Discovery of Potential GSK-3β Allosteric Modulators for Alzheimer's Disease

Guilherme Martins Silva[a,b]*, Vinicius Medeiros Alves[b], Suzane Quintana Gomes[a], Joshua Earl Hochuli[b], Eugene N. Muratov[b,c], Alexander Tropsha[b], Carlos Henrique Tomich de Paula da Silva[a,d]

[a] Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto, SP, Brazil; [b] Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, 27599, Chapel Hill, NC, USA; [c] Department of Pharmaceutical Sciences, Federal University of Paraiba, Joao Pessoa, PB, 58059, Brazil [d] Computational Laboratory of Pharmaceutical Chemistry, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, 14040-903, Ribeirão Preto, SP, Brazil.

* Corresponding author:

Guilherme Martins Silva
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP)
Universidade de São Paulo (USP)
Av. do Café, s/n
14040-901 – Ribeirão Preto, SP, Brazil
Tel.: +55 16 3315 0664
e-mail: silvagm@usp.br

Abstract

Glycogen Synthase Kinase-3 beta (GSK-3β) is a validated target-enzyme associated with Alzheimer's Disease (AD). Usage of allosteric inhibitors of this enzyme represents a valid and promising therapeutic strategy due to their selective and subtle modulation, with a low probability of producing side effects. Nonetheless, only a few GSK-3β allosteric modulators with limited binding affinity have been uncovered so far and published in the public domain. Previous Virtual Screening (VS) studies have not considered such mechanism of action and did not achieve chemical diversity. Therefore, we applied two orthogonal VS workflows by means of shape-based similarity, QSAR, docking, and ADMET filters to select new and diverse GSK-3β allosteric inhibitors. Obtained hits have shown enhanced structural diversity and preliminary results as GSK-3β allosteric inhibitors according to *in vitro* assays. Furthermore, their GSK-3β allosteric inhibition were analyzed by blind docking and pocket coverage studies. These hits can be employed as template molecules for the discovery of more potent inhibitors, with the aim to expand the chemical space of GSK-3β allosteric modulators as promising agents in AD.

Keywords: Alzheimer's disease, GSK-3β allosteric modulators, virtual screening, QSAR, machine learning, shape similarity, docking, ADMET.

1. Introduction

Glycogen Synthase Kinase 3-beta (GSK-3β) is a serine/threonine kinase involved in the phosphorylation of a variety of substrates and, thus, several processes such as insulin signaling, Wnt signaling, neurotransmitter signaling, and microtubule dynamics, among others [1,2]. Abnormal regulation of this target is associated with multiple neurodegenerative diseases, e.g. Alzheimer's disease (AD) [3–5]. Its expression tends to increase with age, leading to hyperphosphorylation of tau proteins, increased production of β-amyloid, and memory impairment [6,7]. Considering this, several GSK-3β inhibitors have been pursued and studied in

recent decades within AD therapeutics. Furthermore, compounds that can perform allosteric inhibition of GSK-3β can have potential advantages like higher selectivity, mild to high potency, and overall minimizing chances of side effects [8,9]. Nevertheless, few chemical classes (or structural scaffolds) of GSK-3β allosteric inhibitors have been uncovered so far.

Computer-Aided Drug Design (CADD) strategies, such as Virtual Screening (VS), have shown potential to reduce time-cost within an usual drug development pipeline [10]. However, achieving high hit rates and chemical diversity is challenging [11,12]. Expanding the chemical diversity of specific inhibitors of a given biological target is particularly difficult - especially when this is a newly validated target. Usually, larger chemical/biology libraries increase the hit rate in VS campaigns [13,14]. However, many studies have succeeded in identifying hits with small libraries [15,16]. This analysis regarding the balance between the database's size and VS goal (diversity/potency/hit rate/cost-benefit) might become subjective, mainly when choosing which methodological approaches are intended to be used in prospective campaigns. In this context, we propose that application of mixed approaches such as structure- and ligand-based (SB and LB) methodologies can be critical for success in VS. We believe that the attempt to integrate both of them within a VS workflow seems to be a reasonable way to achieve diversity of hits.

Recently, a study employed docking and molecular dynamics to screen ~ 617k compounds against GSK-3β [17]. In another study using machine learning, ruboxistaurin was identified as a GSK-3β inhibitor with IC_{50} = 97.3 nM [4]. Pyrazolo [1,5-a]pyrimidin-7-amine derivatives were identified using pharmacophore models, docking, and molecular dynamics from ~ 530k compound's database [18]. Nevertheless, none of these studies identified novel scaffolds nor targeted allosteric binding sites – other than the orthosteric/active one.

Continuing our previous work [19], we propose applying SB and LB approaches to discover new and diverse GSK-3β allosteric inhibitors. We developed two VS workflows: 1) 3D shape-based similarity *plus* docking, and 2) QSAR modeling using machine learning *plus*

docking. We used several databases with ~ 40 million compounds. Moreover, ADMET predictions were employed to select hits with similar properties as Central Nervous System (CNS) agents, avoiding toxicity issues. Obtained hits show enhanced structural diversity and interesting preliminary results as GSK-3β allosteric inhibitors, according to their in vitro assays. In addition, their putative mechanism of GSK-3β allosteric inhibition was analyzed by blind docking and pocket coverage studies. These hits may be employed as template molecules in future hit-to-lead studies to expand the known chemical space of GSK-3β allosteric modulators as promising agents to treat AD.

2. Materials and Methods

2.1. Collection of known GSK-3β allosteric inhibitors

We collected a total of 88 compounds with known experimental allosteric modulation of GSK-3β [9,20–22], as shown in **Table S1**. This collection was initially presented in our previous publication [19]. Worth noting that, out of these 88 compounds, 40 were retrieved from the same publication [9], corresponding to analogs of **compound 1** - VP07, *N*-dodecanoyl-1-ethyl-4-hydroxy-2-oxo-1,2-dihydroquinoline-3-carbohydrazide, with IC_{50} = 2.8 μM (mostly used as reference in our work). In this way, there's a limited chemical diversity among this 88-compound collection. Finally, we emphasize that there is no experimental crystal structure (e.g. PDB complex) of any allosteric inhibitor and GSK-3β reported so far.

2.2. Virtual screening workflows

In this work, we employed two orthogonal workflows for virtual screening, using distinguished sources of databases and varying the emphasis of methodologies employed to screen compounds (**Figure 1**).

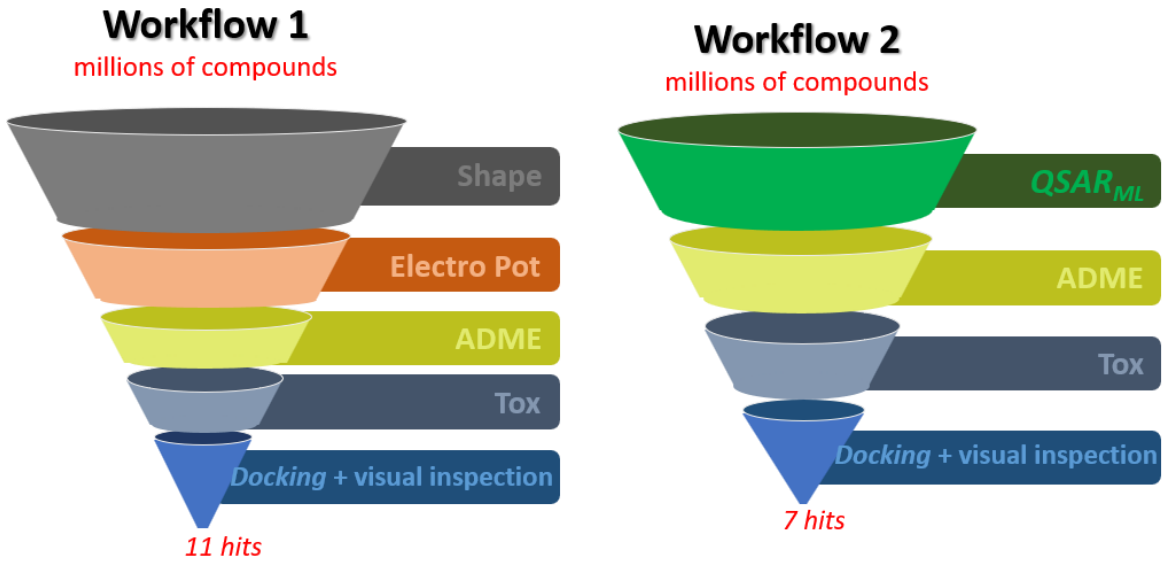

Figure 1. Virtual screening workflows 1 and 2 were applied to achieve diverse hits as GSK-3β allosteric modulators.

