# AI-Designed Family of Dual COX-2/mPGES-1 Inhibitors to Treat Inflammation with Greater Precision and Effectiveness

*Prashant S. Kharkar[1]*, Aarshit Mittal[2]*, Ranbir Saluja[1], Shobhan Shah[2,3]*

[1]Department of Pharmaceutical Sciences and Technology, Institute of Chemical Technology, Nathalal Parekh Marg, Matunga, Mumbai 400 019. India.

[2]Aarth Software Private Limited, Metro Station, 4th Floor, Vasavi MPM Grand, Beside Ameerpet Metro Station, Hyderabad, Telangana, 500073, India

[3]1340 S De Anza Blvd Ste #208, San Jose, CA 95129, USA

* Corresponding author

[1]E-mail: ps.kharkar@ictmumbai.edu.in; Tel: +91 22 3361 2225

[2]E-mail: aarshit@aarth.io; Tel: +91 8588988109



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

## Abstract

Cyclooxygenase-2 (COX-2) selective inhibitors represent a major class of anti-inflammatory drugs, but their clinical utility remains constrained by cardiovascular risks and incomplete efficacy. Dual inhibition of COX-2 and microsomal prostaglandin E synthase-1 (mPGES-1) offers a promising strategy for enhanced safety and effectiveness. However, traditional drug discovery timelines for multi-target inhibitors are prohibitively long. Here, we demonstrate an AI-driven computational pipeline that compressed the design-to-candidate stage from years to one month, generating a novel family of dual COX-2/mPGES-1 inhibitors. We employed a hybrid reinforcement learning-generative adversarial network (RL-GAN) trained on 308 curated COX-2 inhibitors to explore beyond known chemical space, producing ~50,000 de novo molecules. Multi-stage filtering through ADMET prediction models, billion-scale virtual screening using deep learning-enhanced docking protocols, and structure-based design against four COX-2 crystal structures yielded 23 high-confidence candidates. These molecules exhibit predicted COX-2 binding affinities comparable to or exceeding rofecoxib (-9 to -10.5 kcal/mol), with >100-fold predicted selectivity over COX-1 and concurrent mPGES-1 engagement (-8.3 to -10.3 kcal/mol). Computational patent analysis across 44.3 million records suggests structural novelty (Tanimoto <0.8), with 92.9% of candidates dissimilar to known inhibitors. Rigorous validation of computational models, including Boltz-2 predicted COX-1 structure (backbone RMSD 0.541 Å vs. 6Y3C crystal structure), supports prediction reliability. This work demonstrates that AI-guided de novo design can rapidly identify innovative multi-target therapeutics, offering a paradigm shift in early-stage drug discovery efficiency while highlighting the essential role of experimental validation in clinical translation.

**Keywords:** COX-2; mPGES-1; Reinforcement learning; GPU computing; RL-GAN



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

## Introduction

Prostanoids, including prostaglandins (PGs), prostacyclin ($PGI_2$), and thromboxanes (TXs), make up a family of biologically-active lipids derived from arachidonic acid (AA), which is released from cell membrane phospholipids by the action of phospholipase $A_2$ upon tissue injury, inflammation and other signals. These mediators take part in varied physiological and pathological events.[1] The prostanoids biosynthesis pathway thus originates from AA and proceeds to $PGH_2$, with $PGG_2$ being an intermediary. It branches from $PGH_2$ into the major members of the family. Prostaglandin $E_2$ ($PGE_2$), a potent inflammatory mediator, is produced by the action of cyclooxygenase-1 or -2 (COX-1/2) on AA, via $PGG_2$ and $PGH_2$. The terminal step in the biosynthesis of $PGE_2$, i.e., conversion of $PGH_2$ to $PGE_2$, is catalyzed by microsomal prostaglandin synthetase 1 (mPGES-1), making COX-1/2 and mPGES-1 as key targets for traditional and new-age non-steroidal anti-inflammatory drugs (NSAIDs).[2-4] While COX-1 is constitutively expressed and is involved in mainly physiological processes (e.g. protection of gastric mucosa, platelet function), COX-2 is inducible during inflammation and overexpressed in many diseases. Direct inhibition of COX-1/2, thus, leads to inhibition of PG synthesis causing gastric mucosal damage and coagulation dysfunction.[5] Over the past five years (2019–2024), significant progress has been made in designing small-molecule COX-1/2 inhibitors with improved selectivity and new therapeutic applications. Researchers have focused on developing safer anti-inflammatory drugs and repurposing COX inhibitors for diseases like cancer and neurodegenerative disorders.[6-11]

The cardiovascular side effects of inducible COX-2 inhibition stem from the disturbed production of pro-thrombotic $TXA_2$ (produced by platelets, catalyzed by COX-1 in the vasculature), and anti-thrombotic $PGI_2$ (produced by vascular cells, catalyzed by COX-2).[12] Both prostanoids, $PGI_2$ and $TXA_2$, are important for maintaining the vascular tone, controlling blood pressure and blood flow throughout the body. The issues associated with COX-1/2 inhibitor usage led the researchers to explore the utility of mPGES-1 inhibition as a critical strategy for the development



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

of new-age, safer NSAIDs.[13,14] The enzyme mPGES-1, a microsomal, GSH-dependent, and inducible enzyme, was discovered in 1999.[15] It plays crucial roles in inflammation, hyperalgesia, pyresis as well as inflammatory conditions and diseases (including neurological) such as rheumatoid arthritis (RA), osteoarthritis (OA), periodontitis, atherosclerosis, inflammatory kidney damage, stroke, multiple sclerosis and Alzheimer's disease.[14] The inhibition of mPGES-1 will selectively affect $PGE_2$ production, while sparing $PGI_2$, $TXA_2$ and other prostanoids. Several literature reports have lately reviewed the progress in the design, discovery, preclinical and clinical applications of mPGES-1 inibitors.[16-19] Despite the potential and the progress in the drug discovery efforts, none of the mPGES-1 inhibitors have made it to the clinic. The literature cites several reasons for this discrepancy or slower clinical translation. The reasons listed include, but not limited to, the lack of suitable animal models, higher plasma protein binding due to higher lipophilicity, the chemotype-dependence of the inhibitors on binding with the target, poor bioavailability and pharmacokinetics and the redirection of the accumulated intermediate $PGH_2$ (direct result of mPGES-1 inhibition) for the biosynthesis of other prostanoids. A detailed account on the mPGES-1 inhibitor development programs including medicinal chemistry, biology and preclinical efforts at 'big pharma' is given somewhere else.[14] Learning from the obstacles hampering the quicker translation of mPGES-1 inhibitors to the clinic, a new thought process is required to tackle these issues for the development of new-age NSAIDs, particularly for the long-term treatment of inflammatory conditions such as RA and OA.

While the timing of inducible $PGE_2$ production by the prostanoid biosynthesis pathway along with the contribution of cytosolic $PGE_2$ synthase (cPGES) and mPGES-2 are poorly understood, the utility of selective mPGES-1 inhibitors can be debated, particularly for the long-term treatment. In most cells, a brief, immediate/early burst of $PGE_2$ follows $Ca^{2+}$-dependent activation of $cPLA_2$ and pre-existing COX-1; this burst is terminated quickly. A delayed/inducible wave then appears hours later, driven by the transcriptional up-regulation of COX-2 and mPGES-1 in response to cytokines (IL-1β, TNF-α), LPS, growth factors, etc. This temporal division is supported by coupling



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

preferences among terminal synthases: cPGES (PTGES3/p23) couples to COX-1 and mediates the immediate $PGE_2$ response, while mPGES-1 (PTGES) preferentially couples to COX-2 to generate the inducible $PGE_2$ during inflammation.[20,21] The biochemical basis has been shown repeatedly that - 1. cPGES/p23 is a constitutive, GSH-dependent enzyme that works with Hsp90 and is activated by CK2-dependent phosphorylation to convert COX-1-derived $PGH_2$ to $PGE_2$, not COX-2, giving fast $PGE_2$ pulses after acute stimuli. 2. By contrast, mPGES-1 is a Membrane-Associated Proteins in Eicosanoid and Glutathione metabolism (MAPEG) enzyme induced alongside COX-2 by pro-inflammatory cues and becomes the dominant $PGE_2$ source in the late phase.[22,23] Most inflammatory $PGE_2$ is not supplied by mPGES-2; its physiological niche remains uncertain and likely minor compared with mPGES-1.[24] In a modern phase 2 study of a selective mPGES-1 inhibitor, vipoglanstat (GS-248/BI-1029539), once-daily dosing fully inhibited mPGES-1 ex vivo, lowered urinary $PGE_2$ metabolites by ~57%, and raised $PGI_2$ metabolites by ~50%, exactly the intended pharmacodynamic profile. The trial was negative for efficacy in systemic sclerosis-related Raynaud's, but safety/tolerability and the prostanoid shunt were confirmed.[25]

In the present study, we reasoned if dual COX-2/mPGES-1 inhibition strategy would make any therapeutic and/or mechanistic sense and why dialing the selectivity slider (more COX-2 vs more mPGES-1) can be strategically useful across diseases. In most chronic inflammatory settings (arthritides, tumors, vascular injury), COX-2 and mPGES-1 are induced together, creating a high-flux COX-2 leading to higher $PGH_2$, thereby producing more $PGE_2$ via mPGES-1 circuit. Hitting both the supply (COX-2) and the specifier (mPGES-1) is, therefore, rational, rather than either alone.[26] In humans, the selective mPGES-1 inhibitor vipoglanstat lowered $PGE_2$ metabolites while increasing prostacyclin metabolites, the canonical reduced $PGE_2$/ increased $PGI_2$ shunt. In contrast, COX-2 inhibitors reduce both $PGE_2$ and $PGI_2$, a signature linked to cardiovascular hazard. A dual agent lets you tune these opposing effects with one scaffold/dose.[25,27] When mPGES-1 is inhibited (but not COX-2), $PGE_2$ falls and other prostanoids (e.g., $PGI_2$, sometimes



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

$TXB_2$/$PGF_2\alpha$) can rise (shunting). When COX-2 is inhibited, the whole $PGH_2$ pool drops ($PGI_2$ falls too). Combining them (in different ratios) lets one reduce pathogenic $PGE_2$, while containing unwanted shunts and avoiding $PGI_2$ loss.[28] Parallel COX-2/mPGES-1 up-regulation in human cancers and inflamed tissues underscores why dual hits may outperform single-node drugs when flux is extreme.[26] The class effect links coxibs to excess cardiovascular events, plausibly via $PGI_2$ suppression with intact platelet $TXA_2$ (COX-1). A dual profile biased toward mPGES-1 seeks anti-$PGE_2$ efficacy without the $PGI_2$ deficit of pure COX-2 blockade.

The real task is to tune a dual COX-2/mPGES-1 inhibitors in either of the three ways –

1. mPGES-1-dominant (strong mPGES-1, mild COX-2)
2. Balanced dual (moderate mPGES-1, moderate COX-2) or
3. COX-2-dominant (strong COX-2, mild mPGES-1)

Each of these strategies will have its own flavor. Inhibiting only mPGES-1 can redistribute $PGH_2$ to other prostanoids; adding controllable COX-2 inhibition prevents excessive rerouting while retaining $PGI_2$ if COX-2 is not over-suppressed. (Seen in mouse embryo fibroblast[29], mice, and human PD.) A recent report demonstrated multi-step shutdown of AA pathway with simultaneous inhibition of COX-2, 15-lipoxygenase and mPGES-1 with interesting outcomes, demonstrating the usefulness of such strategy.[30] Based on the sound scientific rationale described above, the AI-driven strategic search for dual COX-2/mPGES-1 inhibitor was initiated with a robust, meticulously crafted and curated pipeline incorporating reinforced learning (RL), in silico ADME/T profiling along with carefully designed structure-based virtual screening strategy backed by massive GPU computing power with the sole objective of finding novel chemical matter satisfying the requirements for binding with two sequential targets in the prostanoid biosynthesis pathway. We were flexible in modulating the selectivity element of the designs against one vs the other target. The aim was to come up with interesting, clean designs, with varied selectivity switches, which ultimately would allow us to validate the design hypotheses backed by the



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

experimental screening of the designs moving forward. The Target Product Profile (TPP) is given in Table 1.

## Materials and Methods

### *Computational Infrastructure and Methodological Framework*

All workloads were executed on a high-performance computing (HPC) environment with GPU acceleration (CUDA-based), optimized for deep-learning and molecular simulation tasks. The drug discovery campaign was executed on an end-to-end AI-native platform integrating proprietary OCSR.ai[31] engines with state-of-the-art open-source frameworks.

*Molecular Generation.* It employed a hybrid Reinforcement Learning - Generative Adversarial Networks (RL-GAN) architecture combining generative adversarial training with RL-guided optimization. The approach balances exploration of novel chemical space with targeted optimization toward desired pharmacological properties. De novo molecular design was performed using OCSR.ai Janak™, architecture enabling property-driven generation with multi-parameter reward functions (Figure 1), SMARTS constraints, and integrated patent novelty checking.

*Structure-Based Virtual Screening.* OCSR.ai ProteinLab unified protein structure prediction (Boltz-2, OpenFold2, MMseqs2 MSA pipeline), dual-engine docking (DiffDock, AutoDock Vina 1.2.7, https://pypi.org/project/vina/), and ultra-scale Deep Docking Protocol (DDP) scoring. The deep-learning scoring module evaluated protein-ligand compatibility with integrated ADMET-aware penalties (absorption, CYP liability, hERG, solubility, toxicity), enabling billion-scale screening with multi-objective optimization.

*Cheminformatics and Ligand Processing.* Molecular manipulation, conformer generation, and descriptor calculation employed RDKit 2025.09.1, https://pypi.org/project/rdkit/ and OpenBabel 3.1.1, https://anaconda.org/conda-forge/openbabel)



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

*ADMET and Retrosynthesis.* OCSR.ai Rasayan™ provided template-guided retrosynthetic planning (Figure 2) with vendor catalog integration and patent-aware route optimization. ADMET profiling integrated OCSR.ai Metabolites, KERMT, ChemProp (D-MPNN property prediction), Random Forest, XGBoost, LightGBM, OCSR.ai Qv-GenAI (quantum descriptors), Uni-Mol (https://github.com/deepmodeling/Uni-Mol) and Biomni (https://github.com/snap-stanford/Biomni)

*Database and Chemical Intelligence.* Virtual screening libraries comprised PubChem (119M compounds) (https://pubchem.ncbi.nlm.nih.gov/) and ChEMBL v36 (2.5M compounds) (https://chembl.blogspot.com/2025/09/chembl-36-is-out.html), PrimeKG (https://github.com/-mims-harvard/PrimeKG, https://www.nature.com/articles/s41597-023-01960-3), OCSR.ai Similarity Search (FPSim2, 1.7B Morgan fingerprint index) enabled sub-second Tanimoto queries. Indexing utilized OpenSearch and Google BigQuery.

*Patent Intelligence.* The system integrates Google Patents (via BigQuery) and the SureChEMBL patent database with OCSR.ai's proprietary structure-mining workflow. Powered by fine-tuned DECIMER (https://github.com/Kohulan/DECIMER-Image_Transformer) segmentation and Aarkiv document intelligence, it automates chemical structure extraction and prior-art analysis.

*Machine Learning Infrastructure.* All workflows operated on PyTorch (https://github.com/pytorch/pytorch/wiki/PyTorch-Versions), with CUDA acceleration, Scientific computing utilized NumPy (https://numpy.org/news/), Pandas (https://pandas.pydata.org/docs/whatsnew/index.html), and SciPy https://github.com/scipy/scipy/releases).



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

## List of Softwares used in the present study

| Category | Software |
|---|---|
| **Generation** | OCSR.ai Janak™ (RL-GAN) |
| **Screening** | OCSR.ai ProteinLab (Boltz-2, OpenFold2, MMseqs2, DiffDock, Vina 1.2.7, DDP scoring) |
| **Cheminformatics** | RDKit ecosystem 2025.09.1, OpenBabel 3.1.1, scikit-learn, SASCore, Meeko, RDChiral, MolBloom |
| **ADMET/Retrosynthesis** | OCSR.ai Rasayan, Metabolites, KERMT, ChemProp, Qv-GenAI, Biomni |
| **Databases** | PubChem (119M), ChEMBL v36 (2.5M), OCSR.ai Similarity Search (FPSim2), OpenSearch, Zinc22, PrimeKG |
| **Patent Intelligence** | SureChEMBL, Google Patents, DECIMER Segmentation, Aarkiv, OCSR.ai |
| **Structural Analysis** | PyMOL, GROMACS 2025.3 (AMBER) |
| **ML Stack** | PyTorch, NumPy, Pandas, SciPy, TensorFlow, Route-Distances, Pydantic, PyYAML, h5py, Matplotlib, Seaborn, ONNX |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

***Training Set Curation.*** A curated dataset of COX-2 inhibitors was assembled from ChEMBL v36 (Prostaglandin G/H synthase 2, COX-2 Target ID: CHEMBL230).[32,33] To ensure mechanistic and selectivity relevance, a series of sequential filters was applied (Table 2). First, only compounds with paired COX-1 and COX-2 activity data were retained. To emphasize differential enzyme inhibition, entries exhibiting a >10-fold selectivity for COX-2 over COX-1 were selected. Structural filtration was then applied to enrich the sulfonamide or methylsulfonyl pharmacophore, characteristic of canonical diarylheterocyclic COX-2 inhibitors. The resulting subset was further refined using the Lipinski Rule of Five to maintain drug-like physicochemical properties. Subsequent exclusion of molecules containing toxicophores (nitro groups, azides, Michael acceptors, thiols, epoxides) or PAINS motifs eliminated potential assay artifacts and reactive liabilities (data not shown). Withdrawn COX-2 inhibitors rofecoxib, valdecoxib, and lumiracoxib were removed for obvious reasons. All the operations were performed in OCSR.ai[31], a chemical intelligence platform.

