decreasing the possibilities of chance correlation.[103]

Table 5. Comparison of 1-fold *vs*. 4-fold moving average PTML-LDA models.

PTML Model	Data Series	Obs. Sets[a]	Stat. Parm.	Stat. (%)	Predicted sets		
					n_j	$f(v_{ij})_{pred} = 0$	$f(v_{ij})_{pred} = 1$
1-fold (7 vars)	Training series ($\pi_1 = 0.50$)	$f(v_{ij})_{obs} = 0$	Sp	79.7	13272	10578	2694
		$f(v_{ij})_{obs} = 1$	Sn	68.6	2808	883	1925
		Total	Ac	77.8	16080		
	Validation series	$f(v_{ij})_{obs} = 0$	Sp	80.3	4418	3547	871
		$f(v_{ij})_{obs} = 1$	Sn	69.9	941	283	658
		Total	Ac	78.5	5359		
4-fold (3 vars)	Training series ($\pi_1 = 0.55$)	$f(v_{ij})_{obs} = 0$	Sp	89.2	13272	11845	1427
		$f(v_{ij})_{obs} = 1$	Sn	71.3	2808	807	2001
		Total	Ac	86.1	16080		
	Validation series	$f(v_{ij})_{obs} = 0$	Sp	89.4	4418	3951	467
		$f(v_{ij})_{obs} = 1$	Sn	72.2	941	262	679
		Total	Ac	86.4	5359		

Similarly to previous equation, the input variable $f(v_{ij})_{expt}$ represents the expected value of biological activity for one compound but in different combinations of experimental conditions $c_j = (c_0, c_1, c_2, \ldots c_j \ldots c_{max})$, see analogy with the previous model. The other PT operators (called MMAs) are also MA operators but calculated for multiple conditions at the same time. MMAs operators depend also on the value of the molecular descriptors D_{ki} of type k used to quantify the structure of the i^{-th} drug. However, in this case the average value $\langle D_k(c_j)\rangle$ runs over multiple conditions at the same time. It means the MMAs operators can be calculated as follows: $\Delta D_k({}^a\mathbf{c}_j) = D_{ki} - \langle D_k(c_j)\rangle$. However, it should be noted that in MMAs the $\mathbf{c}_j$ (in boldface) denotes an array of multiple categorical variables, see experimental conditions in **Table 6**. Conversely, in one-condition MAs c_j denotes a single condition.

Table 6. Top-25 most populated assays in the data set.

$\mathbf{c}_j$	c_0	c_1	c_2	c_3	c_4	Assay & Conditions Count						
Assay (a)	Actv. (Units)	Target Org.[a]	Assay Org.[a]	Cell Line[b]	Target ID	Assay	c_0	c_1	c_2	c_3	c_4	Target[c]
1	Potency (nM)	*Hs*	*Hs*	nd	4801	2870	8173	15550	15033	17519	2870	C1
2	Potency (nM)	*Hs*	*Hs*	nd	3468	2283	8173	15550	15033	17519	2283	C7
3	Potency (nM)	nd	nd	nd	612545	1976	8173	2650	2686	17519	2650	nd
4	Potency (nM)	*Hs*	*Hs*	nd	2056	773	8173	15550	15033	17519	773	D_1R
5	IC_{50} (nM)	*Rn*	*Rn*	nd	2564	455	2256	3058	3533	17519	1321	mGluR5
6	EC_{50} (nM)	*Hs*	*Hs*	HEK293	3227	370	4177	15550	15033	1709	1873	mGluR5
7	Activity (%)	*Hs*	*Hs*	CHO	226	363	1500	15550	15033	1922	1592	A_1R
8	EC_{50} (nM)	*Hs*	*Hs*	nd	3772	331	4177	15550	15033	17519	812	mGluR1
9	EC_{50} (nM)	*Hs*	*Hs*	nd	2736	322	4177	15550	15033	17519	807	mGluR4
10	IC_{50} (nM)	*Hs*	*Hs*	HEK293	3227	274	2256	15550	15033	1709	1873	mGluR5
11	EC_{50} (nM)	*Hs*	*Hs*	nd	3227	269	4177	15550	15033	17519	1873	mGluR5
12	Inhibition (%)	*Hs*	*Hs*	nd	226	268	777	15550	15033	17519	1592	A_1R
13	Potency (nM)	*Hs*	*Hs*	nd	1810	266	8173	15550	15033	17519	266	TRHR
14	FC (a.u.)	*Hs*	*Hs*	CHO	226	265	631	15550	15033	1922	1592	A_1R
15	IC_{50} (nM)	*Hs*	*Hs*	nd	3772	229	2256	15550	15033	17519	812	mGluR1
16	EC_{50} (nM)	*Rn*	*Rn*	nd	2564	202	4177	3058	3533	17519	1321	mGluR5
17	IC_{50} (nM)	*Rn*	*Rn*	nd	3067	194	2256	3058	3533	17519	445	mGluR3
18	EC_{50} (nM)	*Rn*	*Rn*	HEK293	2564	191	4177	3058	3533	1709	1321	mGluR5
19	IC_{50} (nM)	*Hs*	*Hs*	nd	3227	189	2256	15550	15033	17519	1873	mGluR5
20	%max (%)	*Hs*	*Hs*	nd	3772	185	1123	15550	15033	17519	812	mGluR1
21	EC_{50} (nM)	*Hs*	*Hs*	nd	216	180	4177	15550	15033	17519	839	M1R
22	Inhibition (%)	*Hs*	*Hs*	CHO	226	172	777	15550	15033	1922	1592	A1R
23	E_{max} (%)	*Hs*	*Hs*	HEK293	3227	166	941	15550	15033	1709	1873	mGluR5
24	IP (nM)	*Hs*	*Hs*	CHO	216	150	204	15550	15033	1922	839	M1R
25	%max (%)	*Rn*	*Rn*	nd	2564	149	1123	3058	3533	17519	1321	mGluR5

[a] *Hs*, *Homo sapiens*; *Rn*, *Rattus norvegicus*. [b] HEK293, Human Embryonic Kidney 293 cells; CHO, Chinese hamster ovary cells. [c] mGluR, Metabotropic glutamate receptor; M1R, Muscarinic acetylcholine receptor 1; A1R, Adenosine A1 receptor; TRHR, Thyrotropin-releasing hormone receptor; C1, Caspase-1; C7, Caspase-7; nd, not determinated.

LDA models are Bayesian methods in the sense that they calculate the posterior probabilities $p(f(v_{ij}) = 1)_{pred}$ taking into consideration the prior probability $p(f(v_{ij}) = 1)_{prior}$.[97] This is the probability with which all compounds presented the desired classification $f(v_{ij})_{pred} = 1$. In this paper the *prior* probability was set as $p(f(v_{ij}) = 1)_{prior} = 0.8$ for the best model found. In fact, the value of χ^2 is 8556.3 with a p-level < 0.05 indicating that the classifier performs a statistically significant separation of both classes ($f(v_{ij})_{obs} = 1$ *vs.* $f(v_{ij})_{obs} = 0$). It can be argued that this kind of multi-condition variable may be less revealing since there seems to be an information lost as all conditions of the same assay are merged together into a single input variable. However, despite this apparently simpler approach, the fact that a lot of information is considered and merged makes this a more powerful model. The activity of a new compound can also be predicted through this model.

Conversely, PT operators, sensible to changes on multiple conditions at the same time, are needed. Thus, the new PT operators are the expected probability $p_j(f(v_{ij}) = 1/\mathbf{c}_j)_{expt}$ and the average values $\langle D_{1,2}(\mathbf{c}_j)\rangle$ for multiple conditions at the same time, see for some examples in **Table 7**.

Table 7. Selected values of 5-fold multi-condition averages.

Multi-condition assays [a]					Multi-condition input parameters [b]			
Activity	Organism	Assay Org.	Cell line	Target	Averages		Count & Probs	
c_0	c_1	c_2	c_3	c_4	$\langle D_1(\mathbf{c}_j)\rangle$	$\langle D_2(\mathbf{c}_j)\rangle$	$n_j(\mathbf{c}_j)$	$f(v_{ij})_{ref}$
Potency (nM)	*H. sapiens*	*H. sapiens*		D_1R	0.49	-51.29	8173	0.140
EC_{50} (nM)	*H. sapiens*	*H. sapiens*	CHO	A_1R	2.80	24.85	4177	0.667
Activity (%)	*R. norvegicus*	*R. norvegicus*	CHO	$mGluR_4$	1.09	-16.76	1500	0.781
EC_{50} (nM)	HHV-5	HHV-5	HEK293	HHV-5 chemokine	0.60	5.20	4177	0.000
Activity (%)	*H. sapiens*	*H. sapiens*	HEK293	CB_1	-0.18	35.10	1500	1.000
%max (%)	*H. sapiens*	*H. sapiens*	CHO-K1	M_4R	-0.65	37.93	1123	0.321
EC_{50} (nM)	*Sus scrofa*	*Sus scrofa*		M_2R	2.26	-17.48	4177	0.417
IC_{50} (nM)	*Mus musculus*	*Mus musculus*		$mGluR_5$	1.26	1.27	2256	0.000
Activity (%)	*R. norvegicus*	*R. norvegicus*	HEK293	NK_2	1.10	-25.98	1500	0.457
Inhibition (%)	*H. sapiens*	*H. sapiens*	HEK293	CB_2	1.72	-12.80	777	0.381

[a] *H. sapiens*, *Homo sapiens*; *R. norvegicus*, *Rattus norvegicus*; HHV-5, Human cytomegalovirus (HHV-5, Human herpesvirus 5); D_1R, Dopamine D_1 receptor; A_1R, Adenosine A_1 receptor; $mGluR_5$ Metabotopic glutamate receptor; HHV-5 chemokine, HHV-5 chemokine receptor; M_4R, Muscarinic acetylcholine receptor M_4; M_2R, Muscarinic acetylcholine receptor M_2; CB_1, Cannabinoid CB_1 receptor; CB_2, Cannabinoid CB_2 receptor; NK_2, Tachykinin receptor 2; HEK293, Human Embryonic Kidney 293 cells; CHO, Chinese hamster

ovary cells; CHO-K1, A subclone of the parental CHO cell line. [b] Expected probability of biological activity $p = p(f(v_{ij})= 1/\mathbf{c}_j)_{ref} = n(f(v_{ij})= 1/\mathbf{c}_j)/ n(\mathbf{c}_j)$, where $n(f(v_{ij})= 1/\mathbf{c}_j)$ is the number of cases with $f(v_{ij})= 1$ and measured under the set of conditions $\mathbf{c}_j$. The denominator is the number of cases $n(\mathbf{c}_j)$ measured under the set of conditions $\mathbf{c}_j$ irrespective of the value of $f(v_{ij})= 1$ or 0. Please, note that multi-condition parameters depend on a vector of conditions $\mathbf{c}_j$ (denoted in **boldface**) and not on a single condition c_j.

Computational study on the 2-furoyl-based MIF-1 peptidomimetics.

