Chemia Naissensis, Vol 7, Issue 1, RESEARCH ARTICLE, 40-58, https://doi.org/10.46793/ChemN7.1.40I

Do machine learning and molecular dynamics reveal key insights into GABA-sulfonamide conjugates as carbonic anhydrase inhibitors?

Budimir S. Ilić[1*]

1- University of Niš, Faculty of Medicine, Department of Chemistry, 18000 Niš, Serbia

Budimir S. Ilić: budimir.ilic@medfak.ni.ac.rs, https://orcid.org/0000-0002-2808-3501

ABSTRACT

Carbonic anhydrase (CA) enzymes are critical to numerous physiological processes, making them valuable therapeutic targets. Aromatic and heterocyclic sulfonamides have demonstrated exceptional inhibitory activity, with significant applications in managing glaucoma, a complex and progressive neurodegenerative condition. This study employs an integrative approach combining machine learning, specifically Multiple Linear Regression (MLR) modeling, with molecular dynamics simulations to investigate a series of γ-aminobutyric acid (GABA)-conjugated sulfonamides. The MLR model effectively identified key structural and physicochemical features governing inhibitory activity against carbonic anhydrase isoforms II and IV, enabling precise predictions of biological efficacy. Molecular dynamics simulations were conducted exclusively on the most active GABA conjugate identified, in complex with CA II and CA IV enzymes. These simulations revealed atomistic details of enzyme-ligand interactions, highlighting critical binding interactions, dynamic stability, and conformational behavior driving potent inhibitory effects. By integrating machine learning techniques and targeted molecular dynamics simulations, this study not only deepens our understanding of sulfonamide activity but also provides a robust foundation for the rational design of next-generation inhibitors with enhanced therapeutic potential against glaucoma.

Keywords: Machine learning, Molecular dynamics, GABA, Sulfonamides, Carbonic anhydrase

* Corresponding author: budimir.ilic@medfak.ni.ac.rs

Introduction

Carbonic anhydrases (CAs) are a family of metalloenzymes that catalyse the reversible hydration of carbon dioxide to bicarbonate and protons, a reaction fundamental to various physiological processes, including respiration, acid-base balance, and electrolyte secretion (Supuran & Scozzafava, 2007). In humans, 15 CA isoforms have been identified, each exhibiting distinct tissue distributions and physiological roles (Supuran & Scozzafava, 2007; Jaitak et al., 2024). Notably, aberrant CA activity is implicated in several pathological conditions, such as glaucoma, epilepsy, and certain cancers, rendering CAs significant therapeutic targets (Supuran, 2021; Naeem et al., 2024). Glaucoma, a neurodegenerative disorder and leading cause of irreversible blindness, is managed by lowering intraocular pressure, with CA inhibitors playing a key role in reducing aqueous humor production (Nocentini & Supuran, 2019). The inhibition of CA II and CA IV isoforms is critical for glaucoma treatment, as supported by the development of novel dual-tail sulfonamide inhibitors that exhibit effective and sustained intraocular pressure reduction (Angeli et al., 2024).

Sulfonamides, particularly aromatic and heterocyclic variants, have been extensively studied as CA inhibitors (Angeli et al., 2023). Inhibition of CAs by sulfonamides occurs through multiple mechanisms, with primary action involving direct binding to the active site zinc ion (Supuran, 2016a; Supuran, 2016b). Furthermore, modifications to their scaffolds with functional groups enable interactions with residues near the active site entrance, facilitating enhanced isoform selectivity and diverse inhibition profiles (Nocentini & Supuran, 2019). The incorporation of γ-aminobutyric acid (GABA) moieties into sulfonamide structures represents a significant advancement in CA inhibitor design (Mincione et al., 1999). This study has shown that GABA-sulfonamide conjugates exhibit different inhibition constants (Ki) for CA II and CA IV, underscoring their potential for isoform-specific inhibition.

Advancements in computational methodologies have revolutionised drug discovery processes (Xu, 2024). Machine learning (ML) techniques, including Multiple Linear Regression (MLR) modeling, are extensively utilised to predict biological activity from chemical structures, thereby facilitating the identification of potential drug candidates (Rodríguez-Pérez & Bajorath, 2021). For example, ML models have demonstrated considerable efficacy in accurately predicting the activity and selectivity profiles of human CA inhibitors, significantly enhancing the efficiency and precision of the drug development pipeline (Tinivella et al., 2021). Complementary to ML approaches, molecular dynamics (MD) simulations provide atomistic insights into the dynamic behaviour of enzyme-inhibitor complexes (Wei & McCammon, 2024). These simulations elucidate the conformational flexibility and stability of such complexes, offering a deeper understanding of binding interactions and inhibition mechanisms (Ilić, 2020; Ilić et al., 2021). Furthermore, a recent study elegantly highlights the pivotal role of molecular dynamics in providing deeper insights into the interactions between CA and sulfonamide inhibitor (Pagnozzi et al., 2022). Integrating ML and MD techniques enables a comprehensive analysis of potential

inhibitors by combining predictive modeling with detailed mechanistic insights (Frasnetti et al., 2024).

This study aims to investigate the inhibitory activity of a series of GABA-sulfonamide conjugates against CA II and CA IV isoforms, leveraging an integrative approach that combines ML-based MLR modeling and MD simulations. By identifying key structural and physicochemical parameters governing inhibitory activity, the research seeks to elucidate the molecular determinants underlying isoform selectivity and binding affinity. These findings are expected to provide a robust foundation for the rational design of next-generation CA inhibitors with enhanced therapeutic efficacy, particularly for the treatment of glaucoma.

Experimental

Machine learning

A dataset comprising 52 sulfonamides featuring the GABA moiety (Figure 1), along with their inhibitory activity values against the CA II and CA IV isoforms, was obtained from the study by Mincione et al. (1999). The dataset was processed using the DTC Lab Tools software, developed by the Drug Theoretics and Cheminformatics Laboratory at Jadavpur University, Kolkata, following the methodology described by Banerjee & Roy (2023). For modeling purposes, the compounds were divided into a training set (41 compounds) and a test set (11 compounds), ensuring a robust dataset partitioning for ML-based MLR modeling. The inhibitory activity values for the CA II and CA IV isoforms, originally reported as Ki values (nM), were converted to pKi values to ensure methodological compatibility with the MLR modeling process. The partitioning of the dataset, along with the pKi values for the training and test sets, is summarized in Table 1.

To develop the MLR model, 484 two-dimensional (2D) molecular descriptors were computed using E-Dragon, the remote version of Dragon, developed by the Milano Chemometrics and QSAR Research Group (Tetko et al., 2005). These descriptors encompassed a wide range of chemical and structural properties, including topological indices, walk and path counts, connectivity indices, functional group counts, and atom-type E-state indices (Todeschini & Consonni, 2009). Subsequently, a rigorous data pretreatment procedure was applied to refine the descriptor dataset. Specifically, descriptors with a standard deviation below 0.0001 were excluded as they lack significant variation across the dataset, rendering them uninformative and unnecessarily increasing computational complexity. Additionally, descriptors with pairwise correlation coefficients exceeding 0.95 were removed to eliminate multicollinearity, as such redundancy can lead to overfitting and compromise the interpretability and reliability of the MLR model. This preprocessing step reduced the initial descriptor pool to 87 variables deemed suitable for inclusion in the MLR modeling process.

