# COMPUTATIONAL STUDY OF SELECTED ANTIMALARIAL AND ANTICANCER ACYLPHLOROGLUCINOLS

By

Neani Tshilande

(11630094)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

In the

Department of Chemistry

Faculty of science, engineering and agriculture

University of Venda

Thohoyandou, Limpopo

South Africa

*Supervisor*

PROFESSOR LILIANA MAMMINO

*Co-supervisor*

Doctor CATERINA GHIO

© University of Venda

# Declaration

I, Neani Tshilande, declare that the thesis for the award of a doctoral degree I submitted has never previously been submitted for a higher degree or diploma at this or any other university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Signature TSHILANDE N Date: 01 March 2024



© University of Venda

# Abstract

Malaria and cancer tend to become drug-resistant few years after a drug is introduced into clinical use. This prompts the search for new molecular structures that are sufficiently different from the drugs for which resistance has developed. Acylphloroglucinols (ACPLs) are natural compounds with several biological activities. They are considered as possible lead structures for developing drugs against degenerative and other diseases. Computational studies are particularly important for biologically active compounds because their activities depend on their molecular properties, and knowing these properties (descriptors) is necessary for drug development.

Representative ACPLs with anticancer and/or antimalarial activities were selected and computationally studied *in vacuo* and in three solvents with different polarities. All the calculations were done with completely relaxed geometry, using the Hartree Fock (HF) and Density Functional Theory (DFT) methods; second-order Moller-Plesset perturbation theory (MP2) calculations were also performed *in vacuo*. Structure-based virtual screening was used to study the interactions between these molecules and relevant proteins. The conformational studies identified conformational preferences and conformers' stabilizing factors, among which intramolecular hydrogen bonds (IHBs) have dominant roles. The highest-occupied molecular orbital (HOMO) - lowest-unoccupied molecular orbital (LUMO) energy gap decreases and the dipole moment increases as the medium polarity increases. The solvent effect depends on the conformer type and the solvent polarity. The molecular docking analysis shows that most of these compounds bind well with the selected proteins and highlights the type of molecule-protein interactions for each case. These results indicate that the selected molecules are interesting for further steps in view of their possible development into antimalarial and anticancer drugs.



© University of Venda

University of Venda
Creating Future Leaders

# Acknowledgments

This thesis is a tribute to my late uncle, Hulisani Samuel Lirumo, whose unwavering encouragement was a guiding light in my academic journey. His belief in my potential was a constant source of inspiration, and I carry his memory with me as I reach this significant milestone.

I am deeply grateful to my promoter, Prof. Liliana Mammino, and my co-promoter, Dr. Caterina Ghio, for their invaluable guidance, patience, and support. Their mentorship has been instrumental in shaping this work and my growth as a researcher. I also extend my thanks to the Department of Chemistry at the University of Venda for providing a supportive environment for my research.

This work would not have been possible without the resources provided by the Centre for High Performance Computing (CHPC, South Africa). I am also grateful for the financial support from the National Research Foundation (DST-NRF Innovation Doctoral Scholarship, grant number 118453) and the S&F - Sasol Inzalo Foundation. Both allowed me to dedicate myself to this research fully.

To my family—mother, sisters, grandmother, uncles, and cousins, I owe a debt of gratitude for your unwavering belief in me. Your encouragement was a constant source of strength, especially during challenging times. I also want to thank my friends who shared this journey with me, particularly Emmanuel Mashile, whose technical support was invaluable. And to my partner, Emmanuel, your unwavering support and belief in me have been a constant source of inspiration.

Lastly, I want to thank everyone who, in their way, supported me along this path. This thesis results from each of your contributions, and I am grateful to have had you all with me.



© University of Venda

# Motivation to do this study

Think about how frustrating it would be if the tools you rely on stopped working overtime. That's what happens with many medicines used to treat severe illnesses like malaria and cancer. At first, these medicines are effective, but eventually, the diseases adapt, making the drugs less helpful or even ineffective. This resistance problem pushes scientists to look for new medicines that can continue to work in the long run.

This study explores new ways to create treatments for malaria and cancer using a group of natural compounds called acylphloroglucinols. These molecules in certain plants have shown potential for fighting diseases. Computer models enable the close study of these molecules to understand their structure and how they might act in the body. Such analysis can help identify if these molecules could interact effectively with targets in the cells affected by malaria or cancer.

If the study shows that these molecules have the right potential, they could be further developed into new medicines. This approach saves time and resources by identifying the most promising candidates without needing to test each one immediately in a lab. In the long term, findings from this research could contribute to effective treatments against these diseases, offering people better chances to get well and stay healthy over time.



© University of Venda

# Acronyms

| | |
|---|---|
| 5-FU | 5-fluorouracile |
| 5Z | quintuple zeta (z) |
| ABC | ATP-binding cassette |
| ACO | ant colony optimization |
| ACPL | acylphloroglucinol |
| ADME | Absorption, Distribution, Metabolism, and Excretion |
| AIM | atoms in molecules |
| ALI | acute lung injury |
| ANN | artificial neural network |
| APC | adenomatous polyposis coli gene |
| ARDS | acute respiratory distress syndrome |
| ASC | apparent surface charge |
| ATP | adenosine triphosphate |
| BBB | blood-brain barrier |
| BCL2 | B-cell lymphoma 2 |
| BCP | bond critical points |
| BCRP | breast cancer resistance protein |
| BE | bond energy |
| BFGS | Broyden-Fletcher-Goldfarb-Shanno |
| BRCA1 | breast cancer gene 1 |
| BRCA2 | breast cancer gene 2 |
| BO | Born-Oppenheimer |
| BSSE | basis set superposition error |
| CAHBs | charge-assisted hydrogen bonds |
| Caco-2 cells | human epithelial colorectal adenocarcinoma cell line |
| cc | correlation consistent |
| CC | coupled cluster |
| CC3 | coupled cluster with triple excitations |
| CDK | cyclin-dependent kinase |
| CCSD | coupled cluster with single and double excitations |
| cc-p | correlation consistent polarized |
| CHA | chemical Hamiltonian approach |
| CI | configuration interaction |
| CISD | configuration interaction Simple singles and doubles |
| CML | chronic myeloid leukemia |
| CNS | central nervous system |
| COX-2 | cyclooxygenase 2 |
| CP-ACPL | cyclic polyketide |
| CPCM | conductor-like polarizable continuum model |
| CR | Coulombic repulsion |
| CSF | classical scoring function |
| D-ACPL | dimeric ACPL |
| DFT | density-functional theory |
| DHFR | dihydrofolate reductase |



© University of Venda

| | |
|---|---|
| DHPS | dihydropteroate synthase |
| DFP | Davidon-Fletcher-Powell |
| DL | deep learning |
| DMC | diffusion Monte Carlo |
| DNA | deoxyribonucleic acid |
| DPC4 | deleted in pancreatic cancer-4 |
| DPCM | dielectric polarizable continuum model |
| DRAMP | drug resistance-associated membrane protein |
| DZ | double zeta (z) |
| DZP | double zeta (z) plus polarization |
| EA | ethacrynic acid |
| ECM | extracellular matrix |
| EG | Euglobals |
| EGFR | epidermal growth factor receptor, erb-B |
| erb-B | epidermal growth factor receptor, EGFR |
| erb-B2 | epidermal growth factor receptor 2, HER2 |
| ERCC1 | excision repair cross-complementing protein |
| ET-743 | ecteinascidin-743 |
| EU | euglobal |
| FDA | Food and Drug Administration |
| FEP | free energy perturbation |
| G | grandinol |
| G6PD | glucose-6-phosphate dehydrogenase |
| GB | generalised-Born |
| GBDT | gradient boosting decision tree |
| GGA | generalized gradient approximation |
| GSH | glutathione |
| GTO | Gaussian-type orbital |
| HB | hydrogen bond |
| HBA | hydrogen bond acceptor |
| HBD | hydrogen bond donor |
| Her-1 | epidermal growth factor receptor 1 |
| HER2 | epidermal growth factor receptor 2, erb-B2 |
| HF | Hartree-Fock |
| HOMO | highest-occupied molecular orbital |
| HPLC | high-pressure liquid chromatography |
| HR | homologous recombination |
| HAS | human serum albumin |
| HTS | high-throughput screening |
| IC | incremental construction |
| $IC_{50}$ | half-maximal inhibitory concentration |
| IEFPC | integral equation formalism polarizable continuum model |
| IHB | intermolecular hydrogen bonds |
| IPCM | isodensity polarizable continuum model |
| IR | infrared |
| LBHB | low-barrier hydrogen bond |



© University of Venda

| | |
|---|---|
| LCAO | linear combination of atomic orbitals |
| LDA | local density approximation |
| LH-RH | luteinizing hormone-releasing hormone |
| LUMO | lowest-unoccupied molecular orbital |
| LUDI | ligand unbiased dynamic search |
| MA | matching algorithm |
| MC | Monte Carlo |
| MCSS | multiple copy simultaneous search |
| MD | molecular dynamics |
| MDCK | Madin-Darby Canine Kidney |
| MDM2 | murine double minute 2 |
| MDR | multidrug resistance |
| MFS | major facilitator superfamily |
| microRNA | micro ribonuicleic acid |
| MLSF | machine learning scoring function |
| MMP | matrix metalloproteinase |
| MO | molecular orbital |
| MP | Møller-Plesset perturbation theory |
| MPPT | Møller-Plesset perturbation theory |
| MP2 | second-order Møller-Plesset perturbation theory |
| MP4 | fourth-order Møller-Plesset perturbation theory |
| MP4 (SDQ) | fourth-order Møller-Plesset perturbation theory |
| MP5 | fifth-order Møller-Plesset perturbation theory |
| MP6 | sixth-order Møller-Plesset perturbation theory |
| mRNA | messenger ribonucleic acid |
| MRP | multidrug resistance-associated protein |
| MT | microtubule |
| MTS1 | multiple tumor suppressor 1 |
| mTOR | mechanistic target of rapamycin |
| MW | molecular weight |
| NADPH | nicotinamide adenine dinucleotide phosphate |
| NBO | natural bond orbital |
| NER | nucleotide excision repair |
| NF-1 | neurofibromatosis type 1 |
| NF-2 | neurofibromatosis type 2 |
| NHEJ | non-homologous end joining |
| NMR | nuclear magnetic resonance |
| NSCLC | non-small cell lung cancer |
| PABA | p-aminobenzoic acid |
| P-ACP | prenylated ACPL |
| PB | Poisson-Boltzmann |
| PCM | polarizable continuum model |
| PCP-ACPL | polycyclic polyprenylated ACPL |
| PDB | Protein Data Bank |
| PDGF | platelet-derived growth factor |
| PE | photoelectron |



© University of Venda

| | |
|---|---|
| PES | potential energy surface |
| Pfcrt | *P. falciparum* chloroquine resistance transporter |
| Pfmdr | *Plasmodium falciparum* multidrug resistance |
| Pfmdr1 | *P. falciparum* multidrug resistance 1 |
| PG-ACPL | phloroglucinol glycosides |
| P-gp | P-glycoprotein |
| PHOA | percentage of predicted human oral absorption |
| PHS | potential hyper surface |
| PI | phosphatidylinositol |
| PP-ACPL | polyprenylated ACPL |
| PTA-ACPL | phloroglucinol-terpene adduct ACPL |
| PTHrP | parathyroid hormone-related protein |
| p53 | tumor protein 53 |
| QMC | quantum Monte Carlo |
| QPlogBB | predicted blood-brain partition coefficient |
| QPlogKhsa | predicted binding affinity to HAS |
| QPPCaco | predicted apparent permeability across Caco-2 |
| QPPMDCK | predicted apparent permeability across MDCK |
| QSAR | Quantitative Structure-Activity Relationships |
| QZ | quadruple zeta (z) |
| QZP | quadruple zeta (z) plus polarization |
| RAHB | resonance-assisted hydrogen bond |
| RB | retinoblastoma protein |
| RBC | red blood cell |
| RF | random forest |
| RNA | ribonucleic acid |
| RND | resistance-nodulation-cell division |
| ROS | reactive oxygen species |
| RTK | receptor tyrosine kinase |
| SAR | structure activity relationship |
| SCF | self-consistent field |
| SCIPCM | self-consistent isodensity polarizable continuum model |
| SOS | save our ship |
| STO | Slater-type orbital |
| spp. | several species |
| SV | split valence |
| SVM | support vector machines |
| SVP | split valence plus polarization |
| SZ | single zeta (z) |
| T-ACPL | trimeric ACPL |
| TDP2 | tyrosyl-DNA phosphodiesterase 2 |
| TKI | tyrosine kinase inhibitor |
| TS | tabu search |
| TZ | triple zeta (z) |
| UGT | uridine 5'-diphospho-glucuronosyltransferase |
| US | United States |



© University of Venda

University of Venda
Creating Future Leaders

| | |
|---|---|
| UV | ultraviolet |
| V | valence (only) |
| VDW | van der Waals |
| VEGF | vascular endothelial growth factor |
| VEGF-1 | vascular endothelial growth factor 1 |
| VEGF-2 | vascular endothelial growth factor 2 |
| VHL | von Hippel-Lindau |
| VMC | variational Monte Carlo |
| WHO | World Health Organization |
| WT1 | Wilms' tumor-1 protein |
| XPE | xeroderma pigmentosum group E binding factor |
| ZPE | zero-point energy |



© University of Venda

# List of figures

**Figures in the text.**





© University of Venda

© University of Venda

© University of Venda

© University of Venda

**Figures in supplementary materials. S before the figure's number, indicates that the figure is in the supplementary materials.**





© University of Venda

© University of Venda

University of Venda
Creating Future Leaders

# List of tables

**Tables in the text.**





© University of Venda

**Tables in supplementary materials. S before the table's number, indicates that the table is in the supplementary materials.**





© University of Venda

© University of Venda

57. **Table S4.35**. Global reactivity descriptors for the calculated conformers of ε- and ζ-isomer euglobal molecules. MP2/6-31G(d,p) results.
58. **Table S4.36**. Comparison of the chemical hardness (η, table a) and chemical softness (S, table b) of the calculated conformers of corresponding ε and ζ euglobal isomers, across the three calculation methods utilised.
59. **Table S4.37**. Comparison of the electronic chemical potential (μ, table a) and the absolute electronegativity (χ, table b) of the calculated conformers of corresponding ε and ζ euglobal isomers, across the three calculation methods utilised.
60. **Table S4.38**. Comparison of the electrophilicity (ω, table a) and the maximum electronic charge ($\Delta N_{max}$, table b) of the calculated conformers of corresponding ε and ζ euglobal isomers, across the three calculation methods utilised.
61. **Table S4.39**. Global reactivity descriptors for the calculated conformers of τ-type model euglobal structures. DFT/B3LYP/6-31 + G(d,p) results.
62. **Table S4.40**. Global reactivity descriptors for the calculated conformers of τ-type model euglobal structures. HF/6-31G(d,p) results.
63. **Table S4.41**. Global reactivity descriptors for the calculated conformers of τ-type model euglobal structures. MP2/6-31G(d,p) results.
64. **Table S4.42**. Dipole moments of the calculated conformers of the euglobal molecules considered in this work.
65. **Table S4.43**. Mulliken atomic charges on the O atoms in the calculated conformers of ε- and ζ-isomer euglobal molecules. DFT/B3LYP/6-31 + G(d,p) results.
66. **Table S4.44**. Mulliken atomic charges on the O atoms in the calculated conformers of ε- and ζ-isomer euglobal molecules. HF/6-31G(d,p) results.
67. **Table S4.45**. Mulliken atomic charges on the O atoms in the calculated conformers of ε- and ζ-isomer euglobal molecules. MP2/6-31G(d,p) results.
68. **Table S4.46**. Ranges of the Mulliken atomic charges on the O atoms in the calculated conformers of ε- and ζ-isomer euglobal molecules, in terms of the intramolecular hydrogen bonds present. DFT/B3LYP/6-31 + G(d,p) results.
69. **Table S4.47**. Ranges of the Mulliken atomic charges on the O atoms in the calculated conformers of ε- and ζ-isomer euglobal molecules, in terms of the intramolecular hydrogen bonds present. MP2/6-31G(d,p) results.
70. **Table S4.48**. Mulliken atomic charges on the H atoms of the OH groups in the calculated conformers of ε- and ζ-isomer euglobal molecules.
71. **Table S4.49**. Mulliken charges on the O atoms in the conformers of the calculated τ-type model euglobal structures. DFT/B3LYP/6-31 + G(d,p) results.
72. **Table S4.50**. Mulliken charges on the O atoms in the conformers of the calculated τ-type model euglobal structures. HF/6-31G(d,p) results.
73. **Table S4.51**. Mulliken charges on the O atoms in the conformers of the calculated τ-type model euglobal structures. MP2/6-31G(d,p) results.
74. **Table S4.52.** Mulliken atomic charges on the H atoms of the OH groups in the conformers of the calculated τ-type model euglobal structures.
75. **Table S4.53**. Relative energies of the calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents, chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn and aq in the column headings). HF/6-31G(d,p) results.
76. **Table S4.54.** Relative energies of the calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents, chloroform, acetonitrile and water



© University of Venda

(respectively denoted as vac, chlrf, actn and aq in the column headings). DFT/B3LYP/6-31+G(d,p results.

77. **Table S4.55.** The solvation free energy ($\Delta G_{solv}$) and its electrostatic component ($G_{el}$) of the calculated conformers of the euglobal molecules considered in this work, in three solvents chloroform, acetonitrile and water (respectively denoted as chlrf, actn, aq in the column headings). HF/6-31G(d,p) results.
78. **Table S4.56.** The solvation free energy ($\Delta G_{solv}$) and its electrostatic component ($G_{el}$) of the calculated conformers of the euglobal molecules considered in this work, in three solvents chloroform, acetonitrile and water (respectively denoted as chlrf, actn, aq in the column headings). DFT/B3LYP/6-31+G(d,p results.
79. **Table S4.57.** HOMO-LUMO energy gap of the calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn, aq in the column headings). HF/6-31G(d,p) results.
80. **Table S4.78.** HOMO-LUMO energy gap of the calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn, aq in the column headings). DFT/B3LYP/6-31+G (d,p) results.
81. **Table S4.59.** Parameters of the O–H···O intramolecular hydrogen bonds in calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn, aq in the column headings).HF/6-31G(d,p) results.
82. **Table S4.60.** Parameters of the O–H···O intramolecular hydrogen bonds in the calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn, aq in the column headings).DFT/B3LYP/6-31+G(d,p) results.
83. **Table S4.61.** Dipole moment of the calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn, aq in the column headings). HF/6-31G(d,p) results.
84. **Table S4.62.** Dipole moment of the calculated conformers of the euglobal molecules considered in this work, *in vacuo* and in three solvents chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn, aq in the column headings). DFT/B3LYP/6-31+G(d,p) results.
85. **Table S4.63**. Binding score energy (kcal $mol^{-1}$) of the euglobal molecules considered docked against molecular targets.
86. **Table S4.64.** Physicochemical properties (Mol MW, QPlogPo/w, Rule of Five) and ADMET (QPPCaco QPPMDCK, QPlogBB, QPlogKhsa, Percent Human Oral Absorption) properties of 38 euglobal molecules calculated using QikProp.
87. **Table S5.1.** Relative energies of the calculated conformers of considered simple ACPLs *in vacuo*.
88. **Table S5.2.** Relative energy corrected for ZPE (sum of electronic and zero-point energies, $\Delta E_{corrected}$, kcal $mol^{-1}$), ZPE correction to the electronic energy ($ZPE_{corr}$, kcal $mol^{-1}$), relative Gibbs free energies (sum of electronic and thermal free energy, $\Delta G_{corrected}$) and its thermal correction ($G_{corr}$), for the conformers of considered simple ACPLs *in vacuo*.



© University of Venda

89. **Table S5.3.** Relative energy corrected for ZPE (sum of electronic and zero-point energies, $\Delta E_{corrected}$, kcal $mol^{-1}$), ZPE correction to the electronic energy ($ZPE_{corr}$, kcal $mol^{-1}$), relative Gibbs free energies (sum of electronic and thermal free energy, $\Delta G_{corrected}$) and its thermal correction ($G_{corr}$), for the conformers of considered simple ACPLs *in vacuo*.
90. **Table S5.4.** Parameters of the O–H···O intramolecular hydrogen bonds in the calculated conformers of considered simple ACPLs *in vacuo*.
91. **Table S5.5.** Vibrational frequencies (harmonic approximation, stretching mode) of the O–H bonds of the calculated conformers of the considered ACPL molecules.
92. **Table S5.6.** Vibrational frequencies (harmonic approximation, stretching mode) of the O–H bonds of the calculated conformers of the considered ACPL molecules.
93. **Table S5.7.** Red-shifts in the calculated vibrational frequencies (harmonic approximation) of the O–H bonds that act as IHB donors in the calculated conformers of the considered ACPL molecules.
94. **Table S5.8.** Red-shifts in the calculated vibrational frequencies (harmonic approximation) of the O–H bonds that act as IHB donors in the calculated conformers of the considered ACPL molecules.
95. **Table S5.9.** HOMO-LUMO energy gap of the calculated conformers of the considered ACPL molecules.
96. **Table S5.10.** Dipole moments of the calculated conformers of the considered ACPL molecules.
97. **Table S5.11.** Relative energies of the calculated conformers of the considered ACPL molecules *in vacuo*, in chloroform, in acetonitrile and in water (respectively denoted as vac, chlrf, actn and aq in the column headings).
98. **Table S5.12.** Relative energies of the calculated conformers of the considered ACPL molecules *in vacuo*, in chloroform, in acetonitrile and in water (respectively denoted as vac, chlrf, actn and aq in the column headings).
99. **Table S5.13.** Solvation free energy ($\Delta G_{solv}$) and its electrostatic ($G_{el}$) and dispersion components for the calculated conformers of the considered ACPL molecules in chloroform, acetonitrile and water (respectively denoted as chlrf, actn, aq in the column headings).
100. **Table S5.14.** Solvation free energy ($\Delta G_{solv}$) and its electrostatic ($G_{el}$) and dispersion components for the calculated conformers of the considered ACPL molecules in chloroform, acetonitrile and water (respectively denoted as chlrf, actn, aq in the column headings).
101. **Table S5.15.** Parameters of the intramolecular hydrogen bonds in the calculated conformers of the considered ACPL molecules in chloroform, acetonitrile and water (respectively denoted as chlrf, actn, and aq in the column headings).
102. **Table S5.16.** Parameters of the intramolecular hydrogen bonds in the calculated conformers of the considered ACPL molecules in chloroform, acetonitrile and water (respectively denoted as chlrf, actn, and aq in the column headings).
103. **Table S5.17.** Length of the IHB (Å) in the calculated conformers of the considered ACPL molecules *in vacuo* and in chloroform, acetonitrile and water (respectively denoted as vac, chlrf, actn, aq in the column headings).



© University of Venda

© University of Venda

University of Venda
Creating Future Leaders

# Table of contents





© University of Venda

University of Venda
Creating Future Leaders





© University of Venda

University of Venda
Creating Future Leaders





© University of Venda

University of Venda
Creating Future Leaders





© University of Venda

© University of Venda

# Conference presentations

N. Tshilande, L. Mammino. XIX RMFQT, Mexican Conference of Theoretical Physical Chemistry (2021), 18-19 November: "Comparación teórica de las propiedades moleculares de floroglucinoles del tipo euglobales" (Poster presentation)

N. Tshilande, L. Mammino. Sasol Virtual Postgraduate Seminar (2022), 2-3 November: "Computational study of selected antimalarial and anti-cancer acylphloroglucinols" (Oral presentation)

N. Tshilande, L. Mammino. 16th CHPC National Conference (2022), 30 November-2 December: "High-performance computing in the computational study of selected antimalarial and anti-cancer acylphloroglucinols" (Micro-talk presentation)

N. Tshilande, L. Mammino. 44th SACI National Convention (2023), 8-13 January: "Computational study of selected anti-cancer acylphloroglucinols *in vacuo* and in solution" (Poster presentation)

N. Tshilande, L. Mammino. SAYAS Annual Young Scientist Conference (2023), 12- 13 June: ‘’Computational study of selected antimalarial and anticancer acylphloroglucinols’’ (Poster Presentation)

N. Tshilande, L. Mammino. SACI NORTH SECTION Young Chemist Symposium, (2023) 23 October: ‘’ Computational Study of Selected Antimalarial and Anticancer Acylphloroglucinols (Oral Presentation, second prize award)



© University of Venda

# Publications

Tshilande, N., & Mammino, L. (2023). Comparison of the Molecular Properties of Euglobals Differing by the Mutual Positions of the Two R–C= O Groups (R= H and $CH_2CH(CH_3)_2$): A Computational Study. Chemistry, 5(4), 2120-2154.

Tshilande, N., & Mammino, L. (2024). Acylphloroglucinols with anticancer and/or antimalarial activities—An overview. Studies in Natural Products Chemistry, 81, 39-77.

Tshilande, N., Mammino, L., & Bilonda, M. K. (2024). The Study of Molecules and Processes in Solution: An Overview of Questions, Approaches and Applications. Computation, 12(4), 78.

Tshilande, N., & Mammino, L. (2025). Computable properties of selected monomeric acylphloroglucinols with anticancer and/or antimalarial activities and first-approximation docking study. Journal of Molecular Modeling, 31(4), 1-27.



© University of Venda

# CHAPTER 1

## 1.0. Background information

Cancer and malaria are major causes of death globally, affecting millions of people each year. Cancer (as described in section 2.1) is a complex disease caused by abnormal changes in the body's cells, leading to uncontrolled growth. It can occur in various tissues and, in some cases, is linked to viral infections. Malaria (as described in section 2.3), caused by *Plasmodium* parasites and transmitted through mosquito bites, continues to be a significant health concern, particularly in tropical regions. Both diseases require better treatments due to their high mortality rates and the limitations of current therapies. Many existing drugs can have severe side effects, and not all are effective in addressing different forms of these illnesses. Another main issue in addressing these diseases is the development of resistance to current treatments. For malaria, resistance arises as the parasite evolves to withstand medications aimed at eradicating it. In cancer, resistance stems from the inherent adaptability of cancer cells. These cells can undergo genetic mutations and utilize multiple mechanisms to survive treatments, including mutating target genes, expelling drugs through efflux pumps, activating alternative survival pathways, and enhancing DNA repair capabilities.

Given these challenges, there is a pressing need for new drugs that can combat drug resistance while causing fewer side effects. Interestingly, some drugs developed for one disease have shown activity against the other. For example, certain antimalarial compounds have demonstrated potential in cancer treatment. Developing treatments that can act on multiple targets and provide broader therapeutic benefits could enhance patient outcomes. Exploring these connections between malaria and cancer treatments offers an opportunity to address critical gaps and improve the management of both diseases.

This thesis focuses on acylphloroglucinols (ACPLs, Fig. 1.1), a group of natural compounds derived from specific plants that have shown potential against both cancer and malaria in previous studies. Their structural properties suggest they could help combat drug resistance, although much about their mechanisms remains to be explored.



© University of Venda

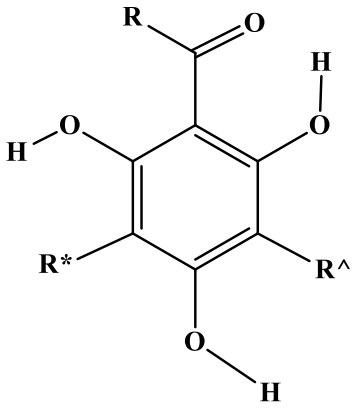



**Figure 1.1**. General structure of acylphloroglucinol.

Advanced computer modelling techniques are used in this research to gain insights into the structural characteristics of acylphloroglucinols and their potential interactions with cells affected by malaria and cancer. This computational approach enables researchers to identify promising candidates for drug development more efficiently, minimizing the need for extensive laboratory testing.

This thesis presents important insights that could lead to more effective treatments for malaria and cancer. Through a better understanding of these natural compounds, one can focus on developing therapies that can effectively combat these diseases while ensuring their long-term efficacy. This study is part of a broader computational investigation of ACPLs at the University of Venda.



© University of Venda

### 1.1. Objectives of the study

The current study has the following objectives:

- Identifying the stable conformers of the selected molecules, considering all the possible geometries obtainable by rotations about relevant single bonds.
- Identifying the conformers' stabilising factors.
- Determining the other computable properties of the selected molecules, such as dipole moments, highest-occupied molecular orbital (HOMO) - lowest-unoccupied molecular orbital (LUMO) energy difference and vibrational frequencies (harmonic approximation).
- Investigating the effect of different solvents with different polarities (to model different biological environments) on the conformational preferences and the other properties of the selected molecules.
- Performing structure-based virtual screening on the selected molecules
    - to structurally evaluate the interactions between the considered ACPLs and multiple anticancer receptors and antimalarial drug targets

### 1.2. Organisation of the material

The present work contains seven chapters:

Chapter 1 provides the background information for this study such as the importance of anti-cancer and antimalarial drugs, the importance of the search for new active drugs with fewer side effects than the current ones, as well as the objectives of this study.

Chapter 2 provides a literature review on malaria and cancer, information on acylphloroglucinols, which is the class of compounds to which the considered molecules belong, and information related to the selected molecules themselves.

Chapter 3 provides the theoretical background of computational approaches to the study of molecules, paying particular attention to the computational methods utilised in the current study and to the molecular properties considered here.

Chapter 4 to 6 report the results obtained and analyse them.

- Chapter 4 presents the result of the computational study of selected EGs *in-vacuo*, in-solution, and docking.



© University of Venda

- Chapter 5 presents the result of the computational study of selected simple ACPLs (with the acylphloroglucinol moiety containing only one acyl group) *in-vacuo*, in-solution, and docking.
- Chapter 6 presents the result of the docking studies of selected dimeric and trimeric ACPLs.

Chapter 7 provides an overall discussion of the results obtained.



© University of Venda

# CHAPTER 2
# LITERATURE REVIEW

## 2.0. Introduction to the chapter

This chapter comprehensively reviews malaria and cancer and provides information on acylphloroglucinols. Sections 2.1-2.2 extensively discuss cancer, its symptoms, available treatments, and anticancer drugs. Section 2.3 overviews malaria, discussing its symptoms, available treatments, and drugs used to combat it. Section 2.4 reviews the issue of drug resistance, a growing concern in the fight against malaria and cancer. Section 2.5 provides information on natural acylphloroglucinols and their potential as antimalarial and anticancer agents, as well as brief information on synthetic acylphloroglucinols. This chapter aims to thoroughly understand these diseases, their treatments, and the promising role of natural compounds in medicine.

## 2.1. Cancer

This section gives an overview of the available information about cancer, its symptoms, treatments, and anticancer drugs.

### 2.1.1. Cancer disease

Out-of-control growth of abnormal cells in a specific organ characterizes over a hundred distinct diseases that constitute cancer [1]. When some cells multiply uncontrollably, it leads to cancer, which grows in a specific body part. Healthy cells undergo copy processes when nearby cells signal them to do so [2]. This acts as a self-regulatory mode, ensuring that cells in each tissue retain the appropriate shape and size. Cancer cells, on the other hand, ignore these signals and instead grow and divide uncontrollably, creating tumors that can invade and damage nearby tissues and organs [2]. Their unique feature allows them to spread from where they first appeared, populating neighboring tissues and creating masses in other areas throughout the body [2]. The growth of these malignant tumors can become fatally aggressive and ultimately destroy vital organs and tissues.

Cancer is caused by a buildup of mutations in certain gene classes that control cell growth and division [3, 4]. Understanding the underlying mechanisms that give rise to human cancer requires



© University of Venda

studying these genes. Genes are located in chromosomes in the cell's nucleus, and they direct the production of specific proteins by dictating the order in which amino acids must be linked. A gene prompts the cell to produce the protein it encodes when activated. Alterations in the gene that codes for the protein can cause changes in the amount of protein produced or its activities, leading to cancer development [2].

Cancer development involves three crucial gene types: proto-oncogenes, tumour suppressor genes, and micro ribonuicleic acid (RNA) genes [2, 5, 6]. While proto-oncogenes typically encourage cell growth and division, they can become oncogenes and create excessive cell growth due to mutations. Tumor suppressor genes regulate cell growth as they prevent uncontrolled cell growth. However, cancer development can arise when tumors form due to mutations in these genes. Lastly, microRNA genes play their part by regulating the expression of other genes, and its improper regulation can lead to cancer development [2, 6].

### 2.1.2. Proto-oncogenes

Proto-oncogenes are a class of genes that, when activated, promote normal cell growth and division [2, 6]. However, when these genes change through amplification, chromosomal rearrangement, or mutation, they can become oncogenes that cause uncontrolled cell growth and division, leading to cancer development [6]. Oncogenic mutations in proto-oncogenes can result in either the overproduction of a growth-stimulating protein or the production of a hyperactive form of the protein [2]. Examples of growth factors or receptor oncogenes include platelet-derived growth factor (PDGF), epidermal growth factor receptor (erb-B or EGFR), epidermal growth factor receptor 2 (HER2 or erb-B2), receptor tyrosine kinase (RTK), and many others. PDGF is associated with gliomas, a type of brain cancer [7]. Breast cancer and glioblastoma, another type of brain cancer, are linked to the erb-B gene that codes for the epidermal growth factor receptor. HER-2, or erb-B2, is associated with ovarian, breast, and salivary gland cancers [2]. The RTK gene codes for a growth factor receptor linked to thyroid cancer.

Another class of oncogenic genes is those that facilitate cytoplasmic relays and are involved in stimulatory signaling pathways. For example, mutations in the ki-ras gene are associated with lung, ovarian, colon, and pancreatic cancers, while N-ras is involved in leukemias [2]. Transcription



© University of Venda

factors that activate genes promoting growth, such as c-myc, N-myc, and L-myc, are also implicated in cancer development. C-myc is involved in the development of leukemias, as well as breast, stomach, and lung cancers [8]. N-myc is associated with neuroblastoma, a type of cancer affecting nerve cells, and glioblastoma [9]. L-myc, on the other hand, is linked to lung cancer [10].

Other types of oncogenic genes code for different types of molecules. For instance, the Bcl-2 gene codes for a protein that ordinarily prevents cells from undergoing apoptosis (programmed cell death). Mutations in this gene are associated with follicular B-cell lymphoma. The Bcl-1 gene is a component of the cell cycle clock (i.e., the regulatory mechanisms that control the progression of a cell through several stages of the cell cycle) that stimulates cyclin D1 and is associated with cancers of the breast, head, and neck. The murine double minute 2 (MDM2) gene codes for an antagonist of the p53 tumor suppressor protein [8], which plays an important role in preventing the formation of cancer cells. Mutations in this gene can result in the inhibition of p53 and the development of cancer [2, 6].

### 2.1.3. Tumor suppressor genes

Tumor suppressor genes play an essential role in preventing cancer by suppressing cell growth and division. Mutations that deactivate these genes can lead to cancer development by removing the brakes that limit cell proliferation. Some of these genes code for proteins found in the cytoplasm, while others code for nuclear proteins. Examples of cytoplasmic tumor suppressor genes include adenomatous polyposis coli gene (APC), deleted in pancreatic cancer-4 (DPC4), neurofibromatosis type 1 (NF-1), and neurofibromatosis type 2 (NF-2). Mutations in APC have been linked to colon and stomach cancer [11]. DPC4 encodes a relay molecule involved in a signaling pathway that regulates cell division and is associated with pancreatic cancer [12]. The NF-1 gene encodes a protein that suppresses the activity of the Ras protein, which is involved in malignancies of the peripheral nervous system. NF-2, on the other hand, is involved with brain cancer, specifically meningioma, ependymoma, and schwannoma [13].

Genes that code for nuclear proteins (i.e., different groups of proteins primarily localized within the cell nucleus, the membrane-bound organelle that houses the cell's genetic material) include multiple tumor suppressor 1 (MTS1), retinoblastoma protein (RB), tumor protein 53 (p53), and



© University of Venda

Wilms' tumor-1 protein (WT1). MTS1 codes for the p16 protein, which slows down the cell cycle clock and is linked to several types of cancer [14]. RB encodes the protein pRB, which acts as a master brake on the cell cycle and is associated with retinoblastoma, bone, bladder, small cell lung, and breast cancer [15]. The p53 gene encodes the p53 protein, which can inhibit cell division and induce abnormal cells to undergo apoptosis and is involved in many different types of cancer [16]. WT1 is associated with Wilms' kidney tumors [17].

Breast cancer gene 1 (BRCA1), breast cancer gene 2 (BRCA2), and von Hippel-Lindau (VHL) are important tumor suppressor genes whose exact location in cells is unknown. Mutations in the BRCA1 gene have been linked to both breast and ovarian cancer, while mutations in BRCA2 are associated with breast cancer [18]. VHL, on the other hand, is associated with renal cell carcinoma [19]. Identifying these genes and their functions is important to developing effective cancer treatments and prevention strategies.

### 2.1.4. Micro ribonuicleic acid genes

Micro ribonucleic acid (microRNA) genes are a unique class of genes that produce short RNA molecules, usually between 21 and 23 nucleotides in length. These RNA molecules play a critical role in controlling gene expression, and unlike proto-oncogenes and tumor suppressor genes, they do not encode proteins [2]. Instead, microRNA genes function by regulating the activity of other genes. Interestingly, microRNA genes are often located in genomic regions frequently involved in chromosomal rearrangements in cancer cells, yet these regions lack conventional oncogenes and tumor suppressor genes. This lack of evidence suggests that microRNA genes may be a key player in the development and progression of various types of cancer.

The activity of microRNA genes depends on their specific targets, which can vary depending on the tissue or cell type. In some cases, microRNA genes can act as tumor suppressors by inhibiting the expression of oncogenes. In contrast, in other cases, they can act as oncogenes by inhibiting the expression of tumor suppressor genes [2]. For example, overexpression of miR-155 has been associated with acute lymphoblastic leukemia and high-grade lymphoma [2], whereas downregulation of miR15a and miR-16-1 leads to the overexpression of B-cell lymphoma 2 (BCL2) [2]. This protein prevents apoptosis and is linked to chronic lymphocytic leukemia.



© University of Venda

Additionally, miR191 is overexpressed in various solid tumors and is believed to be a downstream target of cancer-related signaling pathways [2].

## 2.2. Anticancer drugs

### 2.2.1. Overview of cancer treatments and therapeutic strategies

Cancer is a complex disease that can affect any part of the body, and its treatment typically involves surgery, radiation, and anticancer drugs [20]. Previously, drugs used in cancer treatment were classified as chemotherapy, hormone therapy, or immunotherapy [20]. Chemotherapy is still one of the best ways to treat cancer, and scientists are always looking for new chemotherapy drugs and new ways to use the ones they already have [21]. Chemotherapy drugs are small molecules designed to kill or slow the growth of cancer cells [22], but the challenge lies in designing drugs that can destroy cancer cells while minimizing harm to healthy tissue [23].