After training and validating the ALLOPTML model, we decided to illustrate how to use it by providing a detailed step-by-step guide to calculate the posterior probabilities with a practical example. In so doing, the series of peptidomimetics **4-7(a,b)** were used for this purpose in different assay conditions $\mathbf{c}_j$ (target proteins, cell lines, assay organisms). Firstly, the chemical structures of peptidomimetics **4-7(a,b)** were drawn using ChemDraw software to obtain their SMILE codes. Then, the file containing all SMILE codes was created and uploaded to Virtual Computational Chemistry Laboratory (VCCLAB, http://www.vcclab.org, 2005)[104] in order to calculate the molecular descriptors D_k (LogP, PSA) for each compound. Since the goal was to illustrate how to operate the model only a few assay conditions were selected. After that, the function of reference $f(v_{ij})_{ref}$ (expected probability of activity) values for each assay were obtained from Supporting Information file (SI00.xlsx). Using this file, the expected or average values $<D_k(\mathbf{c}_j)>$ of the molecular descriptors D_k were also obtained for all compounds assayed on these conditions. With these values in hand, we calculated the MMAs operators $\Delta D_k(\mathbf{c}_j)$ for each compound *vs*. every set of experimental conditions selected. Subsequently, $f(v_{ij})_{ref}$ and $\Delta D_k(\mathbf{c}_j)$ values were substituted into the equation of the model to obtain the values of the output function $f(v_{ij})_{calc}$. Lastly, these values were used to calculate the posterior probabilities $p(f(v_{ij})_{pred} = 1)$; with which the compounds may be considered as candidates for assay according to our criteria. The equation used to calculate this probabilities was a sigmoid function $p(f(v_{ij})_{pred} = 1) = 1 /(1 + (\pi_0/\pi_1)\cdot Exp(-f(v_{ij})_{calc})$. Recall that $\pi_1 = 1 - \pi_0$ are the probabilities (defined before to seek the model) with which a compound may considered as active $f(v_{ij})_{pred} = 1$ or non-active and $f(v_{ij})_{pred} = 0$, *a priori*.

In **Table 8** are listed the results obtained for the selected examples including different c_0 = Activity parameter (units) such as Effective 50% concentration EC_{50}(nM), Maximun Effect E_{max}(%), and Inhibition(%) = Inhibition(%). The examples also involve two different organisms (Human and Rat) expressing the protein target and organisms of assays (c_1 and c_2). We also changed the condition c_3 = cell line of assay (HEK293, CHO) and the condition c_4 = target proteins (mGluR5 and mGluR2). Interestingly, the model predicts a similar behavior for all the series of compounds in the same conditions, which is reasonably for a homogeneous series of compounds. The model predicts low probabilities of having interesting EC_{50}(nM) values for mGluR5, but high probabilities of Inhibition(%) for the same receptor. This may be indicative that this set of peptidomimetics display activity on human mGluR5 but perhaps at higher concentrations.

Table 8. PTML predictive study of new compounds.

	c_0	EC_{50}(nM)	E_{max}(%)	$\%_{max}$	Inhibition(%)
	c_1	*H. sapiens*	*H. sapiens*	*R. norvegicus*	*H. sapiens*
Assay	c_2	*H. sapiens*	*H. sapiens*	*R. norvegicus*	*H. sapiens*
	c_3	HEK293	CHO	null	HEK293
	c_4	mGluR5	mGluR2	mGluR2	mGluR5
Cmpnd	**4a**	0.042	0.087	0.404	0.796
	4b	0.039	0.082	0.388	0.785
	5a	0.043	0.090	0.413	0.802
	5b	0.041	0.085	0.397	0.791
	6a	0.041	0.085	0.397	0.791
	6b	0.038	0.078	0.378	0.778
	7a	0.042	0.087	0.403	0.796
	7b	0.040	0.083	0.392	0.787
	MIF-1	0.036	0.076	0.368	0.770

mGluR5 = Metabotropic glutamate receptor 5 (CHEMBL2564 in Rat and CHEMBL3227 in Human);
mGlu2 = Metabotropic glutamate receptor 2 (CHEMBL5137 in Human and CHEMBL2581 in Rat).

3. Conclusions

The use of **2-Fu** for the assembly of MIF-1 peptidomimetics is unprecedented. In this work, a small library of MIF-1 peptidomimetics bearing **2-Fu** as proline surrogate was successfully obtained and pharmacologically evaluated towards D_2R. Remarkably, compound **6a** stands out as D_2R modulator by increasing in the maximal [^{3}H]-NPA response at 10 pM (11 ± 1%), comparable to the effect of MIF-1 (22 ± 9%) at the same concentration with no neurotoxic effect up to 100 μM in cortex neurons of Wistar-Kyoto rat embryos. These findings corroborate that D_2R allosteric binding site can accommodate heteroaromatic rings at N-terminal position of MIF-1, establishing furan as a suitable heterocyclic scaffold to be used as proline surrogate for the assembly of MIF-1 analogues. Consequently, **6a** is considered a valid lead compound for further optimization, paving the way for the discovery of adequate candidates to treat of CNS-related diseases. Exploration of other furan motifs are currently underway in our research group and will be disclosed in due time.

The topology of a complex network of allosteric modulators was constructed and studied delivering a predictive model, ALLOPTML, which shows remarkable specificity Sp = 89.2/89.4% and sensitivity Sn = 71.3/72.2% in training/validation series. The linear model with multi-condition MAs proved to be superior to a linear model with simple condition MAs. To the best of our knowledge, ALLOPTML is the first general-purpose medicinal chemistry tool using PTML model for the multi-output and multi-condition prediction of allosteric compounds, which is expected to be a valuable chemoinformatic tool to assist in the early drug discovery, saving both time and resources.

4. Methods

Computational Methods

ChEMBL Data pre-processing.

Firstly, a large and complex data set of preclinical assays of allosteric modulators from ChEMBL database was downloaded. The data include molecular descriptors of the structure of the drugs D_k (input continuous variables) and several discrete variables for assay specifications. We used a 2D arrange $\mathbf{c}_j$ of 12 different experimental conditions of the assay to summarize all the specifications of the preclinical assays (discrete variables) in the data set. These conditions include, but are not limited to, 38 biological activity parameters $\mathbf{c}_0$ with levels: $c_{0,1}$ = efficacy, $c_{0,2}$ = potency, $c_{0,3}$ = IC_{50}, *etc*. The arrange also includes different organisms $\mathbf{c}_1$, with levels $c_{1,1}$ = human, $c_{1,2}$ = rat, $c_{1,3}$ = mouse, *etc*. Another experimental condition is cellular lines used in the assay $\mathbf{c}_2$, with levels $c_{2,1}$ = CHO, $c_{2,2}$ = HEK293, *etc*., and so on. Next, we used as input variables different PT operators with form of multi-condition moving averages $\Delta D_k(\mathbf{c}_j)$ to train the model.

As we stated above, the first condition $\mathbf{c}_j$ is the type of biological activity measured with values v_{ij} levels (IC_{50}, EC_{50}, *etc*.). The values of v_{ij} compiled are not exact numbers in many cases. In fact, many of them appears in ChEMBL with a >, <, and/or ≈ symbol. Eliminating this data could lead to an important reduction in the large of the dataset with the consequent loss of biologically relevant information. In addition, many of the biological activity values v_{ij} are desired to be as high as possible; e.g., Inhibition(%). However, other values of the biological activity values v_{ij} are desired to be as low as possible; e.g., IC_{50}, EC_{50}, *etc*. That is why we decided to use classification techniques instead of regression methods. Using a classification technique, the original values v_{ij} can be assigned into two one of two possible classes (active or interesting compounds *vs*. not active or control group). In so doing, the real numerical values of v_{ij} were converted into a Boolean variable $f(v_{ij})_{obs}$. In order to construct this variable, two parameters called cutoff(c_0) and desirability $d(c_0)$, were defined.

The cutoff(c_0) is the threshold value to consider v_{ij} as high or low. The desirability gets the values $d(c_0)$ = 1 or -1 for properties that are desired to be as high or as low as possible, respectively. After that, the function $f(v_{ij})_{obs}$ is calculated as $f(v_{ij})_{obs}$ = 1 when v_{ij} > cutoff and $d(c_0)$ = 1 (see **Table 2**). The function is also $f(v_{ij})_{obs}$ = 1 when v_{ij} < cutoff and $d(c_0)$ = -1; $f(v_{ij})_{obs}$ = 0 otherwise. The value $f(v_{ij})_{obs}$ = 1 points to and strong effect of the compound over the target.

PTML linear model. PTML modeling technique is useful to seek predictive models for complex datasets with multiple BD features.[105, 106] We can predict scoring function values $f(v_{ij})_{calc}$ for the i^{-th} compound in the j^{-th} preclinical assay with an specific sub-set of conditions of assay selected out of multiple conditions of assay $\mathbf{c}_j$ = (c_0, c_1, c_2, …c_{jmax}). PT operators similar to Box-Jenkins MA operators are used as input.[71, 79] PTML linear models have the following form:

$$f(v_{ij})_{calc} = a_0 + a_1 \cdot f(v_{ij})_{ref} + \sum_{k=1,j=0}^{k_{max},j_{max}} a_{kj} \cdot \Delta D_k(\mathbf{c}_j) \quad (3)$$

Experimental Methods

Chemistry. All chemicals were of reagent grade and used without further purifications: 2-furoic acid was obtained from Alfa Aesar; *O*-(Benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium tetrafluoroborate (TBTU), H-L-Leu-OMe, H-L-Val-OMe, H-L-Ala-OMe and H-Gly-OMe were obtained from Bachem and *N*,*N*-diisopropylethylamine (DIEA) was obtained from Sigma-Aldrich. All air sensitive reactions were carried out under argon atmosphere. Analytical TLC was carried out on precoated silica gel plates (Merck 60 F_{254}, 0.25 mm) using UV light and an ethanolic solution of phosphomolybdic acid (followed by gentle heating) for visualization. Flash chromatography was performed on silica gel (Merck 60, 230-240 mesh).

Apparatus. Mass spectra were recorded on a LTQ OrbitrapTM XL hybrid mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) controlled by LTQ Tune Plus 2.5.5 and Xcalibur 2.1.0 (Centro de Materiais da Universidade do Porto, CEMUP). The capillary voltage of the electrospray ionization source (ESI) was set to 3.1 kV. The capillary temperature was set at 275 °C. The sheath gas was set at 6 (arbitrary unit as provided by the software settings). The capillary voltage was set at 35 V and the tube lens voltage set at 110 V. ^{1}H- and ^{13}C-NMR spectra were recorded at CEMUP with a Bruker Avance III 400 at 400.15 MHz and 100.62 MHz, respectively. The NMR spectra were calibrated using residual solvents signals ($CDCl_3$: δ_H = 7.26, δ_C = 77.16; CD_3OD: δ_H = 3.31, δ_C = 49.00 and DMSO-d_6: δ_H = 2.50, δ_C = 39.52)[107] and are reported in ppm. The nomenclature used for the assignment of protons and/or carbons for each α-amino acid residue in peptidomimetics was made using a single letter system in subscript for the amino acid residue (Leu: L-leucine; Val: L-valine, Ala: L-alanine; Gly: glycine) and indicating the proton (or group of protons) and/or the carbons in the structures by starting the numeration at the carbonyl carbon of the main chain of each α-amino acid residue. Optical rotations were measured on a JASCO P-2000 thermostated polarimeter using a sodium lamp and are reported as follows: α_D^{θ} expressed in (º) (dm^{-1}) (g^{-1}), in which θ is temperature in Celsius and *c* (g / 100 mL, solvent). Melting points were determined using a STUART Scientific, model SMP1, and are not corrected. Solvents were evaporated in a Büchi rotavapor model R-210.