2.2.1. Workflow 1: Shape and Structure-based Screening

This workflow was mentioned in our previous publication [19], but no experimental validation was described there. This workflow was carried out with emphasis on 3D shape similarity by means of the ROCS software [23,24]. Seventeen databases were screened: ChemBridge CNS [25], ChemBridge CORE Library Stock [25], ChemBridge Diverset CL (DIVERSet^{TM}-CL subcollection) [25], ChemBridge Diverset EXP (DIVERSet^{TM}-EXP subcollection) [25], ChemBridge EXPRESS-Pick Collection Stock [25], eMolecules [26], IBS Bioactive Compounds [27], IBS Natural Compounds [27], IBS Synthetic Compounds [27], Kishida [26], MayBridge (screening collection) [28], Molport Natural [29], Molport Screening [29], NPASS [30], Princeton [31], ZINC CNS [32], ZINC Druglike [33], e ZINC Naturals (natural products collection) [32], reaching more than 39 million molecules.

These databases were prepared individually with FILTER [34], using default parameters and the following additional filter settings: maximum of 2 chiral centers, 5 ring systems, 20 atoms by ring system, 16 rotational bonds, and 55 rigid bonds. Usage of FILTER allows elimination of compounds with undesired drug-like properties, e.g. molecules bearing too many rotatable bonds (high flexibility) and/or rings (exceptionally flexible ones) [35]. In sequence,

OMEGA [34,36] was employed for conformer generation. We generated up to 300 conformers per molecule using the default parameters, except for a strain energy window of up to 9.0 kcal/mol, and Root Mean Square Deviation (RMSD) for conformer identity adjusted to 0.6 Å (thus optimized by our research group [37]). Worth noting that running OMEGA to generate conformers is an essential step during the preparation of databases, since both ROCS and EON software require several conformers of each molecule to perform overlap-similarity search.

In ROCS [23,24], we used our predicted bioactive pose of Compound **1** as a query and applied it to the seventeen prepared databases. It is worth mentioning that such predicted bioactive pose as well as the validation of its use in shape similarity has been previously detailed in our earlier paper [19]. Hence, we virtually screened the databases and analyzed the 5,000 top-ranked molecules according to corresponding ROCS TC (Tanimoto Combo considering shape+color features) indices. In sequence, another screening was carried out considering 3D electrostatic potential similarity using EON [38] – and the same pose of **1** as a query. We selected the 1,000 top-ranked molecules (from each database) considering EON TC indices.

Lastly, compounds were submitted to pharmacokinetic and toxicological (ADMET) predictions using the QikProp and DEREK software, as described in *section 2.3*.

The final methodology applied to this workflow consisted of docking approaches. We applied our previously developed [19] and reliable docking protocol using the Glide software to evaluate the survival compounds' ability to establish consistent intermolecular interactions within the GSK-3β allosteric binding site. Docking simulations are briefly described in *section 2.4*. At last, a visual inspection of remaining compounds was narrowly conducted to select the most promising and diverse compounds.

2.2.2. Workflow 2: QSAR Modeling and Structure-based Screening

In this second screening workflow, emphasis was given to the development and application of QSAR models using machine learning. To develop these models, we used the dataset of 88 GSK-3β allosteric modulators described in *section 2.1*. We considered an IC_{50}

threshold value of 20 µM to sort them into active/inactive, thus obtaining 37 active compounds and 51 inactive ones. **Table S1** shows this dataset of 88 GSK-3β allosteric modulators and their respective IC_{50} values.

As a standard protocol [39,40], chemical data curation was executed by standardizing chemical structures and corresponding biological information using KNIME 4.6.0 [41,42] integrated with ChemAxon Standardizer v.20.8.0 (ChemAxon, Budapest, Hungary, http://www.chemaxon.com). In addition, specific chemotypes, such as aromatic rings and nitro groups, were normalized. Counterions, inorganics, metals, organometallic compounds, and mixtures were removed. Duplicates were identified, analyzed, and removed.

QSAR models were developed employing two sets of molecular descriptors: Morgan fingerprints [43–45] and RDKit whole-molecule descriptors (http://www.rdkit.org) [46,47]. The open-source Morgan fingerprints with 2048 bits and an atom radius of 3 were calculated using the RDKit node implemented in KNIME and also in Python 3.6 (https://www.python.org). RDKit whole-molecule descriptors calculated in KNIME and Python correspond to 117 properties, such as SlogP, SMR, LabuteASA, TPSA, and others. Worth noting that these RDKit properties were selected using low variance and correlation filters; moreover, each descriptor's importance was calculated after the models' generation.

QSAR models were built and rigorously validated, following the best practices for model development and validation [39,48,49]. These models were generated using the Random Forest (RF) algorithm [50] implemented in scikit-learn (http://scikit-learn.org) [51] in Python, and, in KNIME, using the nodes 'Tree Ensemble Learner' and 'Tree Ensemble Predictor' – where the following settings were used: Gini index as a split criterion and 750 decision trees to build each forest.

A 5-fold external cross-validation procedure was followed, randomly dividing the complete set of compounds with known experimental activity into five subsets of equal size. One of these subsets (20 % of all compounds) is set aside as the external validation set, while the

remaining four groups form the modeling set (80 % of all compounds). This procedure is repeated five times, allowing each of the five subsets to be used as an external validation set. Models are built using the modeling set only, and it is essential to emphasize that the model never sees the external validation set.

In addition, 20 rounds of Y-randomization were performed for each data set to ensure that the model performance was not due to chance correlations. The model's domain of applicability (DA) was estimated using the z-cutoff method [52,53].

More than 10 million compounds were screened from two databases: Molport Screening [29] and Enamine Screening Collection [54], prepared with the same data curation protocol described above.

Afterwards, we applied the QSAR model with the best statistical characteristics to filter only compounds predicted as active and within the DA. Such statistical characteristics consisted of correct classification rate (CCR), sensitivity, specificity, positive and negative predictive values (PPV and NPV). Furthermore, the resulting compounds were sorted in descending order according to their confidence/score predictions (i.e., votes of majority of local trees from RF), and the top 40,000 were retrieved.

In sequence, these were submitted to ADMET predictions using QikProp and DEREK and filtered considering the same criteria mentioned in *section 2.3*.

Finally, remaining compounds were filtered using the same docking protocol using Glide (as described in *section 2.4),* followed by a visual inspection of corresponding intermolecular interactions within the allosteric site of GSK-3β.

2.3. ADMET predictions

In order to predict the physicochemical/pharmacokinetic and toxicological (ADMET) properties of compounds, we used QikProp and DEREK, respectively. QikProp criteria to filter most promising compounds were established considering CNS drug-likeness reports [55–57], adapted by us, as follows: MW ≤ 360, PSA ≤ 90 $Å^2$, $^{QP}logP_{ow}$ = -2.0 - 6.5; $^{QP}logBB$ > -0,5;

heteroatoms ≤ 21; human oral absorption > 80%; $^{QP}P_{Caco}$ > 500 nm/s; $^{QP}P_{MDCK}$ > 500 nm/s. DEREK toxicity endpoints (carcinogenicity, genotoxicity, neurotoxicity, cardiotoxicity, hepatotoxicity, among others, both for mammals/bacteria) were predicted and rejected when respective alerts were fired as 'plausible', 'probable' or 'certain'.

2.4. Molecular docking simulations

The GSK-3β structure used here (PDB code 1PYX) was imported from the Protein Data Bank (PDB, https://www.rcsb.org/) and input into the Protein Preparation Wizard software [58,59]. Pre-processing was done by checking the following functions in sequence: assignment of bond orders using the CCD (Chemical Component Dictionary) database [60], addition of hydrogens, and removal of water molecules, cofactors, metals and the native ligand. We used Glide's extra precision (XP) scoring function, did not allow N inversion, kept the Ser236 residue flexible, and used a 10x10x10 Å grid centered at the centroid of GSK-3β allosteric pocket (x = 11.78, y = 13.15, z = 38.33). Note that the development and validation of this docking protocol are further detailed in our previous paper [19]. At last, docking figures were prepared using either Maestro [61] or PyMOL [62] software.

2.5. GSK-3β biological assays

Biological activity assays were carried out at the International Centre for Kinase Profiling, University of Dundee, following the methods described by Davies et al. (2000) [63]. In brief, a radiometric filter binding assay using [γ-^{33}P]ATP (concentration of approximately K_m, five μM), substrate Phospho GS2 YRRAAVPPSPSLSRHSSPHQS*EDEEE (20 μM) was considered. Moreover, assays were conducted with the enzyme GSK-3β in 8 mM Mops, pH 7.0, containing 0.2 mM EDTA, and a compound concentration of 100 μM. All assays were performed in duplicate for 40 min at ambient temperature in 25 μL incubations.

