Journal Pre-proof

Efficient identification of novel anti-glioma lead compounds by machine learning models

Bruno Junior Neves, Jonathan Paulo Agnes, Marcelo do Nascimento Gomes, Marcio Roberto Henriques Donza, Rosângela Mayer Gonçalves, Marina Delgobo, Lauro Ribeiro de Souza Neto, Mario Roberto Senger, Floriano Paes Silva-Junior, Sabrina Baptista Ferreira, Alfeu Zanotto-Filho, Carolina Horta Andrade

PII: S0223-5234(19)31133-X

DOI: https://doi.org/10.1016/j.ejmech.2019.111981

Reference: EJMECH 111981

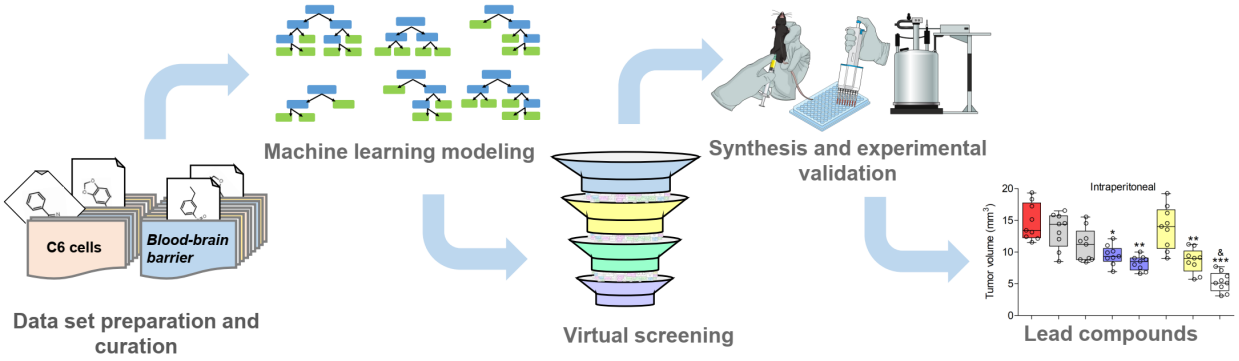
To appear in: *European Journal of Medicinal Chemistry*

Received Date: 3 July 2019

Revised Date: 18 November 2019

Accepted Date: 16 December 2019

Please cite this article as: B.J. Neves, J.P. Agnes, M.d.N. Gomes, M.R. Henriques Donza, Rosâ.Mayer. Gonçalves, M. Delgobo, L. Ribeiro de Souza Neto, M.R. Senger, F.P. Silva-Junior, S.B. Ferreira, A. Zanotto-Filho, C.H. Andrade, Efficient identification of novel anti-glioma lead compounds by machine learning models, *European Journal of Medicinal Chemistry* (2020), doi: https://doi.org/10.1016/j.ejmech.2019.111981.



Efficient Identification of Novel Anti-Glioma Lead Compounds by Machine Learning Models

Bruno Junior Neves[1,2§], Jonathan Paulo Agnes[3§], Marcelo do Nascimento Gomes[2,4], Marcio Roberto Henriques Donza[5], Rosângela Mayer Gonçalves[3], Marina Delgobo[3], Lauro Ribeiro de Souza Neto[6], Mario Roberto Senger[6], Floriano Paes Silva-Junior[6], Sabrina Baptista Ferreira[5], Alfeu Zanotto-Filho[3], and Carolina Horta Andrade[2]**

[1]LabChem – Laboratory of Cheminformatics, Centro Universitário de Anápolis, UniEVANGÉLICA, Anápolis, GO, 75083-515, Brazil.

[2] LabMol – Laboratory for Molecular Modeling and Drug Design, Faculdade de Farmácia, Universidade Federal de Goiás, Goiânia, GO, 74605-510, Brazil.

[3]LabCancer - Laboratório de Farmacologia e Bioquímica do Câncer, Departamento de Farmacologia, Centro de Ciências Biológicas, Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil.

[4]InSiChem Drug Discovery – Universidade Estadual de Goiás, Anápolis, GO, 74643-090, Brazil.

[5]LSOPB – Laboratório de Síntese Orgânica e Prospecção Biológica, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 21949-900, Brazil.

[6] LaBECFar – Laboratório de Bioquímica Experimental e Computacional de Fármacos, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, 21040-900, Brazil.

[§]These authors provided equal contributions to this study and should be considered as co-first authors.

*Corresponding authors: CHA, Tel: +55 62 3209 6451, Fax: +55 62 3209 6037, E-mail: carolina@ufg.br; AZF, Tel: +55 48 37212474, Fax: +55 48 33375479, E-mail: alfeu.zanotto@ufsc.br

ABSTRACT

Glioblastoma multiforme (GBM) is the most devastating and widespread primary central nervous system tumor. Pharmacological treatment of this malignance is limited by the selective permeability of the blood-brain barrier (BBB) and relies on a single drug, temozolomide (TMZ), thus making the discovery of new compounds challenging and urgent. Therefore, aiming to discover new anti-glioma drugs, we developed robust machine learning models for predicting anti-glioma activity and BBB penetration ability of new compounds. Using these models, we prioritized 41 compounds from our in-house library of compounds, for further *in vitro* testing against three glioma cell lines and astrocytes. Subsequently, the most potent and selective compounds were resynthesized and tested *in vivo* using an orthotopic glioma model. This approach revealed two lead candidates, **4m** and **4n**, which efficiently decreased malignant glioma development in mice, probably by inhibiting thioredoxin reductase activity, as shown by our enzymological assays. Moreover, these two compounds did not promote body weight reduction, death of animals, or altered hematological and toxicological markers, making then good candidates for lead optimization as anti-glioma drug candidates.

Keywords: Cancer; glioblastoma; machine learning; predictive modeling; orthotopic glioma model; thioredoxin reductase.

1. Introduction

Glioblastoma multiforme (GBM) or grade IV glioma, is the most commonly occurring and aggressive type of primary central nervous system (CNS) tumor [1,2]. Most of GBM occur particularly in the brain, but they can also appear in cerebellum brainstem and also in spinal cord [3,4]. This malignance can manifest at any age, but is more frequent in adult men with a median age of 64 years [5]. The majority symptoms of GBM include intracranial pressure, headache and focal or progressive neurologic deficits [5]. Consequently, GBM prognosis remains dismal for decades with a median overall survival of ~14 months, with less than 10% of patients surviving beyond five years [6,7].

GBM is particularly difficult to treat due to its characteristic of excessive invasiveness and fast-growing behavior, as well as its particular location and the selective permeability offered by the blood-brain barrier (BBB)[8,9]. At present, chemotherapy is the main postsurgical and adjuvant therapy for GBM, and the alkylating agent temozolomide (TMZ) is the first-line drug frequently combined with radiotherapy [10]. TMZ is absorbed orally and presents favorable toxicity profile compared with older alkylating agents such as carmustine [11]. However, the emergence of temozolomide resistance hampers its use in GBM patients [12,13]. Hence, the discovery of new anti-glioma drugs is urgently needed.

In this context, quantitative machine learning (ML) has exerted profound impact on drug discovery, making it faster and less expensive [14,15]. ML is a growing field of artificial intelligence that uses different statistical techniques to enable computers to learn from various data types without being explicitly [16]. Several ML methods, such as Support Vector Machines [17], Random Forest [18–20], and more recently Deep Neural Networks [21–24], have been utilized for drug discovery. Methodologically, ML tools uses pattern recognition algorithms to discern mathematical relationships between experimental observations of small molecules and extrapolate them to predict biological properties of novel compounds.[25–27] So, ML represents a helpful tool for virtual screening (VS) of new chemicals with desired biological properties.

In this work, we developed ML models and applied them for predicting the anti-proliferative activity against glioma cells and the BBB penetration ability of new compounds from our in-house library. Then, the prioritized compounds were experimentally evaluated *in vitro* against glioma cells and astrocytes, and *in vivo* using an orthotopic glioma model.

2. Results and discussion

The general study design is presented in Fig 1. Briefly, we followed successive steps as follows: (i) data collection, curation, and integration of compounds reported in the literature with activity against C6 glioma cells and BBB penetration; (ii) chemical space analysis of curated datasets; (iii) development and validation of ML models; (iv) VS of an in-house chemical database (1,250 compounds) using the ML models for prioritization of compounds; (v) *in vitro* experimental validation using phenotypic and enzymatic assays; (vi) chemical synthesis of most promising compounds; and (vii) *in vivo* investigation using orthotopic glioma model.

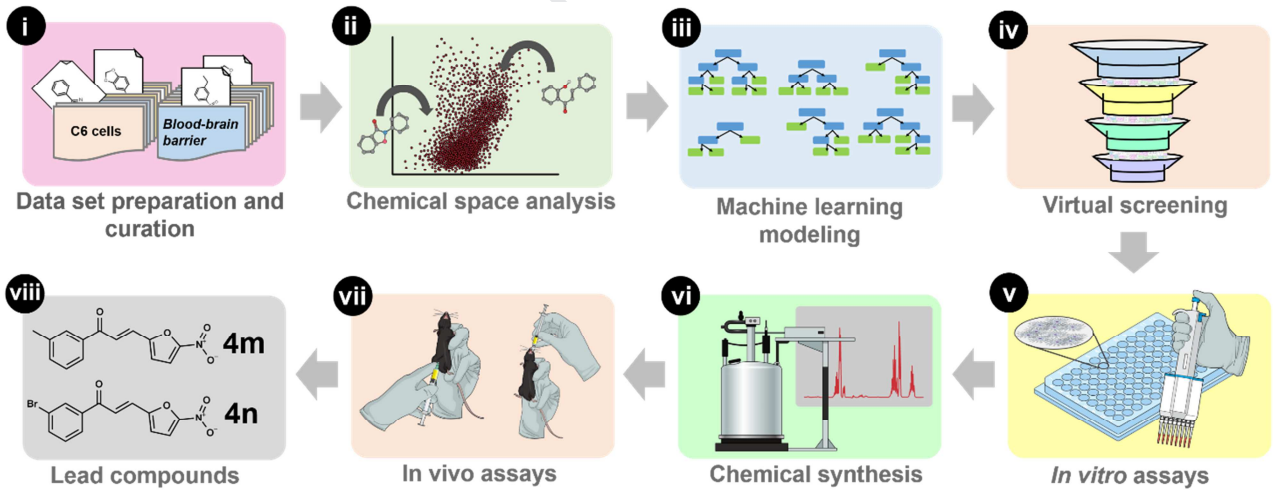


Figure 1. Study design. (i) data collection, curation, and integration of molecules with activity against C6 glioma cells and BBB penetration; (ii) chemical space analysis of curated datasets; (iii) development of binary and continuous QSAR models; (iv) VS of an in-house chemical database (containing 1,250 compounds); (v) *in vitro* experimental validation using phenotypic and enzymatic assays; (vi) chemical synthesis of most promising compounds; (vii) *in vivo* investigation using orthotopic glioma model, and (viii) identification of novel anti-glioma leads.

Initially, two datasets of compounds with anti-glioma (Supplementary File S1) and blood-brain barrier (File S2) bioactivity data were retrieved from the ChEMBL database (ID: CHEMBL614657 [28]) and scientific literature [29–32], respectively. Both datasets were submitted to a rigorous data curation protocol. An activity threshold of 10 μM based on half maximal effective concentration (EC_{50}) against C6 rat glioma cells was defined for discrimination between active and inactive compounds. Compounds with experimental logBB greater than or equal to −1 were labeled as BBB+ (penetrating) and those with logBB below −1 as BBB− (not penetrating). Subsequently, the C6 dataset (97 active and 173 inactive compounds), and BBB dataset (433 BBB− and 1436 BBB+ compounds) were balanced using a linear under-sampling approach [33].