General protocol for peptide coupling – Synthesis of 2(a,b), 4(a,b), and 5(a,b). The appropriate carboxylic acid (1 equiv.) was charged in a round bottom flask and dissolved in anhydrous CH_2Cl_2 (50 mL), under argon atmosphere, followed by iterative addition of DIEA (3 equiv.) and TBTU (1.1 equiv.), with stirring for 30 min at RT. Finally, the appropriate amine (1.2 equiv.) was added and the reaction was left stirring for approximately 1 h. After completion of the reaction (TLC), CH_2Cl_2 was removed *in vacuo* and the crude oil was dissolved with EtOAc (100 mL) and transferred to a separatory funnel. The organic layer was washed with saturated solutions of $NaHCO_3$ (3 x 100 mL) and NaCl (100 mL) and dried over anhydrous Na_2SO_4. After filtration, the solvent was removed *in vacuo* and the crude oil was chromatographed as specified for each peptide product.

General protocol for saponification – Synthesis of pseudodipeptides 3(a,b). Pseudodipeptide methyl ester (1 equiv.) was charged in a round bottom flask, dissolved in MeOH (50 mL), and the solution was cooled at 0 ºC using an ice bath with magnetic stirring. Then, LiOH (3 equiv.) were slowly added during a period of 30 min and left stirring until complete consumption of the starting material (TLC). The solvent was removed *in vacuo* (at 40 ºC) and the crude solid was then dissolved in water (20 mL), cooled at 0 ºC using an ice bath with magnetic stirring. After that, 1 M H_2SO_4 was carefully added until pH = 4. The solvent was distilled off (at 40 ºC) and the crude solid was then triturated with hot $CHCl_3$ and filtered. Removal of the volatiles were performed in a rotatory evaporator to afford the desired compound without further purifications.

General protocol for C-terminal carboxamide – Synthesis of pseudotripeptides 6(a,b) and 7(a,b). Pseudotripeptide methyl ester (1 equiv.) was charged in a reaction flask and dissolved in MeOH p.a. (40 mL). The solution was cooled at 0 ºC using an ice bath and gaseous ammonia (generated *in situ* by mixing NH_4Cl and $Ca(OH)_2$ 2:1 (w/w) at 100 ºC using an oil bath) was bubbled into the reaction flask. The ammonia stream was maintained for 1 h and the reaction was left stirring at room temperature (*ca*. 25 ºC) until completion (*ca*. 48 h, TLC). Removal of the volatiles were performed in a rotatory evaporator to afford the desired compound without the need of chromatographic purification.

Methyl 2-furoyl-L-leucinate (2a). Following the general protocol for peptide coupling, to a solution of **2-Fu** (1.00 g, 8.92 mmol) in anhydrous CH_2Cl_2 (30 mL) was added DIEA (4.66 mL, 26.8 mmol), TBTU (3.15 g, 9.81 mmol) and methyl L-leucinate hydrochloride (1.94 g, 10.7 mmol). After the typical work-up, the crude oil was chromatographed using EtOAc as eluent, affording 2.05 g of **2a** as a white solid. Yield: 96%. **m.p.**: 83–87 ºC. $\mathbf{R_f}$: 0.84 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{24}}$: +8.3 (*c*1, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm: 7.56 – 7.36 (m, 1H, H-5), 7.23 – 7.00 (m, 1H, H-3), 6.77 (d, *J* = 8.0 Hz, 1H, CONH), 6.59 – 6.38 (m, 1H, H-4), 4.82 (dt, *J* = 8.6, 4.9 Hz, 1H, H_{Leu}-2), 3.76 (s, 3H, CO_2CH_3), 1.89 – 1.52 (m, 3H, H_{Leu}-3 + H_{Leu}-4), 1.11 – 0.82 (m, 6H, H_{Leu}-5). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm: 173.4 (C, CO_2CH_3), 158.1 (C, CONH), 147.6 (C, C-2), 144.2 (CH, C-5), 114.8 (CH, C-3), 112.3 (CH, C-4), 52.4 (CH_3, CO_2CH_3), 50.4 (CH, C_{Leu}-2), 41.9 (CH_2, C_{Leu}-3), 25.0 (CH, C_{Leu}-4), 22.9 (CH_3, C_{Leu}-5), 22.0 (CH_3, C_{Leu}-5). **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{12}H_{18}NO_4^+$: 240.12303; Found: 240.12312.

Methyl 2-furoyl-L-valinate (2b). Following the general protocol for peptide coupling, to a solution of **2-Fu** (1.5022 g, 13.403 mmol) in anhydrous CH_2Cl_2 (30 mL) was added DIEA (7.00 mL, 40.2 mmol), TBTU (4.7340 g, 14.743 mmol) and methyl L-valinate hydrochloride (2.6960 g, 16.084 mmol). After the typical work-up, the crude oil was chromatographed using EtOAc as eluent, affording 2.8981 g of **2b** as a white solid. Yield: 96%. **m.p.**: 75–78 ºC. $\mathbf{R_f}$: 0.89 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{21}}$: +9.0 (*c*1.015, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm: 7.45 (dd, *J* = 1.8, 0.8 Hz, 1H, H-5), 7.10 (dd, *J* = 3.5, 0.8 Hz, 1H, H-3), 6.80 (d, *J* = 8.3 Hz, 1H, CONH), 6.48 (dd, *J* = 3.5, 1.8 Hz, 1H, H-4), 4.70 (dd, *J* = 9.0, 5.0 Hz, 1H, H_{Val}-2), 3.74 (s, 3H, CO_2CH_3), 2.24 (dhept, *J* = 6.9, 5.1 Hz, 1H, H_{Val}-3), [0.98 (d, *J* = 6.9 Hz, 3H), 0.95 (d, *J* = 6.9 Hz, 3H), H_{Val}-4]. **^{13}C-NMR and DEPT**

($CDCl_3$, 101 MHz) δ ppm: 172.4 (C, CO_2CH_3), 158.2 (C, CONH), 147.7 (C, C-2), 144.2 (CH, C-5), 114.8 (CH, C-3), 112.3 (CH, C-4), 56.8 (CH_3, CO_2CH_3), 52.3 (CH, C_{Val}-2), 31.6 (CH, C_{Val}-3), 19.1 (CH, C_{Val}-4), 17.9 (CH, C_{Val}-4). **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{11}H_{16}NO_4^+$: 226.10738; Found: 226.10738.

2-Furoyl-L-leucine (3a). Following the general protocol for saponification, to a solution of **2a** (0.8034 g, 3.360 mmol) in MeOH (30 mL) was added LiOH (0.2414 g, 10.08 mmol). After the typical work-up, 0.7038 g of **3a** was obtained as a white solid. Yield: 93%. **m.p.**: 81–84 ºC. $\mathbf{R}_f$: 0.09 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{24}}$: +8.3 (*c*1, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm: 8.95 (br s, 1H, CO_2H), 7.41 (d, *J* = 1.3 Hz, 1H, H-5), 7.10 (d, *J* = 3.4 Hz, 1H, H-3), 7.06 (d, *J* = 7.9 Hz, 1H, CONH), 6.42 (dd, *J* = 3.4, 1.7 Hz, 1H, H-4), 4.76 – 4.61 (m, 1H, H_{Leu}-2), 1.80 – 1.57 (m, 3H, H_{Leu}-3 + H_{Leu}-4), 1.10 – 0.80 (m, 6H, H_{Leu}-5). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm: 177.6 (C, CO_2H), 158.7 (C, CONH), 147.3 (C, C-2), 144.6 (CH, C-5), 115.3 (CH, C-3), 112.3 (CH, C-4), 51.6 (CH, C_{Leu}-2), 41.5 (CH_2, C_{Leu}-3), 25.0 (CH, C_{Leu}-4), 23.1 (CH_3, C_{Leu}-5), 21.8 (CH_3, C_{Leu}-5). **HRMS** (ESI-TOF) *m*/*z*: $[M - H]^-$ Calcd for $C_{11}H_{14}NO_4^-$: 224.09283; Found: 224.09482.

2-Furoyl-L-valine (3b). Following the general protocol for saponification, to a solution of **2b** (0.8019, 3.560 mmol) in MeOH (30 mL) was added LiOH (0.2558 g, 10.68 mmol). After the typical work-up, 0.6843 g of **3b** was obtained as a white solid. Yield: 91%. **m.p.**: 65–70 ºC. $\mathbf{R}_f$: 0.05 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{20}}$: +67.5 (*c*1.09, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm: 10.12 (br s, 1H, CO_2H), 7.42 (d, *J* = 0.9 Hz, 1H, H-5), 7.10 (d, *J* = 3.3 Hz, 1H, H-3), 7.04 (d, *J* = 8.5 Hz, 1H, CONH), 6.42 (dd, *J* = 3.4, 1.7 Hz, 1H, H-4), 4.60 (dd, *J* = 8.5, 5.1 Hz, 1H, H_{Val}-2), 2.37 – 2.13 (m, 1H, H_{Val}-3), 0.95 (d, *J* = 7.1 Hz, 3H, H_{Val}-4), 0.93 (d, *J* = 7.1 Hz, 3H, H_{Val}-4). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm: 176.2 (C, CO_2H), 158.8 (C, CONH), 147.3 (C, C-2), 144.6 (CH, C-5), 115.3 (CH, C-3), 112.3 (CH, C-4), 57.9 (CH, C_{Val}-2), 31.3 (CH, C_{Val}-3), 19.3 (CH, C_{Val}-4), 17.9 (CH, C_{Val}-4). **HRMS** (ESI-TOF) *m*/*z*: $[M - H]^-$ Calcd for $C_{10}H_{12}NO_4^-$: 210.07718; Found: 210.07777.

Methyl 2-furoyl-L-leucylglycinate (4a). Following the general protocol for peptide coupling, to a solution of **3a** (0.8767 g, 3.944 mmol) in anhydrous CH_2Cl_2 (30 mL) was added DIEA (2.06 mL, 11.8 mmol), TBTU (1.3933 g, 4.3393 mmol) and methyl glycinate hydrochloride (0.5942 g, 4.733 mmol). After the typical work-up, the crude oil was chromatographed using EtOAc as eluent, affording 1.1394 g of **4a** as a white solid. Yield: 94%. **m.p.**: 135–136 ºC. $\mathbf{R}_f$: 0.58 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{20}}$: −8.5 (*c*1, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm: 7.44 – 7.40 (m, 1H, H-5), 7.26 – 7.12 (m, 1H, CONH), 7.12 – 7.06 (m, 1H, H-3), 7.04 – 6.86 (m, 1H, CONH), 6.46 (dt, *J* = 4.8, 1.7 Hz, 1H, H-4), 4.82 – 4.64 (m, 1H, H_{Leu}-2), 4.13 – 3.87 (m, 2H, H_{Gly}-2), 3.76 – 3.63 [3.70 (s), 3.69 (s), 3H, CO_2CH_3], 1.87 – 1.53 (m, 3H, H_{Leu}-3 + H_{Leu}-4), 1.04 – 0.79 (m, 6H, H_{Leu}-5). **^{13}C-NMR e DEPT** ($CDCl_3$, 101 MHz) δ ppm: [172.5 (C), 172.4 (C), 170.2 (C), CONH + CO_2CH_3], 158.5 (C, CONH), 147.4 (C, C-2), 144.5 (CH, C-5), 115.0 (CH, C-3), 112.3 (CH, C-4), 52.4 (CH_3, CO_2CH_3), 51.3 (CH, C_{Leu}-2), 41.3 (2 CH_2, C_{Gly}-2 + C_{Leu}-3), 24.8 (CH, C_{Leu}-4), 23.0 (CH_3, C_{Leu}-5), 22.1 (CH_3, C_{Leu}-5). **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{14}H_{21}N_2O_5^+$: 297.14450; Found: 297.14377.