2.6. Other computational analyses

Similarity radar graph showed in **Figure 4** was built considering the respective calculation of MACCS and Morgan molecular fingerprints, computed by RDKit node implemented in KNIME, and expressed in terms of Tanimoto indices.

Blind docking was performed using the AutoDock software [64,65] implemented on PyRx [66]. To reproduce a blind docking previously performed on literature [67] – which closely corroborates our cavity detection and surface mapping evaluation [19] – the following parameters were adopted: centroid x=26.7110, y=5.9133, z=38.2832, a grid of 193 × 166 × 167 Å with a spacing of 0.375 Å, and Lamarckian GA (200 number of GA runs, 200 population size, 2500000 max number of evals). Poses generated for each compound were manually inspected and clustered according to their apparent overlap and/or interactions within the orthosteric and allosteric sites. The best-scored pose was then retrieved from each cluster of poses in each site. The best-scored pose obtained for each compound at the orthosteric and allosteric sites was then submitted to pocket coverage (P.C.) evaluation using the DoGSiteScorer software [68,69].

3. Results and Discussion

3.1. Virtual screening using workflow 1

Workflow 1 consisted mainly of shape-based screening considering the ROCS 3D shape-based similarity between the query – well-validated bioactive pose predicted for Compound **1** [19] – and the prepared VS set. Specifically, we retrieved 5,000 molecules from each one of 17 databases, i.e., 85,000 molecules out of more than 39 million molecules.

The 17 databases were selected considering their chemical diversity, including commercially available compounds, natural products, and pre-filtered collections of drug-like or CNS molecules (see Materials and Methods). Worth noting that some of these databases correspond to previous versions and/or preprocessed databases from our lab that have shown to be an efficient source of hits on previous works regarding virtual screening ([70–72]).

After using ROCS, compounds were filtered using the EON electrostatic potential similarity between them and the same query, resulting in 17,000 compounds. This combination of methodologies (i.e., shape- and electrostatic-filtering) has also achieved successful results, as reported in the literature ([71–73]).

For ADME predictions, emphasis was given to physicochemical and pharmacokinetic properties well-known in the literature to correlate with drugs acting in the CNS [55–57]. Based on that, we adapted and derived some rules to filter out compounds that could most likely become drug candidates in such a context. Therefore, 1,100 molecules survived after QikProp filtering. These were submitted to toxicity prediction using DEREK, and we filtered out those that showed at least 'plausible' predictions for considered endpoints.

At this point, 656 survival compounds were submitted to structure-based screening using our previously developed Glide docking protocol [19]. This docking protocol was thoroughly developed considering that there is no experimentally obtained complex of protein-ligand at the allosteric cavity of GSK-3β reported so far; so, we applied *in silico* studies to reinforce both the allosteric cavity as well as the bioactive pose of existing allosteric modulators of this enzyme. We briefly mention that this protocol has been compared with three other docking software/methodologies after systematically checking the possibilities of grid size, residues flexibilities, scoring functions, consensus between docking poses, and corroboration with surface maps, cavity detection analyses, and molecular dynamics. Moreover, this protocol has been successfully validated by using analysis of the Area Under the Receiver Operating Characteristic (AU-ROC) curves built for the classification of actives and inactives/decoys.

Hence, after applying this docking protocol and careful inspection of interactions between survival compounds within the allosteric cavity of GSK-3β, together with their previously obtained results from ROCS, EON, and ADMET properties, we selected 11 compounds for experimental validation.

3.2. Virtual screening using workflow 2

When searching for GSK-3β modulators in the literature, one may find that there is a considerable amount of ATP competitive (orthosteric) and non-ATP competitive inhibitors. For the latter, however, data concerning which other pocket could host such (allosteric) inhibitors are rarely found. Therefore, we performed an exhaustive paper-mining to compile the dataset of 88 GSK-3β inhibitors potentially acting at the concerned allosteric pocket (see **Table S1**).

Since this dataset consists of 88 'cherry-picked' compounds, curation was conducted without removing duplicates and/or further issues of chemical structure corrections.

As mentioned in methods, we developed three models using three sets of descriptors, i.e., Model 1 using RDKit, Model 2 using Morgan, and Model 3 using RDKit+Morgan. Recall that each model consists of 5 consensus models, each generated using Random Forest and a different combination of training/test sets from the five-fold cross-validation. Their statistical characteristics are summarized in **Table 1** and **Figure 2**.

In Model 1, the most relevant RDKit properties were selected after filtering out those with low variance (< 10 %) and analyzing the correlation between them. A correlation matrix heatmap was generated (**Fig. S1**) considering the 28 properties selected to develop Model 1. Properties selected for modeling did not present more than 90 % correlation with each other. Furthermore, **Fig. S1** also shows each descriptor's importance calculated after Model 1 generation.

All models showed statistical characteristics higher than the usual threshold of 0.6 [74]. Nevertheless, the QSAR model built with RDKit descriptors (Model 1) showed the highest correct classification rate (CCR) of 0.87, a sensitivity of 0.84, a specificity of 0.90, and positive and negative predictive values (PPV and NPV) of 0.86 and 0.88, respectively (see **Fig. 2** and **Table 1**). It also showed the best statistics for the Y-randomization validation compared to Models 2 and 3 (see **Fig. S2**).

This model was thus employed in a virtual screening campaign using the MolPort Screening Compounds and Enamine databases (more than 10 million compounds). The top 40,000 compounds, predicted as active and within the DA, were sorted according to their QSAR confidence/score prediction values, followed by filtration of favorable ADMET properties. This resulted in 76 compounds that were docked using a well-established Glide protocol to inspect preferential interactions within the allosteric pocket of the enzyme. Finally, 7 compounds were selected for purchase.

Table 1. Statistical characteristics of developed machine learning QSAR models.

Model	Algorithm	Descriptors	CCR	Sensitivity	PPV	Specificity	NPV
1	Random Forest	RDKit properties	0.87	0.84	0.86	0.90	0.88
2	Random Forest	Morgan fingerprints	0.83	0.76	0.85	0.90	0.84
3	Random Forest	RDKit+Morgan	0.84	0.78	0.85	0.90	0.85

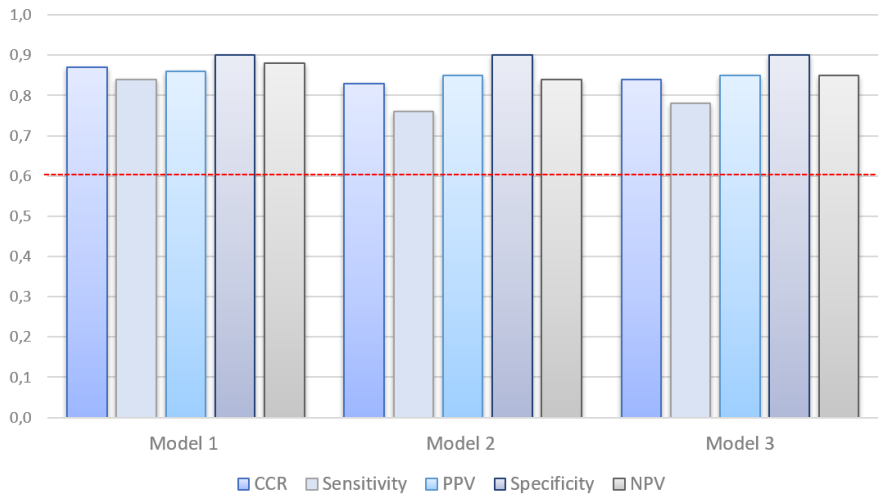

Figure 2. Statistical characteristics of developed machine learning QSAR models.

3.3. Analyzing identified hits

We identified 18 virtual hits from virtual screening workflows 1 and 2 (see **Fig. 3**). These compounds were purchased and experimentally evaluated using the assay methods described earlier. Due to resource limitations, we were able to only run preliminary assays to identify their GSK-3β percentage of inhibition at 100 µM. It is worth noting that the most active compounds used as references in this work have IC_{50} values near 3 µM, which values of inhibition (in %)

would not be too far from those obtained with our hits in similar conditions. Moreover, according to Palomo et al. [9], we should not expect a great inhibition value since the allosteric mechanism is expected to provide a subtle modulation of GSK-3β. Such phenomena should be advantageous because of a smaller chance of producing side effects. In addition, considering the challenging proposition of this work, i.e., the low amount of data available (few known inhibitors as well as no ligand-receptor crystallographic complexes), such hits obtained by virtual screening are promising to expand the chemical space of GSK-3β modulation.