2.1. Chemical space analysis

To visualize the structural diversity of our datasets, we performed a principal component analysis (PCA, see Supplementary Figure S1). The PCA reduces high-dimensional space composed by Molecular ACCess System (MACCS) keys into a smaller number of orthogonal (non-correlated) variables called principal components (PCs), thus making it more manageable and comprehensible by extracting essential information [34,35]. The PCA model with the first two PCs described 34.8% of total data variance. Projecting variables on the planes defined by a PC1 and PC2 allows an interesting chemical space analysis, in which most of the active (green dots) and inactive (yellow dots) compounds from C6 dataset overlap within the same regions of chemical space (defined by PC1 and PC2) of BBB+ (blue dots) and BBB− (red dots) compounds from the BBB dataset. This analysis revealed that multiple compounds active against glioma cells can also potentially penetrate the BBB. Based on these data, we developed predictive computational models for both biological properties in order to select only compounds predicted as active for C6 cells and BBB+.

2.2. Performance of ML models

ML models were built to distinguish active *vs.* inactive compounds for C6 cell line (see Supplementary File S1) and BBB+ *vs.* BBB– compounds (Supplementary File S2). Statistical

characteristics of developed ML models estimated by 5-fold external cross-validation are reported in Table 1. According to the statistical results, the combination of Morgan and FeatMorgan fingerprints (radius 2: FeatMorgan_2, Morgan_2; radius 4: FeatMorgan_4, Morgan_4) with Random Forest algorithm led to predictive binary ML models. Briefly, correct classification rate (CCR) values were ranging between 0.83–0.87; sensitivity (SE) 0.82–0.87; specificity (SP) 0.82–0.87; positive predictive value (PPV) 0.84–0.88; negative predictive value (NPV) 0.81–0.88; and a Cohen's kappa (κ) 0.66–0.73. The model built using Morgan_2 demonstrated the best performance among all other models developed for antiproliferative activity against C6 cells (CCR = 0.87; SE = 0.89; SP = 0.85) and BBB (CCR = 0.85; SE = 0.81; and SP = 0.89).

Table 1. Summarized statistical characteristics of ML models.

Models	CCR	SE	SP	PPV	NPV	κ	Coverage
Glioma (C6 cell line) models							
Morgan_2	0.87	0.89	0.85	0.85	0.88	0.73	0.52
Morgan_4	0.85	0.84	0.86	0.85	0.84	0.69	0.51
FeatMorgan_2	0.85	0.86	0.84	0.84	0.85	0.69	0.59
FeatMorgan_4	0.85	0.86	0.84	0.84	0.85	0.69	0.51
Blood-brain barrier (BBB) models							
Morgan_2	0.85	0.81	0.89	0.88	0.82	0.70	0.59
Morgan_4	0.84	0.81	0.88	0.87	0.82	0.68	0.57
FeatMorgan_2	0.83	0.79	0.87	0.86	0.81	0.66	0.61
FeatMorgan_4	0.84	0.80	0.88	0.87	0.81	0.67	0.57

CCR: correct classification rate; SE: sensitivity; SP: specificity; PPV: positive predictive value; and NPV: negative predictive value; κ: Cohen's kappa; Coverage: percentage of test set compounds within the applicability domain.

2.3. Virtual screening

The virtual screening (VS) was carried out following the workflow presented in Figure 2.

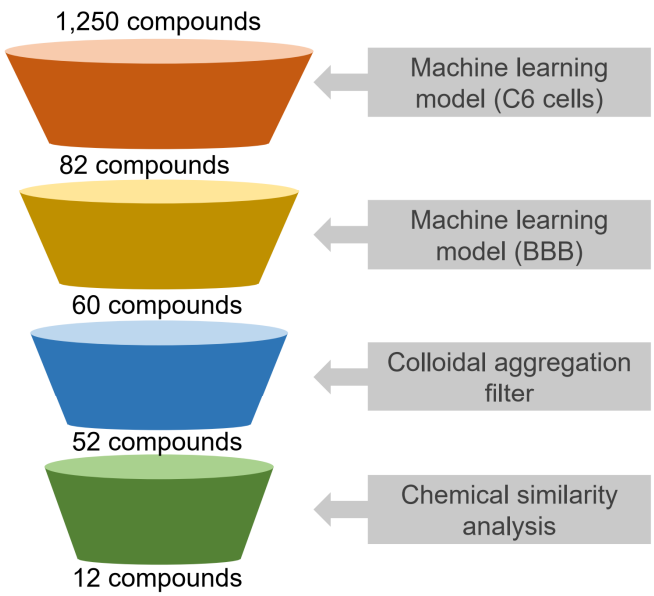


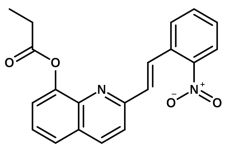
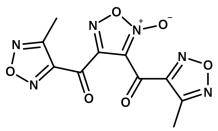
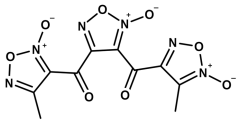
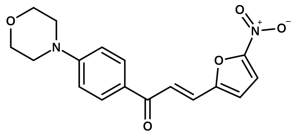
Figure 2. Virtual screening workflow used for identifying new anti-glioma hits. Colloidal aggregation tool was used to filter out molecules that are known to aggregate in experimental assays; chemical similarity analysis and visual inspection were performed to select a subset of structurally diverse virtual hits.

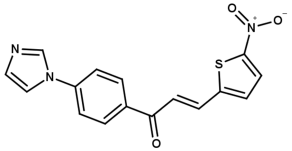
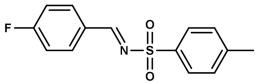
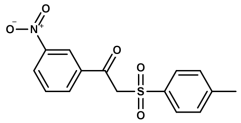
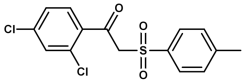
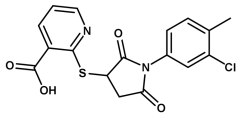
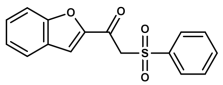
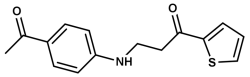
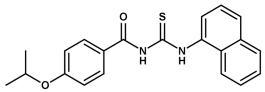
Initially, 1,250 structurally diverse compounds (synthesized or purchased) available in our in-house database were compiled and standardized for VS. Then, the best ML models developed for C6 and BBB were used to prioritize potential anti-glioma compounds. The final selection of hits can be summarized as follows: (i) the compounds predicted as active and BBB+ by the ML models; and (ii) compounds inside the applicability domain (AD) of the ML models. The AD was determined in order to set “reliable” and “unreliable” predictions [36,37]. The predictions were considered reliable when the virtual hits are within the chemical space of the molecules used for training models. Subsequently, a colloidal aggregation tool was used to filter out molecules that are known to aggregate in experimental assays [38,39]. Finally, we performed a chemical similarity analysis to select a subset of structurally diverse virtual hits. At the end of this process, twelve putative hits with model probability >0.65 (Table 2) were selected for biological evaluation.

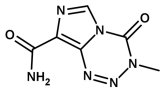
2.4. In vitro antiproliferative activity

The twelve prioritized hits were primarily evaluated *in vitro* against three glioma cell lines (Table 2). The EC_{50} values (see Table 2) indicated that three compounds, 2-[(E)-2-(2-nitrophenyl)ethenyl]quinolin-8-yl propanoate (**1**), (2E)-1-[4-(morpholin-4-yl)phenyl]-3-(5-nitrofuran-2-yl)prop-2-en-1-one (**4**), and (2E)-1-[4-(1H-imidazol-1-yl)phenyl]-3-(5-nitrothiophen-2-yl)prop-2-en-1-one (**5**) were potent at inhibiting the cell growth, showing activities in submicromolar range against C6, U251MG, and U87MG cells.

Table 2. ML probability of selected virtual screening hits, *in vitro* activity against glioma cell lines (C6, U251MG and U87MG), cytotoxicity on astrocytes and selectivity index.

ID	Chemical structure	ML Probability		EC_{50} (µM) ± SD				Selectivity index[b]		
		C6	BBB	C6	U251MG	U87MG	Astrocytes	C6	U251MG	U87MG
1		0.76	0.67	6.3 ± 0.8	10.1 ± 2.2	10.5 ± 2.8	55.2 ± 5.5	8.7	5.4	5.3
2		0.81	0.94	>50	>50	–	–	–	–	–
3		0.82	0.94	>50	>50	–	–	–	–	–
4		0.88	0.96	6.6 ± 2.1	37.8 ± 3.6	9.8 ± 2.9	75.8 ± 8.5	11.5	2	7.7

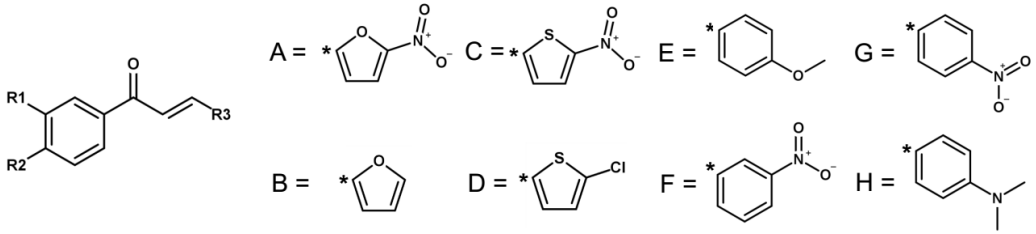
5		0.92	0.94	1.9 ± 1.1	9.4 ± 2.7	10.1 ± 3.4	40.4 ± 3.2	21.3	4.3	4
6		0.68	0.95	48.9 ± 4.4	>50	–	–	–	–	–
7		0.76	0.91	>50	>50	–	–	–	–	–
8		0.70	0.96	>50	>50	–	–	–	–	–
9		0.72	0.68	>50	>50	–	–	–	–	–
10		0.81	0.96	>50	>50	–	–	–	–	–
11		0.80	0.98	46.6 ± 5.2	>50	–	–	–	–	–
12		0.88	0.89	>50	>50	–	–	–	–	–

TMZ[a]		–	–	60.46 ± 3.62	>200	>200	–	–	–	–

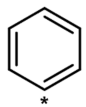
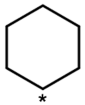
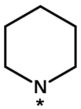
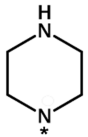
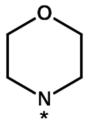
[a]Data retrieved from reference [40]; [b]Selectivity index calculated by astrocyte CC_{50}/glioma cell line EC_{50}; The data are expressed as mean ± SD of three independent assays. Dashed SI values means that cytotoxicity against astrocytes cannot be calculated because compounds did not show activity even at highest concentrations used in the assay.

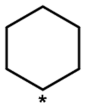
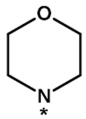
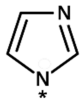
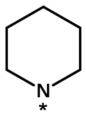
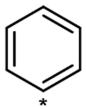
Since heteroaryl chalcones **4** (EC_{50} = 6.6 µM and 9.8 µM for C6 and U87MG, respectively) and **5** (EC_{50} = 1.9 µM, 9.4 µM and 10.1 µM for C6, U251MG and U87MG, respectively) were the most promising anti-glioma hits in experimental assays, a new round of biological assays were performed with 29 structural analogs available on in-house database (Table 3).

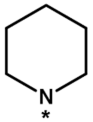
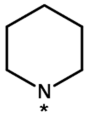
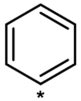
Table 3. *In vitro* cytotoxicity of heteroaryl chalcones against glioma cell lines (C6, U251MG, and U87MG) and primary astrocytes, and thioredoxin reductase activity.