Methyl 2-furoyl-L-valylglycinate (4b). Following the general protocol for peptide coupling, to a solution of **3b** (1.5004 g, 7.1035 mmol) in anhydrous CH_2Cl_2 (30 mL) was added DIEA (3.70 mL, 21.3 mmol), TBTU (2.5089 g, 7.8138 mmol) and methyl glycinate hydrochloride (1.0701 g, 8.5242 mmol). After the typical work-up, the crude oil was chromatographed using EtOAc as eluent, affording 1.5842 g of **4b** as a white solid. Yield: 79%. **m.p.**: 133–136 ºC. $\mathbf{R_f}$: 0.60 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{22}}$: −38.4 (*c*1.015, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm: 7.50 – 7.40 (m, 1H, H-5), 7.32 (br s, 1H, CONH), 7.20 – 7.00 [7.11 (dd, *J* = 6.9, 4.2 Hz), 7.09 (br s), H-3 + CONH], 6.49 (dd, *J* = 3.4, 1.7 Hz, 1H, H-4), 4.63 – 4.53 (m, 1H, H_{Val}-2), 4.20 – 3.90 (m, 2H, H_{Gly}-2), 3.73 (s, 3H, CO_2CH_3), 2.28 – 2.15 (m, 1H, H_{Val}-3), 1.12 – 0.92 (m, 6H, H_{Val}-4). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm: [171.6 (C), 170.2, CO_2CH_3 + CONH], 158.5 (C, CONH), 147.5 (C, C-2), 144.5 (CH, C-5), 114.8 (CH, C-3), 112.2 (CH, C-4), 58.1 (CH, C_{Val}-2), 52.4 (CH_3, CO_2CH_3), 41.2 (CH_2, C_{Gly}-2), 31.5 (CH, C_{Val}-3), 19.3 (CH, C_{Val}-4), 18.3 (CH, C_{Val}-4). **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{13}H_{19}N_2O_5^+$: 283.12885; Found: 283.12878.

Methyl 2-furoyl-L-leucyl-L-alaninate (5a). Following the general protocol for peptide coupling, to a solution of **3a** (1.6678 g, 7.5045 mmol) in anhydrous CH_2Cl_2 (30 mL) was added DIEA (3.92 mL, 22.5 mmol), TBTU (2.6506 g, 8.2550 mmol) and methyl L-alaninate hydrochloride (1.2570 g, 9.0054 mmol). After the typical work-up, the crude oil was chromatographed using EtOAc as eluent, affording 1.9797 g of **5a** as a white solid. Yield: 85%. **m.p.**: 135–136 ºC. $\mathbf{R_f}$: 0.58 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{22}}$: +83.7 (*c*1.16, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm (rotamers present 60:40): [7.43 (dd, *J* = 1.7, 0.8 Hz, minor), 7.41 (dd, *J* = 1.7, 0.8 Hz, major), 1H, H-5], [7.11 (dd, *J* = 3.5, 0.7 Hz, minor), 7.09 (dd, *J* = 3.5, 0.7 Hz, major), 7.07 (br s), 2H, H-3 + CONH], [6.95 (d, *J* = 8.6 Hz, major), 6.91 (d, *J* = 8.6 Hz, minor), 1H, CONH], 6.51 – 6.41 (m, 1H, H-4), 4.78 – 4.66 (m, 1H, H_{Leu}-2), 4.60 – 4.47 (m, 1H, H_{Ala}-2), [3.72 (s, major), 3.66 (s, minor), 3H, CO_2CH_3], 1.78 – 1.59 (m, 3H, H_{Leu}-3 + H_{Leu}-4), [1.40 (d, *J* = 7.2 Hz, minor), 1.36 (d, *J* = 7.2 Hz, major), 3H, H_{Ala}-3], 0.94 – 0.90 (m, 6H, H_{Leu}-5). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm: [173.3 (C), 173.1 (C), 171.8 (C), 171.6 (C), CONH + CO_2CH_3], [158.4 (C), 158.3 (C), CONH], [147.6 (C), 147.5 (C), C-2], 144.4 (CH, C-5), [114.9 (CH), 114.8 (CH), C-3], [112.3 (CH), 112.2 (CH), C-4], 52.5 (CH_3, CO_2CH_3), [51.3 (CH), 51.2 (CH), C_{Leu}-2], [48.2 (CH), 48.2 (CH), C_{Ala}-2], [41.7 (CH_2), 41.5 (CH_2), C_{Leu}-3], [24.9 (CH), 24.8 (CH), C_{Leu}-4], [23.0 (CH_3), 23.0 (CH_3), C_{Ala}-3], [22.2 (CH_3), 22.2 (CH_3), 18.2 (CH_3), 18.0 (CH_3), C_{Leu}-5]. **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{15}H_{23}N_2O_5^+$: 311.16015; Found: 311.15826.

Methyl 2-furoyl-L-valyl-L-alaninate (5b). Following the general protocol for peptide coupling, to a solution of **3b** (0.5012 g, 2.373 mmol) in anhydrous CH_2Cl_2 (30 mL) was added DIEA (1.24 mL, 7.12 mmol), TBTU (0.8381 g, 2.610 mmol) and methyl L-alaninate hydrochloride (0.3975 g, 2.848 mmol). After the typical work-up, the crude oil was chromatographed using EtOAc as eluent, affording 0.6258 g of **5b** as a white solid. Yield: 89%. **m.p.**: 130–132 ºC. $\mathbf{R_f}$: 0.69 in EtOAc. $[\boldsymbol{\alpha}]_{\mathbf{D}}^{\mathbf{20}}$: −8.1 (*c*1.01, DMSO). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm (rotamers present 60:40): 7.44 – 7.40 (m, 1H, H-5), 7.40 – 7.26 [7.38 (d, *J* = 6.6 Hz, major), 7.29 (d, *J* = 6.9 Hz, minor), 1H, CONH], 7.19 – 6.99 [7.13 (d, *J* = 9.0 Hz), H-3 + CONH], 6.50 – 6.39 (m, 1H, H-4), 4.68 – 4.46

(m, 2H, H_{Val}-2 + H_{Ala}-2), [3.71 (s, major), 3.65 (s, minor), 3H, CO_2CH_3], 2.26 – 2.04 (m, 1H, H_{Val}-3), [1.40 (d, *J* = 7.2 Hz, minor), 1.36 (d, *J* = 7.3 Hz, major), 3H, H_{Ala}-3], 1.04 – 0.90 (m, 6H, H_{Val}-4). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm (rotamers present): [173.3 (C), 173.0 (C), 170.9 (C), 170.7 (C), CO_2CH_3 + CONH], [158.3 (C), 158.3 (C), CONH], [147.6 (C), 147.6 (C), C-2], [144.4 (CH), 144.4 (CH), C-5], [114.7 (CH), 114.7 (CH), C-3], [112.2 CH), 112.2 (CH), C-4], [57.8 (CH), 57.8 (CH), C_{Val}-2], [52.4 (CH_3), 52.4 (CH_3), CO_2CH_3], [48.2 (CH), 48.1 (CH), C_{Ala}-2], [31.9 (CH), 31.7 (CH), C_{Val}-3], [19.3 (CH_3), 19.1 (CH_3), 18.3 (CH_3), 18.2 (CH_3), 18.1 (CH_3), 17.8 (CH_3), C_{Val}-4 + C_{Ala}-3]. **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{14}H_{21}N_2O_5^+$: 297.14450; Found: 297.14463.

Methyl 2-furoyl-L-leucylglycinamide (6a). Following the general protocol for the synthesis of primary amide, gaseous ammonia was bubbled into a solution of **4a** (0.3090 g, 1.043 mmol) in MeOH p.a. (20 mL) and the system was left stirring for 48 h at rt. After the typical work-up, 0.2921 g of **6a** was obtained as a beige solid. Yield: quantitative. **m.p.**: 90–95 ºC. **R_f**: 0.05 in EtOAc. $[\alpha]_D^{20}$: +161.2 (*c*1, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm: 8.01 – 7.77 (m, 1H, CONH), 7.49 – 7.32 (m, 2H, H-5 + CONH), 7.12 – 7.03 (m, 1H, H-3), 6.94 (br s, 1H, $CONH_2$), 6.48 (br s, 1H, $CONH_2$), 6.42 (dd, *J* = 3.1, 1.5 Hz, 1H, H-4), 4.74 – 4.50 (m, 1H, H_{Leu}-2), 3.98 (dd, *J* = 16.9, 6.1 Hz, 1H, H_{Gly}-2), 3.87 – 3.66 (m, 1H, H_{Gly}-2), 1.80 – 1.58 (m, 3H, H_{Leu}-3 + H_{Leu}-4), 0.99 – 0.81 (m, 6H, H_{Leu}-5). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm: [173.3 (C), 173.3 (C), 172.5 (C), 159.1 (C), 2 x CONH + $CONH_2$], 147.2 (C, C-2), 144.9 (CH, C-5), 115.1 (CH, C-3), 112.3 (CH, C-4), 52.3 (CH, C_{Leu}-2), 42.9 (CH_2, C_{Leu}-3), 41.0 (CH_2, C_{Gly}-2), 24.9 (CH, C_{Leu}-4), 23.0 (CH_3, C_{Leu}-5), 21.9 (CH_3, C_{Leu}-5). **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{13}H_{20}N_3O_4^+$: 282.14483; Found: 282.14417.

