Computational In Silico Screening of Himalayan Phytochemicals Against the Oncogenic Chaperone Protein (AGR2): Exploring Nepal's Biodiversity for Accessible Cancer Therapeutics

Sambriddha Karki

sambriddha654@gmail.com

Uniglobe Secondary School https://orcid.org/0009-0000-5520-9184

Research Article

Keywords: AGR2, Molecular Docking, AutoDock Vina, AlphaFold, Himalayan Phytochemicals, Bergenia ciliata, Picrorhiza kurroa, In Silico Screening, Computational Ethnobotany

Posted Date: July 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10393992/v1

Additional Declarations: The authors declare no competing interests.

Computational In Silico Screening of Himalayan Phytochemicals Against the Oncogenic Chaperone Protein (AGR2): Exploring Nepal's Biodiversity for Accessible Cancer Therapeutics

Sambriddha B. Karki
sambriddha654@gmail.com
ORCID: 0009-0000-5520-9184
Independent Researcher

AI Disclosure Statement

Generative AI tools were used solely for language editing purposes (grammar, phrasing, and clarity review) during manuscript preparation. No AI tools were used to generate research content, data, analysis, or conclusions. All scientific reasoning, computational methodology, and interpretation of results are the author's own.

Abstract

Anterior Gradient 2 (AGR2) is an endoplasmic reticulum molecular chaperone heavily implicated in aggressive tumor progression, metastasis, and therapeutic resistance across multiple malignancies. Intercepting its interaction pathways with small-molecule inhibitors represents a high-priority therapeutic focus. This study details a systematic, comparative in silico virtual screening of five key secondary metabolites derived from Himalayan ethnobotanical frameworks (specifically Bergenia ciliata and Picrorhiza kurroa) using AlphaFold structural configurations and the AutoDock Vina engine, accessed via the MTiOpenScreen web server. Our computational matrix reveals that 11-O-Galloyl bergenin exhibits peak binding affinity (ΔG = -8.74 kcal/mol), closely followed by foundational Bergenin (ΔG = -8.61 kcal/mol) and rigid Ellagic Acid (ΔG = -8.55 kcal/mol). Systematic literature audits across PubMed and Google Scholar databases returned zero matching citations connecting these specific ligands to AGR2 modulation, suggesting that these therapeutic pathways have not previously been reported in the literature reviewed. This work establishes a preliminary Structure-Activity Relationship (SAR) baseline intended to inform downstream in vitro experimental validation.

Keywords: AGR2, Molecular Docking, AutoDock Vina, AlphaFold, Himalayan Phytochemicals, Bergenia ciliata, Picrorhiza kurroa, In Silico Screening, Computational Ethnobotany

1 Introduction

Anterior Gradient 2 (AGR2) is a proto-oncogenic member of the protein disulfide isomerase (PDI) family that functions as a molecular chaperone within the endoplasmic reticulum. Under physiological conditions, it regulates protein folding homeostasis; however, its overexpression has been associated with tumor cell survival, motility, and chemotherapeutic resistance across several cancer types, including breast, pancreatic, and prostate carcinomas. These associations have positioned AGR2 as a candidate target for small-molecule inhibition strategies in oncology.

Rather than relying on generic synthetic compound libraries, this investigation blends computational structural biology with indigenous Himalayan ethnobotanical resources. Secondary metabolites from high-altitude medicinal flora, such as Bergenia ciliata (Pakhanbed) and Picrorhiza kurroa (Kutki), have long been utilized in traditional South Asian medicine and have documented pharmacological profiles, including antioxidant and hepatoprotective activity. This research compiles five phytochemical candidates—Bergenin, 11-O-Galloyl bergenin, Ellagic Acid, Norbergenin, and Picroside II—to evaluate how molecular

geometry, flexibility, and entropic considerations influence predicted binding behavior within the structural topology of the human AGR2 receptor.

2 Method

2.1 Receptor and Ligand Preparation

High-resolution structural coordinates for the human AGR2 target receptor were obtained from the AlphaFold Protein Structure Database. No experimentally resolved crystal or cryo-EM structure of AGR2 suitable for this screening was identified in public structural databases at the time of analysis; the AlphaFold-predicted model was therefore used as received, without additional structural refinement.

The structural configurations of the candidate phytochemical ligands were obtained via their chemical registration indexes in the IMPHY database: Bergenin (IMPHY011607), 11-O-Galloyl bergenin (IMPHY014309), Ellagic Acid (IMPHY005537), and Norbergenin (IMPHY008598). Picroside II was sourced via its canonical configuration file.