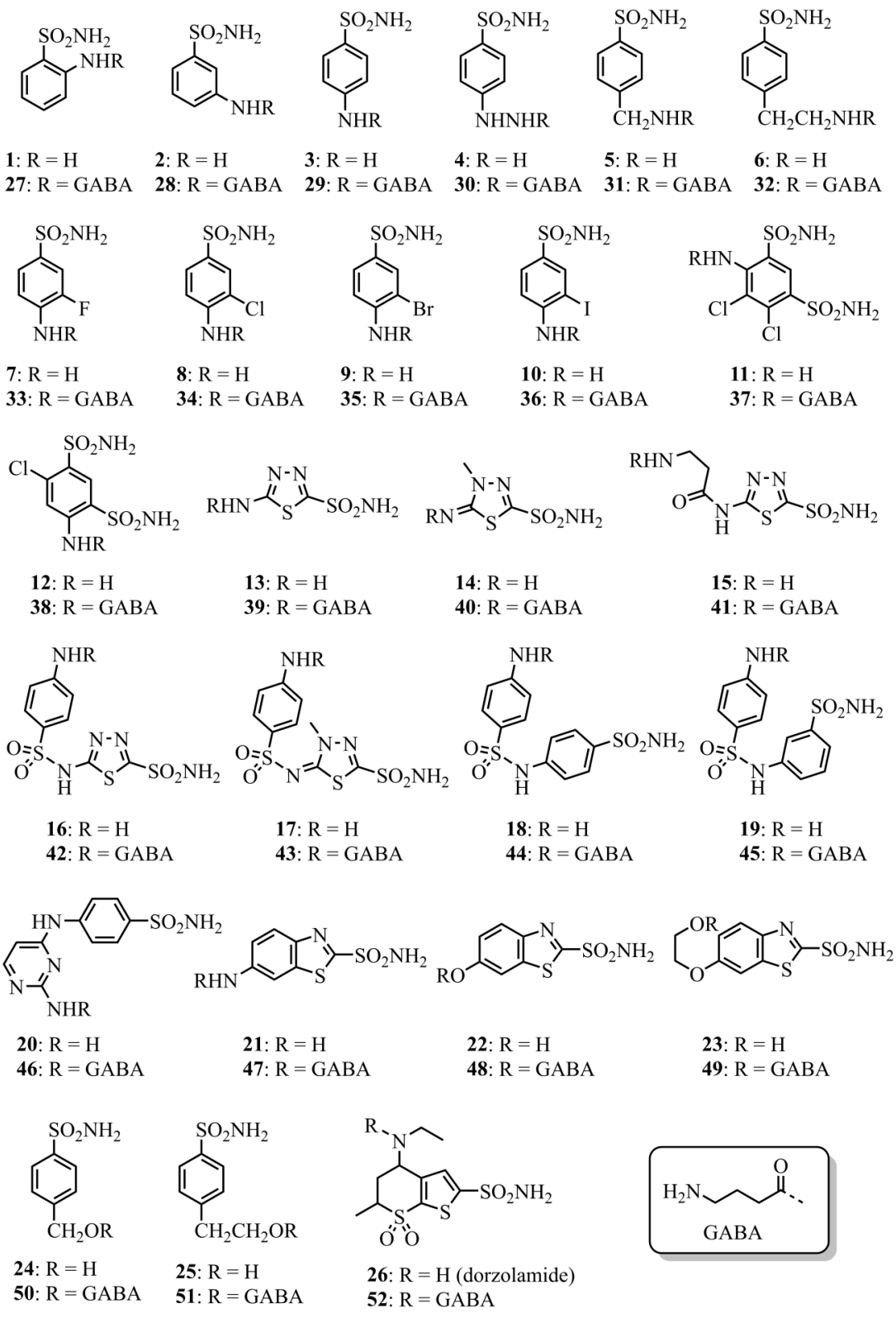

Figure 1. Aromatic and heterocyclic sulfonamides and their GABA conjugates as CA inhibitors

Table 1. Experimental and predicted inhibition constants for CA inhibitors

Inhibitor	CA II isoform inhibition			CA IV isoform inhibition		
	Ki (nM)	**pKi (nM)**	**pKi* (nM)**	**Ki (nM)**	**pKi (nM)**	**pKi* (nM)**
1[#]	295	6.53	6.24	1310	5.88	5.68
2	240	6.62	6.48	2200	5.66	5.74
3	300	6.52	6.65	3000	5.52	5.77
4	320	6.49	6.66	3215	5.49	5.94
5	170	6.77	6.66	2800	5.55	5.74
6	160	6.80	6.67	2450	5.61	5.75
7[#]	60	7.22	7.25	180	6.74	6.43
8	110	6.96	7.25	320	6.49	6.31
9[#]	40	7.40	7.25	66	7.18	6.75
10[#]	70	7.15	7.25	125	6.90	6.78
11	28	7.55	7.72	175	6.76	6.97
12	75	7.12	7.18	160	6.80	6.60
13	60	7.22	7.46	540	6.27	6.34
14	19	7.72	7.62	355	6.45	6.48
15	3	8.52	7.99	125	6.90	7.21
16	2	8.70	8.60	5	8.30	8.06
17[#]	6	8.22	8.58	8	8.10	8.28
18[#]	6	8.22	8.01	50	7.30	7.61
19	9	8.05	7.99	53	7.28	7.52
20[#]	12	7.92	8.06	154	6.81	7.68
21	9	8.05	8.15	19	7.72	7.57
22	8	8.10	8.15	17	7.77	7.71
23[#]	7	8.15	8.29	15	7.82	7.92
24[#]	125	6.90	6.66	560	6.25	5.99
25	110	6.96	6.67	450	6.35	6.03
26	9	8.05	8.28	45	7.35	7.34
27	197	6.71	6.90	243	6.61	6.47
28	182	6.74	7.04	215	6.67	6.54
29[#]	112	6.95	7.09	164	6.79	6.58
30	212	6.67	7.21	305	6.52	6.70
31	32	7.49	7.21	69	7.16	6.58
32	30	7.52	7.22	62	7.21	6.56
33	10	8.00	7.60	38	7.42	7.20
34	31	7.51	7.60	63	7.20	7.06
35	30	7.52	7.60	60	7.22	7.52
36	27	7.57	7.60	56	7.25	7.56
37	10	8.00	8.01	50	7.30	7.44
38	9	8.05	7.54	50	7.30	7.12
39	8	8.10	8.10	43	7.37	7.20
40	10	8.00	8.00	42	7.38	7.52
41	5	8.30	8.36	15	7.82	7.96
42	3	8.52	8.54	9	8.05	7.83
43	4	8.40	8.46	14	7.85	7.94