Methyl 2-furoyl-L-valylglycinamide (6b). Following the general protocol for the synthesis of primary amide, gaseous ammonia was bubbled into a solution of **4b** (0.3819 g, 1.353 mmol) in MeOH p.a. (20 mL) and the system was left stirring for 48 h at rt. After the typical work-up, 0.3611 g of **6b** was obtained as a white solid. Yield: quantitative. **m.p.**: 180–183 ºC. **R_f**: 0.20 in EtOAc. $[\alpha]_D^{21}$: +14.6 (*c*1.025, DMSO). **^{1}H-NMR** (DMSO-d_6, 400 MHz) δ ppm: 8.30 (t, *J* = 5.8 Hz, 1H, CONH), 8.04 (d, *J* = 8.4 Hz, 1H, CONH), 7.87 – 7.82 (m, 1H, H-5), 7.29 – 7.15 [7.23 (br s), 7.21 (d, *J* = 3.5 Hz), 2H, CONHH + H-3], 7.04 (br s, 1H, CONHH), 6.63 (dd, *J* = 3.5, 1.8 Hz, 1H, H-4), 4.25 (dd, *J* = 8.1, 7.7 Hz, 1H, H_{Val}-2), [3.66 (dd, *J* = 16.7, 6.0 Hz), 3.65 (dd, *J* = 16.7, 5.5 Hz), 2H, H_{Gly}-2], 2.18 – 2.04 (m, 1H, H_{Val}-3), [0.91 (d, *J* = 6.8 Hz), 0.90 (d, *J* = 6.7 Hz), 6H, H_{Val}-4]. **^{13}C-NMR and DEPT** (DMSO-d_6, 101 MHz) δ ppm: [171.1 (C), 170.7 (C), 157.8 (C), 2 x CONH + $CONH_2$], 147.4 (C, C-2), 145.2 (CH, C-5), 114.0 (CH, C-3), 111.9 (CH, C-4), 58.3 (CH, C_{Val}-2), 41.8 (CH_2, C_{Gly}-2), 30.1 (CH, C_{Val}-3), [19.3 (CH_3), 18.6 (CH_3), C_{Val}-4]. **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{12}H_{18}N_3O_4^+$: 268.12918; Found: 268.12884.

Methyl 2-furoyl-L-leucyl-L-alaninamide (7a). Following the general protocol for the synthesis of primary amide, gaseous ammonia was bubbled into a solution of **5a** (0.5552 g, 1.788 mmol) in MeOH p.a. (20 mL) and the system was left stirring for 48 h at rt. After the typical work-up, 0.5279 g of **7a** was obtained as a white

solid. Yield: quantitative. **m.p.**: 85–90 ºC. **R_f**: 0.27 in EtOAc. $[\alpha]_D^{23}$: −36.6 (*c*1.015, $CHCl_3$). **^{1}H-NMR** ($CDCl_3$, 400 MHz) δ ppm (rotamers present 70:30): [7.84 (d, *J* = 7.8 Hz, minor), 7.76 (d, *J* = 7.5 Hz, major), 1H, CONH], 7.47 – 7.20 [7.44 (dd, *J* = 1.7, 0.7 Hz), 2H, H-5 + CONH], [7.13 (dd, *J* = 3.5, 0.6 Hz, major), 7.10 (dd, *J* = 3.6, 0.6 Hz, minor), 1H, H-3], [7.02 (br s, minor), 6.94 (br s, major), 1H, CONHH], 6.63 – 6.19 [6.42 (br s, minor), 6.38 (br s, major), 2H, H-4 + CONHH], 4.81 – 4.63 (m, 1H, H_{Leu}-2), [4.53 (p, *J* = 7.0 Hz), 4.52 (p, *J* = 7.1 Hz), 1H, H_{Ala}-2], 1.76 – 1.63 (m, 3H, H_{Leu}-3 + H_{Leu}-4), [1.39 (d, *J* = 7.1 Hz, minor), 1.33 (d, *J* = 7.1 Hz, major), 3H, H_{Ala}-3], 0.98 – 0.87 (m, 6H, H_{Leu}-5). **^{13}C-NMR and DEPT** ($CDCl_3$, 101 MHz) δ ppm (rotamers present): [176.3 (C), 176.2 (C), 173.4 (C), 173.2 (C), 159.8 (C), 159.5 (C), 2 x CONH + $CONH_2$], [148.2 (C), 148.0 (C), C-2], [145.5 (CH), 145.4 (CH), C-5], [115.9 (CH), 115.8 (CH), C-3], 113.1 (CH, C-4), [53.0 (CH), 52.6 (CH), C_{Leu}-2], [49.7 (CH), 49.6 (CH), C_{Ala}-2], [42.3 (CH_2), 42.1 (CH_2), C_{Leu}-3], [25.7 (CH), 25.7 (CH), C_{Leu}-4], [23.8 (CH_3), 23.7 (CH_3), 22.9 (CH_3), 22.8 (CH_3), C_{Leu}-5], [18.8 (CH_3), 18.7 (CH_3), C_{Ala}-3]. **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{14}H_{22}N_3O_4^+$: 296.16048; Found: 296.16033.

Methyl 2-furoyl-L-valyl-L-alaninamide (7b). Following the general protocol for the synthesis of primary amide, gaseous ammonia was bubbled into a solution of **5b** (0.4012 g, 1.353 mmol) in MeOH p.a. (20 mL) and the system was left stirring for 48 h at rt. After the typical work-up, 0.3580 g of **7b** was obtained as a white solid. Yield: 94%. **m.p.**: 202–204 ºC. **R_f**: 0.14 in Et_2O. $[\alpha]_D^{21}$: +4.3 (*c*1.04, DMSO). **^{1}H-NMR** (CD_3OD, 400 MHz) δ ppm (rotamers present 60:40): 7.75 – 7.58 (m, 1H, H-5), 7.25 – 7.07 (m, 1H, H-3), 6.66 – 6.50 [6.59 (dd, *J* = 3.5, 1.8 Hz), 6.58 (dd, *J* = 3.5, 1.8 Hz), 1H, H-4], 4.46 – 4.16 (m, 2H, H_{Val}-2 + H_{Ala}-2), 2.25 – 2.10 (m, 1H, H_{Val}-3), 1.42 – 1.33 [m, 3H, H_{Ala}-3], 1.08 – 0.93 (m, 6H, H_{Val}-4). **^{13}C-NMR and DEPT,** (CD_3OD, 101 MHz) δ ppm (rotamers present): [177.6 (C), 177.3 (C), 173.6 (C), 173.1 (C), 160.9 (C), 160.6 (C), 2 x CONH + $CONH_2$], [148.5 (C), 148.5 (C), C-2], 146.7 (CH, C-5), [115.9 (CH), 115.9 (CH), C-3], [113.1 CH), 113.0 (CH), C-4], [60.8 (CH), 59.9 (CH), C_{Val}-2], [50.1 (CH), 50.0 (CH), C_{Ala}-2], [32.2 (CH), 31.6 (CH), C_{Val}-3], [19.8 (CH_3), 19.6 (CH_3), 19.3 (CH_3), 18.9 (CH_3), 18.3 (CH_3), 18.0 (CH_3), C_{Val}-4 + C_{Ala}-3]. **HRMS** (ESI-TOF) *m*/*z*: $[M + H]^+$ Calcd for $C_{13}H_{20}N_3O_4^+$: 282.14483; Found: 282.14437.

Pharmacology

Dopamine D_2 receptor binding assay

Chinese hamster ovary (CHO) cells expressing short isoform of human D_{2S} receptors were grown in 150 mm petri dishes in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and 2 mM L-Glutamine. When cells were confluent, medium was removed, and cells were washed twice with buffer A (5 mM Tris-HCl pH = 7.4, 2 mM EDTA). Cells were scrapped and homogenized twice in a Polytron. Cell suspension was centrifuged (300 *g*, 10 min, 4 °C). Pellet was discarded, and supernatant was centrifuged (48,400 *g*; 4 °C; 60 min). Pellet was resuspended in buffer B (50 mM Tris-HCl; pH = 7.4) and protein quantity was measured by using Bradford method. The binding of [^{3}H]-NPA to the membrane preparation was assayed in duplicate in 96-well plates. Membranes (30 μg/well) expressing human D_2R were incubated with 0.25 nM

[^{3}H]-NPA and test compounds for 60 min at 25 °C in a 96-well polypropylene microplate with incubation buffer (50 mM Tris-HCl, pH = 7.4; 120 mM NaCl, 5 mM KCl, 4 mM $MgCl_2$, 1 mM EDTA) up to a total volume of 250 μL. Non-specific binding was defined in the presence of 1 μM (+)-butaclamol. After incubation time 200 μL were transferred to a multiscreen FC microplate (Millipore) pre-treated with 0.5% polyethilenimine and samples were filtered and washed 4 times with 250 μL of wash buffer (50 mM Tris-HCl, pH = 7.4; 0.9% NaCl). Filters were dried and 35 μL of scintillation cocktail (Universol) were added to each well and radioactivity was detected in a microplate beta scintillation counter (Microbeta Trilux). Data are expressed as the increase of specific binding following the formula:

$$\%\ \text{Increase} = \frac{(\text{X} - \text{NSB}) \times 100}{\text{BT} - \text{NSB}} - 100$$

where X is the radioactivity detected in the test well; BT is the radioactivity detected when [^{3}H]-NPA was incubated in the absence of any compound and NSB is the radioactivity detected when [^{3}H]-NPA was co-incubated with 10 μM (+)-butaclamol. ANOVA analysis was carried out to evaluate significant differences using SPSS software (V15.0). Statistical significance was set at $P < 0.05$.

Biological Assays.

Animals. Female Wistar-Kyoto (WKY) rats (Iffa-Credo, L'Arbresele, Lyon, France), purchased from Criffa (Barcelona, Spain), were used throughout this study. They were housed (groups of five) in Makrolon cages (Panlab, Barcelona, Spain) on poplar shaving bedding (B&K Universal; G. Jordi, Barcelona, Spain) in a standard experimental animal Bioclean room, illuminated from 8:00 AM to 8:00 PM (12 h/12 h/light/dark cycle) and maintained at a temperature of 22 to 24 °C. Animals had free access to food pellets (B&K Universal), drinking fluid (tap water). All animal experiments performed were conducted in compliance with institutional guidelines and European regulations on the protection of animals (Directive 2010/63/UE), the Spanish Real Decreto 53/2013 (1 February) and the Guide for the Care and Use of Laboratory Animals as adopted and promulgated by the US.

Culture of Rat Cortical Neurons and Glial Cells. Pregnant rats (19−20 days) were killed by CO_2 inhalation. Embryos were immediately extracted from the womb by caesarean section, and their brains were carefully dissected out. Meninges were removed, and a portion of motor cortex was isolated after the dissection of the brain. Fragments obtained from several embryos were subjected to mechanic disintegration. Neurobasal medium supplemented with 2% B-27 was used to seed the cells in 48-well plates at a density of 100 000 cells/mL. Neuronal cultures were allowed to grow for 8−10 days keeping in an incubator (Forma Direct Heat CO_2, Thermo Electron Corporation, Madrid, Spain) under saturated humidity at a partial pressure of 5% CO_2 in air at 37 °C.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:

The full lists of the values of MA and MMA appear in the Supplementary Information file SI00, along with the dataset used and the results of the MMA model for each case, including compound code, molecular descriptors, and the assay conditions. Copies of the 1D (^{1}H, ^{13}C, DEPT-135) and 2D (COSY, HSQC) NMR spectra for all the compounds reported are provided in the file SI01.

Corresponding Authors:

*E-mails: idias@fc.up.pt (I.E.S.-D.); xerardo.garcia@usc.es (X.G.-M.); humberto.gonzalezdiaz@ehu.es (H.G.-D.).

Notes

The authors declare no competing financial interest.