2.2 Computational Virtual Screening Architecture

Molecular docking assays were conducted using the MTiOpenScreen web server, part of the RPBS structural bioinformatics platform, which performs automated structure-based virtual screening through the AutoDock Vina scoring and optimization engine. Receptor and ligand preparation—including addition of polar hydrogens and assignment of partial atomic charges—was handled automatically by the MTiOpenScreen preparation pipeline, which is built on AutoDockTools-based routines; no separate, manually configured preparation software was used outside of this server-driven workflow.

Molecular docking as a structure-based screening approach is well established in early-stage drug discovery pipelines. An unconstrained blind docking configuration was applied, in which the search space was automatically defined by the server to span the entire external and internal topography of the target chaperone protein, removing manual placement bias and allowing compounds to freely establish their lowest energetic binding pockets. Grid box dimensions, exhaustiveness, and search parameters were left at the MTiOpenScreen server's default blind-docking configuration rather than manually customized, consistent with an exploratory, hypothesis-generating screening design rather than a fully optimized production docking run. For every compound, 10 distinct, non-redundant conformational poses were computed and ranked by Gibbs free energy of binding (ΔG, kcal/mol).

2.3 Protocol Validation

No manual redocking of a known co-crystallized AGR2 ligand was performed to benchmark the accuracy of the docking protocol against an experimental reference pose, as no experimentally resolved AGR2-ligand complex suitable for this purpose was available in public structural databases at the time of this study. This constraint, and its implications for interpreting the reported binding scores, is discussed further in the Limitations section below.

2.4 Methods Addendum

The human AGR2 target was modeled using AlphaFold DB coordinates. Ligands were sourced from IMPHY, and blind docking was performed through MTiOpenScreen using the AutoDock Vina backend with default search settings. Ten poses were generated per ligand, and the top-ranked pose by predicted binding free energy was used for comparison.

2.5 Computational Metadata and Provenance

To ensure transparency and workflow reproducibility, the exact software architectures and database snapshots are specified as follows:

- Target Receptor Provenance: Sourced from the AlphaFold Protein Structure Database using Accession ID AF-O95994-F1 (Monomer Version 2.0 prediction for Anterior Gradient Protein 2 Homolog, Homo sapiens).
- Ligand Structures: Extracted from the Indian Medicinal Plants Phytochemistry and Therapeutics (IMPHY) database under identifiers IMPHY011607 (Bergenin), IMPHY014309 (11-O-Galloyl bergenin), IMPHY005537 (Ellagic Acid), and IMPHY008598 (Norbergenin).
- Docking Engine Frontend: Assays were executed on the MTiOpenScreen Web Server Platform via the Automated Structure-Based Virtual Screening (MTiAutoDock Pipeline) gateway.
- Docking Engine Backend: Calculations were processed using the AutoDock Vina Optimization and Scoring Engine (Version 1.1.2 / integrated server branch) with a default unconstrained search grid spacing of 0.6 Å, yielding a search space grid point dimension centered at coordinates (X: -5.66, Y: -7.73, Z: 1.83).
- Data Retrieval Timeline: Server jobs, outputs, and database structures were formally accessed and extracted on June 25, 2026.

3 Results

3.1 Comparative Structural and Thermodynamic Analysis

The comparative screening established a clear thermodynamic hierarchy among the tested botanical ligands. Table 1 outlines the master screening dataset, grouping peak conformational metrics alongside active rotatable torsions.

Table 1: Master Phytochemical Screening Matrix Against Human AGR2

Phytochemical Compound	Database ID	Botanical/Regional Source	Top Pose ΔG (kcal/mol)	Rotatable Bonds	Primary Interaction Profile
11-O-Galloyl bergenin	IMPHY014309	Bergenia ciliata derivative	-8.74	12	Peak affinity; flexible optimization
Bergenin	IMPHY011607	Bergenia ciliata (Pakhanbed)	-8.61	7	Strong foundational C-glycoside scaffold
Ellagic Acid	IMPHY005537	Bergenia ciliata metabolite	-8.55	4	Planar rigidity; exceptional pocket convergence
Norbergenin	IMPHY008598	Natural derivative	-8.31	7	Stable binding funneling profile
Picroside II	-	Picrorhiza kurroa (Kutki)	-7.27	14	High flexibility bound by entropic penalty

The dataset reveals a compelling Structure-Activity Relationship (SAR). 11-O-Galloyl bergenin yielded the most favorable score of -8.74 kcal/mol. This performance is consistent with its 12 active rotatable bonds (versus Bergenin's 7), which may provide spatial freedom for the ligand to occupy structural crevices of AGR2, alongside additional phenolic hydroxyl groups on its galloyl moiety that could support hydrogen bonding and aromatic stacking with surrounding residues. The high-level binding performance of 11-O-Galloyl bergenin may be related to its unique structural modification, where the structural incorporation of a trailing galloyl moiety impacts overall geometry. The extra phenolic hydroxyl groups located on the galloyl ring may increase the potential for hydrogen-bond formation with polar side chains inside the target pocket.