To enhance selectivity learning and mitigate off-target liabilities, a counter-target training panel of 2,500 compounds was assembled. This included 500 COX-1 (ID: CHEMBL221[34]) inhibitors (to encode COX isoform discrimination), 1,000 hERG channel (ID: CHEMBL240[35]) binders (to penalize cardiotoxic scaffolds), 500 CYP2C9 (ID: CHEMBL3397[36]) inhibitors, and 500 CYP3A4 (ID: CHEMBL340[37]) inhibitors (to minimize potential metabolic drug–drug interactions).

***Molecular Generation via Reinforcement Learning (RL).*** A three-stage RL framework integrated with Generative Adversarial Networks (RL-GAN), titled Janak™[31] (molecular generation model inside OCSR.ai) was employed to design novel COX-2-selective chemotypes within the defined chemical space. This multi-objective generative strategy iteratively optimized molecular features for pharmacophoric fidelity, selectivity, and synthetic feasibility.

*Stage 1 (Epochs 0–30): Pharmacophore Learning.* Initially, the learning focus was on defining a COX-2-like chemical space using as reference structures or known binders, e.g., celecoxib and



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

etoricoxib (assigned a relative weight of 1.5x). The next step involved conditioning the model to enforce the presence of pharmacophoric features (e.g., sulfonamide or methylsulfonyl groups), weighted at 2.0x to ensure capture of the canonical COX-2 recognition motif. Global quality constraints were simultaneously applied, enforcing a quantitative estimate of drug-likeness, i.e., QED[38] ≥ 0.8) and a synthetic accessibility (SA) score ≤ 0.7.

*Stage 2 (Epochs 31–60): Multi-Target Balancing.* In the following stage, multi-objective optimization was attempted to retain COX-2 structural fidelity while minimizing off-target binding liabilities. COX-2 similarity weighting was maintained at 1.0x, whereas negative-learning penalties were activated for COX-1 (0.8x), hERG (0.7x), and CYP (0.6x) dissimilarities. The model then gradually evolved to yield molecular representations that preserved COX-2 pharmacophoric alignment while learning to avoid structural determinants linked to cardiotoxicity and metabolic instability.

*Stage 3 (Epochs 61–100): Final Optimization.* The final optimization stage encouraged chemical exploration and selectivity refinement. The COX-2 similarity constraint was relaxed to 0.8x, enabling broader scaffold diversification, while the sulfonamide requirement remained mandatory (weight 2.0x). Negative-learning penalties were simultaneously strengthened for COX-1 (1.0x), hERG (0.9x), and CYP (0.8x) profiles, coupled with maximum drug-likeness enforcement. The model was trained for 10,000 total epochs with a batch size of 256.

The model was constrained to favour synthetically accessible scaffolds, as assessed by internal retrosynthetic feasibility scores[31], as implemented in Rasayan™[31] (retrosynthesis model inside OCSR.ai), and fragment-based reconstructability metrics[33] (as implemented in Janak™ inside OCSR.ai). Rasayan™ is an intelligent retrosynthesis engine that deconstructs a target molecule into feasible reaction pathways and identifies optimal synthetic routes using a data-driven, template-guided search strategy. The system is trained on large-scale reaction knowledge



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

and commercially-available building blocks, enabling it to recommend routes that are not only chemically valid, but also cost-efficient and immediately actionable. By integrating vendor catalogs, Rasayan™ prioritizes starting materials based on price and availability, while its built-in patent intelligence flags routes already claimed in prior art, helping researchers avoid legal and IP obstacles early in the design cycle. The result is a retrosynthesis engine that delivers fast, diverse, economically optimized, and patent-aware synthetic strategies, empowering chemists to move from idea to actionable lab-plan with unprecedented clarity and speed.

***Initial Processing and Deduplication.*** A systematic multi-stage filtering protocol was applied to the generated molecular library to progressively enrich drug-like synthetically accessible, and COX-2-selective compounds. Each stage introduced more stringent physicochemical and target-based criteria, ensuring efficient triage of over 2.5 million initially valid structures into a high-quality, biologically relevant subset. Stage 1 was deduplication using canonical SMILES, followed by basic property filtering using QED, molecular weight, LogP, TPSA, H-bond donors, and acceptors, and rotatable bonds (Stage 2). Stage 3 consisted of strict drug-likeness filters wherein a composite six-filter screen was applied to ensure medicinal chemistry viability. The criteria included: 1. Mandatory sulfonamide/methylsulfonyl group; 2. QED ≥ 0.4 (raised threshold); 3. MW 300-450 Da (optimal COX-2 range); 4. Synthetic accessibility (SA) ≤ 6.0; 5. Absence of toxicophores (seven SMARTS patterns), and 6. ≤1 Lipinski violation.

### *Virtual Screening Campaign Using Deep Docking Protocol (DDP) inside ProteinLab*

A comprehensive structure-based virtual screening campaign was executed using the DDP, a Deep Learning-based Virtual Screening (inside ProteinLab, a part of OCSR.ai)[31], and an AI-accelerated molecular docking and prioritization framework integrating deep RL. The screening protocol was designed to evaluate both on-target efficacy and off-target liability, thereby refining the chemical library toward potent, selective, and pharmacologically safe COX-2 inhibitors.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

*Target Panel Screened.* The primary on-target evaluation was performed against human COX-2 using the rofecoxib-bound crystal structure as the structural template (PDB ID: 5KIR, Resolution: 2.7 Å).[39,40] To quantify isoform selectivity, all hits from initial virtual screening at COX-2 were evaluated against ovine COX-1 using the flurbiprofen-bound structure as reference (PDB ID: 3N8Z, Resolution: 2.90 Å).[41] The negative hits from COX-1 DDP were further screened at human microsomal prostaglandin $E_2$ synthase-1 (mPGES-1) structure (PDB ID: 4YL3[42]), which served as a secondary efficacy target, complementing the COX-2 inhibition profile through downstream prostanoid suppression. To mitigate single-template bias and to sample alternative active-site geometries, a second structure of mPGES-1 (PDB ID: 8PYV)[43] was used. It is a more recent human mPGES-1 model solved with GSH)-analog bound, using extended-wavelength anomalous scattering/natural-SAD phasing to localize heteroatoms precisely. This entry provides a sharp definition of the cofactor binding path and nearby polar network, offering an orthogonal receptor state that is less influenced by an inhibitor's induced fit.

Off-target safety profiling encompassed prostacyclin synthase (PTGIS) for cardiovascular risk mitigation, the hERG $K^+$ channel to address QT-prolongation potential, and CYP2C9 and CYP3A4 to monitor metabolic interaction liabilities. Carbonic anhydrase II (CA-II), a frequent serendipitous binder, was also included as a general off-target screen.

*Human COX-1 Comparative Modeling.* While the original workflow used the low-resolution, ovine COX-1 structure with ligand (PDB ID: 3N8Z), the need for better human COX-1 structure for virtual screening was realized. The lone unliganded human COX-1 crystal structure (PDB ID: 6Y3C, resolution 3.36 Å)[44] was deemed unsuitable for our analyses due to its limited resolution and incomplete active-site definition, necessitating comparative protein modeling based on higher-quality ovine COX-1 templates (3N8Z[41], 3N8X[45]). To ensure direct comparability between COX-1 and COX-2 binding analyses, a high-quality comparative protein model of human COX-1 was predicted. The Boltz-2 structure prediction framework[46-48] was employed using the human COX-1 FASTA sequence. Both ovine COX-1 crystal structures, 3N8Z[41] and 3N8X[45] were used as



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

backbone templates, selected for their high sequence identity (>95%) and excellent resolution of the active-site region, including the canonical catalytic triad (Tyr385, Ser530, Arg120). Template alignment was optimized to preserve the spatial arrangement of the cyclooxygenase channel, membrane-binding helices, and heme-binding domain, ensuring accurate replication of the COX-1 catalytic topology observed in mammalian systems. Next, the predicted structure was superposed onto human COX-2 structure (PDB ID: 5KIR[40]) with rofecoxib (ligand) and Co-heme (cofactor). The COX-2 protein structure was deleted and the ligand and cofactor were retained into their corresponding binding sites in the predicted model. His388 N was manually bonded to the metal centre of heme. The complex was further subjected to energy minimization with the ligand and the bonded cofactor. This complex was used for terminal docking studies with the top-performing 'clean' hits.

***Medicinal Chemistry Optimization***

Medicinal chemistry optimization was performed on the 16,984 candidates emerging from the virtual screening and liability filtering cascade to refine physicochemical, structural, and reactivity profiles consistent with clinically tractable COX-2/mPGES-1 inhibitors. A series of sequential medicinal chemistry filters was applied to balance drug-likeness, permeability, metabolic stability, and synthetic feasibility in line with Lipinski's rule of five and Veber's criteria. The filtering pipeline included molecular property constraints as listed in Table 4. All descriptors were calculated using RDKit v2025.09.1[49], and filters were applied in the following sequence:

1. Molecular Weight
2. Lipophilicity
3. TPSA
4. Rotatable Bonds
5. Aromatic Rings
6. H-Bond Donors and Acceptors



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

7. Structural Alert Removal: Molecules containing halogens (-Cl, -Br, -I), imines, *p*-anilides, or Michael acceptors were excluded to minimize potential reactivity, genotoxicity, or metabolic instability.

At each stage, the number of molecules retained was recorded to quantify attrition and identify major bottlenecks in the optimization sequence.

***Novelty and Patent Analysis***

To evaluate the chemical novelty and intellectual property freedom of the optimized compound library, all molecules passing the medicinal chemistry filters (4,567 compounds) were subjected to multi-tiered structural novelty and patent screening using OCSR.ai's proprietary similarity engine integrated with large-scale public and commercial databases.[31]

*Chemical Novelty Assessment.* Novelty assessment was conducted against a database of ~1.7 billion unique chemical entities drawn from public and proprietary repositories. Structural similarity was calculated using Tanimoto coefficients based on Morgan circular fingerprints[50] (ECFP4, radius = 2, 2048 bits). Two thresholds were applied to capture different levels of structural relatedness:

1. High-stringency filter (similarity ≥ 0.8) to identify close analogs or scaffold reuses;
2. Moderate-stringency filter (similarity ≥ 0.7) to capture remote analogs or substructure-level resemblance.

Compounds below the 0.7 similarity cutoff were classified as chemically novel scaffolds

*Computational Patent Landscape Analysis.* Patent evaluation was performed using the OCSR.ai Patent Intelligence Engine, which cross-references molecular fingerprints against 44.3 million global patent records aggregated from SureChEMBL2.0[51] and Google Patents Public Datasets on BigQuery.[52] The screening detects exact and near-exact structure matches, Markush claims, and chemically encoded patent representations (SMILES, InChI, and substructure forms).



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

*Quality Control.* To ensure the integrity of novelty and IP results, an additional structural quality control (QC) filter was applied to remove molecules with undefined stereochemistry, valence errors, or disconnected fragments. Post-QC molecules were retained for further prioritization in the final generative design stage.

### *ADMET Profiling and Clean Candidate Selection*

The final 4,219 quality-controlled compounds were subjected to an extensive in silico ADMET evaluation to identify chemically tractable, pharmacologically safe, and synthetically feasible candidates suitable for lead optimization. The assessment integrated physicochemical, toxicological, and developability parameters within a unified multi-endpoint prediction framework, trained over Kinetic GROVER Multi-Task (KERMT) graph neural network (GNN)[53] and fine-tuned TxGemma 27B models.[54]

*Properties Evaluated.* Each compound was profiled for:

1. Aqueous solubility (logS)
2. AMES mutagenicity
3. hERG channel inhibition
4. Drug-induced liver injury (DILI), and
5. SAScore (as implemented in RDKit)

*Selection Criteria.* Compounds were classified as clean only if they simultaneously met the following criteria: AMES: Non-mutagenic prediction outcome; hERG: Low-risk classification; DILI: Low to moderate risk (no high-risk predictions permitted); Solubility: logS> –5.0, ensuring adequate aqueous solubility for human intestinal absorption and oral bioavailability.

### *Final Set Expansion for Validation*

To enhance the likelihood of experimental success and preserve scaffold diversity during downstream validation, the final clean compound pool was selectively expanded. The starting library comprised 1,939 candidates that met all ADMET and synthetic-accessibility thresholds



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

established during the previous step. A deliberate, chemistry-driven relaxation strategy was applied to recover borderline yet promising molecules. Compounds exhibiting a single moderate ADMET flag, for example, marginal hERG or DILI risk predictions, were re-included if the associated liability was considered chemically optimizable. Likewise, the aqueous-solubility threshold was slightly relaxed (from logS > -5.0 to ≥ -5.5) to admit analogs with acceptable formulation potential. Finally, all structures that satisfied core drug-likeness filters (Lipinski, Veber, and Ghose compliance) were retained, ensuring preservation of fundamental developability characteristics. Following this controlled expansion, the library increased to 3,921 molecules, representing a rational compromise between predictive purity and exploratory chemical breadth. This curated set was subsequently advanced as the input for the orthogonal docking and validation campaigns.

***Orthogonal Validation Through Molecular Docking***

Physics-based validation employed AutoDock Vina-GPU (version 1.6)[55] with default scoring and optimization parameters, except for the search space dimensions, which were set to 24 × 24 × 24 Å centered on the active-site coordinates. Receptor preparation involved removal of all crystallographic water molecules while retaining the heme cofactor, and conversion of protein structures into a rigid PDBQT format. Ligand preparation was carried out using RDKit's ETKDG algorithm[56] for low-energy conformer generation followed by MMFF94 force-field minimization. The optimized geometries were converted to PDBQT format using Open Babel. Binding-site grid coordinates were defined from the centroid of the co-crystallized ligand in each structure. Docking validation and scoring were performed across total of five protein structures (in stagewise manner, initially three- followed by five-structures), comprising four COX-2 conformations, 5KIR (rofecoxib), 5IKR[57] (mefenamic acid), 4PH9[58] (ibuprofen), and 3NT1[59] (naproxen), and a COX-1 control (3N8Z[60], flurbiprofen).



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

*Gold-Standard Protocol Validation.* Prior to prospective screening, the docking workflow was rigorously validated for pose reproduction and scoring fidelity against crystallographic benchmarks. For each reference complex, the co-crystallized ligand was extracted from the PDB entry, the receptor was converted to an apo form by ligand removal, and the ligand was re-docked using the production parameters intended for the campaign. Performance was assessed by (i) heavy-atom RMSD between the redocked and experimental poses and (ii) agreement between predicted and experimental binding affinities. Success criteria were predefined as: RMSD < 2.0 Å (pose accuracy) and |ΔAffinity| ≤ 1.0 kcal/mol (scoring accuracy).

*Primary Production Screening (Three-Structure Campaign)*. A high-throughput molecular docking campaign was executed to evaluate the full 3,921-molecule validation library against multiple cyclooxygenase structures. The objective was to quantify binding strength, reproducibility, and COX-2/COX-1 selectivity across orthogonal crystallographic templates.

*Campaign design.* Each compound was docked into three validated receptor structures: COX-2 (PDB 5KIR, rofecoxib-bound), COX-2 (PDB 5IKR, mefenamic-acid-bound), and COX-1 (PDB 3N8Z, flurbiprofen-bound). In total, 11,763 docking jobs (3,921x3) were executed using a distributed infrastructure of five GPU servers with ten concurrent workers. The full campaign completed in 96 min with a 100% success rate, demonstrating robust computational throughput and reproducibility.

*High-Selectivity Filtering.* To prioritize ligands with pronounced COX-2 selectivity, a high-selectivity filter was applied, retaining molecules exhibiting a binding energy difference >2.0 kcal.mol$^{-1}$ relative to COX-1 (corresponding to >10-fold selectivity on a thermodynamic scale).

*Conformational Expansion and Ensemble Docking.* To capture the conformational heterogeneity of COX-2, a conformational expansion screening was performed beyond the initial two-structure docking set. The campaign commenced with 2,035 pre-filtered highly-selective ligands. Two additional COX-2 crystal structures, 4PH9 (ibuprofen-bound) and 3NT1 (naproxen-bound), were incorporated, resulting in 4,070 docking jobs (2,035 × 2) executed on AutoDock Vina-GPU v1.6



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

under default parameters. The total computational duration was 33.7 min, with a 100% success rate.

The binding modes in the binding pockets of the studied targets COX-1/2 and mPGES-1 were thoroughly analyzed for interactions with the site residues. The 2D interaction diagrams were generated using the respective enzyme 2D interaction diagrams were generated using LigPlot+ v2.3.[61] The PDB files of the complexes were used for the analysis.

