Identification of Riluzole derivatives as novel calmodulin inhibitors with neuroprotective activity by a joint synthesis, biosensor, and computational guided strategy

Maider Baltasar-Marchueta [a], Leire Llona [a], Sara M-Alicante [b], Iratxe Barbolla [a], Markel Garcia Ibarluzea [c,d], Rafael Ramis [c,d], Ane Miren Salomon [a], Brenda Fundora [a], Ariane Araujo [b], Arantza Muguruza-Montero [b], Eider Nuñez [b], Scarlett Pérez-Olea [a], Christian Villanueva [a], Aritz Leonardo [c,d], Sonia Arrasate [a], Nuria Sotomayor [a], Alvaro Villarroel [b,*], Aitor Bergara [c,d,**], Esther Lete [a,*], Humberto González-Díaz [a,b,e,***]

[a] Department of Organic and Inorganic Chemistry, University of the Basque Country UPV/EHU, Leioa 48940, Spain
[b] Biofisika Institute, CSIC-UPV/EHU, Leioa 48940, Spain
[c] Donostia International Physics Center, Donostia, Spain
[d] Departament of Physics, University of the Basque Country, UPV/EHU, Leioa, Spain
[e] IKERBASQUE, Basque Foundation for Science, Bilbao 48011, Spain

ARTICLE INFO

ABSTRACT

The development of new molecules for the treatment of calmodulin related cardiovascular or neurodegenerative diseases is an interesting goal. In this work, we introduce a novel strategy with four main steps: (1) chemical synthesis of target molecules, (2) Förster Resonance Energy Transfer (FRET) biosensor development and *in vitro* biological assay of new derivatives, (3) Cheminformatics models development and *in vivo* activity prediction, and (4) Docking studies. This strategy is illustrated with a case study. Firstly, a series of 4-substituted Riluzole derivatives **1–3** were synthetized through a strategy that involves the construction of the 4-bromoriluzole framework and its further functionalization via palladium catalysis or organolithium chemistry. Next, a FRET biosensor for monitoring Ca^{2+}-dependent CaM-ligands interactions has been developed and used for the *in vitro* assay of Riluzole derivatives. In particular, the best inhibition (80%) was observed for 4-methoxyphenylriluzole **2b**. Besides, we trained and validated a new Networks Invariant, Information Fusion, Perturbation Theory, and Machine Learning (NIFPTML) model for predicting probability profiles of *in vivo* biological activity parameters in different regions of the brain. Next, we used this model to predict the *in vivo* activity of the compounds experimentally studied *in vitro*. Last, docking study conducted on Riluzole and its derivatives has provided valuable insights into their binding conformations with the target protein, involving calmodulin and the SK4 channel. This new combined strategy may be useful to reduce assay costs (animals, materials, time, and human resources) in the drug discovery process of calmodulin inhibitors.

1. Introduction

Calmodulin (CaM) is an acidic and thermostable low-molecular-weight protein, present in all eukaryotic cell types. As a prototype calcium sensor, CaM is involved in various cellular signaling processes [1, 2] as a mediator, such as the plasticity and synaptic transmission, regulation of enzymatic activities, modulation of ionic channels activities and regulation of gene expression. Changes in the concentration of calcium ions in the internal environment of the cell regulate CaM in three different ways: at a cellular level, directing its subcellular

---

* Corresponding authors.
** Corresponding author at: Donostia International Physics Center, Donostia, Spain.
*** Corresponding author at: Department of Organic and Inorganic Chemistry, University of the Basque Country UPV/EHU, Leioa 48940, Spain
E-mail addresses: alvaro.villarroel@csic.es (A. Villarroel), a.bergara@ehu.eus (A. Bergara), esther.lete@ehu.eus (E. Lete), humberto.gonzalezdiaz@ehu.eus (H. González-Díaz).

distribution; at a molecular level, promoting different binding modes to target proteins; and, lastly, correcting the conformational states of CaM that achieve specific target activation [3].

Among the diseases associated with CaM, the most notable are cardiopathies, calmodulinopathies and neurodegenerative diseases. Cardiovascular diseases encompass a range of conditions affecting the heart and blood vessels, with disruptions in cellular signaling pathways. Ca^{2+}-CaM dependent protein kinase II (CaMKII) has emerged as a potential indicator of cardiomyopathy, with increased activity linked to arrhythmias and mechanical dysfunction, suggesting the possibility of using myocardial CaMKII inhibition as a strategy to reduce cardiac dysfunction [4,5]. Additionally, calmodulinopathies, caused by mutations in CaM genes, pose life-threatening arrhythmia risks, especially among young individuals [6]. Shifting to neurodegenerative diseases, Alzheimer's disease is characterized by beta-amyloid plaques and neurofibrillary tangles, and the CaM hypothesis has revealed CaM binding domains in key amyloid-related proteins [7,8]. Furthermore, beta-secretase (BACE1) and amyloid precursor protein (AβPP) regulation by CaM have been demonstrated. In Parkinson's disease, heightened αCaMKII autophosphorylation levels and increased interaction with NMDA receptor subunits NR2A-NR2B have been observed [8,9]. Finally, in Amyotrophic Lateral Sclerosis (ALS), calcium dysregulation serves as a convergence point for major dysfunctional pathways and critical proteins, shedding light on environmental factors that influence the aetiology and pathology mechanisms or ALS [10].

SK channels regulate cell volume, modulate the immune response [11,12], and mediate the intrinsic excitability of neurons [13], contributing to the hyperpolarization of the action potential. These channels are activated by Ca^{2+} acting via CaM, with four CaM molecules interacting with each of the four subunits that conform the tetrameric channel. Cryo-EM images of the SK4 channel/CaM complex solved in presence and absence of Ca^{2+} show that under both conditions the CaM C-Lobe remains intimately bound to hA, an intracellular alpha helix that follows the S6 transmembrane domain (the S6 bundle crossing towards the intracellular side of the membrane constitutes a gate that control ion flow). In contrast, the N-lobe exhibits large positional variations. Remarkably, the N-lobe was captured bound to the linker between transmembrane domains 4 and 5 (S4S5) when loaded with Ca^{2+}, suggesting that it was pulling down the domain equivalent to the voltage sensor of other potassium channels, leading to widening of the S6 bundle crossing, allowing K^+ flow through the now opened channel [14]. Examination of this structure reveals the existence of a pocket between the N-lobe and the S4S5 linker, where Riluzole can dock. Remarkably, this putative binding pocket in similar to that found in complexes between an isolated portion of the SK2 C-terminus, CaM and Riluzole [15] (structures that are unlikely to be adopted by the full channel [14]). This similarity suggest that Riluzole may affect SK channel function. Indeed, Riluzole increases Ca^{2+} sensitivity of SK2 channels, reducing the concentration of Ca^{2+} required to cause 50% of the maximal activation from ~0.7 to ~0.4 μM [15].

There are currently various treatments available to treat the diseases caused by the loss of calcium regulation. Riluzole remains to be the only drug currently approved by the U.S. Food and Drug Administration (FDA) for the treatment of ALS, which is also related with CaM. Riluzole is associated with a low rate of serum aminotransferase elevations during therapy and has been linked to rare cases of clinically apparent acute liver injury [16,17]. Additionally, Riluzole shows anti-glutamate agent activity with anticonvulsant and neuroprotective properties in animal models of neurodegenerative disease. However, neuroprotective mechanism of Riluzole is still not well-defined, limiting the ability to design new medicines for ALS. Furthermore, Riluzole is also effective in animal models of Parkinson's disease, Huntington's disease and cerebral and retinal ischemia [18]. It also exhibits strong anticonvulsant, neuroprotective and antidepressant effects, as well as sedative properties [19].

Presently, the prevalence of these diseases commented above [20]

are increasing due, among other factors, to the extension of life expectancy. For these reasons, the creation of new and more effective therapeutic strategies has become an urgent goal for medicine. In fact, a large part of the international scientific community has focused its research on the discovery of efficient drugs [21–24] for their treatment. In any case, the heterogeneity of the possible structures to be explored as potential hits is very high, which makes the discovery of new compounds by trial and error very expensive and slow. For instance, ChEMBL, https://www.ebi.ac.uk/chembldb, is one of the biggest databases [25] of chemical compounds with biological activity with more than 11,420,000 activity data for >1295,500 compounds, and 9844 targets. All this information comes from very heterogeneous preclinical assays. Specifically, in the case of pre-clinical assays of compounds for neurodegenerative diseases, ChEMBL [26–28] contains more than >10000 outcomes. Moreover, ChEMBL datasets of neuroprotective compounds cover multiple biological activity parameters [29–31] (Potency, IC_{50}, K_i, K_m, *etc.*), different cell lines, organisms of the protein target, organism of assay, *etc.* Nevertheless, working with these large volumes of information (Big Data) is a very complex task.

This picture, together with some disappointing results in clinical trials, makes the prediction of drug candidates with computational techniques interesting. One possibility is the use of NIFPTML (Network Invariants, Information Fusion, Perturbation Theory, Machine Learning) models to solve this kind of problems in drug discovery. In fact, NIFPTML models have been reported applied to organic synthesis, medicinal chemistry, biotechnology, nanotechnology, biomedical engineering, chemistry of materials or biomolecular systems, among others [32–35]. Nonetheless, there are no reports of general purpose NIFPTML models for the prediction of neuroprotective compounds considering compound structure, target protein, Protein Interaction Networks (PINs) from different brain regions, expression of target proteins in specific brain regions associated with degenerative neurological diseases and assay conditions.

As stated above, Riluzole is an FDA-approved drug for ALS, with limited liver side effects. Its neuroprotective mechanism is still unclear, but it shows promising results towards treating various neurodegenerative diseases and has diverse therapeutic properties, including anticonvulsant and antidepressant effects. Various Riluzole derivatives and analogues have been also described and tested. The effect of the substitution on C-6, changing the trifluoromethoxy group, and on the endocyclic nitrogen atom N-3 has been extensively studied [36]. The substitution on the C-2 amino group has also been reported in the context of brain diseases [37]. More recently, this type of Riluzole analogues have shown improved use-dependent inhibition of skeletal muscle sodium channels [38]. However, the effect of the substitution on C-4 has not been studied so far, beyond the introduction of the trifluoromethoxy group [36]. In order to find novel Calmodulin inhibitors with neuroprotective activity, we decided to modify the benzothiazole scaffold with different functional groups at the C-4 position. We reasoned that the introduction of substituted aromatic rings at C-4, as well as other functional groups, may affect the binding mode in the CaM-SK4 binding pocket. Herein we report the synthesis and *in vitro* assays of 4-substituted Riluzole derivatives **1–3** (Fig. 1) using a new Förster Resonance Energy Transfer (FRET) biosensor for monitoring Ca^{2+}-dependent calmodulin-ligands interactions. Docking analysis has been carried out to elucidate how the introduction of substituents at the C-4 position could affect the binding mode with the target protein. In addition, we have developed the first general-purpose NIFPTML model for the prediction of neuroprotective compounds that considers the above mentioned aspects with regard to CaM related diseases.

2. Results and discussion

2.1. Synthesis of new Riluzole derivatives

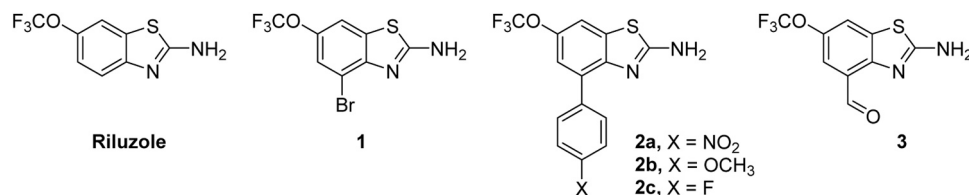Riluzole or 6-(trifluoromethoxy)benzothiazol-2-amine is a derivative

Fig. 1. Structures of Riluzole and its derivatives.

of benzothiazole. The synthesis of their analogues **1–3** was carried out as shown in Scheme 1. First, the 4-bromo derivative **1** was prepared, as the presence of a bromine atom offers an excellent platform to access more complex structures *via* transition-metal catalyzed cross-coupling reactions. The construction of the benzo [*d*]thiazol-2-amine framework was achieved following a procedure similar to that reported for the preparation of Riluzole [36]. Thus, the one-pot reaction of commercially available 2-bromo-4-(trifluoromethoxy)aniline with thiocyanogen, generated *in situ* from potassium thiocyanate and bromine in acetic acid, provided **1** in moderate yield. Then, we continued with the functionalization of the benzothiazole ring *via* Suzuki-Miyaura reaction [39]. Pd (0)-mediated C-4 arylation of **1** was carried out with different boronic acids using Pd(PPh$_3$)$_4$ as catalyst in the presence of sodium carbonate as a base to afford 4-aryl Riluzole derivatives **2a-c** in excellent yields. Finally, a metalation-electrophilsequence allowed the introduction of a carbaldehyde at the C-4 position. Lithiation of **1** via halogen-metal exchange reaction (*n*-BuLi, THF, −78ºC), followed by reaction with *N,N*-dimethylformamide provided aldehyde **3** (Scheme 1).

2.2. Biological assay of Riluzole derivatives

The preparation and use of a biosensor for monitoring Ca^{2+}-dependent interactions between CaM and peptide targets is described in this section. It involves the fusion of CaM with a target peptide and the incorporation of fluorescent proteins as part of the biosensor. FRET was used to monitor Ca^{2+}-dependent CaM interactions with peptide targets. The assay allows investigating the binding of CaM to the target peptide in response to changes in calcium ion concentration. FRET index for the original YC-Nano15 varied from 2.7 ± 0.1 with no Ca^{2+} added, to 27.0 ± 0.1 in the presence of 24 nM free Ca^{2+}, with an EC$_{50}$ of 2.1 ± 0.1 nM. The effect of Riluzole (and related compounds at a final concentration of 100 μM) was negligible on Ca^{2+} sensitivity or FRET amplitude with the exception of compound **2a** which produces an inhibition of the amplitude of 17% (Supplementary Materials Table S1). In contrast, the response to Ca^{2+} for the SK4 biosensor, whose FRET index increased from 2.0 ± 0.5–8.7 ± 0.1, changed in presence of Riluzole. The response to Ca^{2+} presented two resolvable phases: Phase I at low Ca^{2+} concentrations (0–150 nM) and Phase II at higher concentrations (Fig. 2).