ID	Chemical structure			ML probability		EC_{50} (µM) ± SD				TrxR[a] (%)	Selectivity index[b]		
	R1	R2	R3	C6	BBB	C6	U251MG	U87MG	Astrocytes		C6	U251MG	U87MG
4a	H	Br	A	0.85	0.97	2.7 ± 1.1	21.4 ± 4.4	12.6 ± 3.4	85.4 ± 7.4	−	31.6	3.9	6.7
4b	H	I	A	0.86	0.97	2.1 ± 0.7	12.5 ± 4.1	5.2 ± 2.7	68.2 ± 5.2	−	32.5	5.4	13.1
4c	H	SCH_3	A	0.85	0.91	1.6 ± 1.1	14.2 ± 3.3	3.9 ± 1.7	28.4 ± 3.6	56.4	17.7	2	7.3
4d	H	$C(CH_3)_3$	A	0.84	0.88	28.6 ± 5.7	44.5 ± 6.3	−	−	−	−	−	−
4e	H	$(CH_2)_3CH_3$	A	0.65	0.96	1.7 ± 0.8	3.9 ± 2.1	5.8 ± 1.6	41.2 ± 2.8	80.7	24.2	10.6	7.1
4f	H		A	0.86	0.97	3.5± 1.2	22.6 ± 2.5	13.4 ± 2.9	45.3 ± 4.4	−	12.9	2	3.4

4g	H		A	0.86	0.98	16.4 ± 5.1	35.9 ± 3.3	35.4 ± 7.7	65.8 ± 5.7	–	4	1.8	1.8
4h	H		A	0.76	0.96	17.7 ± 4.5	24.6 ± 4.4	–	–	–	–	–	–
4i	H		A	0.80	0.96	5.1± 1.6	4.4 ± 1.5	9.5 ± 3.3	57.2 ± 5.2	–	11.2	13	6
4j	H		A	0.87	0.97	48.7 ± 6.6	>50	–	–	–	–	–	–
4k	H		A	0.73	0.94	1.5 ± 0.1	16.7 ± 6.4	4.9 ± 2.2	44.3 ± 4.2	0	29.5	2.6	9
4l	H		A	0.92	0.87	6.6 ± 2.1	37.8 ± 3.6	9.8 ± 2.9	75.8 ± 8.5	–	11.5	2	7.7
4m	CH_3	H	A	0.81	0.96	2.1 ± 0.5	2.3 ± 0.3	3.3 ± 0.8	52.2 ± 4.8	95.2	24.8	22.7	15.8
4n	Br	H	A	0.81	0.97	1.4 ± 0.9	5.8 ± 1.3	6.8 ± 2.4	35.6 ± 3.2	78.7	25.4	6.1	5.2
5[a]	H	SCH_3	B	0.84	0.97	30.5 ± 6.8	47.1 ± 6.2	–	–	–	–	–	–
5b	H	CH_3	B	0.92	0.87	7.8 + 3.1	23.3 ± 3.4	19.2 ± 3.5	>100	–	–	–	–

5c	H	SCH_3	B	0.90	0.84	14.4 ± 5.1	24 ± 5.6	45.1 ± 5.2	98.0 ± 8.9	–	–	–	–
5d	H	$C(CH_3)_3$	B	0.89	0.79	18.5± 4.1	16.4 ± 5.1	32.3 ± 4.3	88.7 ± 8.3	3.6	–	–	–
5e	H	$(CH_2)_3CH_3$	B	0.68	0.90	8.4 ± 2.1	6.4 ± 3.3	12.3 ± 3.5	>100	60.4	–	–	–
5f	H		B	0.87	0.88	>50	26.4 ± 3.9	–	–	–	–	–	–
5g	H		B	0.92	0.87	23.6 ± 4.1	>50	–	–	–	–	–	–
5h	H		D	0.92	0.98	14.5 ± 3.3	35.1 ± 5.1	–	–	–	–	–	–
5i	H		D	0.84	0.99	36.6 ± 4.4	>50	–	–	–	–	–	–
13	H		E	0.83	0.87	>50	>50	–	–	–	–	–	–
14	Br	H	E	0.87	0.94	22.3 ± 5.1	32.1 ± 4.2	–	–	–	–	–	–
15	I	H	E	0.86	0.93	14.6 ± 3.5	21.4 ± 3.1	13.4 ± 3.1	>100	–	>6.8	>4.7	>7.5

16	H		F	0.88	0.94	32.5 ± 5.2	20.3 ± 3.6	–	–	–	–	–	–
17	H		G	0.91	0.94	>50	>50	–	–	–	–	–	–
18	H		H	0.86	0.95	49.5 ± 6.7	>50	–	–	–	–	–	–

[a]Percent of inhibition of thioredoxin reductase activity at 100 μM; [b]Selectivity index calculated by astrocyte CC_{50}/glioma lineage EC_{50}; The data are expressed as mean ± SD of three independent assays. Dashed SI values means that cytotoxicity against astrocytes cannot be calculated because compounds did not show significant cytotoxicity even at highest concentrations used in the assay.

The most promising compound was the nitrofuran analog (2E)-1-(3-methylphenyl)-3-(5-nitrofuran-2-yl)prop-2-en-1-one (**4m**) with EC_{50} of 2.1 µM, 2.3 µM and 3.3 µM for C6, U251MG and U87MG cells, respectively. The compounds **4c**, **4e**, **4n**, and **4k** were the most active against C6 cells (EC_{50} ~1.5), but lost activity against U251MG and U87MG cells in comparison to **4m**. The positive control TMZ had activity against C6 cells with an EC_{50} value of 60.46 µM, and showed minimal activity in U251MG and U87MG cells at 200 µM [40], probably because TMZ requires a slightly basic pH to spontaneously converts to a bioactive methylating agent [41]. Concentration-response curves used for EC_{50} determination of **4e**, **4m**, and **4n** are shown in Figure 3A. Considering the lipophilicity (LogP calculate ~3.6 for **4m** and 4**n**, and 4.8 for **4e**) of test compounds, we performed cytotoxicity curves in C6 cells in the presence of varying Triton X-100 concentrations (0.001%, and 0.0001%). At 72 h treatment, **4e, 4m** and **4n** cytotoxicity (EC_{50}) was similar to C6 cells without detergent thereby excluding a possible promiscuous colloidal aggregate effect (Figure 3B) as well as indicating that cytotoxicity of the hit compounds is related to specific inhibition. It is noteworthy that Triton-X100 0.01% killed all C6 cells, thus making the EC_{50} estimation unfeasible at this concentration (data not shown).

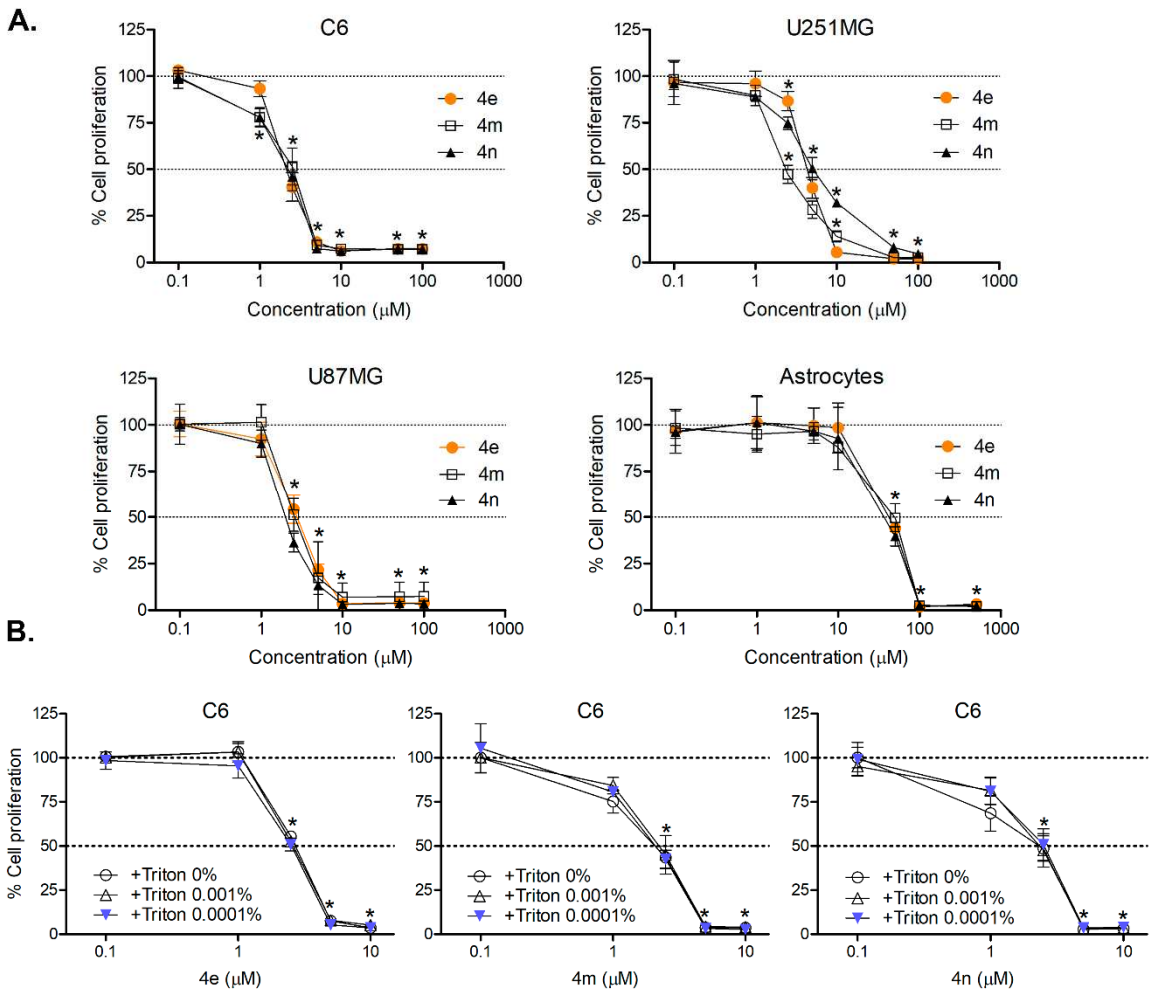


Figure 3. (A) Viability concentration-responses curves for compounds **4e**, **4m**, and **4n** against C6, U251MG, U87MG, and astrocytes after 48h of incubation. (B) Concentration-response curves for test compounds in the presence of Triton X-100 in C6 cells (72 h treatment, sulforhodamine B (SRB) assay). The data are expressed as mean ± SD of three independent assays. * Different from untreated cells ($p<0.05$, ANOVA).

2.5. In vitro Cytotoxicity assays in astrocytes

The hit compounds were also evaluated for their cytotoxicity *in vitro* against astrocytes, as a control for normal glial cells. These compounds demonstrated modest to moderate cytotoxicity on this assay (see Table 3 and Figure 3), with selectivity indexes (SI) ranging between 1.8 and 32.5. The compound **4m** showed the most promising cytotoxicity profile, with SI of 24.8, 22.7 and 15.8 for C6, U251MG and U87MG cells, respectively.

2.6. Inhibition of thioredoxin reductase (TrxR) activity and involvement in thiol homeostasis

Since 5-nitrofuran chalcones display potent antiproliferative activities, we then asked the possible cellular target of these in glioma cells. Chalcones contain an α,β-unsaturated ketone moiety, a key structure for many reported TrxR inhibitors [42,43]. Considering this key feature, we made computational predictions with these molecules using binary ML models developed in-house for TrxR (data not shown). As a result, all investigated compounds were predicted as inhibitors this enzyme. In view of this, the most promising compounds were tested *in vitro* against TrxR. As shown in Table 3, compound **4m** showed the greatest TrxR inhibitory activity (95.2% at 100 µM), followed by **4e** (80.7%) and **4n** (78.7%). In addition, compounds **4e** and **4m** showed dose-dependent decrease of TrxR activity, with IC_{50} values ~25 µM.