**Results and Discussion**

Our previous work on computational repositioning of approved drugs in search of safer NSAIDs, served as a conceptual and methodological foundation for the present work on dual COX-2/mPGES-1 inhibitors.[62] The earlier study employed diclofenac (a non-selective COX inhibitor), celecoxib (a COX-2 selective inhibitor), and a potent mPGES-1 inhibitor as query molecules for ligand- and structure-based virtual screening, with particular emphasis on shape- and electrostatics-based analyses. With this background, the therapeutic target product profile (TPP) established for the present study defines the optimal translational parameters for next-generation dual COX-2/mPGES-1 inhibitors (Table 1). The core design principle was to achieve high COX-2 selectivity (>10-fold over COX-1) to minimize gastrointestinal toxicity commonly associated with non-selective NSAIDs. In parallel, maintenance of positive mPGES-1 binding activity was pursued to ensure downstream suppression of $PGE_2$, thereby complementing COX-2 blockade and achieving broader anti-inflammatory efficacy. Absence of hERG channel liability was treated as a non-negotiable pharmacological filter due to potential cardiotoxicity risk. Drug-likeness and oral bioavailability criteria were embedded into the molecular design to promote favorable ADME behavior, targeting clinical utility through, preferably once-daily oral dosing potential. Equally critical was freedom-to-operate (FTO), ensured through the generation of novel, patent-free chemical scaffolds, thereby providing a secure IP landscape for future development. Collectively, this TPP provided the strategic framework linking molecular design with translational



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

endpoints, guiding compound prioritization toward candidates that are efficacious, safe, orally active, and commercially defensible.

In the current study, that conceptual framework has been expanded from drug repositioning to de novo molecular generation and selectivity engineering. The earlier insight that mPGES-1 inhibition offers a safer downstream alternative to COX blockade, is operationalized here through a reinforcement learning generative approach designed to produce compounds with concurrent COX-2 and mPGES-1 activity, while penalizing COX-1, hERG, and CYP liabilities. Where the earlier work focused on repurposing existing scaffolds for potential mPGES-1 inhibition, the present study advances to designing new, patent-free, dual-active chemotypes validated across both inhibitor- and cofactor-bound mPGES-1 conformations, alongside multiple COX-2 structures. The two studies together thus trace a clear evolutionary arc, from computational repurposing of known drugs to AI-guided generation of safer, mechanistically optimized dual inhibitors that integrate the lessons of COX-2 selectivity and mPGES-1 safety for next-generation anti-inflammatory therapeutics.

***Training Set Curation.*** The initial extraction from ChEMBL comprised 4,004 bioactivity records corresponding to 2,983 unique small molecules annotated with COX-2 inhibitory activity. Assay type included was $IC_{50}$; activity range included 0.1-10,000 nM. Activity-based weighting was applied in the following manner: 1. Highly potent ($IC_{50}$ <100 nM): 892 compounds, weight 3.0; 2. Potent ($IC_{50}$:100-<1,000 nM): 1,456 compounds, weight 2.0, and 3. Active ($IC_{50}$:1-10 μM): 1,656 compounds, weight 1.0. Table 2 lists the sequential filters and the outcome during training set curation.

The hits (1,362 compounds) with paired COX-1 and COX-2 activity data (retained in the very first filter mean) showed mean selectivity: 4,277-fold, and range: 10- to >10,000-fold. All the hits (1,069 compounds) with ≥10-fold selectivity for COX-2 over COX-1 (Filter: selectivity threshold) possessed same pharmacophoric elements (SMARTS patterns: S(=O)(=O)N or S(=O)(=O)C).



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

The final training set comprised 308 compounds, including 305 ChEMBL entries representing 218 unique Murcko frameworks (maximum two entries per Murcko scaffold for representing structural diversity), supplemented with three FDA-approved reference coxibs (celecoxib, etoricoxib, parecoxib). The carefully curated training dataset demonstrated >10-fold COX-2 selectivity, ensuring a pharmacologically coherent dataset for model training. On similar lines, the dataset for selectivity learning and mitigating off-target liabilities was used for multiparameter optimization. These meticulously curated datasets provided a balanced representation of desired COX-2 selectivity and undesirable off-target pharmacology, forming the foundation for subsequent model development and validation.

***Molecular Generation via Reinforcement Learning (RL).*** The integration of RL with GANs (Janak™[33] inside OCSR.ai) represents a powerful paradigm shift in *de novo* drug design and molecular generation.[63] The platform enables researchers to train a fully-customized molecular generative model directly from their own SMILES datasets. By learning scaffold distributions and structure-activity relationships from the input chemistry, the system applies a three-stage RL framework integrated with an RL-GAN architecture to explore, optimize, and expand the desired chemical space. Users can precisely steer generation through property-based reward functions, SMARTS-level structural constraints, and negative datasets that penalize toxic or unwanted chemotypes. Once trained, the model produces a high-volume library of novel, valid, drug-like molecules that can be filtered, ranked, checked for novelty against an integrated patents database, and seamlessly exported for downstream discovery workflows. This hybrid approach addresses critical challenges in generating discrete molecular structures while optimizing desired chemical and biological properties.[64] It integrates adversarial learning with reinforcement-driven molecular optimization to navigate chemical space efficiently toward desired pharmacological properties. Within this framework, the generator network functions as the *actor*, iteratively constructing molecular structures in a policy-driven manner, while the discriminator network



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

operates as the *critic*, distinguishing synthetically feasible and property-aligned molecules from non-optimal candidates. A reward function quantitatively encodes task-specific objectives such as binding affinity, selectivity, or drug-likeness, and serves as the guiding signal for adaptive learning. The policy-gradient mechanism enables gradient propagation through discrete molecular representations, allowing the generator to refine its sampling strategy based on the critic's feedback. Collectively, this closed feedback loop enables *de novo* molecular design to evolve dynamically toward chemical motifs that optimize multi-parameter objectives, bridging generative modeling with rational, property-driven lead discovery. A recently published article highlighted the application of transformer-based architecture, Taiga, for molecular generation with desired properties.[65] A similar article used graph transformer-based RL for scaffold-constrained molecular design.[66]

In Step 1 of the workflow, drug-likeness was quantified using the QED index implemented in RDKit[43], which integrates eight key physicochemical descriptors derived from approved oral drugs, including molecular weight, LogP, H-bond donors and acceptors, rotatable bonds, polar surface area, aromatic ring count, and structural alerts. Each property contributes to a desirability function normalized between 0 and 1, with the composite QED value representing the overall probability of compliance with favorable drug-like space. Molecules exhibiting a QED score ≥0.8 were classified as 'highly drug-like', consistent with thresholds commonly observed among clinically successful small-molecule therapeutics. This stringent cutoff ensured that subsequent analyses focused on candidates with optimal physicochemical balance, reflecting favorable ADMET profiles suitable for progression in medicinal chemistry optimization. In the present study, the stagewise workflow comprising of pharmacophore learning, multi-target balancing and final optimization, as implemented in Janak™[31], yielded a library of approximately 2.56 x $10^6$ virtual molecules, each conforming to the desired pharmacophore, drug-likeness, and selectivity criteria, representing a robust in silico expansion of the COX-2-privileged chemical space.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

***Generation Strategy.*** Molecule generation was performed using a three-stage RL framework designed for progressive optimization of both pharmacophoric and ADMET-driven objectives. Each stage incorporated adaptive weighting of positive and negative learning components to iteratively balance COX-2 selectivity, structural feasibility, and drug-likeness. The staged RL approach effectively guided the molecular generator toward a COX-2-specific and synthetically accessible chemical domain. During Stage 1, the model rapidly converged on the characteristic sulfonamide/methylsulfonyl motif central to COX-2 inhibition. High QED scores (≥ 0.8) and favorable synthetic accessibility (SA ≤ 0.7) confirmed that the pharmacophoric bias successfully constrained the search to drug-like space. The elevated weighting of celecoxib and etoricoxib references established an appropriate structural anchor for model convergence.

By Stage 2, activation of negative-learning components achieved clear differentiation between desired and undesired regions of chemical space. Molecules with structural resemblance to COX-1 inhibitors, hERG blockers, and CYP-binding scaffolds were systematically deprioritized, while COX-2-aligned analogs were retained. The balance of positive and negative reinforcement preserved COX-2 chemical identity while reducing the likelihood of off-target effects, mimicking the nuanced selectivity challenges encountered in NSAID design.

In Stage 3, relaxation of the COX-2 similarity constraint and simultaneous strengthening of off-target penalties promoted controlled scaffold diversification. This transition allowed the model to explore novel ring systems and substituent patterns beyond the classical diarylheterocyclic cores while preserving the essential sulfonamide anchor. The final library of 2.56 million chemically-valid molecules, as verified by RDKit sanitization, demonstrated both synthetic feasibility and pharmacophoric fidelity, representing a robust expansion of the COX-2-privileged space. Overall, the RL-GAN framework successfully captured the structure-activity logic underlying COX-2 selectivity and produced a chemically coherent, experimentally actionable molecular set. The 100% validity rate, balanced property distributions, and stable convergence score (0.50) collectively attest to the success and robustness of the generative learning protocol.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

***Initial Processing and Deduplication.*** During Stage 1 (deduplication), a total of 99,044 redundant entries (3.9%) were removed, yielding 2,460,914 unique molecules for subsequent filtering. In Stage 2 (Basic Property Filtering), filters enforced the presence of a sulfonamide or methylsulfonyl group, ≤1 Lipinski violation, and a minimum QED threshold of 0.3. This basic screen retained 98.1% of the dataset with a mean QED of 0.625. Nearly all molecules (99.1%) contained the targeted sulfonamide motif, resulting in 2,401,236 compounds advancing to the next stage. Strict Drug-Likeness Filter (Stage 3) retained 85.3% of molecules from Stage 2, producing 2,047,122 compounds with QED values ranging from 0.400–0.948, MW from 300.2–450.0 Da, and SA scores between 1.44-6.00. The outcome reflects a well-balanced collection of chemically-diverse, yet synthesizable scaffolds consistent with oral drug-like space.

### *Virtual Screening Campaign Using DDP inside ProteinLab*

At the core of ProteinLab[31] is a deep-learning scoring module that evaluates protein-ligand compatibility using joint structural embeddings of the binding site and candidate molecule. Trained on large-scale interaction data, this model captures key biochemical determinants of binding, including a) Hydrophobic complementarity, b) H-bonding networks, c) Electrostatics fit, d) Shape complementarity, and e) Pocket-specific pharmacophore signatures. Because scoring does not require full docking for every compound, the system can screen millions to billions of molecules in ultra-high throughput manner. To ensure that top-ranked molecules are not only strong binders but viable drug candidates, ProteinLab integrates ADMET-aware reward and penalty terms, adjusting scores based on predicted absorption and permeability, metabolic liability (e.g., CYP interactions), hERG-mediated cardiotoxicity, solubility and physicochemical balance, clearance and bioavailability risk, and toxicophores and safety alerts. This creates a hybrid, multi-objective scoring function that simultaneously optimizes for binding strength, selectivity, developability, and safety.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

ProteinLab[31] therefore, prioritizes ligands that are strong binders, pocket-specific and selective, pharmaceutically developable (no major ADMET liabilities), safe and non-toxic and scalable for downstream medicinal chemistry. This approach goes far beyond traditional docking[67], enabling early triaging of chemical space at massive scale while reducing downstream experimental failure. Following screening and scoring, ProteinLab[31] DDP performs multi-parameter hit ranking, highlighting molecules with the most favorable balance of binding affinity, interaction quality, pharmacokinetic suitability, and safety profile. All results can be exported for molecular dynamics simulations, synthesis and retrosynthetic planning, and in vitro experimental validation.

In the present work, the virtual screening cascade efficiently narrowed the initial 2,047,122 drug-like compounds to a final subset of 16,984 high-confidence candidates through progressive multitarget filtering (Table 3). The initial COX-2 screen identified 238,689 active compounds (11.7% hit rate), establishing the baseline for potency selection. When these compounds were cross-screened against COX-1, all were successfully evaluated for isoform discrimination. However, the selectivity filter proved highly restrictive: only 33,327 compounds (14.0%) achieved the required >10-fold COX-2 preference, revealing that 86% of COX-2 binders lacked sufficient COX-1 discrimination, a key bottleneck in selective inhibitor design. Among these 33,327 selective compounds, mPGES-1 screening retained 33,143 molecules (99.4%), confirming strong overlap between COX-2-active scaffolds and the downstream mPGES-1 binding pocket. The subsequent off-target liability filters systematically eliminated molecules with predicted safety risks. The PTGIS filter removed 10% of candidates for prostacyclin interference, while the hERG filter, the most stringent, eliminated 8,337 molecules (28%) due to potential cardiotoxicity. Further refinement through CYP2C9 (10% removal), CYP3A4 (10% removal), and CA-II (2.2% removal) yielded the final, safety-optimized subset. The target CA II was included in the off-target panel as a known serendipitous binder of primary sulfonamide motifs, which can inadvertently engage the zinc-metalloenzyme active site via coordination of the sulfonamide NH to the catalytic $Zn^{2+}$. For



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

example, previously, we have reported that sulfonyl semicarbazide derivatives, bearing sulfonamide moieties, inhibited human CA II in the low-nanomolar range, underscoring the risk of off-target CA II engagement when primary sulfonamides are used.[68,69]

The critical observation from this campaign was that COX-2/COX-1 selectivity remained the most challenging discriminant in the pipeline, with most actives failing to maintain adequate isoform separation despite favorable COX-2 binding. This aligns with prior structure-based findings showing that subtle variations in the COX-1 side pocket and secondary channel geometry can significantly reduce selectivity even among potent COX-2 ligands. Nevertheless, the final 16,984-compound panel preserved both efficacy (COX-2/mPGES-1) and safety (low off-target liabilities), constituting a refined virtual library suitable for ensemble docking validation and ADMET simulation in subsequent phases (Table 3).

***Medicinal Chemistry Optimization***

Sequential physicochemical filtering led to reduction of the compound set size from 16,984 to 4,567 molecules (26.9% survival), representing progressive enrichment toward chemically robust, drug-like scaffolds suitable for lead optimization. The first two filters, Mol Wt and lipophilicity were moderately restrictive, retaining 91.6% and 82.1% of compounds, respectively. These criteria ensured balanced hydrophobicity compatible with the COX-2 secondary pocket while excluding overly lipophilic analogs prone to metabolic instability. The TPSA filter proved one of the most selective steps, eliminating 3,311 molecules (19.5% loss) and indicating that many candidates possessed excessive polarity or insufficient H-bonding surface area. Aromatic ring limitation further removed 2,442 molecules (14.4% loss), reflecting a reduction in overly rigid or polyaromatic frameworks often associated with poor solubility and cardiotoxic liabilities. Minor attrition occurred in the rotatable bond, H-bond donor, and acceptor filters, collectively maintaining over 60% of the dataset. The final structural alert screening produced the most significant compositional refinement: halogen elimination alone removed 2,814 molecules (16.6% loss),



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

while smaller subsets containing imines (0.2%), *p*-anilides (1.1%), or Michael acceptors (0.1 %) were excluded due to potential reactivity and toxicity concerns.

Cumulatively, the optimization process favored medium-sized, moderately lipophilic, and synthetically accessible sulfonamide/methylsulfonyl-containing scaffolds, consistent with known COX-2 inhibitor pharmacophore dimensions. The major attrition points, halogen elimination, TPSA refinement, and aromatic complexity reduction represent the critical medicinal chemistry constraints shaping the final compound set (Table 4). The resulting 4,567 molecules define a refined, in silico-optimized chemical space possessing balanced polarity, limited aromatic density, and absence of reactive liabilities, ideally positioned for subsequent lead docking validation, SAR exploration, and in vitro profiling.

### *Novelty and Patent Analysis*

The novelty and IP assessment confirmed that the optimized chemical library represented a highly original, patent-free chemical space with strong potential for intellectual property protection and commercial development.

*Novelty Analysis.* At the 0.8 similarity threshold (high stringency), 4,169 compounds (98.8%) were identified as structurally novel, with only 50 compounds (1.2%) showing high similarity to existing molecules in global databases. When the threshold was relaxed to 0.7 (moderate stringency), 3,921 compounds (92.9%) remained novel, while 298 (7.1%) exhibited moderate-level similarity, typically limited to shared sulfonamide or diaryl-heterocyclic motifs common in COX-2 inhibitor scaffolds (Table 5). These results indicate that the RL and medicinal chemistry filters successfully explored underrepresented regions of chemical space, yielding scaffolds with minimal overlap to known NSAID chemotypes.

*Computational Patent Landscape Analysis.* To assess the novelty space of the generated molecules, we performed a comprehensive computational patent search across three major databases: PubChem (>100 million compounds), Google Patents (global filings), and



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

SureChEMBL (>44.3 million patent-extracted structures). Molecular fingerprint similarity analysis used Morgan circular fingerprints (ECFP4, radius=2, 2048 bits) and Pubchem similarity search via REST API with a Tanimoto similarity threshold of 0.8 to identify structurally similar compounds. All 2035 final candidates showed Tanimoto similarity <0.8 to any patent-disclosed structure, suggesting structural novelty in the searched databases. Additionally, substructure searches for common COX-2 pharmacophores (diarylsulfonamide and diaryl-methylsulfonyl motifs) confirmed no exact scaffold matches in the patent literature accessed as of October 2025. This computational analysis represents an initial patent landscape assessment and should not be construed as a comprehensive FTO opinion. Patent coverage includes composition of matter claims, Markush structures, use patents, and formulation claims that may not be captured by structure-similarity searches alone. Our analysis focused on granted patents and published applications available in the searched databases, which may have 18-month publication delays for recent filings and incomplete coverage of Asian patent offices (China, Japan, Korea). Despite these limitations, the computational patent screening provides strong preliminary evidence that our AI-generated molecules occupy novel chemical space distinct from existing COX-2 inhibitor patents, supporting their potential as innovative therapeutic candidates.

*Quality Control and Final Dataset.* After removing low-quality or structurally incomplete entries, 4,219 molecules (92.4% of the input set) passed the quality control filters, forming the final high-confidence, novel, and IP-clean compound library.