Funding

This research was funded by Fundação para a Ciência e Tecnologia (FCT, Portugal), through grants UIDB/50006/2020 (to LAQV-REQUIMTE Research Unit) and for project grants PTDC/BIA-MIB/29059/2017 and PEst-OE/QUI/UI0674/2013. This work was also supported by the Collaborative Project of Genomic Data Integration (CICLOGEN). The authors thank the USEF Drug Screening Platform at University of Santiago de Compostela the support to carry out this work and also acknowledge the support of Ikerbasque, Basque Foundation for Science and the research grants from Ministry of Economy and Competitiveness, MINECO, Spain (FEDER CTQ2016-74881-P) and Basque government (IT1045-16). IESD also thanks FCT for the doctoral grant SFRH/BD/93632/2013 and XGM thanks Xunta de Galicia for financial funding with reference GPC2014/003.

Acknowledgments

Thanks are due to the Portuguese NMR network (RNRMN) for supporting the Laboratory for Structural Elucidation (LAE) of the Materials Centre of the University of Porto (CEMUP). IESD would like to thank Tiago Carrola for the valuable discussions about this work.

Authors Contributions

Conceived and designed the experiments: IESD, HGD. Experimental work: IESD. Development of the model: HGD, SA, JL, VYP, HB. Analysed the data: IESD, XGM, JERB, HGD, OC. Pharmacological Assays: JMB, MIL; Biological assays: DV. Wrote the paper: IESD, XGM, HGD. All authors have given approval to the final version of the manuscript. The authors declare no competing financial interest.

■ Abbreviations Used

^{13}C-NMR, Carbon-13 nuclear magnetic resonance; ^{1}H-NMR, Proton magnetic nuclear resonance; 2-Fu, 2-Furoic acid; A_1R, Adenosine A_1 receptor; Ac, Accuracy; BD, Big data; C1, Caspase-1; C7, Caspase-7; CB_1, Cannabinoid CB_1 receptor; CB_2, Cannabinoid CB_2 receptor; CEMUP, Centro de Materiais da Universidade do Porto; CHO, Chinese hamster ovary cells; CHO-K1, A subclone of the parental CHO cell line; D_1R, Dopamine D_1 receptor; DIEA, *N*-Ethyl-*N*,*N*-diisopropylamine; DMSO-d_6, Deterated dimethylsulfoxide; DOS, Diversity-oriented synthesis; ESI, Electrospray ionization; $GABA_A$, γ-Aminobutyric acid type A; GPCRs, G-protein coupled receptors; HEK293, Human Embryonic Kidney 293 cells; HHV-5, Human cytomegalovirus; HHV-5 chemokine, HHV-5 chemokine receptor; HRMS, High-resolution mass spectrometry; LDA, Linear Discriminant Analysis; M_1R, Muscarinic acetylcholine receptor 1; M_2R, Muscarinic acetylcholine receptor M_2; M_4R, Muscarinic acetylcholine receptor M_4; MA, Moving Average operators; mGluR, Metabotropic glutamate receptor; MIF-1, Melanostatin; ML, Machine Learning; MMA, Multiple Moving Average operators; NAM, Negative allosteric modulator; NK_2, Tachykinin receptor 2; NPA, *N*-propylapomorphine; PAM, Positive allosteric modulator; PSA, Polar Surface Area; PT, Perturbation Theory; PTML, Perturbation Theory and Machine Learning Model; Sn, Sensitivity; Sp, Specificity; TBTU, *O*-(Benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium tetrafluoroborate; TLC, Thin-layer chromatography; TRHR, Thyrotropin-releasing hormone receptor; VCCLAB, Virtual Computational Chemistry Laboratory.

■ References

1. Mishra, A.; Singh, S.; Shukla, S., Physiological and Functional Basis of Dopamine Receptors and Their Role in Neurogenesis: Possible Implication for Parkinson's disease. *J. Exp. Neurosci.* **2018,** *12*, 1-8.

2. Beaulieu, J. M.; Gainetdinov, R. R., The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacol. Rev.* **2011,** *63* (1), 182-217.

3. Platania, C. B. M.; Salomone, S.; Leggio, G. M.; Drago, F.; Bucolo, C., Homology modeling of dopamine D2 and D3 receptors: molecular dynamics refinement and docking evaluation. *PLoS One* **2012,** *7* (9), e44316.

4. Lindgren, N.; Usiello, A.; Goiny, M.; Haycock, J.; Erbs, E.; Greengard, P.; Hokfelt, T.; Borrelli, E.; Fisone, G., Distinct roles of dopamine D2L and D2S receptor isoforms in the regulation of protein phosphorylation at presynaptic and postsynaptic sites. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (7), 4305-4359.

5. Nussinov, R.; Tsai, C. J., Allostery in disease and in drug discovery. *Cell* **2013,** *153* (2), 293-305.

6. Khan, R. S.; Yu, C.; Kastin, A. J.; He, Y.; Ehrensing, R. H.; Hsuchou, H.; Stone, K. P.; Pan, W., Brain Activation by Peptide Pro-Leu-Gly-NH_2 (MIF-1). *Int. J. Pept.* **2010,** *2010*, 537639.

7. Allegretti, M.; Cesta, M. C.; Locati, M., Allosteric Modulation of Chemoattractant Receptors. *Front. Immunol.* **2016,** *7*, 170.

8. 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.; Mishra, R. K., Modulation of agonist binding to human dopamine receptor subtypes by L-prolyl-L-leucyl-glycinamide and a peptidomimetic analog. *J. Pharmacol. Exp. Ther.* **2005,** *315* (3), 1228-1236.

9. Vartak, A. P.; Skoblenick, K.; Thomas, N.; Mishra, R. K.; Johnson, R. L., Allosteric modulation of the dopamine receptor by conformationally constrained type VI beta-turn peptidomimetics of Pro-Leu-Gly-NH2. *J. Med. Chem.* **2007,** *50* (26), 6725-6729.

10. Ferreira da Costa, J.; Caamaño, O.; Fernández, F.; García-Mera, X.; Sampaio-Dias, I. E.; Brea, J. M.; Cadavid, M. I., 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.* **2013,** *69* (Supplement C), 146-158.

11. Sampaio-Dias, I. E.; Silva-Reis, S. C.; García-Mera, X.; Brea, J.; Loza, M. I.; Alves, C. S.; Algarra, M.; Rodríguez-Borges, J. E., Synthesis, Pharmacological, and Biological Evaluation of MIF-1 Picolinoyl Peptidomimetics as Positive Allosteric Modulators of D_2R. *ACS Chem. Neurosci.* **2019,** *10* (8), 3690-3702.

12. Sampaio-Dias, I. E.; Sousa, C. A. D.; García-Mera, X.; da Costa, J. F.; Caamaño, O.; Rodríguez-Borges, J. E., *Org. Biomol. Chem.* **2016,** *14*, 11065-11069.

13. Celis, M. E.; Taleisnik, S.; Walter, R., Regulation of formation and proposed structure of the factor inhibiting the release of melanocyte-stimulating hormone. *Proc. Natl. Acad. Sci. U. S. A.* **1971,** *68* (7), 1428-1433.

14. Srivastava, L. K.; Bajwa, S. B.; Johnson, R. L.; Mishra, R. K., Interaction of l-Prolyl-l-Leucyl Glycinamide with Dopamine D2 Receptor: Evidence for Modulation of Agonist Affinity States in Bovine Striatal Membranes. *J. Neurochem.* **1988,** *50* (3), 960-968.

15. Chiu, S.; Paulose, C. S.; Mishra, R. K., Effect of L-prolyl-L-leucyl-glycinamide (PLG) on neuroleptic-induced catalepsy and dopamine/neuroleptic receptor bindings. *Peptides* **1981,** *2* (1), 105-111.

16. Saitton, S.; Del Tredici, A. L.; Mohell, N.; Vollinga, R. C.; Boström, D.; Kihlberg, J.; Luthman, K., Design, synthesis and evaluation of a PLG tripeptidomimetic based on a pyridine scaffold. *J. Med. Chem.* **2004,** *47* (26), 6595-6602.

17. Reed, G. W.; Olson, G. A.; Olson, R. D., The Tyr-MIF-1 family of peptides. *Neurosci. Biobehav. Rev.* **1994,** *18* (4), 519-525.

18. Mishra, R. K.; Makman, M. H.; Costain, W. J.; Nair, V. D.; Johnson, R. L., Modulation of agonist stimulated adenylyl cyclase and GTPase activity by L-pro-L-leu-glycinamide and its peptidomimetic analogue in rat striatal membranes. *Neurosci. Lett.* **1999,** *269* (1), 21-24.

19. Birtwistle, J.; Baldwin, D., Role of dopamine in schizophrenia and Parkinson's disease. *Br. J. Nurs.* **1998,** *7* (14), 832-834, 836, 838-841.

20. Smith, J. R.; Morgan, M., The effects of prolyl-leucyl-glycine amide on drug-induced rotation in lesioned rats. *Gen. Pharmacol.* **1982,** *13* (3), 203-207.

21. Marcotte, E. R.; Chugh, A.; Mishra, R. K.; Johnson, R. L., Protection Against MPTP Treatment by an Analog of Pro-Leu-Gly-NH2 (PLG, MIF-1). *Peptides* **1998,** *19* (2), 403-406.

22. Barbeau, A.; Roy, M.; Kastin, A. J., Double-blind evaluation of oral L-prolyl-Lleucyl-glycine amide in Parkinson's disease. *Can. Med. Assoc. J.* **1976,** *114* (2), 120-122.

23. Sharma, S.; Paladino, P.; Gabriele, J.; Saeedi, H.; Henry, P.; Chang, M.; Mishra, R. K.; Johnson, R. L., Pro-Leu-glycinamide and its peptidomimetic, PAOPA, attenuate haloperidol induced vacuous chewing movements in rat: A model of human tardive dyskinesia. *Peptides* **2003,** *24* (2), 313-319.

24. Mycroft, F. J.; Bhargava, H. N.; Wei, E. T., Pharmacological activities of the MIF-1 analogues Pro-Leu-Gly, Tyr-Pro-Leu-Gly and pareptide. *Peptides* **1987,** *8* (6), 1051-1055.

25. Chiu, P.; Rajakumar, G.; Chiu, S.; Johnson, R. L.; Mishra, R. K., Mesolimbic and striatal dopamine receptor supersensitivity: Prophylactic and reversal effects of L-prolyl-L-leucyl-glycinamide (PLG). *Peptides* **1985,** *6* (2), 179-183.

26. Reed, L. L.; Johnson, P. L., Solid state conformation of the C-terminal tripeptide of oxytocin, L-Pro-L-Leu-Gly-NH2.0.5H2O. *J. Am. Chem. Soc.* **1973,** *95* (22), 7523-7524.

27. Aizpurua, J. M.; Palomo, C.; Balentová, E.; Jimenez, A.; Andreieff, E.; Sagartzazu-Aizpurua, M.; Miranda, J. I.; Linden, A., Chirality-Driven Folding of Short β-Lactam Pseudopeptides. *J. Org. Chem.* **2013,** *78* (2), 224-237.

28. Johnson, R. L.; Rajakumar, G.; Yu, K. L.; Mishra, R. K., Synthesis of Pro-Leu-Gly-NH2 analogues modified at the prolyl residue and evaluation of their effects on the receptor binding activity of the central dopamine receptor agonist, ADTN. *J. Med. Chem.* **1986,** *29* (10), 2104-2107.