In addition to traditional chemotherapy, targeted therapy has emerged as a promising treatment method [24]. Targeted therapy exploits the dependence of cancer cells on specific oncogenes for their continued growth and survival. Targeted treatment drugs are developed to kill cancer cells while sparing healthy tissue [25], ranging from alkylators and antibiotics to antimetabolites, topoisomerase, and mitotic inhibitors. On the other hand, hormone therapy uses drugs such as steroids, anti-estrogens, anti-androgens, luteinizing hormone-releasing hormone (LH-RH) analogs, and anti-aromatase drugs to prevent the growth of certain hormone-dependent tumors. Immunotherapy uses interferon, interleukin-2, and vaccinations to activate the immune system and help it find and kill cancer cells.

Drugs may also target other components of carcinogenesis, such as the endothelium, extracellular matrix (ECM), immune system, and bone. Depending on the target, anticancer drugs can be categorized as deoxyribonucleic acid (DNA)-targeting, RNA-targeting, or protein-targeting drugs [22]. Figure 2.1 shows the targets where several drugs can disrupt or inhibit part of the biochemical process. Despite the range of available treatments, cancer remains a complex and challenging disease, and ongoing research aims to improve our understanding of its mechanisms and develop more effective therapies.



© University of Venda

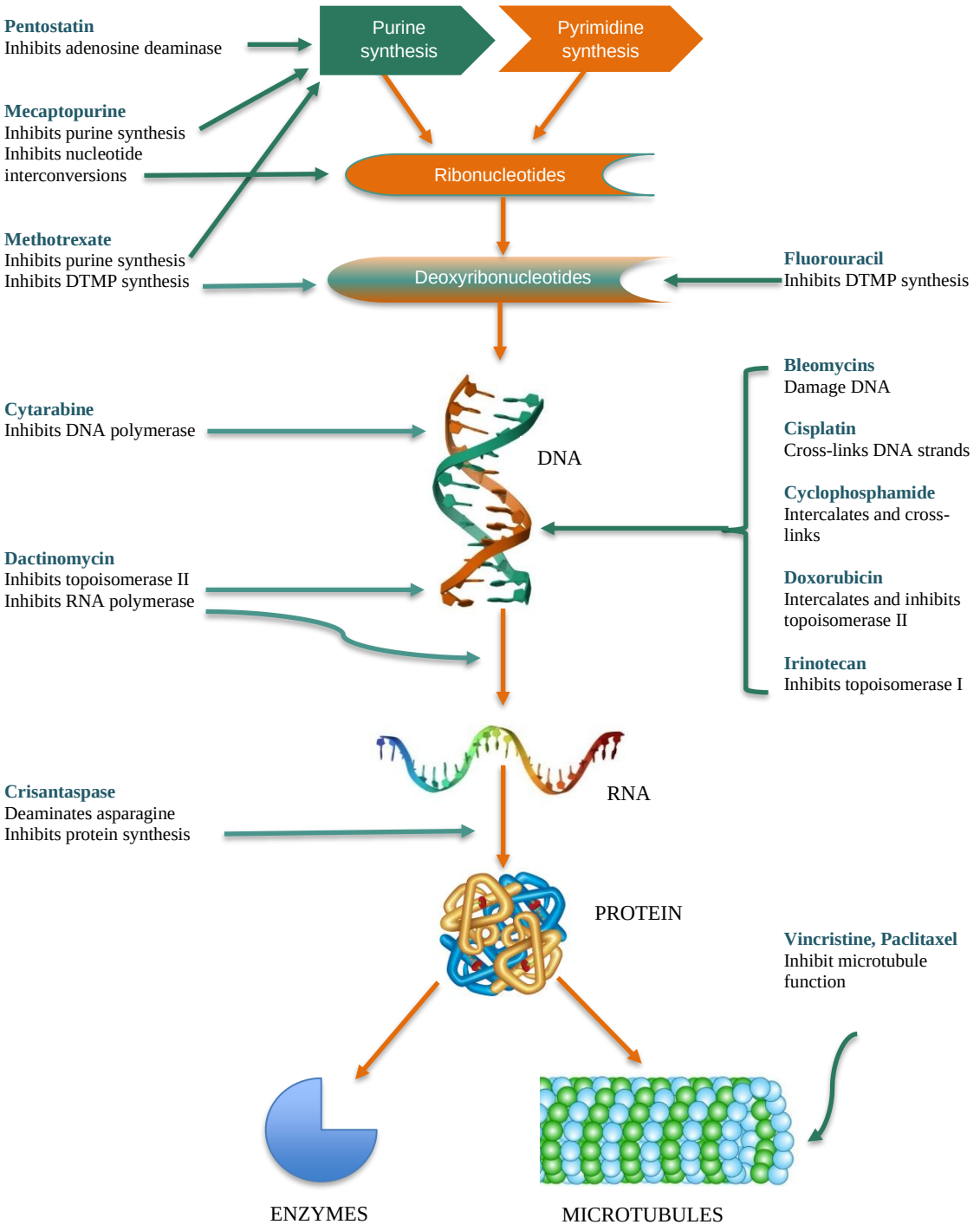



**Figure 2.1. The key stages in the biochemical pathway of the action of anticancer drugs.** Names of these compounds of interest are reported on the left- and right-hand sides**;** the central side shows the targeted stages.



© University of Venda

### 2.2.2. Drugs directed against tumour DNA

Traditional anticancer agents work by damaging DNA directly or indirectly. Direct damage to the DNA can lead to lesions or breaks in the double helix. In contrast, indirect damage can occur through interference with DNA-related proteins or the alteration of gene expression. Most traditional anticancer agents use these modes of action, including chemotherapy drugs and radiation therapy. These treatments aim to target rapidly dividing cancer cells and induce cell death while minimizing damage to healthy cells in the body.

#### 2.2.2.1. Drugs targeting the DNA helix

Alkylating agents (Fig. 2.2) are the first class of drugs effective against cancer and have made significant progress. These compounds form adducts when they attach to DNA strands, causing structural or functional changes. In addition to alkylating agents, intercalating agents have a similar planar polyaromatic system that binds to DNA by inserting itself between base pairs, with a preference for the 5'-pyrimidine-purine-3' positions [26]. Several alkylator classes exist, including nitrogen mustards, nitrosoureas, triazenes, platinum compounds, and antibiotics [20, 26, 27]. Dacarbazine, a member of the triazene class, is used to treat metastatic melanoma and soft tissue sarcomas alone or in combination with other medications. However, myelosuppression, severe nausea, and vomiting are common side effects [28]. Another triazene agent, temozolomide, is a relatively new drug used as a second-line therapy for various types of brain cancer. It shares a similar structure with dacarbazine, but its oral bioavailability and distribution properties, including its ability to enter the central nervous system, make it an attractive option [28].



© University of Venda

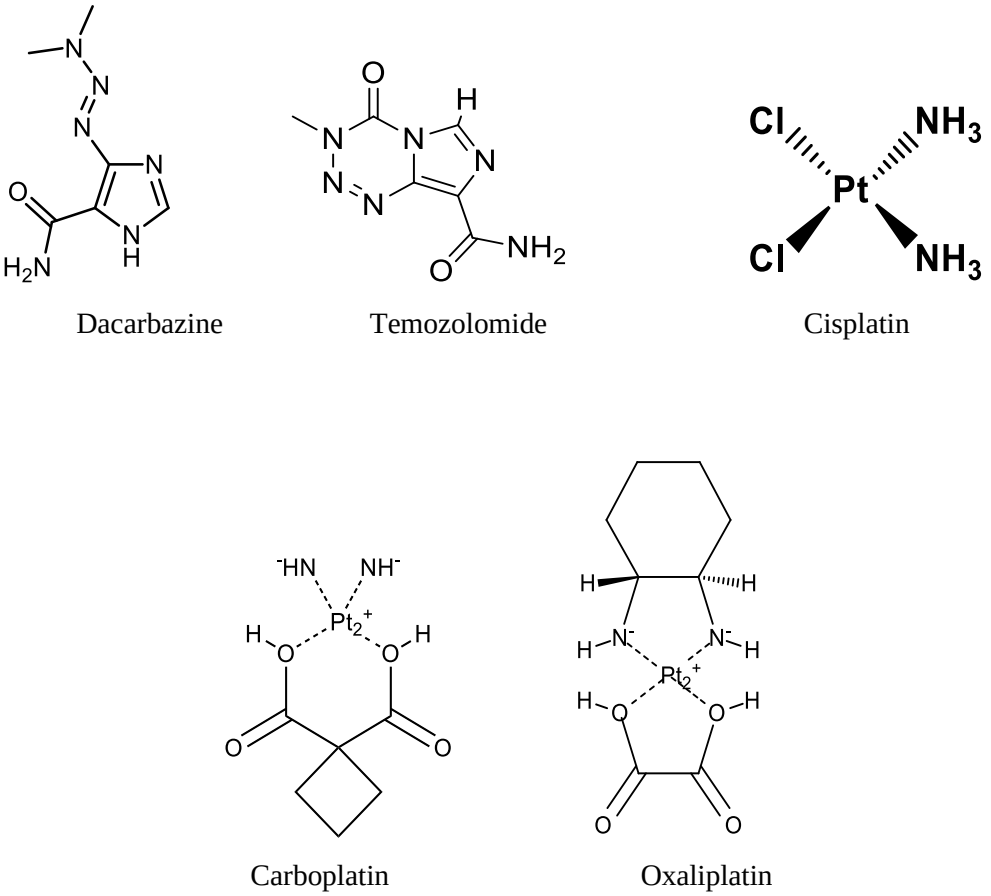



**Figure 2.2. Examples of molecular structures of alkylating agents.**

Platinum complexes, the most frequently used therapeutic class for cancer treatment, have significant activity against many cancers. Cisplatin, the first platinum drug with significant activity against ovarian and testicular malignancies, was followed by carboplatin and oxaliplatin, which were developed to reduce the adverse effects of cisplatin [28]. Unlike alkylating and intercalating agents, platinum compounds form DNA adducts that cause the DNA to bend, leading to cell death. Additionally, platinum compounds can bind to non-DNA targets such as proteins, contributing to their effectiveness. Despite their success, platinum compounds have drawbacks, such as neurotoxicity, nephrotoxicity, and myelosuppression [23].



© University of Venda

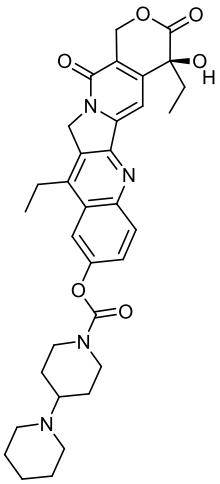
### 2.2.2.2. Drugs targeting DNA-related proteins

Drugs that target protein-DNA complexes are an essential class of chemotherapy drugs that play a critical role in cancer treatment. These drugs include topoisomerase I and II inhibitors, antimetabolites, and ecteinascidin. One of the key advantages of these drugs is that they do not bind to DNA directly but instead target the proteins that interact with DNA. This means they can potentially have a more targeted effect on cancer cells while minimizing damage to healthy cells [20].

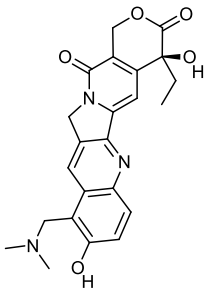
Topoisomerases are important enzymes for DNA metabolism, playing a fundamental role in virtually every stage of the cell division process [29]. There are two types of topoisomerases: topoisomerase I and topoisomerase II. Topoisomerase I temporarily breaks one DNA strand during replication to relieve torsional stress, while topoisomerase II can break both strands and pass one through the other to disentangle them. Irinotecan and topotecan are two derivatives of camptothecin (Fig. 2.3). Camptothecins are considered interfacial inhibitors because they bind at the interface between topoisomerase I and DNA, trapping the macromolecular complex and preventing DNA replication. Several drugs can inhibit the reannealing function of topoisomerase II, including doxorubicin and mitoxantrone [1, 26, 29], by intercalating into DNA and preventing the proper binding of DNA-binding proteins. These drugs can lead to DNA strand breakage and are widely used as anticancer agents. Doxorubicin is effective against various types of cancer, including acute leukemia, lymphomas, and many types of solid tumors. However, it can cause side effects such as nausea, vomiting, myelosuppression, mucositis, alopecia, and cardiotoxicity due to cumulative dosing [28]. Mitoxantrone treats advanced forms of breast cancer, non-lymphoma Hodgkin's in adults, and adult non-lymphocytic leukemia. It shares similar side effects with doxorubicin, and it has less prominent cardiac toxicity [28].



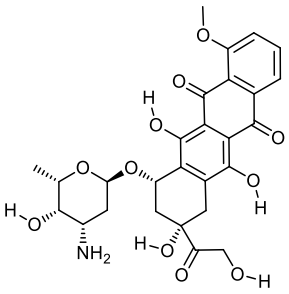
© University of Venda



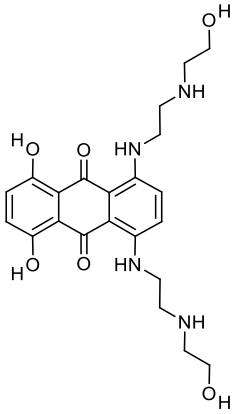
Irinotecan



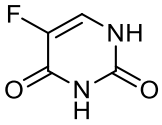
Topotecan



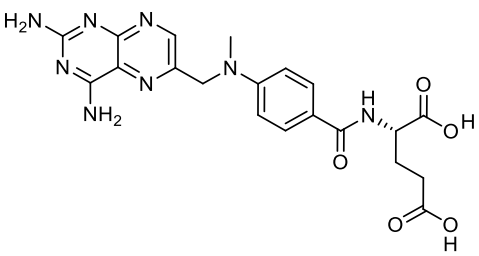
Doxorubicin



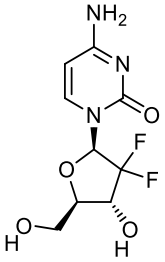
Mitoxantrone



5-Fluorouracil



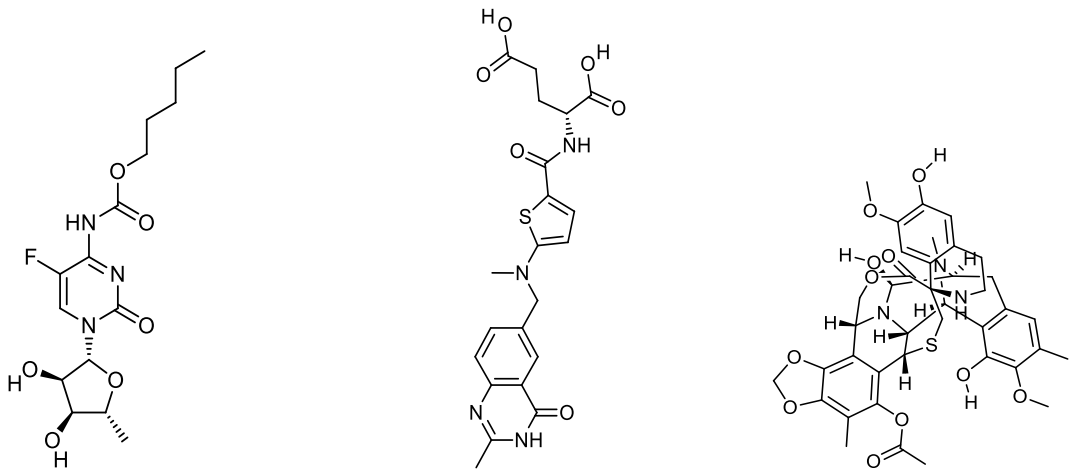
Methotrexate



Gemcitabine



© University of Venda



Capecitabine Raltitrexed Ecteinascidin-743

**Figure 2.3. Examples of molecular structures of topoisomerase inhibitors and antimetabolites.**

Antimetabolites are a class of drugs that disrupt nucleotide metabolism by interfering with the synthesis of DNA and RNA [27, 30]. They were one of the first types of targeted cancer therapies developed, as they specifically target the nucleic acid synthesis processes necessary for cancer cell growth. Antimetabolites are structural analogs of naturally occurring compounds required for DNA and RNA synthesis. They work by competing with these compounds for binding sites on enzymes involved in nucleotide synthesis. Once cells take them up, antimetabolites are incorporated into DNA or RNA, causing errors in replication and transcription [27]. Hence, this mode of action leads to abnormal function and, ultimately, cell death. Some antimetabolites are cell cycle-specific, meaning they only work during a specific cell cycle phase.

In contrast, others are non-specific and can work at any cell cycle phase. Several antimetabolites exist, including antifolates, fluoropyrimidines, raltitrexed, cytarabine, and gemcitabine [27, 30, 31]. Some antimetabolites, such as 5-fluorouracil (5-FU) and methotrexate, have been used for decades to treat many cancers [31]. Other, more recently developed drugs, such as gemcitabine, capecitabine, and raltitrexed, have shown promising results in clinical trials [27]. Gemcitabine has shown promise in treating several types of cancer, including pancreatic, non-small-cell lung, bladder, breast, and ovarian cancers. It disrupts DNA synthesis in cancer cells, preventing their ability to divide and grow. Gemcitabine can be administered intravenously or subcutaneously, and it is typically given in cycles over several weeks.



© University of Venda

Capecitabine is a prodrug that is converted to 5-FU in the body. 5-FU is another antimetabolite used to treat various cancers, including colorectal, breast, and head and neck cancers. Capecitabine is particularly attractive because it increases 5-FU's tumor selectivity while extending 5-FU's time in those tumors [31]. It is typically given orally and is well-tolerated by most patients. Raltitrexed is a more potent thymidylate synthase inhibitor than older drugs like 5-FU [31]. Thymidylate synthase is an enzyme required for DNA synthesis and inhibiting it can prevent cancer cells from dividing and growing. Raltitrexed is currently used to treat colorectal cancer and is effective in combination with other chemotherapy drugs.

One of the advantages of antimetabolites is their specificity for cancer cells. Because they target the metabolic pathways essential for cancer cell growth, they are often more effective against cancer cells than normal cells. However, they can also affect rapidly dividing normal cells in the body, such as those in the bone marrow and gastrointestinal tract. This may cause undesirable side effects, such as nausea, vomiting, and hair loss. However, these side effects are usually manageable with supportive care and dose adjustments. Ecteinascidin-743 (ET-743, is a potent antitumor drug derived from the Caribbean tunicate Ecteinascidia turbinate [26]. It is a DNA-binding agent that has shown promising results in treating various cancer types, including sarcomas, melanoma, breast, ovarian, colon, renal, non-small cell lung, and prostate cancer [20]. ET-743 exhibits its antitumor activity through several mechanisms, including inhibition of DNA repair, induction of apoptosis, and inhibition of transcriptional factors like TCF-1 and Sp1 [20].

While initially thought to function as an alkylating agent, recent studies have suggested that ET-743 interferes with RNA polymerase II-mediated gene transcription [32]. Specifically, it appears to bind to the minor groove of DNA, which can lead to the formation of DNA adducts and ultimately inhibit RNA polymerase II. This inhibitory effect on RNA polymerase II is independent of p53 status, making it a potentially effective treatment option for p53-mutant tumors. ET-743 has been approved for use in Europe for treating advanced soft tissue sarcoma and is currently being evaluated in clinical trials for other cancer types. While ET-743 has shown promising results in early trials, it does have some side effects, including nausea, vomiting, fatigue, and



© University of Venda

myelosuppression. However, these side effects are generally mild and can be managed with appropriate supportive care.

#### 2.2.2.3. Drugs acting on specific genes

Hormonal agents are a class of drugs that can have potent effects on the body's endocrine system. These drugs work by altering the levels of certain hormones in the body, which can, in turn, impact the growth and proliferation of cancer cells. The mechanism of action of hormonal agents is to bind to specific receptors on the surface of cells, which then triggers a series of downstream signaling events that alter gene expression and cellular function. Hormonal agents, such as glucocorticoids, steroids, antihormones, and retinoids, are non-cytotoxic drugs with anticancer effects in specific populations [33]. These drugs modify the expression of genes, a typical mode of action shared by steroid, antihormone, and retinoid agents.

When hormonal agents bind to their corresponding receptor proteins in the cytoplasm or nucleus, they form a hormone-receptor complex. This hormone-receptor complex can activate specific regulatory sequences in DNA, influencing gene expression [20]. For instance, when glucocorticoids or other steroid hormones bind to their receptors, they initiate regulatory activity within DNA, thereby modifying gene expression. Receptors for estrogens and androgens are nuclear-localized transcription factor subtypes controlled by their respective ligands. These receptors are inhibited by antioestrogens and antiandrogens, respectively. Antiestrogens, antiandrogens, and antiprogestins are not cytotoxic but can alter the endocrine environment, potentially inhibiting the growth of hormone-dependent cancers [33].

Gene therapy is a relatively new approach to cancer treatment that involves the delivery of therapeutic genes into cancer cells to modify their genetic makeup. The goal is to correct or replace the defective or missing genes that contribute to cancer development or induce cell death [20]. Unlike hormonal agents, gene therapy targets specific genes or pathways and can treat various types of cancer. Different types of gene therapy include gene replacement therapy, gene editing, and gene silencing. Gene replacement therapy involves introducing a functional copy of a defective gene into cells to replace the non-functional one. Gene editing, on the other hand,



© University of Venda

involves modifying the DNA sequence of a gene to correct mutations or introduce new functions. Gene silencing aims to turn off the expression of a specific gene by introducing molecules that block the expression of the gene [20].

Despite its potential benefits, gene therapy is still a relatively new and experimental approach to cancer treatment, and many challenges still need to be addressed. One of the main challenges is delivering therapeutic genes to the target cells efficiently and safely. Gene therapy can also have off-target effects, and there is a risk of inducing mutations or causing unintended side effects. As such, gene therapy is currently used only in clinical trials and is not yet a standard cancer treatment.

### 2.2.3. Drugs acting on tumor RNA

In oncology, researchers and clinicians constantly seek new ways to target cancer cells with greater precision and efficacy. One approach that has gained attention in recent years is using drugs that target specific RNA molecules. While several anticancer drugs, such as fluoropyrimidines and platinum compounds, have been found to inhibit RNA production, their primary mechanism of action is through binding to DNA [20]. However, a class of drugs called antisense oligonucleotides takes a different approach by targeting specific messenger ribonucleic acid (mRNA) molecules.

Antisense oligonucleotides are designed to bind to and block the translation of specific mRNA molecules often overexpressed in cancer cells, such as bcl-2, myb, p53, mdm2, HER2, and methyltransferase-1[20]. By inhibiting the translation of these mRNA molecules into functional proteins, the target gene's function can be effectively blocked, offering a potential therapeutic benefit in cancers driven by the overexpression of these oncogenes. One such example of an antisense oligonucleotide drug is Angiozyme [20]. This drug goes after the mRNA that makes vascular endothelial growth factor (VEGF) [34], a protein that helps new blood vessels grow. By inhibiting the production of VEGF, Angiozyme can reduce the blood supply to tumors, which can slow their growth or even cause them to shrink.

In addition to their potential as anticancer agents, antisense oligonucleotides have shown promise in treating other diseases, such as genetic disorders, viral infections, and cardiovascular disease



© University of Venda

[35]. As research continues, new RNA-targeting drugs will likely be developed that offer even greater specificity and efficacy in treating cancer and other diseases.

### 2.2.4. Drugs acting on proteins in the tumor cell

Drugs directed against proteins in the tumor cell can target different types of molecules, such as receptors in the tumor membrane, intracellular pathways in tumor cells, and tubulin.

#### 2.2.4.1. Receptors in the tumor membrane

The tumor membrane receptors are targeted through two main approaches: monoclonal antibodies and small molecule inhibitors. Unlike traditional chemotherapy drugs, these targeted therapies have a cytostatic rather than a cytotoxic action, meaning they do not directly kill cancer cells but rather slow down or stop their growth [20].

Monoclonal antibodies have emerged as a promising class of drugs for cancer treatment, offering targeted and specific action against cancer cells. They can inhibit their growth and survival by binding specific proteins on the surface of cancer cells. The primary mode of action of monoclonal antibodies is to inhibit the extracellular region of the receptor, preventing cancer cells from receiving the signals essential for their growth and proliferation. Initially, monoclonal antibodies were developed to target lymphoid antigens such as CD20 and CD52. Rituximab, ibritumomab-I, and tositumomab-I are antibodies that target CD20, while alemtuzumab targets CD52. Recent studies have shown that combining these antibodies with isotopes can enhance their effectiveness [36-38], and they are currently being evaluated as frontline treatments for patients with refractory lymphomas. On the other hand, anti-carcinoma antibodies, such as trastuzumab and cetuximab, are the most used antibodies against carcinomas [20]. Trastuzumab can be used alone or in combination with chemotherapy to treat HER2-positive breast tumors, and its efficacy is being further investigated for other potential uses.

Small molecules are a class of targeted therapies that interact with proteins in cancer cells' cytoplasm or on their membrane's surface. Unlike monoclonal antibodies, they are usually administered orally and can penetrate the cell membrane to reach their targets. These molecules



© University of Venda

inhibit the intracellular domain of an enzyme, typically tyrosine-kinase. Although some chemotherapy drugs are also small molecules, the term is used in this context to distinguish them from monoclonal antibodies. Small molecules usually aim for the epidermal growth factor receptor (EGFR) family. Drugs like gefitinib (Fig. 2.4) target the epidermal growth factor receptor (Her-1) in particular [39]. Several types of cancer, such as non-small cell lung cancer (NSCLC) and cancers of the head and neck, have cells that overexpress this receptor [40]. By inhibiting the activity of Her-1, gefitinib can block the signaling pathways that promote cancer cell growth and survival [20]. Studies have shown that gefitinib can be effective as a monotherapy for these types of cancers [41, 42].

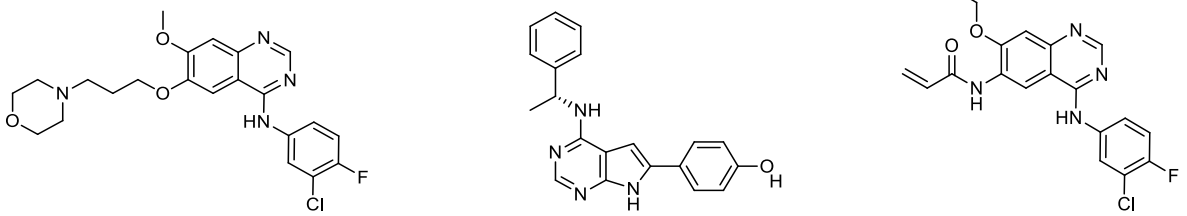



**Figure 2.4. Examples of molecular structures of small molecule inhibitors.**

PKI-166 (Fig. 2.4) is another small molecule that can inhibit the activity of Her-1 and HER2 [20], another member of the EGFR family. Inhibiting the activity of both receptors can provide additional benefits in treating certain types of cancer, such as breast cancer. The drug CI-1033 (Fig. 2.4) can irreversibly inhibit all epidermal growth factor receptors [20]. This means that once the drug binds to the receptor, it can not be easily removed, and the receptor is permanently



© University of Venda

inactivated. This mechanism of action makes CI-1033 particularly effective against cancer cells that have developed resistance to other EGFR inhibitors.

### 2.2.4.2. Intracellular pathways in tumor cells

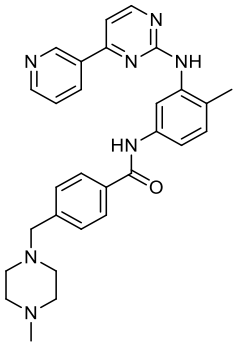
The proliferation of tumor cells is controlled by various signaling pathways activating different intracellular proteins and enzymes. These pathways are activated by growth factors that bind to specific receptors on the cell's surface [43], leading to a cascade of intracellular events that ultimately transport proliferation signals to the nucleus.

The protein kinase pathways are one important group of intracellular pathways involved in tumor cell proliferation [44]. These pathways are activated by the phosphorylation of proteins by kinases, and they play a role in controlling cell development and division, among other cellular activities. Abnormal activation of protein kinase pathways is a common feature of many types of cancer, making them attractive targets for anti-cancer therapies.

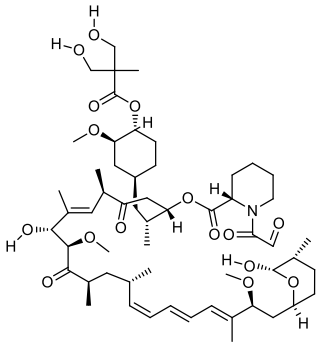
The tyrosine kinase pathway is one example of a protein kinase pathway that plays an important role in tumor cell proliferation [45]. Tyrosine kinases add phosphate groups to tyrosine residues on proteins, activating downstream signaling pathways. Imatinib (Fig. 2.5) is a drug that blocks the tyrosine kinase of bcr/abl and c-kit, which are involved in cell proliferation [20]. It is one of the best treatments for gastrointestinal stromal tumors and chronic myeloid leukemia.



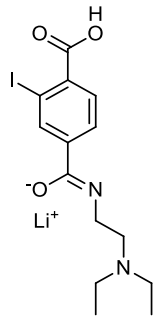
© University of Venda



Imatinib



CCI-779



PKC-412



© University of Venda

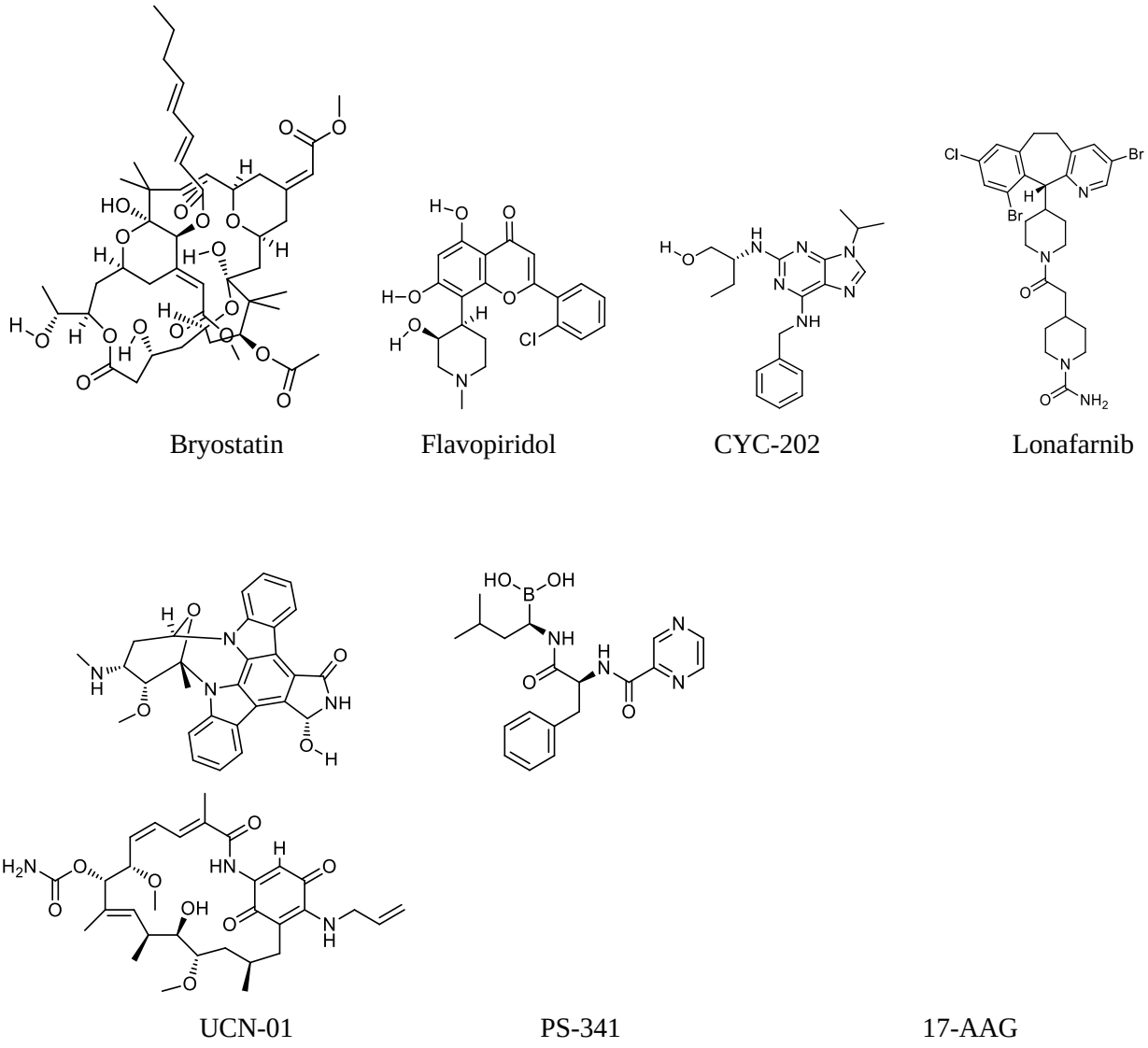



**Figure 2.5. Examples of molecular structures of intracellular pathways (in tumor cells) inhibitors.**

Another group of pathways involved in tumor cell proliferation is the phosphatidylinositol (PI) pathway [46]. The PI pathway is activated by the serine/threonine kinase PI-3K, which phosphorylates membrane lipids, activating downstream signaling pathways. The PI pathway involves many cellular processes, including cell proliferation, survival, and differentiation. PKB/Akt is a downstream effector of the PI pathway, and it connects this pathway to mechanistic target of rapamycin (mTOR), which controls apoptosis and cellular metabolism. Rapamycin



© University of Venda

derivatives like CCI-779 inhibit mTOR and are used as specific pharmaceuticals that work along this pathway.

Protein kinase C, an enzyme family that activates the transcription factor NF-jB, is linked to PI-3K. PKC-412 and bryostatin can inhibit protein-kinase C [47]. The cyclin-dependent kinase (CDK) pathway is another important pathway for cell cycle progression and is often dysregulated in cancer. Cyclin-dependent kinases (CDKs) are enzymes that regulate the activity of cyclins, which control the progression of the cell cycle. Flavopiridol, CYC-202, and UCN-01 (Fig. 2.5) can inhibit CDK activity and block cell cycle progression [48].

Ras is another intracellular pathway important to tumor cell proliferation. The ras protein is a small GTPase activated by the enzyme farnesyl transferase, which is responsible for adding a farnesyl group to the ras protein. Once activated, the ras protein stimulates downstream signaling pathways, including raf and MEK. Inhibitors of farnesyl transferase like lonafarnib and R115,777 [20], as well as inhibitors of raf (BAY 43-9006) and MEK (CI-1040) [20], can inhibit this pathway.

Finally, the proteasome pathway is another intracellular pathway often dysregulated in cancer. The proteasome is a protein-degrading enzyme complex that plays an important role in protein turnover and cellular homeostasis. PS-341 is a drug that blocks the proteasome, preventing protein degradation and leading to apoptosis [20], while geldanamycin compounds like 17-AAG promote heat shock protein 90 breakdown, which prevents protein breakdown.

#### 2.2.4.3. Drugs acting on tubulin

Tubulin is a protein that plays a critical role in the cytoskeleton structure of cells, which is responsible for maintaining cell shape, intracellular transport, and cell division [49]. It comprises two subunits, alpha, and beta-tubulin, which can be assembled to form tubulin dimers. These tubulin dimers can then polymerize to form long microtubules (MTs), essential cytoskeleton components [50].



© University of Venda

The vinca alkaloids affect tubulin by destabilizing MTs [51]. They bind to specific locations on tubulin dimers. This action interferes with the formation of the mitotic spindle during cell division, which is essential for adequately segregating chromosomes. As a result, vinca alkaloids are often used in cancer treatment to inhibit the growth of rapidly dividing cancer cells. Examples of vinca alkaloids include vinblastine, vincristine, and vinorelbine, as shown (Fig. 2.6) [52].



© University of Venda

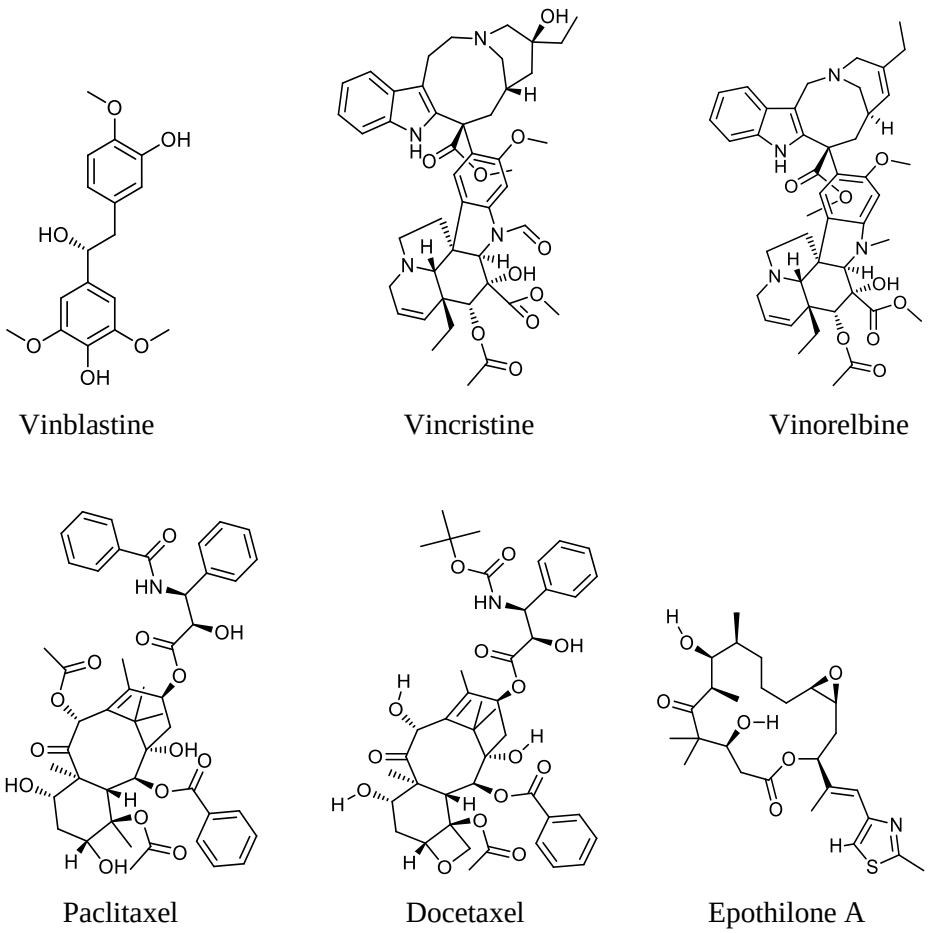



**Figure 2.6. Examples of molecular structures of tubulin inhibitors.**

Taxanes are another class of drugs that affect tubulin by stabilizing MTs [51]. Taxanes bind to a unique site on the beta subunit of tubulin, which prevents the microtubule network from reorganizing and results in cell cycle arrest. Taxanes treat various cancers, including breast, ovarian, and lung cancers. Examples of taxanes include paclitaxel and docetaxel (Fig. 2.6) [51].