Eventually, 5 hits were confirmed to be active, with % values of GSK-3β inhibition of 12.44, 18.41, 14.02, 11.59, and 16.77 for compounds **LCQFGS03**, **04**, **06**, **12,** and **15**, respectively (highlighted in **Table 2**). The first three were obtained by screening workflow 1 and the remaining by workflow 2. Assuming 10 % of inhibition as a threshold, 3/11 and 2/7 of compounds reasonably active were achieved by workflows 1 and 2, respectively, corresponding to a hit rate of ~ 28% from our screening workflows.

Table 2. Compounds obtained by virtual screening and their corresponding percentage of GSK-3β inhibition determined by assays and their physicochemical and ADMET properties.

Compound	GSK-3β % inhibition	VS Workflow	ROCS TC	QSAR score	Glide score	MW	PSA	$^{QP}logP_{o/w}$	$^{QP}logBB$	%HOA
1	-	-	-	-	-6.130	429.25	118.58	5.84	-2.04	94.20
LCQFGS01	0.10	1	0.884	-	-5.919	312.81	48.80	5.14	-0.425	100.00
LCQFGS02	0.01	1	0.917	-	-5.622	269.35	57.19	3.02	-0.357	100.00
LCQFGS03	**12.44**	1	0.844	-	-6.065	351.17	56.01	4.26	-0.094	100.00
LCQFGS04	**18.41**	1	0.938	-	-5.407	287.38	41.11	2.83	0.056	100.00
LCQFGS05	9.98	1	0.947	-	-5.045	287.32	71.22	2.82	-0.238	100.00
LCQFGS06	**14.02**	1	0.833	-	-5.679	308.81	62.67	2.52	-0.100	96.44
LCQFGS07	0.63	1	0.907	-	-4.760	304.40	75.36	2.10	-0.504	90.79
LCQFGS08	8.20	1	0.866	-	-5.355	299.68	35.22	4.4	0.389	100.00
LCQFGS09	4.32	1	0.813	-	-5.171	297.37	41.66	4.19	0.018	100.00
LCQFGS10	4.04	1	0.852	-	-5.005	325.20	73.61	2.67	-0.333	96.01
LCQFGS11	6.65	1	0.858	-	-5.225	276.34	71.36	1.20	-0.248	87.87
LCQFGS12	**11.59**	2	-	0.894	-5.131	349.40	75.98	2.84	-0.317	100.00
LCQFGS13	4.33	2	-	0.839	-6.711	351.40	92.98	3.56	-1.064	100.00
LCQFGS14	1.88	2	-	0.838	-4.865	320.39	69.97	3.69	-0.279	100.00
LCQFGS15	**16.77**	2	-	0.833	-5.041	356.81	81.25	4.26	-0.711	100.00
LCQFGS16	5.76	2	-	0.831	-5.683	341.41	79.89	3.12	-0.522	100.00
LCQFGS17	4.26	2	-	0.820	-5.518	312.79	78.40	2.63	-0.413	95.06
LCQFGS18	7.13	2	-	0.819	-5.716	342.44	60.07	3.06	-0.051	93.46

ROCS TanimotoCombo indexes; QSAR confidence/score prediction values; Glide extra precision (XP) docking score values in kcal/mol; Molecular weight in g/mol; Polar surface area; Logarithm of partition coefficient in 1-octanol/water predicted by QikProp; Logarithm of blood-brain barrier predicted by QikProp; Human oral absorption in %.

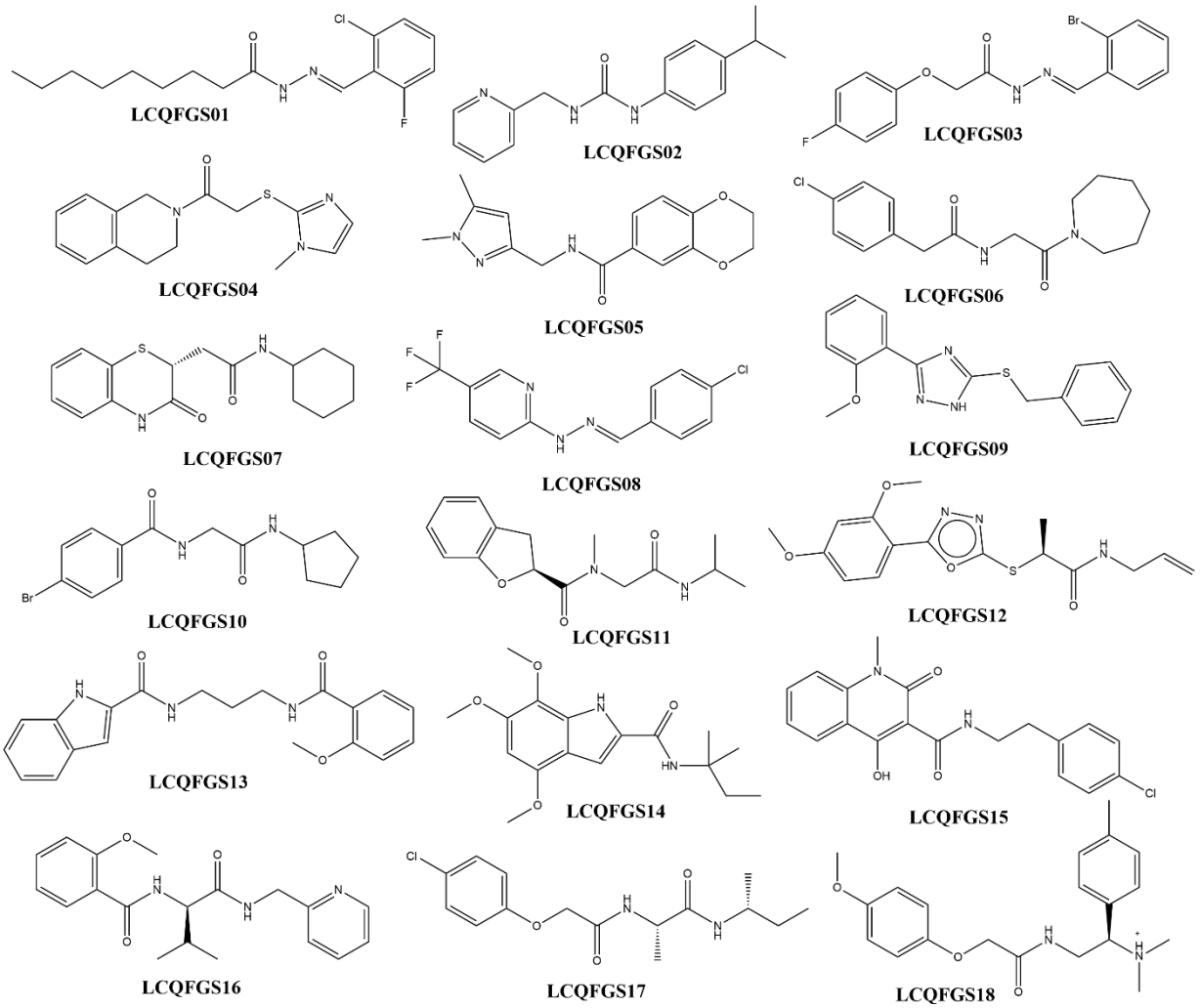

Figure 3. Representation of molecular structures of 18 hits (**LCQFGS01**-18) identified by VS workflows.

Table 2 reports ROCS TC values for compounds obtained from workflow 1, QSAR scores for compounds obtained from workflow 2, and docking scores for all of them. All hits presented medium-to-high values of these three metrics, ranging from 0.813-0.947 for ROCS TC, 0.819-0.894 for QSAR, and -4.760 – -6.130 for docking scores. Apparently, there is no obvious relationship between the value of these three metrics and their corresponding % of GSK-3β inhibition. For instance, **LCQFGS04** showed the highest GSK-3β inhibition but not the highest docking score. Interestingly, compounds **LCQFGS01** and **02** were identified as potential

virtual hits in our previous work [19], however they did not show reasonable GSK-3β inhibition according to the criteria adopted here.

We also employed ADMET models to analyze pharmacokinetic and safety profiles for the top five hits, and compared them to **1** – the known GSK-3β allosteric inhibitor reported in the literature [67]. All five compounds analyzed here presented greater predicted values for human oral absorption as well as ^{QP}logBB values greater than -0.5 (except for **LCQFGS15**), which is a crucial threshold for compounds designed and intended to act at the CNS. Furthermore, their MW and PSA values are within the desired range for AD drugs, as proposed in the literature [55–57].