44	14	7.85	7.86	75	7.12	7.23
45	13	7.89	7.85	67	7.17	7.13
46	11	7.96	8.07	50	7.30	7.55
47	5	8.30	8.40	12	7.92	7.79
48[#]	54	7.27	8.40	12	7.92	8.02
49	4	8.40	8.31	11	7.96	7.88
50	73	7.14	7.21	180	6.74	6.83
51	66	7.18	7.22	155	6.81	6.81
52	6	8.22	8.17	32	7.49	7.63

pKi*****- pKi value predicted by the MLR model
[#]- inhibitors of the test set analyzed using the MLR model

Using the DTC Lab Tools software and following the methodology described by Banerjee & Roy (2023), an extensive modeling process was conducted, resulting in the development of over 225 million MLR models for the CA II isoform and more than 274 million MLR models for the CA IV isoform. The selection of optimal models for both isoforms was based on rigorous validation using multiple statistical metrics. The validation process involved assessments performed on the training set, test set, and through the Y-randomization test. Key statistical metrics used for model evaluation included the coefficient of determination (R^2), adjusted coefficient of determination (R_a^2), cross-validated coefficient of determination (Q^2), scaled $r_{m(training)}^2$ and $\Delta r_{m(training)}^2$ for training set validation. For test set validation, predictive R^2 (R_{pred}^2), Q_{F1}^2, Q_{F2}^2, concordance correlation coefficient (CCC), $r_{m(test)}^2$, and $\Delta r_{m(test)}^2$ were calculated. Furthermore, average R^2, average Q^2, and ${}^{C}Rp^2$ were derived from the results of the Y-randomization test. The developed MLR models were assessed for acceptability based on the criteria recommended by Golbraikh & Tropsha (2002), ensuring their robustness and reliability. Additionally, the predictive quality of the models was evaluated using mean absolute error (MAE)-based criteria and categorized as 'Good', 'Moderate', or 'Bad', as recommended by Roy et al. (2016).

Molecular dynamics simulation

The molecular dynamics simulation of the most potent GABA-sulfonamide conjugate, compound **42**, bound to the CA II (PDB: 4M2U) and CA IV (PDB: 3FW3) isoforms was performed using Desmond molecular dynamics software (version 2018.4). Developed by D. E. Shaw Research in New York, Desmond relied on PDB files obtained from the Protein Data Bank. The simulation protocol, with minor modifications, followed the methodology described by Ilić (2020) and Ilić et al. (2021). The water molecules in the system were modeled using the simple point charge (SPC) solvent model. Chloride ions (Cl^-) were introduced to neutralize the system, ensuring a net zero charge in the simulation box. The final system consisted of approximately 30,000 atoms. Before initiating the MD simulation, the system underwent a six-step relaxation protocol to ensure stability. The relaxed system was then subjected to a 100 ns simulation using a normal pressure-temperature (NPT) ensemble. A Nosé–Hoover thermostat maintained the temperature at 300 K, while a Martyna–Tobias–Klein barostat controlled the pressure at 1.01325 bar. Atomic coordinate data and system energies were recorded at intervals of 1 ps. To evaluate the dynamic behavior and

stability of the inhibitor-isoform complexes, root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyses were conducted over the entire simulation period. These metrics provided insights into the structural fluctuations and conformational stability of the complexes during the simulation.

Results and Discussion

Machine learning

The most effective MLR model for GABA-sulfonamide conjugates targeting CA II isoform is illustrated in Figure 2. The model demonstrates high predictive accuracy, as evidenced by the corresponding correlation parameters in Figure 2, and predicted pKi* values in Table 1. The statistical robustness and reliability of the model are substantiated by its compliance with the criteria recommended by Golbraikh & Tropsha (2002).

pKi*= 3.00709(+/-0.79181) - 0.00018(+/-0.00005) × Wap - 12.81387(+/-4.49435) × PW4 + 0.30171(+/-0.0486) × SRW05 + 0.95161(+/-0.14585) × piID - 0.31387(+/-0.07068) × nCb- + 0.81704(+/-0.11941) × nArX

Training set validation parameters:

R^2 = 0.89052, R_a^2 = 0.8712, Q^2 = 0.85603, $r_{m(training)}^2$ = 0.79956, $\Delta r_{m(training)}^2$ = 0.08739

Test set validation parameters:

R_{pred}^2 = 0.73695, Q_{F1}^2 = 0.55258, Q_{F2}^2 = 0.51286, CCC = 0.82018, $r_{m(test)}^2$ = 0.64512, $\Delta r_{m(test)}^2$ = 0.18128

Y-randomization test results:

average R^2 = 0.233384, average Q^2 = -0.14606, ${}^{C}Rp^2$ = 0.804388

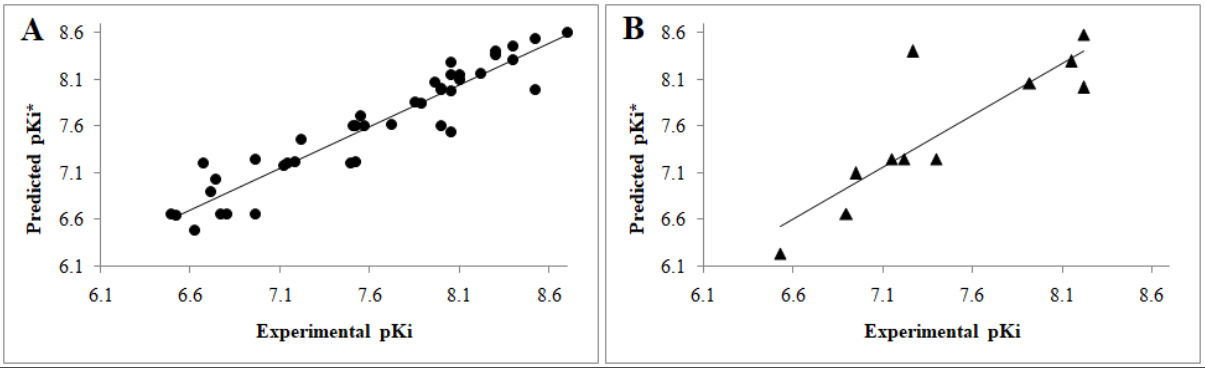

Figure 2. The MLR model, correlation parameters, and plots of experimental versus predicted pKi values for (**A**) the training set and (**B**) the test set of CA II inhibitors