Epothilones are a newer drug class that also affects tubulin by stabilizing microtubules [52, 53]. Epothilones bind to a different site on the beta subunit of tubulin than taxanes [54], resulting in a different microtubule stabilization pattern. Epothilones are being investigated for their potential use in cancer treatment, particularly in cases where cancer cells have become resistant to other tubulin-targeting drugs [55].



© University of Venda

### 2.2.5. Drugs acting on the endothelium and extracellular matrix (ECM)

An inner lining of cells that lines blood vessels (such as arteries, veins, and capillaries) is called the endothelium and plays an important role in vascular function. The ECM is a network of proteins and carbohydrates that surrounds cells and provides structural support. The endothelium and the ECM are potential targets for drugs to treat various diseases, including cardiovascular disease and cancer. Drugs targeting the endothelium inhibit endothelial growth factors or their receptors. However, metalloproteinases are typically targeted by drugs that act on the ECM [56]. Each of these drugs has antiangiogenic properties.

#### 2.2.5.1. Endothelium

The endothelium is very important for controlling blood flow, keeping the tone of the blood vessels, and controlling vascular permeability [57-59]. It also produces a variety of growth factors that are important in angiogenesis, the process by which new blood vessels are formed. Inhibition of endothelial growth factors, such as VEGF and bFGF, can be a potential target for cancer treatment. Thalidomide (Fig. 2.7), a drug initially developed as a sedative, has been shown to suppress the activity of VEGF and bFGF [60]. Carboxyamidotriazole is another drug that explicitly targets VEGF [61]. Interferon-gamma has also been shown to mediate the effect of interferon-alpha on VEGF production in tumor cells [62].



© University of Venda

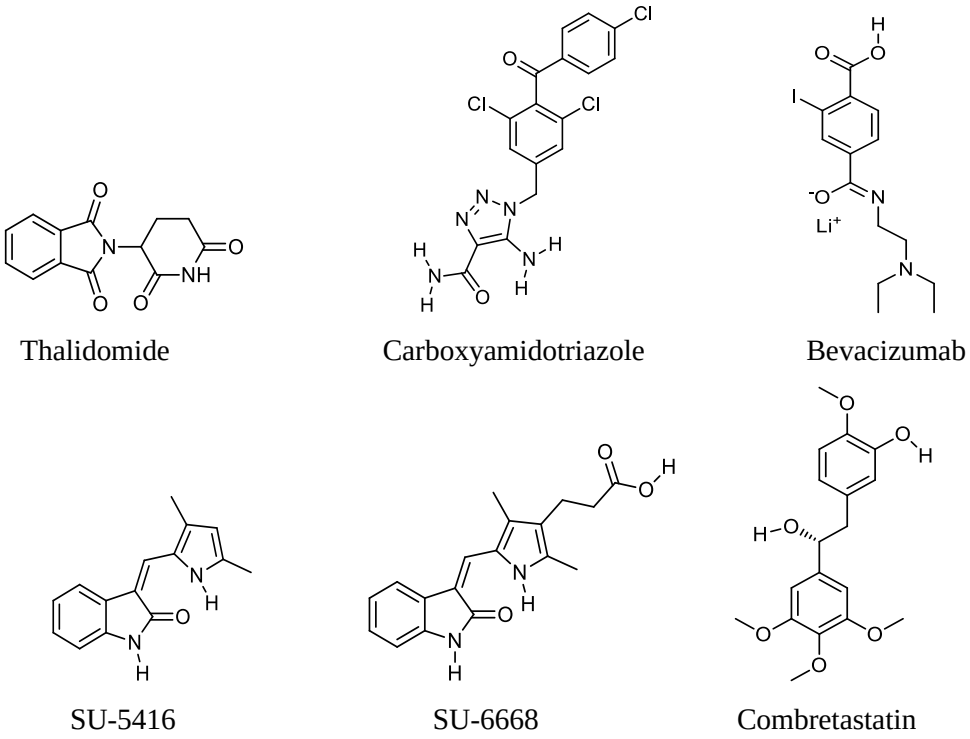



**Figure 2.7. Examples of molecular structures of endothelial growth factor (s) inhibitors.**

Cyclooxygenase 2 (COX-2) inhibitors, a class of drugs commonly used to treat pain and inflammation, have been found to prevent COX-2-mediated endothelial growth stimulation, suggesting their potential use as anti-cancer agents [63,64]. The inhibition of angiogenesis has been demonstrated using the monoclonal antibody bevacizumab, which binds to various VEGF receptors [65-67]. SU-5416 is a small molecule that binds to tyrosine kinase receptors, including vascular endothelial growth factor 1 (VEGFR-1) and vascular endothelial growth factor 2 (VEGFR-2) [65,68], and other receptors, such as c-kit and PDGF receptors. SU-6668 is another small molecule that binds to multiple growth factor receptors, including VEGFR, bFGFR, and PDGFR [69,70].

Combretastatin is a drug that triggers apoptosis by inhibiting the endothelial mitotic spindle, disrupting the formation of new blood vessels, and leading to tumor regression [71, 72].

### 2.2.5.2. Drugs acting on extracellular matrix

The extracellular matrix (ECM) is a sophisticated network of proteins and carbohydrates that surrounds cells and provides structural support to tissues. It plays an important role in cell



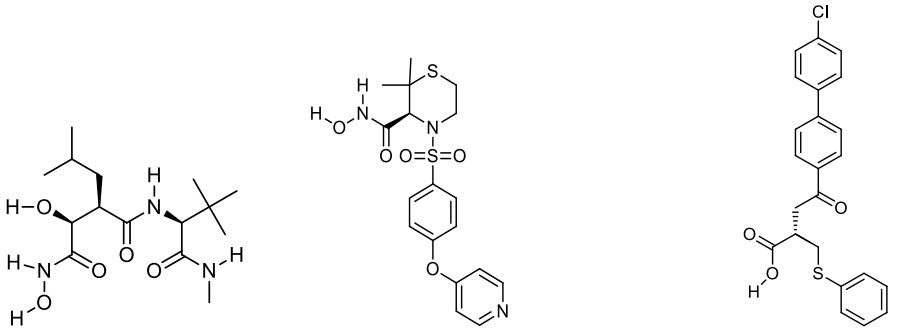
© University of Venda

adhesion, migration, proliferation, differentiation, and survival. In cancer, the ECM is remodeled, leading to increased cell migration, invasion, and angiogenesis, contributing to tumor growth and metastasis. Matrix metalloproteinases (MMPs) are a class of proteolytic enzymes that are important in remodeling the ECM [73]. They are involved in the degradation of various ECM components, including collagen, laminin, and fibronectin [74]. MMPs are upregulated in various cancers and are associated with poor prognosis [75,76]. Activating MMPs in tumors promotes invasion and is essential to develop new blood vessels. Some research suggests that MMPs can also trigger the production of bFGF, VEGF, and insulin growth factors. Several MMP inhibitors are in the testing phase of their development [77]. Enzyme activity inhibitors like marimastat [78-80], prinomastat, and BAY 12-9566 are often synthetic [77]. Neovastat, like other tetracycline derivatives, reduces MMP synthesis, blocks activation, and promotes the breakdown of MMPs [81, 82]. In addition to MMP, additional components of the extracellular matrix, such as integrins, endothelins, and thrombospondins, could be targeted as an anticancer therapy [83, 84].

Several MMP inhibitors are being developed for cancer therapy [77]. These inhibitors target the catalytic domain of MMPs and prevent their activity. Synthetic inhibitors (Fig. 2.8), such as marimastat, prinomastat, and BAY 12-9566 [20], have reduced tumor growth and angiogenesis in preclinical studies. However, their clinical development has been hindered by their poor specificity, toxicity, and lack of efficacy.



© University of Venda



Marimastat Prinomastat BAY 12-9566

**Figure 2.8. Examples of molecular structures of matrix metalloproteinase inhibitors.**

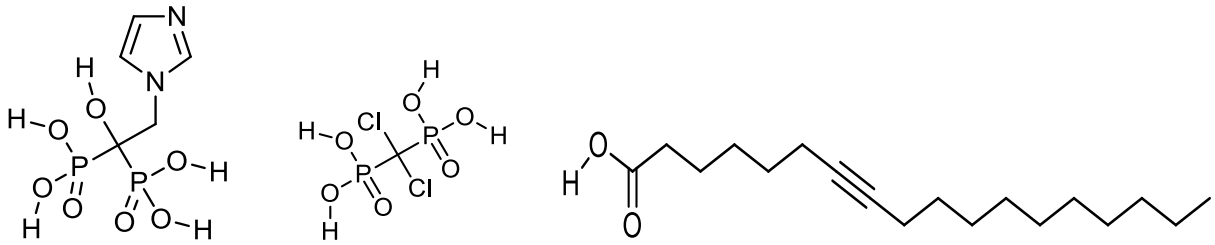
Neovastat, a natural compound derived from shark cartilage, is a promising MMP inhibitor. It reduces MMP synthesis, blocks activation, and promotes the breakdown of MMPs [20]. It has been shown to work in preclinical studies and is now being tested for different cancers in clinical trials. In addition to MMPs, other components of the ECM, such as integrins, endothelins, and thrombospondins, have been identified as potential targets for anticancer therapy [20]. Integrins are transmembrane receptors that mediate cell-ECM interactions and play an important role in tumor progression. Endothelins are vasoactive peptides that promote angiogenesis and tumor growth. Thrombospondins are ECM glycoproteins that have antiangiogenic properties and can inhibit tumor growth. Drugs targeting these components of the ECM are currently being developed and tested in preclinical and clinical studies.

### 2.2.6. Host cell inhibitors and other drugs

In addition to targeting cancer cells, some drugs target the host cells or the tumor's microenvironment. Bisphosphonates such as zoledronic acid and clodronate (Fig. 2.9), for example, are drugs used to treat bone disorders and can also effectively prevent or treat bone metastases in cancer patients [85, 86]. Osteoprotegerin is another drug that targets bone cells and inhibits the formation and activity of osteoclasts, which can prevent bone loss and reduce the risk of skeletal-related events in cancer patients [87].



© University of Venda



Zoledronic Acid Clodronate Osteoprotegerin

**Figure 2.9. Examples of molecular structures of parathyroid hormone-related protein and osteoclasts inhibitors.**

Anti-PTHrP antibodies are designed to inhibit the action of parathyroid hormone-related protein (PTHrP) produced by many cancer cells [88]. They can cause bone resorption and calcium release into the bloodstream, leading to hypercalcemia (elevated blood calcium levels). By blocking PTHrP, these antibodies can prevent or treat hypercalcemia in cancer patients. Interferon and interleukin-2 are cytokines that prompt the body's immune system to defend against cancer cells. The immune system naturally produces a protein called interferon, while interleukin-2 is a synthetic version of a naturally occurring cytokine. These drugs can enhance the immune system's ability to recognize and attack cancer cells [89]. However, their use is limited by significant side effects, and their effectiveness in treating cancer has been modest in clinical trials.

## 2.3. Malaria

This section comprehensively discusses malaria, including its symptoms and available treatment options. Additionally, it discusses various drugs that are effective in combating malaria.

### 2.3.1. Malaria disease

Malaria is a parasitic disease caused by the protozoan genus *Plasmodium* [90-92]. It can be fatal. Among the different Plasmodium species, only six can infect humans under certain conditions, namely *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale wallickeri*, *Plasmodium ovale curtisi*, *Plasmodium malariae*, and *Plasmodium knowlesi*. These species differ in morphology, immunology, geography, behavior, relapse patterns, and drug responses [90]. Malaria is believed to have originated in sub-Saharan Africa and was later spread to other parts of the world through human migration and trade. The disease has been present in human populations



© University of Venda

for thousands of years, and there is evidence of malaria in ancient Egyptian mummies. The ancient Greeks and Romans also described a disease believed to have been malaria. Malaria was also prevalent in parts of Europe during the 19th century until its final eradication through improved sanitation and insecticides. Today, malaria is found in many tropical and subtropical regions worldwide, including sub-Saharan Africa, Southeast Asia, South America, and parts of the Middle East and Oceania. In many regions of the world, malaria is still a significant health concern, particularly in sub-Saharan Africa. *P. falciparum* is the most lethal species responsible for malaria-related deaths, particularly in Africa [90-92]. Also, infected erythrocytes from *P. falciparum* can block small blood vessels in the brain. This can cause cerebral malaria, which is usually fatal, especially in African children [90].

### 2.3.2. Malaria life cycle

Malaria parasites have a complex life cycle that involves both invertebrate and vertebrate hosts, and the expression of highly specialized proteins is essential for each stage of this cycle. The parasite's ability to survive intracellularly and extracellularly, invade host cells, and evade the immune system depends on these proteins. The malaria parasite's life cycle is complex, both in the mosquitoes that serve as vectors and in the vertebrate hosts that become infected [90, 93]. The life cycle of malaria parasites involves multiple stages, beginning with the introduction of sporozoites into the bloodstream, as shown in Figure 2.10. The carriers for these parasites are female *Anopheles* mosquitoes, which deposit them in human hosts during blood meals. The gametocytes, which are male and female forms of the parasites, are transmitted from the mosquito to the human host during this feeding. Upon reaching the liver, the gametocytes infect hepatocytes and replicate asexually, producing schizonts. After a certain number of days, the liver schizonts rupture into the bloodstream and invade red blood cells. The parasite then replicates mitotically, progressing through stages to produce merozoites [90]. This is when the clinical manifestations of the disease begin to appear.

Due to many rounds of asexual reproduction within the human host, the parasite population can increase from 10 to $10^2$ at the time of infection to $10^8$ to $10^{13}$ within a few weeks. The transmission of 10–$10^3$ malaria gametocytes back to the mosquito host occurs when another female mosquito feeds on the blood of an infected human [93]. The immature gametes develop, fuse to create



© University of Venda

zygotes, and proceed through sexual recombination and meiosis to form haploid cells. To complete the life cycle, haploid cells replicate asexually, forming sporozoites that travel to the mosquito's salivary glands [93]. The Plasmodium life cycle restarts each time an infected mosquito bites a person who is vulnerable to the disease.

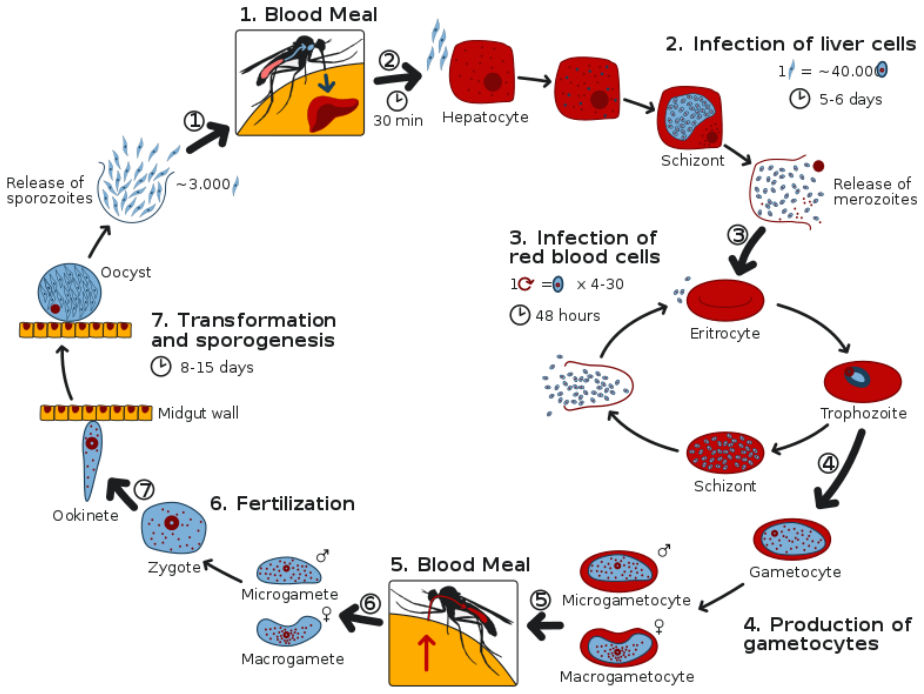



**Figure 2.10: Life Cycle of the Malaria Parasite [94].**

### 2.3.3. Symptoms, cases, and fatalities of malaria

When the parasite enters the human body, it replicates and spreads throughout the bloodstream, eventually invading red blood cells and causing clinical symptoms. The clinical manifestation of malaria in humans results from the complex interaction between the parasite's preprogrammed biology and the human host's pathophysiological response. Parasite infection sets off a sequence of immunological reactions, the nature and severity of which are highly variable between individuals and dependent on age, immune condition, and genetic background [95]. Some individuals may develop a robust immune response to control the parasite and prevent severe



© University of Venda

disease. In contrast, some people may have a weak immune system that lets the parasite grow and spread, causing more severe clinical signs.

The symptoms of malaria can appear within six days to many months after being bitten by an infected mosquito in areas where the disease is common. Some of the most common symptoms of malaria parasite infection include high body temperature, fever, cough, breathing difficulties, joint pain, headache, watery diarrhea, vomiting, and seizures [96, 97]. Not all people infected with malaria will show symptoms, and some may not have any symptoms at all [90]. While malaria can be a serious and potentially life-threatening disease, most infections do not lead to severe complications. However, certain factors can increase the risk of severe malaria, such as being pregnant, having a weakened immune system, or not having acquired immunity to the disease [90, 98]. According to the World Health Organization (WHO), uncomplicated malaria is fever without other clinical or laboratory indicators indicating severe disease or organ failure [99]. This means that a person with uncomplicated malaria is experiencing fever as their primary symptom, without any other signs or symptoms of severe illness [100]. Most cases of uncomplicated malaria can be treated effectively with antimalarial drugs that target the parasite responsible for the infection. The duration of treatment can vary depending on the specific drug used, but in general, most people can expect to recover within a few days of starting treatment.

A severe form of complicated malaria can lead to serious and potentially life-threatening complications [101, 102]. Cerebral malaria is a severe complication that can occur when the parasite infects the brain. It can lead to seizures, confusion, coma, and, in some cases, permanent neurological damage. Malaria can be considered another potential cause of acute lung injury (ALI). The lungs become severely inflamed, leading to a buildup of fluid in the air sacs, making it difficult to breathe. In severe cases, ALI can progress to acute respiratory distress syndrome (ARDS), a life-threatening condition requiring urgent medical attention.

Severe anemia develops when parasites kill more red blood cells (RBCs) than the host's body can replace, leading to a shortage of oxygen. This can cause weakness, fatigue, and other complications. Acute kidney injury can occur when the parasite infects the kidneys, leading to kidney failure. This can result in a buildup of waste products in the body and fluid imbalances.



© University of Venda

Shock is a severe condition that can occur when the body's organs do not receive enough blood flow, leading to low blood pressure and organ failure. It is a life-threatening complication of severe malaria [103]. Complicated malaria is more common in children under five, pregnant women, and those with weakened immune systems.

Malaria is a significant health risk for pregnant women and young children, particularly where the disease is endemic. Pregnant women are particularly vulnerable to the severe effects of malaria, as infection with the *Plasmodium falciparum* parasite can lead to the accumulation of infected red blood cells in the placenta [104], causing a condition known as placental malaria. This can lead to adverse outcomes, including low birth weight, preterm delivery, and increased infant mortality [90].

Women who live in malaria-endemic regions and have been pregnant multiple times may develop some immunity to the disease over time. This is because exposure to the parasite during previous pregnancies can stimulate the production of antibodies that can provide some protection against subsequent infections. However, this immunity is incomplete, and pregnant women are still at risk of severe malaria, particularly during their first pregnancy. It is, therefore, essential to ensure that pregnant women in malaria-endemic regions receive appropriate preventive measures, such as intermittent preventive treatment with antimalarial drugs and the use of insecticide-treated bed nets, to reduce their risk of infection and severe disease.

Malaria remains a major health concern in many parts of the world, with sub-Saharan Africa being the most affected region. In 2019, there were an estimated 229 million cases of malaria worldwide, with an estimated 409,000 deaths [105]. Furthermore, in 2019, children under 5 accounted for 67% of all malaria deaths worldwide [105]. Most cases (94%) and deaths (96%) occurred in the African region. As of 2021, the WHO estimated approximately 241 million malaria cases worldwide, with 627,000 deaths.

In recent years, the number of reported malaria cases has remained relatively constant, with a small increase in 2019 compared to 2018. While the global malaria mortality rate has decreased by 44%



© University of Venda

since 2000 [105, 106], progress has stalled in recent years, with the number of deaths remaining relatively constant since 2016.

Efforts to control and eliminate malaria include the following:

- strategies such as vector control (using insecticide-treated bed nets and indoor residual spraying)
- effective and rapid malaria diagnosis and treatment
- preventative measures for pregnant women.

However, progress in these efforts has been hindered by various factors, including inadequate funding, insecticide resistance, and disruptions to malaria control efforts due to the COVID-19 pandemic [105,107].

The COVID-19 pandemic has significantly impacted the fight against malaria. Many malaria prevention and control programs have been disrupted due to the diversion of resources toward COVID-19. Many people have been unable to access healthcare services due to lockdowns and restrictions on movement.

### 2.3.4. Treatment-antimalarial drug

Malaria can be treated if patients receive timely and appropriate treatment. For decades, natural products, semi-synthetic compounds, and synthetic compounds have been used to treat the disease [108]. Quinoline derivatives, antifolates, naphthoquinones, and artemisinin derivatives are the main antimalarial drugs currently available (Fig. 2.11.). Each drug class has a unique mechanism of action against the malaria parasite and is used for different purposes, such as treatment, prevention, or both.

**Quinolines:** The first antimalarial medication to gain widespread use was quinine, an alkaloid extracted from the bark of the Andean Cinchona tree back in the 17th century [90]. Quinolines and their derivatives mainly target the blood stages of the parasite, while other antimalarial drugs focus on the hepatic stage. These medications bind to the parasite's digestive vacuole and prevent haem polymerization. As a result, the accumulation of haem from hemoglobin breakdown becomes lethal to the parasite, ultimately leading to death. Interestingly, quinine was discovered long before the causes of malaria were understood [90]. However, its effectiveness is limited by its toxicity



© University of Venda

and resistance to P. falciparum. Nowadays, quinine is typically administered in conjunction with another drug to mitigate treatment duration and side effects when treating severe malaria cases.

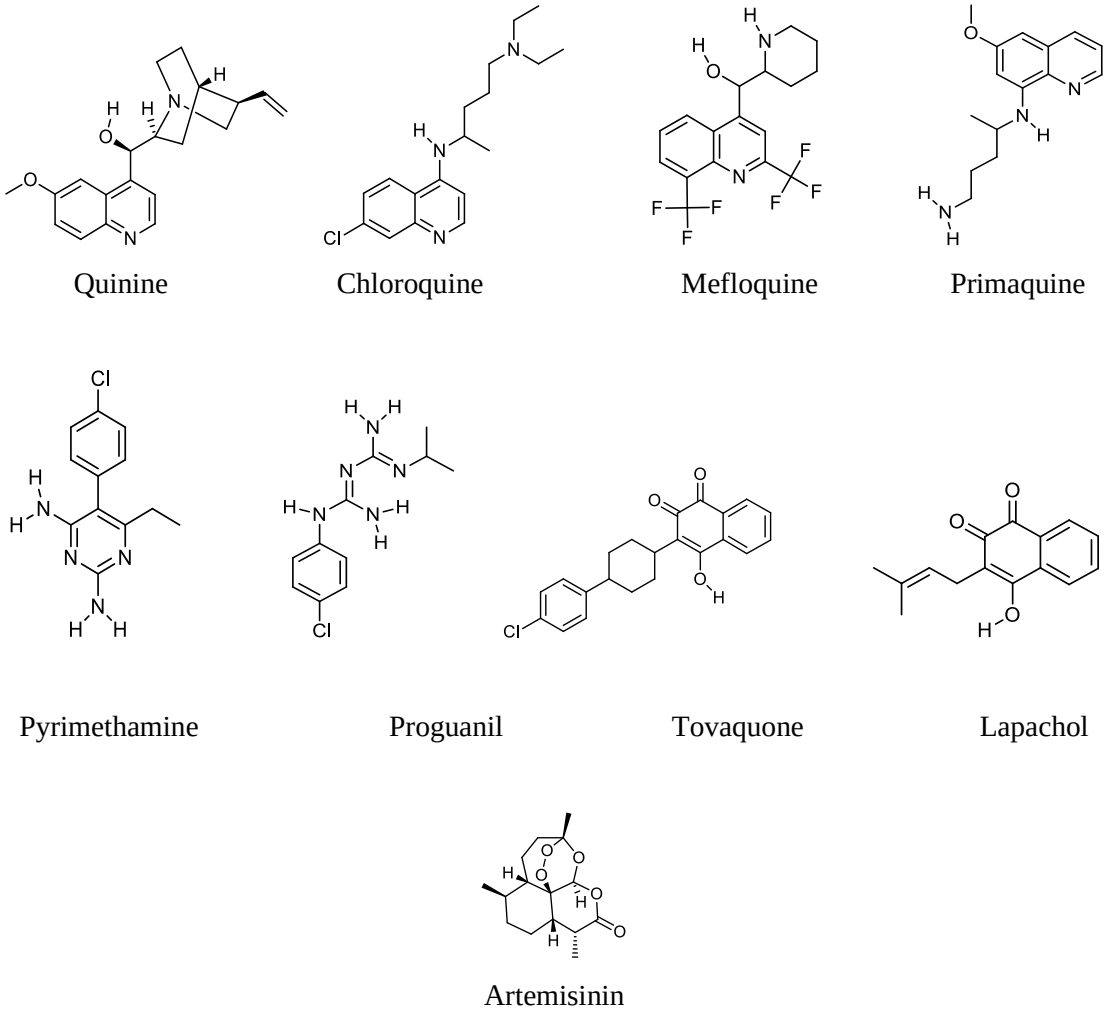



**Figure 2.11. Examples of molecular structures of antimalarial drugs.**

During World War II, the development of chloroquine was a significant step toward combating malaria. This medicine quickly became the preferred treatment option for malaria due to its high effectiveness, low risk of side effects, and affordable cost. However, the irrational use of chloroquine led to P. falciparum's resistance towards these widely used antimalarial medicines [109].



© University of Venda

**Antifolates:** These are a group of antimalarial drugs that work by inhibiting the production of folic acid [110], a vitamin essential for the parasite's growth and replication. Folic acid is required to produce nucleotides and amino acids, which are necessary building blocks for DNA synthesis and protein production. When antifolates inhibit folic acid production, they prevent the parasite from replicating and dividing, eventually leading to death. The molecular structures of antifolates are characterized by the presence of a pteridine ring, a p-aminobenzoic acid (PABA) moiety, and a glutamic acid residue. The pteridine ring is similar to the natural substrate of dihydrofolate reductase (DHFR), the enzyme targeted by antifolates [111]. The PABA moiety is responsible for the drug's selective toxicity by targeting the bacterial or parasitic enzymes involved in folic acid synthesis [112]. The glutamic acid residue helps increase the drug's water solubility.

Antifolates like pyrimethamine and proguanil primarily target the asexual stages of the Plasmodium parasite, specifically during the schizont stage. This is the stage where the parasite replicates inside erythrocytes and hepatocytes. By inhibiting the nuclear division of the parasite during this stage, antifolates can effectively kill the parasite. Pyrimethamine and proguanil target the enzyme DHFR [113, 114], which converts dihydrofolic acid to tetrahydrofolic acid, the body's active form of folic acid. By inhibiting DHFR, the drugs prevent the production of tetrahydrofolic acid, thereby blocking the synthesis of nucleotides and amino acids. This leads to the parasite's death. However, the parasite has developed resistance to antifolates like fansidar due to the overuse and misuse of these drugs. This has led to decreased effectiveness, and they were once less widely used than they are now [115]. Antifolate commercial drugs like fansidar were usually a combination of pyrimethamine and a sulphonamide.

**Naphthoquinones**: These are a class of antimalarial medications derived from natural sources, such as the bark of the cinchona tree, the leaves of the sweet wormwood plant, the roots of the madder plant, and the heartwood of the lapacho tree. Naphthoquinones are characterized by the presence of two fused rings: a naphthalene ring and a quinone ring. The naphthalene ring comprises two fused benzene rings, while the quinone ring is a cyclic-dione with two carbonyl groups. The electron-deficient quinone ring can interact with electron-rich biological molecules such as proteins and enzymes [116]. In malaria parasites, naphthoquinones can inhibit the electron transport chain of the parasite's mitochondria, leading to the depletion of adenosine triphosphate



© University of Venda

(ATP), the parasite's primary energy source. This inhibition can also lead to a buildup of toxic byproducts and ultimately kill the parasite. A lipophilic naphthalene ring in naphthoquinones enables them to easily cross the parasite's cell membrane and accumulate within the parasite, increasing their efficacy [117]. Some examples of naphthoquinones used to treat malaria include atovaquone and lapachol [118]. However, naphthoquinones are not typically used as standalone treatments for malaria; they are often used with other antimalarial drugs.

Atovaquone is an antimalarial drug that targets the *Plasmodium* mitochondria by inhibiting electron transport through the cytochrome bc1 complex [119]. It acts as a ubiquinone analog, preventing cytochrome b components from transferring electrons. This unique mechanism of action makes atovaquone effective against the parasite's sexual stages, preventing malaria from being transmitted from mosquitoes to humans. Combined with proguanil, atovaquone is safe and effective for children and pregnant women [120].

Lapachol has been shown to have antimalarial activity, but its mechanism of action still needs to be fully understood [121]. However, some studies suggest that lapachol may act by inhibiting the enzyme fumarate reductase in the mitochondria of the malaria parasite [122]. This enzyme is needed for the synthesis of ATP, which is the parasite's main energy source. By inhibiting fumarate reductase, lapachol may disrupt the energy production of the parasite, leading to its death. Other studies suggest that lapachol may also interfere with the heme detoxification pathway in the parasite, leading to an accumulation of toxic heme and, ultimately, the parasite's death. However, more research must be done to fully understand how lapachol acts as an antimalarial drug.

Therapeutic drug combinations are currently employed to combat antimalarial drug resistance and extend the efficacy of currently available treatments. For several years, azithromycin and chloroquine have been safe for pregnant women and children. In addition, clinical trials have shown that combining azithromycin and chloroquine is synergistic in treating symptomatic malaria and effective even against chloroquine-resistant strains of *P. falciparum* [123].

**Artemisinin:** Artemisinin is a sesquiterpene lactone and a natural product derived from the sweet wormwood plant (*Artemisia annua*). It is a highly effective antimalarial drug, particularly against



© University of Venda

the deadliest malaria parasite, *Plasmodium falciparum*. Studies show that artemisinin and its derivatives can inhibit the development and maturation of *Plasmodium* gametocytes. This action reduces the number of gametocytes, reducing the probability of transmitting from infected individuals to mosquitoes [124]. This mechanism of action can help to control the spread of malaria. Artemisinin gets activated by heme iron, which is released when the parasite digests hemoglobin in the erythrocytic stage. When the activated artemisinin is cleaved in the parasite's food vacuole, it produces free radicals and reactive oxygen species, which can damage and kill the parasite. Free radicals and reactive oxygen species can interfere with the parasite's calcium ATPase and proteasome [125], which are essential for maintaining the parasite's intracellular environment and protein turnover. As a result of this, the parasite is unable to live and reproduce, which ultimately leads to its death.

According to studies, artemisinin is effective against all stages of the malaria parasite's life cycle in red blood cells. The most significant effect is observed in the early stages of development. However, it is ineffective against the parasite's dormant stages, known as hypnozoites, which can persist in the liver for extended periods and cause disease relapses. To address this limitation, artemisinin is often combined with other medications that target the hypnozoite stage, such as primaquine. In addition, artemisinin is typically used in combination therapy with other antimalarial drugs, as there is a risk of the parasite developing resistance to it if it is used alone. The combination therapy helps to slow down the development of resistance and improve the efficacy of the treatment. Resistance to artemisinin has already appeared in Asia.

Furthermore, it stops the transfer of gametocytes from people to mosquitoes [124]. Cleavage of the artemisinin Endoperoxide-Bridge in the parasite's food vacuole generates free radicals that block the parasite's calcium ATPase and proteasome, leading to the drug's antimalarial action [125]. Artemisinin and its derivatives are effective against strains of malaria that have developed resistance to chloroquine and mefloquine. This has been observed. These drugs are often combined with other medications, such as fansidar and mefloquine. However, resistance to mefloquine is becoming more prevalent, especially in Southeast Asia, where it is still an essential and effective treatment. The emergence of malaria parasites resistant to current treatments highlights the need for more effective and less expensive treatments.



© University of Venda

### 2.3.5. The discovery and development of new antimalarial agents

Finding new drug targets and making new compounds that act on these targets is the focus of finding and developing new antimalarial drugs [126]. Historically, most drugs on the market have focused on targeting the blood stage of the disease, which is responsible for causing symptoms such as fever, chills, and anemia. However, new drugs are also being developed to target other stages of the disease. For instance, drugs targeting the liver stage of the disease are essential for preventing relapses. This is because some malaria parasites can enter the liver and remain inactive for months, only to reappear and produce symptoms months later [127]. Atovaquone and primaquine are two drugs that are effective against the liver stage of the disease and are often used in combination to prevent relapses [128].

It is also important to note that drugs that only target the blood stage of the disease can contribute to the development of drug-resistant strains of malaria. As a result, there is a growing focus on developing drugs that target multiple stages of the disease or have a different mechanism of action altogether. These drugs can be used in preventative medicine to reduce the likelihood of parasites developing resistance. Recent research has focused on developing new treatments for malaria that target the asymptomatic liver-stage sporozoites and hypnozoites without using 8-aminoquinolines or drugs that are susceptible to glucose-6-phosphate dehydrogenase (G6PD) [129]. This is because targeting this stage can help to combat the spread of the parasite. However, research into pharmaceuticals that can inhibit or eliminate the production of gametocytes has been limited due to a lack of quantitative high-throughput assays [130]. Developing such drugs could prevent transmission by effectively destroying mature gametocytes in the human host or preventing the growth of gametocytes into ookinetes and then sporozoites in the mosquito [131].

Chloroquine and sulfadoxine-pyrimethamine are antimalarial drugs that were commonly used in the past to treat malaria. However, their application has been restricted since the malaria parasite has developed the ability to resist the drugs. In addition to their antimalarial effects, it has been demonstrated that these medications stimulate gametocyte synthesis, the sexual form of the malaria parasite that can be transmitted from human to mosquito and back to human, thus perpetuating the infection cycle [132, 133]. The production of gametocytes is a concern because it can increase



© University of Venda

disease transmission, particularly in areas with high malaria transmission rates [132, 133]. This can occur even if the antimalarial drug effectively reduces the number of asexual parasites in the patient's bloodstream, as gametocytes can persist in the body for several weeks after the clearance of asexual parasites.

Furthermore, the use of antimalarial drugs that induce gametocytogenesis can also lead to the spread of drug-resistant strains of the malaria parasite, as the transmission of these strains is not affected by the drug's action on the parasite's asexual form. Therefore, developing alternative antimalarial drugs that do not induce gametocytogenesis and have a lower risk of promoting the spread of drug-resistant strains is a fundamental goal in malaria research. Primaquine is the only known effective gametocytocidal drug, but patients lacking G6PD develop hemolytic anemia after taking primaquine. The potential side effects of primaquine have prompted the search for alternative medications that can stop malaria transmission.

Drugs that target the liver and the transmission stages have the potential to be game-changing, but the lack of high-throughput screens has limited the research efforts [134, 135]. This issue is starting to be resolved by new emerging techniques, which also open new directions thanks to an innovative therapeutic drug with liver-stage activity.

*Plasmodium falciparum* has been extensively studied to identify new therapeutic targets for developing effective antimalarial drugs [136, 137]. Researchers have found that most novel targets are related to specific organelles within the parasite, such as the cytosol, apicoplast, mitochondria, food vacuole, transporter, and plasma membrane.

In the cytosol, drugs target pathways such as folate, glycolysis, shikimic acid, purine biosynthesis, thiol metabolism, and protein kinase [138]. In the apicoplast, drugs target the biosynthesis of fatty acids, isoprenoids, and heme, as well as plastic DNA replication and translation and protein farnesylation [138]. In the mitochondria, drugs inhibit dihydroorotase dehydrogenase and dihydrofolate reductase and target nicotinamide adenine dinucleotide phosphate (NADPH) dehydrogenase. In the food vacuole, drugs inhibit protease and interfere with the formation of hemozoin [138]. Drugs can also inhibit phosphatidylcholine synthesis within the transporter, and



© University of Venda

in the plasma membrane, drugs can inhibit membrane biosynthesis and target new permeability pathways from stopping the supply of nutrients.

### 2.4. Drug resistance and general mechanism

Drug resistance is a phenomenon in which microorganisms or cancer cells become less responsive or entirely unresponsive to the therapeutic effects of a drug, making it harder to treat the disease or infection [139]. It has become a significant problem that complicates the treatment of illnesses and increases the risk of patient mortality. Drug resistance can occur due to various molecular mechanisms, including changes in the target of the drug or alterations in drug metabolism. Drug resistance can be intrinsic or acquired [140], posing significant challenges to effective disease management, particularly multidrug-resistant strains. Multidrug-resistant strains are especially hard to treat because they need longer and more expensive chemotherapy regimens, which could make it harder to control the disease [141]. Understanding how microorganisms and cancer cells become resistant to currently available therapeutic drugs at the molecular level is important for better using presently available drugs and finding new targets for new classes of effective drugs. There are two main types of mechanisms through which drug resistance can occur: intrinsic drug resistance and acquired resistance.

#### 2.4.1. Intrinsic drug resistance

Intrinsic resistance is a type of drug resistance that occurs naturally in microorganisms due to inherent characteristics or features that have evolved over time [142]. Several mechanisms can contribute to the development of intrinsic resistance [139, 142] and are briefly discussed in the following sub-sections.

**Modification of the drug target:** Microorganisms' absorption of an antimicrobial drug is important for its target-oriented activity. Resistance to chemotherapy often arises due to modifications in the expression levels or mutations of the drug's target, which can render it ineffective [142, 143].

**Modification/ inactivation of therapeutic drugs:** Drug inactivation mechanisms are one of the effective protective strategies employed by microbes to reduce the free drug availability for



© University of Venda

binding to its intracellular target, leading to drug resistance [143]. Two distinct classes of enzymes are involved in the general mechanism of drug modification. The first class of enzymes promotes drug degradation by breaking down the drug into smaller, less active compounds. In contrast, the second class catalyzes chemical modifications that alter the drug's structure and reduce its efficacy [142]. The inactivation of therapeutic drugs via enzymatic modification or degradation is discussed in the following paragraphs.