Overall, despite their low % values of GSK-3β inhibition, there was a clear improvement of their ADMET properties. Such properties are well-suited for drug candidates associated with CNS disorders and were planned in advance, during filtering steps for each screening workflow. In view of this, we intend to use these hits to propose structural optimizations that shall result in more potent compounds, and obtain corresponding IC_{50} values of GSK-3β inhibition in the future.

3.3.1. Structure-activity relationships and structural diversity

In order to propose possible interactions that could be responsible for the observed inhibition of 5 selected hits, we hypothesized structure-activity relationships based on their docking poses. We considered three key interactions of GSK-3β allosteric pocket (previously reported on [9,19]): one hydrogen bond with the N-H of the Ser236 backbone; cation-π (or electrostatic/hydrophobic) interaction of aliphatic/aromatic group under the Arg209; and hydrophobic interactions between aliphatic/aromatic group and hydrophobic region defined by Thr330, Pro331, and Arg328.

In **Figure 4**, one may observe that **LCQFGS04** shows a hydrogen bond interaction between its carbonyl C=O and the backbone amide N-H of the residue Ser236. Moreover, it fits

the electrostatic potential surface map of the allosteric cavity, posing its (non)aromatic ring under the Arg209, an expected hotspot for this type of ring.

LCQFGS15 showed the same indispensable hydrogen bond and a hydrogen bond between its hydroxyl OH and the C=O of His173 (**Fig. 4**). Worth noting that this hydroxyl corresponds to a more favorable specie in the corresponding keto-enol tautomerization. It also showed the third interaction between its chloride atom and the NH of Gly210. The latter is interesting (and also previously reported) because it also constrains the aromatic ring under the Arg209. **Figures S3-5** show further details of docking interactions observed for the remaining hits.

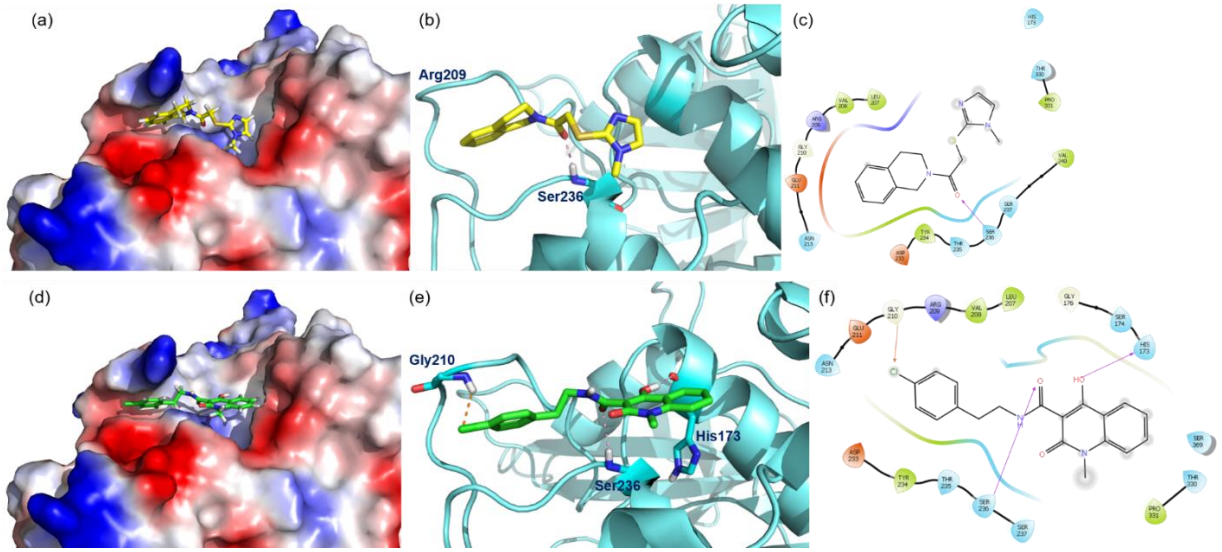

Figure 4. Docking poses (in stick representations) of compounds making enzyme-inhibitor interactions observed on the 3D surface of the GSK-3β allosteric pocket, with respective enzyme key residues detailed in the Ribbons representation, as well as 2D diagrams of such interactions, for the two best hits selected as putative allosteric inhibitors: **LCQFGS04** (a, b, c) and **LCQFGS15** (d, e, f).

In addition, we constructed a similarity radar graph (**Fig. 5**) to illustrate the structural diversity achieved by the 5 hits, built by comparing their chemical similarities in terms of molecular fingerprints (MACCS and Morgan) and corresponding Tanimoto indices.

Considering MACCS fingerprints, compounds **LCQFGS03**, **04**, **06**, **12** and **15** showed the respective values of 0.45, 0.46, 0.44, 0.48, and 0.68. Whereas Morgan fingerprints: 0.14, 0.16, 0.14, 0.10, and 0.38, respectively. Both indicate that all compounds showed small similarity

values in comparison to the reference Compound **1** (most representative GSK-3β allosteric modulator) and, thus, that there was a reasonable increase in structural diversity of hits found. **LCQFGS15** was the only compounds with similarity higher than 0.60 (Morgan) due to the fact that it bears an almost identical quinolone ring when compared to reference **1**.

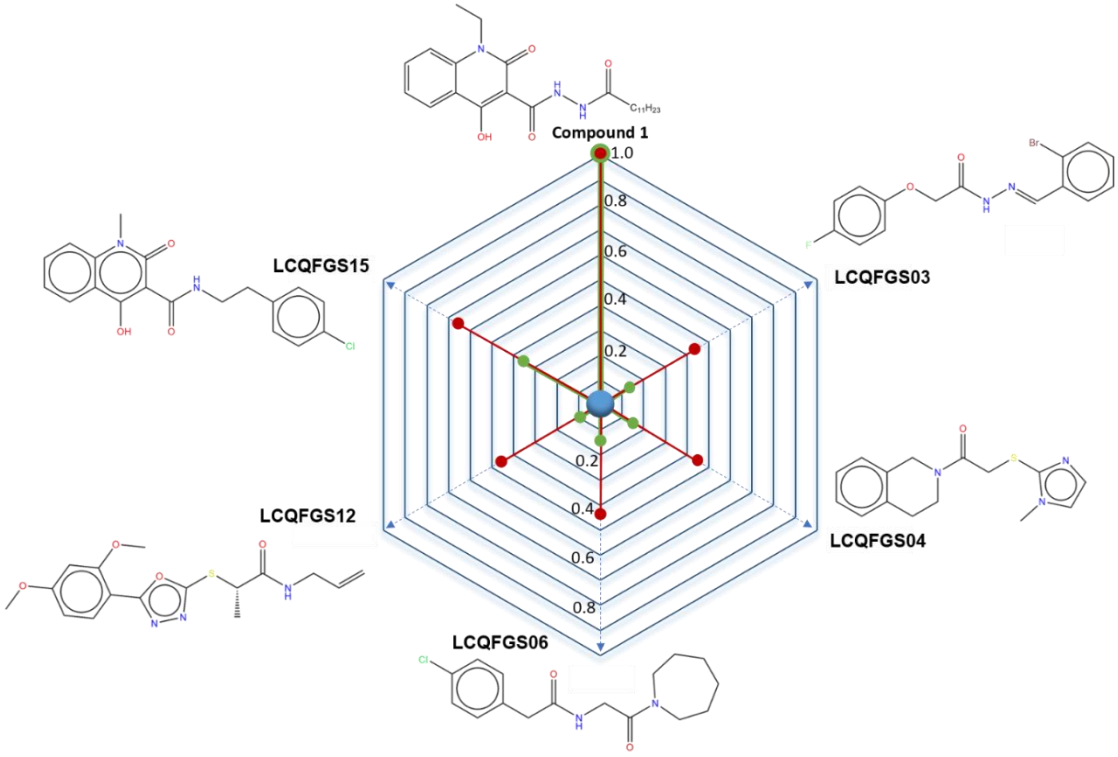

Figure 5. Radar graph showing similarities between the reference Compound **1** and **five LCQFGS hits**, depicted in terms of the calculation of respective Tanimoto indexes, using two similarity fingerprints (MACCS in red and Morgan in green).

3.3.2. In silico reinforcement//validation of allosteric modulation

In order to reinforce the hypothesis that the 5 hits found here might truly interact with allosteric pocket of GSK-3β, we performed an additional blind docking evaluation. For this, we used the AutoDock software, as described in methodologies *(section 2.6)*.