Riluzole clearly affected Phase II, which presented an EC$_{50}$ of 302.4 ± 22.3 nM and 221.3 ± 31.1 nM in the absence and presence of this drug. This response resembles the effect of this compound on SK2 channel function [15] (Supplementary Materials Figure S6B), suggesting that it may be a useful tool for screening CaM acting compounds.

We next performed Ca^{2+} titrations using the SK4 biosensor in the presence of Riluzole analogs. We focused on the effect on Phase II, since the reduced signal to noise ratio of this portion of the response compromised the assessment of effects on Phase I. Contrary to Riluzole, none of the compounds investigated caused a clear change in Ca^{2+} sensitivity on Phase I, however in Phase II, there was a clear reduction (see Table 1). The largest reduction (80%) was observed for compound **2b**.

2.3. NIFPTML model development and application

Predictive NIFPTML methods are used in drug discovery for different diseases. They handle vast datasets like ChEMBL, extracting patterns from diverse compound structures. They streamline drug discovery, reducing time and costs compared to trial-and-error approaches. In this project, these methods were used to predict potential drug candidates considering compound structure, target proteins, Protein Interaction Networks (PINs), protein expression in specific brain regions, and assay conditions. These models can be personalized, matching drugs to patients based on genetic profiles and specific disease characteristics. In essence, NIFPTML offers a data-driven computational solution to expedite the discovery of neuroprotective compounds, potentially improving treatments for CaM-related neurodegenerative diseases.

Eq. 1 shows the proposed NIFPTML-LDA model. As reported by Hill and Lewicki [40] the forward-stepwise strategy of variable selection was applied to detect the more important perturbations on different conditions. The model presented high values of specificity (Sp= 80.52%), sensitivity (Sn = 80.83%), and overall accuracy (Ac = 80.55%), in training series. In addition, the model displayed similar values of Sn, Sp, and Ac in external validation series. Table 2 summarizes these parameters. The value of Chi-square is χ^2 = 146346.05with a p-level < 0.05 indicating that the classifier performs a statistically significant separation of both classes f(v$_{ij}$)$_{obs}$= 0 *vs.* f(v$_{ij}$)$_{obs}$= 1. It is important to mention
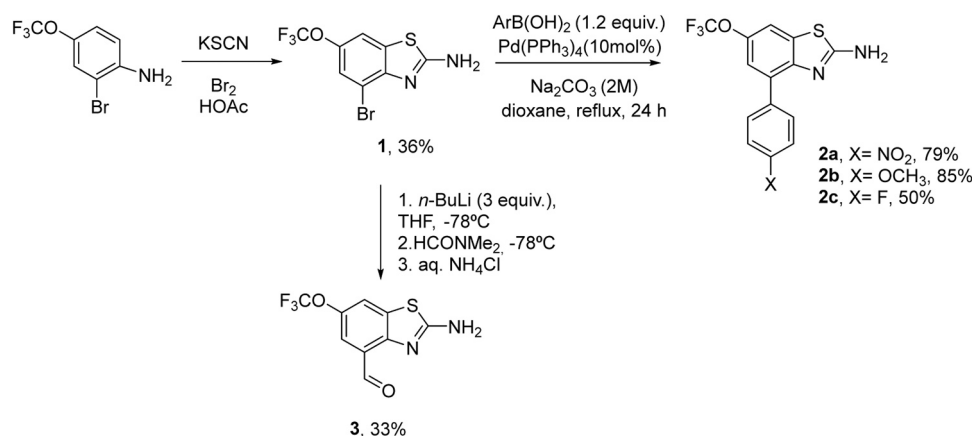
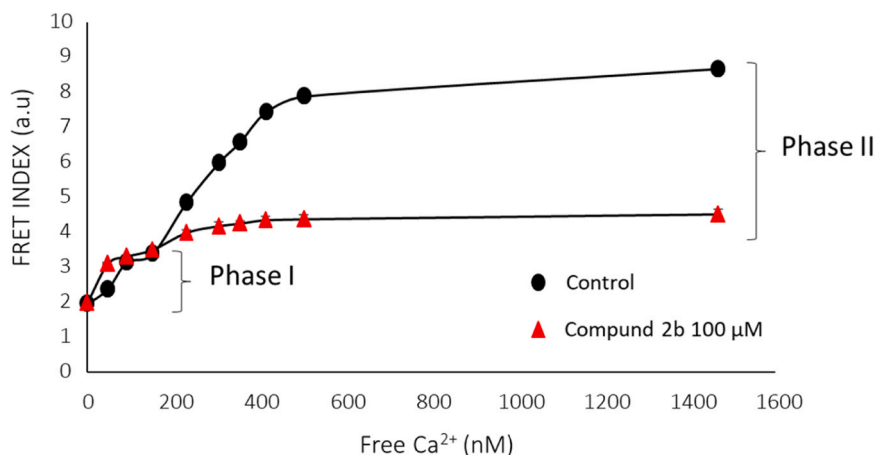**Scheme 1.** Synthesis of 4-substituted Riluzole derivatives **1–3**.

Fig. 2. Ca^{2+} titration in the presence (red) and absence (black) of compound **2b**. A two binding site equation was fitted to the data.

Table 1

Relative amplitude in the presence of the indicated compounds at 100 μM. Mean ± standard error of the mean (SEM) (n ≥ 3). Total inhibition is the relative reduction of the combined Phase I and Phase II amplitudes.

Compound	Total inhibition	Phase I Inhibition	Phase II inhibition
Riluzole	7.9 ± 2.3	- 2.3 ± 2.3	10.7 ± 3.3
1	24.1 ± 4.6	-2.1 ± 2.6	31.2 ± 5.6
2a	34.5 ± 1.6	12.1 ± 7.0	40. 6 ± 1.2
2b	62.5 ± 2.8	-5.3 ± 3.9	80.9 ± 2.6
2c	33.1 ± 1.3	23.7 ± 3.1	35.6 ± 1.5
3	25.2 ± 1.1	3.3 ± 4.7	31.2 ± 1.6

that the obtained values are in the range considered useful for classification models with application in Medicinal Chemistry.

$$
\begin{aligned}
f(v_{ij})_{calc} = & -3.880183 + 15.366209 * f(v_{ij})_{ref} - 17.196505 \\
& * Sh_2(Drug_i) + 1.287039 * Sh_1(Prot_y) - 2.784924 \\
& * Sh_2(Prot_y) - 2.140295 * Sh_4(Prot_y) - 8.984754 \\
& * Sh_5(Prot_y) + 48.607304 * Sh_0(PIN) + 22.6191659 \\
& * \Delta Sh_2(Drug_i, c_j) - 2.320339 * \Delta Sh_1(Prot_y, c_j) - 21.381213 \\
& * \Delta Sh_0(PIN, c_j)
\end{aligned}
$$

(1)

$n = 794108 \chi^2 = 146346.05 p < 0,05$

The first value that appears in the equation of this IFPTML-LDA classification model is the intercept, in addition, the equation contains the reference function, and the Perturbation Theory Operators (PTO) that are added depending on the boundary conditions (ΔSh_k, c_j), as well as the structural descriptors (Sh_k) with the corresponding predicted coefficients. A detailed explanation about all the input variables analyzed is shown in Table 3.

The specific molecular descriptor of the drug included in the model is the min-max scaled Shannon's entropy used to measure polar surface area features of the drug. In addition, the model included 4 variables related to the descriptors of the target protein sequences and a variable associated with the PIN molecular descriptor. All these variables were described in terms of Shannon's entropy and were determined as explained in the materials and methods section. In addition, the model included the PTOs corresponding to the description of the drug, the target protein, and the PIN. The output of the model $f(v_{ij})_{calc}$ is a scoring function of the value v_{ij} of biological activity of the ith drug that interacts with a protein expressed in one of the regions of the cerebral cortex associated with degenerative neurological diseases in the different combinations of conditions of assay c_j. For a Linear Discriminant Analysis (LDA) model $f(v_{ij})_{calc}$ does not take values of 0–1 and it does not count as probability. Nevertheless, for a given value of $f(v_{ij})_{calc}$ the LDA algorithm can analyzes the respective values of posterior probabilities p ($f(v_{ij})_{pred} = 1$). The LDA algorithm uses the Mahalanobis's distance metric to calculate these probabilities [41,42]. By calculating p(f $(v_{ij})_{pred} = 1$), whether the compound is active with $f(v_{ij})_{pred} = 1$ (when p(f $(v_{ij})_{pred} = 1) > 0.5$) or not can be observed. By counting the number of cases with $f(v_{ij})_{pred} = 1 = f(v_{ij})_{obs} = 1$ or -1 (correct classifications) vs. f $(v_{ij})_{pred} \neq f(v_{ij})_{obs}$ (incorrect classification), the specificity and sensitivity of the model was determined [43].

The model could be used to predict the activity of a new compound. This is why it is essential to apply PT operators that are sensitive to fluctuations in several conditions at the same time. Those refer to the expected probability, p(f(v_{ij}, c_j)$_{expt} = 1$, and the average values $\langle D_k(c_j) \rangle$ for multiple combinations of conditions at the same time. It should be noted that these values change for different activities, like Zone of Inhibition, MIC_{50}, LD_{50}, Activity (%), etc. As a consequence, the model has the ability to predict several activity parameters for a given compound.

2.3.1. Comparison with other models

Several cheminformatics models have been reported in the literature for the discovery of compounds that present biological activity against neurodegenerative diseases. Table 4 reports a comparison between the

Table 2

Classification matrix. Statistical parameters to characterize the built IFPTML-LDA model.

Data Set	Observed Classification	Stat. Param.[a]	Pred. Stats.	Predicted Classification		
				n_j	$f(v_{ij})_{pred} = 0$	$f(v_{ij})_{pred} = 1$
Train	$f(v_{ij})_{obs} = 0$	Sp (%)	80.52	443348	432035	104548
	$f(v_{ij})_{obs} = 1$	Sn (%)	80.83	152234	11313	47686
	Total	Ac (%)	80.55	595582		
Val.	$f(v_{ij})_{obs} = 0$	Sp (%)	80.67	148018	144268	34576
	$f(v_{ij})_{obs} = 1$	Sn (%)	80.95	50508	3750	15932
	Total	Ac (%)	80.69	198526		

[a] Statistical parameters of the model: Sp: Specificity; Sn: Sensitivity; Ac: Accuracy; n: the number of cases used to train the model.

Table 3

Input variables in the best model found.

Label (c_j)	Operator Formula	Operator Information
Biological activity (c_0)	$f(v_{ij})_{ref} = n(f(v_{ij}) = 1, c_0)/n_{c_0,a}$	A priori probability that a compound that interacts with a protein expressed in one of the brain regions of interest is active, for the specific condition c_0.
Drug's Chemical structure	$Sh_2(Drug_i) = -p(D_{02}(Drug_i)) \cdot \log[p(D_{02}(Drug_i))]$	It accounts for variability of chemical structure information of the drugs in terms of lipophilicity expressed as TPSA surface area.
Protein Sequence	$Sh_k(Prot_y) = -p(D_k(Prot_y)) \cdot \log[p(D_k(Prot_y))]$	It accounts for variability of the sequence of the target protein. The information has been quantified to the neighboring aa of the sequence at a distance of X amino acids. (X= 1, 2, 4 or 5).
PIN	$Sh_k(PIN) = -\sum_{y=1}^{y=ymax} p_k(Prot_y) \cdot \log[p_k(Prot_y)]$	It accounts for variability on the structure of the PIN of the brain cortex region.
PTO of the molecular descriptor PSA (c_j)	$\Delta Sh_{02}(Drug_i, c_j) = Sh_{02}(Drug_i) - \langle Sh_{02}(c_j) \rangle$	It is a PTO and by itself calculates how much the value of the descriptor $Sh_{02}(Drug_i)$ of a compound deviates from the average of said value in the database, for the conditions c_j.
PTO of protein sequences expressed in the brain (c_j)	$\Delta Sh_{01}(Prot_y, c_j) = Sh_{01}(Prot_y) - \langle Sh_{01}(c_j) \rangle$	It is a PTO and by itself calculates how much the value of the descriptor $Sh_{01}(Prot_y)$ of a compound deviates from the average of said value in the database, for the conditions c_j.
PTO of PIN (c_j)	$\Delta Sh_{00}(PIN, c_j) = Sh_{00}(PIN) - \langle Sh_{00}(c_j) \rangle$	It is a PTO and by itself calculates how much the value of the descriptor $Sh_{00}(PIN_t)$ of a compound deviates from the average of said value in the database, for the conditions c_j.