**Enzymatic degradation:** Mycobacteria possess a range of enzymes that can specifically target and break down antibiotics belonging to different classes, such as β-lactams, chloramphenicol, aminoglycosides, or macrolides [142].

**Enzymatic modification:** Enzyme-mediated target modification is a mechanism that pathogens increasingly use to develop drug resistance.

**High detoxication ability:** Some mycobacteria, as well as other bacteria like Streptomyces spp (i.e., several species), can develop natural resistance mechanisms to protect themselves from potential predators or enemies. One of these mechanisms involves the production of harmful chemicals by the bacteria themselves

**Low drug delivery:** Low drug delivery into the target site can also occur due to efflux pumps and protein channels that pump out drugs from the bacterial or cancer cell. These efflux pumps have been identified in many cancer cells, and they work by expelling the drug out of the cell before it can bind to its target and exert its therapeutic effect. This can lead to suboptimal drug concentrations at the site of action, making the drug less effective and promoting the development of drug resistance [142].

**Cell cycle effects:** Chemotherapy drugs are designed to target and kill rapidly dividing cancer cells. However, they can also harm normal cells that divide quickly, such as those in the bone marrow, hair follicles, and digestive tract [144, 145]. The toxicity of these normal cells limits the maximum safe dose of chemotherapy drugs that can be given to a patient [142]. Intrinsic drug resistance can develop when cancer cells adapt to the toxic effects of chemotherapy by slowing



© University of Venda

down their division rate, making them less susceptible to the drug's cytotoxic effects. As a result, higher doses or alternative treatment strategies may be required to overcome this drug resistance. Slow-growing solid tumors are often more resistant to chemotherapy because the cells are in a non-dividing or resting state, known as the G0 phase [146]. Chemotherapy is designed to target rapidly dividing cells, which are more vulnerable to the drugs, so cells in the G0 phase are less affected by the treatment. Additionally, slow-growing tumors may have mutations that make them less responsive to chemotherapy or more adept at repairing damage caused by the drugs. As a result, these tumors may become resistant to the treatment over time, limiting the effectiveness of chemotherapy.

**Chemically induced adaptive change:** Chemically induced adaptive change refers to a type of drug resistance caused by introducing a drug or toxic substance into a cell, which sets off a series of metabolic events that result in the cell becoming resistant to further treatment with the same or similar compounds. This type of resistance is typically temporary and can be overcome when the drug or toxic substance is no longer present.

**Stress response:** A stress response is a cellular resistance induced by various environmental stressors. These stressors include pH and osmotic shock changes, ultraviolet radiation, heat, trauma, viral infection, hypoxia, and oxidative stress. When exposed to these stressors, cells can activate genetic pathways that help them adapt and survive under stressful conditions. However, these same pathways can also provide resistance to drugs [142], making the cells less susceptible to the effects of chemotherapy or antibiotics. In prokaryotes, four primary stress-induced regulons activate stress response pathways. These include the SOS (save our ship) response, the heat-shock response, the oxyR network, and the adaptive response to alkylating chemicals.

### 2.4.2. Acquired resistance

Acquired drug resistance is a phenomenon where microorganisms, including bacteria, viruses, fungi, and cancer cells, develop the ability to resist the effects of previously effective drugs in treating the infection or disease. This resistance can occur through several mechanisms, including changes in the target of the drug, alterations in drug transport or metabolism, or the development of alternative pathways that bypass the drug's action. The development of acquired drug resistance



© University of Venda

is a significant challenge in treating infectious diseases and cancer. It can significantly reduce the effectiveness of drugs and increase the risk of treatment failure. Some factors contributing to acquired drug resistance include genetic mutations, horizontal gene transfer, and environmental pressures, such as exposure to antibiotics or chemotherapeutic agents [139].

**Chromosomal-based genetic alteration:** Chromosomal-based genetic alterations are a common mechanism that promotes the development of resistance to antibiotics [147]. These alterations can occur through mutations or genetic modifications that affect the drug target or other components of the microorganism's metabolic pathways.

**Genomic duplication:** Genomic duplication, also known as polyploidization, is a process in which the entire genome of a cell is duplicated, resulting in the cell having more than two sets of chromosomes. This process is common in many eukaryotic organisms, particularly plants, but rare in animals. In cancer cells, genomic duplication can lead to increased genetic instability and contribute to drug resistance. The duplicated genome can lead to an overexpression of genes, including those involved in biosynthesis and transport, which can affect the efficacy of drugs. For example, the overexpression of drug transporters can reduce the concentration of the drug within the cell and decrease its effectiveness [148].

In addition to genomic duplication, amplification and overexpression of specific genes can contribute to drug resistance in cancer cells. This can occur through various mechanisms, such as gene mutations or changes in gene expression levels.

**Enzymatic approach to drug modification:** Enzymatic modification of drugs is a common mechanism that can lead to drug resistance. β-lactamases are enzymes that break down β-lactam antibiotics, such as penicillin and cephalosporins, by cleaving the β-lactam ring in their structure. This enzymatic activity can lead to drug resistance [149], as the antibiotic can no longer bind to its target and inhibit bacterial cell wall synthesis.

Some β-lactamases are encoded by plasmids and transposons, which can be transferred between bacteria and contribute to adaptive resistance. These enzymes can be classified into two categories



© University of Venda

based on their structural and functional characteristics: those with serine at the active catalytic site (classes A, C, and D) and those with zinc-dependent metalloenzymes (class B). EDTA can inhibit class B enzymes and are responsible for the hydrolysis of carbapenems, a class of antibiotics often used as a last resort for treating multidrug-resistant infections.

**Modulated drug targets:** In modulated drug targets, the cells or organisms can develop mechanisms to counteract the effects of the drug and maintain their normal function. There are several mechanisms by which drug resistance can develop in modulated drug targets. For example, mutations in the target molecule can alter its structure or function, making it less susceptible to the drug. Alternatively, the cells can upregulate the target molecule's expression, reducing the drug's effectiveness.

**Efflux mechanisms and membrane permeability channel:** Efflux mechanisms are one of the mechanisms by which cells can develop resistance to drugs. These mechanisms involve removing drugs from the cell by specialized proteins called efflux pumps. Efflux pumps are present in the cellular membranes of many cells, including bacteria, fungi, and human cells.

Efflux pumps use energy to pump drugs out of the cell, preventing them from reaching their intended targets and reducing their effectiveness [150]. This resistance mechanism is important in bacterial infections, as it allows bacteria to expel antibiotics and other drugs from the cell, making it difficult to treat infections.

There are several families of efflux pumps, including the ATP-binding cassette (ABC) transporter family, the major facilitator superfamily (MFS), and the resistance-nodulation-cell division (RND) family. These pumps bind to the drug molecule and use energy from ATP to pump the drug out of the cell or use an electrochemical gradient.

In addition to efflux mechanisms, drug resistance can also develop due to changes in membrane permeability channels. Membrane permeability channels transport various molecules across the



© University of Venda

cellular membrane, including drugs. Alterations in the expression or function of these channels can affect the ability of drugs to enter the cell and reach their targets.

### 2.4.3. Drug resistance for cancer

#### 2.4.3.1. Tumor drug resistance

Chemotherapy resistance is a significant challenge in cancer treatment [151], as it can limit the effectiveness of drugs that target cancer cells. Chemotherapy resistance can be either intrinsic—present in the tumor cells from the beginning or acquired, meaning it develops over time in response to treatment. Intrinsic resistance can arise due to different factors, including genetic abnormalities in tumor cells and differences in a patient's genetic makeup and metabolism. Acquired resistance, conversely, can occur due to the selective pressure of chemotherapy, leading to the survival and growth of resistant tumor cells. Understanding the mechanisms of chemotherapy resistance is critical for developing strategies to overcome it. Several molecular pathways are involved in developing drug resistance in tumors, including altered membrane transport, genetic responses, enhanced DNA repair, alterations in target molecules, etc.

##### 2.4.3.1.1. Altered membrane transport

Drug resistance-associated membrane proteins (DRAMPs) [152] play an important role in multidrug resistance (MDR) by directly expelling drug molecules out of cells, reducing intracellular accumulation, or indirectly altering drug accumulation through physicochemical processes. The presence of these proteins can affect the absorption, distribution, and excretion of various pharmacological chemicals. DRAMPs are classified into two main types: the ABC transporter superfamily [153], responsible for removing hydrophobic chemotherapy agents from tumor cells and reducing their net intracellular accumulation, ultimately decreasing efficacy, and solute carrier transporters, which increase resistance to chemotherapy by reducing the uptake of water-soluble anticancer drugs [154].

About 48 genes in humans code for ABC transporters [155], and ABC transporters are known to help move substances, including anticancer drugs, through the body. The three major classes of



© University of Venda

ABC transporters involved in multidrug resistance are P-glycoprotein, ABCG2, and multidrug resistance-associated proteins (MRPs).

**P-glycoprotein:** The MDR1 (ABCB1) gene produces the P-glycoprotein (P-gp) protein, which plays a critical role in drug transport across cell membranes. P-gp is a membrane-bound transporter in many tissues, including the liver, intestine, and blood-brain barrier. P-gp acts as a drug efflux pump, actively pumping drugs out of cells, thereby reducing their intracellular concentration. Overexpression of P-gp, due to increased expression of the ABCB1 gene, has been linked to multidrug resistance (MDR) in different types of cancer, including multiple myeloma and leukemia [139]. This overexpression of P-gp leads to the reduced effectiveness of chemotherapy drugs, as they are actively transported out of the cancer cells [139]. The ABCB1 gene has also been shown to be amplified in MDR murine melanoma cells, further demonstrating the relationship between ABCB1 overexpression and drug resistance [156]. The broad substrate selectivity of P-gp, which includes many chemotherapy drugs such as anthracyclines, vinca alkaloids, taxanes, and epipodophyllotoxins, is due to the complex structure of the P-gp protein and its ability to interact with a wide range of structurally diverse compounds. The correlation between ABCB1 overexpression and chemotherapeutic failure highlights the need for alternative treatment strategies to overcome MDR. This may involve developing new drugs that are not substrates for P-gp or using combination therapies that target multiple mechanisms of drug resistance.

**ABCG2:** ABCG2 is a gene that codes for a protein part of membrane transporters' ATP-binding cassette (ABC) superfamily. It is also called breast cancer resistance protein (BCRP) [127]. This protein is primarily expressed in the liver, small intestine, and breast tissue. It plays a critical role in protecting cells from toxins and drugs by pumping them out of cells and into the extracellular space. Research has shown that ABCG2 is highly expressed in specific cancer cells, including breast cancer cells [139]. It can contribute to drug resistance by pumping out chemotherapeutic agents before they can kill cancer cells. This can lead to the failure of chemotherapy and the progression of cancer. Several molecularly targeted anticancer drugs, such as Gefitinib and Imatinib, are substrates for ABCG2. When ABCG2 is overexpressed in cancer cells, it can pump out these drugs, reducing their effectiveness and contributing to drug resistance. It is also known



© University of Venda

that ABCG2-mediated drug resistance affects camptothecins, a class of anticancer medications used to treat various types of cancer, including breast cancer [157, 158].

**Multidrug resistance-associated protein (MRP)**: ABCC1 (MRP1) is a member of the MRP family that plays a role in drug resistance by pumping out chemotherapeutic agents from cells. Overexpression of MRP1 has been observed in several types of cancer and has been associated with drug resistance. The transfer of unmodified chemotherapeutic agents via MRP1 requires reduced glutathione (GSH) [159], and the use of a peptidomimetic glutathione conjugate of ethacrynic acid (EA), GS-EA, has been shown to prevent MRP1-mediated efflux of drugs [139] and enhance the cytotoxic effects of chemotherapy.

#### 2.4.3.1.2. Genetic responses

In some cancer cases, mutant p53 can render treatments that damage DNA ineffective against some tumors. This is because mutant p53 may not be able to trigger apoptosis in cells with sustained DNA damage. In other words, the cancer cells with mutant p53 can continue to grow and divide despite DNA damage caused by chemotherapy or radiation therapy. Furthermore, deleting the p53 gene has been linked to multidrug resistance in numerous cancer cases [160-163]. Cancer cells with deleted p53 genes may be resistant to multiple types of chemotherapy drugs, making it even more challenging to treat these tumors.

Lower levels of p53 expression in human breast cancer cells have been found to change the sensitivity of those cells to chemotherapy drugs, such as paclitaxel and 5-fluorouracil [164, 165]. In addition, increased resistance to several molecularly targeted medications has been linked to gene amplification of specific genes involved in drug resistance. Gene rearrangements in the chromosomes of drug-resistant cancer cell lines and DNA from drug-resistant leukemia patients have also been associated with the first activation or increased expression of the MDR1 gene, which encodes the P-glycoprotein transporter and is a significant contributor to drug resistance in cancer cells. Understanding these genetic responses is essential in developing new strategies to overcome drug resistance in cancer.



© University of Venda

#### 2.4.3.1.3. Enhanced DNA repair

One of the ways that cancer cells can become resistant to chemotherapy is by enhancing their ability to repair DNA damage caused by chemotherapeutic agents [166]. By enhancing DNA repair mechanisms, cancer cells can repair the damage caused by chemotherapy more quickly and effectively than normal cells, making them more resistant to treatment. Cancer cells can upregulate the expression of DNA repair proteins, such as the xeroderma pigmentosum group E binding factor (XPE) [166], which is involved in nucleotide excision repair (NER) of cisplatin-DNA adducts. Other DNA repair pathways, such as homologous recombination (HR) and non-homologous end joining (NHEJ), can also be upregulated in cancer cells, allowing them to repair DNA damage more efficiently and resist the effects of chemotherapy. DNA-binding protein (excision repair cross-complementing protein, ERCC1) is another DNA repair protein upregulated in cancer cells resistant to platinum-based chemotherapy drugs, including cisplatin and carboplatin. ERCC1 is involved in the NER pathway and is critical in repairing DNA damage induced by platinum-based drugs. Its upregulation in cancer cells is thought to contribute to the development of drug resistance by enhancing DNA repair and reducing the amount of drug-induced DNA damage that would otherwise lead to cell death.

#### 2.4.3.1.4. Metabolic effects

How the body breaks down drugs and other xenobiotic substances can significantly influence their effectiveness. The body utilizes drug-metabolizing enzymes to break down drugs, and these enzymes play an important role in determining the effectiveness of chemotherapy. Chemotherapy resistance can occur when drug-metabolizing enzymes are overexpressed, leading to increased drug breakdown and elimination from the body. This results in lower drug concentrations in the tumor tissue and decreased effectiveness.

Several drug-metabolizing enzymes, including cytochrome P450, glutathione S-transferase, and uridine 5'-diphospho-glucuronosyltransferase (UDP), have been implicated in chemotherapy resistance. These enzymes can break down a broad range of chemotherapy drugs, such as cisplatin, doxorubicin, and paclitaxel. Genetic variations in drug-metabolizing enzymes can also impact



© University of Venda

chemotherapy response. Specific individuals may have genetic polymorphisms that reduce enzyme activity, leading to higher drug concentrations in the body and increased toxicity.

#### 2.4.3.1.5. Alterations in target molecules

Alterations in the drug target are a common mechanism underlying the development of drug resistance. Mutations in the gene encoding the drug target can alter the structure or function of the target protein, making it less susceptible to the drug's action. In the case of topoisomerase II, mutations that alter its nuclear location can lead to resistance to drugs like etoposide that target the enzyme. Loss of the drug-sensitive gene or chromosomal rearrangements can also affect the function of multiple biochemical pathways, leading to changes in the drug's mechanism of action and eventual resistance. One clinical example of resistance caused by alterations in target molecules is seen in treating chronic myeloid leukemia (CML) with tyrosine kinase inhibitors (TKIs) such as imatinib [167]. The Philadelphia chromosome causes CML, a genetic abnormality that produces the BCR-ABL fusion protein, which drives the growth of leukemia cells. Imatinib targets the BCR-ABL protein and effectively induces remission in CML patients. However, some patients can resist imatinib due to mutations in the BCR-ABL gene that alter the drug's binding site, making it less effective.

Tumor size can also affect drug efficacy, as larger tumors may have limited blood supply and reduced access to the drug [168]. In treating NSCLC with epidermal growth factor receptor (EGFR) inhibitors such as gefitinib and erlotinib, EGFR mutations in NSCLC cells make them dependent on EGFR signaling for growth and survival. EGFR inhibitors block this signaling and can induce tumor regression. However, some NSCLC patients can develop resistance to EGFR inhibitors due to the emergence of a secondary EGFR mutation (T790M) that alters the binding site of the drug and reduces its effectiveness.

#### 2.4.3.1.6. Growth factors

Cancer drug resistance has been linked to growth factors like EGFR and HER2. For example, in NSCLC, mutations in the EGFR gene can cause cancer cells to become immune to EGFR-targeted



© University of Venda

drugs like gefitinib and erlotinib. Additionally, amplifying the HER2 gene in breast cancer can lead to resistance to chemotherapy drugs such as doxorubicin and cyclophosphamide. Inhibitors of these growth factors, such as trastuzumab for HER2-positive breast cancer, have been developed to overcome drug resistance.

### 2.4.3.2. Drug resistance for malaria

*Plasmodium falciparum*, *Plasmodium vivax*, and *Plasmodium malariae* [142] malaria are serious health risks that are difficult to treat due to drug resistance. Resistance to antimalarial drugs is a complex phenomenon that involves various genetic and environmental factors. Full adherence to prescribed treatment is vital to prevent the development of drug resistance, but other factors can contribute to the development of drug resistance in malaria. Extrinsic factors such as limited access and availability of pharmaceuticals due to financial constraints, distribution statistics, and user resistance can result in inadequate medical treatment. Intrinsic factors include the parasite's cellular, molecular, and clinical characteristics, as well as its species-based innate resistance, drug activity spectrum, and pharmacological response to the parasite's stage susceptibility.

The emergence of drug resistance in parasites typically involves two stages: the initial emergence of drug tolerance, followed by the spread of resistant parasites in the population. Subtherapeutic drug exposure, which results in the death of the susceptible lot, promotes the emergence of drug-resistant parasites and contributes to the spread of drug resistance. Subtherapeutic drug exposure refers to using a drug at a lower dose than required to entirely kill the parasite, resulting in the survival of some parasites that have acquired resistance to the drug. Parasites with this resistance are more likely to reproduce and spread, which can result in the establishment and spread of drug-resistant parasite strains. This is why it is important to use antiparasitic drugs at the correct dose and in combination with other drugs and to prevent the overuse or misuse of these drugs to slow down the development and spread of drug resistance.

During the initial stage, the parasite population may contain a few individuals that have acquired genetic mutations that make them less susceptible to the drug. These parasites are said to be "tolerant" to the drug. If drug treatment continues, these tolerant parasites may survive and



© University of Venda

reproduce, passing their resistance genes to their offspring. Over time, the proportion of resistant parasites in the population increases, leading to the spread of resistance. The spread of drug resistance can be influenced by several factors, including the intensity of drug use, the frequency of drug-resistance genes in the parasite population, and the parasite's ability to be transmitted to new hosts. In malaria, the spread of drug resistance has been facilitated by using suboptimal drug regimens, which provide selective pressure for the emergence of resistant parasites and the widespread distribution of resistant parasites through mosquito vectors.

#### 2.4.3.2.1. Genetics of antimalarial resistance

Resistance to antiparasitic drugs can result from mutations in genes encoding drug targets and genes encoding efflux transporters that remove the drug from the parasite before it can have its intended effect [142, 169]. When exposed to a drug, a parasite may experience a fitness loss, reducing its ability to survive and reproduce. However, mutations can arise in the parasite's DNA, allowing it to tolerate the drug better and maintain its fitness. These mutations can occur spontaneously or be selected during treatment, leading to the development of drug resistance. The exact genetic changes that confer resistance can vary depending on the parasite species and the drug being used. However, they often involve alterations to the genes that control the target of the drug or the mechanisms of drug uptake and efflux.

Here are some brief explanations of the mode of action for some of the most extensively studied antimalarial resistance genes:

***P. falciparum* chloroquine resistance transporter (Pfcrt):** This gene codes for a transporter protein responsible for moving chloroquine out of the parasite's digestive vacuole, which interferes with the parasite's ability to break down hemoglobin [170]. Mutations in this gene can prevent chloroquine from binding to the transporter, resulting in decreased drug accumulation in the digestive vacuole and reduced drug efficacy. The exact mechanism by which Pfcrt mutations lead to chloroquine resistance is still not fully understood. However, it is thought to involve the altered drug transport into the parasite's digestive vacuole.



© University of Venda

***P. falciparum* multidrug resistance 1 (Pfmdr1):** This gene codes for a transporter protein involved in the export of a variety of drugs out of the parasite's cell. Mutations in this gene can increase the parasite's ability to pump out drugs, leading to decreased drug accumulation and reduced drug efficacy [170].

**Dihydrofolate reductase (DHFR):** This gene codes for an enzyme involved in synthesizing nucleotides, essential for DNA replication. Drugs that target DHFR, such as pyrimethamine and sulfadoxine, interfere with nucleotide synthesis and inhibit parasite growth **[**171]. This gene's mutation can reduce drug binding and increase the enzyme's activity, decreasing drug efficacy.

**Dihydropteroate synthase (DHPS):** This gene codes for an enzyme involved in folate synthesis essential for nucleotide synthesis. Drugs targeting DHPS, such as sulfadoxine, inhibit folate synthesis and parasite growth. This gene's mutation can reduce drug binding and increase the enzyme's activity, decreasing drug efficacy.

**PfK13:** This gene encodes the kelch protein, which is involved in the artemisinin resistance of P. falciparum. Mutations in this gene are associated with delayed parasite clearance after artemisinin treatment. The mechanism by which PfK13 mutations lead to artemisinin resistance is still not fully understood, but it is thought to involve altered protein-protein interactions in the parasite.

**Cytochrome b:** This gene encodes a protein involved in the electron transport chain in the parasite's mitochondria. Mutations in this gene are associated with resistance to atovaquone, a drug that targets the mitochondrial respiratory chain. The mutations affect the binding affinity of the protein for the drug, leading to reduced drug efficacy.

Each gene works through a different mechanism to influence drug efficacy, and understanding these mechanisms is important for developing new drugs and combatting resistance. Host immunity is vital in modulating spontaneous genetic changes in malaria parasites. Malaria parasites have evolved complex mechanisms to evade the host immune system, which include antigenic variation, sequestration, and suppression of immune responses. However, the host



© University of Venda

immune system also can recognize and eliminate the parasite, which puts selective pressure on the parasite population.

***P. falciparum* multidrug resistance (Pfmdr):** Genomic duplication in the Pfmdr gene is a known mechanism of antimalarial resistance [172]. The Pfmdr gene is located on chromosome 5 of *P. falciparum* and encodes a membrane-bound transporter protein. This protein plays an important role in regulating the intracellular concentration of drugs by mediating their influx and efflux. The duplication of the Pfmdr gene has been associated with decreased susceptibility to several antimalarial drugs, including mefloquine, chloroquine, and artemisinin. This is because the overexpression of Pfmdr leads to increased efflux of these drugs, reducing their intracellular concentration and efficacy.

Mutations in specific regions of the Pfmdr gene have also been associated with drug resistance. For example, the N86Y mutation in Pfmdr has been linked to resistance to mefloquine, halofantrine, and amodiaquine [172], while the Y184F mutation is associated with decreased susceptibility to artemisinin-based drugs. The combination of gene amplification and mutations in Pfmdr can result in even higher levels of drug resistance.

### 2.5. Acylphloroglucinols (ACPLs)

This section reviews studies on ACPL compounds with anticancer and/or antimalarial effects, including information on their molecular characteristics, natural sources, and activities. In addition, computational studies were also included. A comprehensive overview of studies on ACPLs with anticancer and/or antimalarial activities has been presented in [173], and only a summary is reported here.

#### 2.5.1. Structures and molecular properties of acylphloroglucinols

Phloroglucinol (1,3,5-trihydroxybenzene) derivatives known as acylphloroglucinols (ACPLs), shown in Fig. 2.12, are characterized by the presence of one or more acyl (R–C=O) groups. ACPLs are the largest class of naturally occurring phloroglucinol derivatives [174]. Many biologically active ACPLs have medicinal benefits such as antibacterial, antifungal, antioxidant, anticancer, antiplasmodial, antiviral effects, and more [174]. This has led to extensive research in various



© University of Venda

fields, ranging from ACPLs' potential as lead structures for drugs designed to treat degenerative illnesses [174], to more recent studies on their pharmacological abilities against the SARS-CoV-2 virus [175]. As a result, the search for feasible synthetic pathways has become more active [174]. The study of specific subclasses of ACPLs primarily involves investigating the effects of structural alterations, such as substituting certain groups with others, adding specific substituents, or hydrogenating C=C double bonds found in prenyl chains. These modifications provide insight into the significance of individual structural components, including the OH groups, the acyl group, the prenyl chains, and the presence of fused rings in the acylphloroglucinol moiety. Additionally, these changes may lead to the development of molecular structures with enhanced activity.

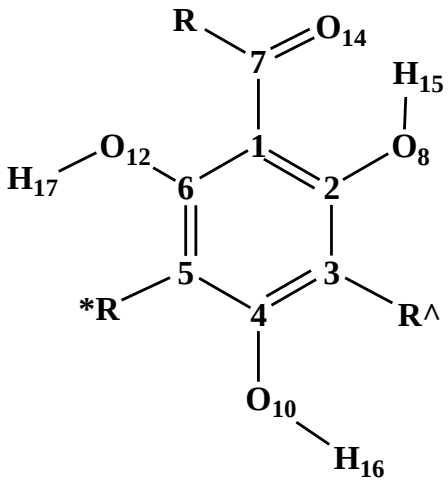



**Figure 2.12. The general structure of acylphloroglucinols and atom-numbering utilized in this study.**
The carbon atoms are denoted by their numbers. When R^ ≠ H, its first C atom is given the number 9; when R* ≠ H, its first C atom is given the number 11.

Based on their structural characteristics, ACPLs can be categorized into several subclasses. Classifications of ACPLs in terms of their structural features may refer to the number of phloroglucinol units or the types of substituents, as shown in Fig. 2.13. The ACPLs considered in



© University of Venda

this study can be classified into the following subclasses, which will hereafter be denoted with the acronyms shown in parentheses: simple ACPLs, polyprenylated ACPLs (PP-ACPLs), prenylated ACPLs (P-ACPLs), phloroglucinol glycosides (PG-ACPLs), phloroglucinol-terpene adducts (PTA-ACPLs), dimeric ACPLS (D-ACPLs), trimeric ACPLs (T-ACPLs), cyclic polyketides (CP-ACPLs), polycyclic polyprenylated ACPLs (PCP-ACPLs). It is also to be noted that some molecules may be considered under more than one group; for instance, iriflophenone 4-glucoside can be viewed as a simple ACPL and as a PG-ACPL; since it is the only PG-ACPL considered in this study, it is expedient to group it with the simple ACPL (obviously also mentioning that it is a PG-ACPL).



© University of Venda

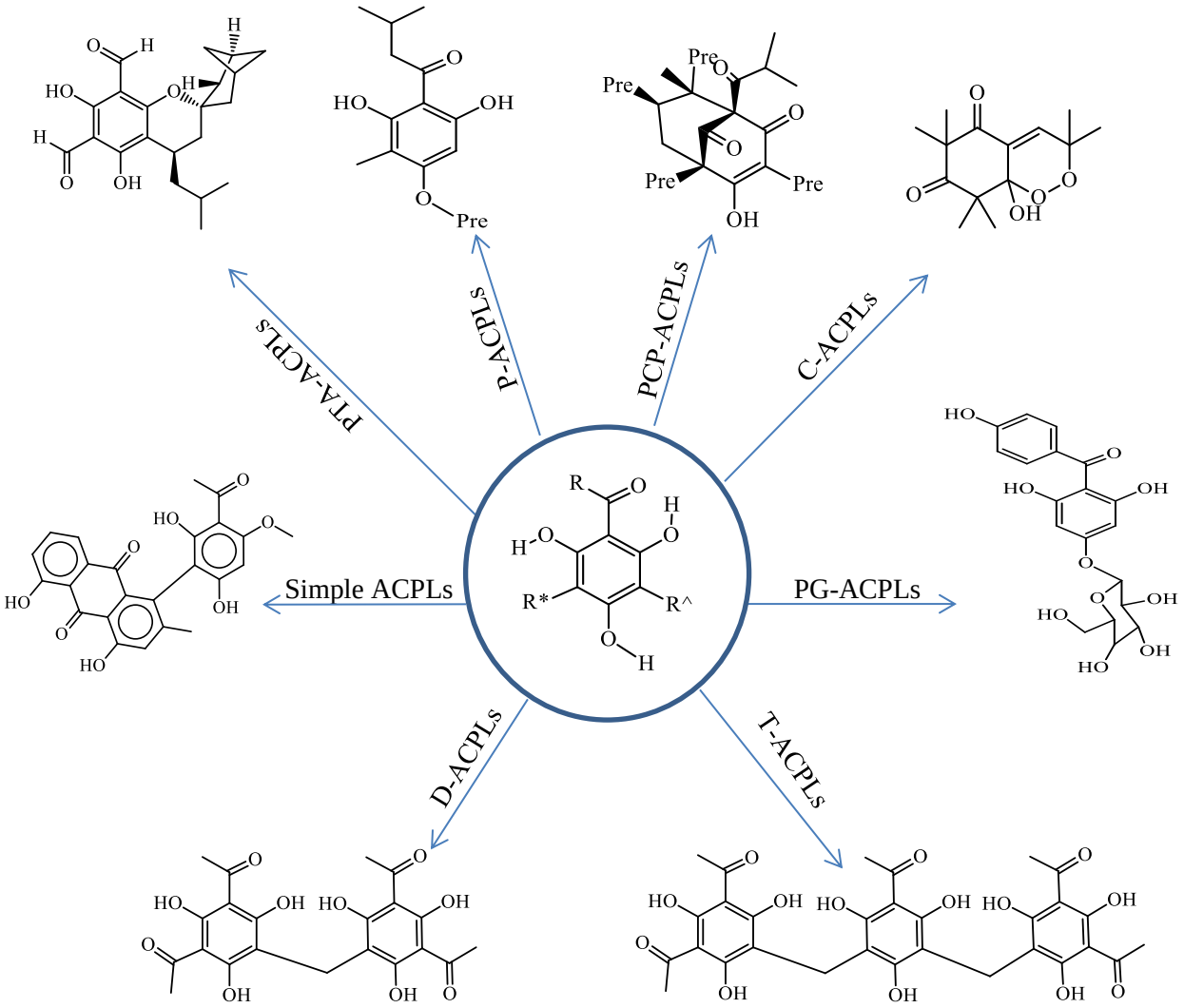



**Figure 2.13. Examples of molecular structures of subclasses of acylphloroglucinols.**

Some ACPLs contain multiple monomeric units linked by methylene bridges (Fig. 2.13.), enhancing biological activity. Their molecules may contain-two, three, or more monomeric units [176].

The molecular properties of active compounds determine their biological activity, so understanding these properties is important for understanding the biological effects of different compounds. The molecular properties of ACPLs will be briefly outlined to gain a better



© University of Venda

understanding, focusing on their conformational preferences. Intramolecular hydrogen bonds (IHBs) are especially important for biologically active molecules [177], as they can significantly impact important mechanisms such as selective binding, molecular recognition, and anticancer activity. Therefore, these bonds will be given special attention.

In ACPLs with at least one phenol OH ortho to the acyl group, low-energy conformers are characterized by an IHB between a neighboring OH and the $sp^2$ O of the acyl group. Figure 2.14 shows how the IHB forms between H15 and O14 (O8–H15···O14 IHB). If the acyl group is oriented to the other side, the IHB forms between H17 and O14 (O12–H17···O14 IHB). Such an IHB is conveniently called the 'first IHB' [178-180].

Some ACPLs contain two acyl groups, and then they can form two IHBs of this type. For example, euglobals (EG) in section 4 have lower energy conformers, including two first intramolecular hydrogen bonds (IHBs) for each group. H-bond donor or acceptor groups in R, R*, or R^ can form additional IHBs [181]. The first IHBs are often the strongest in ACPLs because the acceptor is a $sp^2$ O, and the ring they close gives them resonance-assisted hydrogen bond characteristics [182]. These bonds are present in crystal structures and when the molecule is in an aqueous solution.



© University of Venda

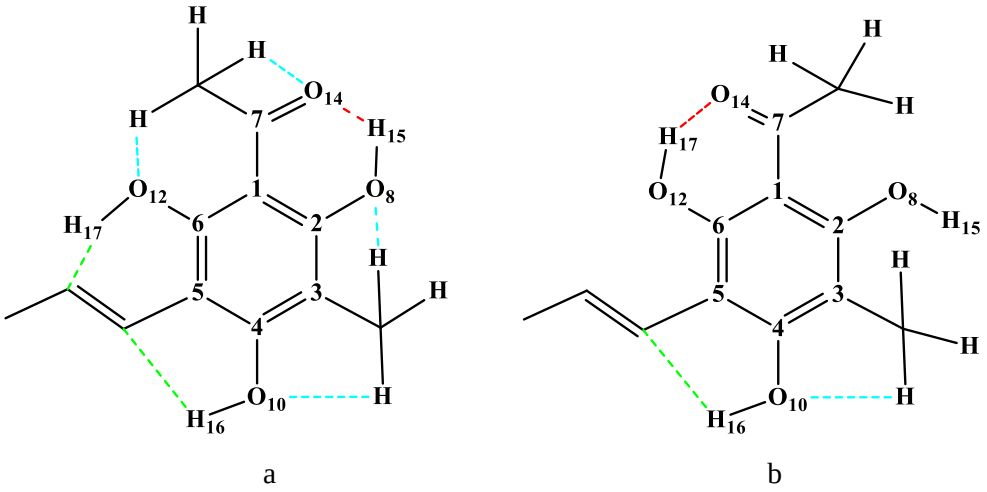



**Figure 2.14. Intramolecular hydrogen bonds (IHBs) in simple acylphloroglucinols.**

A model molecular structure is used as an illustration. The IHBs between the O of the acyl group and a neighboring OH are indicated by dashed red segments; dashed light blue segments show the weaker IHBs between C-H and neighboring O atoms, the one between O–H and π bond. a) shows a case in which the first IHB forms between H15 and O14, and b) shows a case in which the first IHB forms between H17 and O14.

IHBs that involve weaker C-H bonds (Fig. 2.14) are always present and may influence the orientation of nearby substituents [183, 184]. In ACPLs where the R, R^, or R* groups have π bonds or systems, the O–H···π IHBs (Fig. 2.14) can also form. These may determine the orientation of the substituent chains, such as prenyl or geranyl chains. All IHBs have a stabilizing effect, with the strength of the effect increasing with the strength of the IHB. The O-H···O IHBs are the most dominant stabilizing factor and have a significant influence on the conformational preferences of ACPLs, as well as other molecular properties. If a molecule has multiple possible IHBs, the lower energy conformers contain the maximum number of simultaneous IHBs. All lower energy conformers of ACPLs with more than one acylphloroglucinol unit have IHBs that link the monomers, with some conformers containing two or more such IHBs.

ACPLs can be stabilized by several factors, including the positioning of OH groups in monomeric ACPLs [180]. For ACPLs with multiple units, each individual monomer's conformation and orientation are important for stabilization [185-187]. In cases where ACPLs have more than two units, the mutual orientation of methylene bridges becomes significant [187]. In cases where one of the substituents is another ring connected to the acylphloroglucinol moiety by a methylene



© University of Venda

bridge and contains O atoms or OH groups, as seen in arzanol [188], similar things can be observed. When interacting with a biological target, IHB acceptors or donors among the IHB group could potentially become involved in intermolecular hydrogen bonds. The benzene ring can also interact with the target through O–H···π interactions or O–H···π H-bonding, particularly if the target contains aromatic rings susceptible to stacking interactions.

### 2.5.2. Overview of available information on acylphloroglucinols with anticancer and/or antimalarial activities

#### 2.5.2.1. Considering ACPLs in terms of specific biological activities

This study focuses on ACPLs with potential antimalarial and/or anticancer activities. While many compounds exhibit activities against both diseases, some may have only one reported activity, possibly due to a lack of testing for the other. So far, ACPLs have not been used clinically for malaria or cancer treatment, and resistance has not yet emerged towards them. Some ACPLs are considered lead compounds for more potent drugs, suggesting potential roles for antimalarial or anticancer applications.

The organization of ACPLs in this study prioritizes biological activity as the primary criterion, with structural features considered secondarily. The subdivision into subclasses aims to align with Quantitative Structure-Activity Relationships (QSAR) criteria (QSAR uses activity as a primary criterion and structural analogies as an instrumental criterion.). The molecular structures of 149 ACPLs are presented, denoted by numbers for easy reference. Relevant information, including natural sources, pharmacological activities and their determination and targets, and computational study results, is presented for each compound. Computational studies are emphasized for their role in elucidating drug mechanisms and facilitating drug design.

#### 2.5.2.2. Acylphloroglucinols for which both anticancer and antimalarial activities are reported

Compounds **1-31** have been reported to exhibit both antimalarial and anticancer properties. Most of these compounds (compounds **1-26,** Fig. 2.15) are polycyclic polyprenylated ACPLs (PCP-ACPLs). These molecules contain multiple prenyl chains attached to the ACPL moiety, along with

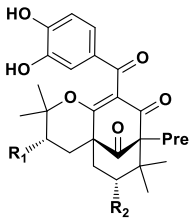


© University of Venda

additional rings. Although the ACPL moiety is altered from its "canonical" structure depicted in Fig. 2.12, it remains recognizable. Additionally, compounds **1-23** can also be classified as benzophenones due to their benzoyl-phloroglucinol skeleton. The molecules of compounds **27-31** (Fig. 2.15) contain a "canonical" ACPL moiety with an attached ring system. As there is only one "canonical" ACPL moiety present, they can be classified as "simple ACPLs" (although no operational definition for "simple ACPLs" has been found in available literature).



© University of Venda



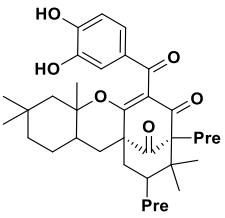
**1**: $R_1$ = prenyl, $R_2$ = prenyl

**2**: $R_1$ = $CH_2CH_2C(CH_2)(CH_3)$, $R_2$ = prenyl

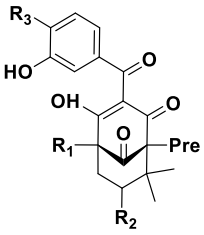
**3**: $R_1$ = $CHCHC(CH_3)_2OH$, $R_2$ = prenyl

**4**: $R_1$ = $CH_2CH_2C(CH_3)_2OH$, $R_2$ = prenyl

**5** and **6**: $R_1$ = prenyl, $R_2$ = $CH_2CHOHC(CH_3)_2OH$



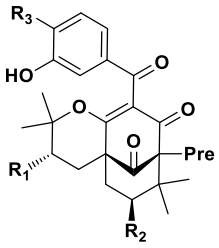
**7**.