It is acknowledgeable that the orthosteric pocket possesses a greater volume than the allosteric pocket. However, considering that the allosteric pocket was reported by the DogSiteScorer software as split into two independent pockets *(pocs 2 & 3)*, the sum of their druggability scores exceeds that of the orthosteric pocket (*poc 1*) (see **Figure 6**). As calculated by DogSiteScorer, the orthosteric pocket *(poc 1)* has a volume of 930.96 $Å^3$ and a druggability

score of 0.764, while the allosteric pocket (constituted by *pocs 2 & 3*) has a volume of 464.29 $Å^3$ (286.58 & 177.71 $Å^3$) and a druggability score of 0.772 (0.550 & 0.222). This may suggest two things: (i) that the allosteric pocket might present a druggability comparable to the orthosteric site (corroborating the previous cavity detection study [19]); (ii) that, proportionally, the pocket occupancy (or coverage, in terms of volume) of compounds might represent a good measure and indication of compound's preference to interact with a given pocket. With this in mind, we analyzed blind docking results with focus on the 'pocket coverage' parameter – along with docking scores -, which should indicate a reasonable metric to depict which pocket is most likely to host the selected compounds.

Since we know that Compound **1** is an actual allosteric modulator that binds to the concerned allosteric pocket of GSK-3β, we used its results as a reference to compare with data obtained for our hits. In view of this, both docking score and pocket coverage values for **1** were greater for its allosteric pose rather than for the orthosteric (see **Figure 6**). Furthermore, results summarized in **Table 3** show that all 5 hits showed greater docking score values for the allosteric pocket, except for **LCQFGS04** and **15**. However, considering their pocket coverage values we observe that all of them showed preference towards the allosteric pocket *(pocs 2 & 3)*.

Table 3. Top-ranked blind-docking poses obtained for each compound, corresponding docking scores as well as pocket coverage values at the orthosteric pocket *(poc 1)* and allosteric pocket *(pocs 2 & 3)* of the enzyme GSK-3β.

Compound	**Orthosteric pocket/pose**		**Allosteric pocket/pose**	
	Score	**P.C. *(poc 1)***	**Score**	**P.C. *(pocs 2 & 3)***
1	-6.91	47.32	**-7.24**	**53.6 & 75.99**
LCQFGS03	-6.98	32.36	**-7.17**	**42.39 & 82.70**
LCQFGS04	**-6.59**	24.60	-6.30	**0.03 & 85.44**
LCQFGS06	-7.19	34.33	**-7.31**	**46.35 & 72.99**
LCQFGS12	-6.13	25.52	**-6.36**	**19.32 & 86.81**
LCQFGS15	**-7.62**	31.32	-7.47	**26.98 & 91.83**

Autodock blind-docking score values, in kcal/mol; P.C. = Pocket coverage values calculated for each corresponding pose, at each poc (pocket), predicted by using the DoGSiteScorer software.

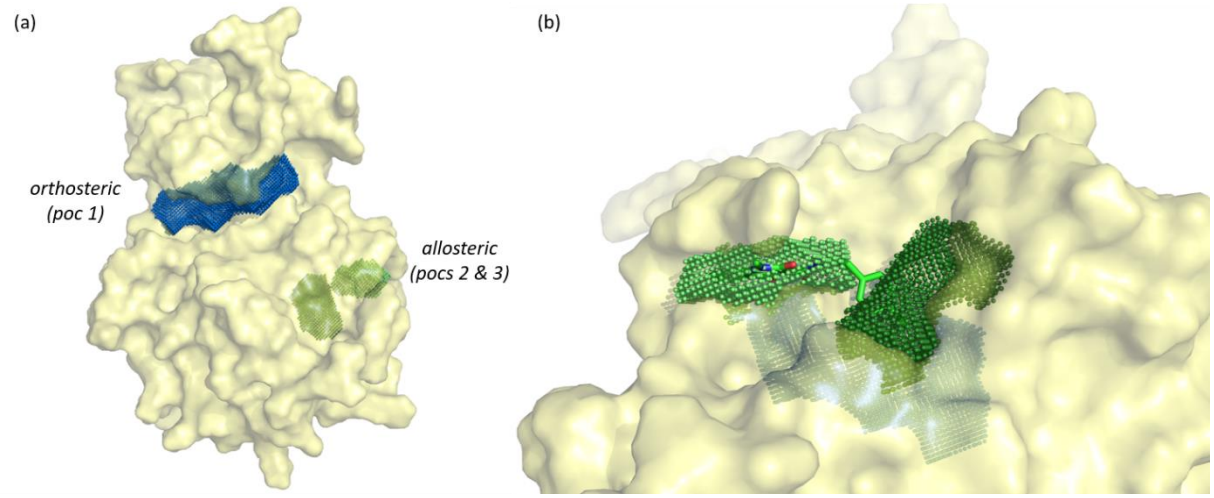

Figure 6. (a) Othosteric (*poc 1*; blue spheres) and allosteric (*pocs 2 & 3*; green spheres) pockets predicted on the GSK-3β surface, by using the DoGSiteScorer software. (b) pocket coverage for the best blind-docking pose of compound **LCQFGS15** (in stick representation) within the GSK-3β allosteric pocket.

4. Conclusions

We have observed that varying the breadth of methodologies in virtual screening studies, and looking for a consensus among them could be a promising strategy that should increasingly be explored and applied to find and optimize new hits. The studies conducted here demonstrate that it is possible to obtain hits with wide structural diversity by mixing LBVS (using shape-based similarity and QSAR models) and SBVS (docking) methods in different workflows. This was possible even with a limited amount of data as starting material for VS workflows and a modest number of molecules within screening databases (in comparison to currently available billions of molecules).

Although the GSK-3β inhibition data reported here could benefit from further determination of IC_{50} values, our results already prove valuable in terms of structure-activity relationship investigation and structural diversity. In fact, increasing the structural diversity of known GSK-3β allosteric inhibitors was a primary goal of this study, which was successfully achieved (as demonstrated in *section 3.3.1)*.

Moreover, compounds were screened for favorable predictions of ADMET properties (and suitable values) in the context of CNS drug candidates. In this way, the five hits reported

here are attractive for continuation of our studies, including lead optimization proposals, with a focus on maintaining favorable physicochemical/pharmacokinetic parameters as well as increasing their potencies, i.e., biological activities as GSK-3β allosteric modulators.

Author contribution
G.M.S. conceptualized, evolved methodologies and research goals, analyzed data results, wrote and edited the manuscript - original draft; S.Q.G., V.M.A and J.E.H evolved methodologies and critically evaluated data results; E.N.M., A.T. and C.H.T.P.S supervised and critically evaluated work results; All authors revised the manuscript.

Acknowledgements
The authors acknowledge *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)*, *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)*, and *Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)* for financial support and fellowships.

Declaration of Competing Interest
AT, VMA, and ENM are co-founders of Predictive, LLC, which develops computational methodologies and software for toxicity prediction. All other authors declare they have nothing to disclose.

References

[1] A. Martinez, C. Gil, D.I. Perez, Glycogen synthase kinase 3 inhibitors in the next horizon for Alzheimer's disease treatment, Int. J. Alzheimers. Dis. 2011 (2011). https://doi.org/10.4061/2011/280502.

[2] C.L. Benn, L.A. Dawson, Clinically Precedented Protein Kinases: Rationale for Their Use in Neurodegenerative Disease, Front. Aging Neurosci. 12 (2020). https://doi.org/10.3389/fnagi.2020.00242.

[3] R. Shukla, N.S. Munjal, T.R. Singh, Identification of novel small molecules against GSK3β for Alzheimer's disease using chemoinformatics approach, J. Mol. Graph. Model. 91 (2019) 91–104. https://doi.org/10.1016/j.jmgm.2019.06.008.

[4] P.A. Vignaux, E. Minerali, D.H. Foil, A.C. Puhl, S. Ekins, Machine learning for discovery of GSK3β inhibitors, ACS Omega. 5 (2020) 26551–26561. https://doi.org/10.1021/acsomega.0c03302.

[5] A. Martinez, D.I. Perez, C. Gil, Lessons Learnt from Glycogen Synthase Kinase 3 Inhibitors Development for Alzheimer ' s Disease, Curr. Top. Med. Chem. 13 (2013) 1808–1819. https://doi.org/10.2174/15680266113139990138.

[6] S. Matsunaga, H. Fujishiro, H. Takechi, Efficacy and safety of glycogen synthase kinase 3 inhibitors for Alzheimer's disease: A systematic review and meta-analysis, J. Alzheimer's Dis. 69 (2019) 1031–1039. https://doi.org/10.3233/JAD-190256.