$c_j = (c_0, c_1, c_2, c_3)$, c_0 = biological activity (Property (unit) measured to quantify the biological activity of each compound), c_1 = cell line, c_2 = brain region, c_3 = organism of assay.

present model and some of them. In this comparative study, we included 14 models, most of which (64%) are based on heterogeneous series of compounds (models 2 and 3, models 5–11). However, two models were based on heterocycles (models 12 and 13), another specifically in β-amyloid aggregation-inhibiting heterocycles (model 4), and another one on structurally diverse succinimides (model 14). Regarding the number of cases, eight models can be found that include hundreds of cases, which represent 57%, two of the models include less than one hundred and fifty cases, while the rest (21%) include much larger amounts. It should be noted that the model reported in this paper fits a very complex and notably larger data set of cases compared to the other models (n> 794100). Regarding the complexity of the models, most of

them include less than 10 variables, such as models 1, 5, 7 and 10; while other models, such as 2, 4, 11 and 13, include between 10 and 13 variables. There is a great variety of techniques constructing the reported models. Some of the most repeated techniques are the LDA, used in models 2, 9 and 10; the ANNs used in models 5, 7 and 11; and the SVM that is reported in models 1, 12, 13. Regarding the validation techniques, in most cases cross validation is used (8 models, 57%), only three models employ external validation (21%) and three other models use both validation techniques (21%). The reported models have been built to find compounds with biological activity that can be used in the treatment of various neurodegenerative diseases. Most of them are focused on the search for drugs for the treatment of Alzheimer's (5 models), for example, models 2, 3, 4, 7 and 9; although there are models that search for compounds against other diseases such as Parkinson's (model 10), amyotrophic lateral sclerosis (model 8) and Huntington's disease (model 1). There are also some described models that pursue active compounds against various diseases, such as models 5, 6, 11, 12, 13, like the model reported here. On the other hand, only 3 models found in the literature are multi output. In fact, the model reported in this study is multi output; that is, it can predict more than one type of biological activity (MIC, IC_{50}, MBC, *etc.*). On the other hand, a distinctive characteristic of our model is that it considers the PIN, and, therefore, the brain region in which the target proteins of the compounds are expressed, which is not considered by any of the reported models. Finally, most of the reported models are multi-target models, like the model presented in this study.

2.3.2. IFPTML prediction of new Riluzole derivatives vs. calmodulin

The prediction was carried out successfully and 6 different biological activities were calculated for the 6 compounds. The 1692 assays were composed by combining the following assay components: 14 different proteins, 1 tested cell, 6 brain regions, and 5 organism types. Moreover, the assays types consisted of ADMET (A), Functional (F) and Binding (B) assays. The different results for $\Delta f(v_{ij};v_{CaMj})_{calc}$ (%) are displayed on Fig. 3 as a visual representation. Binding assays against CaM were selected for this results' diagram, because those are the assays carried out on the experimental section (Supplementary Materials Table S5). Additionally, results can be seen depending on the different brain regions, including: Superior Frontal Gyrus represented in yellow color, Prefrontal Cortex represented in navy blue, Middle Temporal Gyrus represented in green, Hippocampus represented in orange color, Entorhinal Cortex represented in pink and Visual Cortex represented in light blue. Furthermore, Riluzole's $f(v_{ij};v_{yj})_{calc}$ values are represented with a circle, whereas the $f(v_{ij};v_{yj})_{calc}$ values of the proposed derivatives are represented with a square. The combining outcomes were characterized in two different ways. On the one hand, when the circle is inside the square, it means that the derivative resulted in an equal or better response than Riluzole. On the other hand, when the circle is outside the square, it means that the derivative resulted in a worse response than the Riluzole. The searched outcome consist in the compound has the same or better response than Riluzole. This is because the proteins included in this prediction are composed of CaM and its kinases, so the aim is to observe if the derivative has better binding capacity than Riluzole. Taking that into account, the green and red colors were chosen depending on the response sought, with green being the satisfactory result.

These results showed that compounds **1** and **2c** resulted on equal or better response than Riluzole on the next brain regions: Superior Frontal Gyrus, Prefrontal Cortex, Middle Temporal Gyrus, Hippocampus and Entorhinal Cortex and Visual Cortex. These compounds were predicted to have equal or better capacity than the Riluzole to bind to CaM.

2.3.3. IFPTML prediction of new Riluzole derivatives vs. human brain proteome

The prediction was carried out successfully. In total 369 different biological activities were calculated for the 6 compounds. The 78120

Table 4
Comparison to other Chemoinformatic models of neuroprotective compounds.

m[a]	CT[a]	n[b]	Var.[b]	Tech.[c]	Val.[d]	ND[e]	MO[f]	Net.[g]	MT[h]	Acc.[i]	Ref.
0	HS	794108	10	LDA	i	MD	Yes	Yes	Yes	80.6	here
1	HS (TCM)	-	8	MLR SVM	i, ii	HD	No	No	Yes	-	[44]
2	HS	1432	12	LDA	i	AD	No	No	No	88.3	[45]
3	HS	18741	-	NB RP	i, ii	AD	No	No	Yes		[46]
4	HBAAI	252	13	PLS	i, ii	AD	No	No	No		[47]
5	HS	121	5	ANN AMT	i	AD P	No	No	No		[48]
6	HS	134	-	PRKDEM	ii	MD	Yes	No	Yes		[49]
7	HS	213	4	BMLR ANN	ii	AD	No	No	Yes		[50]
8	HS	946	-	PM-HDE	ii	ALS	No	No	No		[51]
9	HS	3381	5	LDA	i	AD	Yes	No	Yes		[52]
10	HS	230	6	LDA, QDA MLP	ii	P	No	No	Yes		[53]
11	HS	550	10	ANN	ii	P AD	No	No	No		[54]
12	HC	449	-	SVM	ii	MD	No	No	No		[55]
13	HC	413	10	SVM-RFE	ii	MD	No	No	Yes		[56]
14	SDS	327	-	MNN	ii	E	Yes	No	Yes		[57]

[a] CT = Compound Type, P = Peptide, HS = Heterogeneous Series of compounds, HS (TCM) = Heterogeneous series of compounds present is Traditional Chinese Medicine database, HBAAI = Heterocyclic β-Amyloid Aggregation Inhibitors, HC = Heterocycles, SDS = Structurally Diverse Succinimides.

[b] Total number of cases in training and/or validation series and Vars. = Variables in the model.

[c] Technique: LDA = Linear Discriminant Analysis, ANN = Artificial Neural Network, MLR = Multiple Linear Regression, SVM = Support Vector Machine, NB = Naive Bayesian, RP = Recursive Partitioning, PLS = Partial Least Squares regression technique, AMT = Advanced Multilinear Techniques, PRKDEM = Parzen–Rosenblatt kernel Density Estimation Method, QDA = Quadratic Linear Discriminant Analysis, MLP = Multi-Layer Perceptron, SVM-RFE = SVM-based Recursive Feature Elimination, MNN = Modular Neural Network.

[d] Validation methods: i) external predicting series, ii) cross-validation,

[e] ND = Neurodegenerative Disease, MD = Multiple Diseases (five or more), HD = Huntington's Disease, AD = Alzheimer's Disease, P = Parkinson's, ALS = Amyotrophic Lateral Sclerosis, E = Epilepsy,

[f] MO = Multi Output: multi-output models are those able to predict more than one type of biological activity (MIC, IC_{50}, MBC, *etc.*).

[g] Net. = PINs: Models that consider the PIN.

[h] MT = MultiTarget models,

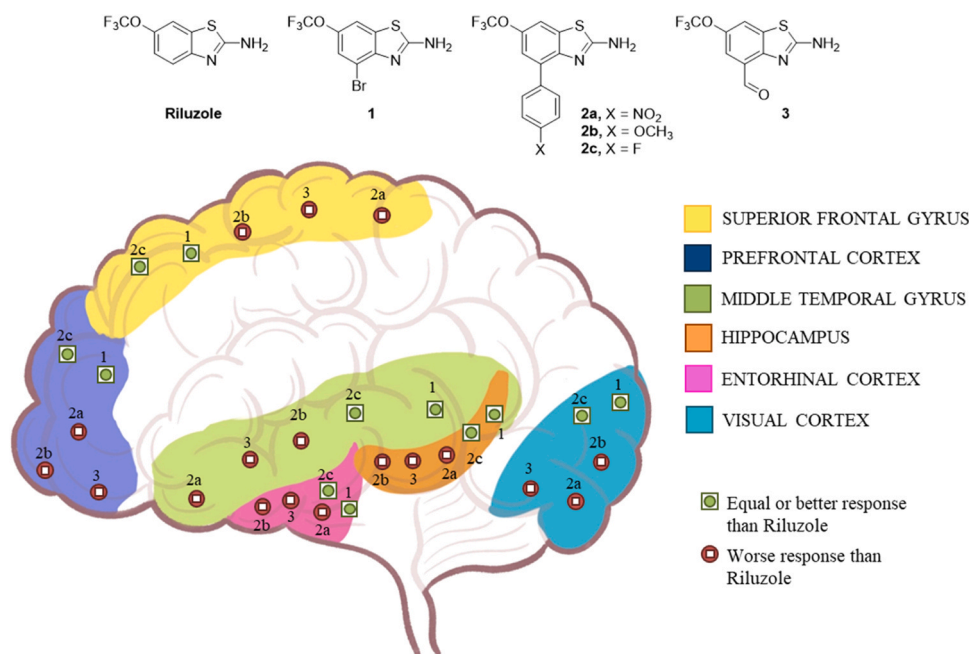[i] Acc. = Accuracy of the model in training series.

Fig. 3. Prediction results of Riluzole derivatives *vs.* CaM protein.

assays were by composed combining the following assay components: 124 different genes, 134 different proteins, 70 tested cells, 6 brain regions and 28 organism types. Moreover, the assays types consisted of ADMET (A), Functional (F), Binding (B) and Unclassified (U) for those assays with no specified type. The different results for $\Delta f(v_{ij};v_{CaMj})_{calc}$ (%) are shown on Fig. 4 as a visual representation. Binding assay types were selected for this results' diagram, because those are the assays carried out on the experimental section (Supplementary Materials

Table S10). Additionally, results can be seen depending on the different brain regions (the color code indicated in Fig. 3 has also been used). Furthermore, Riluzole's $f(v_{Riluzolej}; v_{yj})_{calc}$ values are represented with a circle, whereas the $f(v_{ij}; v_{yj})_{calc}$ values of the derivatives are represented with a square. The combining outcomes were characterized in two different ways: on the one hand, when the circle is inside the square, it means that the derivative resulted in an equal or worse response than the Riluzole; and, on the other hand, when the circle is outside the square, that means that the derivative resulted in a better response than the Riluzole. The searched outcome consists in the compound having an equal or worse response than Riluzole. This is because the proteins included in this prediction are composed of proteins that are not related to CaM, and, thus, there is not point on searching the binding result among these proteins and the drug. Taking that into account, the green and red colors were chosen depending on the searched response, being the green one the satisfactory result.

There results showed that compounds **1–3** resulted on equal or worse response than Riluzole on the next brain regions: Superior Frontal Gyrus, Prefrontal Cortex, Middle Temporal Gyrus, Hippocampus and Entorhinal Cortex. Nevertheless, on the Visual Cortex, only **1** and **2c** showed equal or worse response than Riluzole. These compounds were predicted to have less capacity than the Riluzole to bind to those proteins.

2.4. Molecular docking

Among the computational techniques currently used in modern drug development pipelines, docking is one of the most prominent and useful ones. Docking algorithms are able to determine the binding mode of a ligand in a receptor by finding the orientation of the ligand that optimizes a scoring function, which in the context of docking, represents the affinity between receptor and ligand. The main power of docking lies in its efficiency, as it allows to evaluate large libraries of candidate drugs in a fraction of the time and cost that an experimental screening would require. However, in combination with high quality 3-dimensional structures of the receptor protein, it can also be used to elucidate the way in which the ligand interacts with its receptor at an atomic resolution, and thus provides vital information for rational drug design. In this study we report the Docking analysis of the present series of compounds.

The best conformation for Riluzole obtained from Glide docking calculations is shown in Fig. 5. The bound conformation is highly similar to the structure previously reported in [15], characterized by Riluzole's -OCF$_3$ group buried in the hydrophobic pocket formed between CaM and the SK4 channel, and the amino group forming a hydrogen bond with residue E54 at the opening of the pocket. However, unlike the SK2 structure, where the border of the binding pocket is formed by amino acids Q470, N474 and L480 from the SK2 C-terminus, in the SK4 structure, this border is instead formed by amino acids A184 and L185 of the S$_{45}$A helix.

The binding conformations of derivatives **2a-c** (Fig. 5E) are very similar to each other, and also share many similarities with Riluzole. The -OCF$_3$ of these derivatives lies buried in the binding pocket similar to Riluzole, in contact with CaM's hydrophobic residues such as F19, I27, L32, I63 and F68. However, in the case of these derivatives, the amino group at the head of the benzothiazole core points towards SK4 residues A184 and L185, and the opening of the border is instead occupied by the different substituents added to these derivatives.

Similarly, the conformations of the **1** and **3** (Fig. 5D) derivatives also show high similarity with one another. Instead of forming a hydrogen bond with residue E54 like Riluzole, the amino group in these ligands interacts with the E83 residue of the CaM C-lobe directly in front of the binding pocket opening. As a result, the positioning of these ligands in the hydrophobic core is not as deep as Riluzole and the other derivatives studied in this work.