### *ADMET Profiling and Clean Candidate Selection*

*ADMET Prediction Benchmark (KERMT)*. A total of eight ADMET prediction models were trained using the KERMT framework with transfer learning from GROVER-Large and scaffold-based splits for robust generalization. KERMT[53], a pretrained GNN model for molecular property prediction, is an enhanced reimplementation of the GROVER model.[71] The KERMT implementation uses PyTorch Distributed Data Parallel for distributed pretraining, automates



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

hyperparameter tuning, and accelerates finetuning and prediction using cuik-molmaker.[70] The models demonstrated strong predictive performance across physicochemical, efficacy, and toxicity endpoints, achieving high classification Area Under the Receiver Operating Characteristic Curve, i.e., AUROC and low regression error on held-out scaffolds (Table 6). These results confirmed the reliability of our ADMET engine for virtual screening and early risk assessment. These benchmark results highlight the effectiveness of our KERMT-based ADMET models in capturing diverse biochemical properties and safety liabilities, enabling accurate multi-parameter prioritization during hit triage. Similar prediction benchmark using fine-tuned TxGemma 27B model is not given.

Application of these filters yielded 1,939 clean candidates (46% of 4,219) that satisfied all ADMET and physicochemical thresholds. The selected subset demonstrated favorable developability, with a mean SAScore of 2.68, indicative of high synthetic accessibility. Notably, 75.7% of these candidates exhibited SAScore <3.0, denoting relative ease of synthesis using conventional medicinal chemistry workflows. This clean candidate pool represents the most promising balance of COX-2 selectivity, drug-likeness, and developability, forming the foundation for subsequent docking validation, binding-pose analysis, and lead prioritization studies.

***Final Set Expansion for Validation***

The strategic expansion of the clean candidate set yielded a two-fold increase in chemical diversity while maintaining overall pharmacokinetic integrity. Inclusion of compounds with a single moderate ADMET deviation proved particularly valuable: several borderline cases exhibited sub-threshold risks (e.g., modest solubility or weak CYP binding) that could be mitigated through minor structural optimization or formulation adjustment. Relaxation of the solubility constraint contributed to a richer distribution of hydrophilic and lipophilic chemotypes, broadening the physicochemical space explored during validation. Importantly, the re-admitted molecules continued to satisfy core drug-likeness criteria, confirming that the expansion did not compromise



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

medicinal chemistry tractability. This approach provided a balanced validation library (n = 3,921), large enough to capture structural and pharmacophoric diversity yet, sufficiently filtered to maintain synthetic feasibility and safety confidence. The expanded dataset thus offered a robust foundation for the subsequent stages in the workflow.

### *Orthogonal Validation Through Molecular Docking*

*Gold-Standard Protocol Validation.* Five COX isoform complexes (four COX-2 structures representing distinct binding site conformations and one COX-1) spanning distinct chemotypes were evaluated for protocol validation. The results are summarized in Table 7. The protocol achieved a 5/5 (100%) success rate against the dual criteria, with a mean RMSD of 0.762 Å. Affinity agreement was exact for 3/5 cases and within 0.5 kcal/mol for 2/5 cases, confirming suitability of the scoring and sampling settings for prospective application. During early testing of 3NT1, use of web-tool-supplied ligand coordinates introduced 60-80 Å displacements, underscoring the necessity of generating the reference ligand coordinates directly from the co-crystal structure. All validations reported herein utilized crystallographically extracted ligand poses to prevent coordinate misregistration.

*Primary Production Screening (Three-Structure Campaign)*.

*COX-2 binding performance.* When the best score from the two COX-2 structures was considered per ligand, the mean predicted affinity was −8.14 kcal/mol (median = −8.20 kcal/mol; range = −12.70 to −3.00; SD = 1.15). These results indicate strong and consistent engagement of the COX-2 active site across structurally diverse scaffolds.

*COX-1 binding performance.* For the COX-1 reference structure (3N8Z), the mean affinity was −5.74 kcal · $mol^{-1}$ (median = −5.60; range = −9.30 to −0.70; SD = 1.08), confirming lower binding strength relative to COX-2 and supporting the enrichment of selective chemotypes in the generated library.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

*Selectivity analysis.* Selectivity ($\Delta G_{sel}$) was calculated as:

$$\Delta G_{sel} = E_{bind}(COX-1) - E_{bind}(COX-2)_{best}$$

where positive values indicate COX-2 preference. The mean selectivity was 2.40 kcal/mol (median = 2.10; maximum = 9.2). Table 8 summarized the isoform-selective performance in three-structure campaign. Overall, 85.7% of the screened library exhibited measurable COX-2 bias, and over half (54.3%) were strongly COX-2-selective, validating the effectiveness of the prior reinforcement-learning and filtering stages in sculpting a COX-2-privileged chemical space.

*High-Selectivity Filtering*

From an initial 3,921 docked compounds, 2,035 molecules (51.9% retention) satisfied this criterion and advanced to ensemble docking. Application of the filter resulted in marked enrichment of selectivity and affinity metrics: the mean COX-2 selectivity increased from 2.40 to 3.50 kcal.mol$^{-1}$, while the mean COX-2 binding energy improved from -8.14 to -8.88 kcal.mol$^{-1}$. This 2,035-molecule refined dataset formed the basis for subsequent conformational expansion and cross-structure screening.

*Conformational Expansion and Ensemble Docking*

The total computational duration with additional two structures was 33.7 min, with a 100% success rate. For 4PH9, the mean binding energy was –8.02 kcal mol$^{-1}$ (median = –8.00; range = –11.50 to –4.40 kcal.mol$^{-1}$). For 3NT1, the mean binding energy was –7.50 kcal.mol$^{-1}$ (median = –7.50; range = –10.10 to –3.20 kcal.mol$^{-1}$). Ensemble docking across all four COX-2 conformations (5KIR, 5IKR, 4PH9, 3NT1) improved overall affinity and selectivity metrics. A total of 259 molecules (12.7%) displayed enhanced binding in the newly added structures, 156 molecules (7.7%) in 4PH9 and 139 (6.8%) in 3NT1. The mean best COX-2 affinity improved from –8.46 to –8.93 kcal.mol$^{-1}$ (+0.47 kcal.mol$^{-1}$), while the mean selectivity index (COX-1 vs COX-2) increased from 3.08 to 3.55 kcal mol$^{-1}$ (+15.3%), reaching a maximum selectivity of 9.2 kcal.mol$^{-1}$.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

*Final Characterization of the Validated COX-2 Ligand Set*

Binding-mode distribution (Table 9) revealed that nearly half of all ligands (49.0%) preferentially bound to the 5IKR (mefenamic acid) conformation, while 38.4% showed highest affinity for 5KIR (rofecoxib). The remaining molecules favored the 4PH9 (ibuprofen) and 3NT1 (naproxen) structures, accounting for 7.7% and 4.9%, respectively. This distribution indicated that the mefenamic-acid-like pocket geometry (5IKR) represented the dominant conformational preference within the screened COX-2 ensemble, whereas the ibuprofen- and naproxen-bound conformations contributed additional but smaller subsets of ligand-specific adaptability.

The dominance of the 5IKR (mefenamic acid-bound) conformation underscores its pharmacophoric complementarity and pocket flexibility relative to the chemical space of the screened library. This structure represents an open COX-2 active-site conformation, characterized by an expanded hydrophobic subpocket adjacent to Tyr385 and Ser530, an arrangement that favors diarylheterocyclic scaffolds and bulky arylacetic acid derivatives. The 5KIR (rofecoxib) structure, in turn, captures a more restricted yet deeply hydrophobic channel, suggesting strong π–π and van der Waals stabilization for compact coxib-like cores. In contrast, the 4PH9 (ibuprofen) and 3NT1 (naproxen) conformations contributed smaller subsets, consistent with their more constrained, ligand-locked pocket geometries that are less permissive to diverse scaffolds. Collectively, this conformational profiling demonstrates that while the mefenamic acid- and rofecoxib-bound COX-2 states dominate ligand recognition, inclusion of ibuprofen- and naproxen-bound conformations enhances the ensemble's coverage of ligand-induced adaptability, thereby improving detection of alternative binding modes and selectivity determinants.

*Consensus and Conformation-Specific Binding Profiles*

Ligands were further classified based on their binding consistency across the COX-2 conformational ensemble. Consensus binders, defined as molecules maintaining strong affinity ($\Delta G < -7$ kcal.mol$^{-1}$) across all four COX-2 structures, accounted for 624 compounds (30.7%) of



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

the validated set. These ligands likely recognize core pharmacophoric features conserved among the COX-2 active states, suggesting robust scaffold adaptability and stable anchoring interactions, typically involving the Arg120-Tyr355-Ser530 triad and hydrophobic channel residues. In contrast, 187 molecules (9.2%) exhibited conformation-specific binding, displaying a >2 kcal.mol$^{-1}$ preference for one COX-2 structure over the others. Such ligands preferentially stabilized unique pocket geometries, reflecting induced-fit adaptability or allosteric microstate sensitivity. These molecules are of mechanistic interest, as they may reveal structure-selective inhibition patterns linked to conformational dynamics and isoform selectivity.

The integrated RL, multi-target optimization, and orthogonal docking workflow produced a robust portfolio of highly selective, novel, and synthetically tractable COX-2 inhibitor candidates. Quantitative performance metrics across validation, potency, selectivity, novelty, and safety endpoints highlight the overall success and translational readiness of the modeling framework.

*Orthogonal validation performance.* The docking and scoring pipeline demonstrated a 51.9% orthogonal validation success rate, substantially exceeding the 10-20 % benchmark typical of large-scale virtual screening campaigns. This high reproducibility reflects the strong alignment between the generative model's latent chemical features and experimentally validated COX-2 pharmacophoric space.

*Potency and selectivity.* A total of 23 molecules displayed binding energies exceeding that of rofecoxib (−10.1 kcal/mol) across COX-2 conformations, underscoring the model's capacity to enrich for highly potent analogues. The maximum selectivity differential reached 9.2 kcal/mol, corresponding to an estimated >10,000-fold preference for COX-2 over COX-1, affirming the success of negative-learning reinforcement against COX-1 similarity. On average, the generated ligands exhibited a mean selectivity margin of 2.4 kcal/mol, consistent with pharmacologically meaningful COX-2 bias.

*Novelty and intellectual property freedom.* The final library achieved 92.9 % structural novelty at a 0.7 Tanimoto similarity threshold and 100% patent freedom, indicating that the generative model



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

explored chemically uncharted and commercially viable regions of COX-2 inhibitor space. This degree of novelty surpasses typical de-novo design campaigns, where redundancy with known scaffolds often exceeds 20-30%.

*Safety and developability.* In silico toxicity profiling demonstrated that 88% of compounds were predicted to have low hERG cardiotoxic risk, while 83% presented low DILI propensity, signifying strong safety alignment relative to historical NSAID benchmarks. The candidates also exhibited excellent synthetic tractability, with 75.7% of the library displaying SAScores <3.0, corresponding to straightforward laboratory synthesis routes.

*Consensus binding robustness.* Importantly, 624 molecules maintained consistently high affinity across all COX-2 conformations (rofecoxib-like, mefenamic-acid-like, and ibuprofen-like), confirming that the learned pharmacophore features are conformation-agnostic and transferable across multiple binding-site geometries. Collectively, these results validate the efficacy, generalizability, and translational potential of the proposed AI-driven molecular design platform, establishing it as a scalable strategy for generating clinically relevant, safe, and novel COX-2-selective inhibitors.

*Comparative docking of top multi-target candidates.* The Boltz-2 predicted and further modeled COX-1 structure (addition of ligand and cofactor) was used for docking analyses of top-performing hits (23 hits). Boltz-2 achieved an overall confidence score of 91.6%, indicating very high accuracy (CA ± pLDDT <2.2 ± 0.34 Å). The Predicted TM-score (pTM) for complex folding was 94.6%, demonstrating excellent structural alignment, while the interface predicted TM-score (ipTM) was 0.0%, suggesting no interfacial dependency in this model. The quality of the Boltz-2-predicted human COX-1 structure was rigorously validated through dual-crystal structure comparison. First, alignment with human COX-1 crystal structure 6Y3C (100% sequence identity) yielded a backbone RMSD of 0.541 Å and active site RMSD of 0.681 Å, demonstrating excellent structural accuracy. Next, to validate functional conservation across species, we performed sequence-independent spatial alignment with ovine COX-1 crystal structure 3N8Z (>90% sequence identity).



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

The binding pocket geometry was remarkably conserved with pocket center displacement 0.744 Å, volume similarity 97.4% and shape similarity 84.8%. All five critical catalytic residues spatially superimposed within 1.35 Å (ARG96, TYR331, TYR361, TRP363, SER506) before energy minimization. Flurbiprofen (3N8Z) showed several favorable contacts with the predicted structure, with only one minor geometric adjustment required (1.71 Å O-$NH_2$ distance, forming a favorable salt bridge). The convergence of validation results confirmed that the Boltz-2 prediction accurately captured the essential COX-1 active site architecture, validating its use for virtual screening and selectivity calculations. Figure 6 shows the ribbon diagram of Boltz-2 assisted predicted COX-1 structure. These results confirmed a highly reliable monomeric structure prediction with strong backbone confidence and minimal spatial deviation from known crystallographic templates. Table 10 lists five of the top-performers along with their docking scores for COX-2 (two structures), mPGES-1 (two structures) and predicted COX-1 structures.

*Comparative Binding Analysis of **MAI_2**, **MAI_4**, **MAI_7**, **MAI_8** and **MAI_9** at COX-2*

Across the dataset, a clear preference for COX-2 and mPGES-1 binding was observed, consistent with the pharmacophoric bias introduced during generative modeling and reinforcement training. The mean predicted binding energies for COX-2 ranged from -7.2 to -10.5 kcal/mol, with 5KIR consistently yielding stronger interactions than 4PH9, indicating the deeper, inhibitor-bound conformation captured by the rofecoxib pocket. **MAI_2** demonstrated the highest COX-2 affinity (-10.5 kcal/mol), surpassing the reference ligand rofecoxib (-10.1 kcal/mol). **MAI_4** and **MAI_7**) both maintained >-9.6 kcal/mol docking scores.

All ligands demonstrated deep penetration into the COX channel (PDB ID: 5KIR), forming key interactions with residues Arg120, Tyr355, Ser353, Leu352, Phe518, Gln192, and Tyr385, the canonical pharmacophoric triad responsible for COX-2 inhibitor stabilization (Table 11, Figure X). The sulfonamide/methyl-sulfonyl groups acted as polar anchors, engaging Ser353 and Gln192 through strong H-bonds (2.7–3.1 Å), while their aromatic scaffolds aligned along the hydrophobic corridor (Val349, Leu359, Ala527, Phe518), mirroring the orientation of classical coxibs such as



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

rofecoxib and celecoxib. **MAI_2** maintained the classical COX-2 binding pose, forming H-bonds with Ser353, Leu352, Gln192, and Tyr385. Its orientation enabled π–π stacking with Phe518 and hydrophobic stabilization by Val349 and Ala527. The combination of polar precision and hydrophobic enclosure yielded a balanced and stable interaction comparable to celecoxib-like binding. **MAI_4** formed an extended polar network involving Arg120, Ser353, Leu352, Gln192, and Tyr385. Additional stabilization came from Phe518 and Tyr355, indicating a dual H-bonding pattern bridging both the entrance (Arg120/Tyr355) and catalytic core (Ser353/Tyr385). The binding energy (-9.6 kcal/mol) and multi-residue engagement suggested enhanced pocket retention and selectivity.

A compact orientation was observed for **MAI_7**, anchored primarily by Ser353, Leu352, and Gln192 through H-bonding. The ligand also engaged Tyr385 via π–π interactions and Val523 and Phe518 through hydrophobic stabilization. This simpler yet effective interaction scheme indicated efficient occupation of the hydrophobic channel with minimal steric strain. **MAI_8** exhibited a multi-residue H-bonding network spanning Arg120, Ser353, Leu352, Gln192, Tyr355, and Tyr385. The ligand's sulfonyl oxygen engaged the Arg120-Tyr355 polar cluster, while its aromatic ring extended into the secondary hydrophobic pocket (Phe518, Val349, Ala527). This dual-pocket occupancy mimics the rofecoxib binding pattern and rationalizes its superior affinity. **MAI_9** displayed the most extensive binding footprint, integrating polar interactions with Arg120, Ser353, Leu352, Gln192, and Tyr385 alongside hydrophobic engagement from Val349, Leu359, Ala527, and Phe518. The aromatic sulfonamide core was deeply buried, forming a network like high-potency diarylheterocyclic inhibitors. This balanced polar-hydrophobic interaction pattern corresponds with its predicted strong dual inhibition.

Across all five complexes, Ser353, Leu352, Gln192, and Tyr385 emerged as the core anchoring residues, consistent with the conserved pharmacophore of selective COX-2 inhibitors. The involvement of Arg120 and Tyr355 in **MAI_4**, **MAI_8**, and **MAI_9** interactions indicated



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

additional entrance-site anchoring, often associated with improved residence time and reduced off-target COX-1 affinity.

*Comparative Binding Analysis of **MAI_2**, **MAI_4**, **MAI_7**, **MAI_8** and **MAI_9** at mPGES-1*.