**8**: $R_1$ = (S)-isogeranyl, $R_2$ = prenyl, $R_3$ = OH

**9**: $R_1$ = (S)-isogeranyl, $R_{12}$ = β-prenyl, $R_3$ = OH

**10**: $R_1$ = isogeranyl, $R_2$ = prenyl, $R_3$ = H



**11**: $R_1$ = prenyl, $R_2$ = prenyl, $R_3$ = OH

**12**: $R_1$ = prenyl, $R_2$ = prenyl, $R_3$ = H

**13**: $R_1$ = prenyl, $R_2$ = $CHCHC(CH_2)(CH_3)$, $R_3$ = OH

**14**: $R_1$ = prenyl, $R_2$ = $CH_2CHC(CH_3)(CH_2$-prenyl), $R_3$ = OH

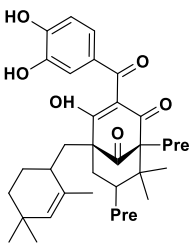
**15**: $R_1$ = prenyl, $R_2$ = $CHCHC(CH_3)_2OH$, $R_3$ = OH

**16**: $R_1$ = $CHCHC(CH_3)_2OH$, $R_2$ = prenyl, $R_3$ = OH

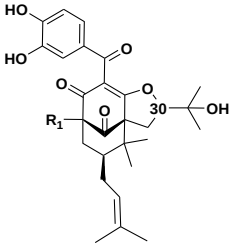
**17**, **18**: $R_1$ = $CH_2CHOHC(CH_3)_2OH$, $R_2$ = prenyl, $R_3$ = OH

**16**: $R_1$ = $CHCHC(CH_3)_2OH$, $R_2$ = prenyl, $R_3$ = OH

**17**, **18**: $R_1$ = $CH_2CHOHC(CH_3)_2OH$, $R_2$ = prenyl, $R_3$ = OH

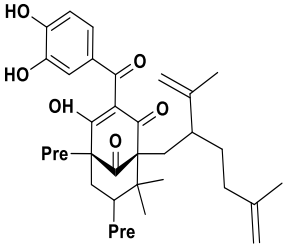


**19**

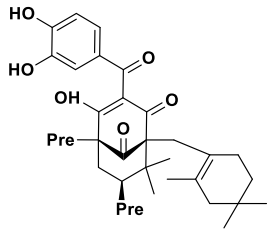


**20**: 30 α, $R_1$ = prenyl

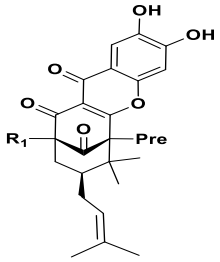
**21**: 30 β, $R_1$ = prenyl



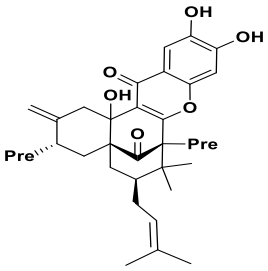
**22**



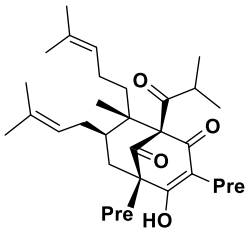
**23**



**24:** $R_1$ = (S)-isogeranyl



**25**



**26**

**Figure 2.15. Molecular structures of polycyclic polyprenylated acylphloroglucinols having both anticancer and antimalarial activities.**



© University of Venda

ACPLs such as compounds **1-7, 11-18, 20, 21,** and **24-27** provide instances where the first IHB can not be formed. Compounds **1-25** can have an O–H···O IHB between the two consecutive OHs on the benzoyl ring. The existence of prenyl chains in molecules **1-27** does not facilitate the formation of O–H···π IHBs, as either no OH is close enough to the prenyl chain to enable IHB formation, or the available OH is in a position to form the first IHB, which is much stronger. Consequently, PCP-ACPLs have fewer opportunities for IHB creation than other ACPLs, despite having multiple H-bond acceptors such as the strong $sp^2$ O atoms and weaker acceptors like the O in the heterocycles and the π bond in the prenyl chains that are available for intermolecular H-bonding with biological targets.

Compounds **1-26** are isolated from Clusiaceae or Hypericaceae floras [189, 190]. Among them, compounds **1-11** and **23-25** (the former plucked from Moronobea coccinea's trunk latex) combat the chloroquine-resistant FcB1 strain of *P. falciparum* and damage human MRC-5 cells when tested in vitro. Compounds with heterocycles fused to the ACPL moiety demonstrate heightened effectiveness [189, 191]. The half-maximal inhibitory concentration ($IC_{50}$) values of each individual molecule for both activities can be found in [189]. The compounds containing a tetrahydropyrane ring (**1–7, 11–18, 20, 21, 24, 25**) have recorded the best $IC_{50}$ values for antimalarial activity, ranging from 2.1-9.0 µM. Meanwhile, compounds (**8–10, 19, 22, 23**) have $IC_{50}$ values that exceed 10 µM [189]. The significance of the proximity of two OH groups on the benzoyl ring is shown when comparing the antiplasmodial functions of garcinol **(8**) and dehydroxygarcinol (**10**). Interestingly, this feature is also relevant for the antioxidant activities of polyphenolic compounds, given the IHB they form and the stabilizing effect on radicals that are formed when activated. The lowest cytotoxicity is found in compounds that have side chains that have been hydroxylated (**3-6**).

Compounds **12–22**, derived from *Symphonia globulifera* root barks, demonstrate stronger antiplasmodial activity against the FcB1 strain, with $IC_{50}$ values ranging from 2.1 to 10.1 µM. They also exhibit moderate cytotoxicity towards the MRC-5 cell line, displaying $IC_{50}$ values ranging from 3.7 to 19.5 µM [191].



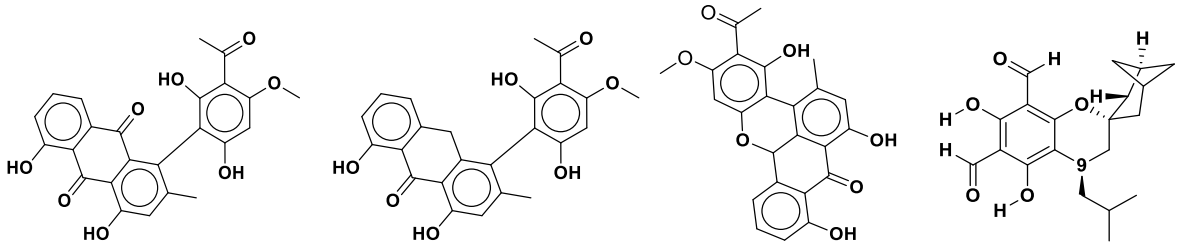
© University of Venda

Hyperforin (**26**), a *Hypericum perforatum L* (St. John's wort) bioactive constituent, exhibits several biological activities, including antidepressant, antiviral, anticancer, and antimalarial. It is known as a promising lead compound due to its ability to prevent and control cancer spread and metastasis. Administered orally, regular consumption of St. John's wort has been associated with a reduced cancer risk by modifying the genotoxic effects of carcinogens. Hyperforin's mechanisms involve decreasing inflammatory mediators, regulating reactive oxygen species (ROS) production, maintaining pH balance, and countering malignant phenotypes. In vivo studies confirm its inhibitory effects on various human and rat tumor cell lines, displaying $IC_{50}$ values within the 3 to 15 µM range. Hyperforin exhibits modest antimalarial activity, inhibiting *Plasmodium* growth at micromolar concentrations ($IC_{50}$ of 2.1 µM). Notably, its antimalarial efficacy is independent of phenol-like sensitivity to auto-oxidation and prenyl double bonds. Like monensin, a prominent antimalarial, hyperforin affects several monoamine neural transporters. In vitro and in vivo tests have shown that certain hyperforin-related acylphloroglucinols possess antiplasmodial activity. When comparing derivatives, structure-activity relationship (SAR) investigations have demonstrated that a prenyl chain double bond is crucial for anticancer activity but does not affect antimalarial activity [192].

Knipholone (**27**) and its derivatives, knipholone anthrone (**28**) and knipholone cyclooxanthrone (**29**) (Fig. 2.16), originally isolated from *Kniphofia foliosa*, an Ethiopian medicinal plant, exhibit antiplasmodial activity against the 3D7 strain of *P. falciparum* ($IC_{50}$ values: 1.9 µM for knipholone, 0.7 µM for knipholone anthrone) [176, 193]. Knipholone demonstrates substantial cytotoxicity against Jurkat, HEK293, and SH-SY5Y cells, with approximately 62–95% growth inhibition at 50 µM. Knipholone cyclooxanthrone exhibits activity against chloroquine-resistant strains of P. falciparum, with an $IC_{50}$ value of 6.13 µM [194]. In certain media, knipholone anthrone readily oxidizes to form knipholone, suggesting a potential influence on measured biological activity. The planar anthraquinone moiety in **27–29** suggests a plausible DNA intercalation mechanism.



© University of Venda



**27** **28** **29** **30**: 9β **31**: 9α

**Figure 2.16. Molecular structures of knipholone and related structures (27-29), and robustadials (30, 31).**

In addition to using spectral analysis, the determination of their structure involved converting them into known compounds: knipholone [195], knipholone anthrone [196], and knipholone cyclooxanthrone [194]. The molecules of these compounds contain an anthraquinone moiety attached to the ACPL moiety in a *meta* position to the acyl group. It has been proposed that this coupling may occur through the oxidative coupling of chrysophanol with acetylphloroglucinol [176]. Knipholone and knipholone anthrone are effective against the 3D7 strain of P. falciparum, with $IC_{50}$ values of 1.9 and 0.7 µM, respectively [193]. In addition, knipholone exhibits significant cytotoxicity against Jurkat, HEK293, and SH-SY5Y cells, causing growth inhibition of about 62-95% at 50 µM. On the other hand, knipholone anthrone reduces cell viability by 26%, 48%, and 70% for the same three strains, respectively [193].

Knipholone cyclooxanthrone has been found to be effective against strains of *P. falciparum* that are resistant to chloroquine, as indicated by its $IC_{50}$ value of 6.13 µM [194]. However, in certain media, the oxidation of knipholone anthrone can result in the formation of knipholone and other decomposition products, which may influence the biological activity measured in those media [193]. The anthraquinone moiety that is planar in **27-29** suggests that their activities may be due to their ability to intercalate DNA.

The robustadials A (**30**) and B **(31**), which are diastereomeric compounds, are adducts of phloroglucinol-terpene ACPLs (P-TA-ACPLs). As a result of their ACPL moiety containing two



© University of Venda

CHO acyl groups (at C1 and C5), the robustadials are part of the euglobals subclass [176]. They can also be considered cyclo-adducts of the jensenone ACPL with β-pinene [176]. Robustadials are present in *Eucalyptus robusta* [199], a plant used in traditional Chinese medicine to treat bacterial infections, dysentery, and malaria. They are also found in the fruits of *Eucalyptus globulus Labill*. Robustadials have been shown to have potent antimalarial activity against *Plasmodium berghei in vivo* [199]. Moreover, they exhibit anticancer activity by selectively inhibiting tyrosyl-DNA phosphodiesterase 2 (TDP2), which is responsible for repairing topoisomerase 2-mediated DNA damage that results in cancer cell drug resistance. The inhibitory activity of robustadials A and B has EC50 values of 17 and 42 μM, respectively [197]. Molecular docking studies indicate that the interaction between robustadials and the TDP2 enzyme involves two H-bonds: one (2.7 Å) between O12···H17 and the Arg231 residue of TDP2 and the other (2.9 Å) between the O atom of the pyran ring and the Arg266 residue.

According to molecular docking studies, it is suggested that the TDP2 enzyme interacts with robustadials via two hydrogen bonds [reference]. The first bond (2.7 Å) is formed between O12···H17 and the Arg231 residue of TDP2, while the second bond (2.9 Å) is formed between the O atom within the pyran ring and the Arg266 residue. A hydrophobic interaction exists between the isobutyl group attached to the pyran ring and the Ala309 and Ile307 residues [197]. Remarkable synergistic effects have been observed between robustadial A and the anticancer drug etoposide in specific types of human cancer cell lines, including non-small cell lung cancer (A549), prostate cancer (DU145), breast cancer (MCF-7), colorectal adenocarcinoma (HCT-116), and chicken lymphoma (DT40) [197]. This data suggests robustadial A-based TDP2 selective inhibitors may hold significant potential for future developments [197].

#### 2.5.2.3. Acylphloroglucinols for which antimalarial activity is chiefly reported

In literature, compounds **32-50** (Fig. 2.17) have been extensively studied for their potential antimalarial properties. Within this group, compound 1-(2,6-dihydroxy-3-methyl-4-(3-methylbut-2-enyloxy)phenyl)-3-methylbutan-1-one **(32**), which is the simplest in structure with a prenyl chain in place of H16, has attracted particular attention. Obtained from *Hypericum calycinum*'s



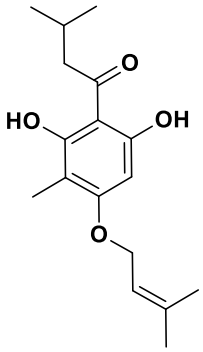
© University of Venda

aerial parts through petroleum ether extraction, compound **32** demonstrated significant antimalarial activity in vitro, presenting an $IC_{50}$ of 0.88 µg $mL^{-1}$ against *P. falciparum* [176].

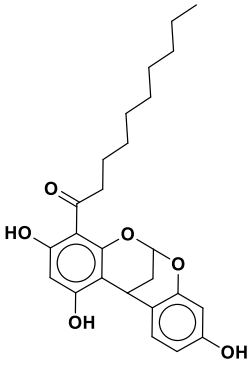
Myristicyclins A **(33)** and B **(34)** are structural isomers phloroglucinol-terpene adducts that were isolated from an extract of *Horsfieldia spicat* from Papua New Guinea [200, 201]. They possess antimalarial activity against the ring, trophonezoite, and schizont stages of *P. falciparum*, with $IC_{50}$ values of 35, 43, and 54 µM for each of these stages respectively for myristicyclin A and 10.0, 6.6, and 7.9 µM respectively for myristicyclin B [200]. However, the potential use of myristicyclin B as a therapeutic drug is limited by the fact that it causes hemolysis at 230 µM [200]. Extensive studies have been performed to determine the key structural attributes contributing to myristicyclins A and B antimalarial actions. These studies have also broadened to examine the effects of other similar compounds, such as A-type procyanidins and myristinin A [200]. A-type procyanidins incorporate the polycyclic core of myristicyclins A and B; they exhibit various properties, including antioxidant, anticancer, anti-bacterial, and antiparasitic. Myristinin A, on the other hand, has been proven to induce damage to DNA and hinder DNA polymerase β at an $IC_{50}$ of approximately 1 µM [202]. The *P. falciparum* parasite can be targeted with compounds designed to regulate mammalian cell signaling pathways. Within erythrocytes, which have an oxidative environment, the parasite resides for a significant amount of time. Proanthocyanidins may interfere with crucial pathways in the parasite's metabolism by creating and then neutralizing radicals, known as reactive oxygen species (ROS). Compounds related to myristinin A may affect DNA damage and cell signaling by potentially involving ROS. Interestingly, malabaricone A, which has a 2-acyl-1,3-dihydroxyphenyl group, has demonstrated activity against P. falciparum, suggesting the possible contributions of the acyl group and the phenol OHs.



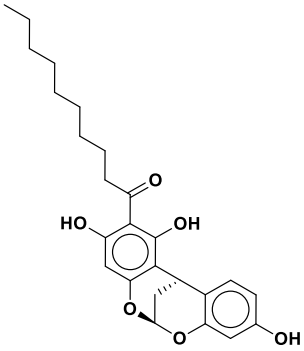
© University of Venda



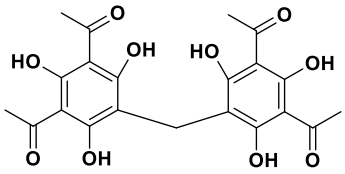
**32**



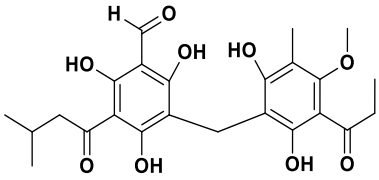
**33**



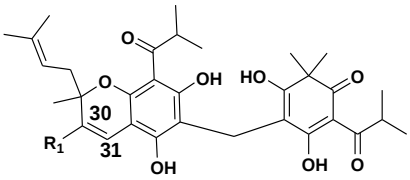
**34**



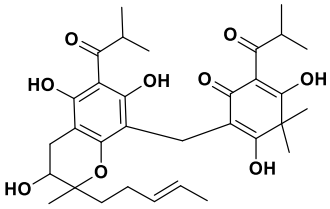
**35**



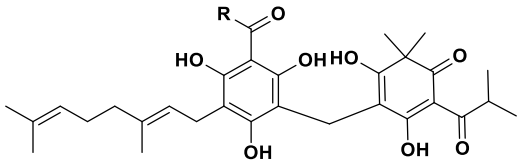
**36**



**37**: $R_1$=H **38**: $R_1$ = OH Δ30, 31

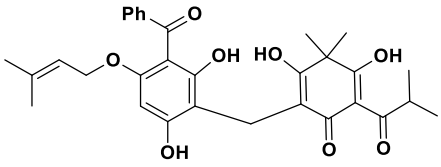


**39**

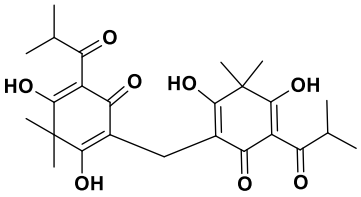


**40**: R= $CH(CH_3)_2$

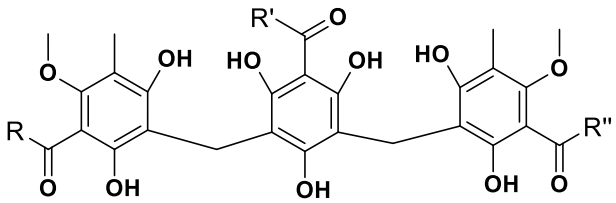
**41**: R = Ph



**42**



**43**





© University of Venda

**44**: R = R″ = $CH(CH_3)_2$, R′ = $CH(CH_3)CH_2CH_3$

**45**: R = R″ = $CH_2CH_2CH_3$, R′ = $CH(CH_3)CH_2CH_3$

**46**: R = R′ = R″ = $CH_2CH_2CH_3$

**47**: R = $CH(CH_3)_2$, R′ = $CH(CH_3)CH_2CH_3$, R″ = $CH_3$

**48**: R = R″ = $CH_3$, R' = $CH(CH_3)CH_2CH_3$

**49**: R = $CH_3$, R' = R″ = $CH_2CH_2CH_3$

**50**: R = R' = R″ = $CH(CH_3)_2$

---

**Figure 2.17. Molecular structures of acylphloroglucinols for which antimalarial activity is reported chiefly.**

Dimeric ACPLs **35-43** are ACPLs that have two acylphloroglucinol moieties linked with a methylene bridge. Bis(2,4-diacetylphloroglucyl)-methane (**35**), an antibiotic and antimalarial compound, is of microbial origin and contains additional acyl groups at C5 and C3′. Its isolation was from the culture fluid of *Pseudomonas aurantiaca* [176]. Robustaol A (**36**), exhibiting antagonistic activity against *Plasmodium berghei* [203], was isolated from the leaves of Myrtaceae's *Eucalyptus robusta Sm*. In Chinese traditional medicine, this plant prepares remedies for malaria treatment.

From *Hypericum japonicum* (a plant that can be found across Asia, Oceania, and North America [204, 205]), several compounds have been extracted and utilized in traditional Chinese medicine. Sarothralens B **(37,** japonicine B**)**, C **(38)**, D **(39)**, A **(40)** and G **(41),** alongside sarothralin (**42,** japonicine C) and japonicine A **(43)** [204-208] have been isolated from this plant and are believed to treat tumors, relieve fever or internal heat, induce haemostasis and detumescence. These are not the only useful medicinal properties of *Hypericum japonicum*, as it has a diverse range of compounds in varying chemical classes [186]; a computational study has determined the computable molecular properties of numerous dimeric ACPLs. The study included compounds **33**, **34**, **35**, **38**, and **41**, and findings on their conformational preferences, computed energetics, IHB parameters, and other molecular properties can be found in [186]. These properties can serve as descriptors for QSAR studies. The compounds outlined in [186] were labeled with systematic acronyms, and here are the acronyms that pertain to compounds **33**, **34**, **35**, **38**, and **41**: D2-B5,5′, D11-B5-M5′-ET6′, D5-KT2′-pyr,5,6, D5- KT2′-M,5′,5′-J5 and D5-KT2,2′-M,5,5,5′,5′. These acronyms can help quickly identify the compounds when accessing them [186].



© University of Venda

Agrimols A-G (**44-50**) are trimeric ACPLs typically found in *Agrimonia pilosa*. This plant is widespread in China and commonly utilized in traditional Chinese medicine to cure various illnesses [209, 210]. These ACPLs have exhibited potent antimalarial activity [176]. Agrimol B (**45**) is also included in a recipe designed to encourage cancer cells to enter a dormant G0 state, preventing the recurrence of the disease if those cancer cells remain in the G0 state. Proliferation is prevented by cancer cells that remain in this state. The computational study analyzed the molecular characteristics of many trimeric ACPLs and their dimeric counterparts. The compounds **44-50**, denoted by the acronyms T14-M5,3"-ET6,2", T13-M5,3"-ET6,2", T9-M5,3"-ET6,2", T11-M5,3"-ET6,2", T3-M5,3"-ET6,2", T7-M5,3"-ET6,2", and T9-ET6,2", were included in the study [211].

#### 2.5.2.4. Acylphloroglucinols for which anticancer activity is chiefly reported

Extensive investigations into the mechanisms of action for ACPLs, which have great potential for use in anticancer applications, have led to their recognition. The significance of carbonyl groups in the active molecules of phloroglucinol derivatives, known for their potential anti-tumor properties, is emphasized in a QSAR analysis. These carbonylic oxygens are crucial for the activity of the molecules due to their ability to bind with biological targets, a significant feature for ACPLs because of the acyl group present. However, studies suggest that charge transfer is unlikely to affect the interaction between ACPLs and target molecules.

Compounds **51–119** (Fig. 2.18–20) are associated with anticancer benefits. Polycyclic polyprenylated ACPLs (PCP-ACPLs) make up most compounds 51–58 (Fig. 2.18) and are commonly found in various plants. Several PCP-ACPLs have shown promise in combating cancer [212], leading to their growing popularity. Meanwhile, compounds **59-62** are P-ACPLs. Hyperatomarin (**51**), a bicyclic PCP-ACPL, comes from the aerial components of *Hypericum annulatum Moris subsp*. Annulatum, having two enol and keto tautomeric forms that exist in equilibrium [212, 213]. Certain tumour cell lines are extremely vulnerable to the strong cytotoxic effects of this compound, with $IC_{50}$ values varying from 0.14 to 15.7 μM [212, 214]. The performance was comparable to the established anti-cancer drug daunorubicin, with some cell lines

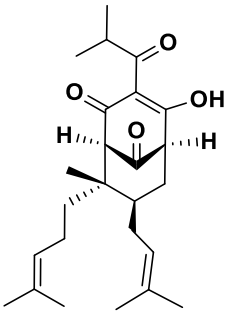


© University of Venda

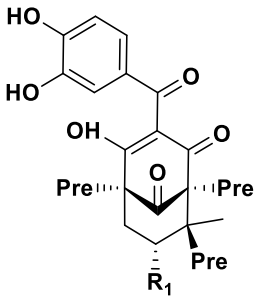
being outclassed [214]. The procedure for cellular death via apoptosis has been triggered due to its cytotoxic effects. One of the suggested mechanisms behind the activation of apoptosis is caspase activation, as suggested by recent testing [214].



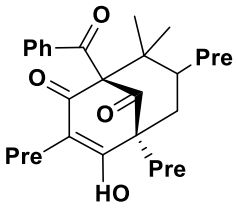
© University of Venda



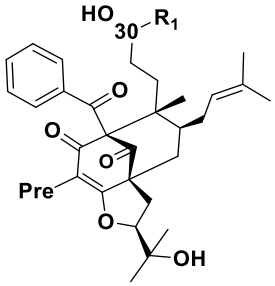
**51**



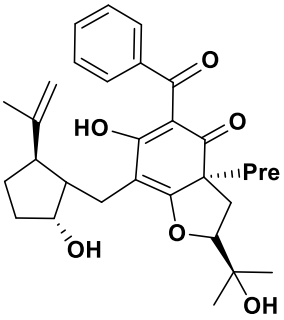
**52**: $R_1$ = geranyl
**53**: $R_1$ = prenyl



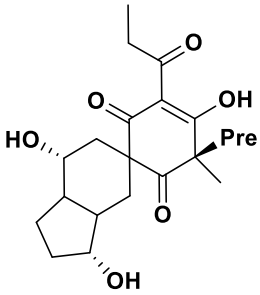
**54**



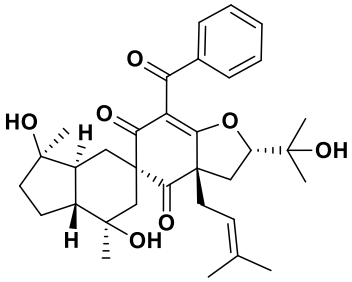
**55**: $R_1$ = $C(CH_2)CH_3$, 30S
**56**: $R_1$ = $C(CH_2)CH_3$, 30R
**57**: $R_1$ = $C(CH_3)_2OH$, 30S
**58**: $R_1$ = $C(CH_3)_2OH$, 30R



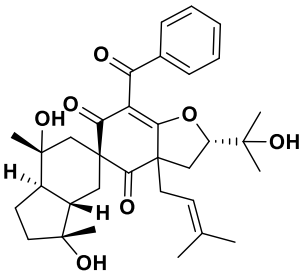
**59**



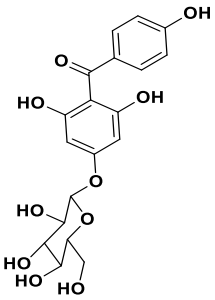
**60**



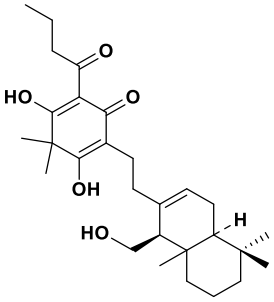
**61**



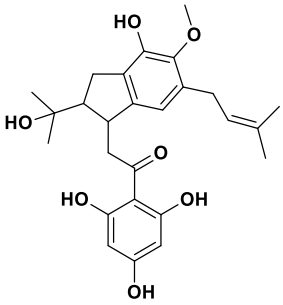
**62**



**63**



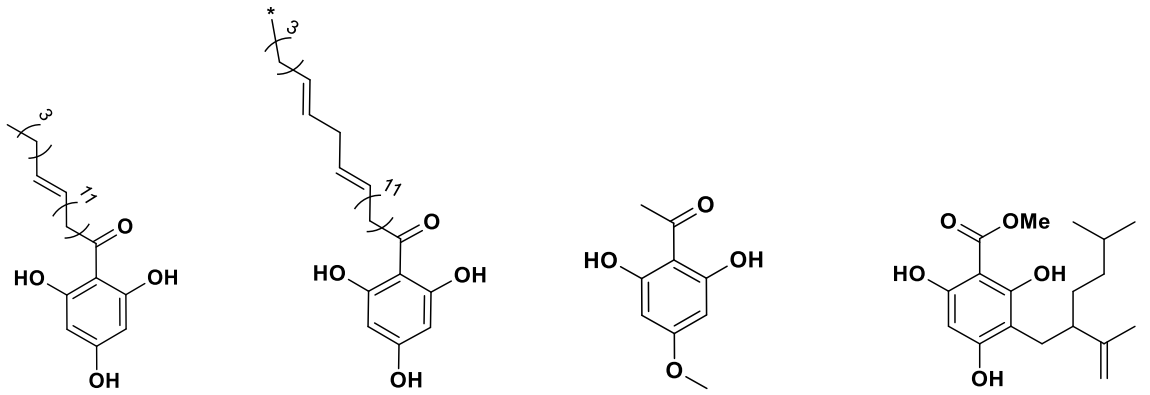
**64**



**65**



© University of Venda




66 67 68 69

**Figure 2.18. Molecular structures of polyprenylated, prenylated and simple acylphloroglucinols with anticancer activity.**

Garcinia species' edible fruits contain two bicyclic ACPLs known as oblongifolin C (**52**, Ob-C) and guttiferone K (53, Gt-K) [215]. The only difference between their molecules is C4, where a prenyl chain exists in **52** and a geranyl chain in 53 [215]. These two compounds are effective against colorectal cancer cells and reduce cancer cell proliferation in vitro, as well as tumor growth and metastasis in vivo [215, 216]. They cause cancer cells to die by inducing apoptosis and inhibiting autophagy flux [215, 216]. When used together, they exhibit significant synergism that enhances cell apoptosis, inhibits autophagy flux, and promotes the generation of ROS in HCT116 colon cancer cells, which may further encourage cell death through apoptosis and autophagy [216].

The analysis of their activity at the molecular level [215, 217] investigated their interactions with the HSPA8 heat shock protein. HSPA8 acts as a molecular chaperone in autophagy and in protein quality control, import, and immunity [215, 218]. Both compounds can strongly bind to HSPA8. However, compound **53** can form a more stable protein-drug complex, with an interaction energy of -98.1 kcal $mol^{-1}$kcal $mol^{-1}$ (-86.7 kcal $mol^{-1}$ for compound **52**). The same trend was observed for the enol tautomers of the compounds, with calculated interaction energy of -104.4 kcal $mol^{-1}$ for **53** (enol) and -96.3 kcal $mol^{-1}$ for **52** (enol). The two compounds interact with the same hydrophobic cleft found in HSPA8, but there are differences in the way they interact. Compound **53** is completely buried in the groove delimited by two alpha-helices, while the longer geranyl chain of compound **52** sticks out of the groove and points toward the exterior of the protein surface



© University of Venda

[215]. The drug-protein interactions discovered include van der Waals (VDW) contacts, conventional H-bonds, and H-bonds with a π bond as acceptor and H-bonds with C–H as donor [215].

Nemorosone (**54**) is an isolated PP-ACPL found in brown *Cuban propolis* [219, 220]. It's in vitro action against various human cancer cell lines [220-222] has attracted much attention. Moreover, it exhibits antiestrogenic activity, reducing cell proliferation caused by 17-β-estradiol (E2) [219]. Hyperpatulones A–F (**55–60**) are peroxide PP-ACPLs obtained from *Hypericum patulum* (*Guttiferae*), a plant that grows extensively in Southwest China and is traditionally utilized to decrease heat, cool blood, relax tendons, and treat anxiety and bacterial and viral infections [223, 224]. These peroxide PP-ACPLs are noted for their cytotoxic activity, with hyperpatulone E being particularly effective against HepG-2, HeLa, and A549 cell lines ($IC_{50}$ values of 9.52, 11.87, and 12.63 μM respectively) [223, 224]. Sampsonols A (**61**) and B (**62**) are among other PP-ACPLs isolated from the aerial portions of *Hypericum sampsonii* [225]. These compounds display significant cytotoxicity against four human tumor cell lines, with IC50 values ranging from 13 to 28 μM [225].

The ACPLs, compounds **63-69**, can be classified as simple due to their relatively simple chemical structure. Iriflophenone 4-glucoside (**63**) is a phloroglucinol-glucoside found in the rhizomes of *Davallia solida* [226]. According to studies, it has a moderate binding affinity to the C-terminal cytosolic domain of P-glycoprotein, a major pharmacological target in treating multidrug-resistant cancer. This protein is a crucial pharmacological target for treating cancer that has become multidrug-resistant, as it promotes resistance by expelling drugs from the cell, thus limiting their efficacy. The binding affinity increases by 6- to 10-fold for the aglycone derivative of iriflophenone 4-glucoside and other related compounds, as per studies [226].

Dryofragin (**64**) is a compound that has been isolated from the plant Dryopteris fragrans (L.) Schott, and it is structurally connected to filicinic acid [227]. According to reports, Dryofragin induces apoptosis in human breast cancer MCF-7 cells through a ROS-mediated mitochondrial pathway [227]. The compound has two phenolic OH groups, where O8H15 is switched out with a keto O, and these two groups can have different effects in regulating apoptosis [176]. Dryofragin



© University of Venda

has demonstrated its ability to inhibit the growth of MCF-7 cells over a 24-, 48-, and 72-hour period, with $IC_{50}$ values of 76.10 µM, 37.51 µM, and 27.26 µM respectively [176].

Antiarone J (**65**) was isolated from the root bark of *Antiaris toxicaria* using methanol extraction [227]. The ACPL moiety of the compound contains a prenylated indane group, and this moiety plays a role in the acyl group's R. The compound has cytotoxic effects that target colon 38 and L1210 strains [227, 228]. In dried fruits of *Protorhus thouvenotii*, two compounds, Thouvenols A (**66**) and B (**67**), were isolated [227]. Both compounds were shown to have cytotoxicity in vitro against the A2780 ovarian cancer cell line with $IC_{50}$ of 11 µg $mL^{-1}$ [227].

Compound **68**, isolated from *Mallotus japonicus*, is one of the simplest ACPL structures. It has a methyl group replacing H16, and R is a $CH_3$ group. The compound has exhibited cytotoxic properties [227].

*Sophora flavescent*, a plant widely found in China, Japan, Korea, Russia, and India, has been traditionally employed in Chinese medicine for its medicinal properties to treat health conditions such as pruritus, dysentery, trichomonas vaginitis, eczema, and pyogenic skin infections [229-231]. Flavescenol A (**69**), obtained from the plant's roots, has been identified as an ACPL that is lavandulylated, marking the first instances of this type of compound. Flavescenol A displays moderate cytotoxicity against HepG2 cell lines, with an $IC_{50}$ of 29.74 µM, while its related compound, flavescenol B, is inactive [231].

The Euglobals (EUs) are predominantly found in plants belonging to the myrtaceae family, and so far, 31 EUs have been identified [176]. The molecular structures of different EUs are shown in Figure 2.19, with each EU reported to possess only anticancer properties. These EUs include euglobal III (**70**), Ia1 (**71**), Ia2 (**72**), IIc (**73**), T1 (**74**), Ib (**75**), BI1 (76), Ic (**77**), IIa (**78**), IIb (**79**), and G1–G12 (**80–91**), IVa (**92**), IVb (**93**), VII (**94**), V (**95**), In-1 (**96**), In-2 (**97**), and In-3 (**98**) [176]. The remaining two EUs, 30 and 31, have been discussed in section 2.5.1, as they exhibit both antimalarial and anticancer activity. In terms of molecular composition, EUs can be viewed as adducts of chroman ring formation involving acylphloroglucinol monoterpene or sesquiterpene. The ACPL moiety in EUs contains two acyl groups, with R being either an isobutyl group or an H



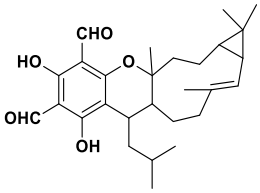
© University of Venda

atom. Notably, the same molecule can not possess two isobutyl groups, although it is possible for two H atoms to be present simultaneously. As such, there exist three potential structure types for the acyl phloroglucinol moiety in conjunction with a given EU.

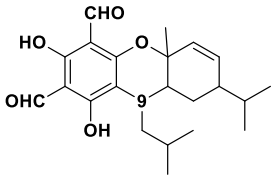
The results of computational studies [232, 233] indicate that the reversal of acyl group positions, such as those seen in molecules **73** and **74** or 80 and 81, where R equals H and isobutyl, can have a significant impact on several molecular properties. These properties include relative energies, characteristics of the IHBs, HOMO-LUMO energy gap, dipole moment, and IR vibrational frequencies, as well as frequency changes due to the IHBs. Recently, EUs have garnered attention for their ability to potently inhibit the activation of the Epstein-Barr virus [234]. All the EUs listed previously have demonstrated this activity and may play a significant role in cancer prevention, as infection by this virus increases the risk of certain cancer types. EU-G1 has a stronger inhibitory effect than glycyrrhetic acid, a well-known anti-tumor promoter, while EU-III exhibits similar effects to glycyrrhetic acid [234, 235, 236]. Additionally, tests [235] have shown that the ACPL moiety is essential for their inhibition activity, as is the presence of two acyl groups. EUs have been found to exhibit several biological activities, including but not limited to antileishmanial, antimalarial, antibacterial, antitrypanosomal, and granulation inhibition [176]. Recently, the association of the Epstein-Barr virus with multiple sclerosis has been confirmed, leading to the possibility of new roles for EUs [237].