Table 2. Calculated molecular descriptors of CA inhibitors

Inhibitor	Wap	PW4	SRW05	piID	nCb-	nArX	ZM1V	Qindex	SaasC	SddssS	SsCl
1[#]	383	0.15	0	6.18	2	0	194	8	0.15	-3.66	0
2	383	0.13	0	6.16	2	0	194	8	0.43	-3.61	0
3	383	0.12	0	6.15	2	0	194	8	0.59	-3.58	0
4	490	0.13	0	6.31	2	0	210	8	0.70	-3.60	0
5	490	0.13	0	6.31	2	0	190	8	1.13	-3.55	0
6	618	0.14	0	6.45	2	0	194	8	1.22	-3.55	0
7[#]	480	0.12	0	6.33	3	1	250	9	-1.18	-3.84	0
8	480	0.12	0	6.33	3	1	250	9	0.47	-3.68	5.57
9[#]	480	0.12	0	6.33	3	1	250	9	1.04	-3.63	0
10[#]	480	0.12	0	6.33	3	1	250	9	1.31	-3.60	0
11	1139	0.16	0	7.20	5	2	430	14	-2.66	-8.49	11.20
12	986	0.15	0	7.08	4	1	374	13	-1.61	-8.28	5.56
13	273	0.13	2.40	5.69	0	0	244	8	-0.17	-3.72	0
14	353	0.13	2.40	5.89	0	0	236	9	0.13	-3.70	0
15	864	0.14	2.40	6.52	0	0	312	9	-0.22	-3.85	0
16	2965	0.14	2.40	8.27	2	0	436	16	0.33	-7.99	0
17[#]	3327	0.15	2.40	8.36	2	0	446	17	0.34	-8.02	0
18[#]	3469	0.14	0	9.13	4	0	408	16	0.41	-7.71	0
19	3469	0.15	0	9.17	4	0	408	16	0.14	-7.88	0
20[#]	2665	0.15	0	8.43	2	0	310	13	1.39	-3.67	0
21	1550	0.15	2.40	8.00	3	0	278	13	0.49	-3.71	0
22	1550	0.15	2.40	8.00	3	0	294	13	-0.07	-3.75	0
23[#]	2658	0.15	2.40	8.36	3	0	338	13	0.43	-3.77	0
24[#]	490	0.13	0	6.31	2	0	214	8	0.71	-3.61	0
25	618	0.14	0	6.45	2	0	218	8	0.98	-3.60	0
26	3300	0.17	2.40	7.67	0	0	350	17	0.66	-7.30	0
27	1280	0.15	0	7.07	2	0	266	9	0.13	-3.83	0
28	1280	0.14	0	7.00	2	0	266	9	0.40	-3.74	0
29[#]	1280	0.13	0	6.97	2	0	266	9	0.57	-3.68	0
30	1524	0.13	0	7.05	2	0	282	9	0.60	-3.68	0
31	1524	0.13	0	7.05	2	0	270	9	0.89	-3.66	0
32	1795	0.13	0	7.13	2	0	274	9	1.04	-3.64	0
33	1471	0.14	0	7.11	3	1	322	10	-1.25	-3.95	0
34	1471	0.14	0	7.11	3	1	322	10	0.39	-3.79	5.87
35	1471	0.14	0	7.11	3	1	322	10	0.97	-3.74	0
36	1471	0.14	0	7.11	3	1	322	10	1.24	-3.71	0
37	2636	0.16	0	7.80	5	2	502	15	-2.84	-8.77	11.70
38	2389	0.15	0	7.71	4	1	446	14	-1.74	-8.55	5.77
39	1040	0.14	2.40	6.61	0	0	316	9	-0.20	-3.85	0
40	1208	0.14	2.40	6.61	0	0	326	10	0	-3.85	0
41	2208	0.13	2.40	7.06	0	0	392	10	-0.34	-3.92	0
42	6516	0.15	2.40	8.91	2	0	508	17	0.31	-8.15	0
43	7198	0.15	2.40	9.00	2	0	518	18	0.33	-8.17	0
44	7424	0.14	0	9.77	4	0	480	17	0.33	-7.86	0

45	7424	0.15	0	9.81	4	0	480	17	0.06	-8.04	0
46	6074	0.15	0	9.11	2	0	382	14	1.29	-3.73	0
47	4018	0.15	2.40	8.70	3	0	350	14	0.48	-3.80	0
48[#]	4018	0.15	2.40	8.70	3	0	370	14	0.20	-3.82	0
49	6005	0.15	2.40	8.90	3	0	414	14	0.40	-3.81	0
50	1524	0.13	0	7.05	2	0	290	9	0.75	-3.68	0
51	1795	0.13	0	7.13	2	0	294	9	0.96	-3.66	0
52	6291	0.17	2.40	8.15	0	0	436	19	0.17	-7.64	0

[#]- inhibitors of the test set analyzed using the MLR model

Furthermore, the quality of the model was evaluated using the MAE-based criteria proposed by Roy et al. (2016), and an MAE value of 0.18025 confirms that the model satisfies the requirements for classification as a 'Good' model. This categorization underscores its applicability in reliably predicting the pKi values of novel GABA-sulfonamide conjugates, paving the way for rational design of potent CA II inhibitors. To provide deeper insights into the molecular determinants of the observed inhibitory activity, the calculated values of the descriptors used in the MLR model (Wap, PW4, SRW05, piID, nCb-, nArX) are presented in Table 2 for each GABA-sulfonamide conjugate. The descriptors in the MLR model and their directional contributions reveal key design principles for optimizing inhibitory potency against the CA II isoform (Figure 2).

The negative contribution of Wap (All-Path Wiener Index) to the inhibitory activity suggests that highly connected or compact molecular structures diminish the effectiveness of GABA-sulfonamide conjugates as CA II inhibitors. This likely occurs because overly condensed or less branched molecules may hinder proper alignment within the enzyme's active site or reduce binding flexibility. To enhance spatial compatibility and binding efficiency, introducing moderate branching or elongation into the molecular structure is recommended, allowing for improved interaction with the enzyme's active site.

The negative contribution of PW4 (Path/Walk 4 - Randić Shape Index) indicates that more complex molecular shapes over short paths of four bonds adversely affect inhibitory activity. This suggests that excessive local branching or structural irregularities may disrupt proper docking or cause steric clashes within the enzyme's active site. To address this, simplifying local molecular shapes by minimizing unnecessary branching or bulky substituents near the core structure is recommended to enhance binding efficiency and inhibition.

The positive contribution of SRW05 (Self-Returning Walk Count of Order 5) underscores the significance of local symmetry and structural stability in enhancing inhibitory activity. Molecules capable of maintaining symmetry over 5-bond paths are likely to stabilize interactions with the zinc ion and surrounding residues within the enzyme's catalytic center. To optimize this effect, it is advisable to preserve or enhance molecular symmetry, particularly in regions directly involved in enzyme binding.

The positive contribution of piID (Conventional Bond Order ID Number) indicates that stronger or more polarizable bonds, such as sulfonamide bonds, play a crucial role in enhancing inhibitory activity. This enhancement likely arises from improved coordination with the zinc ion or other critical interactions within the enzyme's active site. To optimize binding affinity, the design should prioritize molecules featuring strong, polar bonds, particularly high-quality sulfonamide groups.