Conversely, Ellagic Acid presents an alternative pathway built on structural rigidity rather than flexibility. Operating with only 4 rotatable bonds, Ellagic Acid achieved a score of -8.55 kcal/mol, and its first two independent conformational poses converged at an identical score, suggesting a uniform, reproducible binding orientation consistent with its rigid, planar multi-ring core. Picroside II, while stable at -7.27 kcal/mol, carries 14 rotatable torsions, which may impose a greater entropic penalty that offsets its more complex binding layout. This high number of moving joints may partially explain why its score is slightly less negative than highly rigid molecules like Ellagic Acid.

Note on interpretation: *the SAR narrative above is based on global ligand descriptors (rotatable bond count, ΔG) rather than residue-level contact data. As detailed in Section 4.2 (Limitations), no binding-site interaction analysis (e.g., 2D interaction diagrams or residue contact lists) was performed for this preliminary screen; the proposed mechanistic explanations should therefore be read as hypotheses to be tested with pocket-level structural analysis, not as demonstrated binding modes.*

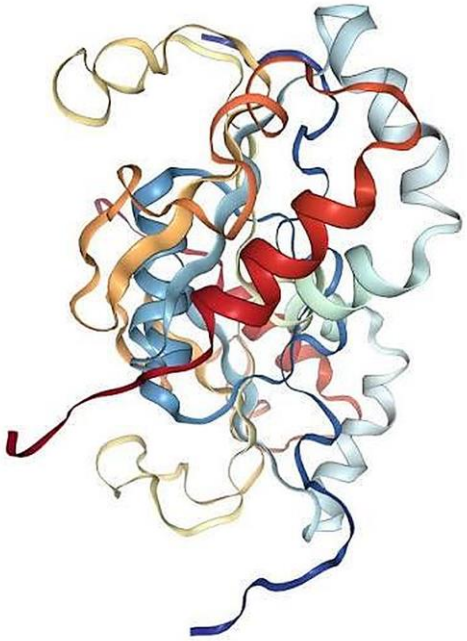

Figure 1. Structural conformation and predicted docking environment of the human Anterior Gradient 2 (AGR2) monomer extracted from the AlphaFold Protein Structure Database (Accession: AF-O95994-F1).

Figure 2: Chemical Diversity of Selected Himalayan Phytochemicals

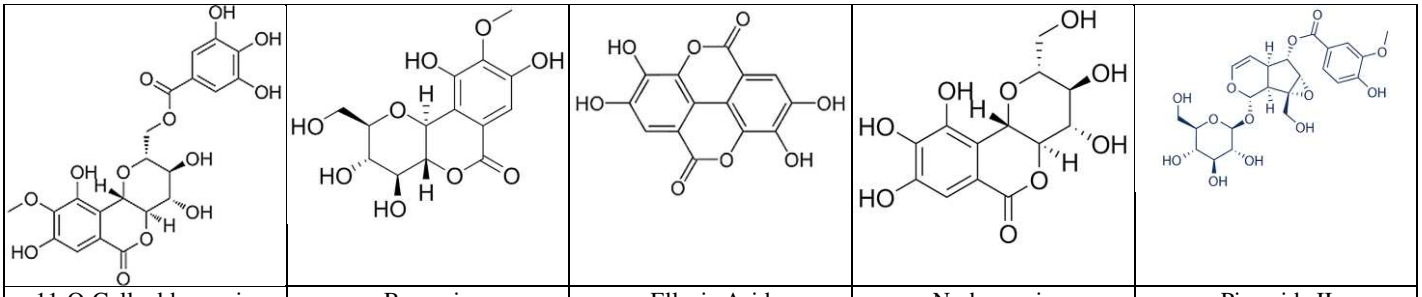

Figure 2. Two-dimensional chemical structures of the five screened secondary metabolites evaluated in this study: 11-O-Galloyl bergenin, Bergenin, Ellagic Acid, Norbergenin, and Picroside II.