© University of Venda

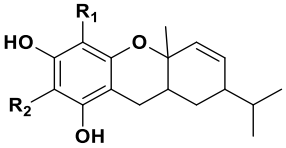



**70**



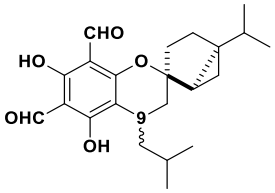

**71**: 9β

**72**: 9α



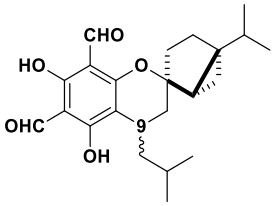

**73**: $R_1$ = CHO, $R_2$ = $COCH_2CH(CH_3)_2$

**74**: $R_1$ = $COCH_2CH(CH_3)_2$, $R_2$ =CHO



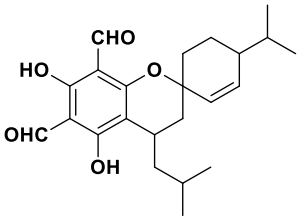

**75**: 9β

**76**: 9α

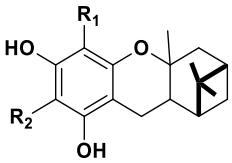



**77**: 9α

**78**: 9β

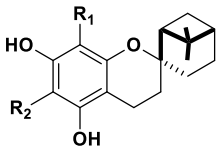



**79**



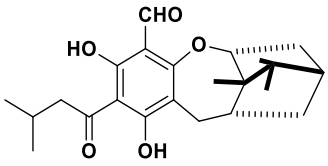

**80**: $R_1$ = $COCH_2CH(CH_3)_2$, $R_2$ =CHO

**81**: $R_1$ = CHO, $R_2$ = $COCH_2CH(CH_3)_2$

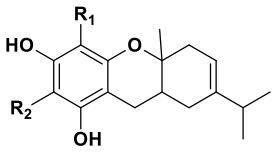



**82**: $R_1$ = CHO, $R_2$ = $COCH_2CH(CH_3)_2$

**83**: $R_1$ = $COCH_2CH(CH_3)_2$, $R_2$ =CHO

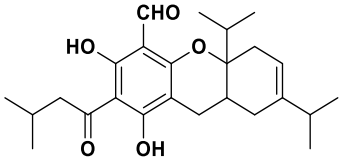



**84**

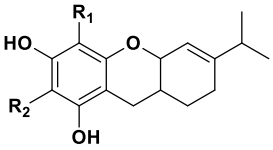



**85**: $R_1$ = CHO, $R_2$ = $COCH_2CH(CH_3)_2$

**86**: $R_1$ = $COCH_2CH(CH_3)_2$, $R_2$ =CHO

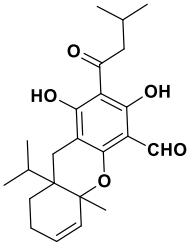



**87**




**88**: $R_1$ = CHO, $R_2$ = $COCH_2CH(CH_3)_2$

**89**: $R_1$ = $COCH_2CH(CH_3)_2$, $R_2$ =CHO




**90**



© University of Venda

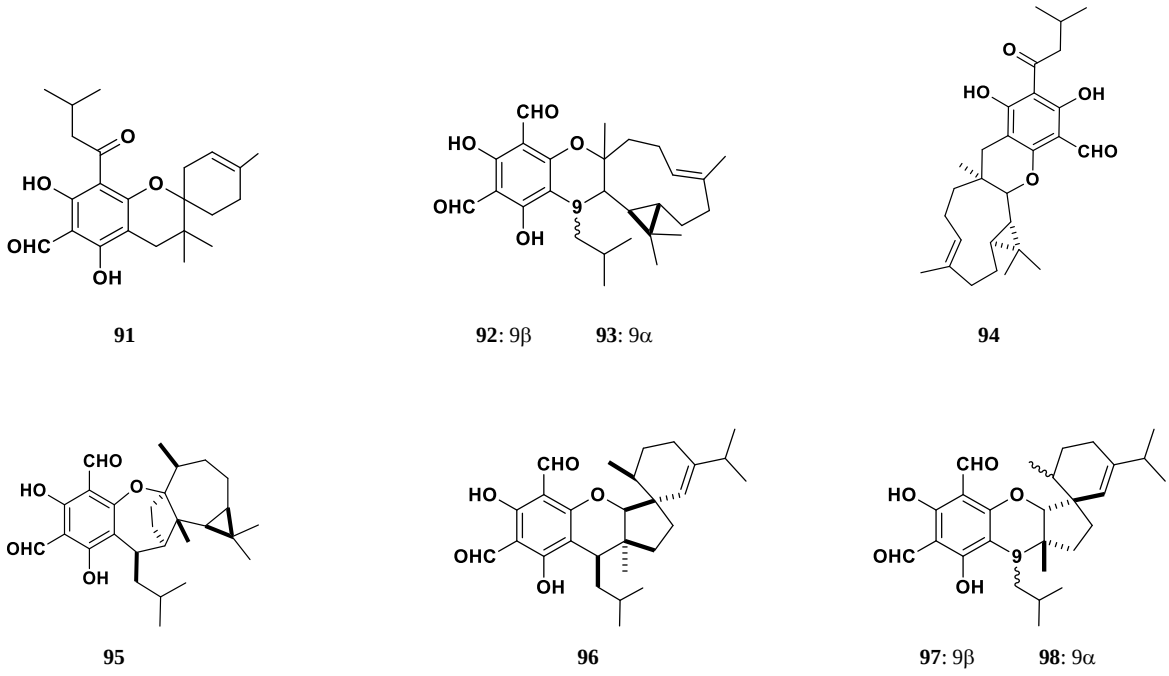



**Figure 2.19. Molecular structures of euglobals.**

Dimeric ACPLs, specifically compounds **99-126** (Fig. 2.20), have been identified as having a benzopyran ring skeleton, with drummondins A-F (**99-104**) being isolated from *Hypericum drummondii*. Drummondins E and F have a prenyl chain attached to one of the monomeric units. These compounds exhibit cytotoxic activity when tested on cultured P-388, KB, and various human cancer cell lines such as breast, colon, lung, and melanoma [238-240]. They also possess antimicrobial activity comparable to, or even greater than, that of streptomycin [176].

A plant commonly found in Japan, *Mallotus japonicus (Euphorbiaceae*), has bark traditionally used in treating ulcers and cancer and leaves used for boils [241]. Mallotojaponin (105) and its analogues (106–122) were isolated from this plant. These compounds exhibit cytotoxic properties against KB and L-51784 cell lines in cultures, moderately inhibit the growth of Ehrlich carcinoma in mice, and exhibit antiherpetic properties [241-247]. Mallotojaponin had strong cytotoxicity against all tumor cells, leading to a marked life-span prolongation in mice with L51784 leukemia [176].

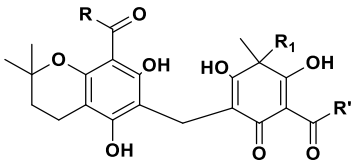


© University of Venda

Prenylated dimeric ACPLs, specifically Acrovestone (**123**) and demethylacrovestone (**124)**, were unearthed from *Acronychia pedunculata* and *Acronychia laurifolia* barks [248, 249]. Diarrhoea, ulcers, cough, asthma, itchiness, rheumatism, and other indispositions have been treated through the utilization of Acronychia pedunculate, which is widely present in areas such as Southern China and Indo-Malayan [248]. In terms of cytotoxicity, both compounds have demonstrated activity. Acrovestone, in particular, has been shown to possess potent cytotoxicity, with 100% inhibition at 0.5 µg $mL^{-1}$ in a KB tissue culture assay. Moreover, it exhibited significant cytotoxicity against A-549, P-388, and L-1210 cells, with ED50s of 0.98, 3.28, and 2.95 µg mL−1, respectively [248-250], making it a promising candidate for further study.



© University of Venda

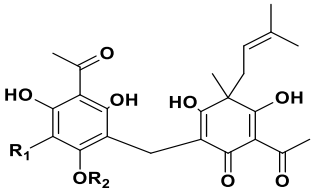


**99**: R = R′ = $CH_2CH_3$, $R_1$ = $CH_3$

**100**: R = $CH_3$, $R_1$= $CH_2CH_3$, $R_1$= $CH_3$

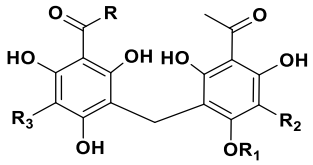
**101**: R = R′= $R_1$= $CH_3$

**102**: R = R′= $CH_3$, $R_1$= prenyl



**103**: $R_1$= $OCH_2CHC(CH_3)_2$, $R_2$ = H

**104**: $R_1$ = H, $R_2$= prenyl



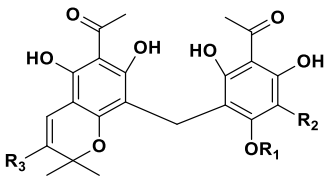
**105**: R = $R_1$ = $R_2$ = $CH_3$, $R_3$ = prenyl

**106**: R = $CH_2CH_2CH_3$, $R_1$ = $R_2$ = $CH_3$, $R_3$ = prenyl

**107**: R = $CH(CH_3)_2$, $R_1$ = $R_2$ = $CH_3$, $R_3$ = prenyl

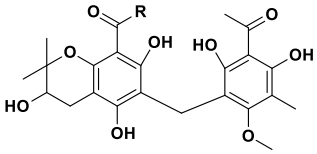
**108**: R = $CH_3$, $R_1$ = H, $R_2$ = $OCH_3$, $R_3$ =$CH_2CHOHC(CH_2)CH_3$

**109**: R = $CH_2CH_2CH_3$, $R_1$ = H, $R_2$ = $OCH_3$, $R_3$ = $CH_2CHOHC(CH_2)CH_3$



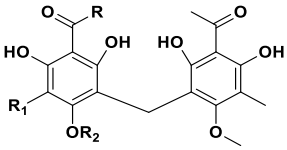
**110**: $R_1$= H, $R_2$ = $OCH_3$, $R_3$= H

**111**: $R_1$ = $OCH_3$, $R_2$ = $CH_3$, $R_3$ = OH



**112**: R = $CH(CH_3)_2$

**113**: R = $CH_3$

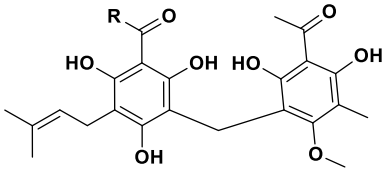


**114**: R = $CH_3$, $R_1$ = $CH_2CHOHC(CH_3)_3$, $R_2$ =H

**115**: R = $R_1$ = $R_2$ = $CH_3$

**116**: R = $CH(CH_3)_2$ $R_1$ = $CH_2CHOHC(CH_2)CH_3$, $R_2$ = H

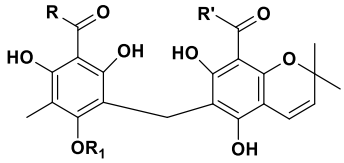
**117**: R = $CH_3$, $R_1$ = $CH_2CHOHC(CH_2)CH_3$, $R_2$ = H



**118**: R = $CH_2CH_2CH_3$

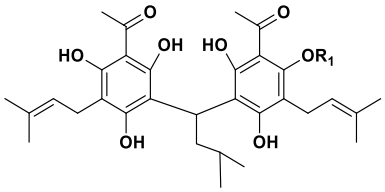
**119**: R = $CH_3$

**120**: R = $R_2$ = $CH_3$, $R_1$ = $CH(CH_3)_2$



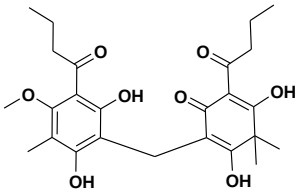
**121**: R = R′ = $R_1$ = $CH_3$

**122**: R = $R_1$= $CH_3$, R′ = $CH_2CH_2CH_3$

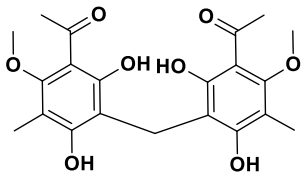


**123**: $R_1$ = H

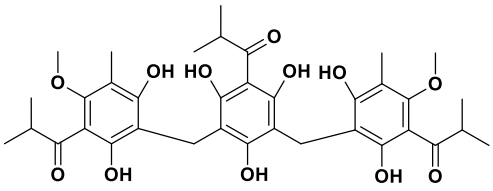
**124**: $R_1$ = $CH_3$



**125**



**126**



**127**

**Figure 2.20. Molecular structures of several dimeric and one trimeric acylphloroglucinols having anticancer activity.**



© University of Venda

Aspidin BB (**125**) isolated from *Dryopteris fragrans (L.) Schott*, exhibits strong inhibition of human ovarian cancer cell line (HO-8910) growth at 24, 48, and 72 hours with IC50 values of 68.81 lM, 25.79 lM, and 15.02 lM, respectively [251]. Cell cycle arrest in the S phase and apoptosis induction through the mitochondrial pathway also occur in these cells [251]. Traditional medicinal practices in Ethiopia use the female flowers of Hagenia abyssinica to treat parasitic worm infections [176]. From these flowers, a dimeric ACPL known as Kosin (**126**) and a trimeric ACPL known as Protokosin (**127**, Fig. 2.20) have been isolated [252, 253]. In vitro and in vivo studies show that both compounds exhibited cytotoxic properties against transplantable murine colonic adenocarcinomas with varying growth characteristics and morphology. Significantly decreased colony formation was also observed in vitro [252, 253]. It's important to note that compounds with potent antioxidant activity are crucial for cancer prevention because they manage the excessive formation of radical species (reactive oxygen species, ROS).

#### 2.5.2.5. Synthetic active acylphloroglucinols based on naturally occurring ones – two examples.

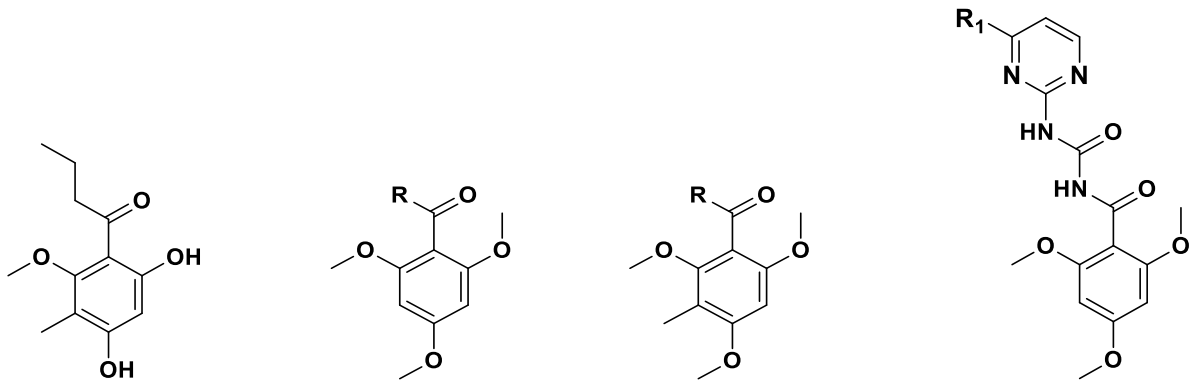
Designing new potent molecular structures mostly starts with naturally occurring compounds with pharmacological properties. Fundamental knowledge of a molecule's specific parts responsible for a certain activity is crucial for creating active molecules.

The active group in pseudoaspidinol (**128**, Fig. 2.21) has been confirmed to be the acyl group through numerous studies. To confirm this further, a range of derivatives were synthesized and analyzed. The derivatives varied in the R chain of the acyl group and replaced all OHs of the phloroglucinol moiety with methoxy groups [254-256], which prevents the formation of the first IHB. The study used acronyms and progressive numbers to denote them, as utilized in Figure 2.12. Aspidinol and pseudoaspidinol are related structures in two series, A2-A9, and B2-B9, respectively. The former lacks substituents at C3 and C4, while the latter has a methyl at C3. In addition, structures C1-C3 (**145-147**) were created by substituting R with an aminopyrimidine unit, a common feature of tyrosine kinase inhibitors. These other derivatives of aspidinol were also synthesized [257]. These synthesized compounds were tested in vitro for their effects on four human cancer cell lines (A-549, MCF-7, HeLa, and HepG2) using the MTT assay [257]. Each compound showed certain antiproliferative effects on these cell lines, with significant inhibition



© University of Venda

observed in the MCF-7 line. Compound C2 stood out with an IC50 value of 18.49 μM, surpassing the standard chemotherapeutic agent, 5-fluorouracil. Additionally, cell apoptosis assays indicated that C2 induced apoptosis in a concentration-dependent way. Molecular docking analyses further illustrated the likely interactions between the active compounds and target proteins 4I22 and 1OG5 binding sites.



**128**

**129**: R=H
**130**: R=$CH_3$
**131**: R=$CH_2CH_3$
**132**: R=$CH_2CH_2CH_3$
**133**. R=$CH(CH_3)_2$
**134**: R=$CH_2CH_2CH_2CH_3$
**135**: R=$CH_2CH(CH_3)_2$
**136**: R=$C(CH_3)_3$

**137**: R=H
**138**: R=$CH_3$
**139**: R=$CH_2CH_3$
**140**: R=$CH_2CH_2CH_3$
**141:** R=$CH(CH_3)_2$
**142**: R=$CH_2CH_2CH_2CH_3$
**143**: R=$CH_2CH(CH_3)_2$
**144**: R=$C(CH_3)_3$

**145**: $R_1$=H
**146**: $R_1$=$CH_3$
**147**: $R_1$=Cl

**Figure 2.21. Molecular structures of pseudoaspidinol (128) and synthetic acylphloroglucinols are based on it and have anticancer activity.**

Two proteins, the epidermal growth factor receptor (EGFR) and the cytochrome P4502C9 enzyme (CYP2C9), with binding sites Met793 and Phe476, respectively [258, 259], have been the focus of a molecular docking study due to their high expression in breast cancer. The study found that the synthesized APCLs and afatinib all possess the ability to bind strongly to Met793 and Phe476, with the C1-C3 compounds showing greater binding energy (−6.6959, −6.8147, and −6.8208 kcal $mol^{-1}$ for EGFR, and −7.1243, −7.6265, and −7.1304 kcal $mol^{-1}$ for CYP2C9) [257]. The data



© University of Venda

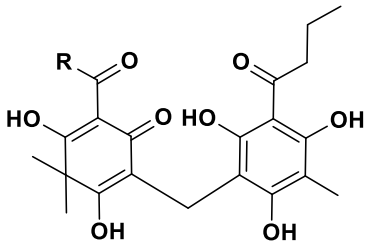
suggests that compounds C1-C3 may act by inhibiting the functional polymorphism of CYP2C9, thus regulating and intervening in its metabolic process to inhibit excessive proliferation and restore normal cell apoptosis [257]. These findings also provide insights into the nature of drug-protein interactions. The data obtained from the study also showed insights into the characteristics of the protein-drug interactions. In the case of protein EGFR, C2 interacts with it by forming hydrogen bonds between the amide group's oxygen and nitrogen atoms and Met793 and conjugating the methylpyrimidine group with Leu718. Meanwhile, in the interaction between C2 and protein 1OG5, the aminopyrimidine group's nitrogen atoms form hydrogen bonds with Ser365, and there is a π-π interaction between the methylpyrimidine group and Phe476 [257].

The synthesis of two dimeric ACPLs, A3 (**148**) and A5 (**149**) (Fig. 2.22), with the potential to treat cancer, was done by utilizing the scaffolds of flavaspidic acid AB and albaspidin AA (which were originally extracted from Dryopteris fragrans (L.) Schott). In [260], the antitumor activity results indicate that compound A5 showed the strongest effect against A549 cells, while compound A3 was most effective against HepG2 and MCF-7 cells, demonstrating a concentration-dependent response that surpassed the efficacy of the positive control, 5-fluorouracil (5-FU). Additionally, both compounds A3 and A5 exhibited lower cytotoxicity toward normal cells, highlighting their potential as selective anticancer agents. A number of proteins, including Bcl-2 (1PQ1,2 W3 L,4MAN), Bcl-w (2Y6 W), Bcl-xl (2YXJ), Mcl-1 (5FDO,3KZO), Caspase-9 (1JXQ), Caspase-7 (4FDL), Caspase-3 (1SHJ), and fatty acid synthase (2PX6), were docked with these two compounds [260]. The results indicate that **148** have greater binding energy with Bcl-xl, while **149** have greater binding energy with Mcl-1. In the binding site of Bcl-xl, **148** forms H-bond interactions with GLY-138, TYR-195, and ASN-197, and A5 forms H-bonds with GLY-138, TYR-195, and GLU-96 of 2YXJ. With regards to binding energy, **148** demonstrated a higher affinity with Bcl-xl, whereas Mcl-1 exhibited greater affinity with **149**. Interestingly, the results indicated that GLY-138, TYR-195, and ASN-197 were the points of H-bond contact in the binding site of Bcl-xl for **148**. Conversely, in the binding site of Bcl-xl, A5 forms H-bonds with GLY-138, TYR-195, and GLU-96 of 2YXJ [260]. The computable molecular properties of monomeric ACPL have been investigated by computational studies as previously described and analyzed in [180]; compounds **128**, **130** (corresponding to compound **62**), **132**, **133**, **138,** and **140** were included in this study and are mentioned in [180] by the following abbreviations respectively: E1, BB-ET4,



© University of Venda

EE1-ET4, EE2 - ET4, B-ET4, and E1-ET4. Many derivatives of naturally occurring bicyclic polyprenylated ACPL have been analyzed.




**148**: R=$CH_2CH_2CH_3$

**149**: R=$CH_3CH(CH_3)_2$

**Figure 2.22. Molecular structures of synthetic dimeric acylphloroglucinols with structures based on the scaffolds of flavaspidic acid AB and albaspidin AA and having anticancer activity**.



© University of Venda

## 2.6. References

1. American Association for Cancer Research. (n.d.). Patients, caregivers, and advocates. Retrieved from https://www.aacr.org/patients-caregivers/ (Accessed: October 1, 2023)
2. Weinberg, R. A. (1996). How cancer arises. *Scientific American*, 275(3), 62-70.
3. Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, 144(5), 646-674.
4. Garraway, L. A. (2013). Genomics-driven oncology: Framework for an emerging paradigm. *Journal of Clinical Oncology*, 31(15), 1806-1814.
5. Stanford Health Care. (n.d.). What causes cancer? Retrieved from https://stanfordhealthcare.org/medical-conditions/cancer/cancer/cancer-causes.html (Accessed: October 5, 2023)
6. Abbott, A. (2016). Scientists bust myth that our bodies have more bacteria than human cells. *Nature*, 10.
7. De Groot, J. F., Fuller, G., Kumar, A. J., Piao, Y., Eterovic, K., Ji, Y., & Conrad, C. A. (2010). Tumor invasion after treatment of glioblastoma with bevacizumab: Radiographic and pathologic correlation in humans and mice. *Neuro-Oncology*, 12(3), 233-242.
8. Stasevich, E. M., Murashko, M. M., Zinevich, L. S., Demin, D. E., & Schwartz, A. M. (2021). The role of non-coding RNAs in the regulation of the proto-oncogene MYC in different types of cancer. *Biomedicines*, 9(8), 921.
9. Rickman, D. S., Schulte, J. H., & Eilers, M. (2018). The expanding world of N-MYC–driven tumors. *Cancer Discovery*, 8(2), 150-163.
10. Kumimoto, H., Hamajima, N., Nishimoto, Y., Matsuo, K., Shinoda, M., Hatooka, S., & Ishizaki, K. (2002). L-myc genotype is associated with different susceptibility to lung cancer in smokers. *Japanese Journal of Cancer Research*, 93(1), 1-5.
11. Samowitz, W. S., Slattery, M. L., Sweeney, C., Herrick, J., Wolff, R. K., & Albertsen, H. (2007). APC mutations and other genetic and epigenetic changes in colon cancer. *Molecular Cancer Research*, 5(2), 165-170.
12. Hruban, R. H., Offerhaus, G. J. A., Kern, S. E., Goggins, M., Wilentz, R. E., & Yeo, C. J. (1998). Tumor-suppressor genes in pancreatic cancer. *Journal of Hepato-Biliary-Pancreatic Surgery*, 5, 383-391.



© University of Venda

13. Roy, I., & Ghosh, G. (2021). Aberrant signaling pathways in cancer cells: Application of nanomaterials. *Journal of Cellular Signaling*, 2(4), 281-299.
14. Sherr, C. J. (1996). Cancer cell cycles. *Science*, 274(5293), 1672-1677.
15. Uchida, C. (2012). The retinoblastoma protein: Functions beyond the G1-S regulator. *Current Drug Targets*, 13(13), 1622-1632.
16. Benchimol, S. (2001). p53-dependent pathways of apoptosis. *Cell Death and Differentiation*, 8(11), 1049-1051.
17. Pritchard-Jones, K. (1999). The Wilms tumor gene, WT1, in normal and abnormal nephrogenesis. Pediatric Nephrology, 13, 620-625.
18. Brown, M. A. (1997). Tumor suppressor genes. *Advances in Genetics*, 36, 45.
19. Kumar, A., Kumari, N., Gupta, V., & Prasad, R. (2018). Renal cell carcinoma: Molecular aspects. *Indian Journal of Clinical Biochemistry*, 33, 246-254.
20. Espinosa, E., Zamora, P., Feliu, J., & Barón, M. G. (2003). Classification of anticancer drugs—a new system based on therapeutic targets. *Cancer Treatment Reviews*, 29(6), 515-523.
21. Meegan, M. J., & O'Boyle, N. M. (2019). Special issue "anticancer drugs." *Pharmaceuticals*, 12(3), 13.
22. Zhong, L., Li, Y., Xiong, L., Wang, W., Wu, M., Yuan, T., Yang, W., Tian, C., Miao, Z., Wang, T., & Yang, S. (2021). Small molecules in targeted cancer therapy: Advances, challenges, and future perspectives. *Signal Transduction and Targeted Therapy*, 6(1), 1-48.
23. Pratt, W. B. (1994). The anticancer drugs. Oxford University Press, USA.
24. Baudino, A. T. (2015). Targeted cancer therapy: The next generation of cancer treatment. *Current Drug Discovery Technologies*, 12(1), 3-20.
25. Begg, A. C., Stewart, F. A., & Vens, C. (2011). Strategies to improve radiotherapy with targeted drugs. *Nature Reviews Cancer*, 11(4), 239-253.
26. Brana, M. F., Cacho, M., Gradillas, A., Pascual-Teresa, B. D., & Ramos, A. (2001). Intercalators as anticancer drugs. *Current Pharmaceutical Design*, 7(17), 1745-1780.
27. Avendaño, C., & Menendez, J. C. (2015). *Medicinal chemistry of anticancer drugs*. Elsevier.
28. Nussbaumer, S., Bonnabry, P., Veuthey, J. L., & Fleury-Souverain, S. (2011). Analysis of anticancer drugs: a review. *Talanta*, 85(5), 2265-2289.
29. Cummings, J., & Smyth, J. F. (1993). DNA topoisomerase I and II as targets for rational design of new anticancer drugs. *Annals of Oncology*, 4(7), 533-543.



© University of Venda

30. Peters, G. J. (2014). Novel developments in the use of antimetabolites. *Nucleosides, Nucleotides and Nucleic Acids*, 33(4-6), 358-374.
31. Kaye, S. B. (1998). New antimetabolites in cancer chemotherapy and their clinical impact. *British Journal of Cancer*, 78(3), 1-7.
32. Larsen, A. K., Galmarini, C. M., & D'Incalci, M. (2016). Unique features of trabectedin mechanism of action. *Cancer Chemotherapy and Pharmacology*, 77, 663-671.
33. Pommier, Y. (2006). Topoisomerase I inhibitors: Camptothecins and beyond. *Nature Reviews Cancer*, 6(10), 789-802.
34. What is VEGF? - News-Medical.net. (n.d.). Retrieved from https://www.news-medical.net/life-sciences/What-is-VEGF.aspx (Accessed: October 7, 2023)
35. Galderisi, U., Cascino, A., & Giordano, A. (1999). Antisense oligonucleotides as therapeutic agents. *Journal of Cellular Physiology*, 181(2), 251-257.
36. Davis, T. A., Grillo-Lopez, A. J., White, C. A., McLaughlin, P., Czuczman, M. S., Link, B. K., et al. (2000). Rituximab anti-CD20 monoclonal antibody therapy in non-Hodgkin's lymphoma: Safety and efficacy of re-treatment. *Journal of Clinical Oncology*, 18(17), 3135–3143.
37. Vose, J. M., Wahl, R. L., Saleh, M., Rohatiner, A. Z., Knox, S. J., Radford, J. A., et al. (2000). Multicenter phase II study of iodine-131 tositumomab for chemotherapy-relapsed/refractory low-grade and transformed low-grade B-cell non-Hodgkin's lymphomas. *Journal of Clinical Oncology*, 18(6), 1316–1323.
38. Lundin, J., Kimby, E., Bjorkholm, M., Broliden, P. A., Celsing, F., Hjalmar, V., et al. (2002). Phase II trial of subcutaneous anti-CD52 monoclonal antibody alemtuzumab (Campath-1H) as first-line treatment for patients with B-cell chronic lymphocytic leukemia (B-CLL). *Blood*, 100(3), 768–773.
39. Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, 458(7239), 719-724.
40. Tremelimumab granted Orphan Drug Designation by US FDA for treatment of malignant mesothelioma. (n.d.). Retrieved from https://britishreporter.com/tremelimumab-granted-orphan-drug-designation-by-us-fda-for-treatment-of-malignant-mesothelioma/ (Accessed: October 8, 2023)



© University of Venda

41. Wu, J. Y., Shih, J. Y., Yang, C. H., Chen, K. Y., Ho, C. C., Yu, C. J., & Yang, P. C. (2010). Second-line treatments after first-line gefitinib therapy in advanced nonsmall cell lung cancer. *International Journal of Cancer*, 126(1), 247-255.
42. Feng, L., Li, W., Chao, Y., Huan, Q., Lu, F., Yi, W., ... & Shougang, Z. (2021). Synergistic inhibition of renal fibrosis by nintedanib and gefitinib in a murine model of obstructive nephropathy. *Kidney Diseases*, 7(1), 34-49.
43. Bodnar, M., Guerrero, P., Perez-Carrasco, R., & Piotrowska, M. J. (2016). Deterministic and stochastic study for a microscopic angiogenesis model: Applications to the Lewis Lung carcinoma. *PLoS One*, 11(5), e0155553.
44. Traxler, P., Bold, G., Buchdunger, E., Caravatti, G., Furet, P., Manley, P., ... & Zimmermann, J. (2001). Tyrosine kinase inhibitors: From rational design to clinical trials. *Medicinal Research Reviews*, 21(6), 499-512.
45. Paul, M. K., & Mukhopadhyay, A. K. (2004). Tyrosine kinase–role and significance in cancer. *International Journal of Medical Sciences*, 1(2), 101.
46. Vivanco, I., & Sawyers, C. L. (2002). The phosphatidylinositol 3-kinase–AKT pathway in human cancer. *Nature Reviews Cancer*, 2(7), 489-501.
47. Jirousek, M. R., & Goekjian, P. G. (2001). Protein kinase C inhibitors as novel anticancer drugs. *Expert Opinion on Investigational Drugs*, 10(12), 2117-2140.
48. Kong, N., Fotouhi, N., Wovkulich, P. M., & Roberts, J. (2003). Cell cycle inhibitors for the treatment of cancer. *Drugs Future*, 28(9), 881.
49. Janke, C., & Magiera, M. M. (2020). The tubulin code and its role in controlling microtubule properties and functions. *Nature Reviews Molecular Cell Biology*, 21(6), 307-326.
50. Janke, C., & Bulinski, J. C. (2011). Post-translational regulation of the microtubule cytoskeleton: Mechanisms and functions. *Nature Reviews Molecular Cell Biology*, 12(12), 773-786.
51. Gilbar, P.J., & Carrington, C.V. (2006). The incidence of extravasation of vinca alkaloids supplied in syringes or mini-bags. *Journal of Oncology Pharmacy Practice*, 12(2), 113-118.
52. Fojo, T., & Menefee, M. (2007). Mechanisms of multidrug resistance: The potential role of microtubule-stabilizing agents. *Annals of Oncology*, 18, v3-v8.
53. Goodin, S. (2008). Novel cytotoxic agents: Epothilones. *American Journal of Health-System Pharmacy*, 65(10 Supplement 3), S10-S15.



© University of Venda

54. Perez, E.A. (2009). Microtubule inhibitors: Differentiating tubulin-inhibiting agents based on mechanisms of action, clinical activity, and resistance. *Molecular Cancer Therapeutics*, 8(8), 2086-2095.
55. Fumoleau, P., Coudert, B., Isambert, N., & Ferrant, E. (2007). Novel tubulin-targeting agents: Anticancer activity and pharmacologic profile of epothilones and related analogues. *Annals of Oncology*, 18, v9-v15.
56. Vihinen, P., Ala-aho, R., & Kahari, V.M. (2005). Matrix metalloproteinases as therapeutic targets in cancer. Current Cancer Drug Targets, 5(3), 203-220.
57. Monahan-Earley, R., Dvorak, A.M., & Aird, W.C. (2013). Evolutionary origins of the blood vascular system and endothelium. *Journal of Thrombosis and Haemostasis*, 11, 46-66.
58. Zhang, W., Zhang, Y.S., Bakht, S.M., Aleman, J., Shin, S.R., Yue, K., ... & Khademhosseini, A. (2016). Elastomeric free-form blood vessels for interconnecting organs on chip systems. *Lab on a Chip*, 16(9), 1579-1586.
59. Wang, X.Y., Pei, Y., Xie, M., Jin, Z.H., Xiao, Y.S., Wang, Y., ... & Huang, W.H. (2015). An artificial blood vessel implanted three-dimensional microsystem for modeling transvascular migration of tumor cells. *Lab on a Chip*, 15(4), 1178-1187.
60. Fanelli, M., Sarmiento, R., Gattuso, D., Carillio, G., Capaccetti, B., Vacca, A., Roccaro, A.M., & Gasparini, G. (2003). Thalidomide: A new anticancer drug? *Expert Opinion on Investigational Drugs*, 12(7), 1211-1225.
61. Afzal, A., Caballero, S., Palii, S.S., Jurczyk, S., Pardue, M., Geroski, D., Edelhauser, H., Hochhaus, G., Kim, M., Franklin, A., & Shapiro, G. (2010). Targeting retinal and choroid neovascularization using the small molecule inhibitor carboxyamidotriazole. *Brain Research Bulletin*, 81(2-3), 320-326.
62. Zhang, T., Sun, H.C., Zhou, H.Y., Luo, J.T., Zhang, B.L., Wang, P., Wang, L., Qin, L.X., Ren, N., Ye, S.L., & Li, Q. (2010). Interferon alpha inhibits hepatocellular carcinoma growth through inducing apoptosis and interfering with adhesion of tumor endothelial cells. *Cancer Letters*, 290(2), 204-210.
63. Leahy, K.M., Ornberg, R.L., Wang, Y., Zweifel, B.S., Koki, A.T., & Masferrer, J.L. (2002). Cyclooxygenase-2 inhibition by celecoxib reduces proliferation and induces apoptosis in angiogenic endothelial cells in vivo. *Cancer Research*, 62(3), 625-631.



© University of Venda

64. Wang, Z., Fuentes, C.F., & Shapshay, S.M. (2002). Antiangiogenic and chemopreventive activities of celecoxib in oral carcinoma cells. *Laryngoscope*, 112(5), 839-843.
65. Figg, W.D., Kruger, E.A., Price, D.K., Kim, S., & Dahut, W.D. (2002). Inhibition of angiogenesis: Treatment options for patients with metastatic prostate cancer. *Investigational New Drugs*, 20(2), 183-194.
66. Bevacizumab. (2002). Anti-VEGF monoclonal antibody, Avastin, Rhumab-VEGF. *Drugs R D*, 3(1), 28-30.
67. Chen, H.X., Gore-Langton, R.E., & Cheson, B.D. (2001). Clinical trials referral resource: Current clinical trials of the anti-VEGF monoclonal antibody bevacizumab. *Oncology* (Huntington), 15(8), 1017, 1020, 1023-1026.
68. Stopeck, A., Sheldon, M., Vahedian, M., Cropp, G., Gosalia, R., & Hannah, A. (2002). Results of a Phase I dose-escalating study of the antiangiogenic agent, SU5416, in patients with advanced malignancies. *Clinical Cancer Research*, 8(9), 2798-2805.
69. Fabbro, D., & Manley, P. W. (2001). Su-6668. SUGEN. *Current Opinion in Investigational Drugs* (London, England: 2000), 2(8), 1142-1148.
70. Hoekman, K. (2001). SU6668, a multitargeted angiogenesis inhibitor. *Cancer Journal* (Sudbury, Mass.), 7, S134-138.
71. Dowlati, A., Robertson, K., Cooney, M., Petros, W.P., Stratford, M., Jesberger, J., et al. (2002). A phase I pharmacokinetic and translational study of the novel vascular targeting agent combretastatin A-4 phosphate on a single-dose intravenous schedule in patients with advanced cancer. *Cancer Research*, 62(12), 3408–3416.
72. Ohno, T., Kawano, K., Sasaki, A., Aramaki, M., Tahara, K., Etoh, T., et al. (2002). Antitumor and antivascular effects of AC-7700, a combretastatin A-4 derivative, against rat liver cancer. *International Journal of Clinical Oncology*, 7(3), 171–176.
73. Nagase, H., & Woessner Jr, J.F. (1999). Matrix metalloproteinases. *Journal of Biological Chemistry*, 274(31), 21491–21494.
74. Sternlicht, M.D., & Werb, Z. (2001). How matrix metalloproteinases regulate cell behavior. *Annual Review of Cell and Developmental Biology*, 17(1), 463-516.
75. Overall, C.M., & López-Otín, C. (2002). Strategies for MMP inhibition in cancer: Innovations for the post-trial era. *Nature Reviews Cancer*, 2(9), 657-672.



© University of Venda

76. Egeblad, M., & Werb, Z. (2002). New functions for the matrix metalloproteinases in cancer progression. *Nature Reviews Cancer*, 2(3), 161-174.
77. Hoekstra, R., Eskens, F.A., & Verweij, J. (2001). Matrix metalloproteinase inhibitors: Current developments and future perspectives. *Oncologist*, 6(5), 415–427.
78. Bramhall, S.R., Hallissey, M.T., Whiting, J., Scholefield, J., Tierney, G., Stuart, R.C., et al. (2002). Marimastat as maintenance therapy for patients with advanced gastric cancer: A randomised trial. *British Journal of Cancer*, 86(12), 1864–1870.
79. Bramhall, S.R., Schulz, J., Nemunaitis, J., Brown, P.D., Baillet, M., & Buckels, J.A. (2002). A double-blind placebo-controlled, randomised study comparing gemcitabine and marimastat with gemcitabine and placebo as first line therapy in patients with advanced pancreatic cancer. *British Journal of Cancer*, 87(2), 161–167.
80. Groves, M.D., Puduvalli, V.K., Hess, K.R., Jaeckle, K.A., Peterson, P., Yung, W.K., et al. (2002). Phase II trial of temozolomide plus the matrix metalloproteinase inhibitor, marimastat, in recurrent and progressive glioblastoma multiforme. *Journal of Clinical Oncology*, 20(5), 1383–1388.
81. Batist, G., Patenaude, F., Champagne, P., Croteau, D., Levinton, C., Hariton, C., et al. (2002). Neovastat (AE-941) in refractory renal cell carcinoma patients: Report of a phase II trial with two dose levels. *Annals of Oncology*, 13(8), 1259–1263.
82. Falardeau, P., Champagne, P., Poyet, P., Hariton, C., & Dupont, E. (2001). Neovastat, a naturally occurring multifunctional antiangiogenic drug, in phase III clinical trials. *Seminars in Oncology*, 28(6), 620–625.
83. Gutheil, J.C., Campbell, T.N., Pierce, P.R., Watkins, J.D., Huse, W.D., Bodkin, D.J., et al. (2000). Targeted antiangiogenic therapy for cancer using Vitaxin: A humanized monoclonal antibody to the integrin avb3. *Clinical Cancer Research*, 6(8), 3056–3061.
84. Patel, S.R., Jenkins, J., Papadopolous, N., Burgess, M.A., Plager, C., Gutterman, J., et al. (2001). Pilot study of vitaxin—an angiogenesis inhibitor—in patients with advanced leiomyosarcomas. *Cancer*, 92(5), 1347–1348.
85. Coleman, R. E. (2002). *Treatment of Skeletal Complications of Cancer with Zoledronic Acid (Zometa)*. Saunders.
86. Body, J.J., & Mancini, I. (2002). Bisphosphonates for cancer patients: Why, how, and when? *Supportive Care in Cancer*, 10, 399–407.