The negative contribution of nCb- (Number of Substituted Benzene $C(sp^2)$) suggests that an increased number of substituents on the benzene ring can introduce steric hindrance or disrupt the electronic balance required for effective binding. To mitigate these issues and improve docking efficiency, it is advisable to limit the number of substituents on the benzene ring, ensuring the molecule remains compact and well-suited for interaction with the active site.

The positive contribution of nArX (Number of X on Aromatic Ring) highlights the beneficial impact of functionalizing the aromatic ring with suitable groups, such as -OH, -Cl, or $-CH_3$, on inhibitory activity. These modifications likely enhance interactions with the enzyme through improved hydrogen bonding, dipole interactions, or hydrophobic effects. To optimize these benefits, it is recommended to introduce functional groups that strengthen specific interactions while avoiding excessive bulk or disruptions to the molecule's electronic balance.

The MLR model developed for GABA-sulfonamide conjugates targeting the CA IV isoform is both robust and highly predictive. Its accuracy is demonstrated through the correlation parameters shown in Figure 3 and the predicted pKi* values detailed in Table 1. Statistical validation confirms the model’s reliability, meeting the criteria established by Golbraikh & Tropsha (2002), while an MAE value of 0.22307, categorizes it as a 'Good' model. The descriptor values (ZM1V, Qindex, Wap, SaasC, SddssS, SsCl), which serve as the foundation of the MLR model, are presented in Table 2 to provide a detailed analysis of molecular factors influencing inhibitory activity against the CA IV isoform.

The positive contribution of ZM1V (Zero-Order Molecular Connectivity Index) indicates that greater connectivity in the molecular graph plays a critical role in enhancing binding affinity. Molecules with well-connected atomic structures are likely to align their functional groups more effectively, facilitating optimal interactions with the CA IV active site. To maximize inhibitory activity, the design should focus on creating conjugates with dense and well-connected molecular scaffolds.

The positive contribution of the Q-index (Quadratic Index) highlights the role of molecular complexity and quadratic contributions in enhancing inhibitory activity. Higher Q-index values suggest that conjugates with carefully distributed atomic connections are structurally well-suited to fit the CA IV binding pocket. To optimize activity, it is important to maintain molecular complexity while avoiding excessive branching, ensuring an effective balance between structure and functionality.

The negative contribution of Wap suggests that fewer or less accessible molecular paths enhance inhibitory activity, indicating that excessive molecular flexibility or extended networks can reduce binding efficiency by introducing conformational instability. To optimize binding to CA IV, designing compact and rigid molecules is recommended, as these structures promote stability and improve interaction with the active site.

pKi*= 2.83126(+/-0.34758) + 0.01346(+/-0.00133) × ZM1V + 0.23813(+/-0.03506) × Qindex - 0.00038(+/-0.00005) × Wap + 0.10417(+/-0.08338) × SaasC + 0.41842(+/-0.04781) × SddssS - 0.06358(+/-0.02217) × SsCl

Training set validation parameters:

R^2 = 0.89695, R_a^2 = 0.87876, Q^2 = 0.86014, $r_{m(training)}^2$ = 0.80295, $\Delta r_{m(training)}^2$ = 0.08654

Test set validation parameters:

R_{pred}^2 = 0.83926, Q_{F1}^2 = 0.71703, Q_{F2}^2 = 0.71468, CCC = 0.8907, $r_{m(test)}^2$ = 0.67378, $\Delta r_{m(test)}^2$ = 0.14716

Y-randomization test results:

average R^2 = 0.224263, average Q^2 = -0.13319, $^{C}Rp^2$ = 0.822077

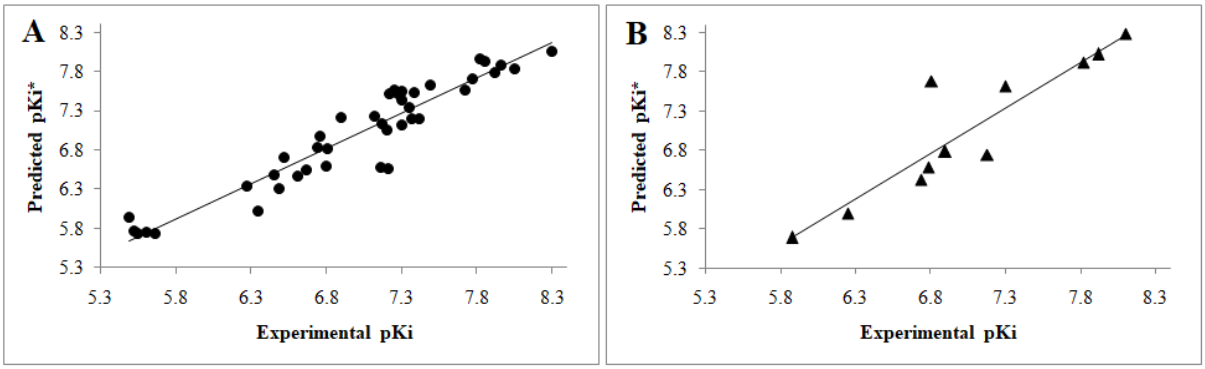

Figure 3. The MLR model, correlation parameters, and plots of experimental versus predicted pKi values for (**A**) the training set and (**B**) the test set of CA IV inhibitors

The positive contribution of SaasC (Sum of Atomic Electronegativity for Carbon Atoms) underscores the importance of the carbon framework, particularly in the GABA backbone and sulfonamide group, in stabilizing interactions with CA IV through hydrogen bonding or hydrophobic effects. A strong carbon-based backbone is essential for enhancing inhibitory activity.

The positive contribution of SddssS (Sum of Electronic Contributions for Sulfur Atoms with Specific Bonding Patterns) highlights the crucial role of sulfur atoms in the sulfonamide group. This descriptor emphasizes that these atoms contribute significantly to coordinating the zinc ion in the CA IV active site. To optimize inhibitory activity, the design should prioritize the inclusion

of sulfur atoms within the sulfonamide moiety, as they are essential structural components for effective binding.

The negative contribution of SsCl (Sum of Electronic Contributions for Chlorine Atoms) indicates that minimal electronic influence from chlorine atoms enhances inhibitory activity. This suggests that excessive electron-withdrawing effects or steric hindrance caused by chlorine substitution can disrupt optimal interactions with the CA IV active site. To maximize activity, chlorine should be used sparingly to modulate properties without introducing excessive halogen substitution.

Molecular dynamics simulation

Understanding 2D and 3D molecular properties is critical in drug development to achieve a detailed characterization of molecular behavior and interactions. While 2D descriptors capture essential topological and electronic features, 3D properties such as spatial orientation and conformational flexibility are indispensable for accurate modeling of binding affinities, steric effects, and molecular dynamics (Cho & Choi, 2019). By leveraging 3D molecular characteristics, molecular dynamics reveals dynamic conformational changes, binding mechanisms, and interaction networks, offering a powerful framework for optimizing drug candidates (De Vivo et al., 2016; Do et al., 2018).