29. Johnson, R. L.; Rajakumar, G.; Yu, K. L.; Mishra, R. K., Synthesis of Pro-Leu-Gly-NH2 analogs modified at the prolyl residue and evaluation of their effects on the receptor binding activity of the central dopamine receptor agonist, ADTN. *J. Med. Chem.* **1986,** *29* (10), 2104-2107.

30. Sampaio-Dias, I. E.; Sousa, C. A. D.; Silva-Reis, S. C.; Ribeiro, S.; García-Mera, X.; Rodríguez-Borges, J. E., Highly efficient one-pot assembly of peptides by double chemoselective coupling. *Org. Biomol. Chem.* **2017,** *15* (36), 7533-7542.

31. Liang, X.; Haynes, B. S.; Montoya, A., Acid-Catalyzed Ring Opening of Furan in Aqueous Solution. *Energy Fuels* **2018,** *32* (4), 4139-4148.

32. Open-source cheminformatics; http://zinc15.docking.org/patterns/home (accessed July 10, 2020).

33. Christopoulos, A.; Kenakin, T., G protein-coupled receptor allosterism and complexing. *Pharmacol. Rev.* **2002,** *54* (2), 323-374.

34. Bhagwanth, S.; Mishra, R. K.; Johnson, R. L., Development of peptidomimetic ligands of Pro-Leu-Gly-NH(2) as allosteric modulators of the dopamine D(2) receptor. *Beilstein J. Org. Chem.* **2013,** *9*, 204-214.

35. Alonso, N.; Caamaño, O.; Romero-Duran, F. J.; Luan, F.; Cordeiro, M. N. D. S.; Yanez, M.; González-Díaz, H.; García-Mera, X., Model for High-Throughput Screening of Multitarget Drugs in Chemical Neurosciences: Synthesis, Assay, and Theoretic Study of Rasagiline Carbamates. *ACS Chem. Neurosci.* **2013,** *4* (10), 1393-1403.

36. Luan, F.; Cordeiro, M. N. D. S.; Alonso, N.; García-Mera, X.; Caamaño, O.; Romero-Duran, F. J.; Yanez, M.; González-Díaz, H., TOPS-MODE model of multiplexing neuroprotective effects of drugs and experimental-theoretic study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases. *Bioorg. Med. Chem.* **2013,** *21* (7), 1870-1879.

37. Ricart, K. C.; Fiszman, M. L., Hydrogen Peroxide-Induced Neurotoxicity in Cultured Cortical Cells Grown in Serum-Free and Serum-Containing Media. *Neurochem. Res.* **2001,** *26* (7), 801-808.

38. Schneider, G., Computational medicinal chemistry. *Future Med. Chem.* **2011,** *3* (4), 393-394.

39. Csermely, P.; Nussinov, R.; Szilagyi, A., From allosteric drugs to allo-network drugs: state of the art and trends of design, synthesis and computational methods. *Curr. Top. Med. Chem.* **2013,** *13* (1), 2-4.

40. Szilagyi, A.; Nussinov, R.; Csermely, P., Allo-network drugs: extension of the allosteric drug concept to protein- protein interaction and signaling networks. *Curr. Top. Med. Chem.* **2013,** *13* (1), 64-77.

41. Csermely, P.; Korcsmaros, T.; Kiss, H. J.; London, G.; Nussinov, R., Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **2013,** *138* (3), 333-408.

42. Taylor, P. J., *Comprehensive Medicinal Chemistry*. 1990 ed.; Pergamon Press: Oxford, 1990; Vol. 4, p 241-294.

43. Wang, J.; Yun, D.; Yao, J.; Fu, W.; Huang, F.; Chen, L.; Wei, T.; Yu, C.; Xu, H.; Zhou, X.; Huang, Y.; Wu, J.; Qiu, P.; Li, W., Design, synthesis and QSAR study of novel isatin analogues inspired Michael acceptor as potential anticancer compounds. *Eur. J. Med. Chem.* **2018,** *144*, 493-503.

44. Pogorzelska, A.; Slawinski, J.; Zolnowska, B.; Szafranski, K.; Kawiak, A.; Chojnacki, J.; Ulenberg, S.; Zielinska, J.; Baczek, T., Novel 2-(2-alkylthiobenzenesulfonyl)-3-(phenylprop-2-ynylideneamino)guanidine derivatives as potent anticancer agents - Synthesis, molecular structure, QSAR studies and metabolic stability. *Eur. J. Med. Chem.* **2017,** *138*, 357-370.

45. Slawinski, J.; Szafranski, K.; Pogorzelska, A.; Zolnowska, B.; Kawiak, A.; Macur, K.; Belka, M.; Baczek, T., Novel 2-benzylthio-5-(1,3,4-oxadiazol-2-yl)benzenesulfonamides with anticancer activity: Synthesis, QSAR study, and metabolic stability. *Eur. J. Med. Chem.* **2017,** *132*, 236-248.

46. Murahari, M.; Kharkar, P. S.; Lonikar, N.; Mayur, Y. C., Design, synthesis, biological evaluation, molecular docking and QSAR studies of 2,4-dimethylacridones as anticancer agents. *Eur. J. Med. Chem.* **2017,** *130*, 154-170.

47. Ruddarraju, R. R.; Murugulla, A. C.; Kotla, R.; Chandra Babu Tirumalasetty, M.; Wudayagiri, R.; Donthabakthuni, S.; Maroju, R.; Baburao, K.; Parasa, L. S., Design, synthesis, anticancer, antimicrobial activities and molecular docking studies of theophylline containing acetylenes and theophylline containing 1,2,3-triazoles with variant nucleoside derivatives. *Eur. J. Med. Chem.* **2016,** *123*, 379-396.

48. Singh, H.; Kumar, R.; Singh, S.; Chaudhary, K.; Gautam, A.; Raghava, G. P., Prediction of anticancer molecules using hybrid model developed on molecules screened against NCI-60 cancer cell lines. *BMC Cancer* **2016,** *16*, 77.

49. Pingaew, R.; Prachayasittikul, V.; Worachartcheewan, A.; Nantasenamat, C.; Prachayasittikul, S.; Ruchirawat, S.; Prachayasittikul, V., Novel 1,4-naphthoquinone-based sulfonamides: Synthesis, QSAR, anticancer and antimalarial studies. *Eur. J. Med. Chem.* **2015,** *103*, 446-459.

50. Anwer, Z.; Gupta, S. P., A QSAR study on some series of anticancer tyrosine kinase inhibitors. *Med. Chem.* **2013,** *9* (2), 203-212.

51. Toropov, A. A.; Toropova, A. P.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J., SMILES-based QSAR approaches for carcinogenicity and anticancer activity: comparison of correlation weights for identical SMILES attributes. *Anticancer Agents Med. Chem.* **2011,** *11* (10), 974-982.

52. Huang, X. Y.; Shan, Z. J.; Zhai, H. L.; Li, L. N.; Zhang, X. Y., Molecular design of anticancer drug leads based on three-dimensional quantitative structure-activity relationship. *J. Chem. Inf. Model.* **2011,** *51* (8), 1999-2006.

53. Benfenati, E.; Toropov, A. A.; Toropova, A. P.; Manganaro, A.; Gonella Diaza, R., coral Software: QSAR for Anticancer Agents. *Chem. Biol. Drug Des.* **2011,** *77* (6), 471-476.

54. González-Díaz, H.; Bonet, I.; Teran, C.; De Clercq, E.; Bello, R.; Garcia, M. M.; Santana, L.; Uriarte, E., ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds. *Eur. J. Med. Chem.* **2007,** *42* (5), 580-585.

55. González-Díaz, H.; Vina, D.; Santana, L.; de Clercq, E.; Uriarte, E., Stochastic entropy QSAR for the in silico discovery of anticancer compounds: prediction, synthesis, and in vitro assay of new purine carbanucleosides. *Bioorg. Med. Chem.* **2006,** *14* (4), 1095-1107.

56. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernadez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A., Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design I: discovery of anticancer compounds. *J. Mol. Model.* **2003,** *9* (6), 395-407.

57. Jung, M.; Kim, H.; Kim, M., Chemical genomics strategy for the discovery of new anticancer agents. *Curr. Med. Chem.* **2003,** *10* (9), 757-762.

58. Shi, L. M.; Fan, Y.; Myers, T. G.; O'Connor, P. M.; Paull, K. D.; Friend, S. H.; Weinstein, J. N., Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR study of anticancer ellipticine analogues. *J. Chem. Inf. Comput. Sci.* **1998,** *38* (2), 189-199.

59. Han, L.; Cui, J.; Lin, H.; Ji, Z.; Cao, Z.; Li, Y.; Chen, Y., Recent progresses in the application of machine learning approach for predicting protein functional class independent of sequence similarity. *Proteomics* **2006,** *6* (14), 4023-4037.

60. Chou, K. C.; Shen, H. B., Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS ONE* **2010,** *5* (6), e11335.

61. Chou, K. C.; Shen, H. B., Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat. Protoc.* **2008,** *3* (2), 153-162.

62. Cai, Y. D.; Chou, K. C., Predicting enzyme subclass by functional domain composition and pseudo amino acid composition. *J. Proteome Res.* **2005,** *4* (3), 967-971.

63. Shen, H. B.; Chou, K. C., QuatIdent: a web server for identifying protein quaternary structural attribute by fusing functional domain and sequential evolution information. *J. Proteome Res.* **2009,** *8* (3), 1577-1584.

64. Chou, K. C.; Shen, H. B., Large-scale predictions of gram-negative bacterial protein subcellular locations. *J. Proteome Res.* **2006,** *5* (12), 3420-3428.

65. Rodriguez-Soca, Y.; Munteanu, C. R.; Dorado, J.; Pazos, A.; Prado-Prado, F. J.; González-Díaz, H., Trypano-PPI: a web server for prediction of unique targets in trypanosome proteome by using electrostatic parameters of protein-protein interactions. *J. Proteome Res.* **2010,** *9* (2), 1182-1190.

66. Munteanu, C. R.; Vazquez, J. M.; Dorado, J.; Sierra, A. P.; Sanchez-Gonzalez, A.; Prado-Prado, F. J.; González-Díaz, H., Complex network spectral moments for ATCUN motif DNA cleavage: first predictive study on proteins of human pathogen parasites. *J. Proteome Res.* **2009,** *8* (11), 5219-5228.

67. González-Díaz, H.; Saiz-Urra, L.; Molina, R.; Santana, L.; Uriarte, E., A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. *J. Proteome Res.* **2007,** *6* (2), 904-908.

68. González-Díaz, H.; Prado-Prado, F.; García-Mera, X.; Alonso, N.; Abeijon, P.; Caamaño, O.; Yanez, M.; Munteanu, C. R.; Pazos, A.; Dea-Ayuela, M. A.; Gomez-Munoz, M. T.; Garijo, M. M.; Sansano, J.; Ubeira, F. M., MIND-BEST: Web server for drugs and target discovery; design, synthesis, and assay of MAO-B inhibitors and theoretical-experimental study of G3PDH protein from Trichomonas gallinae. *J. Proteome Res.* **2011,** *10* (4), 1698-1718.