[7] M.P. Davies, R. Benitez, C. Perez, S. Jakupovic, P. Welsby, K. Rzepecka, J. Alder, C. Davidson, A. Martinez, J.M. Hayes, Structure-Based Design of Potent Selective Nanomolar Type-II Inhibitors of Glycogen Synthase Kinase-3β, J. Med. Chem. 64 (2021) 1497–1509. https://doi.org/10.1021/acs.jmedchem.0c01568.

[8] A. Bidon-Chanal, A. Fuertes, D. Alonso, D.I. Pérez, A. Martínez, F.J. Luque, M. Medina, Evidence for a new binding mode to GSK-3: Allosteric regulation by the marine compound palinurin, Eur. J. Med. Chem. 60 (2013) 479–489. https://doi.org/10.1016/j.ejmech.2012.12.014.

[9] V. Palomo, D.I. Perez, C. Roca, C. Anderson, N. Rodríguez-Muela, C. Perez, J.A. Morales-Garcia, J.A. Reyes, N.E. Campillo, A.M. Perez-Castillo, L.L. Rubin, L. Timchenko, C. Gil, A. Martinez, Subtly Modulating Glycogen Synthase Kinase 3 β: Allosteric Inhibitor Development and Their Potential for the Treatment of Chronic Diseases, J. Med. Chem. 60 (2017) 4983–5001. https://doi.org/10.1021/acs.jmedchem.7b00395.

[10] K.H. Bleicher, H.J. Böhm, K. Müller, A.I. Alanine, Hit and lead generation: Beyond high-throughput screening, Nat. Rev. Drug Discov. 2 (2003) 369–378. https://doi.org/10.1038/nrd1086.

[11] E. Lionta, G. Spyrou, D.K. Vassilatis, Z. Cournia, Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances, Curr. Top. Med. Chem.

14 (2014) 1923–1938. https://doi.org/10.2174/1568026614666140929124445.

[12] C. Gorgulla, A. Jayaraj, K. Fackeldey, H. Arthanari, Emerging frontiers in virtual drug discovery: From quantum mechanical methods to deep learning approaches, Curr. Opin. Chem. Biol. 69 (2022) 102156. https://doi.org/10.1016/j.cbpa.2022.102156.

[13] D.E. Gloriam, Bigger is better in virtual drug screens, Nature. 566 (2019) 193–194. https://doi.org/10.1038/d41586-019-00145-6.

[14] C. Gorgulla, A. Boeszoermenyi, Z.-F. Wang, P.D. Fischer, P.W. Coote, K.M. Padmanabha Das, Y.S. Malets, D.S. Radchenko, Y.S. Moroz, D.A. Scott, K. Fackeldey, M. Hoffmann, I. Iavniuk, G. Wagner, H. Arthanari, An open-source drug discovery platform enables ultra-large virtual screens, Nature. 580 (2020) 663–668. https://doi.org/10.1038/s41586-020-2117-z.

[15] D.E. Clark, Virtual Screening: Is Bigger Always Better? Or Can Small Be Beautiful?, J. Chem. Inf. Model. 60 (2020) 4120–4123. https://doi.org/10.1021/acs.jcim.0c00101.

[16] V.M. Alves, T. Bobrowski, C.C. Melo-Filho, D. Korn, S. Auerbach, C. Schmitt, E.N. Muratov, A. Tropsha, QSAR Modeling of SARS-CoV M pro Inhibitors Identifies Sufugolix, Cenicriviroc, Proglumetacin, and other Drugs as Candidates for Repurposing against SARS-CoV-2, Mol. Inform. 40 (2021) 2000113. https://doi.org/10.1002/minf.202000113.

[17] N.D. Elangovan, A.K. Dhanabalan, K. Gunasekaran, R. Kandimalla, D. Sankarganesh, Screening of potential drug for Alzheimer's disease: a computational study with GSK-3 β inhibition through virtual screening, docking, and molecular dynamics simulation, J. Biomol. Struct. Dyn. 39 (2021) 7065–7079. https://doi.org/10.1080/07391102.2020.1805362.

[18] Y. Lee, S.-B. Yoon, H. Hong, H.Y. Kim, D. Jung, B.-S. Moon, W.-K. Park, S. Lee, H. Kwon, J. Park, H. Cho, Discovery of GSK3β Inhibitors through In Silico Prediction-

and-Experiment Cycling Strategy, and Biological Evaluation, Molecules. 27 (2022) 3825. https://doi.org/10.3390/molecules27123825.

[19] G.M. Silva, R.S. Borges, K.L.B. Santos, L.B. Federico, I.A.G. Francischini, S.Q. Gomes, M.P. Barcelos, R.C. Silva, C.B.R. Santos, C.H.T.P. Silva, Revisiting the Proposition of Binding Pockets and Bioactive Poses for GSK-3β Allosteric Modulators Addressed to Neurodegenerative Diseases, Int. J. Mol. Sci. 2021, Vol. 22, Page 8252. 22 (2021) 8252. https://doi.org/10.3390/IJMS22158252.

[20] S. Brogi, A. Ramunno, L. Savi, G. Chemi, G. Alfano, A. Pecorelli, E. Pambianchi, P. Galatello, G. Compagnoni, F. Focher, G. Biamonti, G. Valacchi, S. Butini, S. Gemma, G. Campiani, M. Brindisi, First dual AK/GSK-3β inhibitors endowed with antioxidant properties as multifunctional, potential neuroprotective agents, Eur. J. Med. Chem. 138 (2017) 438–457. https://doi.org/10.1016/j.ejmech.2017.06.017.

[21] P. Zhang, H.-R.R. Hu, S.-H.H. Bian, Z.-H.H. Huang, Y. Chu, D.-Y.Y. Ye, Design, synthesis and biological evaluation of benzothiazepinones (BTZs) as novel non-ATP competitive inhibitors of glycogen synthase kinase-3β (GSK-3β), Eur. J. Med. Chem. 61 (2013) 95–103. https://doi.org/10.1016/j.ejmech.2012.09.021.

[22] P. Zhang, S. Li, Y. Gao, W. Lu, K. Huang, D. Ye, X. Li, Y. Chu, Novel benzothiazinones (BTOs) as allosteric modulator or substrate competitive inhibitor of glycogen synthase kinase 3β (GSK-3β) with cellular activity of promoting glucose uptake, Bioorganic Med. Chem. Lett. 24 (2014) 5639–5643. https://doi.org/10.1016/j.bmcl.2014.10.078.

[23] OpenEye, Scientific Software. ROCS 3.4.0.4, (2020).

[24] P.C.D.P.C.D. Hawkins, A.G.G. Skillman, A. Nicholls, Comparison of shape-matching and docking as virtual screening tools, J. Med. Chem. 50 (2007) 74–82. https://doi.org/10.1021/jm0603365.

[25] ChemBridge, The gold standard in small molecule screening libraries and building blocks (https://www.chembridge.com/screening_libraries/index.php), (2020).

[26] OpenEye, Scientific Software (https://www.eyesopen.com/), (2020).

[27] IBS, InterBioScreen - Databases, (2020). https://www.ibscreen.com/bases.

[28] Maybridge, The Maybridge Screening Collection, (2014). https://www.alfa.com/pt/maybridge-pre-plated-screening-compounds-and-fragment-libraries/ (accessed February 15, 2023).

[29] Molport, Chemical Libraries & Screening Compound Database, (2020). https://www.molport.com/shop/screeening-compound-database.

[30] X. Zeng, P. Zhang, W. He, C. Qin, S. Chen, L. Tao, Y. Wang, Y. Tan, D. Gao, B. Wang, Z. Chen, W. Chen, Y.Y. Jiang, Y.Z. Chen, NPASS: Natural product activity and species source database for natural product research, discovery and tool development, Nucleic Acids Res. (2018). https://doi.org/10.1093/nar/gkx1026.

[31] Princeton, Chemistry Princeton University Library, (2015). https://library.princeton.edu/databases/subject/chemistry.

[32] J.J. Irwin, B.K. Shoichet, ZINC--a free database of commercially available compounds for virtual screening., J. Chem. Inf. Model. 45 (2005) 177–82. https://doi.org/10.1021/ci049714+.

[33] T. Sterling, J.J. Irwin, ZINC 15 - Ligand Discovery for Everyone, J. Chem. Inf. Model. 55 (2015) 2324–2337. https://doi.org/10.1021/acs.jcim.5b00559.

[34] OpenEye, Scientific Software. OMEGA 4.0.0.4, (2020).

[35] OpenEye, Scientific Software. Documentation - Applications 2020.0.4, (2020). https://docs.eyesopen.com/applications/index.html (accessed February 7, 2020).