All five molecules exhibited strong mPGES-1 binding, with 8PYV structure producing more favorable scores (−8.849 to −10.31 kcal/mol) than 4YL3 (−8.658 to −8.935 kcal/mol). This supported the rationale for incorporating 8PYV, the GSH-analog–bound conformation, in the screening pipeline, as it captured the polar cofactor interface absent in the inhibitor-bound 4YL3 structure. All five hits anchored within the GSH-binding channel (Figure X), confirming that the generative framework successfully captured the pharmacophoric logic of mPGES-1 inhibition. While each hit shared the core sulfonamide-based H-bond triad (Arg73–Glu77–Ser127), distinct interaction expansions were observed across the series, corresponding to incremental improvements in stability and affinity (Table 12). **MAI_2** served as the baseline, establishing canonical GSH-mimetic interactions with Ser127(Chain A), Tyr130(Chain A), and Arg38(Chain C). **MAI_4** exhibited enhanced anchoring via an extended polar triad (Arg73–Glu77–Tyr117) and a denser π–hydrophobic framework, leading to improved orientation and binding strength. **MAI_7** formed a more compact electrostatic network centered on Arg38–Arg73–Glu77, favoring strong charge-driven stabilization but fewer secondary H-bonds, consistent with its slightly reduced energy compared to **MAI_4**. **MAI_8** achieved optimal polar coverage, engaging His72(C), Asp75(C), Arg73(A), Ser127(A), and Glu77(A), yielding the lowest docking energy (−10.31 kcal/mol). MAI_9 extended this motif by incorporating His53(C), Asp49(C), and Tyr117(A) interactions, resulting in the widest H-bond network (seven interactions) and strong hydrophobic burial, stabilizing the ligand throughout the catalytic channel. Figure 7 shows the ribbon diagram of trimeric mPGES-1 crystal structure used for virtual screening with representative hit (two copies) bound at the dimer interface.

H-Bond density correlated linearly with docking energy improvement, with **MAI_8** and **MAI_9** showing the most favorable balance between electrostatic anchoring and hydrophobic



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

complementarity. Residue recurrence analysis identified Arg73, Glu77, and Ser127 as critical interaction nodes across all high-affinity ligands, underscoring their essential role in inhibitor stabilization. The inclusion of Tyr117 and His72/His53 in **MAI_4**, **MAI_8**, and **MAI_9** complexes suggested an expanded cofactor-recognition mimicry, possibly enhancing both affinity and kinetic residence time. Hydrophobic residues (Leu69, Phe44, Leu39, Ala123) consistently formed a lipophilic envelope that reinforced pocket occupancy and shape complementarity, supporting the dual amphiphilic character necessary for potent inhibition. Collectively, these results demonstrated clear progression in interaction complexity and energetic stability from **MAI_2** to **MAI_9**. The higher-ranked ligands (**MAI_8** and **MAI_9**) not only anchored the canonical Arg73–Glu77–Ser127 triad but also recruited additional residues (His72, Asp75, His53, Tyr117) to form a continuous polar belt across the catalytic entrance.

*Comparative Binding Analysis of Top Hits at COX-1*

Predicted COX-1 affinities (-4.3 to -5.8 kcal/mol) were uniformly weaker across all compounds, yielding selectivity differentials of 4-5 kcal/mol relative to COX-2. this corresponds to >1,000-fold preference for COX-2 binding, aligning with the intended outcome of negative-learning reinforcement against COX-1 similarity. The reduced interaction energies confirmed minimal engagement of the Arg120-Tyr355-Glu524 catalytic triad, validating the designed COX-2 selectivity.

A strong correlation ($r^2 \approx 0.86$) was observed between the COX-2 (5KIR) and mPGES-1 (8PYV) docking energies, reflecting consistent pharmacophoric overlap and complementarity between these functionally linked enzymes. The top-performing molecules (**MAI_2**, **MAI_4**, **MAI_9**) demonstrated concurrent potency against both targets, suggesting a dual inhibitory mechanism capable of blocking prostaglandin synthesis at both the COX and terminal synthase levels. Based on docking energetics and selectivity indices, **MAI_2** emerged as the lead candidate with optimal multi-target engagement (COX-2: −10.5; mPGES-1: −9.59; COX-1: −4.8 kcal/mol). The second-tier analogs, **MAI_4** and **MAI_8**, exhibited similar selectivity and may represent



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

structural variants with alternative H-bond geometries near the sulfonamide region. The molecular structures of the hits are given in Figure 3. Figures 4 and 5 show the 2D ligand-receptor interaction maps between all five top-performers and COX-2 (5KIR) and mPGES-1 (8PYV), respectively.

*Computational Prediction Uncertainty and Statistical Considerations.* All computational predictions in this study are subject to inherent methodological uncertainty. Molecular docking predictions typically exhibit 2-3 kcal/mol standard error in binding energy estimation when compared to experimental measurements, arising from approximations in scoring functions, protein flexibility, solvation modeling, and entropy estimation. Consequently, our reported binding energies should be interpreted as relative ranking metrics rather than absolute affinity predictions. For selectivity calculations based on binding energy differentials ($\Delta\Delta G = \Delta G_COX\text{-}1 - \Delta G_COX\text{-}2$), this uncertainty propagates. A selectivity differential of 9.2 kcal/mol, while corresponding to an estimated >10,000-fold preference under ideal conditions (assuming $\Delta\Delta G = RT \ln(K_COX\text{-}1/K_COX\text{-}2)$), should be interpreted as high confidence for 100-fold selectivity given typical docking error margins. Selectivity differentials of 4-5 kcal/mol correspond to estimated >1,000-fold preference, again with the caveat that absolute fold-change values require experimental validation.

ADMET classification models report point-estimates (binary predictions or probabilities) based on training set performance. Model performance metrics (AUROC 0.79-0.899) indicate good but not perfect classification accuracy, with expected false positive/negative rates of 10-20% depending on the endpoint. Predictions should be viewed as computational risk stratification rather than definitive toxicity assessments. Throughout this work, we report mean docking scores, with standard deviations where multiple binding modes or protein conformations were assessed. Statistical comparisons between groups (e.g., consensus vs. non-consensus binders) were performed using non-parametric tests where appropriate, though we emphasize that computational predictions ultimately require experimental validation for clinical translation. The



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

value of this AI-driven approach lies in efficient exploration and prioritization of chemical space, reducing the search space for subsequent experimental investigation from billions to tens of molecules.

## Conclusions

The present computational campaign represented a comprehensive AI-driven effort toward the discovery of dual COX-2/mPGES-1 inhibitors with improved efficacy and safety over existing NSAIDs. Using the OCSR.ai Deep Docking Platform (DDP), a total of 2.56 million de novo–generated molecules were systematically evaluated across eight pharmacologically relevant targets, integrating both on-target efficacy and off-target safety filters. This multi-stage screening cascade resulted in the identification of 16,984 COX-2 active compounds, of which 14.0% displayed significant COX-2 selectivity over COX-1 and 99.4% demonstrated concurrent mPGES-1 activity, a critical attribute for enhanced prostanoid modulation. Through successive rounds of physicochemical, structural, and toxicophore filtering, a refined library of 4,219 drug-like candidates was obtained, exhibiting excellent compliance with Lipinski's Rule of Five, synthetic accessibility, and safety parameters. Among these, 1,939 clean leads (46%) fulfilled stringent ADMET criteria, non-mutagenic, hERG-safe, DILI-safe, and synthetically feasible (mean SAScore: 2.68). Notably, all surviving compounds were 100% novel, patent-free, and resided within an unexplored chemical space beyond PubChem.

Compared to existing COX-2 inhibitors such as celecoxib, rofecoxib, and lumiracoxib, the generated leads offer several differentiating advantages: absence of halogenated or furanone scaffolds, integrated cardiotoxicity and hepatotoxicity risk screening, and inclusion of downstream prostanoid modulation via mPGES-1 inhibition, a mechanistic advancement over single-target NSAIDs. The study established a robust computational pipeline combining RL-based generative modeling with multi-target virtual screening and in silico ADMET validation, achieving a >99.9% attrition rate and yielding only the most promising synthesis-ready candidates. The campaign's



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

success highlights the value of negative learning integration (to suppress off-target features such as COX-1, hERG, and CYP binding), and large-scale patent-space filtering (over 44 million patents) to ensure FTO, an innovation rarely implemented in early-stage AI drug discovery. The findings also underscore biologically relevant insights: the high overlap between COX-2 and mPGES-1 actives (99.4%) supports the feasibility of dual inhibition strategies, while analysis of sulfonamide derivatives confirms that selectivity arises primarily from COX-1 avoidance rather than enhanced COX-2 binding.

In conclusion, this campaign provides a first-in-class, AI-driven blueprint for discovering multi-target, safety-optimized NSAIDs within a fraction of the traditional timeline. The successfully identified chemical series not only meet pharmacological benchmarks for dual COX-2/mPGES-1 inhibition but also demonstrate the practical maturity of reinforcement learning–based molecular design as a translational discovery engine. These findings establish a new foundation for advancing next-generation anti-inflammatory therapeutics with improved efficacy-safety balance, synthetic tractability, and strong IP positioning.

## Conflict of Interest

The authors declare no competing financial interests. This study was conducted independently as part of an academic–industry collaborative drug discovery initiative aimed at the development of next-generation NSAIDs with improved safety and efficacy profiles. All data generated are intended for academic dissemination and future translational validation under appropriate ethical and institutional guidelines.

## Acknowledgements

PSK is thankful to Prof. Aniruddha B. Pandit, Vice Chancellor, Institute of Chemical Technology, Mumbai, for making facilities available during the present study. Authors thank Ms. Jitsu Sheth, Aarth Software, for her help and encouragement during the present work.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Abbreviations**

ADMET: Absorption, Distribution, Metabolism, Excretion, Toxicity

CA-II: Carbonic Anhydrase II (off-target enzyme)

cLogP: Calculated logarithm of partition coefficient (octanol/water), measure of lipophilicity

COX-1/COX-2: Cyclooxygenase-1 and -2

CYP: Cytochrome P450

DILI: Drug-Induced Liver Injury

DDP: Deep Docking Platform (OCSR.ai)

FTO: Freedom-to-Operate

HBA/HBD: H-Bond Acceptors/Donors

hERG Human Ether-à-go-go-Related Gene

$IC_{50}$: Half-maximal Inhibitory Concentration

mPGES-1: Microsomal Prostaglandin E Synthase-1

PAINS: Pan-Assay Interference Compounds

$PGE_2$: Prostaglandin $E_2$

PTGIS: Prostaglandin $I_2$ Synthase

QED: Quantitative Estimate of Drug-likeness

SAScore: Synthetic Accessibility Score

TPSA: Topological Polar Surface Area



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

## References


1. Wang B, Wu L, Chen J, Dong L, Chen C, Wen Z, Hu J, Fleming I, Wang DW. Metabolism pathways of arachidonic acids: mechanisms and potential therapeutic targets. Signal Transduct Target Ther. 2021;6(1):94. doi: 10.1038/s41392-020-00443-w.
2. Biringer RG. The enzymology of the human prostanoid pathway. Mol Biol Rep. 2020;47(6):4569-4586. doi: 10.1007/s11033-020-05526-z.
3. Ju Z, Li M, Xu J, Howell DC, Li Z, Chen FE. Recent development on COX-2 inhibitors as promising anti-inflammatory agents: The past 10 years. Acta Pharm Sin B. 2022;12(6):2790-2807. doi: 10.1016/j.apsb.2022.01.002.
4. Kassab AE. Recent advances in targeting COX-2 for cancer therapy: a review. RSC Med Chem. 2025 Apr 21. doi: 10.1039/d5md00196j.
5. Mazaleuskaya LL, Ricciotti E. Druggable prostanoid pathway. Adv Exp Med Biol. 2020;1274:29-54. doi: 10.1007/978-3-030-50621-6_3.
6. Ghosh N, Chaki R, Mandal V, Mandal SC. COX-2 as a target for cancer chemotherapy. Pharmacol Rep. 2010;62(2):233-44. doi: 10.1016/s1734-1140(10)70262-0.
7. Stiller CO, Hjemdahl P. Lessons from 20 years with COX-2 inhibitors: Importance of dose-response considerations and fair play in comparative trials. J Intern Med. 2022;292(4):557-574. doi: 10.1111/joim.13505.
8. Rodrigues P, Bangali H, Hammoud A, Mustafa YF, Al-Hetty HRAK, Alkhafaji AT, Deorari MM, Al-Taee MM, Zabibah RS, Alsalamy A. COX 2-inhibitors; a thorough and updated survey into combinational therapies in cancers. Med Oncol. 2024;41(1):41. doi: 10.1007/s12032-023-02256-7.
9. Zhang F, Zhu G, Li Y, Qi Y, Wang Z, Li W. Dual-target inhibitors based on COX-2: a review from medicinal chemistry perspectives. Future Med Chem. 2023;15(23):2209-2233. doi: 10.4155/fmc-2023-0192.




https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

10. Tajdari M, Peyrovinasab A, Bayanati M, Ismail Mahboubi Rabbani M, Abdolghaffari AH, Zarghi A. Dual COX-2/TNF-α inhibitors as promising anti-inflammatory and cancer chemopreventive agents: A review. Iran J Pharm Res. 2024;23(1):e151312. doi: 10.5812/ijpr-151312.
11. Hiskens MI, Schneiders AG, Fenning AS. Selective COX-2 inhibitors as neuroprotective agents in traumatic brain injury. Biomedicines. 2024;12(8):1930. doi: 10.3390/biomedicines12081930.
12. Dogné JM, Supuran CT, Pratico D. Adverse cardiovascular effects of the coxibs. J Med Chem. 2005;48(7):2251-7. doi: 10.1021/jm0402059.
13. Steinmetz-Späh J, Jakobsson PJ. The anti-inflammatory and vasoprotective properties of mPGES-1 inhibition offer promising therapeutic potential. Expert Opin Ther Targets. 2023;27(11):1115-1123. doi: 10.1080/14728222.2023.2285785.
14. Koeberle A, Werz O. Perspective of microsomal prostaglandin E2 synthase-1 as drug target in inflammation-related disorders. Biochem Pharmacol. 2015;98(1):1-15. doi: 10.1016/j.bcp.2015.06.022.
15. Jakobsson PJ, Thorén S, Morgenstern R, Samuelsson B. Identification of human prostaglandin E synthase: a microsomal, glutathione-dependent, inducible enzyme, constituting a potential novel drug target. Proc Natl Acad Sci U S A. 1999; 96(13):7220-5. doi: 10.1073/pnas.96.13.7220.
16. Bergqvist F, Morgenstern R, Jakobsson PJ. A review on mPGES-1 inhibitors: From preclinical studies to clinical applications. Prostaglandins Other Lipid Mediat. 2020;147:106383. doi: 10.1016/j.prostaglandins.2019.106383.
17. Wang M, FitzGerald GA. Cardiovascular biology of microsomal prostaglandin E synthase-1. Trends Cardiovasc Med. 2010;20(6):189-95. doi: 10.1016/j.tcm.2011.04.002.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

18. Bahia MS, Katare YK, Silakari O, Vyas B, Silakari P. Inhibitors of microsomal prostaglandin E2 synthase-1 enzyme as emerging anti-inflammatory candidates. Med Res Rev. 2014;34(4):825-55. doi: 10.1002/med.21306.
19. Huang Q, Lai T, Wang Q, Luo L. mPGES-1 Inhibitor discovery based on computer-aided screening: Pharmacophore models, molecular docking, ADMET, and MD simulations. Molecules. 2023;28(16):6059. doi: 10.3390/molecules28166059.
20. Ricciotti E, FitzGerald GA. Prostaglandins and inflammation. Arterioscler Thromb Vasc Biol. 2011;31(5):986-1000. doi: 10.1161/ATVBAHA.110.207449.
21. Nørregaard R, Kwon TH, Frøkiær J. Physiology and pathophysiology of cyclooxygenase-2 and prostaglandin E2 in the kidney. Kidney Res Clin Pract. 2015;34(4):194-200. doi: 10.1016/j.krcp.2015.10.004.
22. Kobayashi T, Nakatani Y, Tanioka T, Tsujimoto M, Nakajo S, Nakaya K, Murakami M, Kudo I. Regulation of cytosolic prostaglandin E synthase by phosphorylation. Biochem J. 2004;381(Pt 1):59-69. doi: 10.1042/BJ20040118.
23. de Vries S, Benes V, Naarmann-de Vries IS, Rücklé C, Zarnack K, Marx G, Ostareck DH, Ostareck-Lederer A. P23 acts as functional RBP in the macrophage inflammation response. Front Mol Biosci. 2021;8:625608. doi: 10.3389/fmolb.2021.625608.
24. Li K, Zhao J, Wang M, Niu L, Wang Y, Li Y, Zheng Y. The roles of various prostaglandins in fibrosis: A review. Biomolecules. 2021;11(6):789. doi: 10.3390/biom11060789.
25. Tornling G, Edenius C, Pauling JD, Denton CP, Olsson A, Kowalski J, Murray A, Anderson M, Bhat S, Del Galdo F, Hall F, Korkosz M, Krasowska D, Olas J, Smith V, van Laar JM, Vonk MC, Wojteczek A, Herrick AL. A phase 2 trial investigating the efficacy and safety of the mPGES-1 inhibitor vipoglanstat in systemic sclerosis-related Raynaud's. Rheumatology (Oxford). 2025;64(2):704-713. doi: 10.1093/rheumatology/keae049.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

26. Takeda H, Miyoshi H, Tamai Y, Oshima M, Taketo MM. Simultaneous expression of COX-2 and mPGES-1 in mouse gastrointestinal hamartomas. Br J Cancer. 2004;90(3):701-4. doi: 10.1038/sj.bjc.6601584.
27. Patrono C. Cardiovascular effects of cyclooxygenase-2 inhibitors: a mechanistic and clinical perspective. Br J Clin Pharmacol. 2016;82(4):957-64. doi: 10.1111/bcp.13048.
28. Bergqvist F, Ossipova E, Idborg H, Raouf J, Checa A, Englund K, Englund P, Khoonsari PE, Kultima K, Wheelock CE, Larsson K, Korotkova M, Jakobsson PJ. Inhibition of mPGES-1 or COX-2 Results in Different Proteomic and Lipidomic Profiles in A549 Lung Cancer Cells. Front Pharmacol. 2019;10:636.
29. Kapoor M, Kojima F, Qian M, Yang L, Crofford LJ. Shunting of prostanoid biosynthesis in microsomal prostaglandin E synthase-1 null embryo fibroblasts: regulatory effects on inducible nitric oxide synthase expression and nitrite synthesis. FASEB J. 2006;20(13):2387-9. doi: 10.1096/fj.06-6366fje.
30. Elzahhar PA, Orioli R, Hassan NW, Gobbi S, Belluti F, Labib HF, El-Yazbi AF, Nassra R, Belal ASF, Bisi A. Chromone-based small molecules for multistep shutdown of arachidonate pathway: Simultaneous inhibition of COX-2, 15-LOX and mPGES-1 enzymes. Eur J Med Chem. 2024;266:116138. doi: 10.1016/j.ejmech.2024.116138.
31. OCSR.ai The World's Most Powerful Compound Intelligence Engine. https://ocsr.ai/ (Accessed on October 26, 2025).
32. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, Magariños MP, Mosquera JF, Mutowo P, Nowotka M, Gordillo-Marañón M, Hunter F, Junco L, Mugumbate G, Rodriguez-Lopez M, Atkinson F, Bosc N, Radoux CJ, Segura-Cabrera A, Hersey A, Leach AR. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1):D930-D940. doi: 10.1093/nar/gky1075.
33. Prostaglandin G/H synthase 2. https://www.ebi.ac.uk/chembl/explore/target/CHEMBL230 (Accessed on October 26, 2025).