To validate the conformations obtained from this docking study, BPMD simulations were run for every ligand and frames from the last 2 ns of each metadynamics simulation were collected. The stability of the docked conformations was measured in terms of the "CompScore" [58]. This score is computed from two contributions: the "PoseScore", which is computed from the RMSD between the ligand heavy atoms in each frame of the simulation and the reference docked configuration after aligning the protein C$_\alpha$ atoms, and the "ContactScore", which computes the ratio of native non-covalent interactions that are preserved during the simulation. The final CompScore decreases as the average RMSD becomes smaller and as the ratio of conserved contacts increases, so smaller numbers are indicative of more stable conformations. The precise score is computed as follows:
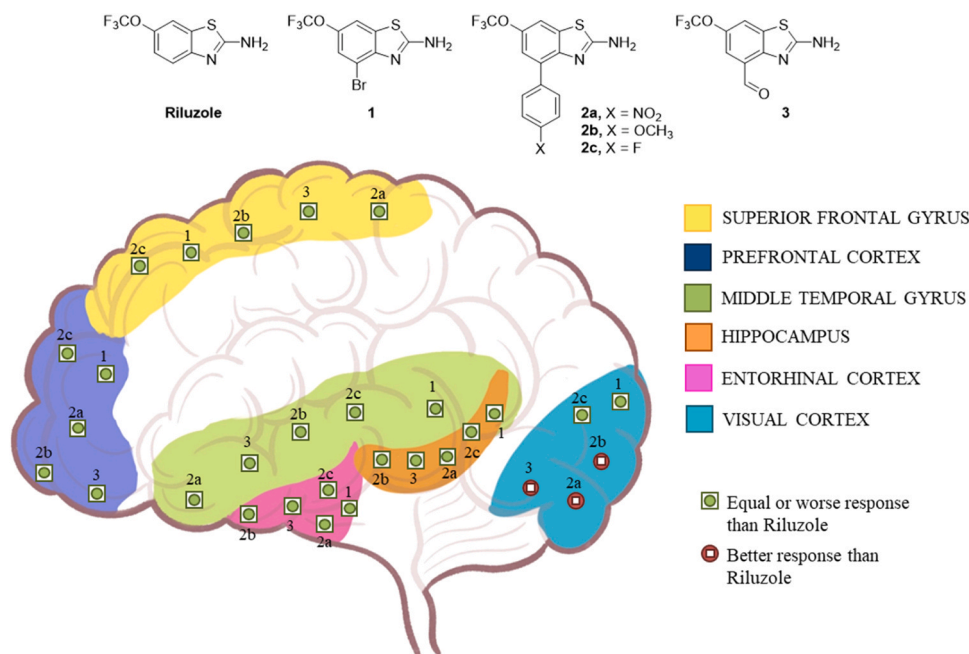

Fig. 4. Prediction results of Riluzole derivatives *vs.* full human brain proteome (excluding CaM).
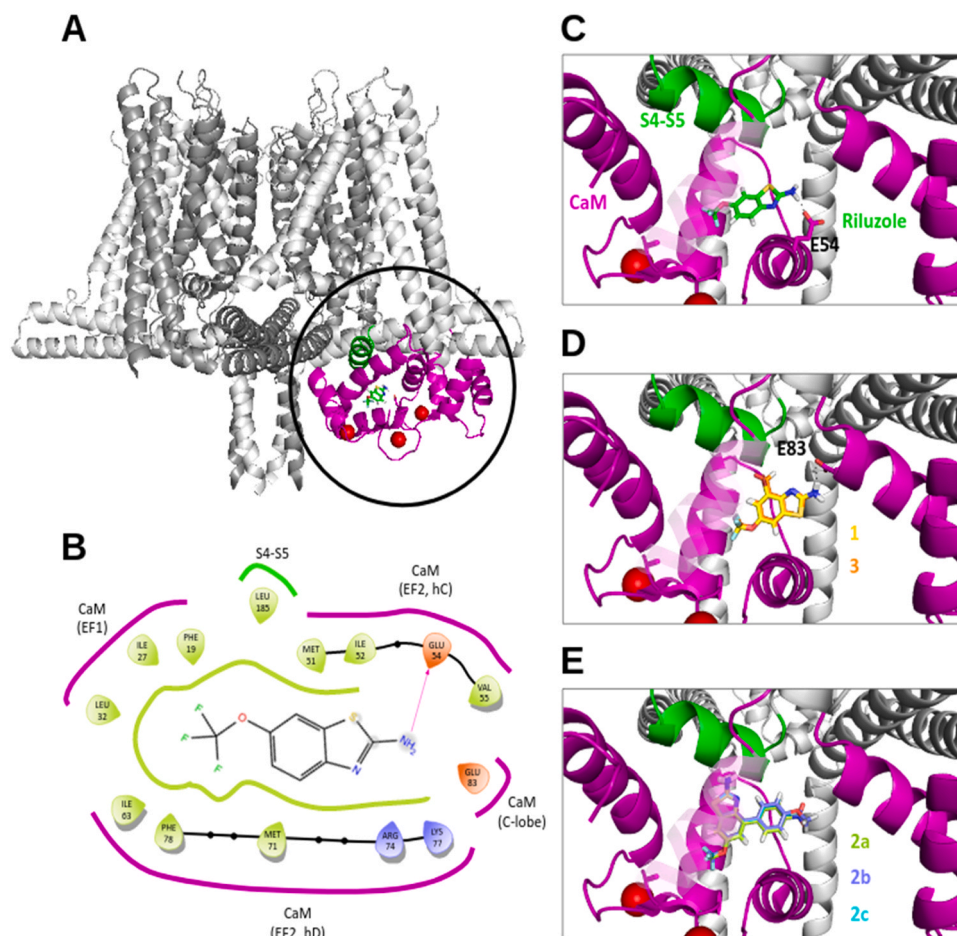
Fig. 5. Docking of Riluzole derivatives: (A) Localization of the binding site in the SK4 structure, which lies in the interface between CaM and the S45A helix. (B) Schematic representation of the interactions between Riluzole and the amino acids in its environment. (C) Best binding pose of Riluzole determined through docking, characterized by a -OCF₃ group buried in the hydrophobic pocket of the binding site, and a hydrogen bond formed between its amino group and amino acid E54. (D) Best binding poses of derivatives **1** and **3**, which share similar optimal poses, with a hydrogen bond between their amino group and amino acid E83. (E) Best binding poses of derivatives **2a**, **2b** and **2c**, which also share similar optimal binding poses.

$$CompScore = PoseScore - 5xContactScore \qquad (2)$$

The CompScore obtained by averaging over the last 2 ns from each of the 10 simulations for Riluzole and its derivatives is given in Fig. 6. Based on the computed metrics, all ligands display behavior of stable conformations, with average RMSD below the 2 Å threshold that is often used to differentiate stable and unstable poses [58,59], and high ratios of conserved native interactions. Overall, CompScores measured for all ligands are low enough to classify all analyzed conformations as stable.

2.5. Overall result's summary

Table 5 presents a comprehensive summary of the findings derived from the current research, denoted by a ✓ symbol when demonstrating results deemed acceptable or comparable to those of Riluzole, and marked as × otherwise. This table encompasses the overarching outcomes of the synthesis of Riluzole derivatives, the results of the biosensor assay employing CaM, the predictive model's findings, and the docking results. Concerning the synthesis aspect, all synthesized molecules were successfully produced. Furthermore, in the context of the biological assay methods, derivatives **1** and **2b** emerged as the most promising when compared to Riluzole. In the NIFPTML study, it was predicted that derivatives **1** and **1c** exhibited activity scores similar to Riluzole within critical brain regions associated with neurodegenerative diseases, such as the Frontal and Temporal regions. Additionally, the NIFPTML model predicted a relatively high degree of selectivity for

CaM, as evidenced by the model's projection of low average affinity scores for 134 other proteins across various brain regions, as determined through binding assays. Furthermore, the results gleaned from the docking study indicated that all synthesized molecules exhibited favorable binding conformations with the CaM protein.

3. Conclusions

In conclusion, this research represents a significant step towards the development of new general purpose strategy for the screening of lead compounds towards the treatment of CaM-related diseases, particularly neurodegenerative conditions. The strategy involves four main stages: (1) chemical synthesis, (2) FRET biosensor development and *in vitro* biological assay of new compounds, (3) Cheminformatics model development and prediction of *in vivo* assays compounds, and (4) Docking studies of drug-target binding. Riluzole was used as the bechmark compound for the case of study presented, even when the methodology is of general use. In this case, palladium and lithium-mediated reactions were succesfully used for the synthesis of the series of five new Riluzole derivatives. Additionally, the *in vitro* evaluation of these derivatives against the CaM protein revealed promising results. Notably, compound **2b** demonstrated the highest inhibition, reducing CaM activity by 80%. Furthermore, the implementation of the NIFPTML algorithm has proven to be a powerful tool for predicting various biological activity parameters, not only against CaM proteins but also related kinases and other
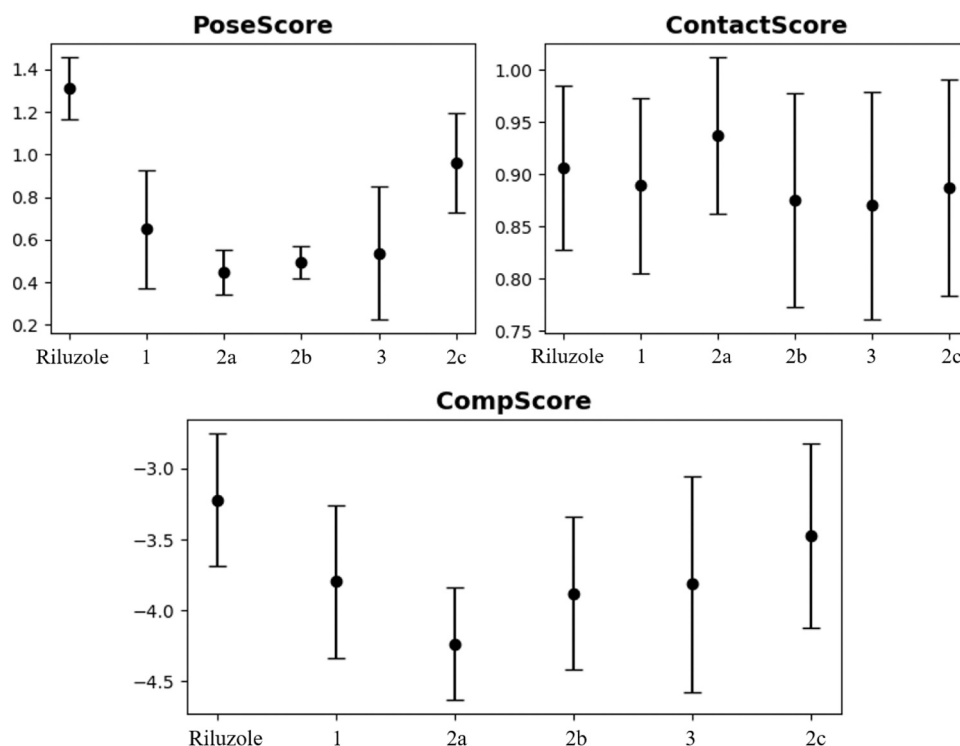
Fig. 6. Mean and standard deviation of the PoseScore, ContactScore and CompScore measured for the last 2 ns of the metadynamics simulations, for Riluzole and all its derivatives.

Table 5
Overall summary of all results.

Compound	Synthesis Results	Biological Assay (*In vitro*)	Predictive Study (*In vivo*)			Docking and MD Study (*In vitro*)
			CaM A.R.	O.P. VCX	O.P. O.R.	
Riluzole	✓	✓	✓	✓	✓	✓
1	✓	✓	✓	✓	✓	✓
2a	✓	✗	✗	✓	✓	✓
2b	✓	✓	✗	✓	✓	✓
2c	✓	✗	✓	✓	✓	✓
3	✓	✗	✗	✓	✓	✓

[a]A.R. = All Brain Regions, O.P. = Other 134 proteins without including CaM, O. R. = Other Brain Regions, without including VCX.

proteins. It marks the first instance where a comprehensive approach integrates vital elements such as drug structure, protein sequence, and PINs in various brain regions, test conditions, and the specific expression profiles of these proteins in brain regions linked to degenerative neurological diseases. This innovative model, boasting high specificity and sensitivity values, offers the potential to significantly streamline the drug discovery process, reducing both time and resource costs. Lastly, the docking study of Riluzole and its derivatives with the CaM-SK4 channel complex has provided valuable insights into their binding conformations, shedding light on potential modes of interaction. Derivatives **1–3** exhibit distinctive binding conformations related to Riluzole. These findings collectively contribute to the advancement of therapeutic strategies targeting CaM-related diseases and underscore the potential of innovative computational and machine learning approaches in drug discovery.

4. Experimental section

4.1. General synthesis of Riluzole derivatives 1–3

All commercial chemicals were reagent grade and were used without further purification unless otherwise specified. Palladium catalysts were commercially available, and were used without further purification: Pd (PPh$_3$)$_4$: 99% purity. All solvents used in reactions were anhydrous and purified according to standard procedures. All air- or moisture-sensitive reactions were performed under argon; the glassware was dried (130 ℃) and purged with argon. TLC was carried out with 0.2 mm-thick silica gel Merck F254 plates. Visualization was accomplished by UV light (λ = 254 nm and 360 nm). Flash column chromatographic separations and purifications were performed on silica Flash P60 (Silicycle), 230–400 mesh ASTM. Final compounds were purified to ≥95% purity as assessed by ^{1}H NMR spectra and analytical liquid chromatography. Melting points were measured in a Büchi B-540 apparatus in unsealed capillary tubes. IR spectra were obtained using Attenuated Total Reflection (ATR) in a JASCO FT/IR 4100 in the interval between 4000 and 400 cm^{-1} with a 4 cm^{-1} resolution. Only characteristic bands are given in each case. ^{1}H and ^{13}C NMR spectra were recorded at 20–25 ℃ on either a Bruker AC-300 spectrometer (300 MHz for ^{1}H and 75.5 MHz for ^{13}C) and on a Bruker AC-500 spectrometer (500 MHz for ^{1}H and 125.7 MHz for ^{13}C). Chemical shifts are reported in parts per million (ppm) relative to an internal solvent reference. Recorded peaks are listed in the order multiplicity (s, singlet; d, doublet; dd, doublet of doublets; m, multiplet), coupling constants, and number of protons. Assignments of individual ^{13}C and ^{1}H resonances are supported by DEPT experiments and 2D correlation experiments (COSY, HSQCed or HMBC) when necessary. High resolution mass spectra (HRMS) were performed by the Mass Spectrometry General Service at the University of the Basque Country using an ultra-performance liquid chromatograph (Acquity UPLC, Waters Chromatography.), in tandem with a QTOF mass spectrometer (SYNAPT G2 HDMS, Waters Chromatography), with an electrospray ionization source in a positive mode.