[36] P.C.D. Hawkins, A.G. Skillman, G.L. Warren, B.A. Ellingson, M.T. Stahl, Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures

from the Protein Databank and Cambridge Structural Database, J. Chem. Inf. Model. 50 (2010) 572–584. https://doi.org/10.1021/ci100031x.

[37] C.H.T. de P. da Silva, C.A. Taft, 3D descriptors calculation and conformational search to investigate potential bioactive conformations, with application in 3D-QSAR and virtual screening in drug design, J. Biomol. Struct. Dyn. 35 (2017) 2966–2974. https://doi.org/10.1080/07391102.2016.1237382.

[38] OpenEye, Scientific Software. EON 2.3.3.4, (2020).

[39] D. Fourches, E. Muratov, A. Tropsha, Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research, J. Chem. Inf. Model. 50 (2010) 1189–1204. https://doi.org/10.1021/ci100176x.

[40] D. Fourches, E. Muratov, A. Tropsha, Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation, J. Chem. Inf. Model. 56 (2016) 1243–1252. https://doi.org/10.1021/acs.jcim.6b00129.

[41] M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, KNIME: The Konstanz Information Miner, in: Stud. Classif. Data Anal. Knowl. Organ., 2008: pp. 319–326. https://doi.org/10.1007/978-3-540-78246-9_38.

[42] KNIME AG, KNIME Analytics Platform 4.3.1, (2021).

[43] J. Figueras, Morgan revisited, J. Chem. Inf. Comput. Sci. 33 (1993) 717–718. https://doi.org/10.1021/ci00015a009.

[44] H.L. Morgan, The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service., J. Chem. Doc. 5 (1965) 107–113. https://doi.org/10.1021/c160017a018.

[45] D. Rogers, M. Hahn, Extended-Connectivity Fingerprints, J. Chem. Inf. Model. 50 (2010) 742–754. https://doi.org/10.1021/ci100050t.

[46] RDKit: Open-source cheminformatics, http://www.rdkit.org, (2021).

[47] S. Riniker, G.A. Landrum, Open-source platform to benchmark fingerprints for ligand-based virtual screening, J. Cheminform. 5 (2013) 26. https://doi.org/10.1186/1758-2946-5-26.

[48] E.N. Muratov, J. Bajorath, R.P. Sheridan, I. V. Tetko, D. Filimonov, V. Poroikov, T.I. Oprea, I.I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtalolo, D. Fourches, Y. Cohen, A. Aspuru-Guzik, D.A. Winkler, D. Agrafiotis, A. Cherkasov, A. Tropsha, QSAR without borders, Chem. Soc. Rev. 49 (2020) 3525–3564. https://doi.org/10.1039/D0CS00098A.

[49] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR Modeling: Where Have You Been? Where Are You Going To?, J. Med. Chem. 57 (2014) 4977–5010. https://doi.org/10.1021/jm4004285.

[50] L. Breiman, Random Forests, Mach. Learn. 2001 451. 45 (2001) 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324.

[51] F. Pedregosa FABIANPEDREGOSA, V. Michel, O. Grisel OLIVIERGRISEL, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot andÉdouardand, A. Duchesnay, Fré. Duchesnay EDOUARDDUCHESNAY, Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[52] A. Tropsha, A. Golbraikh, Predictive QSAR modeling workflow, model applicability domains, and virtual screening., Curr. Pharm. Des. 13 (2007) 3494–504.

[53] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, A Novel Automated Lazy Learning QSAR (ALL-QSAR) Approach: Method Development, Applications, and Virtual Screening of Chemical Databases Using Validated ALL-QSAR Models, J. Chem. Inf. Model. 46 (2006) 1984–1995. https://doi.org/10.1021/ci060132x.

[54] Enamine, Screening Collection, (2021). https://enamine.net/compound-collections/screening-collection.

[55] A.K. Ghose, T. Herbertz, R.L. Hudkins, B.D. Dorsey, J.P. Mallamo, Knowledge-based, central nervous system (CNS) lead selection and lead optimization for CNS drug discovery, ACS Chem. Neurosci. 3 (2012) 50–68. https://doi.org/10.1021/cn200100h.

[56] Z. Rankovic, CNS Physicochemical Property Space Shaped by a Diverse Set of Molecules with Experimentally Determined Exposure in the Mouse Brain, J. Med. Chem. 60 (2017) 5943–5954. https://doi.org/10.1021/acs.jmedchem.6b01469.

[57] T.T. Wager, X. Hou, P.R. Verhoest, A. Villalobos, Moving beyond rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties, ACS Chem. Neurosci. 1 (2010) 435–449. https://doi.org/10.1021/cn100008c.

[58] Schrödinger, Protein Preparation Wizard - Suite 2015-2, (2018).

[59] 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.

[60] J.D. Westbrook, C. Shao, Z. Feng, M. Zhuravleva, S. Velankar, J. Young, The chemical component dictionary: Complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank, Bioinformatics. (2015). https://doi.org/10.1093/bioinformatics/btu789.

[61] Schrödinger, Maestro Schrödinger Suite 2020-1, (2020).

[62] Schrödinger LLC., The Pymol Molecular Graphics System 1.3, (2010).

[63] S.P. Davies, H. Reddy, M. Caivano, P. Cohen, Specificity and mechanism of action of some commonly used protein kinase inhibitors., Biochem. J. 351 (2000) 95. https://doi.org/10.1042/0264-6021:3510095.

[64] G.M. Morris, D.S. Goodsell, M.E. Pique, W. “Lindy” Lindstrom, R. Huey, S. Forli, W.E. Hart, S. Halliday, R. Belew, A.J. Olson, Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity, J. Comput. Chem. (2009).

[65] The Scripps Research Institute, Autodock 4 (http://autodock.scripps.edu/), (2020).

[66] S. Dallakyan, A.J. Olson, Small-Molecule Library Screening by Docking with PyRx, in: Methods Mol. Biol., Methods Mol Biol, 2015: pp. 243–250. https://doi.org/10.1007/978-1-4939-2269-7_19.

[67] V. Palomo, I. Soteras, D.I. Perez, C. Perez, C. Gil, N.E. Campillo, A. Martinez, Exploring the binding sites of glycogen synthase kinase 3. identification and characterization of allosteric modulation cavities, J. Med. Chem. 54 (2011) 8461–8470. https://doi.org/10.1021/jm200996g.

[68] A. Volkamer, D. Kuhn, T. Grombacher, F. Rippmann, M. Rarey, Combining global and local measures for structure-based druggability predictions, J. Chem. Inf. Model. 52 (2012) 360–372. https://doi.org/10.1021/ci200454v.

[69] A. Volkamer, A. Griewel, T. Grombacher, M. Rarey, Analyzing the topology of active sites: On the prediction of pockets and subpockets, J. Chem. Inf. Model. 50 (2010) 2041–2052. https://doi.org/10.1021/ci100241y.

[70] R.P. Rodrigues, C.H.T.P. da Silva, Discovery of potential neurodegenerative inhibitors in Alzheimer’s disease by casein kinase 1 structure-based virtual screening, Med. Chem. Res. 26 (2017) 3274–3285. https://doi.org/10.1007/s00044-017-2020-9.

[71] L.B. Federico, G.M. Silva, A. de Fraga Dias, F. Figueiró, A.M.O. Battastini, C.B.R. dos Santos, L.T. Costa, J.M.C. Rosa, C.H.T. de Paula da Silva, Identification of novel αβ-tubulin modulators with antiproliferative activity directed to cancer therapy using ligand and structure-based virtual screening, Int. J. Biol. Macromol. 165 (2020) 3040–3050. https://doi.org/10.1016/j.ijbiomac.2020.10.136.

[72] J.R. de Almeida, M. Figueiro, W.P. Almeida, C.H.T. de Paula da Silva, Discovery of novel dual acetylcholinesterase inhibitors with antifibrillogenic activity related to Alzheimer's disease, Future Med. Chem. 10 (2018) 1037–1053. https://doi.org/10.4155/fmc-2017-0201.

[73] L.B. Federico, G.M. Silva, S.Q. Gomes, I.A.G. Francischini, M.P. Barcelos, C.B.R. dos Santos, L.T. Costa, J.M. Campos Rosa, C.H.T. de Paula da Silva, Potential colchicine binding site inhibitors unraveled by virtual screening, molecular dynamics and MM/PBSA, Comput. Biol. Med. 137 (2021) 104817. https://doi.org/10.1016/j.compbiomed.2021.104817.

[74] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, Mol. Inform. 29 (2010) 476–488. https://doi.org/10.1002/minf.201000061.