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

34. Prostaglandin G/H synthase 1. https://www.ebi.ac.uk/chembl/explore/target/CHEMBL221 (Accessed on October 26, 2025).

35. Voltage-gated inwardly rectifying potassium channel KCNH2. https://www.ebi.ac.uk/chembl/explore/target/CHEMBL240 (Accessed on October 26, 2025).

36. Cytochrome P450 2C9. https://www.ebi.ac.uk/chembl/explore/target/CHEMBL3397 (Accessed on October 26, 2025).

37. Cytochrome P450 3A4. https://www.ebi.ac.uk/chembl/explore/target/CHEMBL340. (Accessed on October 26, 2025).

38. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. Nat Chem. 2012;4(2):90-8. doi: 10.1038/nchem.1243.

39. wwPDB consortium. (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucleic Acids Res 47: D520-D528 doi: https://doi.org/10.1093/nar/gky949

40. PDB ID: 5KIR. DOI: https://doi.org/10.2210/pdb5KIR/pdb. Orlando BJ, Malkowski MG. Crystal structure of rofecoxib bound to human cyclooxygenase-2. Acta Crystallogr F Struct Biol Commun. 2016;72(Pt 10):772-776. doi: 10.1107/S2053230X16014230.

41. PDB ID: 3N8Z. DOI: https://doi.org/10.2210/pdb3N8Z/pdb. Sidhu RS, Lee JY, Yuan C, Smith WL. Comparison of cyclooxygenase-1 crystal structures: cross-talk between monomers comprising cyclooxygenase-1 homodimers. Biochemistry. 2010;49(33):7069-79. doi: 10.1021/bi1003298

42. PDB ID: 4YL3. DOI: https://doi.org/10.2210/pdb4YL3/pdb. Luz JG, Antonysamy S, Kuklish SL, Condon B, Lee MR, Allison D, Yu XP, Chandrasekhar S, Backer R, Zhang A, Russell M, Chang SS, Harvey A, Sloan AV, Fisher MJ. Crystal Structures of mPGES-1 Inhibitor Complexes Form a Basis for the Rational Design of Potent Analgesic and Anti-Inflammatory Therapeutics. J Med Chem. 2015;58(11):4727-37. doi: 10.1021/acs.jmedchem.5b00330



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

43. PDB ID: 8PYV. Doi: https://doi.org/10.2210/pdb8PYV/pdb. El Omari K, Duman R, Mykhaylyk V, Orr CM, Latimer-Smith M, Winter G, Grama V, Qu F, Bountra K, Kwong HS, Romano M, Reis RI, Vogeley L, Vecchia L, Owen CD, Wittmann S, Renner M, Senda M, Matsugaki N, Kawano Y, Bowden TA, Moraes I, Grimes JM, Mancini EJ, Walsh MA, Guzzo CR, Owens RJ, Jones EY, Brown DG, Stuart DI, Beis K, Wagner A. Experimental phasing opportunities for macromolecular crystallography at very long wavelengths. Commun Chem. 2023;6(1):219. doi: 10.1038/s42004-023-01014-0.
44. PDB ID: 6Y3C. Doi: https://doi.org/10.2210/pdb6Y3C/pdb. Miciaccia M, Belviso BD, Iaselli M, Cingolani G, Ferorelli S, Cappellari M, Loguercio Polosa P, Perrone MG, Caliandro R, Scilimati A. Three-dimensional structure of human cyclooxygenase (hCOX)-1. Sci Rep. 2021;11(1):4312. doi: 10.1038/s41598-021-83438-z.
45. PDB ID: 3N8X. Doi: https://doi.org/10.2210/pdb3N8X/pdb. Sidhu RS, Lee JY, Yuan C, Smith WL. Comparison of cyclooxygenase-1 crystal structures: cross-talk between monomers comprising cyclooxygenase-1 homodimers. Biochemistry. 2010;49(33):7069-79. doi: 10.1021/bi1003298.
46. Passaro S, Corso G, Wohlwend J, Reveiz M, Thaler S, Somnath VR, Getz N, Portnoi T, Roy J, Stark H, Kwabi-Addo D, Beaini D, Jaakkola T, Barzilay R. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. bioRxiv [Preprint]. 2025:2025.06.14.659707. doi: 10.1101/2025.06.14.659707.
47. Wohlwend J, Corso G, Passaro S, Getz N, Reveiz M, Leidal K, Swiderski W, Atkinson L, Portnoi T, Chinn I, Silterra J, Jaakkola T, Barzilay R. Boltz-1 Democratizing Biomolecular Interaction Modeling. bioRxiv [Preprint]. 2025:2024.11.19.624167. doi: 10.1101/2024.11.19.624167.
48. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022 Jun;19(6):679-682. doi: 10.1038/s41592-022-01488-1.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

49. RDKit: Open-source cheminformatics. https://www.rdkit.org (Accessed on October 26, 2025).
50. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50(5):742-54. doi: 10.1021/ci100050t.
51. SureChEMBL2.0. https://www.surechembl.org/ (Accessed on October 26, 2025).
52. Google Patents Public Data, provided by IFI CLAIMS Patent Services, is a worldwide bibliographic and US full-text dataset of patent publications. https://console.cloud.google.com/projectselector2/bigquery?p=patents-public-data&d=patents&page=dataset&supportedpurview=project (Accessed on October 26, 2025).
53. NVIDIA-Digital-Bio/KERMT. https://github.com/NVIDIA-Digital-Bio/KERMT?tab=readme-ov-file (Accessed on October 26, 2025)
54. TxGemma. https://developers.google.com/health-ai-developer-foundations/txgemma (Accessed on October 26, 2025)
55. a. Ding J, Tang S, Mei Z, Wang L, Huang Q, Hu H, Ling M, Wu J. Vina-GPU 2.0: Further accelerating AutoDock Vina and its derivatives with graphics processing units. J Chem Inf Model. 2023;63(7):1982-1998. doi: 10.1021/acs.jcim.2c01504; b. Tang S, Chen R, Lin M, Lin Q, Zhu Y, Ding J, Hu H, Ling M, Wu J. Accelerating AutoDock Vina with GPUs. Molecules. 2022;27(9):3041. doi: 10.3390/molecules27093041; c. O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, J Comput Chem. 2010;31:455-461. doi: 10.1002/jcc.21334.
56. ETKDG and distance constraints. https://greglandrum.github.io/rdkit-blog/posts/2021-02-22-etkdg-and-distance-constraints.html (Accessed on October 26, 2025)
57. PDB ID: 5IKR. Doi: https://doi.org/10.2210/pdb5IKR/pdb. Orlando BJ, Malkowski MG. Substrate-selective inhibition of cyclooxygeanse-2 by fenamic acid derivatives is dependent on peroxide tone. J Biol Chem. 2016;291(29):15069-81. doi: 10.1074/jbc.M116.725713.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

58. PDB ID: 4PH9 Doi: https://doi.org/10.2210/pdb4PH9/pdb. Duggan KC, Walters MJ, Musee J, Harp JM, Kiefer JR, Oates JA, Marnett LJ. Molecular basis for cyclooxygenase inhibition by the non-steroidal anti-inflammatory drug naproxen. J Biol Chem. 2010 Nov 5;285(45):34950-9. doi: 10.1074/jbc.M110.162982.
59. PDB ID: 3NT1 Doi: https://doi.org/10.2210/pdb3NT1/pdb. Duggan KC, Walters MJ, Musee J, Harp JM, Kiefer JR, Oates JA, Marnett LJ. Molecular basis for cyclooxygenase inhibition by the non-steroidal anti-inflammatory drug naproxen. J Biol Chem. 2010 Nov 5;285(45):34950-9. doi: 10.1074/jbc.M110.162982.
60. PDB ID: 3N8Z. Doi: https://doi.org/10.2210/pdb3N8Z/pdb. Sidhu RS, Lee JY, Yuan C, Smith WL. Comparison of cyclooxygenase-1 crystal structures: cross-talk between monomers comprising cyclooxygenase-1 homodimers. Biochemistry. 2010;49(33):7069-79. doi: 10.1021/bi1003298.
61. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model. 2011;51(10):2778-86. doi: 10.1021/ci200227u.
62. Kharkar PD, Borhade S, Dandi A, Warrier S. In search of novel anti-inflammatory agents: Computational repositioning of approved drugs. J Comput Sci. 2015;10:217-224.
63. Meyers J, Fabian B, Brown N. De novo molecular design and generative models. Drug Discov Today. 2021;26(11):2707-2715. doi: 10.1016/j.drudis.2021.05.019.
64. Pang C, Qiao J, Zeng X, Zou Q, Wei L. Deep generative models in *de novo* drug molecule generation. J Chem Inf Model. 2024;64(7):2174-2194. doi: 10.1021/acs.jcim.3c01496.
65. Mazuz E, Shtar G, Shapira B, Rokach L. Molecule generation using transformers and policy gradient reinforcement learning. Sci Rep. 2023;13(1):8799. doi: 10.1038/s41598-023-35648-w.



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

66. Liu X, Ye K, van Vlijmen HWT, IJzerman AP, van Westen GJP. DrugEx v3: scaffold-constrained drug design with graph transformer-based reinforcement learning. J Cheminform. 2023;15(1):24. doi: 10.1186/s13321-023-00694-z.
67. Vittorio S, Lunghini F, Morerio P, Gadioli D, Orlandini S, Silva P, Jan Martinovic, Pedretti A, Bonanni D, Del Bue A, Palermo G, Vistoli G, Beccari AR. Addressing docking pose selection with structure-based deep learning: Recent advances, challenges and opportunities. Comput Struct Biotechnol J. 2024 May 18;23:2141-2151. doi: 10.1016/j.csbj.2024.05.024. Erratum in: Comput Struct Biotechnol J. 2024 Jul 17;23:2963. doi: 10.1016/j.csbj.2024.07.001.
68. Pichake J, Kharkar PS, Ceruso M, Supuran CT, Toraskar MP. Carbonic anhydrase inhibitors: design, synthesis, and biological evaluation of novel sulfonyl semicarbazide derivatives. ACS Med Chem Lett. 2014;5(7):793-6. doi: 10.1021/ml500140t.
69. Singasane N, Kharkar PS, Ceruso M, Supuran CT, Toraskar MP. Inhibition of carbonic anhydrase isoforms I, II, IX and XII with Schiff's bases incorporating iminoureido moieties. J Enzyme Inhib Med Chem. 2015;30(6):901-7. doi: 10.3109/14756366.2014.986118.
70. Rong Y, Bian Y, Xu T, Xie W, Wei Y, Huang W, Huang J. Self-supervised graph transformer on large-scale molecular data. doi: arXiv:2007.02835
71. NVIDIA cuik-molmaker. https://github.com/NVIDIA-Digital-Bio/cuik-molmaker ((Accessed on October 26, 2025)



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

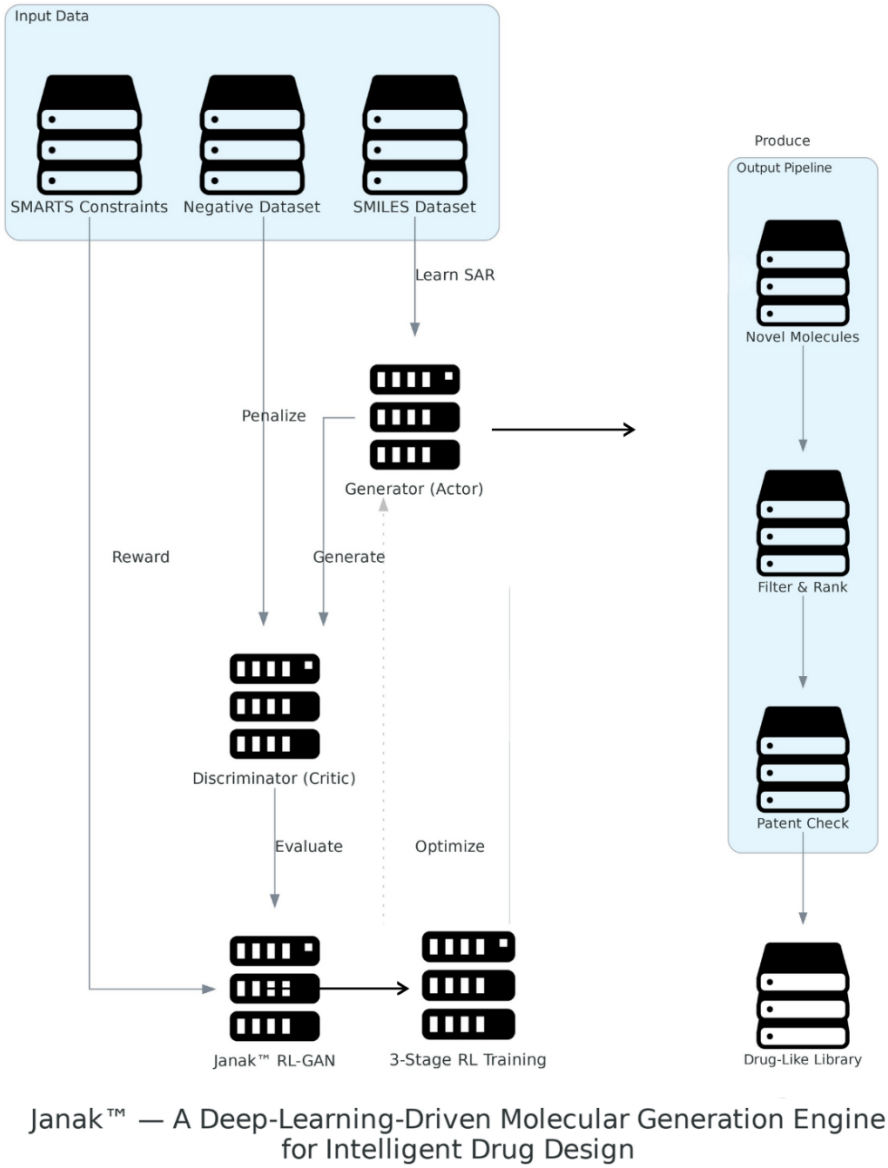



**Figure 1.** Overview of Janak™, a deep-learning-driven molecular generation engine that designs novel, synthesizable, and property-optimized compounds for intelligent drug discovery