The molecular dynamics simulation of the most potent GABA-sulfonamide conjugate, compound **42**, in complex with the CA II isoform over a 100 ns trajectory unveiled significant details regarding the stability and interaction dynamics of the complex (Figure 4). The RMSD and RMSF values for Cα atoms, side chains, and heavy atoms consistently remained below the 2 Å threshold, in agreement with Liu & Kokubo (2017). These findings highlight minimal structural rearrangements and conformational changes, confirming the stability of the compound **42**/CA II complex. The interactions observed throughout the 100 ns molecular dynamics simulation underscored the critical roles of Glu69, His94, His119, Phe131, Thr199, Thr200, and Pro201 in the formation and stabilization of the compound **42**/CA II complex (Figure 4). The involvement of these residues in CA enzyme inhibition is thoroughly documented (Supuran & Scozzafava, 2007; Supuran, 2016a; Jaitak et al., 2024; Naeem et al., 2024). Hydrogen bonding is a central feature of ligand stabilization, as demonstrated in Figure 4A. Thr199 and His119 are critical contributors to these interactions, underscoring their importance in maintaining stability. Glu69 further supports the ligand through a combination of hydrogen bonding and ionic interactions, highlighting the role of charge complementarity. Additional stabilization arises from hydrophobic contacts with His94 and Phe131, as well as water-mediated bridges involving Thr200 and Pro201. As observed in Figure 4C, compound **42** stabilizes the complex through strong hydrogen bonds formed by the sulfonamide group via its oxygen atom with Thr199 and its nitrogen atom with His119. The amino group of the GABA tail interacts with Glu69 through hydrogen bonding and ionic interactions, playing a critical role in electrostatic stabilization. Hydrophobic interactions occur between the benzene ring and Phe131, as well as the thiadiazole ring and His94, strengthening the molecular interface through van der Waals forces. Additionally, the nitrogen

atoms of the thiadiazole ring form water-mediated bridges with Thr200 and Pro201, providing further stability to the complex.

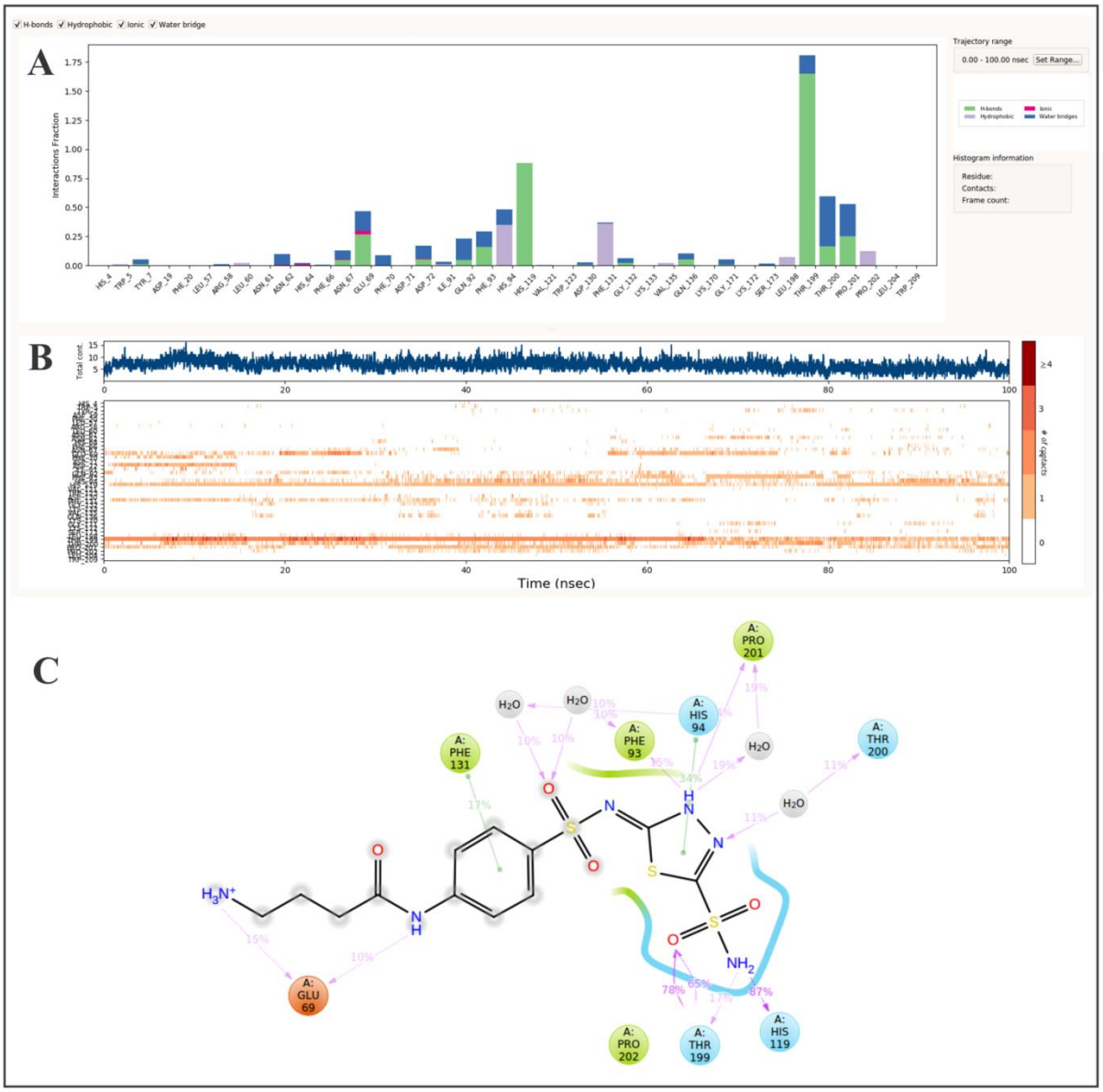

Figure 4. Interaction analysis of the GABA-sulfonamide conjugate **42** with CA II isoform: frequency distribution (**A**), interaction timeline (**B**), and spatial contact map (**C**) from a 100 ns molecular dynamics simulation

Based on the results, the design of novel GABA-sulfonamide conjugates targeting the CA II isoform can focus on replacing the thiadiazole ring with a heterocyclic scaffold capable of forming direct hydrogen bonds with Thr200 and Pro201. This substitution could eliminate reliance on water-mediated interactions, thereby enhancing binding affinity and the activity of the compounds (Fischer & Riedl, 2013). Additionally, the incorporation of hydrophobic substituents, such as

methyl or ethyl groups, on the benzene ring could strengthen van der Waals interactions with Phe131, further contributing to increased activity.