69. Concu, R.; Dea-Ayuela, M. A.; Perez-Montoto, L. G.; Bolas-Fernandez, F.; Prado-Prado, F. J.; Podda, G.; Uriarte, E.; Ubeira, F. M.; González-Díaz, H., 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.* **2009,** *8* (9), 4372-4382.

70. Aguero-Chapin, G.; Varona-Santos, J.; de la Riva, G. A.; Antunes, A.; Gonzalez-Vlla, T.; Uriarte, E.; González-Díaz, H., Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from Coffea arabica and prediction of a new sequence. *J. Proteome Res.* **2009,** *8* (4), 2122-2128.

71. Martinez, S. G.; Tenorio-Borroto, E.; Barbabosa Pliego, A.; Diaz-Albiter, H.; Vazquez-Chagoyan, J. C.; González-Díaz, H., PTML Model for Proteome Mining of B-cell Epitopes and Theoretic-Experimental Study of Bm86 Protein Sequences from Colima Mexico. *J. Proteome Res.* **2017**, 16 (11), 4093-4103.

72. Fernández, M.; Caballero, F.; Fernández, L.; Abreu, J. I.; Acosta, G., Classification of conformational stability of protein mutants from 3D pseudo-folding graph representation of protein sequences using support vector machines. *Proteins* **2008,** *70* (1), 167-175.

73. Fernández, L.; Caballero, J.; Abreu, J. I.; Fernández, M., Amino Acid Sequence Autocorrelation Vectors and Bayesian-Regularized Genetic Neural Networks for Modeling Protein Conformational Stability: Gene V Protein Mutants. *Proteins* **2007,** *67*, 834-852.

74. Caballero, J.; Fernandez, L.; Abreu, J. I.; Fernandez, M., 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.* **2006,** *46* (3), 1255-1268.

75. 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.; Besada, V., Isoelectric point optimization using peptide descriptors and support vector machines. *J. Proteomics* **2012,** *75* (7), 2269-2274.

76. Greener, J. G.; Sternberg, M. J., AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformatics* **2015,** *16*, 335.

77. 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.; Leach, A. R., The ChEMBL database in 2017. *Nucleic Acids Res.* **2017,** *45* (D1), D945.

78. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P., ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012,** *40* (Database issue), D1100-D1107.

79. González-Díaz, H.; Arrasate, S.; Gomez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.; Ruso, J. M., 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.* **2013,** *13* (14), 1713-1741.

80. Blazquez-Barbadillo, C.; Aranzamendi, E.; Coya, E.; Lete, E.; Sotomayor, N.; González-Díaz, H., Perturbation theory model of reactivity and enantioselectivity of palladium-catalyzed Heck-Heck cascade reactions. *RSC Adv.* **2016,** *6* (45), 38602-38610.

81. Casanola-Martin, G. M.; Le-Thi-Thu, H.; Perez-Gimenez, F.; Marrero-Ponce, Y.; Merino-Sanjuan, M.; Abad, C.; González-Díaz, H., 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.* **2016,** *17* (3), 220-227.

82. Romero-Duran, F. J.; Alonso, N.; Yanez, M.; Caamaño, O.; García-Mera, X.; González-Díaz, H., Brain-inspired cheminformatics of drug-target brain interactome, synthesis, and assay of TVP1022 derivatives. *Neuropharmacol.* **2016,** *103*, 270-278.

83. Kleandrova, V. V.; Luan, F.; González-Díaz, H.; Ruso, J. M.; Speck-Planche, A.; Cordeiro, M. N. D. S., 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.* **2014,** *48* (24), 14686-14694.

84. Luan, F.; Kleandrova, V. V.; González-Díaz, H.; Ruso, J. M.; Melo, A.; Speck-Planche, A.; Cordeiro, M. N. D. S., Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale* **2014,** *6* (18), 10623-10630.

85. Speck-Planche, A.; Cordeiro, M., Fragment-based in silico modeling of multi-target inhibitors against breast cancer-related proteins. *Mol. Divers.* **2017,** *21* (3), 511-523.

86. Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S., Unified multi-target approach for the rational in silico design of anti-bladder cancer agents. *Anticancer Agents Med. Chem.* **2013,** *13* (5), 791-800.

87. Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S., Chemoinformatics in multi-target drug discovery for anti-cancer therapy: in silico design of potent and versatile anti-brain tumor agents. *Anticancer Agents Med. Chem.* **2012,** *12* (6), 678-685.

88. Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S., Rational drug design for anti-cancer chemotherapy: multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. *Bioorg. Med. Chem.* **2012,** *20* (15), 4848-4855.

89. Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S., Chemoinformatics in anti-cancer chemotherapy: multi-target QSAR model for the in silico discovery of anti-breast cancer agents. *Eur. J. Pharm. Sci.* **2012,** *47* (1), 273-279.

90. Cordeiro, M. N. D. S.; Speck-Planche, A., Computer-aided drug design, synthesis and evaluation of new anti-cancer drugs. *Curr. Top. Med. Chem.* **2012,** *12* (24), 2703-2704.

91. Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S., Multi-target drug discovery in anti-cancer therapy: fragment-based approach toward the design of potent and versatile anti-prostate cancer agents. *Bioorg. Med. Chem.* **2011,** *19* (21), 6239-6244.

92. Speck-Planche, A.; Cordeiro, M. N. D. S. Advanced In Silico Approaches for Drug Discovery: Mining Information from Multiple Biological and Chemical Data Through mtk- QSBER and pt-QSPR Strategies. *Curr. Med. Chem.* **2017,** *24*, 1687-1704.

93. Collier, G.; Ortiz, V., Emerging computational approaches for the study of protein allostery. *Arch. Biochem. Biophys.* **2013,** *538* (1), 6-15.

94. Liu, Y.; Tang, S.; Fernandez-Lozano, C.; Munteanu, C. R.; Pazos, A.; Yu. Y.Z.; Tan, Z.; González-Díaz, H., Experimental study and Random Forest prediction model of microbiome cell surface hydrophobicity. *Expert Syst. Appl.* **2017,** *72*, 306-316.

95. González-Díaz, H.; Herrera-Ibata, D. M.; Duardo-Sanchez, A.; Munteanu, C. R.; Orbegozo-Medina, R. A.; Pazos, A., ANN multiscale model of anti-HIV drugs activity vs AIDS prevalence in the US at county level based on information indices of molecular graphs and social networks. *J. Chem. Inf. Model.* **2014,** *54* (3), 744-755.

96. Casanola-Martin, G. M.; Le-Thi-Thu, H.; Perez-Gimenez, F.; Marrero-Ponce, Y.; Merino-Sanjuan, M.; Abad, C.; González-Díaz, H., Multi-output model with Box-Jenkins operators of linear indices to predict multi-target inhibitors of ubiquitin-proteasome pathway. *Mol. Divers.* **2015,** *19* (2), 347-356.

97. Hill, T.; Lewicki, P., *STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*. StatSoft: Tulsa, 2006 Vol. 1, p 813.

98. Young, R. C.; Mitchell, R. C.; Brown, T. H.; Ganellin, C. R.; Griffiths, R.; Jones, M.; Rana, K. K.; Saunders, D.; Smith, I. R.; Sore, N. E.; et al., Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists. *J. Med. Chem.* **1988,** *31* (3), 656-671.

99. Friden, M.; Winiwarter, S.; Jerndal, G.; Bengtsson, O.; Wan, H.; Bredberg, U.; Hammarlund-Udenaes, M.; Antonsson, M., Structure-brain exposure relationships in rat and human using a novel data set of unbound drug concentrations in brain interstitial and cerebrospinal fluids. *J. Med. Chem.* **2009,** *52* (20), 6233-6243.

100. Hitchcock, S. A.; Pennington, L. D., Structure-brain exposure relationships. *J. Med. Chem.* **2006,** *49* (26), 7559-7583.

101.Garcia, I.; Fall, Y.; García-Mera, X.; Prado-Prado, F., Theoretical study of GSK-3alpha: neural networks QSAR studies for the design of new inhibitors using 2D descriptors. *Mol. Divers.* **2011,** *15* (4), 947-955.

102.Marrero-Ponce, Y.; Siverio-Mota, D.; Galvez-Llompart, M.; Recio, M. C.; Giner, R. M.; Garcia-Domenech, R.; Torrens, F.; Aran, V. J.; Cordero-Maldonado, M. L.; Esguera, C. V.; de Witte, P. A.; Crawford, A. D., Discovery of novel anti-inflammatory drug-like compounds by aligning in silico and in vivo screening: the nitroindazolinone chemotype. *Eur. J. Med. Chem.* **2011,** *46* (12), 5736-5753.

103.Kuligowski, J.; Perez-Guaita, D.; Escobar, J.; de la Guardia, M.; Vento, M.; Ferrer, A.; Quintas, G., Evaluation of the effect of chance correlations on variable selection using Partial Least Squares-Discriminant Analysis. *Talanta* **2013,** *116*, 835-840.

104.Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V., Virtual computational chemistry laboratory - design and description. *J. Comput. Aided Mol. Des.* **2005,** *19* (6), 453-463.

105.González-Díaz, H.; Perez-Montoto, L. G.; Ubeira, F. M., 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,** *2014*, 768515.

106.González-Díaz, H.; Arrasate, S.; Gomez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.; Ruso, J. M., 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.* **2013,** *13* (14), 1713-1741.

107.Gottlieb, H. E.; Kotlyar, V.; Nudelman, A., NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* **1997,** *62* (21), 7512-7515.

For Table of Contents Use Only

Synthesis, Pharmacological and Biological Evaluation of 2-Furoyl-based MIF-1 Peptidomimetics, and the Development of a General-Purpose Model for Allosteric Modulators (ALLOPTML)

Ivo E. Sampaio-Dias[1,*], José E. Rodríguez-Borges[1], Víctor Yáñez-Pérez[2], Sonia Arrasate[3], Javier Llorente[3,4], José M. Brea[5], Harbil Bediaga[2,6], Dolores Viñas[7], María Isabel Loza[5], Olga Caamaño[8], Xerardo García-Mera[8,*], and Humberto González-Díaz[2,9,10,*]

[1] *LAQV/REQUIMTE, Dept. of Chemistry and Biochemistry, University of Porto, 4169-007 Porto, Portugal.*

[2] *Dept. of Organic Chemistry II, University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[3] *Dept. of Pharmacology, University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[4] *Dept. of Pharmacology, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[5] *Innopharma Screening Platform. Biofarma Research group. Centre of Research in Molecular Medicine and Chronic Diseases CIMUS, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[6] *Dept. of Physical Chemistry, University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[7] *Dept. of Pharmacology, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[8] *Dept. of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Santiago de Compostela, Spain.*

[9] *Basque Center for Biophysics (CSIC UPV/EHU), University of Basque Country (UPV-EHU), 48940, Leioa, Spain.*

[10] *IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain.*

Corresponding Authors:

*humberto.gonzalezdiaz@ehu.es (H.G.-D.); xerardo.garcia@usc.es (X.G.-M.); idias@fc.up.pt (I.E.S.-D.).

Table of Contents Graphic