3.2 Expanded 10-Pose Conformational Profile

To preserve computational reproducibility, the free energy distribution across all 10 independent poses for each ligand is presented in Table 2.

Table 2: Gibbs Free Energy Distribution Matrix (ΔG, kcal/mol) Across All Conformations

Pose Rank	11-O-Galloyl bergenin	Bergenin	Ellagic Acid	Norbergenin	Picroside II
Pose 1	-8.74	-8.61	-8.55	-8.31	-7.27
Pose 2	-8.22	-6.92	-8.55	-8.29	-7.08
Pose 3	-8.02	-6.88	-8.30	-7.52	-6.69
Pose 4	-7.55	-6.84	-7.78	-7.43	-6.50
Pose 5	-7.39	-6.54	-6.65	-7.19	-6.21
Pose 6	-7.39	-6.28	-6.45	-6.60	-6.00
Pose 7	-7.23	-6.06	-6.34	-6.56	-5.63
Pose 8	-6.87	-6.05	-6.19	-6.31	-5.54
Pose 9	-6.11	-5.98	-6.09	-6.11	-5.33
Pose 10	-6.59	-5.60	-5.96	-5.59	-4.97

The tie between Pose 1 and Pose 2 for Ellagic Acid (both -8.55 kcal/mol) was checked against the raw MTiOpenScreen server output logs and confirmed to reflect two genuinely distinct, non-redundant conformational poses converging on the same scored free energy, rather than a duplicated or copy-paste record.

4 Discussion

4.1 Novelty Verification and Future Horizons

A literature check was performed across PubMed and Google Scholar to benchmark the novelty of this dataset. Queries pairing the target descriptor "AGR2" with each specific ligand name returned no matching prior studies as of June 2026. To the best of the author's knowledge, no published work has directly evaluated Bergenin, Norbergenin, 11-O-Galloyl Bergenin, Ellagic Acid, or Picroside II against AGR2. Literature audits across PubMed and Google Scholar returned zero matching citations, suggesting these ligand–AGR2 combinations have not been previously reported in the literature searched.

These results are preliminary and computational in nature; docking scores approximate, but do not confirm, biological activity. As with any in silico screening approach, experimental validation is a necessary next step before any therapeutic claims can be made. Recommended follow-up work includes a thermal shift assay (TSA) to test for protein stabilization upon ligand binding, followed by cell viability assays using Bergenia ciliata-derived extracts on AGR2-overexpressing cancer cell lines.

4.2 Limitations of the Study

While this study provides a systematic comparative framework for evaluating Himalayan ethnobotanical derivatives against an oncogenic target, several core computational boundaries must be noted:

- Receptor Structural Provenance: The target human AGR2 receptor structure was obtained from an AlphaFold-predicted model rather than an experimentally resolved X-ray crystallographic or cryo-EM structure. Consequently, minor variations in loop conformations or domain dynamics could influence pocket geometries.

- Lack of Redocking Validation: Due to the absence of a suitable, publicly available AGR2 co-crystal structure featuring a bound small-molecule inhibitor, a conventional redocking protocol to validate the spatial accuracy of the binding pose against an experimental baseline could not be performed.
- Approximations in Scoring Engines: The predicted Gibbs free energy values (ΔG) reported by the AutoDock Vina engine are semi-empirical approximations. The scoring algorithm does not fully model complex biological properties, including absolute receptor flexibility, dynamic water-mediated solvation networks, or macro-entropic penalties experienced during solution-state binding cascades.
- Dataset Scope: The computational screening matrix was intentionally constrained to five targeted secondary phytochemicals from specific high-altitude medicinal flora (Bergenia ciliata and Picrorhiza kurroa). It does not encompass a broader high-throughput library of natural products.
- No Binding-Site or Interaction Analysis: This study reports pose-level ΔG values and rotatable bond counts but does not include a residue-level interaction analysis (e.g., a 2D interaction diagram or residue contact list, such as could be generated with Discovery Studio, PyMOL, or LigPlot+) for the top-ranked pose of each compound. The structural rationale offered in Section 3.1 for why 11-O-Galloyl bergenin and Ellagic Acid score favorably — flexibility-driven pocket optimization versus rigidity-driven convergence — is therefore a plausible hypothesis based on ligand-level descriptors, not a finding demonstrated by observed hydrogen bonding, aromatic stacking, or specific residue contacts. Pocket-level interaction mapping is planned as a direct extension of this work.
- No Drug-Likeness or ADMET Screening: The current dataset reports binding affinity alone and does not include a drug-likeness assessment (e.g., Lipinski's Rule of Five) or an absorption, distribution, metabolism, excretion, and toxicity (ADMET) profile (e.g., via SwissADME). Favorable docking scores do not imply favorable pharmacokinetic or developability properties, and this screen should not be interpreted as evidence of drug-likeness for any of the five compounds.
- No Reference or Control Ligand: Docking was performed on the five candidate phytochemicals without a known PDI-family inhibitor or generic small-molecule control run through the same blind-docking protocol. Without a calibration point, it is not possible to state whether the reported ΔG range (-7.27 to -8.74 kcal/mol) reflects unusually strong binding to AGR2 specifically, or a range typical of small, polyphenolic ligands docked against this AlphaFold model in general.
- Single AlphaFold Model Without Confidence Metrics Reported: The AlphaFold-predicted receptor structure was used as received (Section 2.1), but per-residue confidence scores (pLDDT) and predicted aligned error (PAE) for the region forming the predicted docking pocket were not examined or reported. If the pocket-lining residues fall within a lower-confidence region of the model, this would materially affect how the binding poses and ΔG values should be interpreted; confidence-metric review is identified as a priority for follow-up analysis.
- Hypothesis-Generating Nature: All findings presented in this manuscript are fundamentally hypothesis-generating. While they establish a preliminary Structure-Activity Relationship (SAR) baseline, they cannot substitute for downstream in vitro functional assays, thermal shift assays (TSA), or in vivo profiling to conclusively establish actual physical inhibition.