4.1.1. Synthesis of 4-bromo-6-(trifluoromethoxy)benzo[d]thiazol-2-amine (1)

KSCN (1.27 g, 13.07 mmol) and a solution of Br_2 (0.17 mL, 3.27 mmol) in HAcO (2 mL) were added to a solution of 2-bromo-4-(trifluoromethoxy)aniline (0.84 g, 3.27 mmol) in HAcO (7 mL) at room temperature and the mixture was stirred for 16 h. The reaction was quenched with saturated $NaHCO_3$ (150 mL) and the aqueous phase extracted with AcOEt (2 ×50 mL). The combined organic extracts were washed with saturated NaCl (2 ×50 mL), dried over anhydrous Na_2SO_4 and concentrated *in vacuo*. Purification by column chromatography (silica gel, petroleum ether/AcOEt 7:3) afforded **1** as a pale yellow solid (368.2 mg, 36%): mp. (petroleum ether/ AcOEt): 196–199 °C. IR (ATR): 3434, 3060, 1633, 1531, 1443 cm^{-1}. ^{1}H NMR (300 MHz, DMSO-d_6): δ 7.46 (s, 1 H), 7.82 (s, 1 H), 8.04 (br s, 2 H). ^{3}C{^{1}H} NMR (75.5 MHz, DMSO-d_6): δ 109.9, 114.5, 120.6 (q, J = 256.1 Hz), 122.5, 132.3, 142.0 (q, J = 2.2 Hz), 150.7, 168.7. MS (ESI) m/z (rel intensity): 315 (MH$^+$ + 2, 100), 313 (MH$^+$, 97). HRMS (ESI-TOF): calcd for $C_8H_5BrF_3N_2OS$ [MH$^+$]: 312.9258; found, 312.9265.

4.1.2. Suzuki-Miyaura reaction. Synthesis of 4-aryl-6-(trifluoromethoxy) benzo[d]thiazol-2-amines 2a-c General procedure

To a solution of 4-bromo-6-(trifluoromethoxy)benzo [d]thiazol-2-amine (**1**) (1 mmol) in dry dioxane, the adequate boronic acid (1.2 mmol), a solution of Na_2CO_3 2 M (6 mmol), and Pd(PPh$_3$)$_4$ (0.1 mmol) were sequentially added. The reaction mixture was stirred under reflux for 24 h and, then, the solvent was removed and the resulting mixture was dissolved in AcOEt (10 mL). The organic phase washed with water (2 ×10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated under vacuum. Purification by column chromatography (silica gel, petroleum ether/AcOEt 7:3) afforded the corresponding 4-arylbenzothiazol-2-amines **2a-c**.

4.1.3. 4-(4-nitrophenyl)-6-(trifluoromethoxy)benzo[d]thiazol-2-amine (2a)

According to General Procedure, 4-bromo-6-(trifluoromethoxy) benzo [d]thiazol-2-amine (**1**) (80.1 mg, 0.26 mmol) in dry dioxane (10 mL) was treated with (4-nitrophenyl)boronic acid (51.3 mg, 0.31 mmol), a solution of Na_2CO_3 2 M (0.8 mL, 1.56 mmol), and Pd (PPh$_3$)$_4$ (29.6 mg, 0.025 mmol). After work-up, purification by column chromatography provided **2a** (71.6 mg, 79%) as a pale yellow solid: m. p. (petroleum ether/ AcOEt): 246–248°C. IR (ATR): 3398, 3282, 3147, 2920, 1537, 1347 cm^{-1}. ^{1}H NMR (300 MHz, DMSO-d_6): δ 7.42 (d, J = 1.8 Hz, 1 H), 7.88–7.91 (m, 3 H), 8.10 (d, J = 8.2 Hz, 2 H), 8.30 (d, J = 8.2 Hz, 2 H). ^{13}C{^{1}H} NMR (75.5 MHz, DMSO-d_6): δ 115.4, 119.6, 120.8 (q, J = 255.9 Hz), 123.6, 128.3, 130.9, 133.8, 142.6 (q, J = 2.2 Hz), 145.1, 146.9, 149.9, 168.8. MS (ESI) m/z (rel intensity): 356 (MH$^+$, 76). HRMS (ESI-TOF): calcd for $C_{14}H_9F_3N_3O_3S$ [MH$^+$]: 356.031; found, 356.0326.

4.1.4. 4-(4-methoxyphenyl)-6-(trifluoromethoxy)benzo[d]thiazol-2-amine (2b)

According to General Procedure 4-bromo-6-(trifluoromethoxy)benzo [d]thiazol-2-amine (**1**) (72.4 mg, 0.23 mmol) in dry dioxane (10 mL) was treated with (4-methoxyphenyl)boronic acid (42.2 mg, 0.28 mmol), a solution of Na_2CO_3 2 M (0.7 mL, 1.38 mmol), and Pd(PPh$_3$)$_4$ (26.7 mg, 0.023 mmol). After work-up, purification by column chromatography provided **2b** (66.9 mg, 85%) as a pale yellow solid: m. p. (petroleum ether/ AcOEt): 194–195 °C. IR (ATR): 3416, 3289, 3108, 2924, 1606, 1510, 1442 cm^{-1}. ^{1}H NMR (300 MHz, DMSO-d_6): δ 3.81 (s, 3 H), 6.98–7.03 (m, 2 H), 7.23–7.24 (m, 1 H), 7.71–7.79 (m, 5 H). ^{13}C{^{1}H} NMR (75.5 MHz, DMSO-d_6): δ 55.6, 113.2, 114.0, 118.8, 120.8 (q, J = 255.3 Hz), 130.7, 130.8, 130.9, 133.3, 142.8 (q, J = 2.2 Hz), 149.5, 159.2, 167.7. MS (ESI) m/z (rel intensity): 341 (MH$^+$, 100), 318 (13). HRMS (ESI-TOF): calcd for $C_{15}H_{12}F_3N_2O_2S$ [MH$^+$]: 341.0572; found, 341.0574.

4.1.5. 4-(4-fluorophenyl)-6-(trifluoromethoxy)benzo[d]thiazol-2-amine (2c)

According to General Procedure, (trifluoromethoxy)benzo [d]thiazol-2-amine (**1**) (45.6 mg, 0.15 mmol) in dry dioxane (10 mL) was treated with (4-fluorophenyl)boronic acid (24.4 mg, 0.17 mmol), a solution of Na_2CO_3 2 M (0.44 mL, 0.87 mmol), and Pd(PPh$_3$)$_4$ (16.8 mg, 0.015 mmol). After work-up, purification by column chromatography provided **2c** (23.9 mg, 50%) as a white solid: IR (ATR): IR (ATR): 3446, 3285, 2921, 2854, 1640, 1506, 1446 cm^{-1}. ^{1}H NMR (300 MHz, DMSO-d_6): δ 7.27–7.28 (m, 2 H), 7.76–7.75 (m, 5 H). ^{13}C{^{1}H} NMR (75.5 MHz, DMSO-d_6): δ 114.0, 115.3 (d, J = 21.1 Hz), 119.2, 120.8 (q, J = 256.7 Hz), 129.9, 131.7 (d, J = 8.3 Hz), 133.4, 134.9 (d, J = 3.0 Hz), 142.7 (q, J = 2.2 Hz), 149.6, 161.5 (d, J = 245.0 Hz), 168.1. MS (ESI) m/z (rel intensity): 329 (MH$^+$, 100). HRMS (ESI-TOF): calcd for $C_{14}H_9F_4N_2OS$ [MH$^+$] 329.0372; found, 329.0379.

4.1.6. Synthesis of 2-amino-6-(trifluoromethoxy)benzo [d]thiazole-4-carbaldehyde (3)

n-Butyllithium (1.10 mL, 1.81 mmol, 1.6 M solution in hexane) was added to a solution of 4-bromo-6-(trifluoromethoxy)benzo [d]thiazol-2-amine (**1**) (195.4 mg, 0.62 mmol) in THF (3 mL) at −78 °C, and the mixture was stirred at that temperature for 30 min. Then, *N,N*-dimethylformamide (0.24 mL, 3.12 mmol) was slowly added at −78 °C and the mixture was stirred at −78 °C for 1 h. The reaction was quenched with a saturated aqueous solution of NH$_4$Cl (10 mL). The mixture was gradually warmed to room temperature and extracted with AcOEt (2 ×10 mL). The combined organic extracts were washed with brine, dried over anhydrous Na_2SO_4 and concentrated *in vacuo*. The crude material was purified by flash column chromatography (silica gel, petroleum ether/AcOEt 1/1) to afford **3** (54.4 mg, 33%) as a yellow solid: IR (ATR): 3378, 3123, 2970, 2924, 1739, 1680, 1527, 1209 cm^{-1}. ^{1}H NMR (300 MHz, DMSO-d_6): δ 7.50 (s, 1 H), 8.13 (s, 1 H), 8.30 (br s, 2 H), 10.55 (s, 1 H). ^{13}C{^{1}H} NMR (75.5 MHz, DMSO-d_6): δ 115.9, 120.7 (q, J = 256.0 Hz), 120.8, 124.1, 135.5, 142.1 (q, J = 2.2 Hz), 154.9, 170.9, 189.9. MS (ESI) m/z (rel intensity): 263 (MH$^+$, 100). HRMS (ESI-TOF): calcd for $C_9H_6F_3N_2O_2S$ [MH$^+$] 263.0102; found, 263.0100.

4.2. Biological assay of Riluzole derivatives

Ca^{2+}-dependent CaM interactions with peptide targets can be monitored using biosensors that consist of CaM fused to a target peptide, flanked by donor and acceptor fluorescent proteins attached to the N- and C-termini. The YC-Nano15 biosensor [60], which follows this design, was used as template. The M13 sequence was replaced by that of the S4S5 linker from the SK4 channel sequence (Fig. 7A), and cloned into a pProEX-HTc plasmid (Invitrogen) that introduces a 6xHis tag at the N-terminus. The construct was transformed in BL21(DE3) cells (Novagen) by electroporation. Cells were grown at 37 °C in 1 L of LB medium containing ampicillin until an A600 between 0.6 and 0.8 was reached. The expression of the fusion proteins was induced using 0.5 mM IPTG O/N at 20 °C. Cells were then harvested by centrifugation at 9000 g for 9 min and re-suspended in 40 mL Buffer A (KCl 120 mM, K-HEPES 50 mM [pH 7.4], NaCl 5 mM, DTT 500 μM, PMSF 1 mM, protease inhibitor EDTA free: Roche, Ref. 04693132001). After lysis by sonication (15 s ON, 15 s OFF, 25 cycles, 7.5 μm), the slurry was centrifuged at 30,000 g for 30 min, the supernatant was filtered (0.20 μm) and transferred to a clean tube. The complex was affinity purified from the supernatant using a His-Trap-talon column and equilibrated with fluorescence buffer (KCl 120 mM, Hepes 50 mM, NaCl 5 mM, EGTA 5 mM). Size-exclusion chromatography was performed using Superdex 200 pg 26/60 column (GE Healthcare, ref. 28–9893) pre-equilibrated with fluorescence buffer. A proper volume of 1 M EGTA was added to obtain a final concentration of 100 mM to the fractions containing the soluble monomeric protein and dialyzed O/N against fluorescence buffer. Subsequently, the soluble fraction was centrifuged at 14,000 g for 10 min to remove any aggregate. Drugs were dissolved in
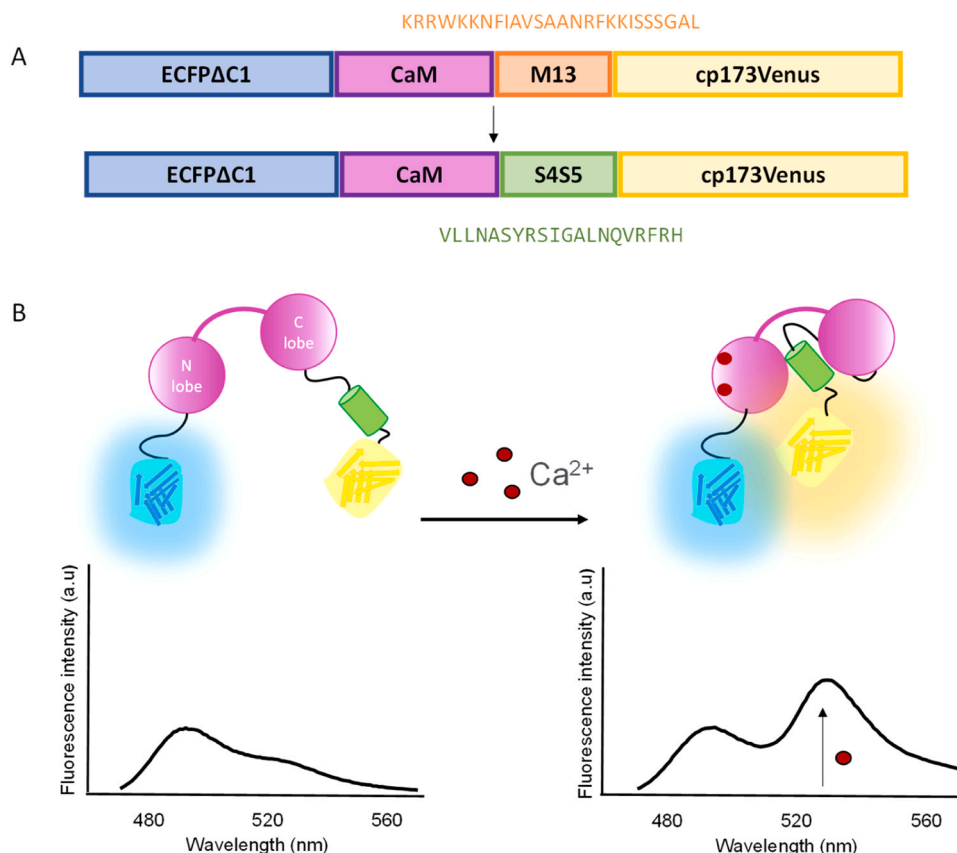
Fig. 7. Schematic representation of the biosensor. (A) Design of the SK4 biosensor (S4S5 target sequence) based on the YC-Nano15 biosensor (M13 target sequence). (B) Presumed conformational changes and subsequent emission response. In the absence of Ca^{2+}, the two fluorophores are far apart in space so there is little energy transfer from the donor to the acceptor. It is expected that the N-lobe will bind the S4S5 target peptide when it becomes Ca^{2+} loaded, moving both fluorophores close together, increasing the energy transfer.