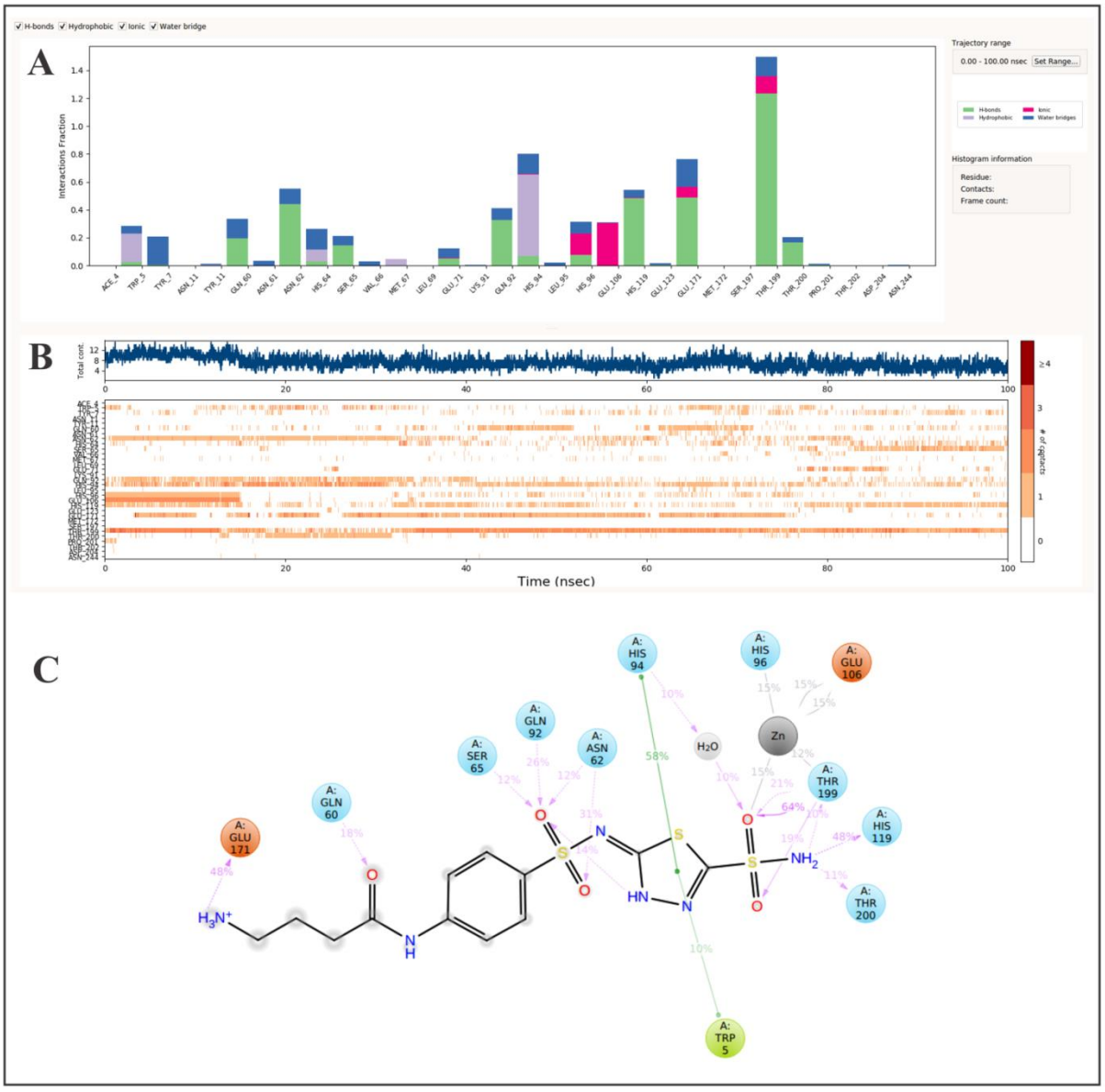

Figure 5. Interaction analysis of the GABA-sulfonamide conjugate **42** with CA IV isoform: frequency distribution (**A**), interaction timeline (**B**), and spatial contact map (**C**) from a 100 ns molecular dynamics simulation

The molecular dynamics simulation of compound **42** with the CA IV isoform revealed stability, with RMSD and RMSF values consistently below 2 Å (Figure 5). Key interactions were observed with residues Trp5, Gln60, Asn62, His64, Ser65, Gln92, His94, His96, Glu106, His119, Glu171, Thr199, and Thr200, emphasizing their roles in CA IV enzyme inhibition, as previously reported (Supuran & Scozzafava, 2007; Supuran, 2016a; Jaitak et al., 2024; Naeem et al., 2024).

The contributions of distinct interaction types are quantified in Figure 5, highlighting their roles in stabilizing compound **42**. The sulfonamide group in compound **42** forms strong hydrogen bonds with Thr199 via its oxygen atoms and with His119 and Thr200 through its nitrogen atom, while also coordinating to the zinc ion via an oxygen atom. A water-mediated bridge connects the sulfonamide group oxygen to His94, contributing to the stabilization of the complex. In contrast, the sulfonamide moiety, located between the benzene and thiadiazole rings, exhibits significant hydrogen bonding interactions with Asn62, Ser65, and Gln92 through its sulfonyl group, highlighting its distinct role in binding. The amino group of the GABA tail engages in strong hydrogen bonds with Glu171 and also demonstrates a degree of ionic interactions. Furthermore, the carbamoyl group of the GABA fragment interacts through hydrogen bonds with Gln60. The thiadiazole ring, as part of the structure, participates in persistent hydrophobic interactions with His94, while additional hydrophobic interactions are observed with Trp5, emphasizing the overall stabilizing contributions of compound **42** within the binding site. From the perspective of rational design, one potential approach for developing more potent inhibitors containing the GABA fragment toward the CA IV isoform is the replacement of the sulfonamide group with a hydroxamate, which would enhance coordination with the zinc ion and optimize hydrogen bonding (Di Fiore et al., 2012; Hou et al., 2021).

Conclusion

This study provides a comprehensive analysis of GABA-sulfonamide conjugates as effective CA inhibitors, specifically targeting the CA II and CA IV isoforms. By integrating ML-based MLR modeling with MD simulations, key structural and physicochemical determinants of inhibitory activity were identified. The developed MLR models demonstrated exceptional predictive accuracy and statistical robustness, enabling the identification of critical molecular descriptors that influence CA inhibition. Simultaneously, MD simulations offered atomistic insights into the dynamic stability and specific binding interactions of the most potent GABA-sulfonamide conjugate within enzyme-ligand complexes. The findings highlight the importance of molecular connectivity, symmetry, and functional group modifications in optimizing inhibitor design. Structural modifications, such as substituting the thiadiazole ring or sulfonamide group, could further enhance binding affinity and isoform selectivity. This integrative approach enhances our mechanistic understanding of sulfonamide-based CA inhibitors while providing a robust framework for the rational design of next-generation inhibitors with improved therapeutic efficacy, particularly for the treatment of glaucoma, a progressive neurodegenerative disorder. Future research focusing on scaffold optimization and novel functional groups will likely yield inhibitors with enhanced isoform selectivity and potency.