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

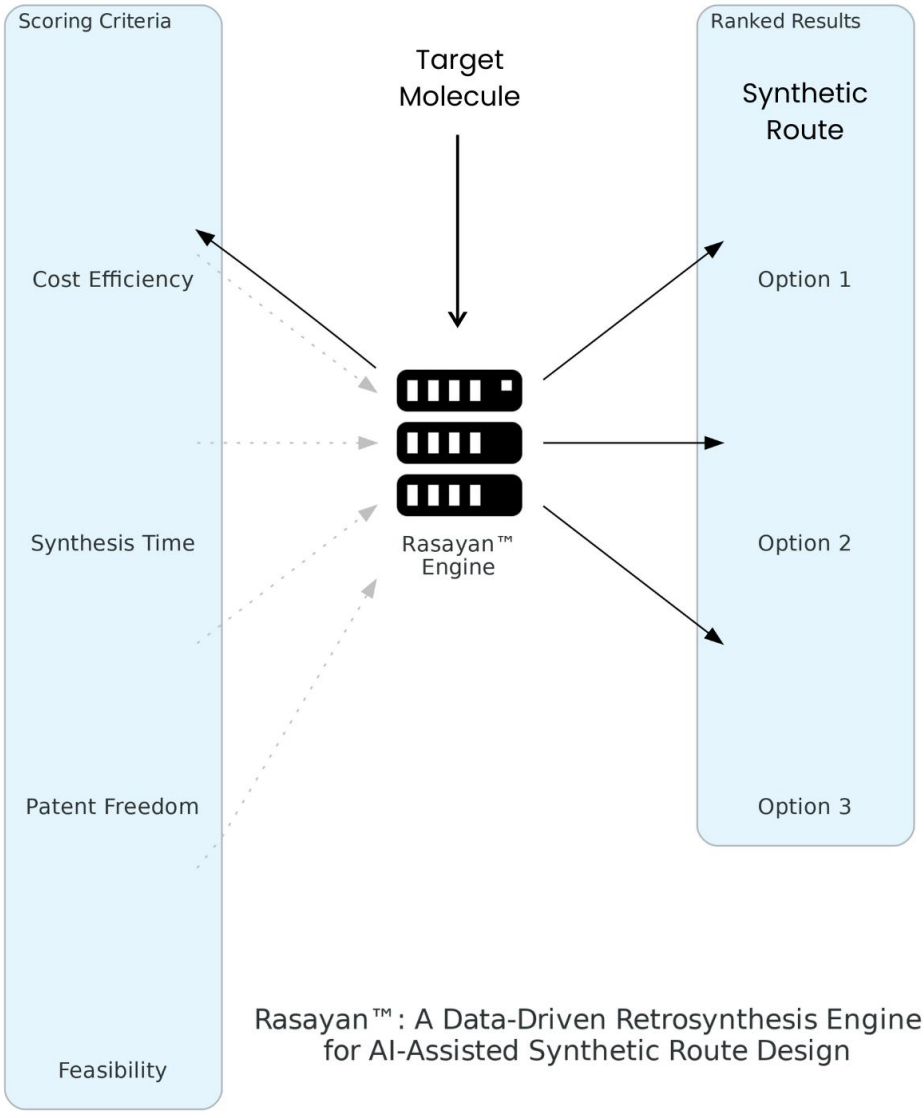



**Figure 2.** Architecture of Rasayan™, a data-driven retrosynthesis engine that predicts, ranks, and optimizes feasible synthetic routes using AI-assisted reaction modeling, cost evaluation, and patent-aware route selection



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

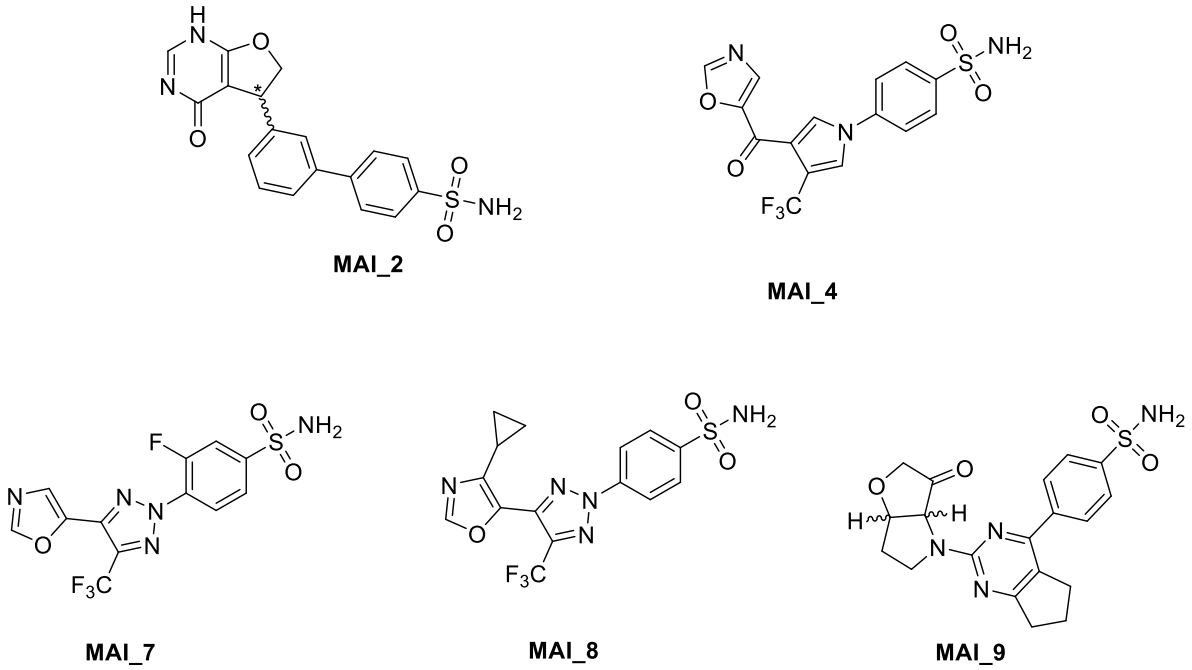
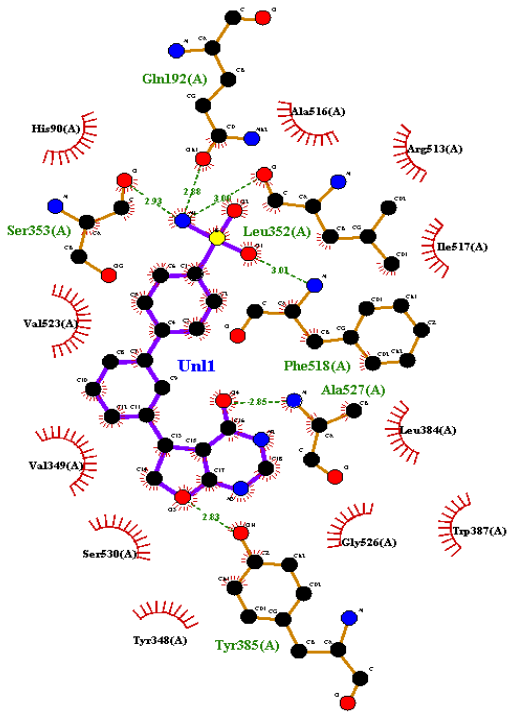



**Figure 3.** Molecular structures of top five hits obtained by rigorous AI-enabled computational discovery pipeline

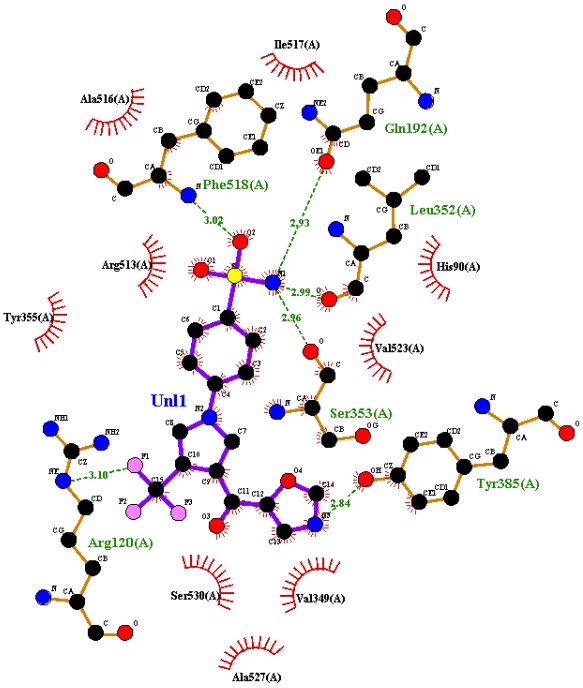


https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

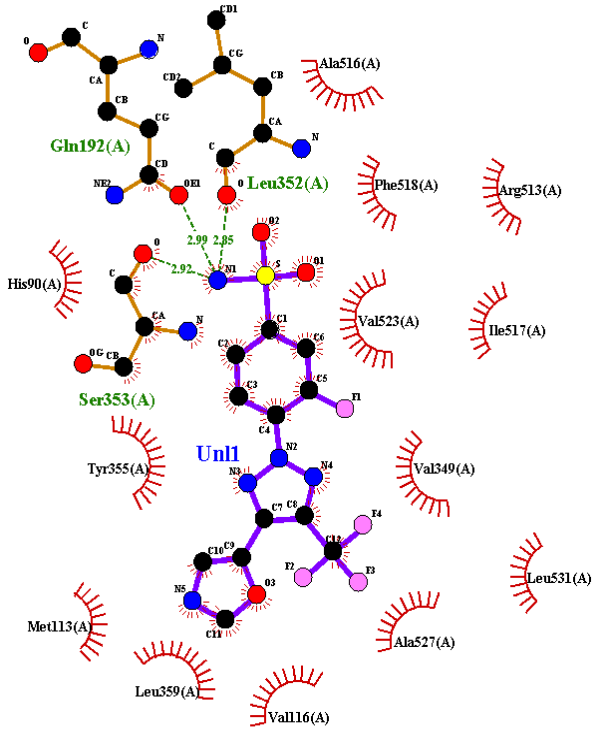
a



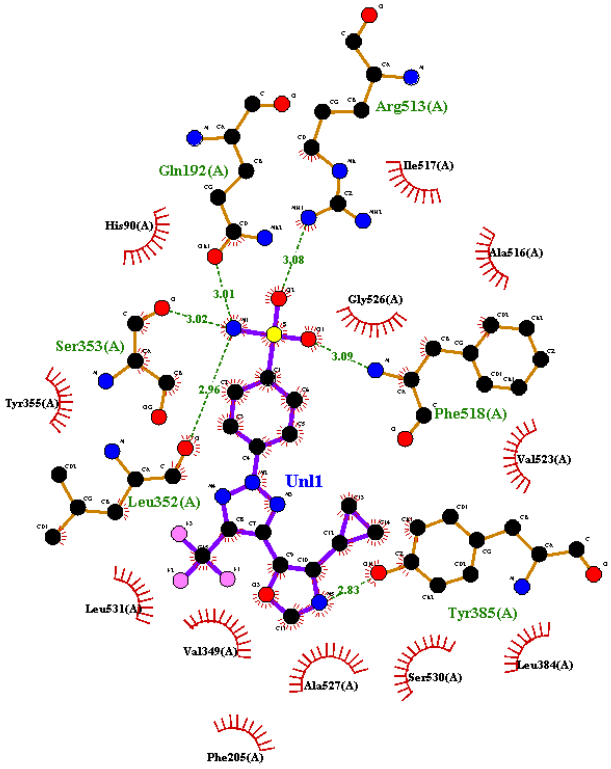

b.




c.




d.






https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

e.

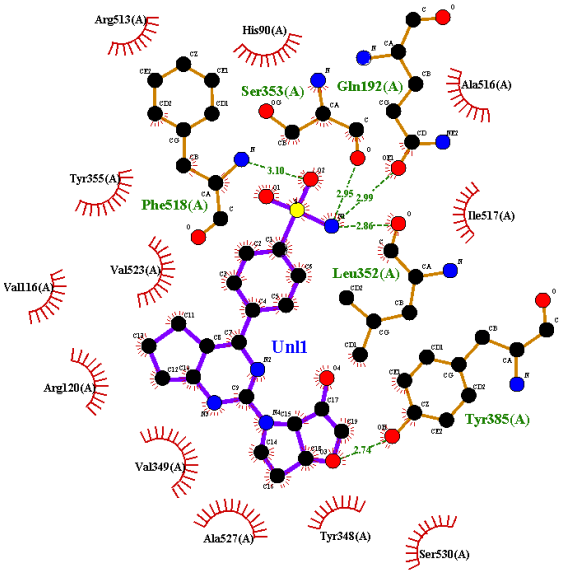
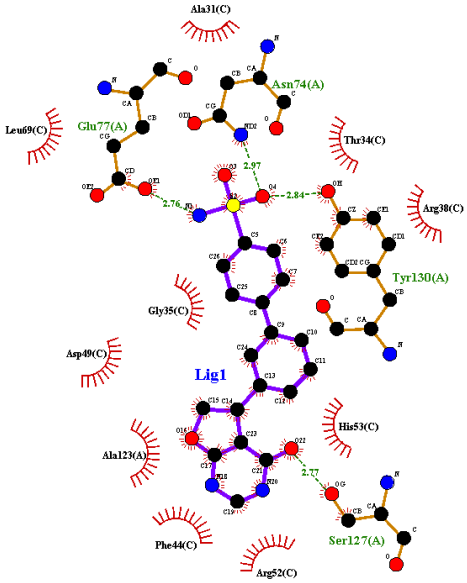



**Figure 4.** Representative 2D ligand-protein interaction maps for top-five dual COX-2/mPGES-1 inhibitor candidates – a. **MAI_2**, b. **MAI_4**, c. **MAI_7**, d. **MAI_8** and e. **MAI_9** docked within the rofecoxib-bound human COX-2 structure (PDB ID: 5KIR)

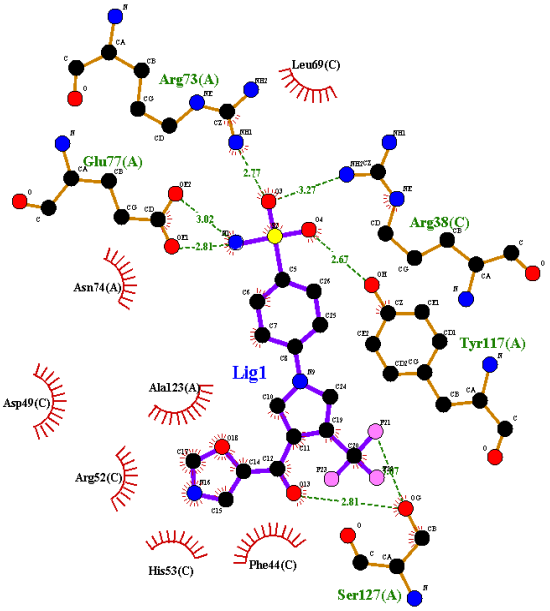


https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

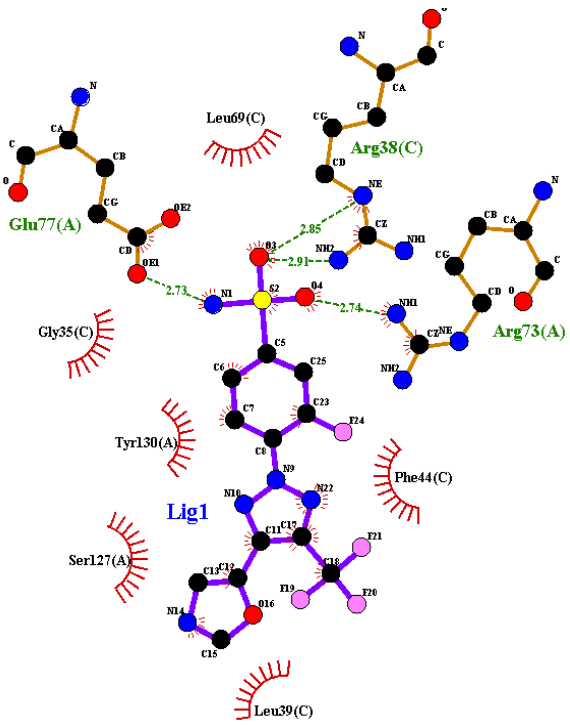
a.



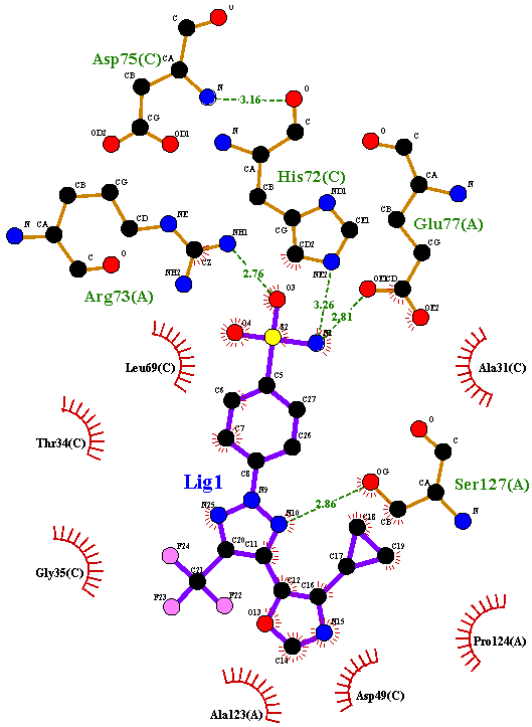

b.




c.




d.






https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

e.