DMSO at 10 mM, and for titration, were added at a final concentration of 100 μM with 1% DMSO. At this concentration, DMSO has negligible effects on the spectra emission of the biosensors used in this study. FRET measurements were performed in the absence and presence of free Ca^{2+} concentration (between 0 and 1600 nM). Fura-2 was used to estimate free Ca^{2+} concentration, following the manufacturer's instructions (Invitrogen).

4.2.1. Förster resonance energy transfer

FRET changes were monitored using a Fluoromax-3 fluorimeter. The emission spectra were collected after excitation at 435 nm (which correspond to wavelength for maximal absorbance for the donor CFP) from 450 to 570 nm. A FRET index was computed as the emission ratio at 475 nm and 525 nm, which corresponds to the wavelength for peak emission for the donor (CFP) and acceptor (YFP) fluorescent proteins, respectively.

4.3. NIFPTML methods

NIFPTML analysis involves four phases: Network Invariants (NI) calculation, Information Fusion (IF) process, Perturbation Theory (PT) variability quantification, and Artificial Intelligence and/or Machine Learning (AI/mL) algorithms training, validation, and use. Firstly, NI-Phase involves system conceptualization, definition of subsystems, calculation of structural descriptors for each subsystem, *etc.* In addition, IF-Phase involves the steps of data gathering, data curation, and data pre-processing. In fact, system conceptualization is conceptual decomposition of the system in different sub-systems easy to study. This involves the definition of the objective function to be fitted by the model,

the outputs variables, and the definition of the input variables used to describe the system. In this case, the system can be theoretically divided into three subsystems: drug information related to assays, protein data referred to assays, and PIN information. In this concept, in data gathering process the examination of databases is carried out. On the other hand, data curation represents the automatic and/or manual verification of all data dealing with missing, contradictory, or duplicated cases. Lastly, as final part of IF-phase, in data pre-processing all the transformations of the original data and calculation of structural parameters useful to codify the structure of each subsystem (drug, proteins, PINs, in this example) are carried out. Besides, PT-Phases refers to the calculation of functions of reference and PTOs used to quantify all the perturbations/variability on the input variables for all subsystems of the query system with respect to conditions or labels for the systems of reference. Finally, mL-Phase involves the training and validation of different mL models [32–35,41,42,61]. In Fig. 8, we illustrate the general workflow used in this Research Article to obtain the NIFPTML model (Supplementary Material Section 3).

4.3.1. Riluzole, CaM and other proteins: case of study

As commented in the previous section, the output of the model f(v_{ij}; v_{yj})$_{calc}$ is a scoring function of the value v_{ij} of biological activity of the ith drug that interacts with a yth protein expressed in one of the regions of the cerebral cortex associated with degenerative neurological diseases in the different combinations of assay conditions, c_j. In this context, it was decided to use this model to predict the output of different compounds, Riluzole and its derivatives **1**, **2a-c**, and **3**, against CaM and its related kinases. The goal was to see if Riluzole and these derivatives bonded with CaM and not with other proteins, since that protein is the one
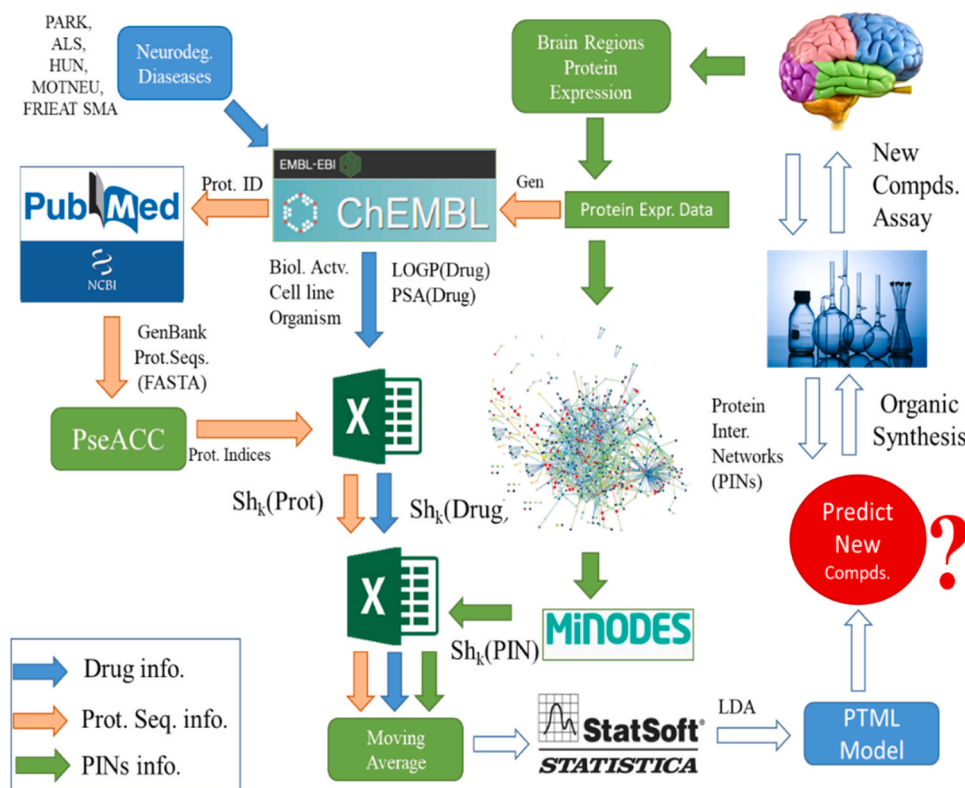
Fig. 8. General workflow used in the development of the predictive model.

related with neurodegenerative diseases as explained in the introduction section. To facilitate these predictions, our data were divided into two distinct databases. The first database contained comprehensive information concerning CaM protein and its related kinases (Supplementary Materials Table S4). On the other hand, the second database compiled data pertaining to other proteins associated with neurodegenerative diseases (Supplementary Materials Table S9). Notably, the data utilized for predicting the molecules' affinities with other proteins were drawn from the same preclinical assays employed in developing our predictive model. Consequently, we needed only to construct the first database from the ground up, given its unique focus on CaM and its associated kinases.

Firstly, the assay's data was obtained from ChEMBL [25] database. Besides, to further prepare the dataset, all assays that were duplicated were deleted taking into account the following boundary conditions (c_j): the tested proteins (c_1) and cell lines (c_3); the brain region where the protein was expressed (c_4); and, lastly, the assay type (c_7) including the organism tested (c_8). Then, the drug information was added to the database. To do so, the data was duplicated 6 times, one for each proposed drug (Supplementary Materials Table S4). The molecular descriptors of the 6 molecules, including LogP and PSA, were searched using the DRAGON [62] software (Supplementary Materials Table S7 and Table S12). DRAGON descriptor is generated by the DRAGON from the particular substructures of given modules. The detail information about the DRAGON descriptors can be obtained at the following website: http://www.talete.mi.it/products/dragon_description.htm. The molecular descriptors were then converted into Shannon's entropy [63] structural information measures for the chemical compounds $Sh_k(Drug_i)$. In addition, the data was also repeated another 6 times to take into account each brain region: Superior Frontal Gyrus, Prefrontal Cortex, Middle Temporal Gyrus, Hippocampus, Entorhinal Cortex, Visual Cortex. For the PIN information, whether the CaM was expressed or not in the 6 brain regions was searched, since that protein was not on the list. In this context, we observed that CaM is expressed in specific regions

of both the human brain, including the Entorhinal Cortex [64], Middle Temporal Gyrus [64] and Hippocampus [65], and the mouse brain, encompassing the Prefrontal Cortex [66], Superior Frontal Gyrus [66], and Visual Cortex [67]. We then made the following determination: if CaM was expressed in the human brain, we assigned a probability of p (PIN_t) equal to 1. Conversely, if CaM was expressed in the mouse brain, we assigned a probability of p(PIN_t) equal to 0.75. Additionally, PTOs [68] were calculated (Supplementary Materials Table S6 and Table S11), and the boundary conditions that were considered to accomplish that were the following ones: c_0= biological activity; c_3= cell name; c_4= brain region; c_6= target organism. Once all the data was gathered, the output of the model was applied, and, thus, the different scoring functions v_{ij} of the drugs (**1–3**) were reached. In Fig. 9, the general process of this part is displayed.

Moreover, to be able to compare with the Riluzole's results, the relative value was calculated for each database. In the first database, where the main protein is CaM, $\Delta f(v_{ij};v_{CaMj})_{calc}$ (%) was calculated by substituting the outputs of the NIFPML model for the derivative f(v_{ij}; $v_{CaMj})_{calc}$ and Riluzole f($v_{Riluzolej}$; $v_{CaMj})_{calc}$ into Eq. 3. Whereas on the second database, where different proteins about neurological disease were collected, $\Delta f(v_{Riluzolej};v_{yj})_{calc}$ (%) was calculated. We calculated it by substituting the outputs of the NIFPTML model for the derivative f(v_{ij}; $v_{yj})_{calc}$ and Riluzole's f($v_{Riluzolej}$;$v_{yj})_{calc}$ *vs.* 134 different proteins in different n^{th} assays into Eq. 4.

$$\Delta f\left(v_{ij};v_{CaMj}\right)_{calc}(\%) = 100 \cdot \frac{\left[f\left(v_{ij};v_{CaMj}\right)_{calc} - f\left(v_{riluzolej};v_{CaMj}\right)_{calc}\right]}{f\left(v_{riluzolej};v_{CaMj}\right)_{calc}} \quad (3)$$

$$\Delta f\left(v_{ij};v_{yj}\right)_{calc}(\%)_{avg} = \frac{100}{n_{assay}} \cdot \sum_{n=1}^{n_{assay}} \frac{\left[f\left(v_{ij};v_{yj}\right)_{calc} - f\left(v_{riluzolej};v_{yj}\right)_{calc}\right]}{f\left(v_{riluzolej};v_{yj}\right)_{calc}}$$

$$= \frac{1}{n_{assay}} \cdot \sum_{n=1}^{n_{assay}} \Delta f\left(v_{ij};v_{yj}\right)\% \quad (4)$$
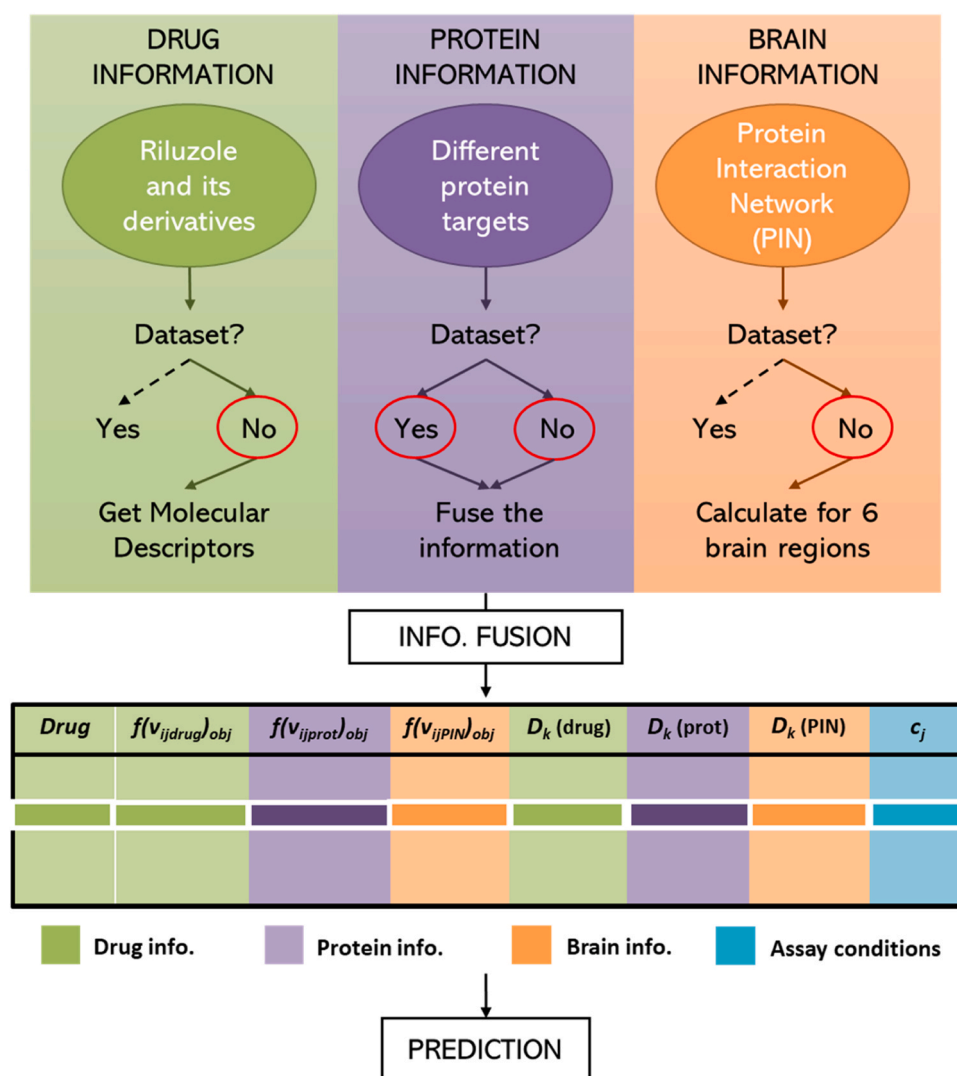
Fig. 9. Workflow of NIFPTML prediction of new Riluzole derivatives.