Compounds with α,β-unsaturated carbonyl system (i.e. chalcones) have been reported to form covalent bonds with cysteines [44]. Consequently, it has been suggested that chalcones are pan-assay interference compounds (PAINS) [45] due to reactivity under assay conditions. In order to evaluate the role of thiol residues in **4n**, **4m** and **4e** cytotoxicity, we pre-incubated C6 glioma cells with 2 mM N-acetyl-cysteine (NAC) for 1 h prior to test compounds treatments. The classical TrxR inhibitor, auranofin (Au), was used as a positive control. Cell viability assays showed that NAC abrogated the cytotoxicity of our test compounds as well as auranofin (Fig. 4A). Using a cell-free *in vitro* system, we observed that **4n**, **4m** and **4e** (10 to 500 µM) were not alkylating agents in presence of NAC and bovine serum albumin (BSA) at physiological conditions (pH 7.4, 37 ºC) (Figures 4B and 4C). The alkylating agent N-ethylmaleimide (NEM) was used as a positive control for cysteine alkylation, leading to thiol depletion (R-SH) at 50 µM (Figs. 4B and 4C). These results indicate that **4n**, **4m** and **4e** are PAINS even though thiol homeostasis and inhibition of TrxR may play a role, at least in part, in 5-nitrofuran chalcones cytotoxicity.

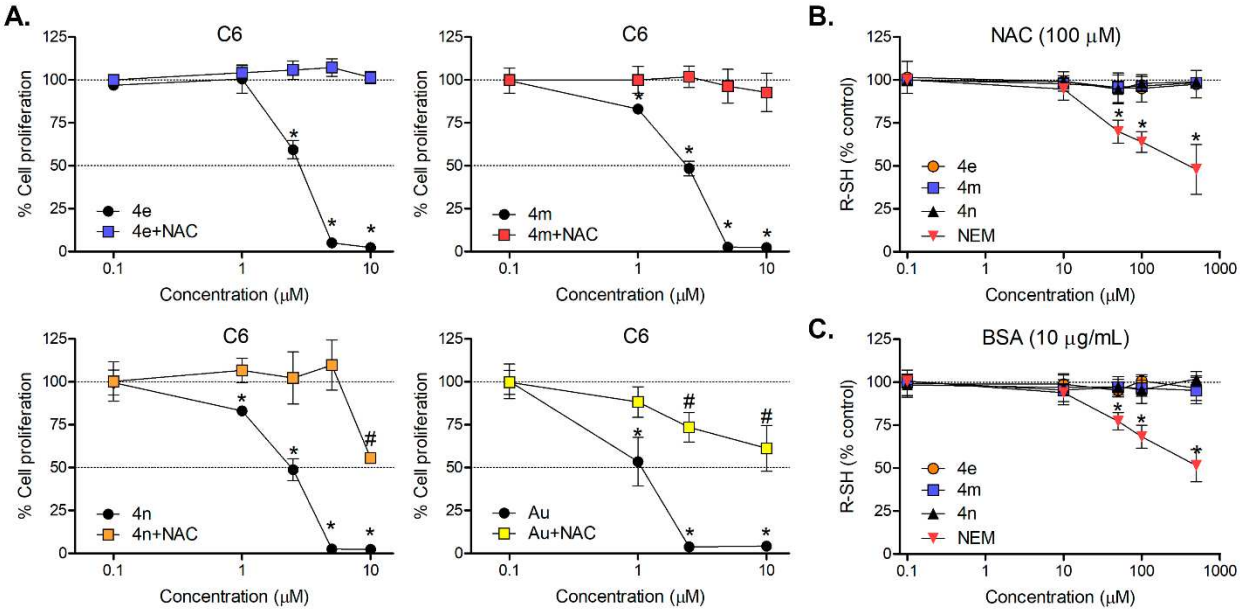


Figure 4. (A) Effect of NAC pretreatment (2 mM, 2 h) on 4e, 4m, 4n and auranofin (Au) cytotoxicity in C6 glioma cells incubated for 72 h as assessed by SRB assay. (B-C) *In vitro* reactivity of 4e, 4m and 4n with reduced thiol residues (R-SH) of NAC and BSA in cell-free assay as determined by 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) assay. In B and C, the thiol-alkylating agent NEM was used as a positive control for thiol alkylation/depletion. *different from untreated cells or control group (NAC or albumin alone, in B and C graphs); #different from untreated and test compound-treated cells at equivalent concentrations (1-way ANOVA, post-hoc Tukey; $p<0.05$).

2.7. Structure-activity relationships (SAR)

Based on the experimental results, we derived structure-activity relationships (SAR) rules to reveal the molecular substituents favorable and unfavorable for anti-glioma activity (Figure 5). The information revealed by the SAR allowed us to derive the following rules: (i) halogen atom and methyl in R1 position increases the activity; (ii) aliphatic six-member rings and hydrophobic groups with primary to tertiary carbons in R2 position increases the activity; (iii) 5-nitrofuran ring in R3 position increases the activity; (iv) aromatic rings or hydrophobic groups with quaternary carbons in R2 position decreases the

activity; (v) 5-nitrothiophene, furan, 5-chlorothiophene, and six-member aromatic rings in the R3 position decreases the activity in glioma cells.

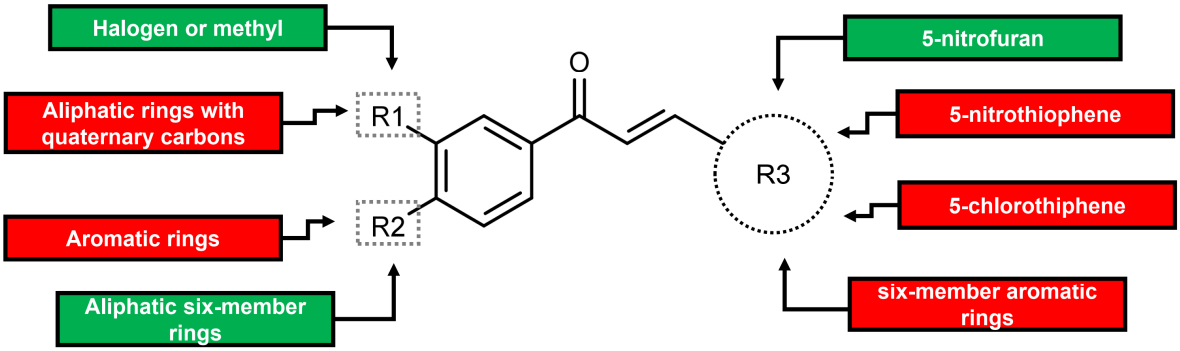
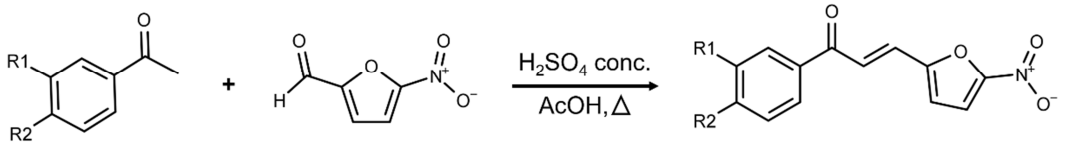


Figure 5. Derived SAR rules for chalcones with antiproliferative activity against glioma cells. Substituents inside green boxes increase the activity, whereas substituents in red boxes decrease the activity.

2.8. Synthesis of lead candidates

The most promising hit compounds (i.e., **4e**, **4m**, and **4n**) identified in antiproliferative assay were selected for *in vivo* testing. Considering the insufficient amount of compound mass for *in vivo* assays, we resynthesized them again in larger scale using Claisen-Schmidt condensation using acetic acid as solvent and sulfuric acid (H_2SO_4) as a catalyst [46]. For the resynthesis of desired chalcones in large quantities required for *in vivo* testing it was necessary an optimization study and improvement of the yields obtained in the methodology previously described by Dr. Andrade´s research group. Gomes and co-workers [46] reported on previous study that standard condition under basic medium could not be used because the starting materials are alkali-sensitive. Thus, the resynthesis of our test chalcones was carried out in acidic medium (Table 4). Table 4 lists the optimized conditions employed for the resynthesis of **4n**, under reflux at 100 ºC, reduced evaporate pressure (as previously reported), isolation with cooled water, H_2SO_4 catalyst reduced, and microwave.

Table 4. Claisen-Schmidt condensation under different conditions.



Entry	Method*	Yield (%)	Reaction time (h)
1	Reflux + reduced evaporate pressure	31%	30
2	Reflux + isolation with cooled water	87%	20
3	Reflux + 0.5 H_2SO_4 catalyst	56 %	30
4	Microwave	37 %	02

*Conditions we employed for the resynthesis of compound 4n.

The protocol used in entry 1 was used in the previous work where after consumption of all starting materials (monitored by TLC), the reaction mixture is concentrated under reduced pressure to remove the acetic acid, thus obtaining the desired chalcone. The problem encountered with the increase in scale is that a drastic decrease in yield is observed because of the higher amount of sulfuric acid causing the degradation of the formed product. As we may notice from Table 4, under reflux at 100 ºC and pressure reduced the reaction pathways had other disadvantages beside low yield (31%), the long reaction time (30h) Under microwave irradiation (entry 4), the reaction time decreased substantially (2h), but yield (37%) remained similar. Also explained by the presence of the acid medium is observed a large amount of degradation of material. The reduction of the amount of sulfuric acid (entry 3) was not satisfactory because there was no total consumption of the starting materials with the same reaction time as the entry 1. The protocol used in entry 2 was the better choice had the higher yield (87%), where after consumption of all starting materials (monitored by TLC), the reaction mixture was poured into cooled water, precipitating the desired chalcone. Taking into consideration the aspects mentioned above, especially those related to yield, we decided to extend the protocol entry 2 for **4e** and **4m** resynthesis.

2.9. In vivo anti-glioma activity

Many prior *in vitro* studies have found potential candidates to treat gliomas, but most of them did not include animal testing or failed in *in vivo* studies. In addition, most of prior studies use

immunocompromised mice [47] and subcutaneous xenografts [48,49], which lack important components of tumor microenvironment and blood barrier restrictions, respectively. In this study, we confirmed the *in vivo* anti-glioma effects of resynthesized compounds (**4e**, **4m**, and **4n)** using a very aggressive model of glioma in immunocompetent mice orthotopically implanted. The overall design of the study, treatment groups, route of injection, and short-term/long-term treatment schedules are described in Figure 6A. Briefly, we first implanted GL261 cells into the right hemisphere of mice brain. Seven days after implantation, vehicle or test compounds were intraperitoneally or orally administrated every day for 10 days. The alkylating agent temozolomide (TMZ) was used as a positive control for anti-glioma activity in mice. During the study, the treated animals did not show any visible toxic effects or mortality and had no significant difference in body weights compared to the negative control group (Figure 6B).

The treatment with compounds by gavage (Figure 6C) does not significantly decrease (p >0.05) the tumor volume compared to negative control group (vehicle consisted of saline 0.9%, DMSO 10%, and Tween-80 0.3%), indicating the poor oral absorption of these molecules. However, the tumor volume of mice treated intraperitonially with **4m** (100 mg/kg) and **4n** (100 mg/kg) significantly reduced by 43.8% and 41.3% (Table S1 and Figure 6D), respectively, compared to vehicle-treated mice. These results indicate that the compounds are able to cross the BBB, confirming our ML predictions. In addition, no significant differences in tumor volume were observed between **4m** (100 mg/kg) and **4n** (100 mg/kg) treatment and the positive control group (TMZ at 20 mg/kg, p > 0.05).