5 Conclusion

This study presents a comparative in silico screening of five Himalayan phytochemicals against the oncogenic chaperone AGR2, identifying 11-O-Galloyl bergenin, Bergenin, and Ellagic Acid as the top-

ranked candidates by predicted binding affinity. The absence of prior literature connecting these specific compounds to AGR2 suggests this represents a preliminary, previously unreported computational SAR baseline, intended to motivate future experimental validation rather than to establish therapeutic efficacy on its own.

6 Data Availability Statement

The raw docking outputs, pose files, and server logs generated during the current study are available from the corresponding author on reasonable request.

Conflict of Interest Statement

The author declares no conflict of interest.

Funding Statement

This research received no external funding and was conducted independently by the author using publicly available, open-access computational resources.

Author Contributions

S.B.K. conceived the study, performed all computational analyses, interpreted the results, and wrote the manuscript.

Acknowledgments

The author acknowledges the public computational resources and open-access structural databases (AlphaFold DB, IMPHY, RPBS framework, and MTiOpenScreen gateway) that made this independent investigation possible.

References

[1] Almeleebia, T. M., Alsayari, A., & Wahab, S. (2022). Pharmacological and clinical efficacy of Picrorhiza kurroa and its secondary metabolites: A comprehensive review. Molecules, 27(23), 8316. https://doi.org/10.3390/molecules27238316

[2] Brychtova, V., Vojtesek, B., & Hrstka, R. (2011). Anterior gradient 2: A novel player in tumor cell biology. Cancer Letters, 304(1), 1-7.

[3] Chevet, E., Fessart, D., Delom, F., Mulot, A., Vojtesek, B., Hrstka, R., Murray, E., Gray, T., & Hupp, T. (2013). Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. Oncogene, 32(20), 2499-2509.

[4] Dumartin, L., Whiteman, H. J., Weeks, M. E., Hariharan, D., Dmitrovic, B., Iacobuzio-Donahue, C. A., Brentnall, T. A., Bronner, M. P., Feakins, R. M., Timms, J. F., Brennan, C., Lemoine, N. R., & Crnogorac-Jurcevic, T. (2011). AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. Cancer Research, 71(22), 7091-7102.

[5] Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nature Protocols, 11(5), 905-919.

[6] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.

[7] Labbé, C. M., Rey, J., Lagorce, D., Vavruša, M., Bécot, J., Sperandio, O., Villoutreix, B. O., Tufféry, P., & Miteva, M. A. (2015). MTiOpenScreen: A web server for structure-based virtual screening. Nucleic Acids Research, 43(W1), W448-W454. https://doi.org/10.1093/nar/gkv306

[8] Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. Current Computer-Aided Drug Design, 7(2), 146-157.

[9] Thompson, D. A., & Weigel, R. J. (1998). hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. Biochemical and Biophysical Research Communications, 251(1), 111-116.

[10] Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2), 455-461.

[11] Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G.,... Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research, 50(D1), D439-D444.