4.4. Docking Methods

We make use of docking for the purpose of better understanding the mechanism of action of Riluzole, and to assess the validity of the Riluzole derivatives proposed in this work as alternatives to Riluzole. In particular, the study is in the context of SK ion channels, which are a small family of four Ca^{2+} operated channels, and which contain CaM binding sites and are thus of interest in this study.

The crystal structure of Riluzole bound at the interface between CaM's N-lobe and the C-terminus of the SK2 ion channel was resolved by Cho *et al.* [15] (PDB ID 5v02). However, this crystal structure was resolved using only the R396–Q487 fragment of the SK2 channel in complex with CaM. More recently, the full structure of the SK4 ion channel, which is structurally very similar to the SK2 channel, in complex with CaM was resolved through Cryo-EM [69] (PDB ID 6CNN). In this new structure, the CaM N-lobe was instead found to recognize the $S_{45}A$ helix of the SK4 channel, and the interface between them was hypothesized to form the pocket where Riluzole binds.

Therefore, in this work we perform docking and molecular dynamics simulations with Riluzole, to see if binding poses similar to the ones reported [15] can be obtained for this new SK4 binding pocket. Then, the proposed set of Riluzole derivatives are docked at the same CaM-SK4 interface to examine whether stable conformations of these derivatives can be obtained in this binding pocket and thus investigate the viability

of these molecules as alternatives to Riluzole.

Before performing docking calculations, the geometry of the ligands was optimized through force field minimization using the OPLS4 [70] force field. The ligand and protein preprocessing steps were performed with LigPrep [69] and the Protein Preparation Wizard [71], respectively. The grid defining the search space of the docking calculation was defined as a 15 Å x 15 Å x 15 Å cube centered around the known binding site of Riluzole in CaM. Docking was then performed using Glide [72] on SP mode.

To validate the stability of the binding poses predicted by the docking calculations, molecular dynamics simulations were performed starting from the predicted configurations. Specifically, the binding pose metadynamics (BPMD) procedure outlined in [59] was used, in which, starting from the predicted bound structure, multiple short metadynamics simulations are run with the root-mean-square deviation (RMSD) between the ligand heavy atoms and the reference structure as the collective variable. The goal of these simulations is not to fully converge the free energy landscape, but instead to observe the extent to which the biasing force causes the ligand to fluctuate, with the assumption that unstable ligands will tend to fluctuate more than sfscheme ones.

The starting structures for MD simulation were prepared using the CHARMM-GUI [73] input generator. The bound structures were solvated in a box of TIP3P water molecules and neutralized with K^+ and Cl^-

ions. For the parameterization of the protein and the ligands, the FF14SB [74] and GAFF2 [75] force fields were used, respectively. The system was then minimized and equilibrated for 500 ps in the NVT ensemble, using the OpenMM [76] simulation engine. During equilibration, protein and ligand heavy atoms were restrained using a harmonic potential with a force constant of 5 kcal/mol.

The BPMD simulations were run and analyzed using the implementation in [58]. For every ligand, 10 independent simulations were run for 10 ns each, all of them in the NVT ensemble with a Langevin integrator, a timestep of 4 fs, a temperature of 300 K and a friction coefficient of 1 ps^{-1}. For metadynamics, the RMSD collective variable was biased through Gaussian potentials with hill height and width of 0.3 kcal/mol and 0.02 Å, respectively.

CRediT authorship contribution statement

Maider Baltasar-Marchueta: Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Leire Llona:** Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Sara M. Alicante:** Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Iratxe Barbolla:** Methodology, Formal analysis, Investigation; **Markel García Ibarluzea:** Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Rafael Remis:** Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Ane Miren Salomon:** Investigation; **Brenda Fundora:** Methodology, Investigation, Data Curation, Formal analysis; **Ariane Araujo:** Investigation; **Arantza Muguruza-Montero:** Investigation; **Eider Nuñez:** Investigation; **Scarlett Pérez Olea:** Formal analysis, Investigation, Data Curation; **Christian Villanueva:** Formal analysis, Investigation, Data Curation; **Aritz Leonardo:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Resources, Supervision, Project administration, Funding acquisition; **Sonia Arrasate:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Resources, Supervision, Project administration, Funding acquisition; **Nuria Sotomayor:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Resources, Supervision, Project administration, Funding acquisition; **Alvaro Villarroel:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Resources, Supervision, Project administration, Funding acquisition; **Aitor Bergara:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Resources, Supervision, Project administration, Funding acquisition; **Esther Lete:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Resources, Supervision, Project administration, Funding acquisition; **Humberto González Díaz:** Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Resources, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopha.2024.116602.

References

[1] E.R. Chin, Role of Ca^{2+}/calmodulin-dependent kinases in skeletal muscle plasticity, J. Appl. Physiol. 99 (2005) 414–423, https://doi.org/10.1152/japplphysiol.00015.2005.

[2] M. Zhang, C. Abrams, L. Wang, A. Gizzi, L. He, R. Lin, Y. Chen, P.J. Loll, J. M. Pascal, J. Zhang, Structural basis for calmodulin as a dynamic calcium sensor, J. Struct. 20 (2012) 911–923, https://doi.org/10.1016/j.str.2012.03.019.

[3] D. Chin, A.R. Means, Calmodulin: a prototypical calcium sensor, Trends Cell. Biol. 10 (2000) 322–328, https://doi.org/10.1016/S0962-8924(00)01800-6.

[4] Q. Liu, S. Hirono, I. Moriguchi, Quantitative structure-activity relationships for calmodulin inhibitors, Chem. Pharm. Bull. 38 (1990) 2184–2189, https://doi.org/10.1248/cpb.38.2184.

[5] M.S.C. Khoo, J. Li, M.V. Singh, Y. Yang, P. Kannankeril, Y. Wu, C.E. Grueter, X. Guan, C.V. Oddis, R. Zhang, L. Mendes, G. Ni, E. C. Madu, J. Yang, M. Bass, R. J. Gomez, B.E. Wadzinski, E.N. Olson, R.J. Colbran, M.E. Anderson, Death, Cardiac dysfunction, and arrhythmias are increased by calmodulin kinase II in calcineurin cardiomyopathy, Circ 114 (2006) 1352–1359, https://doi.org/10.1161/CIRCULATIONAHA.106.644583.

[6] M.-C. Kotta, L. Sala, A. Ghidoni, B. Badone, C. Ronchi, G. Parati, A. Zaza, L. Crotti, Calmodulinopathy: a novel, life-threatening clinical entity affecting the young, Front. Cardiovasc. Med. 5 (2018) 175, https://doi.org/10.3389/fcvm.2018.00175.

[7] D.H. O'Day, Calmodulin binding proteins and Alzheimer's disease: biomarkers, regulatory enzymes and receptors that are regulated by calmodulin, Int. J. Mol. Sci. 21 (2020) 7344, https://doi.org/10.3390/ijms21197344.

[8] S. Mustaly-Kalimi, W. Gallegos, R.A. Marr, A. Gilman-Sachs, D.A. Peterson, I. Sekler, G.E. Stutzmann, Protein mishandling and impaired lysosomal proteolysis generated through calcium dysregulation in Alzheimer's disease, Proc. Natl. Acad. Sci. U. S. A. 119 (2022) e2211999119, https://doi.org/10.1073/pnas.2211999119.

[9] B. Picconi, Abnormal Ca^{2+} calmodulin-dependent protein kinase II function mediates synaptic and motor deficits in experimental Parkinsonism, J. Neurosci. 24 (2004) 5283–5291, https://doi.org/10.1523/JNEUROSCI.1224-04.2004.

[10] S.S. Leal, C.M. Gomes, Calcium dysregulation links ALS defective proteins and motor neuron selective vulnerability, Front. Cell. Neurosci. 9 (2015), https://doi.org/10.3389/fncel.2015.00225.

[11] N.J. Logsdon, J. Kang, J.A. Togo, E.P. Christian, J. Aiyar, A Novel Gene, *hKCa4*, encodes the calcium-activated potassium channel in Human T Lymphocytes, J. Biol. Chem. 272 (1997) 32723–32726, https://doi.org/10.1074/jbc.272.52.32723.

[12] D.H. Vandorpe, B.E. Shmukler, L. Jiang, B. Lim, J. Maylie, J.P. Adelman, L. De Franceschi, M.D. Cappellini, C. Brugnara, S.L. Alper, cDNA cloning and functional characterization of the mouse Ca^{2+}-gated K^+ Channel, mIK1, J. Biol. Chem. 273 (1998) 21542–21553, https://doi.org/10.1074/jbc.273.34.21542.

[13] J.P. Adelman, J. Maylie, P. Sah, Small-conductance Ca^{2+}-activated K^+ channels: form and function, Annu. Rev. Physiol. 74 (2012) 245–269, https://doi.org/10.1146/annurev-physiol-020911-153336.

[14] C.-H. Lee, R. MacKinnon, Activation mechanism of a human SK-calmodulin channel complex elucidated by cryo-EM structures, Sci 360 (2018) 508–513, https://doi.org/10.1126/science.aas9466.

[15] L.T.-Y. Cho, A.J. Alexandrou, R. Torella, J. Knafels, J. Hobbs, T. Taylor, A. Loucif, A. Konopacka, S. Bell, E.B. Stevens, J. Pandit, R. Horst, J.M. Withka, D.C. Pryde, S. Liu, G.T. Young, An intracellular allosteric modulator binding pocket in SK2 Ion Channels is shared by multiple chemotypes, e3, Struct 26 (2018) 533–544, https://doi.org/10.1016/j.str.2018.02.017.

[16] LiverTox, RiluzoleLiverTox, Riluzole. LiverTox: clinical and research information on drug-induced liver injury., Bethesda (MD)2012.

[17] M.K. Jaiswal, Riluzole and edaravone: a tale of two amyotrophic lateral sclerosis drugs, Med. Res. Rev. 39 (2019) 733–748, https://doi.org/10.1002/med.21528.

[18] P. Müller, A. Draguhn, A.V. Egorov, Persistent sodium current modulates axonal excitability in CA1 pyramidal neurons, J. Neurochem. 146 (2018) 446–458, https://doi.org/10.1111/jnc.14479.

[19] E.J. Calabrese, V. Calabrese, J. Giordano, Demonstrated hormetic mechanisms putatively subserve riluzole-induced effects in neuroprotection against amyotrophic lateral sclerosis (ALS): implications for research and clinical practice, Ageing Res. Rev. 67 (2021) 101273, https://doi.org/10.1016/j.arr.2021.101273.

[20] D. Gerovska, H. Irizar, D. Otaegi, I. Ferrer, A. López De Munain, M.J. Araúzo-Bravo, Genealogy of the neurodegenerative diseases based on a meta-analysis of age-stratified incidence data, Sci. Rep. 10 (2020) 18923, https://doi.org/10.1038/s41598-020-75014-8.

[21] E. Trushina, C.T. McMurray, Oxidative stress and mitochondrial dysfunction in neurodegenerative diseases, Neurosci 145 (2007) 1233–1248, https://doi.org/10.1016/j.neuroscience.2006.10.056.

[22] R.F. Allegri, A. Guekht, Cerebrolysin improves symptoms and delays progression in patients with Alzheimer's disease and vascular dementia, Drugs Today (Barc.) 48 (2012) 25–41, https://doi.org/10.1358/dot.2012.48(suppl.a).1739721.

[23] D.K. Simon, C.M. Tanner, P. Brundin, Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology, Clin. Geriatr. Med. 36 (2020) 1–12, https://doi.org/10.1016/j.cger.2019.08.002.

[24] H.R. Morris, A.J. Waite, N.M. Williams, J.W. Neal, D.J. Blake, Recent advances in the genetics of the ALS-FTLD complex, Curr. Neurol. Neurosci. Rep. 12 (2012) 243–250, https://doi.org/10.1007/s11910-012-0268-5.

[25] A. Gaulton, A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magariños, J.P. Overington, G. Papadatos, I. Smit, A.R. Leach, The ChEMBL database in 2017, Nucleic Acids Res. 45 (2017) D945–D954, https://doi.org/10.1093/nar/gkw1074.

[26] F.K. Brown, Chemoinformatics: what is it and how does it impact drug discovery, Annu. Rep. Med. Chem. 33 (1998) 365–384, https://doi.org/10.1016/S0065-7743(08)61100-8.

[27] J.L. Faulon, A. Bender. Handbook of Chemoinformatics Algorithms, first ed., , CRC press, 2010.

[28] D.W. Howells, E.S. Sena, V. O'Collins, M.R. Macleod, Improving the efficiency of the development of drugs for stroke, Int. J. Stroke 7 (2012) 371–377, https://doi.org/10.1111/j.1747-4949.2012.00805.x.

[29] H. Chen, C.Z. Wang, C. Ding, C. Wild, B. Copits, G.T. Swanson, K.M. Johnson, J. Zhou, A combined bioinformatics and chemoinformatics approach for developing asymmetric bivalent AMPA receptor positive allosteric modulators as neuroprotective agents, Chem. Med. Chem. 8 (2013) 226–230, https://doi.org/10.1002/cmdc.201200554.

[30] M. González-Medina, J.J. Naveja, N. Sánchez-Cruz, J.L. Medina-Franco, Open chemoinformatic resources to explore the structure, properties and chemical space of molecules, RSC Adv. 7 (2017) 54153–54163, https://doi.org/10.1039/C7RA11831G.

[31] L.L. Ferreira, A.D. Andricopulo, From chemoinformatics to deep learning: an open road to drug discovery, Future Med. Chem. 11 (2019) 371–374, https://doi.org/10.4155/fmc-2018-0449.

[32] M.P. González, A.M. Helguera, H.G. Díaz, A TOPS-MODE approach to predict permeability coefficients, Polymer 45 (2004) 2073–2079, https://doi.org/10.1016/j.polymer.2003.12.014.

[33] D. Nocedo-Mena, C. Cornelio, M. del R. Camacho-Corona, E. Garza-González, N. Waksman de Torres, S. Arrasate, N. Sotomayor, E. Lete, H. González-Díaz, Modeling antibacterial activity with machine learning and fusion of chemical structure information with microorganism metabolic networks, J. Chem. Inf. Model. 59 (2019) 1109–1120, https://doi.org/10.1021/acs.jcim.9b00034.