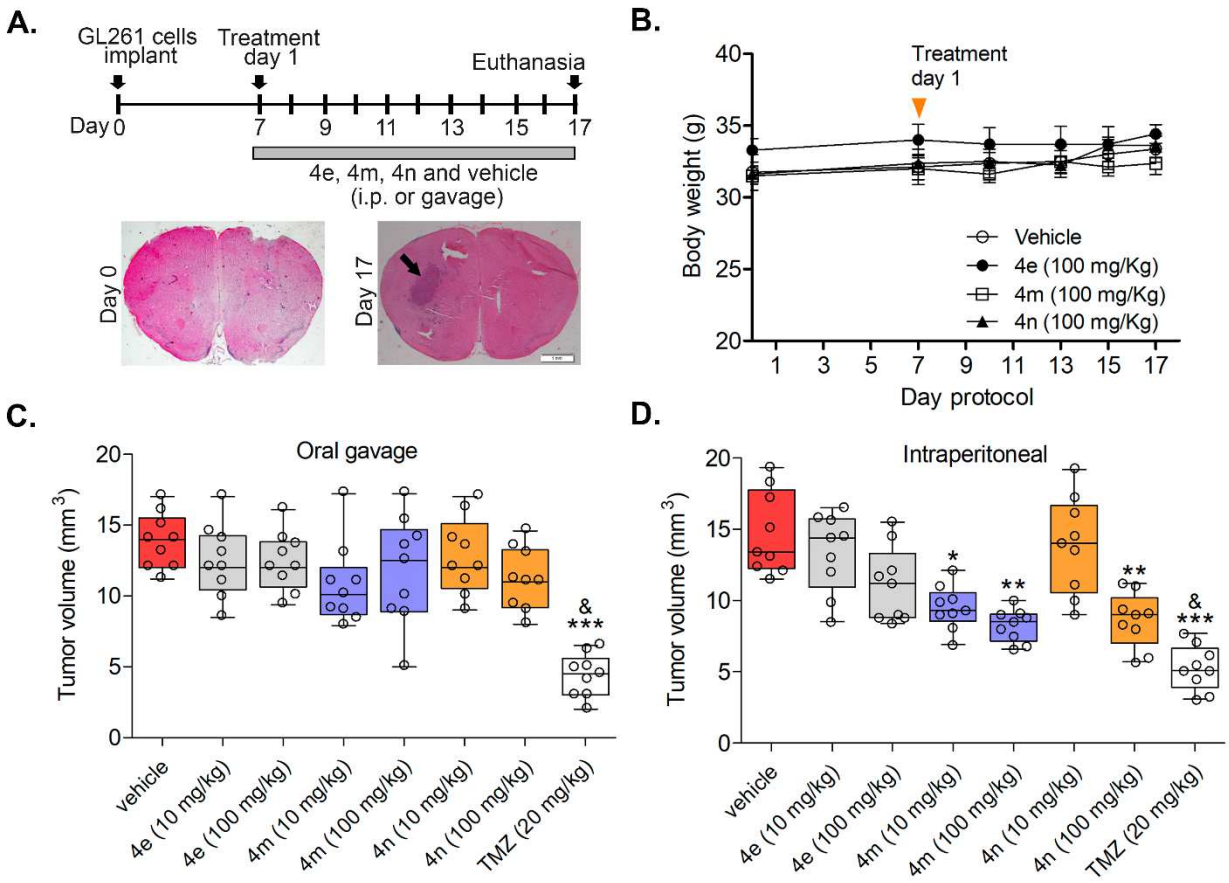


Figure 6. Preclinical testing of **4e**, **4m**, and **4n** in brain implanted GL261 gliomas. (A) Animal study design: from treatment day 1 (7th protocol day), the test compounds were administered once a day for a total 10 days. TMZ was administrated in alternate days (protocol days 7, 9, 11, 13 and 15th. (B) Delta body weight change (final – initial) in C57BL/6 mice across the different treatments; (C) GL261 tumor volumes (mm3) quantification of the mice treated by gavage at the 17th day; (D) GL261 tumor volumes (mm3) quantification of the mice treated by intraperitoneal injection at the 17th day. Asterisks denote differences from vehicle treated mice (*p<0.05; **p<0.01; ***p<0.001); &different from test compounds at 10 mg/kg (1-way ANOVA, post-hoc Tukey).

2.10. Toxicity studies

For toxicity studies, the blood samples were collected at the end of treatment of glioma implanted mice and various enzyme markers and blood parameters were estimated (Table 5). Treatment with **4m**, **4e** and **4n** at 100 mg/kg did not alter the levels of enzymatic markers for hepatotoxicity, i.e. alanine aminotransferase (ALT) and alkaline phosphatase (ALP), and nephrotoxicity (creatinine). The glucose

levels remained unchanged. In addition, these compounds did not promote hematological toxicity in white blood cells counts and hematocrit when compared to vehicle-treated mice. These results are encouraging since TMZ promotes toxicity through the significant reduction in the number of circulating immune cells and increases in the levels of enzyme markers for hepatotoxicity and nephrotoxicity [50].

Table 5. Serum and blood markers of toxicity at the end of treatments in glioma bearing mice.

Parameters	**Concentration ± SD**			
	Untreated	**4e**	**4m**	**4n**
ALT (U/L)	7.6 ± 4.2	19.8 ± 18.6	6.2 ± 3.5	9.9 ± 4.4
ALP (U/L)	19.1 ± 8.9	22.2 ± 8.7	20.3 ± 7.8	20.6 ± 6.1
Creatinine (mg/dL)	0.4 ± 0.3	0.3 ± 0.2	0.4 ± 0.1	0.5 ± 0.4
Glucose (mg/dL)	108 ± 13	94 ± 14	108 ± 16	109 ± 15
WBC ($x10^3$/µL)	8.6 ± 1.6	9.5 ± 1.3	8.4 ± 2.5	9.2 ± 2.6
Hematocrit (%)	44.5 ± 1.8	45.5 ± 2.3	46.0 ± 2.5	45.3 ± 2.9

ALT: alanine aminotransferase; ALP: alkaline phosphatase; WBC: white blood cells.

3. Conclusions

We have developed robust machine learning models for the identification of new compounds able to penetrate BBB and active glioma cells. The ML models were applied for virtual screening of our in-house database of compounds. As a result, forty-one potential anti-glioma hits were prioritized and tested *in vitro* against three glioma cell lines and astrocytes. Among them, compounds **4e**, **4m**, and **4n** were the best candidates from the chalcone series, presenting high potency at submicromolar range (EC_{50} of 1.4–6.8 µM) and moderate cytotoxicity against astrocytes. Then, SAR rules revealed that compounds containing halogen atom and methyl group in R1 position or aliphatic six-member rings and hydrophobic groups with primary to tertiary carbons in R2; and containing 5-nitrofuran ring in R3 position were the most potent. Enzymatic assays indicated that inhibition of TrxR may be at least one of the biological targets of 5-nitrofuran chalcones. In addition, orthogonal *in vitro* assays excluded the

possible promiscuous colloidal aggregate and alkylating effect of test compounds, indicating that the cytotoxic effect of the hit compounds is not related to promiscuous assay-interference. Subsequently, we confirmed the *in vivo* anti-cancer effects of resynthesized compounds (**4e**, **4m**, and **4n**) using mice orthotopic glioma model. The treatment of mice with **4m** and **4n** efficiently decreased glioma growth without promoting body weight reduction, death of animals, or altering hematological and toxicological markers. To summarize, the machine learning models developed in this study allowed us to discover two new lead compounds, which are new chemical scaffolds for developing novel anti-glioma drug candidates.

4. Experimental section

4.1. Computational

4.1.1. Datasets

In this study, a dataset of compounds containing bioactivity data for C6 cell line was extracted from ChEMBL database (https://www.ebi.ac.uk/chembl/; ID: CHEMBL614657) [28], while a dataset of compounds with BBB penetration data was selected from a number of publications [29–32]. A brief description of the datasets is presented below.

- C6 dataset: 376 compounds with EC_{50} data. Based on a threshold of 10 μM, it consisted of 144 active compounds ($EC_{50} \leq 10$ μM) and 232 inactive compounds ($EC_{50} > 10$ μM);
- BBB dataset: 2,053 compounds with LogBB data. Based on a threshold of −1, it consisted of 1,570 BBB+ compounds (if LogBB ≥ −1: penetrate) and 483 BBB− compounds (if LogBB < −1: not penetrate).

4.1.2. Data Curation

All chemical structures and correspondent biological information were carefully standardized using Standardizer v.16.9.5.0 (ChemAxon, Budapest, Hungary, http://www.chemaxon.com) according to the protocols proposed by Fourches and colleagues [51–53]. Briefly, specific chemotypes such as nitro

groups and aromatic rings were normalized. In addition, explicit hydrogens were added, whereas organometallic compounds, mixtures, polymers, and salts were removed. Then, we performed the analysis and exclusion of duplicates as follows: (*i*) if the reported outcomes of the duplicates were the same (e.g. active *vs* active, inactive *vs* inactive, etc.), one entry would be retained in the dataset and the other excluded; and (*ii*) if duplicates presented discordance in biological activity (e.g. active *vs* inactive, BBB+ *vs* BBB-), both entries would be excluded. Consequently, 81 duplicates within the C6 dataset and 102 duplicates within the BBB were identified and removed from original datasets. Furthermore, a high concordance was observed between duplicate records of C6 dataset (82.7%), and BBB dataset (85.3%), revealing the high quality of these datasets.

4.1.3. Dataset Balancing and chemical space analysis

The curated datasets (C6: 97 actives and 173 inactives; BBB: 433 BBB− and 1436 BBB+ compounds) were balanced using a linear under-sampling approach [33]. The linear under-sampling strategy calculates the Euclidean distances between each compound in majority class and whole set of minority class are measured using *k*-nearest neighbor (*k*-NN) algorithm [54]. Then, the samples on majority classes of C6 and BBB datasets were linearly extracted over the whole set by using *k*-distances and used to generate balanced datasets (Supplementary Files S1 and S2, respectively). Finally, a chemical space analysis of balanced datasets was generated combining PCA and MACCS keys and employing the KNIME workspace v.3.2 [55,56].

4.1.4. Molecular fingerprints

Morgan and FeatMorgan fingerprints were calculated in the open-source cheminformatics software RDKit (http://www.rdkit.org [57]) executed on Python v.3.6 (https://www.python.org) [58]. Both fingerprints were generated with radius 2−4 and bit vector of 2,048 bits. Morgan and FeatMorgan are circular fingerprints built by applying the Morgan algorithm to a set of user-supplied 2D chemical structures [59,60]. The fingerprint generation process systematically records the neighborhood of each non-hydrogen atom into multiple circular layers up to a stablished radius. The Morgan captures highly

specific atomic information enabling the representation of a large set of precisely defined structural features [59], whereas FeatMorgan uses functional features (i.e., hydrogen-bond donor and acceptors, aromatic, halogen, basic and acid groups) [61]. Subsequently, these atom-centered substructural features are interpreted as indexes of bits in a huge virtual bit string. Each position in this bit string accounts for the presence or absence of a specific fragment feature [59,60].

4.1.5. Machine learning models

ML models were developed using Random Forest algorithm implemented in Scikit-learn v.0.19.2 (http://scikit-learn.org/) package available on Python v.3.6. The grid search was done using 50−500 estimators (intervals of 25 trees), number of features (Morgan or FeatMorgan bits) ranging from 6.6% to 100% along the bit vector of 2,048 bits (Morgan and FeatMorgan). The Cohen's kappa (κ) was used as scoring function of the estimator.

4.1.6. 5-fold external cross-validation (5FECV)

The full dataset of compounds was randomly divided into five subsets of equal size; then one of these subsets (20% of all compounds) is set aside as an external validation set and the remaining four sets together form the training set (80% of the full set). Models were built using the training set while the compounds in momentary external set (fold) were employed to evaluation of predictive performance. ML models were developed five times, allowing each of the five subsets to be used as a momentary external validation set.

4.1.7. Performance of ML models

The predictive performance of ML models was evaluated using SE, SP, CCR, PPV, NPV and κ. These metrics were calculated as follows:

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad (1)$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \qquad (2)$$

$$\mathrm{CCR} = \frac{\mathrm{SE} + \mathrm{SP}}{2} \qquad (3)$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \qquad (4)$$

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} \qquad (5)$$

Here, TP and TN represent the number of true positives and true negatives, respectively, while FP and FN represent the number of false positives and false negatives, respectively. Statistical parameters higher than 0.65 denote that model is predictive.