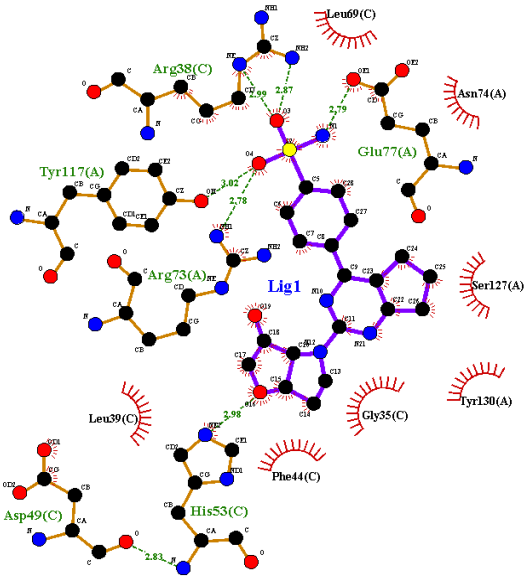



**Figure 5.** Representative 2D ligand-protein interaction maps for top-five dual COX-2/mPGES-1 inhibitor candidates, a. **MAI_2**, b. **MAI_4**, c. **MAI_7**, d. **MAI_8** and e. **MAI_9** docked into the human mPGES-1 GSH-binding conformation (PDB ID: 8PYV)



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

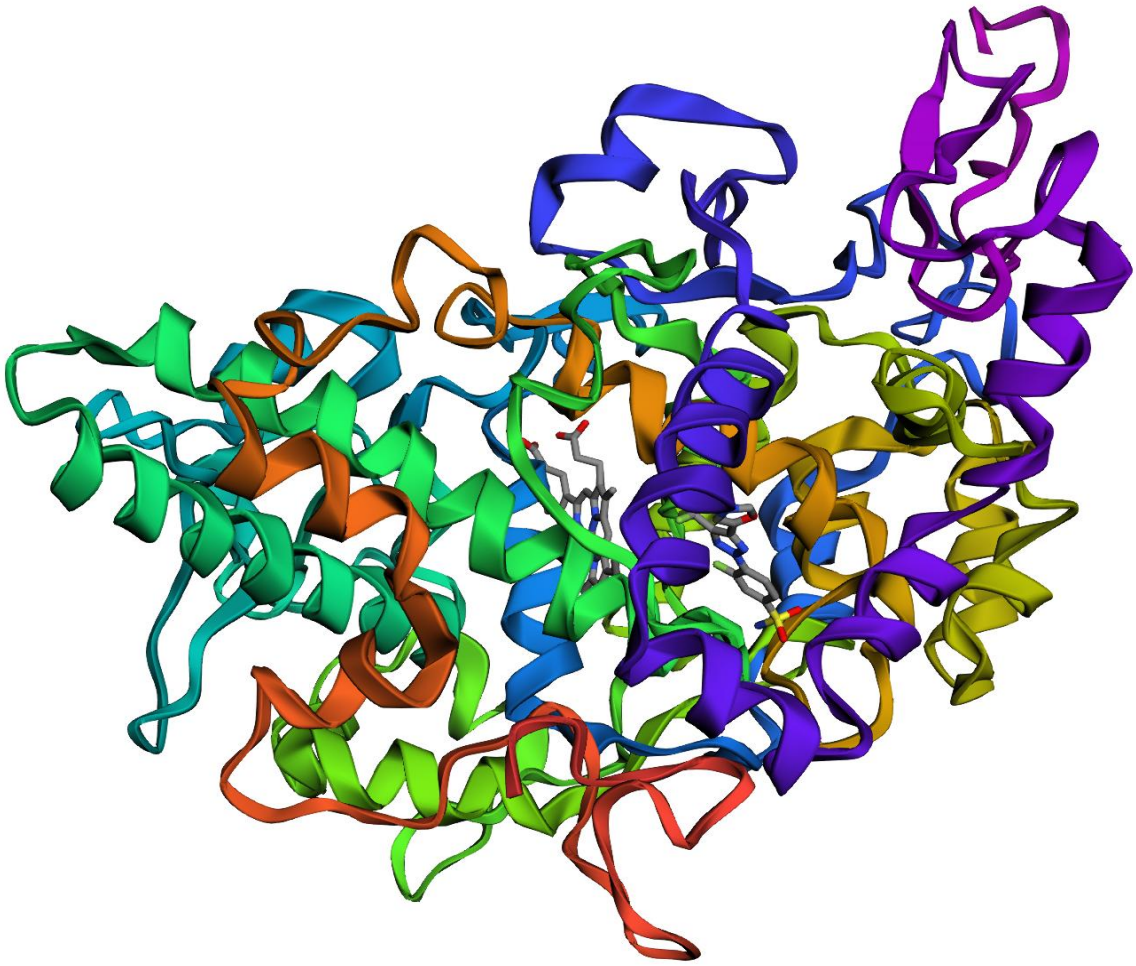


**Figure 6.** Ribbon diagram showing Boltz-2 assisted predicted COX-1 structure with representative hit (from top 23) bound in the COX-pocket and heme at the cofactor-binding site



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

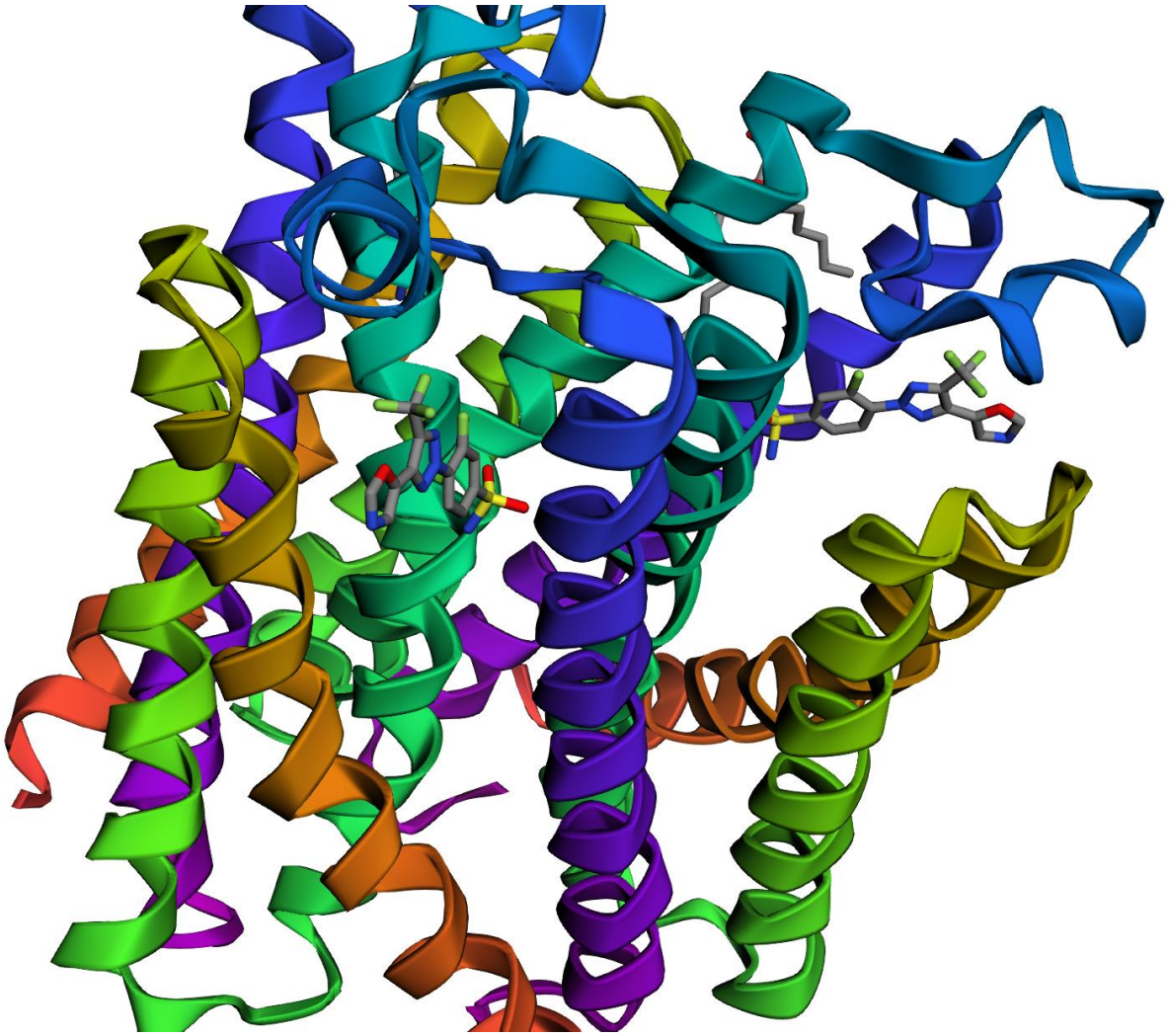


**Figure 7.** Ribbon diagram showing trimeric mPGES-1crystal structure (used for virtual screening) with representative hit (two copies) (from top 23) bound in the pocket at the dimer interfaces



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 1.** Therapeutic TPP for proposed dual COX-2/mPGES-1 inhibitors

| Parameter | Specification | Rationale |
|---|---|---|
| COX-2 selectivity | >10-fold vs COX-1 | Minimize GI toxicity |
| mPGES-1 activity | Positive binding | Enhanced efficacy |
| Cardiovascular safety | No hERG liability | Learn from rofecoxib withdrawal |
| Oral bioavailability | Drug-like properties | Clinical utility |
| Freedom-to-operate | Novel structures | IP protection |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 2**. Application of sequential filters and their outcome on training set curation

| Filter | Criteria | Pass Rate (%) | Final Count |
|---|---|---|---|
| Dual data requirement | Both COX-2 and COX-1 $IC_{50}$ measured | 34.0 | 1,362 |
| Selectivity threshold | COX-2/COX-1 selectivity >10-fold | 78.5 | 1,069 |
| Pharmacophore | Sulfonamide/methylsulfonyl present | 100 | 1,069 |
| Drug-likeness | Lipinski Rule of 5 (≤1 violation) | 97.2 | 1,324 |
| PAINS removal | No pan-assay interference compounds | 93.9 | 1,279 |
| Toxicophore screen | No reactive groups | 82.5 | 1,123 |
| Withdrawn drug exclusion | Remove rofecoxib, valdecoxib, lumiracoxib | - | - |
| **Final training set** | **Maximum 2 per scaffold** | - | **308** |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 3.** Summary of virtual screening campaign using DDP inside ProteinLab

| Stage | Input | Output | Removed | Key Finding |
|---|---|---|---|---|
| Initial COX-2 screen | 2,047,122 | 238,689 | 1,808,433 | 11.7% hit rate |
| COX-1 screening | 238,689 | 238,689 | 0 | All scored for selectivity |
| Selectivity filter (>10-fold) | 238,689 | 33,327 | 205,362 | Only 14% achieved selectivity |
| mPGES-1 co-target | 33,327 | 33,143 | 184 | 99.4% dual activity |
| PTGIS liability | 33,143 | 29,817 | 3,326 | 10% removed |
| hERG liability | 29,817 | 21,480 | 8,337 | 28% removed (most restrictive) |
| CYP2C9 liability | 21,480 | 19,319 | 2,161 | 10% removed |
| CYP3A4 liability | 19,319 | 17,372 | 1,947 | 10% removed |
| CA-II liability | 17,372 | 16,984 | 388 | 2.2% removed |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 4.** Summary of results in medicinal chemistry optimization stage

| Filter | Criterion | Removed | Remaining | Survival |
|---|---|---|---|---|
| Molecular weight | 320-470 Da | 1,424 | 15,560 | 91.6% |
| Lipophilicity | cLogP ≤3.8 | 1,619 | 13,941 | 82.1% |
| TPSA | 60-120 Å² | 3,311 | 10,630 | 62.6% |
| Rotatable bonds | ≤8 | 6 | 10,624 | 62.6% |
| Aromatic rings | ≤3 | 2,442 | 8,182 | 48.2% |
| H-bond donors | ≤2 | 254 | 7,928 | 46.7% |
| H-bond acceptors | 4-9 | 301 | 7,627 | 44.9% |
| Halogens (-Cl, -Br, -I) | None | 2,814 | 4,813 | 28.3% |
| Imines | None | 40 | 4,773 | 28.1% |
| *p*-anilides | None | 189 | 4,584 | 27.0% |
| Michael acceptors | None | 17 | 4,567 | 26.9% |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 5.** Results of chemical novelty assessment

| Filter/Analysis | Criterion | Novel Compounds | Known Compounds | Novelty (%) |
|---|---|---|---|---|
| High-stringency novelty | Similarity ≥0.8 | 4,169 | 50 | 98.8 |
| Moderate-stringency novelty | Similarity ≥0.7 | 3,921 | 298 | 92.9 |
| Patent search | 44.3M global patents | 4,219 | 0 | 100 |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 6.** Performance summary of ADMET prediction models trained using KERMT framework across multiple prediction tasks

| Endpoint | Task | Metric | Performance |
|---|---|---|---|
| AMES | Classification | AUROC[a] | 0.88 |
| DILI | Classification | AUROC | 0.79 |
| hERG | Classification | AUROC | 0.899 |
| Cardiotoxicity | Classification | AUROC | 0.823 |
| *p*Ka | Regression | RMSE[b] | 1.51 ($R^2$ = 0.80) |
| logS | Regression | RMSE | 1.09 ($R^2$ = 0.74) |
| COX-1 (*p*IC50) | Regression | RMSE | 0.603 |
| COX-2 (*p*IC50) | Regression | RMSE | 0.775 |

[a]Area Under the Receiver Operating Characteristic Curve
[b]Root Mean Square Error



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 7.** Summary of protocol validation using distinct COX isoform structures

| PDB | Protein | Ligand | RMSD (Å) | Experimental Structure Affinity | Predicted Affinity |
|---|---|---|---|---|---|
| 5KIR | COX-2 | Rofecoxib | 0.817 | -10.1 | -10.1 |
| 5IKR | COX-2 | Mefenamic acid | 1.800 | -8.9 | -8.4 |
| 3N8Z | COX-1 | Flurbiprofen | 0.106 | -8.9 | -8.4 |
| 4PH9 | COX-2 | Ibuprofen | 0.839 | -7.4 | -7.4 |
| 3NT1 | COX-2 | Naproxen | 0.250 | -8.6 | -8.6 |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 8.** Selectivity distribution among the COX isoforms in three-structure campaign

| Category | $\Delta G_{sel}$ Range (kcal/mol) | Count | % of total |
|---|---|---|---|
| COX-1-selective | < 0 | 32 | 0.8 |
| Non-selective | 0–1 | 528 | 13.5 |
| Moderate COX-2 | 1–2 | 1,231 | 31.4 |
| Highly COX-2-selective | > 2 | 2,130 | 54.3 |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 9.** Summary of binding-mode distribution obtained following conformational expansion

| Structure | Molecules preferring | % of Total |
|---|---|---|
| 5IKR (Mefenamic acid) | 997 | 49.0 |
| 5KIR (Rofecoxib) | 782 | 38.4 |
| 4PH9 (Ibuprofen) | 156 | 7.7 |
| 3NT1 (Naproxen) | 100 | 4.9 |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 10.** Docking analyses of top-five hits against COX isoforms and mPGES-1

| Code | COX-2_ 4PH9 | COX-2_ 5KIR | mPGES-1_ 4YL3 | mPGES-1_8PYV | COX-1_ Predicted |
|---|---|---|---|---|---|
| **MAI_2** | -9 | -10.5 | -8.94 | -9.59 | -4.8 |
| **MAI_4** | -8.5 | -9.6 | -8.78 | -9.51 | -5.3 |
| **MAI_7** | -7.4 | -9.6 | -8.66 | -8.85 | -5.8 |
| **MAI_8** | -7.2 | -8.8 | -8.29 | -10.31 | -4.6 |
| **MAI_9** | -7.5 | -9.4 | -8.76 | -9.99 | -4.3 |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 11.** Comparative interaction summary for top-five hits at COX-2 (PDB ID: 5KIR)

| Code | Docking Energy (kcal/mol) | Key H-Bond Residues | Key Hydrophobic Residues | Mechanistic Inference |
|---|---|---|---|---|
| **MAI_2** | -10.5 | Ser353, Leu352, Gln192, Tyr385 | Phe518, Val349, Ala527 | Canonical COX-2 pose, strong sulfonamide anchoring |
| **MAI_4** | -9.6 | Arg120, Ser353, Gln192, Tyr355, Tyr385 | Phe518, Val523 | Extended H-bonding network, dual-pocket anchoring |
| **MAI_7** | -9.6 | Ser353, Leu352, Gln192, Tyr385 | Val523, Phe518 | Compact binding, efficient hydrophobic fit |
| **MAI_8** | -8.8 | Arg120, Ser353, Gln192, Tyr355, Tyr385 | Phe518, Val349, Ala527 | Deep dual-pocket engagement, highest affinity |
| **MAI_9** | -9.4 | Arg120, Ser353, Leu352, Gln192, Tyr385 | Val349, Leu359, Phe518, Ala527 | Broad coverage, rofecoxib-like orientation |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

**Table 12.** Comparative interaction summary for top-five hits at mPGES-1 (PDB ID: 8PYV)

| Code | Docking Energy (kcal/mol) | No. of H-Bonds | Key Polar Residues Involved | Hydrophobic Contacts | Mechanistic Inference |
|---|---|---|---|---|---|
| **MAI_2** | -9.59 | 4 | Arg38, Glu77, Tyr130, Ser127 | Phe44, Leu69 | GSH-site mimic, baseline binding |
| **MAI_4** | -9.51 | 5 | Arg73, Glu77, Tyr117, Ser127, Arg38 | Leu69, Phe44, His53 | Extended triad anchoring, deeper channel fit |
| **MAI_7** | -8.85 | 4 | Arg38, Arg73, Glu77, Ser127 | Phe44, Leu39, Leu69 | Compact, charge-driven binding |
| **MAI_8** | -10.31 | 6 | Arg73, His72, Asp75, Glu77, Ser127 | Leu69, Pro124, Ala123 | Broad polar network, strongest binding |
| **MAI_9** | -9.99 | 7 | Arg38, Arg73, Glu77, His53, Asp49, Tyr117, Ser127 | Phe44, Leu39, Leu69 | Full-site engagement, cofactor mimic, most extensive stabilization |



https://doi.org/10.26434/chemrxiv-2025-49xk0 **ORCID:** https://orcid.org/0009-0005-0488-0337 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0