© University of Venda

87. Body, J.J., Greip, P., Coleman, R.E., et al. (2003). A phase I study of AMGN-0007, a recombinant osteoprotegerin construct, in patients with multiple myeloma or breast carcinoma related bone metastases. *Cancer*, 97(Suppl 3), 887–892.
88. Onuma, E., Sato, K., Saito, H., Tsunenari, T., Ishii, K., Esaki, K., Yabuta, N., Wakahara, Y., Yamada-Okabe, H., & Ogata, E. (2004). Generation of a humanized monoclonal antibody against human parathyroid hormone-related protein and its efficacy against humoral hypercalcemia of malignancy. *Anticancer Research*, 24(5A), 2665-2674.
89. Brunda, M.J., Bellantoni, D., & Sulich, V. (1987). In vivo anti-tumor activity of combinations of interferon alpha and interleukin-2 in a murine model: Correlation of efficacy with the induction of cytotoxic cells resembling natural killer cells. *International Journal of Cancer*, 40(3), 365-371.
90. Tuteja, R. (2007). Malaria—An overview. The FEBS Journal, 274(18), 4670-4679.
91. Eisenstein, M. (2012). Drug development: Holding out for reinforcements. Nature, 484(7395), S16-S18.
92. Maxmen, A. (2012). Public health: Death at the doorstep. *Nature*, 484(7395), S19-S21.
93. Chang, H.H., Moss, E.L., Park, D.J., Ndiaye, D., Mboup, S., Volkman, S.K., Sabeti, P.C., Wirth, D.F., Neafsey, D.E., & Hartl, D.L. (2013). Malaria life cycle intensifies both natural selection and random genetic drift. *Proceedings of the National Academy of Sciences*, 110(50), 20129-20134.
94. https://upload.wikimedia.org/wikipedia/commons/thumb/2/24/Life_Cycle_of_the_Malaria_Parasite.svg/1280px-Life_Cycle_of_the_Malaria_Parasite.svg.png (Accessed: August, 19, 2022)
95. Parasitic Infections of the Liver - Health Hearty. https://healthhearty.com/parasitic-infections-of-liver (Accessed: August, 14, 2022)
96. Christenson, J.C. (2016). Common Infections. *Global Child Health*, 641.
97. Handbook, W. A. P. (2012). Management of Severe Malaria. *WHO*.
98. Rowe, J.A., Moulds, J.M., Newbold, C.I., Miller, L.H., & Marsh, K. (1991). Plasmodium falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Journal of Experimental Medicine*, 173(4), 961-969.
99. Milner, D.A. (2018). Malaria Pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, 8(1).



© University of Venda

100. World Health Organization. (2015). World Health Statistics 2015. Retrieved from https://www.who.int/docs/default-source/gho-documents/world-health-statistic-reports/world-health-statistics-2015.pdf (Accessed: November 5, 2023)

101. Ashley, E. A., Pyae Phyo, A., & Woodrow, C. J. (2018). Malaria. *Lancet*, 391(10130), 1608-1621.

102. Plewes, K., Leopold, S. J., Kingston, H. W. F., & Dondorp, A. M. (2019). Malaria: What's New in the Management of Malaria? *Infect Dis Clin North Am*, 33(1), 39-60.

103. Kuethe, F., Pfeifer, R., Rummler, S., Bauer, K., Kamvissi, V., & Pfister, W. (2009). Treatment of a patient with shock complicating severe falciparum malaria: a case report. *Cases Journal*, 2, 1-6.

104. Beeson, J. G., Amin, N., Kanjala, M., & Rogerson, S. J. (2002). Selective accumulation of mature asexual stages of Plasmodium falciparum-infected erythrocytes in the placenta. *Infection and Immunity*, 70(10), 5412-5415.

105. World Health Organization. (2022). Malaria. Geneva: World Health Organization. Available from: https://www.who.int/health-topics/malaria#tab=tab_1 (Accessed: December 12, 2022)

106. World Health Organization. (2018). World Malaria Report. Geneva, Switzerland: *WHO*.

107. World Health Organization. (2020). The potential impact of health service disruptions on the burden of malaria: a modelling analysis for countries in sub-Saharan Africa. Geneva: *World Health Organization*.

108. Burrows, J. N., Burlot, E., Campo, B., Cherbuin, S., Jeanneret, F., Kamber, J., ... & Spangenberg, T. (2014). Antimalarial drug discovery–the path towards eradication. *Parasitology*, 141(1), 128-139. doi:10.1017/S0031182013000826

109. Ridley, R. G. (2002). Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature*, 415, 686-693.

110. Nzila, A. (2006). The past, present and future of antifolates in the treatment of Plasmodium falciparum infection. *Journal of Antimicrobial Chemotherapy*, 57(6), 1043-1054.

111. Cody, V., & Schwalbe, C. H. (2006). Structural characteristics of antifolate dihydrofolate reductase enzyme interactions. Crystallography Reviews, 12(4), 301-333.



© University of Venda

112. Fernández-Villa, D., Aguilar, M. R., & Rojo, L. (2019). Folic acid antagonists: antimicrobial and immunomodulating mechanisms and applications. *International Journal of Molecular Sciences*, 20(20), 4996.

113. Nzila-Mounda, A., Mberu, E. K., Sibley, C. H., Plowe, C. V., Winstanley, P. A., & Watkins, W. M. (1998). Kenyan Plasmodium falciparum field isolates: correlation between pyrimethamine and chlorcycloguanil activity in vitro and point mutations in the dihydrofolate reductase domain. *Antimicrobial Agents and Chemotherapy*, 42(1), 164-169.

114. Anderson, A. C. (2005). Targeting DHFR in parasitic protozoa. *Drug Discovery Today*, 10(2), 121-128.

115. Antony, H. A., & Parija, S. C. (2016). Antimalarial drug resistance: an overview. *Tropical Parasitology*, 6(1), 30. doi:10.4103/2229-5070.175081

116. Sathiyajith, C., Shaikh, R. R., Han, Q., Zhang, Y., Meguellati, K., & Yang, Y. W. (2017). Biological and related applications of pillar [n] arenes. *Chemical Communications*, 53(4), 677-696.

117. Rani, R., Sethi, K., Gupta, S., Varma, R. S., & Kumar, R. (2022). Mechanism of Action and Implication of Naphthoquinone as Potent Anti-trypanosomal Drugs. *Current Topics in Medicinal Chemistry*, 22(25), 2087-2105.

118. Sharma, A., Santos, I. O., Gaur, P., Ferreira, V. F., Garcia, C. R., & Da Rocha, D. R. (2013). Addition of thiols to o-quinone methide: new 2-hydroxy-3-phenylsulfanylmethyl [1, 4] naphthoquinones and their activity against the human malaria parasite Plasmodium falciparum (3D7). *European Journal of Medicinal Chemistry*, 59, 48-53.

119. Painter, H. J., Morrisey, J. M., Mather, M. W., & Vaidya, A. B. (2007). Specific role of mitochondrial electron transport in blood-stage Plasmodium falciparum. *Nature*, 446(7131), 88-91.

120. Goodman, C. D., Buchanan, H. D., & McFadden, G. I. (2013). Is the mitochondrion a good malaria drug target? *Trends in Parasitology*, 29(6), 284-286. doi:10.1016/j.pt.2013.04.007

121. Hussain, H., Krohn, K., Ahmad, V. U., Miana, G. A., & Green, I. R. (2007). Lapachol: an overview. *Arkivoc*, 2(1), 145-171.

122. Santos, S. S., de Araujo, R. V., Giarolla, J., El Seoud, O., & Ferreira, E. I. (2020). Searching for drugs for Chagas disease, leishmaniasis and schistosomiasis: a review. *International Journal of Antimicrobial Agents*, 55(4), 105906.



© University of Venda

123. Pereira, M. R., Henrich, P. P., Sidhu, A. B., Johnson, D., Hardink, J., Van Deusen, J., ... & Fidock, D. A. (2011). Antimalarial activity of amodiaquine and its primary metabolite, N-desethylamodiaquine, against pre-erythrocytic and erythrocytic stages of Plasmodium falciparum. *Antimicrobial Agents and Chemotherapy*, 55(6), 3115-3120. doi:10.1128/AAC.01739-10

124. Aderibigbe, B. A. (2017). Design of drug delivery systems containing artemisinin and its derivatives. *Molecules*, 22(2), 323. doi:10.3390/molecules22020323

125. Liu, C. X. (2017). Discovery and development of artemisinin and related compounds. *Chinese Herbal Medicines*, 9(2), 101-114. doi:10.1016/S1674-6384(17)60084-

126. Antimalarial activity of medicinal plants from the Democratic Republic .... Retrieved from https://balistarling.afphila.com/science/article/pii/S0378874115002342 (Accessed: November 5, 2023)

127. Infected by plasmodium malaria parasites. Medical search. Frequent .... Retrieved from https://lookformedical.com/en/faq/infected-by-plasmodium-malaria-parasites (Accessed: November 5, 2023)

128. Fernando, D., Rodrigo, C., & Rajapakse, S. (2011). Primaquine in vivax malaria: an update and review on management issues. *Malaria Journal*, 10(1), 1-12.

129. Wells, T. N., Burrows, J. N., & Baird, J. K. (2010). Targeting the hypnozoite reservoir of Plasmodium vivax: the hidden obstacle to malaria elimination. *Trends in Parasitology*, 26(3), 145-151.

130. Alonso, P. L., Brown, G., Arevalo-Herrera, M., Binka, F., Chitnis, C., Collins, F., ... & Tanner, M. (2011). A research agenda to underpin malaria eradication. *PLoS Medicine*, 8(1), e1000406.

131. Delves, M. J. (2012). Plasmodium cell biology should inform strategies used in the development of antimalarial transmission-blocking drugs. *Future Medicinal Chemistry*, 4(18), 2251-2263.

132. Buckling, A., Ranford-Cartwright, L. C., Miles, A., & Read, A. F. (1999). Chloroquine increases Plasmodium falciparum gametocytogenesis in vitro. *Parasitology*, 118(4), 339-346.

133. Puta, C., & Manyando, C. (1997). Enhanced gametocyte production in Fansidar-treated Plasmodium falciparum malaria patients: implications for malaria transmission control programmes. *Tropical Medicine & International Health*, 2(3), 227-229.



© University of Venda

134. Calderón, F., Ballell, L., Barros, D., Bilbe, G., Cammack, N., Gabarró, R., ... & Fairlamb, A. H. (2021). Tres Cantos Open Lab: celebrating a decade of innovation in collaboration to combat endemic infectious diseases. *Nature Reviews Drug Discovery*, 20(11), 799-800.

135. Jeong, C. G., Dal Negro, G., Getsios, S., & Ekert, J. E. (2019). Application of complex in vitro models (CIVMs) in drug discovery for safety testing and disease modeling. *In Microfluidic Cell Culture Systems* (pp. 121-158). Elsevier.

136. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., ... & Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 419(6906), 498-511.

137. Belete, T. M. (2020). Recent progress in the development of new antimalarial drugs with novel targets. *Drug Design, Development and Therapy*, 3875-3889.

138. Mittal, P., & Gautam, R. K. (2022). Current Therapies and New Drug Targets for the Future Drug Development of Drug Resistant Malaria. *Drug Development for Malaria: Novel Approaches for Prevention and Treatment*, 133-150.

139. Laxminarayan, R., & Heymann, D. L. (2012). Challenges of drug resistance in the developing world. *BMJ*, 344. doi:10.1136/bmj.e1567

140. Zahreddine, H., & Borden, K. L. (2013). Mechanisms and insights into drug resistance in cancer. *Frontiers in Pharmacology*, 4, 28. doi:10.3389/fphar.2013.00028

141. Nasiri, M. J., Haeili, M., Ghazi, M., Goudarzi, H., Pormohammad, A., Imani Fooladi, A. A., & Feizabadi, M. M. (2017). New insights into the intrinsic and acquired drug resistance mechanisms in mycobacteria. *Frontiers in Microbiology*, 8, 681. doi:10.3389/fmicb.2017.00681

142. Ray, S., Das, S., & Suar, M. (2017). Molecular mechanism of drug resistance. *Drug resistance in bacteria, fungi, malaria, and cancer*, 47-110.

143. Longley, D. B., & Johnston, P. G. (2005). Molecular mechanisms of drug resistance. The Journal of Pathology: *A Journal of the Pathological Society of Great Britain and Ireland*, 205(2), 275-292. https://doi.org/10.1002/path.1706

144. Sharma, J. (2011). Now You Can Also Publish Your Article Online. *Pharmatutor*, 20(0).

145. Baldo, B. A., & Pham, N. H. (2013). Adverse reactions to targeted and non-targeted chemotherapeutic drugs with emphasis on hypersensitivity responses and the invasive metastatic switch. *Cancer and Metastasis Reviews*, 32, 723-761.



© University of Venda

146. Perry, S. (1973). Cell Proliferation Characteristics and Cancer Chemotherapy. *In Temporal Aspects of Therapeutics* (pp. 129-141). Springer US.

147. Ray, S., Das, S., & Suar, M. (2017). Molecular mechanism of drug resistance. *In Drug Resistance in Bacteria, Fungi, Malaria, and Cancer* (pp. 47-110). Springer.

148. DeGorter, M. K., Xia, C. Q., Yang, J. J., & Kim, R. B. (2012). Drug transporters in drug efficacy and toxicity. *Annual Review of Pharmacology and Toxicology*, 52(1), 249-273.

149. Zango, U. U., Ibrahim, M., Shawai, S. A. A., & Shamsuddin, I. M. (2019). A review on β-lactam antibiotic drug resistance. *MOJ Drug Design & Development*, 3(2), 52-58.

150. Pagès, J. M., & Amaral, L. (2009). Mechanisms of drug efflux and strategies to combat them: Challenging the efflux pump of Gram-negative bacteria. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1794(5), 826-833.

151. Sui, X., Chen, R., Wang, Z., Huang, Z., Kong, N., Zhang, M., Han, W., Lou, F., Yang, J., Zhang, Q., & Wang, X. (2013). Autophagy and chemotherapy resistance: A promising therapeutic target for cancer treatment. *Cell Death & Disease*, 4(10), e838.

152. Sherlach, K. S., & Roepe, P. D. (2014). Drug resistance associated membrane proteins. *Frontiers in Physiology*, 5, 108.

153. Muriithi, W., Macharia, L. W., Heming, C. P., Echevarria, J. L., Nyachieo, A., Niemeyer Filho, P., & Neto, V. M. (2020). ABC transporters and the hallmarks of cancer: Roles in cancer aggressiveness beyond multidrug resistance. *Cancer Biology & Medicine*, 17(2), 253.

154. Li, Q., & Shu, Y. (2014). Role of solute carriers in response to anticancer drugs. *Molecular and Cellular Therapies*, 2(1), 1-14.

155. Ueda, K. (2011). ABC proteins protect the human body and maintain optimal health. *Bioscience, Biotechnology, and Biochemistry*, 75(3), 401-409.

156. Capranico, G., De Isabella, P., Castelli, C., Supino, R., Parmiani, G., & Zunino, F. (1989). P-glycoprotein gene amplification and expression in multidrug-resistant murine P388 and B16 cell lines. *British Journal of Cancer*, 59(5), 682-685.

157. Bates, S. E., Robey, R., Miyake, K., Rao, K., Ross, D. D., & Litman, T. (2001). The role of half-transporters in multidrug resistance. *Journal of Bioenergetics and Biomembranes*, 33(6), 503-511.

158. Ross, D. D., Yang, W., Abruzzo, L. V., Dalton, W. S., Schneider, E., Lage, H., & Doyle, L. A. (1999). Atypical multidrug resistance: Breast cancer resistance protein messenger RNA



© University of Venda

expression in mitoxantrone-selected cell lines. *Journal of the National Cancer Institute*, 91(5), 429-433.

159. Chang, X. B. (2007). A molecular understanding of ATP-dependent solute transport by multidrug resistance-associated protein MRP1. *Cancer Metastasis Reviews*, 26(1), 15-37. https://doi.org/10.1007/s10555-007-9041-7

160. Wang, Q., & Beck, W. T. (1998). Transcriptional suppression of multidrug resistance-associated protein (MRP) gene expression by wild-type p53. *Cancer Research*, 58(24), 5762-5769.

161. Tsou, S. H., Hou, M. H., Hsu, L. C., Chen, T. M., & Chen, Y. H. (2016). Gain-of-function p53 mutant with 21-bp deletion confers susceptibility to multidrug resistance in MCF-7 cells. *International Journal of Molecular Medicine*, 37(1), 233-242.

162. Borst, P., Jonkers, J., & Rottenberg, S. (2007). What makes tumors multidrug resistant? *Cell Cycle*, 6(22), 2782-2787.

163. Keshelava, N., Zuo, J. J., Chen, P., Waidyaratne, S. N., Luna, M. C., Gomer, C. J., & Reynolds, C. P. (2001). Loss of p53 function confers high-level multidrug resistance in neuroblastoma cell lines. *Cancer Research*, 61(16), 6185-6193.

164. Coley, H. M. (2008). Mechanisms and strategies to overcome chemotherapy resistance in metastatic breast cancer. *Cancer Treatment Reviews*, 34(4), 378-390.

165. Rantanen, V., Engblom, P., Raitanen, M., Hietanen, S., Haarala, M., & Grenman, S. (2002). Mutations of TP53 do not correlate with the sensitivity to paclitaxel — A study using 27 gynaecological cancer cell lines. *European Journal of Cancer*, 38(13), 1783-1791.

166. Abad, E., Graifer, D., & Lyakhovich, A. (2020). DNA damage response and resistance of cancer stem cells. *Cancer Letters*, 474, 106-117.

167. Bavaro, L., Martelli, M., Cavo, M., & Soverini, S. (2019). Mechanisms of disease progression and resistance to tyrosine kinase inhibitor therapy in chronic myeloid leukemia: An update. *International Journal of Molecular Sciences*, 20(24), 6141.

168. Fang, J., Nakamura, H., & Maeda, H. (2011). The EPR effect: Unique features of tumor blood vessels for drug delivery, factors involved, and limitations and augmentation of the effect. *Advanced Drug Delivery Reviews*, 63(3), 136-151.

169. Wernsdorfer, W. (1991). The development and spread of drug-resistant malaria. *Parasitology Today*, 7(9), 297-303.



© University of Venda

170. Wellems, T. E., & Plowe, C. V. (2001). Chloroquine-resistant malaria. Journal of Infectious Diseases, 184, 770–776. https://doi.org/10.1086/322858

171. Cowman, A. F., Morry, M. J., Biggs, B. A., Cross, G. A., & Foote, S. J. (1988). Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, 85(23), 9109–9113.

172. Sidhu, A. B. S., Valderramos, S. G., & Fidock, D. A. (2005). pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in Plasmodium falciparum. *Molecular Microbiology*, 57(4), 913–926. https://doi.org/10.1111/j.1365-2958.2005.04729.x

173. Tshilande, N., & Mammino, L. (2021). Acylphloroglucinols with anticancer and/or antimalarial activities – an overview. *Journal of AOAC International*, 104(2), 510-522.

174. Singh, I. P., Sidana, J., Bharatea, S. B., & Foley, W. J. (2010). Phloroglucinol compounds of natural origin: Synthetic aspects. *Natural Product Reports*, 27(3), 393-416.

175. Khan, F., Tabassum, N., Bamunuarachchi, N. I., & Kim, Y. M. (2022). Phloroglucinol and its derivatives: Antimicrobial properties toward microbial pathogens. *Journal of Agricultural and Food Chemistry*, 70(16), 4817-4838.

176. Singh, I. P., & Bharate, S. B. (2006). Phloroglucinol compounds of natural origin. *Natural Product Reports*, 23(4), 558-591.

177. Mammino, L. (2019). Effects of complexation with a metal ion on the intramolecular hydrogen bonds in acylphloroglucinols. *Theoretical Chemistry Accounts*, 138(8), 103.

178. Mammino, L., & Kabanda, M. M. (2007). Model structures for the study of acylated phloroglucinols and computational study of the caespitate molecule. *Journal of Molecular Structure*, 805(1-3), 39-52.

179. Mammino, L., & Kabanda, M. M. (2009). A study of the intramolecular hydrogen bond in acylphloroglucinols. *Journal of Molecular Structure*, 901(1-3), 210-219.

180. Kabanda, M. M., & Mammino, L. (2012). The conformational preferences of acylphloroglucinols – a promising class of biologically active compounds. *International Journal of Quantum Chemistry*, 112(14), 3691-3702.



© University of Venda

181. Mammino, L., & Kabanda, M. M. (2013). The role of additional O-H…O intramolecular hydrogen bonds for acylphloroglucinols' conformational preferences *in vacuo* and in solution. *Molecular Simulation*, 39(1), 1-13.

182. Bertolasi, V., Gilli, P., Ferretti, V., & Gilli, G. (1991). Evidence for resonance-assisted hydrogen bonding: Crystal structure and spectroscopic parameters in 1,3-diaryl-1,3-propanedione enols. *Journal of the American Chemical Society*, 113(13), 4917-4925.

183. Mammino, L. (2019). Intramolecular hydrogen bonding patterns, conformational preferences, and molecular properties of dimeric acylphloroglucinols: An ab initio and DFT study. *Journal of Molecular Structure*, 1176, 488-500.

184. Kabanda, M. M., & Mammino, L. (2012). Computational study of weaker intramolecular hydrogen bonds stabilizing acylphloroglucinols. *International Journal of Quantum Chemistry*, 112(14), 2650-2658.

185. Mammino, L. (2021). Correlation effects in trimeric acylphloroglucinols. *Computation*, 9(11), 121. https://doi.org/10.3390/computation9110121

186. Mammino, L. (2019). Intramolecular hydrogen bonding patterns, conformational preferences, and molecular properties of dimeric acylphloroglucinols. *Journal of Molecular Structure*, 1176, 488-500.

187. Mammino, L. (2022). Conformational preferences and intramolecular hydrogen bonding patterns of tetraflavaspidic acid BBBB – a tetrameric acylphloroglucinol. *Physical Sciences Reviews*, https://doi.org/10.1515/psr-2021-0239

188. Mammino, L. (2018). Adducts of arzanol with explicit water molecules: an ab initio and dft study. In *Concepts, Methods and Applications of Quantum Systems in Chemistry and Physics: Selected proceedings of QSCP-XXI (Vancouver, BC, Canada, July 2016)* (pp. 281-304). Springer International Publishing.

189. Marti, G., Eparvier, V., Moretti, C., Susplugas, S., Prado, S., Grellier, P., Retailleau, P., Guéritte, F., & Litaudon, M. (2009). Antiplasmodial benzophenones from the trunk latex of Moronobea coccinea (Clusiaceae). *Phytochemistry*, 70(1), 75-85.

190. Menegazzi, M., Masiello, P., & Novelli, M. (2021). Anti-tumor activity of Hypericum perforatum L. and hyperforin through modulation of inflammatory signaling, ROS generation and proton dynamics. *Antioxidants*, 10(1), 18.



© University of Venda

191. Marti, G., Eparvier, V., Moretti, C., Prado, S., Grellier, P., Hue, N., Thoison, O., Delpech, B., Guéritte, F., & Litaudon, M. (2010). Antiplasmodial benzophenone derivatives from the root barks of Symphonia globulifera (Clusiaceae). *Phytochemistry*, 71(9-10), 964-974.

192. Verotta, L., Giovanni, A., Bombardellic, E., & Brund, R. (2007). In vitro antimalarial activity of hyperforin, a prenylated acylphloroglucinol. A structure–activity study. *Bioorganic & Medicinal Chemistry Letters*, 17(6), 1544-1548.

193. Feilcke, R., Arnouk, G., Raphane, B., Richard, K., Tietjen, I., Andrae-Marobela, K., Erdmann, F., Schipper, S., Becker, K., Arnold, N., & Frolov, A. (2019). Biological activity and stability analyses of knipholone anthrone, a phenyl anthraquinone derivative isolated from Kniphofia foliosa Hochst. *Journal of Pharmaceutical and Biomedical Analysis*, 174, 277-285.

194. Abdissa, N., Induli, M., Akala, H. M., Heydenreich, M., Midiwo, J. O., Ndakala, A., & Yenesew, A. (2013). Knipholone cyclooxanthrone and an anthraquinone dimer with antiplasmodial activities from the roots of Kniphofia foliosa. *Phytochemistry Letters*, 6(2), 241-245.

195. Dagne, E., & Steglich, W. (1984). Knipholone: A unique anthraquinone derivative from Kniphofia foliosa. *Phytochemistry*, 23(8), 1729-1731.

196. Dagne, E., & Yenesew, A. (1993). Knipholone anthrone from Kniphofia. *Phytochemistry*, 34(5), 1440-1441.

197. Zhang, Y., He, X. Z., Yang, H., & Liu, H. Y. (2021). Robustadial A and B from Eucalyptus globulus Labill. and their anticancer activity as selective tyrosyl-DNA phosphodiesterase 2 inhibitors. *Phytotherapy Research*, 35(9), 5282-5289.

198. Decosterd, L. A., Hoffmann, E., Kyburz, R., Bray, D., & Hostettmann, K. (1991). A new phloroglucinol derivative from Hypericum calycinum with antifungal and in vitro antimalarial activity. *Planta Medica*, 57(6), 548-551.

199. Xu, R., Snyder, J. K., & Nakanishi, K. (1984). Robustadials A and B from Eucalyptus robusta. *Journal of the American Chemical Society*, 106(3), 734-736. https://doi.org/10.1021/ja00315a046

200. Lu, Z., Van Wagoner, R. M., Pond, C. D., Pole, A. R., Jensen, J. B., Blankenship, D., Grimberg, B. T., Kiapranis, R., Matainaho, T. K., Barrows, L. R., & Ireland, C. M. (2014). Myristicyclins A and B: Antimalarial procyanidins from Horsfieldia spicata from Papua New Guinea. *Organic Letters*, 16(2), 346-349. https://doi.org/10.1021/ol403228q



© University of Venda

201. Myristicyclins B. (n.d.). Google Books. https://books.google.co.za/books?id=CBXYCwAAQBAJ&pg=PR7 (Accessed: December 10, 2023)

202. Maloney, D. J., Deng, J.-Z., Starck, S. R., Gao, Z., & Hecht, S. M. (2005). (+)-Myristinin A, a naturally occurring DNA polymerase beta inhibitor and potent DNA-damaging agent. *Journal of the American Chemical Society*, 127(12), 4140-4141. https://doi.org/10.1021/ja0425738

203. Guo-Wei, Q., Zheng-Xiong, C., Hong-Cheng, W., & Ming-Kun, Q. (1981). The structure and synthesis of robustaol A. *Acta Chimica Sinica*, 39(1), 83-89.

204. Liu, L.-S., Liu, M.-H., & He, J.-Y. (2014). Hypericum japonicum Thunb. ex Murray: Phytochemistry, pharmacology, quality control and pharmacokinetics of an important herbal medicine. *Molecules*, 19(8), 10733-10754. https://doi.org/10.3390/molecules190810733

205. Gu, G. M., Feng, S. Z., & Wang, X. Y. (1988). Antimalarial constituents of Hypericum japonicum Thunb.: Isolation and structure of japonicins A, B, C, and D. *Acta Chimica Sinica*, 46, 246-251.

206. Peron, G., Pant, D. R., Shrestha, S. S., Rajbhandary, S., & D'Acqua, S. (2020). An integrated LC-ESI-MSn and high-resolution LC-ESI-QTOF approach for the identification of phloroglucinols from Nepalese Hypericum japonicum. *Molecules*, 25(5937), 1-13. https://doi.org/10.3390/molecules25245937

207. Ishiguro, K., Yamaki, M., Kashihara, M., & Takagi, S. (1986). Sarothralen A and B: New antibiotic compounds from Hypericum japonicum. *Planta Medica*, 52(4), 288-290. https://doi.org/10.1055/s-2007-969126

208. Ishiguro, K., Yamaki, M., Kashihara, M., Takagi, S., & Isoi, K. (1990). Sarothralin G: A new antimicrobial compound from Hypericum japonicum. *Planta Medica*, 56(3), 274-276. https://doi.org/10.1055/s-2006-960960

209. Cheng, C. (1978). Studies on the active principles of Shianhotsao—II. The structures of Agrimol A, B, D, and E. *Acta Chimica Sinica*, 36, 35-41.

210. Le, Q. U., Joshi, R. K., Lay, H. L., & Wu, M. C. (2018). Agrimonia pilosa Ledeb: Phytochemistry, ethnopharmacology, pharmacology of an important traditional herbal medicine. *Journal of Pharmacognosy and Phytochemistry*, 7, 3202-3211.



© University of Venda

211. Mammino, L. (2021). Correlation effects in trimeric acylphloroglucinols. *Computation*, 9(11), 121. https://doi.org/10.3390/computation9110121

212. Biljali, S., Momekov, G., Nedialkov, P., Zheleva-Dimitrova, D., Kitanov, G., Momekova, D., Stoyanov, N., & Karaivanova, M. (2012). In vitro investigation of the antiproliferative and proapoptotic effects of hyperatomarin – A bicyclic prenylated acylphloroglucinol from Hypericum annulatum Moris Subsp. Annulatum against human tumor and endothelial cells. *Journal of Pharmaceutical Technology and Drug Research*, 1(1), 1-9.

213. Šavikin-Fodulović, K., Aljančić, I., Vajs, V., Menković, N., Macura, S., Gojgić, G., & Milosavljević, S. (2003). Hyperatomarin, an antibacterial prenylated phloroglucinol from Hypericum atomarium ssp. degenii. *Journal of Natural Products*, 66(9), 1236-1238.

214. Momekov, G., Ferdinandov, D., Zheleva-Dimitrova, D., Nedialkov, P., Girreser, U., & Kitanov, G. (2008). Cytotoxic effects of hyperatomarin, a prenylated phloroglucinol from Hypericum annulatum Moris Subsp. Annulatum, in a panel of malignant cell lines. P*hytomedicine*, 15(11), 1010-1015.

215. Vergoten, G., & Bailly, C. (2021). Anticancer properties and mechanism of action of oblongifolin C, guttiferone K, and related polyprenylated acylphloroglucinols. *Natural Products and Bioprospecting*, 11, 629–641. https://doi.org/10.1007/s13659-021-00298-2

216. Li, H., Meng, X. X., Zhang, L., Zhang, B. J., Liu, X. Y., Fu, W. W., Tan, H. S., Lao, Y. Z., & Xu, H. X. (2017). Oblongifolin C and guttiferone K extracted from Garcinia yunnanensis fruit synergistically induce apoptosis in human colorectal cancer cells in vitro. *Acta Pharmacologica Sinica*, 38, 252–263. https://doi.org/10.1038/aps.2016.72

217. Han, L., Xu, D., Xi, Z., Wu, M., Nik Nabil, W. N., Zhang, J., Sui, H., Fu, W., Zhou, H., Lao, Y., Xu, G., Guo, C., & Xu, H. (2020). The natural compound oblongifolin C exhibits anticancer activity by inhibiting HSPA8 and cathepsin B in vitro. *Frontiers in Pharmacology*, 11, 1-14. https://doi.org/10.3389/fphar.2020.569193

218. Stricher, F., Macri, C., Ruff, M., & Muller, S. (2013). HSPA8/HSC70 chaperone protein: Structure, function, and chemical targeting. *Autophagy*, 9(11), 1937–1954. https://doi.org/10.4161/auto.26448

219. Camargo, M. S., Prieto, A. M., Resende, F. A., Boldrin, P. K., Cardoso, C. R. P., Fernández, M. F., Molina-Molina, J. M., Olea, N., Vilegas, W., Cuesta-Rubio, O., & Varanda, E. A. (2013). Evaluation of estrogenic, antiestrogenic and genotoxic activity of nemorosone, the



© University of Venda

major compound found in brown Cuban propolis. *BMC Complementary and Alternative Medicine*, 13, 201. https://doi.org/10.1186/1472-6882-13-201

220. Cuesta-Rubio, F. O., Velez-Castro, H., Frontana-Uribe, B., & Cardenas, J. (2001). Nemorosone, the major constituent of floral resins of Clusia rosea. *Phytochemistry*, 57(20), 279-283. https://doi.org/10.1016/S0031-9422(01)00093-7

221. Pagano, B., Pavone, M., Piccinelli, A., Rastrelli, L., Cuesta-Rubio, O., Mattia, C., & Barone, V. (2008). Structural and conformational investigation of nemorosone: A combined X-ray and quantum mechanical study. *Chemical Physics Letters*, 462(4), 158-163. https://doi.org/10.1016/j.cplett.2008.07.015

222. Pardo-Andreu, G. L., Nuñez-Figueredo, Y., Tudella, V. G., Cuesta-Rubio, O., Rodrigues, F. P., Pestana, C. R., Uyemura, S. A., Leopoldino, A. M., Alberici, L. C., & Curti, C. (2011). The anti-cancer agent guttiferone-A permeabilizes mitochondrial membrane: Ensuing energetic and oxidative stress implications. *Toxicology and Applied Pharmacology*, 253(3), 282-289. https://doi.org/10.1016/j.taap.2011.03.014

223. Wu, Z. N., Niu, Q. W., Zhang, Y. B., Luo, D., Li, Q. G., Li, Y. Y., Kuang, G. K., He, L. J., Wang, G. C., & Li, Y. L. (2019). Hyperpatulones A–F, polycyclic polyprenylated acylphloroglucinols from Hypericum patulum and their cytotoxic activities. *RSC Advances*, 9, 7961-7966. https://doi.org/10.1039/C8RA10199A

224. Ao, Z., Liu, Y. Y., Lin, Y. L., Chen, X., Chen, K., Kong, L. Y., & Luo, J. G. (2020). Hyperpatulones A and B, two new peroxide polyprenylated acylphloroglucinols from the leaves of Hypericum patulum. *Tetrahedron Letters*, 61, 151385. https://doi.org/10.1016/j.tetlet.2020.151385

225. Xin, W. B., Man, X. H., Zheng, C. J., Jia, M., Jiang, Y. P., Zhao, X. X., Jin, G. L., Mao, Z. J., & Huang, H. Q. (2012). Prenylated phloroglucinol derivatives from Hypericum sampsonii. *Fitoterapia*, 83(8), 1540-1547. https://doi.org/10.1016/j.fitote.2012.08.008

226. Rancon, S., Chaboud, A., Darbour, N., Comte, G., Bayet, C., Simon, P. N., Raynaud, J., Pietro, A. D., Cabalion, P., & Barron, D. (2001). Natural and synthetic benzophenones: Interaction with the cytosolic binding domain of P-glycoprotein. *Phytochemistry*, 57(4), 553-557. https://doi.org/10.1016/S0031-9422(01)00057-3

227. Zhang, Y., Luo, M., Zu, Y., Fu, Y., Gu, C., & Wang, W. (2012). Dryofragin, a phloroglucinol derivative, induces apoptosis in human breast cancer MCF-7 cells through



© University of Venda

ROS-mediated mitochondrial pathway. *Chemico-Biological Interactions*, 199(2), 129-136. https://doi.org/10.1016/j.cbi.2012.06.008

228. Nomura, T., Fukai, T., & Hano, Y. (2003). Chemistry and biological activities of isoprenylated flavonoids from medicinal plants (Moraceous plants and Glycyrrhiza species). In *Studies in Natural Products Chemistry* (pp. 199-256). Elsevier. https://doi.org/10.1016/S1572-5995(03)80106-5

229. Zhang, Y. B., Luo, D., Yang, L., Cheng, W., He, L. J., Kuang, G. K., Li, M. M., Li, Y. L., & Wang, G. C. (2018). Matrine-type alkaloids from the roots of Sophora flavescens and their antiviral activities against the hepatitis B virus. *Journal of Natural Products*, 81(10), 2259-2265. https://doi.org/10.1021/acs.jnatprod.8b00169

230. Li, G. X., Du, X. Y., Xie, Y. Q., Liu, D., Li, R. T., & Zhang, Z. J. (2021). Flavescenols A and B, two lavandulylated acylphloroglucinols from Sophora flavescens. *Phytochemistry Letters*, 43, 179-183. https://doi.org/10.1016/j.phytol.2021.01.016

231. Zhang, Y. B., Zhan, L. Q., Li, G. Q., Wang, F., Wang, Y., Li, Y. L., Ye, W. C., & Wang, G. C. (2016). Dimeric matrine-type alkaloids from the roots of Sophora flavescens and their anti-hepatitis B virus activities. *Journal of Organic Chemistry*, 81(15), 6273-6280. https://doi.org/10.1021/acs.joc.6b00648

232. Tshilande, N., & Mammino, L. (2021, November 18-19). Theoretical comparison of the molecular properties of euglobals – a subclass of acylphloroglucinols. Paper presented at the XIX Reunión Mexicana de Fisicoquímica Teórica, online.

233. Mammino, L. (2015). Ab initio and DFT study of chinesin I and chinesin II. *Current Physical Chemistry*, 5(3), 274-293. https://doi.org/10.2174/1877946805666150827200459

234. Coppen, J. J. W. (Ed.). (2021). Eucalyptus: The genus Eucalyptus. CRC Press. https://books.google.co.za/books?id=0dRlDMvlhQ0C&pg=PA286&lpg=PA286&dq=Eucalyptus:+The+Genus+Eucalyptus,+Edited+By+John+J.W.+Coppen+Euglobal+Iii+Glycyrrhetic+Acid&source=bl&ots=rHKwQl8bvI&sig=ACfU3U2W0OANK688YiPU7PA70V7X2zzhEQ&hl=en&sa=X&ved=2ahUKEwih-LX_y8j3AhVlQvEDHZaeBv8Q6AF6BAgCEAM (Accessed: November 19, 2022)

235. Takasaki, M., Konoshima, T., Fujitani, K., Yoshida, S., Nishimura, H., Tokuda, H., Nishino, H., Iwashima, A., & Kozuka, M. (1990). Inhibitors of skin-tumor promotion. VIII: Inhibitory effects of euglobals and their related compounds on Epstein-Barr virus activation.



© University of Venda

*Chemical and Pharmaceutical Bulletin*, 38(10), 2737-2739. https://doi.org/10.1248/cpb.38.2737

236. Takasaki, M., Konoshima, T., Fujitani, K., Yoshida, S., Nishimura, H., Tokuda, H., ... & Kozuka, M. (1990). Inhibitors of Skin-Tumor Promotion. VIII.: Inhibitory Effects of Euglobals and Their Related Compounds on Epstein-Barr Virus Activation.(1). *Chemical and pharmaceutical bulletin*, *38*(10), 2737-2739.