In addition to the above model evaluation metrics, κ was used to measure the agreement between model predictions and experimental data [62]. This statistical parameter is calculated by the following equations:

$$\Pr(a) = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{N}} \qquad (6)$$

$$\Pr(\mathrm{e}) = \frac{(\mathrm{TP} + \mathrm{FP})\ \mathrm{x}\ (\mathrm{TP} + \mathrm{FN}) + (\mathrm{TN} + \mathrm{FN})\ \mathrm{x}\ (\mathrm{TN} + \mathrm{FP})}{\mathrm{N}} \qquad (7)$$

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \qquad (8)$$

Here, Pr(a) represents the relative observed agreement between the predicted classification of the model and the known classification, and Pr(e) is the hypothetical probability of chance agreement. In the end, κ analysis returns values between −1.0 (no agreement) and 1.0 (complete agreement), but values between 0.60 and 1.0 denote that the model is predictive.

4.1.8. Applicability domain

The AD was estimated as a distance threshold (D_T) between a compound under prediction and the closest nearest neighbors in training set. The following equation was used for calculation of distance threshold [63]:

$$D_T = \bar{y} + Z\sigma \qquad (9)$$

In which $\bar{y}$ is the average Euclidean distance of the k nearest neighbors within the modeling set, σ is the standard deviation of these Euclidean distances, and Z is an arbitrary parameter to control the

significance level. We set the default value of this parameter Z at 0.5. If the compound distance exceeded the threshold, the prediction was considered to be less trustworthy [64].

4.1.9. Virtual screening

Developed ML models were used for VS of an in-house library of compounds aiming to identify new potential anti-glioma compounds, which could be potentially penetrate BBB. Initially, compounds had their antiproliferative activity and BBB penetration ability predicted by our ML models. Subsequently, compounds were filtered using a aggregator advisor tool to identify molecules that are known-to aggregate in experimental assays [38,39]. Finally, pairwise Tanimoto coefficients between virtual hits were calculated to select a subset of structurally diverse virtual hits. The selected virtual hits were them proceed to *in vitro* experimental evaluation.

4.2. Cell experiments and chemistry

4.3. Reagents

TMZ, propidium iodide, DMEM medium, BSA, dimethyl sulfoxide (DMSO), SRB, DTNB, NAC, 1X antibiotic/antimycotic solution, TrxR assay kit, Au, Triton X-100, and Tween-80 were purchased from Sigma-Aldrich (St. Louis, MO, USA).

4.3.1. Cell cultures

The C6, U87MG and U251MG cell lines were obtained from American Type Culture Collection (ATCC; Rockville, Maryland, USA); GL261 cells were kindly provided by Dr. Braganhol (Universidade Federal de Ciências da Saúde de Porto Alegre, RS, Brazil). Cell lines were used to a maximum of 30 passages. The cells were grown in DMEM supplemented with 10% fetal bovine serum plus 1X antibiotic/antimycotic solution (Sigma-Aldrich) in a humidified incubator at 37 ºC. Glioma cells were treated at a 40-50% confluence. Primary astrocytes were isolated from cortex of 2-days old Wistar rats by mechanical dissociation with Ca^{+2}/Mg^{+2} – free Hank's balanced salt solution and plated in Poly-L-Lysine-coated 96-well plates. Astrocytes were maintained in high-glucose DMEM plus antibiotics in a

humidified incubator, and treated after reaching a 90% confluence (12-15 days) [65]. The test compounds were dissolved in DMSO at 50 to 100 mM concentrations in order to achieve a maximum 0.1% DMSO final concentration in the cell cultures.

4.3.2. In vitro antiproliferative activity

SRB assays were performed to screen cytotoxicity of the test compounds identified herein. Briefly, glioma cells and astrocytes were plated onto 96-well plates, treated with test drugs in the 0.1 to 100 uM range for 72 h. Afterwards the cells were fixed with 100 µL of ice-cold 40% trichloroacetic acid for 1 h at 4°C. Plates were then washed five times with cold water and left to dry. SRB solution (50 µL; 0.4% SRB in 1% acetic acid) was added to each well and incubated for 30 min. The cells were then washed 4 times with 1% acetic acid and air dried. Then, 100 µL of 10 mM Tris-base at pH 10.5 was added to each well to solubilize the dye. The plates were gently shaken for 20 min and the absorbance was read at 510 nm in a microplate reader. Cell numbers were calculated as the percentage absorbance (% cell proliferation) compared to the absorbance of vehicle treated (DMSO) controls. In some assays, NAC (2 mM) pretreatments were performed for 1 h prior to test compounds incubation.

4.3.3. Controls for nonspecific inhibition

Aggregation-based cytotoxicity were evaluated in C6 cells incubated with test compounds in culture medium containing no detergent versus 0.0001%, 0.001% and 0.01% Triton X-100. SBR assays were carried out after 72 h incubation. The fundamental concept is that increasing the amount of detergent will increase the EC_{50} value of an aggregator, and will have no effect on a reversible/competitive inhibitor [66]. For evaluation of alkylating effect of test compounds with cysteine residues, NAC (100 and 500 µM) was incubated with test compounds ranging from 10 to 500 µM concentrations in 10 mM phosphate buffered saline (pH 7.4) for 2 h at 37 C in the dark. In addition, we incubated test compounds with BSA (10 µg/mL) at the same conditions above described [67]. At the end of incubations, reduced thiol residues in NAC and BSA were determined by reaction with DTNB in the presence of boric acid buffer (100 mM boric acid, 0.2 mM EDTA, pH 8.5) for 1 h at room temperature, and then read at 412

nm [68]. The thiol-alkylating agent NEM was used as a positive control for thiol alkylation/depletion. Data are expressed as % R-SH compared to control group (NAC or BSA alone).

4.3.4. Thioredoxin reductase assay

The reducing activity of TrxR (Sigma T9698) on 5,5'-Dithiobis-(2-nitrobenzoic acid) (DTNB, Sigma D8130) was performed on 96-well microplates with a final volume of 100 µL at 37 °C. The reaction medium contained 100 mM potassium phosphate buffer, pH 7.0; 0.2 $mg.mL^{-1}$ of NADPH, 2 $mg.mL^{-1}$ of DTNB. The amount of TrxR used in the assay was 0.01 $mg.mL^{-1}$. The FlexStation 3 multi-mode plate reader (Molecular Devices, CA) was used. The activity was measured following the appearance of 5-thio-2-nitrobenzoic acid (TNB) having an absorbance peak at the wavelength of 412 nm and a molar extinction coefficient of 13.6 $mM^{-1}.cm^{-1}$ [68]. Before starting the reaction reading, the TrxR was incubated for 15 minutes with the test compounds, including auranofin (positive control) at 100 µM or DMSO, used as negative control. The reaction was started with the addition of NADPH. The assays were performed in duplicate and in the absence of the enzyme to discount the contribution of the spontaneous reaction between the substrates. The compounds showing more than 50% of inhibition were further analyzed and had their IC_{50} determined using height dilution series (1 µM – 1000 µM).

4.3.5. Chemical synthesis

Microwave synthesis was performed in microwave Anton Paar, Monowave 300 model. The progress of all reactions was monitored on Silicycle 60 F-254 silica gel plates 0.25 mm using ethyl acetate/n-hexane (v/v) as eluent system. Spots were visualized by irradiation with ultraviolet light (254 nm) or with sulfuric vanillin solution or ninhydrin solution followed by heating. Melting points were determined using Fisatom digital. ^{1}H and ^{13}C NMR spectra were recorded on Bruker of 11.74 Tesla spectrometer at 400,13 or 500,13 MHz for ^{1}H and 125,76 MHz for ^{13}C with spectral large of 10.0 ppm for ^{1}H and 240 ppm for ^{13}C using $CDCl_3$ as solvent and reference. Chemical shifts are given in parts per million (ppm) (δ relative to residual solvent peak for ^{1}H and ^{13}C). Spectra mass was performed on a

Compact Bruker in LCMS-2020. IR spectra were recorded on a Nicolet 6700-FTIR Thermo Scientific model (medium, sweep of 4000 to 400 cm^{-1}).

4.3.5.1. (2E)-1-(4-butylphenyl)-3-(5-nitrofuran-2-yl)prop-2-en-1-one (4e)

Brown solid; yield 79% (807.7 mg, 2.70 mmol); melting point 90º C). ^{1}H NMR ($CDCl_3$, 500 MHz): δ = 7.98 (d, 2H, *J* = 8.2 Hz), 7.76 (d, 1H, *J* = 15.4 Hz), 7.54 (d, 1H, *J* = 15.4 Hz), 7.37 (d, 1H, *J* = 3.7 Hz), 7.33 (d, 2H, *J* = 8.2 Hz), 6.83 (d, 1H, *J* = 3.7 Hz), 2.7 (t, 2H, *J* = 7.7 Hz), 1.6 (q, 2H, *J* = 7.4 Hz; 7,7 Hz), 1.3 (s, 2H, *J* = 7.4Hz), 0,9 (t, 3H, *J* = 7.4 Hz); ^{13}C NMR ($CDCl_3$, 125 MHz): δ = 188.1, 153.2, 149.63, 134.8, 128.9, 128.8, 127.7, 125.2, 116.3, 113.2, 35.7, 33.2, 22.3, 13.9; IR (neat): 3393, 2974, 1777, 1021, 777, 570, 470 cm^{-1}; HRMS-ESI (m/z): calcd for $C_{17}H_{17}NO4$[M+H]+: 299.1158, found: 300.1234.

4.3.5.2. (2E)-1-(3-methylphenyl)-3-(5-nitrofuran-2-yl)prop-2-en-1-one (4m)

Brown solid; yield 84% (430.8mg, 1.67 mmol); melting point 144º C. ^{1}H NMR ($CDCl_3$, 400 MHz): δ = 7.86 (d, 2H, *J* = 6.9 Hz), 7.76 (d, 1H, *J* = 15.5 Hz), 7.54 (d, 1H, *J* = 15.5 Hz), 7.4 (d, 2H, *J* = 7.6 Hz), 7.38 (d, 1H, *J* = 3.8 Hz), 6.84 (d, 1H *J* = 3.8 Hz), 2.46 (s, 3H); ^{13}C NMR ($CDCl_3$, 100 MHz): δ = 188.8, 153.2, 138.8, 137.2, 134.4, 129.1, 128.7, 127.9, 125.9, 125.2, 116.3, 113.1, 21.3; IR (neat): 3393, 2974, 1777, 1021, 777, 570, 470 cm^{-1}; HRMS-ESI (m/z): calcd for $C_{14}H_{11}NO_4$[M+H]+: 257.0688, found: 258.0764.

4.3.5.3. (2E)-1-(3-bromophenyl)-3-(5-nitrofuran-2-yl)prop-2-en-1-one (4n).

Brown solid; yield 87% (148 mg, 0.45 mmol); melting point 146º C. ^{1}H NMR ($CDCl_3$, 500 MHz): δ = 8.17 (t, 1H, *J* = 1.7 Hz), 7.98 (d, 1H, *J* = 7.8 Hz), 7.75 (d, 1H, *J* = 7.8 Hz), 7.68 (d, 1H, *J* = 15.4 Hz), 7.56 (d, 1H, *J* = 15.4 Hz), 7.42 (t, 1H, *J* = 7.8 Hz), 7.39 (d, 1H, *J* = 3.8 Hz), 6.87 (d, 1H, *J* = 3.8 Hz); ^{13}C NMR ($CDCl_3$, 125 MHz): δ = 187.3, 152.7, 138.9, 136.4, 131.6, 130.4, 128.8; IR (neat): 3393, 2974, 1777, 1021, 777, 570, 470 cm^{-1}; HRMS-ESI (m/z): calcd for $C_{13}H_8BrNO_4$[M+H]+: 320.9637, found: 321.9701.

