Title: Integrating Ethnobotanical, Phytochemical and Structural Information for In Silico Prioritization of Bioactive Compounds from Medicinal Plants of Roraima

Author(s): Rhilary Herielle Gomes Pereira (1); Kalil Anthony Carvalho Davi (2); Fabiano Siqueira de Almeida (3); Iane Santana Morais (4). Karla Santana Morais (1, 2*)

Affiliation(s): (1) Post-Graduate Program in Biodiversity and Biotechnology of the Legal Amazon (Bionorte), Boa Vista, Roraima, Brazil; (2) Federal Institute of Roraima (IFRR), Boa Vista, Roraima, Brazil; (3) Federal Institute of Paraíba (IFPB), João Pessoa, Paraíba, Brazil; (4) State Secretariat of Health of the Federal District (SES-DF), Brasília, Federal District, Brazil.

Abstract: Roraima, located in the far north of the Brazilian Amazon, combines high biological diversity with a robust tradition of medicinal plant use from indigenous people, offering a relevant context for the integration of heterogeneous biological information. Nevertheless, the transition from ethnopharmacological knowledge to reproducible molecular prioritization remains limited, particularly in the context of neglected diseases. The present study introduces an *in silico* workflow developed to integrate ethnobotanical use records, phytochemical evidence, and protein–ligand structural interaction data, with the aim of supporting the prioritization of candidate bioactive compounds derived from medicinal plants used in Roraima. Three species–compound–target systems were selected on the basis of primary phytochemical and pharmacological literature: *Aspidosperma excelsum*–aspidodasycarpine–PfLDH (malaria), *Himatanthus articulatus*–plumericin–InhA (tuberculosis), and *Anacardium occidentale*–anacardic acid triene (15:3)–trypanothione reductase (leishmaniasis). Three-dimensional structures of ligands and proteins were retrieved from PubChem and the Protein Data Bank, respectively. Receptors and ligands were prepared using Meeko, and molecular docking was performed with AutoDock Vina. Predicted interactions were compared against reference ligands and evaluated through redocking analyses to enhance traceability and internal consistency. All three compounds exhibited negative binding energies, supporting their prioritization for subsequent investigations. Aspidodasycarpine yielded a docking score of −5.844 kcal/mol against PfLDH, a value closely approaching that of the reference ligand chloroquine (−5.866 kcal/mol) under the same protocol conditions. Anacardic acid triene (15:3) demonstrated the most favorable score (−8.182 kcal/mol) against trypanothione reductase, but remaining inferior to the native cofactor FAD. Redocking analyses reproduced the expected local binding behavior, with the 4TZK system exhibiting the most consistent performance. Although not centered on formal information-theoretic metrics, this study addresses a relevant data integration challenge in computational biology: how to extract actionable molecular signal from heterogeneous ethnobotanical, phytochemical, and structural datasets under conditions of uncertainty. This approach contributes to the initial ranking of therapeutic hypotheses and underscores the value of multimodal evidence integration for systematically neglected disease research.

Keywords: molecular docking (1); keyword multimodal data integration (2); bioactive compound prioritization (3).

Author Profiles(s): Rhilary Herielle Gomes Pereira [ORCID: 0000-0002-4910-9221]; Fabiano Siqueira de Almeida [ORCID: 0000-0003-4191-1399]; Karla Santana Morais [ORCID: 0009-0000-0611-5480]