Acknowledgement

This work was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 451-03-65/2024-03/200113). I am deeply grateful to D. E. Shaw Research for generously providing access to the Desmond software package, which was essential for conducting this research.

Conflict-of-Interest Statement

The auth

Informed Consent Statement

Informed consent was obtained from all individual participants included in the study.

Human and Animal Rights Statement

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Angeli, A., Chelli, I., Lucarini, L., Sgambellone, S., Marri, S., Villano, S., Ferraroni, M., De Luca, V., Capasso, C., Carta, F., & Supuran, C. T. (2024). Novel carbonic anhydrase inhibitors with dual-tail core sulfonamide show potent and lasting effects for glaucoma therapy. Journal of Medicinal Chemistry, 67, 3066–3089.

Angeli, A., Paoletti, N., & Supuran, C. T. (2023). Five-membered heterocyclic sulfonamides as carbonic anhydrase inhibitors. Molecules, 28, 1–23.

Banerjee, A., & Roy, K. (2023). Machine-learning-based similarity meets traditional QSAR: “q-RASAR” for the enhancement of the external predictivity and detection of prediction confidence outliers in an hERG toxicity dataset. Chemometrics and Intelligent Laboratory Systems, 237, 104829.

Cho, H., & Choi, I. S. (2019). Enhanced deep-learning prediction of molecular properties via augmentation of bond topology. ChemMedChem, 14, 1604–1609.

De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. Journal of Medicinal Chemistry, 59, 4035–4061.

Di Fiore, A., Maresca, A., Supuran, C. T., & De Simone, G. (2012). Hydroxamate represents a versatile zinc binding group for the development of new carbonic anhydrase inhibitors. Chemical Communications, 48, 8838–8840.

Do, P. C., Lee, E. H., & Le, L. (2018). Steered molecular dynamics simulation in rational drug design. Journal of Chemical Information and Modeling, 58, 1473–1482.

Fischer, T., & Riedl, R. (2013). Strategic targeting of multiple water-mediated interactions: a concise and rational structure-based design approach to potent and selective MMP-13 inhibitors. ChemMedChem, 8, 1457–1461.

Frasnetti, E., Cucchi, I., Pavoni, S., Frigerio, F., Cinquini, F., Serapian, S. A., Pavarino, L. F., & Colombo, G. (2024). Integrating molecular dynamics and machine learning algorithms to predict the functional profile of kinase ligands. Journal of Chemical Theory and Computation, 20, 9209–9229.

Golbraikh, A., & Tropsha, A. (2002). Beware of q^2! Journal of Molecular Graphics and Modelling, 20, 269–276.

Hou, R., He, Y., Yan, G., Hou, S., Xie, Z., & Liao, C. (2021). Zinc enzymes in medicinal chemistry. European Journal of Medicinal Chemistry, 226, 113877.

Ilić, B. S. (2020). Molecular dynamics simulations of ASC09, ritonavir, lopinavir and darunavir with the COVID-19 protease. Chemia Naissensis, 3, 155–166.

Ilić, B. S., Gajić, M., Bondžić, B. P., Džambaski, Z., Kocić, G., & Šmelcerović, A. (2021). Deoxyribonuclease I inhibitory properties, molecular docking and molecular dynamics simulations of 1-(pyrrolidin-2-yl)propan-2-one derivatives. Chemistry & Biodiversity, 18, 1–8.

Jaitak, A., Kumari, K., Kounder, S., & Monga, V. (2024). Carbonic anhydrases: Moiety appended derivatives, medicinal and pharmacological implications. Bioorganic & Medicinal Chemistry, 114, 117933.

Liu, K., & Kokubo, H. (2017). Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: A cross-docking study. Journal of Chemical Information and Modeling, 57, 2514–2522.

Mincione, G., Menabuoni, L., Briganti, F., Mincione, F., Scozzafava, A., & Supuran, C. T. (1999). Carbonic anhydrase inhibitors. Part 79. Synthesis of topically acting sulfonamides incorporating GABA moieties in their molecule, with long-lasting intraocular pressure-lowering properties. European Journal of Pharmaceutical Sciences, 9, 185–199.

Naeem, N., Sadiq, A., Othman, G. A., Yassin, H. M., & Mughal, E. U. (2024). Exploring heterocyclic scaffolds in carbonic anhydrase inhibition: a decade of structural and therapeutic insights. RSC Advances, 14, 35769–35970.

Nocentini, A., & Supuran, C. T. (2019). Adrenergic agonists and antagonists as antiglaucoma agents: a literature and patent review (2013-2019). Expert Opinion on Therapeutic Patents, 29, 805–815.

Pagnozzi, D., Pala, N., Biosa, G., Dallocchio, R., Dessì, A., Singh, P. K., Rogolino, D., Di Fiore, A., De Simone, G., Supuran, C. T., & Sechi, M. (2022). Interaction studies between carbonic anhydrase and a sulfonamide inhibitor by experimental and theoretical approaches. ACS Medicinal Chemistry Letters, 13, 271–277.

Rodríguez-Pérez, R., & Bajorath, J. (2021). Explainable machine learning for property predictions in compound optimization. Journal of Medicinal Chemistry, 64, 17744–17752.

Roy, K., Das, R. N., Ambure, P., & Aher, R. B. (2016). Be aware of error measures. Further studies on validation of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems, 152, 18–33.

Supuran, C. T. (2016a). How many carbonic anhydrase inhibition mechanisms exist? Journal of Enzyme Inhibition and Medicinal Chemistry, 31, 345–360.

Supuran, C. T. (2016b). Structure and function of carbonic anhydrases. Biochemical Journal, 473, 2023–2032.

Supuran, C. T. (2021). Emerging role of carbonic anhydrase inhibitors. Clinical Science, 135, 1233–1249.

Supuran, C. T., & Scozzafava, A. (2007). Carbonic anhydrases as targets for medicinal chemistry. Bioorganic & Medicinal Chemistry, 15, 4336–4350.

Tetko, I. V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V. A., Radchenko, E. V., Zefirov, N. S., Makarenko, A. S., Tanchuk, V. Y., & Prokopenko, V. V. (2005). Virtual computational chemistry laboratory - design and description. Journal of Computer - Aided Molecular Design, 19, 453–463.

Tinivella, A., Pinzi, L., & Rastelli, G. (2021). Prediction of activity and selectivity profiles of human carbonic anhydrase inhibitors using machine learning classification models. Journal of Cheminformatics, 13, 1–15.

Todeschini, R., & Consonni, V. (2009). Molecular descriptors for chemoinformatics. Volumes I & II, (2nd ed.). Wiley-VCH Verlag GmbH & Co. KGaA.

Wei, H., & McCammon, J. A. (2024). Structure and dynamics in drug discovery. npj Drug Discovery, 1, 1–8.

Xu, W. (2024). Current status of computational approaches for small molecule drug discovery. Journal of Medicinal Chemistry, 67, 18633–18636.