4.3.6. In vivo GL261 glioma growth experiments

GL261 cell culture and tumor implantation: Animal experiments were carried out in accordance with the Animal Research: Reporting of *In Vivo* Experiments (ARRIVE) guidelines, the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978) and institutional guidelines. Mice were obtained from our Institutional Animal Core Facility, and the animal studies were approved by Instructional Animal Care and Use Committee (IACUC, protocol nº 7049120618). The animals were allowed food and water ad libitum. Orthotopic transplantation of GL261 glioma cell line in C57BL/6 mice was carried out as previously described [69,70]. GL261 cells were cultured to subconfluence, trypsinized, washed in DMEM without serum, and resuspended in serum/antibiotics-free DMEM for inoculation. Briefly, GL261 cells ($5x10^4$ in 2 µL DMEM) were injected into the right hemisphere of 60 days-old C57BL/6 mice (males and females, 17-25 g) previously anesthetized by ketamine/xylazine (90/10 mg/Kg, intraperitoneal.) using a 10 µl Hamilton microsyringe coupled with an infusion pump set at 1uL/min (coordinates to the bregma: 2.5 mm posterior, 2.5 mm lateral, and 2.3 mm depth).

Pharmacological treatments and tumor volume quantification: After 7 days for tumor establishment, the animals were grouped (n=9) as follows: vehicle; **4e** (10 mg/Kg and 100 mg/Kg), **4m** (10 mg/Kg and 100 mg/Kg), **4n** (10 mg/Kg and 100 mg/Kg), and TMZ (20 mg/Kg). Test compounds were administered once a day for a total 10 days. TMZ was administrated in alternate days (protocol days 7, 9, 11, 13 and 15th). Oral (200 uL/gavage) and intraperitoneal (100 uL by intraperitoneal injection) administration protocols were carried out for comparison. All test compounds were dissolved in DMSO (10% final), followed by dilution in 0.9% NaCl containing 0.3% Tween-80. Vehicle consisted of 0.9% saline containing 10% DMSO and 0.3% Tween-80. The mice were euthanized after 10 days treatment. The brain was removed, fixed with 10% paraformaldehyde and paraffin embedded. For tumor volume quantification, three H&E-stained coronal sections (5 µM thick) were prepared from each brain/animal. Images were captured and tumor area with each brain slice was estimated using the ImageJ® software. Tumor volume (mm^3) was calculated by sum of the segmented areas as previously described [71].

4.3.7. Toxicological markers

At the end of treatments, the animals were euthanized by ketamine/xylazine (100 mg/Kg, intraperitoneal) followed by cardiac puncture. Whole blood samples were collected in EDTA tubes, and part was harvested without anticoagulants for serum separation by centrifugation (1,300 xg, 10 min), and stored at -80 ºC. Serum activity of alanine aminotransferase (ALT) and alkaline phosphatase (ALP), and glucose and creatinine levels were quantified by Labtest Liquiform commercial kits per manufacturer's instructions (Labtest diagnostica, Brazil). Haemogram analyses were carried out in fresh blood samples collected in EDTA using an ABX Micros 60, HORIBA ABX Diagnostics equipment (Montpellier, France).

4.3.8. Statistical analysis

Data were expressed as average ± SD. Data were analyzed by One-way ANOVA followed by Tukey post-hoc test at a $p<0.05$ cut-off for significance (GraphPad Prism 5).

Abbreviations

AD: applicability domain

ALT: alanine aminotransferase

ALP: alkaline phosphatase

Au: auranofin

BBB: blood-brain barrier

BSA: bovine serum albumin

CCR: correct classification rate

CNS: central nervous system

DMEM: Dulbecco's modified eagle medium

D_T: distance threshold

DTNB: 5,5'-dithiobis-2-nitrobenzoic acid

EC_{50}: half maximal effective concentration

EDTA: ethylenediamine tetraacetic acid

FP: false positives

FN: false negatives

GBM: glioblastoma multiforme

H_2SO_4: sulfuric acid

k-NN: *k*-nearest neighbor

ML: machine learning

NAC: N-acetyl-cysteine

NPV: negative predictive value

NEM: N-ethylmaleimide

NMR: nuclear magnetic resonance

LCMS: liquid chromatography–mass spectrometry

PAINS: pan-assay interference compounds

PCA: principal component analysis

PCs: principal components

PPV: positive predictive value

SD: standard deviation

SE: sensitivity

SI: selectivity index

SP: specificity

SRB: sulforhodamine B

R-OH: thiol

TP: true positives

TN: true negatives

VS: virtual screening

κ: Cohen's kappa

Funding sources

This work has been funded by the National Counsel of Technological and Scientific Development (CNPq), the State of Goiás Research Foundation (FAPEG) and the State of Rio de Janeiro Research Foundation (FAPERJ). C.H.A., A.Z.F., F.P.S-Jr and S.B.F. are CNPq productivity fellows. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Notes

The authors declare no competing financial interest.

Acknowledgment

The authors would like to thank Brazilian funding agencies, CNPq, CAPES, FAPEG and FAPERJ for financial support and fellowships. We are also grateful to ChemAxon Inc. for providing academic license of their software.

References

[1] T.F. Cloughesy, W.K. Cavenee, P.S. Mischel, Glioblastoma: From Molecular Pathology to Targeted Treatment, Annu. Rev. Pathol. Mech. Dis. 9 (2014) 1–25. doi:10.1146/annurev-pathol-011110-130324.

[2] K. Aldape, G. Zadeh, S. Mansouri, G. Reifenberger, A. von Deimling, Glioblastoma: pathology, molecular mechanisms and markers, Acta Neuropathol. 129 (2015) 829–848. doi:10.1007/s00401-015-1432-1.

[3] M. Davis, Glioblastoma: Overview of Disease and Treatment, Clin. J. Oncol. Nurs. 20 (2016) S2–S8. doi:10.1188/16.CJON.S1.2-8.

[4] M. Weller, W. Wick, K. Aldape, M. Brada, M. Berger, S.M. Pfister, R. Nishikawa, M. Rosenthal, P.Y. Wen, R. Stupp, G. Reifenberger, Glioma., Nat. Rev. Dis. Prim. 1 (2015) 15017. doi:10.1038/nrdp.2015.17.

[5] J.P. Thakkar, T.A. Dolecek, C. Horbinski, Q.T. Ostrom, D.D. Lightner, J.S. Barnholtz-Sloan, J.L.

Villano, Epidemiologic and Molecular Prognostic Review of Glioblastoma, Cancer Epidemiol. Biomarkers Prev. 23 (2014) 1985–1996. doi:10.1158/1055-9965.EPI-14-0275.

[6] B.M. Alexander, T.F. Cloughesy, Adult Glioblastoma, J. Clin. Oncol. 35 (2017) 2402–2409. doi:10.1200/JCO.2017.73.0119.

[7] A. Omuro, Glioblastoma and Other Malignant Gliomas, JAMA. 310 (2013) 1842. doi:10.1001/jama.2013.280319.

[8] S. V. Bhujbal, P. de Vos, S.P. Niclou, Drug and cell encapsulation: Alternative delivery options for the treatment of malignant brain tumors, Adv. Drug Deliv. Rev. 67–68 (2014) 142–153. doi:10.1016/j.addr.2014.01.010.

[9] C. Shen, X. Wang, Z. Zheng, C. Gao, X. Chen, S. Zhao, Z. Dai, Doxorubicin and indocyanine green loaded superparamagnetic iron oxide nanoparticles with PEGylated phospholipid coating for magnetic resonance with fluorescence imaging and chemotherapy of glioma, Int. J. Nanomedicine. Volume 14 (2018) 101–117. doi:10.2147/IJN.S173954.

[10] H.S. Friedman, T. Kerby, H. Calvert, Temozolomide and treatment of malignant glioma, Clin. Cancer Res. 6 (2000) 2585–2597.

[11] R. Stupp, W.P. Mason, M.J. van den Bent, M. Weller, B. Fisher, M.J.B. Taphoorn, K. Belanger, A.A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R.C. Janzer, S.K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J.G. Cairncross, E. Eisenhauer, R.O. Mirimanoff, Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma, N. Engl. J. Med. 352 (2005) 987–996. doi:10.1056/NEJMoa043330.

[12] J. Wang, K. Sai, F. Chen, Z. Chen, miR-181b modulates glioma cell sensitivity to temozolomide by targeting MEK1, Cancer Chemother. Pharmacol. 72 (2013) 147–158. doi:10.1007/s00280-013-2180-3.

[13] Q. Pan, X. Yang, H. Wang, X. Dong, W. Wang, Y. LI, J. LI, Chemoresistance to Temozolomide in Human Glioma Cell Line U251 is Associated with Increased Activity of O 6-methylguanine-DNA Methyltransferase and Can be Overcome by Metronomic Temozolomide Regimen, Cell

Biochem. Biophys. 62 (2012) 185–191. doi:10.1007/s12013-011-9280-7.

[14] B.J. Neves, R.C. Braga, C.C. Melo-Filho, J.T. Moreira-Filho, E.N. Muratov, C.H. Andrade, QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery, Front. Pharmacol. 9 (2018) 1–7. doi:10.3389/fphar.2018.01275.

[15] 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. doi:10.1021/jm4004285.

[16] S. Ekins, A.C. Puhl, K.M. Zorn, T.R. Lane, D.P. Russo, J.J. Klein, A.J. Hickey, A.M. Clark, Exploiting machine learning for end-to-end drug discovery and development, Nat. Mater. 18 (2019) 435–441. doi:10.1038/s41563-019-0338-z.

[17] K. Heikamp, J. Bajorath, Support vector machines for drug discovery, Expert Opin. Drug Discov. 9 (2014) 93–104. doi:10.1517/17460441.2014.866943.

[18] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R.P. Sheridan, B.P. Feuston, Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling, J. Chem. Inf. Comput. Sci. 43 (2003) 1947–1958. doi:10.1021/ci034160g.

[19] V.M. Alves, E. Muratov, D. Fourches, J. Strickland, N. Kleinstreuer, C.H. Andrade, A. Tropsha, Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify potentially hazardous compounds, Toxicol. Appl. Pharmacol. 284 (2015) 262–272. doi:10.1016/j.taap.2014.12.014.

[20] V.M. Alves, E. Muratov, D. Fourches, J. Strickland, N. Kleinstreuer, C.H. Andrade, A. Tropsha, Predicting chemically-induced skin reactions. Part II: QSAR models of skin permeability and the relationships between skin permeability and skin sensitization, Toxicol Appl Pharmacol. 284 (2015) 273–280. doi:10.1016/j.taap.2014.12.013.

[21] Y. Uesawa, Quantitative structure–activity relationship analysis using deep learning based on a

novel molecular image input technique, Bioorg. Med. Chem. Lett. 28 (2018) 3400–3403. doi:10.1016/j.bmcl.2018.08.032.

[22] I.I. Baskin, D. Winkler, I. V. Tetko, A renaissance of neural networks in drug discovery, Expert Opin. Drug Discov. 11 (2016) 785–795. doi:10.1080/17460441.2016.1201262.

[23] A. Korotcov, V. Tkachenko, D.P. Russo, S. Ekins, Comparison of Deep Learning With Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets, Mol. Pharm. 14 (2017) 4462–4475. doi:10.1021/acs.molpharmaceut.7b00578.

[24] F. Ghasemi, A. Mehridehnavi, A. Pérez-Garrido, H. Pérez-Sánchez, Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks, Drug Discov. Today. 23 (2018) 1784–1790. doi:10.1016/j.drudis.2018.06.016.

[25] D. Dobchev, G. Pillai, M. Karelson, In Silico Machine Learning Methods in Drug Development, Curr. Top. Med. Chem. 14 (2014) 1913–1922. doi:10.2174/1568026614666140929124203.

[26] A.N. Lima, E.A. Philot, G.H.G. Trossini, L.P.B. Scott, V.G. Maltarollo, K.M. Honorio, Use of machine learning approaches for novel drug discovery, Expert Opin. Drug Discov. 11 (2016) 225–239. doi:10.1517/17460441.2016.1146250.

[27] Y.-C. Lo, S.E. Rensi, W. Torng, R.B. Altman, Machine learning in chemoinformatics and drug discovery, Drug Discov. Today. 23 (2018) 1538–1546. doi:10.1016/j.drudis.2018.05.010.