[34] H. Bediaga, S. Arrasate, H. González-Díaz, PTML Combinatorial Model of ChEMBL Compounds Assays for Multiple Types of Cancer, ACS Comb. Sci. 20 (2018) 621–632, https://doi.org/10.1021/acscombsci.8b00090.

[35] S.G. Martínez-Arzate, E. Tenorio-Borroto, A. Barbabosa Pliego, H.M. Díaz-Albiter, J.C. Vázquez-Chagoyán, H. González-Díaz, PTML model for proteome mining of b-cell epitopes and theoretical–experimental study of Bm86 protein sequences from Colima, Mexico, J. Proteome Res. 16 (2017) 4093–4103, https://doi.org/10.1021/acs.jproteome.7b00477.

[36] P. Jimonet, F. Audiau, M. Barreau, J.-C. Blanchard, A. Boireau, Y. Bour, M.-A. Coléno, A. Doble, G. Doerflinger, C. Do Huu, M.-H. Donat, J.M. Duchesne, P. Ganil, C. Guérémy, E. Honoré, B. Just, R. Kerphirique, S. Gontier, P. Hubert, P. M. Laduron, J. Le Blevec, M. Meunier, J.-M. Miquet, C. Nemecek, M. Pasquet, O. Piot, J. Pratt, J. Rataud, M. Reibaud, J.-M. Stutzmann, S. Mignani, Riluzole Series. Synthesis and *in vivo* "antiglutamate" activity of 6-Substituted-2-benzo-thiazolamines and 3-Substituted-2-imino-benzothiazolines, J. Med. Chem. 42 (1999) 2828–2843, https://doi.org/10.1021/jm980202u.

[37] M. Anzini, A. Chelini, A. Mancini, A. Cappelli, M. Frosini, L. Ricci, M. Valoti, J. Magistretti, L. Castelli, A. Giordani, F. Makovec, S. Vomero, Synthesis and Biological Evaluation of Amidine, Guanidine, and Thiourea Derivatives of 2-Amino-(6-Trifluoromethoxy)Benzothiazole as Neuroprotective Agents Potentially Useful in Brain Diseases. J. Med. Chem. 53 (2010) 734–744, https://doi.org/10.1021/jm901375r.

[38] A. Farinato, M.M. Cavalluzzi, C. Altamura, C. Campanale, P. Laghetti, I. Saltarella, P. Delre, A. Barbault, N. Tarantino, G. Milani, N.P. Rotondo, L.C. Mannelli, C. Ghelardini, S. Pierno, G.F. Mangiatordi, G. Lentini, J.-F. Desaphy, Development of Riluzole Analogs with Improved Use-Dependent Inhibition of Skeletal Muscle Sodium Channels, ACS Med. Chem. Lett. 14 (2023) 999–1008, https://doi.org/10.1021/acsmedchemlett.3c00224.

[39] G.M. Duginome, T. Ohmura, Arylboronic acid cross coupling reactions, Science of Synthesis. Cross Coupling and Heck-Type Reactions 1, Georg Thieme Verlag, Thieme, Stuttgart, 2013, https://doi.org/10.1055/sos-SD-207-00057.

[40] T. Hill, P. Lewicki, Statistics: methods and applications: a comprehensive reference for science, industry, and data mining, StatSoft, Inc, 2006.

[41] L. Simón-Vidal, O. García-Calvo, U. Oteo, S. Arrasate, E. Lete, N. Sotomayor, H. González-Díaz, Perturbation-Theory and Machine Learning (PTML) model for high-throughput screening of Parham reactions: experimental and theoretical studies, J. Chem. Inf. Model. 58 (2018) 1384–1396, https://doi.org/10.1021/acs.jcim.8b00286.

[42] V. Blay, T. Yokoi, H. González-Díaz, Perturbation theory–machine learning study of zeolite materials desilication, J. Chem. Inf. Model. 58 (2018) 2414–2419, https://doi.org/10.1021/acs.jcim.8b00383.

[43] S. Xiang, F. Nie, C. Zhang, Learning a Mahalanobis distance metric for data clustering and classification, Pattern Recognit. 41 (2008) 3600–3612, https://doi.org/10.1016/j.patcog.2008.05.018.

[44] W. Dai, H.-Y. Chen, C.Y.-C. Chen, A network pharmacology-based approach to investigate the novel TCM formula against Huntington's disease and validated by support vector machine model, Evid. Based Complement. Altern. Med. 2018 (2018) 1–29, https://doi.org/10.1155/2018/6020197.

[45] P. Ambure, J. Bhat, T. Puzyn, K. Roy, Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: an *in-silico* approach, J. Biomol. Struct. Dyn. 37 (2019) 1282–1306, https://doi.org/10.1080/07391102.2018.1456975.

[46] J. Fang, Y. Li, R. Liu, X. Pang, C. Li, R. Yang, Y. He, W. Lian, A.-L. Liu, G.-H. Du, Discovery of multitarget-directed ligands against Alzheimer's disease through systematic prediction of chemical–protein interactions, J. Chem. Inf. Model. 55 (2015) 149–164, https://doi.org/10.1021/ci500574n.

[47] V. Kumar, P.K. Ojha, A. Saha, K. Roy, Cheminformatic modelling of β-amyloid aggregation inhibitory activity against Alzheimer's disease, Comput. Biol. Med. 118 (2020) 103658, https://doi.org/10.1016/j.compbiomed.2020.103658.

[48] L. Ivanova, M. Karelson, D. Dobchev, Identification of natural compounds against neurodegenerative diseases using *in silico* techniques, Mol 23 (2018) 1847, https://doi.org/10.3390/molecules23081847.

[49] K. Nikolic, L. Mavridis, O.M. Bautista-Aguilera, J. Marco-Contelles, H. Stark, M. do Carmo Carreiras, I. Rossi, P. Massarelli, D. Agbaba, R.R. Ramsay, J.B.O. Mitchell, Predicting targets of compounds against neurological diseases using cheminformatic methodology, J. Comput. Aided Mol. Des. 29 (2015) 183–198, https://doi.org/10.1007/s10822-014-9816-1.

[50] L. Ivanova, M. Karelson, D.A. Dobchev, Multitarget approach to drug candidates against Alzheimer's disease related to AChE, SERT, BACE1 and GSK3β protein targets, Mol 25 (2020) 1846, https://doi.org/10.3390/molecules25081846.

[51] T. Hidaka, K. Imamura, T. Hioki, T. Takagi, Y. Giga, M.-H. Giga, Y. Nishimura, Y. Kawahara, S. Hayashi, T. Niki, M. Fushimi, H. Inoue, Prediction of compound bioactivities using heat-diffusion equation, Patterns 1 (2020) 100140, https://doi.org/10.1016/j.patter.2020.100140.

[52] I. García, Y. Fall, G. Gómez, H. González-Díaz, First computational chemistry multi-target model for anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors *in vitro*, *in vivo*, and in different cellular lines, Mol. Divers. 15 (2011) 561–567, https://doi.org/10.1007/s11030-010-9280-3.

[53] E. Briand, R. Thomsen, K. Linnet, H.B. Rasmussen, S. Brunak, O. Taboureau, Combined ensemble docking and machine learning in identification of therapeutic agents with potential inhibitory effect on human CES1, Mol 24 (2019) 2747, https://doi.org/10.3390/molecules24152747.

[54] A. Speck-Planche, V. V. Kleandrova, QSAR and molecular docking techniques for the discovery of potent monoamine oxidase B inhibitors: computer-aided generation of new rasagiline bioisosteres, Curr. Top. Med. Chem. 12 (2012) 1734–1747, https://doi.org/10.2174/156802612803989282.

[55] D. Wang, Z. Li, Y. Liu, M. Chen, N. Chen, Z. Zuo, D.-X. Kong, Identification of novel monoamine oxidase selective inhibitors employing a hierarchical ligand-based virtual screening strategy, Future Med. Chem. 11 (2019) 801–816, https://doi.org/10.4155/fmc-2018-0596.

[56] M. Shahid, M. Shahzad Cheema, A. Klenner, E. Younesi, M. Hofmann-Apitius, SVM based descriptor selection and classification of neurodegenerative disease drugs for pharmacological modeling, Mol. Inf. 32 (2013) 241–249, https://doi.org/10.1002/minf.201200116.

[57] D. Antanasijević, J. Antanasijević, N. Trišović, G. Ušćumlić, V. Pocajt, From classification to regression multitasking QSAR modeling using a novel modular neural network: simultaneous prediction of anticonvulsant activity and neurotoxicity of succinimides, Mol. Pharm. 14 (2017) 4476–4484, https://doi.org/10.1021/acs.molpharmaceut.7b00582.

[58] D. Lukauskis, M.L. Samways, S. Aureli, B.P. Cossins, R.D. Taylor, F.L. Gervasio, Open binding pose metadynamics: an effective approach for the ranking of protein–ligand binding poses, J. Chem. Inf. Model. 62 (2022) 6209–6216, https://doi.org/10.1021/acs.jcim.2c01142.

[59] L. Fusani, D.S. Palmer, D.O. Somers, I.D. Wall, Exploring ligand stability in protein crystal structures using binding pose metadynamics, J. Chem. Inf. Model. 60 (2020) 1528–1539, https://doi.org/10.1021/acs.jcim.9b00843.

[60] K. Horikawa, Y. Yamada, T. Matsuda, K. Kobayashi, M. Hashimoto, T. Matsu-ura, A. Miyawaki, T. Michikawa, K. Mikoshiba, T. Nagai, Spontaneous network activity visualized by ultrasensitive Ca^{2+} indicators, yellow Cameleon-Nano, Nat. Methods 7 (2010) 729–732, https://doi.org/10.1038/nmeth.1488.

[61] J. Ferreira da Costa, D. Silva, O. Caamaño, J.M. Brea, M.I. Loza, C.R. Munteanu, A. Pazos, X. García-Mera, H. González-Díaz, Perturbation theory/machine learning model of ChEMBL data for dopamine targets: docking, synthesis, and assay of new

L-Prolyl-L-leucyl-glycinamide peptidomimetics, ACS Chem. Neurosci. 9 (2018) 2572–2587, https://doi.org/10.1021/acschemneuro.8b00083.

[62] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, DRAGON Professional version, v5.3 (Version 5.3), J.P.S. Farah – Instituto de Química/USP, 2005. ⟨http://www.talete.mi.it/products/dragon_description.htm⟩.

[63] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[64] X. Fang, W. Tang, F. Yang, W. Lu, J. Cai, J. Ni, J. Zhang, W. Tang, T. Li, D.-F. Zhang, C. Zhang, A comprehensive analysis of the CaMK2A gene and susceptibility to Alzheimer's Disease in the Han Chinese population, Front. Aging Neurosci. 11 (2019) 84, https://doi.org/10.3389/fnagi.2019.00084.

[65] V. Cimini, S. Van Noorden, C. Terlizzi, G.G. Altobelli, Calcium/calmodulin-dependent kinases in the hypothalamus, pituitary, and pineal gland: an overview, Int. J. Endocrinol. 2022 (2022) 1–9, https://doi.org/10.1155/2022/1103346.

[66] J. Cao, X. Liu, J.-X. Liu, S. Zhao, Y.-X. Guo, G.-Y. Wang, X.-L. Wang, Inhibition of glutamatergic neurons in layer II/III of the medial prefrontal cortex alleviates paclitaxel-induced neuropathic pain and anxiety, Eur. J. Pharmacol. 936 (2022) 175351, https://doi.org/10.1016/j.ejphar.2022.175351.

[67] A. Lilja, G. Didio, J. Hong, W.D. Heo, E. Castrén, J. Umemori, Optical activation of TrkB (E281A) in excitatory and inhibitory neurons of the mouse visual cortex, Int. J. Mol. Sci. 23 (2022) 10249, https://doi.org/10.3390/ijms231810249.

[68] H. González-Díaz, S. Arrasate, A. Gomez-SanJuan, N. Sotomayor, E. Lete, L. Besada-Porto, J.M. Ruso, General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry, Curr. Top. Med. Chem. 13 (2013) 1713–1741, https://doi.org/10.2174/1568026611313140011.

[69] Schrödinger, Ligprep, v3.0 (Version 3.0), (2014). ⟨https://newsite.schrodinger.com/platform/products/ligprep/⟩.

[70] C. Lu, C. Wu, D. Ghoreishi, W. Chen, L. Wang, W. Damm, G.A. Ross, M.K. Dahlgren, E. Russell, C.D. Von Bargen, R. Abel, R.A. Friesner, E.D. Harder, OPLS4: improving force field accuracy on challenging regimes of chemical space, J. Chem. Theory Comput. 17 (2021) 4291–4300, https://doi.org/10.1021/acs.jctc.1c00302.

[71] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, J. Comput. Aided Mol. Des. 27 (2013) 221–234, https://doi.org/10.1007/s10822-013-9644-8.

[72] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749, https://doi.org/10.1021/jm0306430.

[73] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, J. Comput. Chem. 29 (2008) 1859–1865, https://doi.org/10.1002/jcc.20945.

[74] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, Protein.: Str. Func. Bioinform. 78 (2010) 1950–1958, https://doi.org/10.1002/prot.22711.

[75] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174, https://doi.org/10.1002/jcc.20035.

[76] P. Eastman, J. Swails, J.D. Chodera, R.T. McGibbon, Y. Zhao, K.A. Beauchamp, L.-P. Wang, A.C. Simmonett, M.P. Harrigan, C.D. Stern, R.P. Wiewiora, B.R. Brooks, V.S. Pande, OpenMM 7: rapid development of high performance algorithms for molecular dynamics, PLoS Comput. Biol. 13 (2017) e1005659, https://doi.org/10.1371/journal.pcbi.1005659.