[28] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J.P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery., Nucleic Acids Res. 40 (2012) D1100–D1107. doi:10.1093/nar/gkr777.

[29] L. Zhang, H. Zhu, T.I. Oprea, A. Golbraikh, A. Tropsha, QSAR Modeling of the Blood–Brain Barrier Permeability for Diverse Organic Compounds, Pharm. Res. 25 (2008) 1902–1914. doi:10.1007/s11095-008-9609-0.

[30] Y.H. Zhao, M.H. Abraham, A. Ibrahim, P. V. Fish, S. Cole, M.L. Lewis, M.J. de Groot, D.P. Reynolds, Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes, J. Chem. Inf. Model. 47 (2007) 170–175. doi:10.1021/ci600312d.

[31] H. Li, C.W. Yap, C.Y. Ung, Y. Xue, Z.W. Cao, Y.Z. Chen, Effect of Selection of Molecular Descriptors on the Prediction of Blood−Brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods, J. Chem. Inf. Model. 45 (2005) 1376–1384. doi:10.1021/ci050135u.

[32] I.F. Martins, A.L. Teixeira, L. Pinheiro, A.O. Falcao, A Bayesian Approach to in Silico Blood-Brain Barrier Penetration Modeling, J. Chem. Inf. Model. 52 (2012) 1686–1697. doi:10.1021/ci300124c.

[33] B.J. Neves, R.F. Dantas, M.R. Senger, C.C. Melo-Filho, W.C.G. Valente, A.C.M. de Almeida, J.M. Rezende-Neto, E.F.C. Lima, R. Paveley, N. Furnham, E. Muratov, L. Kamentsky, A.E. Carpenter, R.C. Braga, F.P. Silva-Junior, C.H. Andrade, Discovery of New Anti-Schistosomal Hits by Integration of QSAR-Based Virtual Screening and High Content Screening, J. Med. Chem. 59 (2016) 7075–7088. doi:10.1021/acs.jmedchem.5b02038.

[34] H. Lachance, S. Wetzel, K. Kumar, H. Waldmann, Charting, navigating, and populating natural product chemical space for drug discovery., J. Med. Chem. 55 (2012) 5989–6001.

[35] R.S. Bon, H. Waldmann, Bioactivity-guided navigation of chemical space., Acc. Chem. Res. 43 (2010) 1103–14. doi:10.1021/ar100014h.

[36] D. Gadaleta, G.F. Mangiatordi, M. Catto, A. Carotti, O. Nicolotti, Applicability Domain for QSAR Models, Int. J. Quant. Struct. Relationships. 1 (2016) 45–63. doi:10.4018/IJQSPR.2016010102.

[37] T.I. Netzeva, A. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D. Roberts, T. Schultz, D.W. Stanton, J.J.M. van de Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52., Altern. Lab. Anim. 33 (2005) 155–73.

[38] J.J. Irwin, D. Duan, H. Torosyan, A.K. Doak, K.T. Ziebart, T. Sterling, G. Tumanian, B.K. Shoichet, An Aggregation Advisor for Ligand Discovery., J. Med. Chem. 58 (2015) 7076–87.

doi:10.1021/acs.jmedchem.5b01105.

[39] S.C. Owen, A.K. Doak, P. Wassam, M.S. Shoichet, B.K. Shoichet, Colloidal Aggregation Affects the Efficacy of Anticancer Drugs in Cell Culture, ACS Chem. Biol. 7 (2012) 1429–1435.

[40] X. Ye, S. Yu, Y. Liang, H. Huang, X.-Y. Lian, Z. Zhang, Bioactive triterpenoid saponins and phenolic compounds against glioma cells, Bioorg. Med. Chem. Lett. 24 (2014) 5157–5163. doi:10.1016/j.bmcl.2014.09.087.

[41] G. Koukourakis, V. Kouloulias, G. Zacharias, C. Papadimitriou, P. Pantelakos, G. Maravelis, A. Fotineas, I. Beli, D. Chaldeopoulos, J. Kouvaris, Temozolomide with Radiation Therapy in High Grade Brain Gliomas: Pharmaceuticals Considerations and Efficacy;A Review Article, Molecules. 14 (2009) 1561–1577. doi:10.3390/molecules14041561.

[42] B. Zhang, D. Duan, C. Ge, J. Yao, Y. Liu, X. Li, J. Fang, Synthesis of Xanthohumol Analogues and Discovery of Potent Thioredoxin Reductase Inhibitor as Potential Anticancer Agent, J. Med. Chem. 58 (2015) 1795–1805. doi:10.1021/jm5016507.

[43] H.-L. Ng, X. Ma, E.-H. Chew, W.-K. Chui, Design, Synthesis, and Biological Evaluation of Coupled Bioactive Scaffolds as Potential Anticancer Agents for Dual Targeting of Dihydrofolate Reductase and Thioredoxin Reductase, J. Med. Chem. 60 (2017) 1734–1745. doi:10.1021/acs.jmedchem.6b01253.

[44] P.A. Jackson, J.C. Widen, D.A. Harki, K.M. Brummond, Covalent Modifiers: A Chemical Perspective on the Reactivity of α,β-Unsaturated Carbonyls with Thiols via Hetero-Michael Addition Reactions, J. Med. Chem. 60 (2017) 839–885. doi:10.1021/acs.jmedchem.6b00788.

[45] J.B. Baell, G. a. Holloway, New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays, J. Med. Chem. 53 (2010) 2719–2740. doi:10.1021/jm901137j.

[46] M.N. Gomes, R.C. Braga, E.M. Grzelak, B.J. Neves, E. Muratov, R. Ma, L.L. Klein, S. Cho, G.R. Oliveira, S.G. Franzblau, C.H. Andrade, QSAR-driven design, synthesis and discovery of potent chalcone derivatives with antitubercular activity, Eur. J. Med. Chem. 137 (2017) 126–138.

doi:10.1016/j.ejmech.2017.05.026.

[47] V. Mégalizzi, C. Decaestecker, O. Debeir, S. Spiegl-Kreinecker, W. Berger, F. Lefranc, R.E. Kast, R. Kiss, Screening of anti-glioma effects induced by sigma-1 receptor ligands: Potential new use for old anti-psychiatric medicines, Eur. J. Cancer. 45 (2009) 2893–2905. doi:10.1016/j.ejca.2009.07.011.

[48] G. Gravina, A. Mancini, A. Colapietro, S. Delle Monache, R. Sferra, F. Vitale, L. Cristiano, S. Martellucci, F. Marampon, V. Mattei, F. Beirinckx, P. Pujuguet, L. Saniere, G. Lorenzon, E. van der Aar, C. Festuccia, The Small Molecule Ephrin Receptor Inhibitor, GLPG1790, Reduces Renewal Capabilities of Cancer Stem Cells, Showing Anti-Tumour Efficacy on Preclinical Glioblastoma Models, Cancers (Basel). 11 (2019) 359. doi:10.3390/cancers11030359.

[49] W.-S. Wu, C.-C. Chien, K.-H. Liu, Y.-C. Chen, W.-T. Chiu, Evodiamine Prevents Glioma Growth, Induces Glioblastoma Cell Apoptosis and Cell Cycle Arrest through JNK Activation, Am. J. Chin. Med. 45 (2017) 879–899. doi:10.1142/S0192415X17500471.

[50] S. Kumari, S.M. Ahsan, J.M. Kumar, A.K. Kondapi, N.M. Rao, Overcoming blood brain barrier with a dual purpose Temozolomide loaded Lactoferrin nanoparticles for combating glioma (SERP-17-12433), Sci. Rep. 7 (2017) 6602. doi:10.1038/s41598-017-06888-4.

[51] 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) 1189–204. doi:10.1021/ci100176x.

[52] D. Fourches, E. Muratov, A. Tropsha, Curation of chemogenomics data., Nat. Chem. Biol. 11 (2015) 535–535. doi:10.1038/nchembio.1881.

[53] 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. doi:10.1021/acs.jcim.6b00129.

[54] N. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat. 46 (1992) 175–185. doi:10.1080/00031305.1992.10475879.

[55] S. Beisken, T. Meinl, B. Wiswedel, L.F. de Figueiredo, M. Berthold, C. Steinbeck, KNIME-

CDK: Workflow-driven cheminformatics, BMC Bioinformatics. 14 (2013) 257. doi:10.1186/1471-2105-14-257.

[56] A. Fillbrunn, C. Dietz, J. Pfeuffer, R. Rahn, G.A. Landrum, M.R. Berthold, KNIME for reproducible cross-domain analysis of life science data, J. Biotechnol. 261 (2017) 149–156. doi:10.1016/j.jbiotec.2017.07.028.

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

[58] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32. doi:10.1023/A:1010933404324.

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

[60] 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. doi:10.1021/c160017a018.

[61] Gobbi, Poppinger, Genetic optimization of combinatorial libraries, Biotechnol. Bioeng. 61 (1998) 47–54. doi:10099495.

[62] J. Cohen, A coefficient of agreement of nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46. doi:10.1177/001316446002000104.

[63] A. Golbraikh, M. Shen, Z. Xiao, Y.-D. Xiao, K.-H. Lee, A. Tropsha, Rational selection of training and test sets for the development of validated QSAR models., J. Comput. Aided. Mol. Des. 17 (2003) 241–253. doi:10.1023/A:1025386326946.

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

[65] A. Zanotto-Filho, E. Braganhol, R. Schröder, L.H.T. de Souza, R.J.S. Dalmolin, M.A.B. Pasquali, D.P. Gelain, A.M.O. Battastini, J.C.F. Moreira, NFκB inhibitors induce cell death in glioblastomas, Biochem. Pharmacol. 81 (2011) 412–424. doi:10.1016/j.bcp.2010.10.014.

[66] B.Y. Feng, B.K. Shoichet, A detergent-based assay for the detection of promiscuous inhibitors,

Nat. Protoc. 1 (2006) 550–553. doi:10.1038/nprot.2006.77.

[67] J.L. Dahlin, J.W.M. Nissink, J.M. Strasser, S. Francis, L. Higgins, H. Zhou, Z. Zhang, M.A. Walters, PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS, J. Med. Chem. 58 (2015) 2091–2113. doi:10.1021/jm5019093.

[68] G.L. Ellman, Tissue sulfhydryl groups, Arch. Biochem. Biophys. 82 (1959) 70–77. doi:10.1016/0003-9861(59)90090-6.

[69] E.W. Newcomb, S. Demaria, Y. Lukyanov, Y. Shao, T. Schnee, N. Kawashima, L. Lan, J.K. Dewyngaert, D. Zagzag, W.H. McBride, S.C. Formenti, The Combination of Ionizing Radiation and Peripheral Vaccination Produces Long-term Survival of Mice Bearing Established Invasive GL261 Gliomas, Clin. Cancer Res. 12 (2006) 4730–4737. doi:10.1158/1078-0432.CCR-06-0593.

[70] T. Szatmari, K. Lumniczky, S. Desaknai, S. Trajcevski, E.J. Hidvegi, H. Hamada, G. Safrany, Detailed characterization of the mouse glioma 261 tumor model for experimental glioblastoma therapy, Cancer Sci. 97 (2006) 546–553. doi:10.1111/j.1349-7006.2006.00208.x.

[71] A. Zanotto-Filho, E. Braganhol, M.I. Edelweiss, G.A. Behr, R. Zanin, R. Schroder, A. Simoes-Pires, A.M.O. Battastini, J.C.F. Moreira, The curry spice curcumin selectively inhibits cancer cells growth in vitro and in preclinical model of glioblastoma., J. Nutr. Biochem. 23 (2012) 591–601. doi:10.1016/j.jnutbio.2011.02.015.

Highlights

- ML models were developed to predict of anti-glioma activity and BBB penetration
- New hits with antiproliferative activity were identified by virtual screening
- Three hits presented high potency and moderate cytotoxicity
- Compounds were able to inhibit TrxR enzyme
- Two lead compounds stopped the malignant glioma *in vivo* without promoting toxicity