237. Takasaki, M., Konoshima, T., Kozuka, M., & Tokuda, H. (1995). Anti-tumor-promoting activities of euglobals from Eucalyptus plants. *Biological and pharmaceutical bulletin*, *18*(3), 435-438.

238. Aloisi, F., & Salvetti, M. (2022). Epstein-Barr virus and multiple sclerosis: supporting causality. *The Lancet Neurology*, 21(4), 300-301.

239. Jayasurya, H., McChesney, J.D., Swanson, S.M., & Pezzuto, J.M. (1989). Antimicrobial and cytotoxic activity of rottlerin-type compounds from *Hypericum drummondii*. *Journal of Natural Products*, 52(2), 325–331.

240. Jayasurya, H., Clark, A.M., & McChesney, J.D. (1991). New antimicrobial filicinic acid derivatives from *Hypericum drummondii*. *Journal of Natural Products*, 54(5), 1314–1320.

241. Arisawa, M., Fujita, A., Suzuki, R., Hayashi, T., Morita, N., Kawano, N., & Koshimura, S. (1985). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, Part I. *Journal of Natural Products*, 48(3), 455–459.

242. Arisawa, M., Fujita, A., Saga, M., Hayashi, T., Morita, N., Kawano, N., & Koshimura, S. (1986). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, Part II. *Journal of Natural Products*, 49(2), 298–302.

243. Fujita, A., Hayashi, T., Arisawa, M., Shimizu, M., & Morita, N. (1988). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, Part III. *Journal of Natural Products*, 51(4), 708–712.

244. Arisawa, M., Fujita, A., Morita, N. (1990). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, V. Three new phloroglucinol derivatives, butyrylmallotochromanol, isobutyrylmallotochromanol, and mallotojaponol. *Journal of Natural Products*, 53(3), 638–643.



© University of Venda

245. Arisawa, M., Fujita, A., Hayashi, T., Morita, N., Kikuchi, T., & Tezuka, Y. (1990). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, IV. *Chemical & Pharmaceutical Bulletin*, 38(3), 698–700.

246. Arisawa, M., Fujita, A., Morita, N., & Koshimura, S. (1990). Cytotoxic and antitumor constituents in pericarps of *Mallotus japonicus*. *Planta Medica*, 56(4), 377–379.

247. Arisawa, M., Fujita, A., Hayashi, T., Hayashi, K., Ochiai, H., & Morita, N. (1990). Cytotoxic and antiherpetic activity of phloroglucinol derivatives from *Mallotus japonicus* (Euphorbiaceae). *Chemical & Pharmaceutical Bulletin*, 38(6), 1624–1626.

248. Wu, T.-S., Wang, M.L., Jong, T.T., McPhail, A.T., McPhail, D.R., & Lee, K.H. (1989). X-ray crystal structure of acrovestone, a cytotoxic principle from *Acronychia pedunculata*. *Journal of Natural Products*, 52(6), 1284–1289.

249. De Silva, L.B., Herath, W.M., Liyange, C., Kumar, V., Ahmad, V.U., & Sultana, A. (1991). Demethylacrovestone from *Acronychia pedunculata* fruits. *Phytochemistry*, 30(5), 1709–1710.

250. Su, C.-R., Kuo, P.-C., Wang, M.-L., Liou, M.-J., Damu, A.G., & Wu, T.-S. (2003). Acetophenone derivatives from *Acronychia pedunculata*. *Journal of Natural Products*, 66(7), 990–993.

251. Sun, Y., Mu, F., Li, C., Wang, W., Luo, M., & Fu, Y. (2013). Aspidin BB, a phloroglucinol derivative, induces cell cycle arrest and apoptosis in human ovarian HO-8910 cells. *Chemico-Biological Interactions*, 204, 88–97.

252. Woldemariam, T.Z., Fell, A.F., & Linley, P.A. (1990). Chromatographic and spectroscopic studies on the constituents in male and female flowers of *Hagenia abyssinica*. *Journal of Pharmaceutical and Biomedical Analysis*, 8(8–12), 859–865.

253. Woldemariam, T.Z., Fell, A.F., Linley, P.A., Bibby, M.C., & Phillips, M.R. (1992). Evaluation of the anti-tumour action and acute toxicity of kosins from *Hagenia abyssinica*. *Journal of Pharmaceutical and Biomedical Analysis*, 10(8), 555–560.

254. Liu, Z., Zhao, D., Jiang, S., Xue, B., Zhang, Y., & Yan, X. (2018). Anticancer phenolics from *Dryopteris fragrans* (L.) Schott. *Molecules*, 23, 680. https://doi.org/10.3390/molecules23030680 (Accessed: October 15, 2022).



© University of Venda

255. Li, N., Khan, S.I., Qiu, S., & Li, X.C. (2018). Synthesis and anti-inflammatory activities of phloroglucinol-based derivatives. *Molecules*, 23, 3232. https://doi.org/10.3390/molecules23123232 (Accessed: October 12, 2022).

256. Henry, G.E., Campbell, M.S., Zelinsky, A.A., Liu, Y., Bowen-Forbes, C.S., & Li, L. (2009). Bioactive acylphloroglucinols from *Hypericum densiflorum*. *Phytotherapy Research*, 23, 1759–1762.

257. Zhang, F., Lai, Q., Lai, W., Li, M., Jin, X., & Ye, L. (2022). Phloroglucinol derivatives as anti-tumor agents: synthesis, biological activity evaluation and molecular docking studies. *Medical Chemistry Research*, 31(1), 165–176.

258. Gajiwala, K.S., Feng, J., Ferre, R., Ryan, K., Brodsky, O., & Weinrich, S. (2013). Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. *Structure*, 21, 209–219.

259. Williams, P.A., Cosme, J., Ward, A., Angove, H.C., Vinković, D.M., & Jhoti, H. (2003). Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature*, 424, 464–468.

260. Yan, P., Lai, Q., Li, M., Jin, X., Wie, G., Chen, W., & Ye, L. (2021). New anticancer agents: design, synthesis, biological activity, and molecular docking of bicyclic phloroglucinol derivatives. *ChemistrySelect*, 6(7), 1453–1457.

261. Cho, M.-Y., Park, S.-Y., Park, S., Lee, Y. R., Han, G.-D., Kim, J.-A., & Geranyl, M. (2012). Geranyl derivative of phloroacetophenone induces cancer cell-specific apoptosis through bax-mediated mitochondrial pathway in MCF-7 human breast cancer cells. *Biological & Pharmaceutical Bulletin*, 35(1), 98–104.

262. Rancon, S., Chaboud, A., Darbour, N., Comte, G., Bayet, C., Simon, P.-N., Raynaud, J., Pietro, A. D., Cabalion, P., & Barron, D. (2001). Natural and synthetic benzophenones: Interaction with the cytosolic binding domain of P-glycoprotein. *Phytochemistry*, 57(4), 553–557.

263. Lehne, G. (2000). P-glycoprotein as a drug target in the treatment of multidrug-resistant cancer. *Current Drug Targets*, 1(1), 85–99.

264. Xin, A. W.-B., Man, B. X.-H., Zheng, A. C.-J., Jia, A. M., Jiang, A. Y.-P., Zhao, A. X.-X., Jin, A. G.-L., Mao, C. Z.-J., Huang, D. H.-Q., & Lu, Q.-P. (2012). Prenylated phloroglucinol derivatives from *Hypericum sampsonii*. *Fitoterapia*, 83(8), 1540–1547.



© University of Venda

265. Takasaki, M., Konoshima, T., Fujitani, K., Yoshida, S., Nishimura, H., Tokuda, H., Nishino, H., Iwashima, A., & Kozuka, M. (1990). Inhibitors of skin-tumor promotion. VIII: Inhibitory effects of euglobals and their related compounds on Epstein-Barr virus activation. *Chemical & Pharmaceutical Bulletin*, 38(10), 2737–2739.

266. Takasaki, M., Konoshima, T., Kozuka, M., & Tokuda, H. (1995). Anti-tumor-promoting activities of euglobals from eucalyptus plants. *Biological & Pharmaceutical Bulletin*, 18(3), 435–438.

267. Aloisi, F., & Salvetti, M. (2022). Epstein-Barr virus and multiple sclerosis: Supporting causality. *Lancet Neurology*, 21(4), 300–301.

268. Jayasurya, H., & McChesney, J. D. (1988). Drummondins A–C: Three novel rottlerin-type antibiotics from *Hypericum drummondii*. *Journal of the Chemical Society, Chemical Communications*, 24, 1592–1593.

269. Jayasurya, H., McChesney, J. D., Swanson, S. M., & Pezzuto, J. M. (1989). Antimicrobial and cytotoxic activity of rottlerin-type compounds from *Hypericum drummondii*. *Journal of Natural Products*, 52(2), 325–331.

270. Jayasurya, H., Clark, A. M., & McChesney, J. D. (1991). New antimicrobial filicinic acid derivatives from *Hypericum drummondii*. *Journal of Natural Products*, 54(5), 1314–1320.

271. Arisawa, M., Fujita, A., Suzuki, R., Hayashi, T., Morita, N., Kawano, N., & Koshimura, S. (1985). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, Part I. *Journal of Natural Products*, 48(3), 455–459.

272. Arisawa, M., Fujita, A., Saga, M., Hayashi, T., Morita, N., Kawano, N., & Koshimura, S. (1986). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, Part II. *Journal of Natural Products*, 49(2), 298–302.

273. Fujita, A., Hayashi, T., Arisawa, M., Shimizu, M., & Morita, N. (1988). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, Part III. *Journal of Natural Products*, 51(4), 708–712.

274. Arisawa, M., Fujita, A., Morita, N. (1990). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, V: Three new phloroglucinol derivatives, butyrylmallotochromanol, isobutyrylmallotochromanol, and mallotojaponol. *Journal of Natural Products*, 53(3), 638–643.



© University of Venda

275. Arisawa, M., Fujita, A., Hayashi, T., Morita, N., Kikuchi, T., & Tezuka, Y. (1990). Studies on cytotoxic constituents in pericarps of *Mallotus japonicus*, IV. *Chemical & Pharmaceutical Bulletin*, 38(3), 698–700.

276. Arisawa, M., Fujita, A., Morita, N., & Koshimura, S. (1990). Cytotoxic and antitumor constituents in pericarps of *Mallotus japonicus*. *Planta Medica*, 56(4), 377–379.

277. Arisawa, M., Fujita, A., Hayashi, T., Hayashi, K., Ochiai, H., & Morita, N. (1990). Cytotoxic and antiherpetic activity of phloroglucinol derivatives from *Mallotus japonicus* (Euphorbiaceae). *Chemical & Pharmaceutical Bulletin*, 38(6), 1624–1626.

278. Wu, T.-S., Wang, M.L., Jong, T.T., McPhail, A.T., McPhail, D.R., & Lee, K.H. (1989). X-ray crystal structure of acrovestone, a cytotoxic principle from *Acronychia pedunculata*. *Journal of Natural Products*, 52(6), 1284–1289.

279. De Silva, L.B., Herath, W.M., Liyange, C., Kumar, V., Ahmad, V.U., & Sultana, A. (1991). Demethylacrovestone from *Acronychia pedunculata* fruits. *Phytochemistry*, 30(5), 1709–1710.

280. Su, C.-R., Kuo, P.-C., Wang, M.-L., Liou, M.-J., Damu, A.G., & Wu, T.-S. (2003). Acetophenone derivatives from *Acronychia pedunculata*. *Journal of Natural Products*, 66(7), 990–993.

281. Sun, Y., Mu, F., Li, C., Wang, W., Luo, M., & Fu, Y. (2013). Aspidin BB, a phloroglucinol derivative, induces cell cycle arrest and apoptosis in human ovarian HO-8910 cells. *Chemico-Biological Interactions*, 204, 88–97.

282. Woldemariam, T.Z., Fell, A.F., & Linley, P.A. (1990). Chromatographic and spectroscopic studies on the constituents in male and female flowers of *Hagenia abyssinica*. *Journal of Pharmaceutical and Biomedical Analysis*, 8(8–12), 859–865.

283. Woldemariam, T.Z., Fell, A.F., Linley, P.A., Bibby, M.C., & Phillips, M.R. (1992). Evaluation of the anti-tumour action and acute toxicity of kosins from *Hagenia abyssinica*. *Journal of Pharmaceutical and Biomedical Analysis*, 10(8), 555–560.

284. Thomsen, H., Reider, K., Franke, K., Wessjohann, L.A., Keiser, J., Dagne, E., & Arnold, N. (2012). Characterization of constituents and anthelmintic properties of *Hagenia abyssinica*. *Scientific Reports*, 80(2), 433–446.



© University of Venda

285. Liu, Z., Zhao, D., Jiang, S., Xue, B., Zhang, Y., & Yan, X. (2018). Anticancer phenolics from *Dryopteris fragrans* (L.) Schott. *Molecules*, 23, 680. https://doi.org/10.3390/molecules23030680.

286. Li, N., Khan, S.I., Qiu, S., & Li, X.C. (2018). Synthesis and anti-inflammatory activities of phloroglucinol-based derivatives. *Molecules*, 23, 3232. https://doi.org/10.3390/molecules23123232.

287. Henry, G.E., Campbell, M.S., Zelinsky, A.A., Liu, Y., Bowen-Forbes, C.S., & Li, L. (2009). Bioactive acylphloroglucinols from *Hypericum densiflorum*. *Phytotherapy Research*, 23, 1759–1762.

288. Zhang, F., Lai, Q., Lai, W., Li, M., Jin, X., & Ye, L. (2022). Phloroglucinol derivatives as anti-tumor agents: Synthesis, biological activity evaluation and molecular docking studies. *Medicinal Chemistry Research*, 31(1), 165–176.

289. Gajiwala, K.S., Feng, J., Ferre, R., Ryan, K., Brodsky, O., & Weinrich, S. (2013). Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. *Structure*, 21, 209–219.

290. Williams, P.A., Cosme, J., Ward, A., Angove, H.C., Vinković, D.M., & Jhoti, H. (2003). Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature*, 424, 464–468.

291. Yan, P., Lai, Q., Li, M., Jin, X., Wie, G., Chen, W., & Ye, L. (2021). New anticancer agents: Design, synthesis, biological activity, and molecular docking of bicyclic phloroglucinol derivatives. *Chemistry - A European Journal*, 6(7), 1453–1457.



© University of Venda

# CHAPTER 3
# THEORETICAL BACKGROUND

## 3.0. Introduction to chapter

Computational studies of molecules can provide valuable information about their important characteristics, such as geometries, energies, molecular orbitals, dipole moments, etc. This is especially significant when studying molecules that play a role in biological processes and impact living organisms. Understanding the molecular properties that contribute to biological activity enhances our understanding of the fundamental principles governing life. This chapter first focuses on the molecular properties of interest and then explores computational techniques used to investigate molecules, including their theoretical foundations and considerations for accuracy in computational predictions since theoretical results inherently depend on the approximations and models used.

## 3.1. Molecular Geometry

### 3.1.1. Molecular geometry parameters

Molecules consist of stable groups of electrons and nuclei held together by chemical bonds [1-3]. Thus, a chemical bond refers to the attractive force that holds atoms or ions together within a molecule. The characteristics of chemical bonds, such as their energies and the properties of the molecules they form, can be effectively determined and quantified using a range of theoretical and computational approaches.

The atoms within a molecule are arranged in a characteristic sequence, and their arrangement extends into three-dimensional space. Experimental techniques [4-5], particularly spectroscopy, are used to determine the structure of a molecule, which involves the sequential arrangement of its atoms. The three-dimensional structure of a molecule is described utilizing three geometric parameters: bond lengths, bond angles, and torsion angles.

The spatial arrangement of constituent atoms is an important factor in understanding the properties of a molecule, as it plays a fundamental role in determining its chemical and physical characteristics [6]. To better understand these parameters, we can use cyclopropene (as depicted



© University of Venda

in Figure 3.1) as an example. Cyclopropene is a molecule consisting of three carbon atoms and three hydrogen atoms arranged in a ring. It can be used to demonstrate the meaning of bond lengths, bond angles, and torsion angles in molecular geometry.

### 3.1.1.1. The bond length

The bond length is the distance between the nuclei of two atoms bonded together [7], and it is typically expressed in Angstroms (Å), which is equal to $10^{-10}$ meters, or in picometers (pm), which is equal to $10^{-9}$ meters. For instance, the distance is shown as a green dashed line in Figure 3.1. The considered bond lengths in cyclopropene are C1–C2, C1–C3, C2–C3, C1–H4, C2–H6, C2–H7, and C3–H5. The bond length is influenced by various factors, including the types and sizes of the atoms involved, the presence of surrounding functional groups, the electronic configuration of the atoms, hydrogen bonding, and solvent effects.

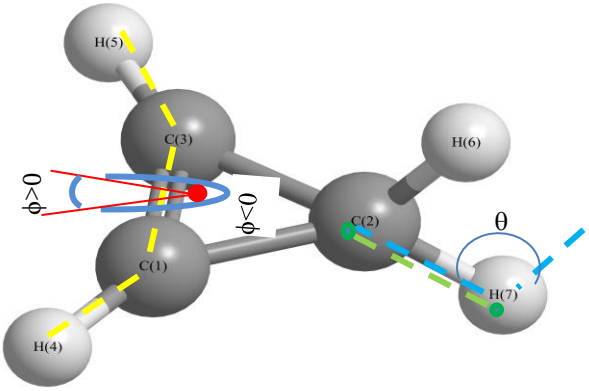



**Figure 3.1. A demonstration of what the terms used to describe molecular geometry mean.**
As an illustration, the cyclopropene molecule is considered. There are seven different bond lengths, and they are as follows: C1–C2, C1–C3, C2–C3, C1–H4, C2–H6, C2–H7, C3–H5. There are ten bond angles, listed as follows: C1–C2–C3, H6–C2–H7, C1–C2–H7, C1–C2–H6, C2–C1–C3, C2–C1–H4, C3–C1–H4, C1–C3–C2, H5–C3–C1, H5–C3–C2. There are five torsion angles, also known as dihedral angles, and they are as follows: H4-C1-C3–H5, H4-C1-C2–H6, H4-C1-C2–H7, H5-C3-C2-H7, H5-C3-C2-H6. The segments that are dashed are representative lines that show the bond length (shown in green), the bond angle (shown in blue), and the torsion angle (shown in yellow).



© University of Venda

### 3.1.1.2. Bond angle

The bond angle is the geometric angle formed by three consecutive atoms, where two successive bonds originate from the same atom, and it is an important parameter that describes the spatial arrangement of the atoms in the molecule. When dealing with more complex molecules that contain a large number of atoms, the bond angles are specified for each set of three atoms that are bonded in succession. For instance, the bond angle in the cyclopropene molecule, are denoted by the notation C1–C2–C3, H6–C2–H7, C1–C2–H7, C1–C2–H6, C2–C1–C3, C2–C1–H4, C3–C1–H4, C1–C3–C2, H5–C3–C1, and H5–C3–C2.

The bond angle in a molecule is influenced by several factors, including the electronic geometry of the molecule, the presence of lone pairs of electrons on the central atom, steric hindrance, hydrogen bonding and solvent effects.

### 3.1.1.3. Dihedral angle

The dihedral angle, also known as the torsion angle, is an angle between two planes in a molecule. It is defined by four atoms that are bonded in sequence, with the angle being measured between the planes formed by the first three atoms and the last three atoms in the chain [8]. This angle is defined for groups of four atoms that are subsequently bonded into a chain. When dealing with molecules that contain a large number of atoms, dihedral angles are determined for each set of four atoms that are bonded in sequence. For instance, in the cyclopropene molecule, one may take into consideration the four atoms in the sequence H4–C1–C3–H5 (shown by the yellow dashed segment line in Fig 3.1). The dihedral angle would be the angle formed between the plane containing atoms H4, C1, and C3 and the plane containing atoms C1, C3, and H5. The angle can be defined as the angle that exists between the C3–H5 and C1–H4 bonds when those bonds are projected into the plane that bisects the C1–C3 bond (the red lines represent the projections of the C3–H5 and C1–H4 bonds into the plane that bisects the C1–C3 bond). This angle can range from -180 degrees to +180 degrees and is usually denoted by the Greek letter such as phi ($\phi$). The sign of the angle depends on the orientation of the planes relative to each other. For example, the sign of the dihedral angle, which is represented by the letter $\omega$ in the figure, changes depending on which side of the plane (C3-H5 or C1-H4 side) is being observed. If the observer is on the C3-H5



© University of Venda

side of the plane, the dihedral angle will be positive, while if the observer is on the C1-H4 side of the plane, the dihedral angle will be negative. Other dihedral angles in the cyclopropene molecule are denoted by H4-C1-C3–H5, H4-C1-C2–H6, H4-C1-C2–H7, H5-C3-C2-H7, and H5-C3-C2-H6.

The value of the dihedral angle is influenced by several factors, such as steric interactions, electronic effects, conformational energy, hydrogen bonding, ring strain, and solvent effects. These factors can change the stability and energy of the molecule, leading to changes in the dihedral angle and also in the other parameters.

### 3.1.1.4. Internal coordinates

Internal coordinates are a set of geometrical parameters (bond lengths, bond angles, and torsion angles) that describe the position of each atom in a molecule relative to the other atoms in the same molecule [9]. They can also be used to construct molecular models and simulate the behavior of molecules in chemical reactions. Internal coordinates are typically used in contrast to cartesian coordinates, which describe the position of each atom in a molecule relative to a fixed origin in three-dimensional space, typically represented as x, y, and z coordinates. Internal coordinates can be more challenging to visualize and manipulate than cartesian coordinates. Therefore, a combination of both coordinate systems may be used depending on the specific application and the complexity of the molecule being studied [10].

The Z-matrix notation is a commonly used way to represent the internal coordinates of a molecule. The Z-matrix makes a hierarchical list of the atoms in the molecule [11]. Each atom is described by its distance from a reference atom, its angle with two reference atoms, and its dihedral angle with three reference atoms. The reference atoms can be either the preceding atoms in the Z-matrix or a fixed point in space. The Z-matrix notation is useful because it makes it easy to describe the internal coordinates of a molecule clearly and concisely. Additionally, the Z-matrix can be easily converted into cartesian coordinates for mathematical calculations. However, it is worth noting that the Z-matrix notation can become quite lengthy for large and complex molecules, and other methods may be preferred in these cases [12]. The Z-matrix is a way of representing the structure



© University of Venda

of a molecule using a table by showing how atoms are arranged in a molecule. For example, a Z-matrix of the ethene molecule ($C_2H_4$) is shown in Table 3.1 [13]:

Table 3.1. Illustration of the Z-matrix of the ethene molecule.

| Atom number | Atom type | Distance (Å) | Bond angle (º) | Torsion angle (º) |
|---|---|---|---|---|
| 1 | C | | | |
| 2 | C | 1.31 | | |
| 3 | H | 1.07 | 121.5 | |
| 4 | H | 1.07 | 121.5 | 180.0 |
| 5 | H | 1.07 | 121.5 | 180.0 |
| 6 | H | 1.07 | 121.5 | 180.0 |

Each row of the table represents an atom. The first and second columns show the number and type of the atom, respectively. In this example, the molecule is ethene, which has two carbon atoms (numbered 1 and 2) and four hydrogen atoms (numbered 3 to 6). The third column shows the distance between the atom and its nearest neighbor. For example, atom 2 is 1.31 atomic units away from atom 1, which is approximately the length of a carbon-carbon double bond. The distance between atom 3 and atom 1 is 1.07 angstroms, which is approximately the length of a C-H bond. The fourth and fifth columns show which atoms are involved in the bond and torsion angles, respectively. The fourth column of the Z-matrix shows the bond angle between atoms. For example, the bond angle between atoms 6, 2, and 1 is 121.5 degrees. The bond angle between atoms 4, 1, and 2 is also 121.5 degrees. The fifth column of the Z-matrix shows the torsion angle between atoms. For example, the torsion angle between atoms 6, 2, 1, and 4 is 180 degrees. This means that the hydrogen atoms and carbon atoms in the molecule lie in the same plane, and the molecule has a planar structure.

### 3.1.2. Molecular conformations and potential energy surfaces

By rotating around a single bond, a molecule can switch from one geometry to another of its possible configurations, as shown in Figure 3.2 [14]. There are minima on the surface of the



© University of Venda

potential energy that correspond to some of these geometries. A potential energy surface (PES) shows how the energy of a molecule is related to its geometry. This can be done either graphically or mathematically.

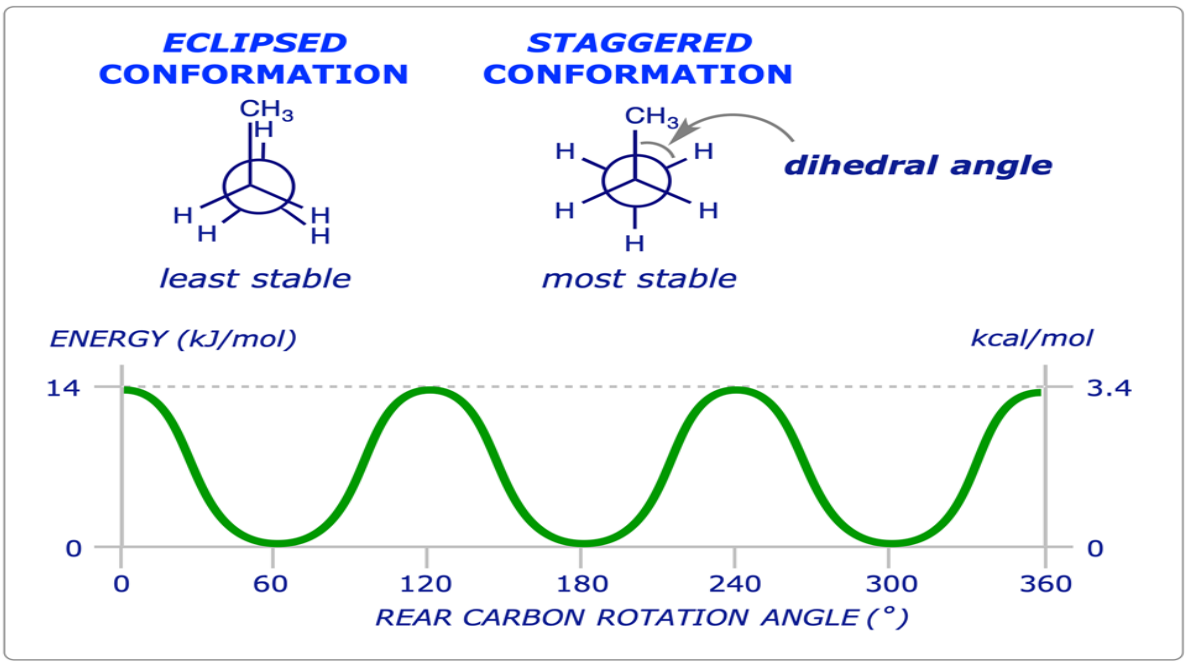



**Figure 3.2: A Conformation illustration.** Propane can take on either an eclipsed (shown on the left) or staggered (shown on the right) conformation, which is determined by the direction in which it rotates around the single carbon-carbon bond. When viewed end-on in a Newman projection, the conformation that has the least amount of energy and is, therefore, the most stable is the one in which all of the bonds are as far apart from each other as is physically possible. The conformation with the highest energy is the one that is the least stable (eclipsed). This occurs when any two adjacent carbon atoms have their six bonds (five C-H and one C-C) in the closest possible proximity. Every other conformation can be found somewhere between these two limits [14].

The degrees of freedom (internal coordinates) of a molecule can be expressed as either 3N−6N or 3N−5N, depending on the type of molecule considered [15]. For a linear molecule with N atoms, the degrees of freedom are 3N−5N, while for a nonlinear molecule, they are 3N−6. Each internal coordinate can be varied to obtain the PES [16]. The PES would need one more coordinate to represent the energy in addition to the 3N − 6 internal coordinates. However, since we can only express schematics in two or three dimensions, we can only change one coordinate at a time to get a two-dimensional PES and we can only change two coordinates at a time to get a three-dimensional PES. This is because we can only express schematics in two or three dimensions. In



© University of Venda

a molecular system, atoms are not static; they are in constant motion. The positions of atoms in the system can be represented by a vector r, which may include both Cartesian coordinates and inter-atomic distances and angles. The value of V(r) represents the potential energy of the system as a function of atomic positions, and it is calculated for all relevant configurations of r. While V(r) captures the potential energy, it does not include the kinetic energy associated with atomic motion. Since atoms are in constant motion, their kinetic energy must also be considered when calculating the total energy of the system. Thus, the potential energy surface reflects the energy landscape of the system, illustrating how the potential energy changes with the configuration of atoms. However, to fully understand the total energy of the system, both the potential and kinetic energy must be considered. The PES specifically focuses on the potential energy because it provides insight into the stable configurations and reaction pathways of the system.

To calculate the PES, we must first find a way to minimize the amount of energy associated with a specific coordinate. In light of this fact, the first derivative of the energy with respect to the selected coordinate needs to have a value of zero for us to be able to arrive at the PES. Obtaining stationary points is possible with this method. The stationary points may correspond to one of three distinct types (fig. 3.3): a minimum, a saddle point, or a maximum, depending on the context in which they are used. The point known as the saddle point is the position that lies in the midst of a minimum and a maximum. The geometries that are deemed to be in equilibrium correspond to a minimum in the set of values being studied. The saddle points of the first order are a representation of the transition stages (the transition from one geometry to another or the transition state during a chemical reaction if one is considering how the energy changes during a reaction). Adjusting parameters like bond angles or dihedral angles by a predetermined amount can generate a PES.

Geometries that correspond to minima on the PES of a molecule are known as conformers. In the case of a reaction PES, the minima correspond to the reactants and products. Each molecular configuration, whether a conformer, reactant, or product, is associated with unique arrangements of atoms that reflect stable configurations on the PES.



© University of Venda

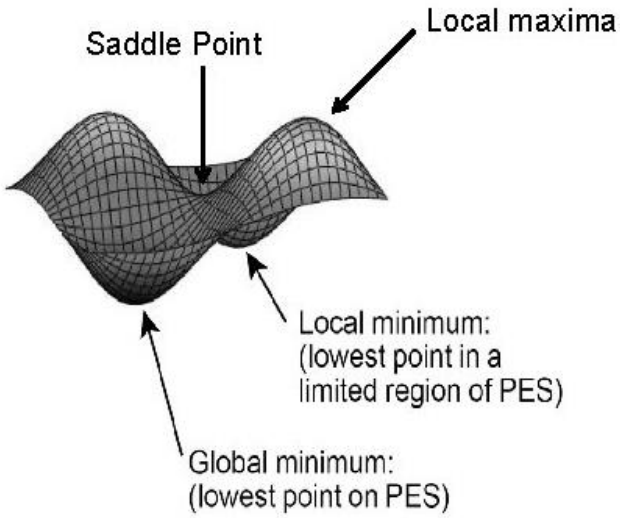



**Figure 3.3: An illustration of stationary points on the PES.** Both maxima and minima are considered stationary points because the first derivative of PES is equal to zero. The minima can not have any imaginary frequencies, and they have to show the structure of equilibrium for either the reactant or the product. The saddle point is equivalent to the transition structure; transition states should have a single imaginary frequency.

Molecules typically adopt conformations that minimize potential energy, resulting in a greater population of molecules in these low-energy states. However, classical dynamics simulations often show that many molecules exist at classical turning points rather than at energy minima, highlighting the dynamic nature of molecular behavior. The relative populations of different conformations are determined by the Boltzmann distribution, which states that the likelihood of a molecule occupying a given conformation decreases exponentially with increasing energy. This relationship can be expressed mathematically as:

$$N_i \propto e^{-\frac{E_i}{kT}} \ , \tag{3.1}$$

where $E_i$ is the energy of the conformation, $k$ is the Boltzmann constant, and $T$ is the absolute temperature. As temperature rises, more molecules can access higher-energy conformations, allowing for a significant presence of these states alongside the more stable, low-energy conformations. The behavior of a molecular ensemble is largely determined by the various



© University of Venda

conformations of its molecules. The properties of a molecule are typically evaluated based on the conformers with the lowest energy, as these are the most abundant. Different conformations may correlate with varying activities of the molecule, but those at the lowest energy states are generally the most relevant for its biological or chemical behavior. Therefore, before investigating other molecular properties, it is crucial to identify the conformations in which the molecule is most stable. This process is known as a "conformational search," which aims to explore and determine the energetically favorable conformers of the molecule.

### 3.2. Approaches to conformational analysis

As mentioned in section 3.1.2, achieving energy minimization involves moving the molecule from its initial (trial) geometry to the geometry that offers the lowest possible local energy level. To accomplish this, two primary conformational search methods are commonly used: the systematic search in torsional space and random search methods, which include approaches like Metropolis Monte Carlo [17]. Both are discussed in this study.

The systematic conformational search uses a methodology that is, in principle, simple to understand and perform. A grid search that is consistent throughout torsional space is created as a result of the generation of new structures by using every conceivable combination of torsional angle values at a predetermined amount (angle increment). When there is only one dimension present (one rotatable bond) or when there are two dimensions present, torsional driving is an option that can be utilized (two rotatable bonds). The dihedral value is kept in its initial state after adding a user-defined value to each torsional angle. At the same time, the molecule's overall energy is minimized in relation to its various degrees of freedom. This occurs after the user-defined value is added to each torsional angle. The output of a calculation that makes use of torsional driving is referred to as a potential energy curve or a potential energy (conformational) map. This map shows areas with the lowest possible energy and energy barriers and pathways that lead to different conformational states. Because it takes a significant amount of time to perform this kind of calculation, it is only ever used from time to time for problems that involve more than two rotatable bonds. Because of the difficulties described above in dealing with conformational or torsional interactions, torsional energy minimization is not usually done for molecules with more than two rotatable bonds.



© University of Venda

All conformational search methods aim to identify the lowest-energy conformation for a specific molecule given all of its torsional angles. Conformational search methods aim to identify energy minima when more than two rotatable bonds are involved. In random conformational search methods, random numbers are used to determine how many torsional angles are to be incremented. In this method, the conformations generated are ranked according to the energy values determined using a force field. This is done to find the best conformation for the given input. The trial conformation generated in this manner by the computer program is then subjected to an energy minimization process, and the conformation produced as a result is compared with those that have already been stored. If it is a new conformation, the storage will be updated to include it. The process is carried out an extremely high number of times, and the thoroughness of the search can be determined by counting the number of times each stored conformation has been located. It is worth mentioning that the computer program searches in a very deep way, which makes it far more likely to locate a new conformation than would be the case if some quick heuristic approaches were adopted. This method has the advantage that it can be stopped and started over at any time. If needed, the results from multiple runs can be combined into a single set of findings to have a more in-depth conversation about conformational search methods.

### 3.3. Computational approaches for the study of molecules

When it comes to the study of molecules from a theoretical perspective, there are several primary ways of approaching different methods that can be utilized to calculate the conformational properties of molecules. The two well-known primary approaches are molecular mechanics, force field methods, and quantum mechanical methods. The following sections will give an overview of molecular mechanics, providing an insight into molecular mechanics models. Molecular mechanics and quantum chemical models are also compared and contrasted.

#### 3.3.1. Molecular mechanics methods

As previously stated, molecular mechanics methods are also known as force field approaches. Classical physics equations are used to calculate bond stretching, angle bending, torsional energy, and other non-bonded properties. This yields various interactions and energies, which are also referred to as force fields [18]. A force field is a mathematical equation that shows how the energy



© University of Venda

of a system depends on the positions of its particles [19]. It comprises a set of parameters and an analytical form of the interatomic potential energy. The parameters are usually derived from quantum mechanical calculations (ab initio or semi-empirical calculations) or experimental data such as neutron spectroscopy, infrared, Raman, nuclear magnetic resonance (NMR), neutron, X-ray, and electron diffraction, and so on [19].

Molecular mechanics uses the term "bonded atoms" to describe molecules that have been deformed from their idealized geometry by van der Waals and Coulombic interactions that are not bonded [20]. In molecular mechanics, a molecule is viewed as a group of atoms held together by classical forces. Potential energy functions of structural features such as bond lengths, bond angles, torsional (dihedral) angles, and so on are used to describe these forces. Based on this theory, atoms are assumed to be "hard" spheres connected through spring networks and have separate force constants [17]. In most cases, electrons are not depicted directly. However, there are a few notable exceptions to this rule. The molecular mechanics energy of a molecule is expressed as the sum of contributions that arise from deviations from idealized geometrical parameters [20]. These include distortions in bond lengths (stretching), bond angles (bending), and torsion angles (torsional strain). In addition, it incorporates contributions from non-bonded interactions, such as van der Waals forces and Coulombic (electrostatic) interactions. This energy is commonly referred to as "strain energy,"

$$E^{strain} = \sum_{i}^{bonds} E_i^{stretch} + \sum_{i}^{bond\ angles} E_i^{bend} + \sum_{i}^{torsion\ angles} E_i^{torsion} + \sum_{i}^{non-bonded\ atoms} \sum_{j} E_{ij}^{non-bonded}, \quad (3.2)$$

and it is written as the sum of the structural energy (first three terms) and the Coulomb energy (last term), thus, representing the inherent strain in a real molecule when compared to an idealized, stress-free structure [20].

In equation 3.2, the first three summations are due to deviations in bond lengths, bond angles, and torsional angles of bonded atoms and can also be expressed as,



© University of Venda

$$E_{bonded} = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} \sum_{n=1}^{6} k_{\emptyset,n}(1 - \cos(n\emptyset - \delta_n)) + \sum_{improper\ torsions} k_\varphi(\varphi - \varphi_0)^2 \quad (3.3)$$

where $b_0$ is the standard bond length used as a reference

$K_b$ is a bond force constant

$\theta_0$ represents a reference valence angle

$K_\theta$ is an angle force constant

$\varphi_0$ is an improper torsion angle reference value

Kφ is an improper torsion force constant.

Hence, information about bonding is considered "part of the input" to a molecular mechanics calculation, whereas in a quantum chemical calculation, this information is considered "part of the output." The expression in equation 3.2 illustrates a type of potential energy function referred to as a Class I additive potential energy function. Most of the biomolecular force fields used today share the terms in equation 3.2 that make up the energy and are part of the equation. These fields include, among others, CHARMM, AMBER, GROMOS, and OPLS [21].

The final sum in equation 3.2 is found by adding up the contributions of all the atoms in the system that are not covalently bound to each other, that is,

$$E_{nonbonded} = \sum_{nonbonded\ pairs\ ij} \frac{q_i q_j}{4\pi D r_{ij}} + \sum_{nonbonded\ pairs\ ij} \varepsilon_{ij} \left[ \left( \frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \quad (3.4)$$

where $n$ is a torsion multiplicity

$\delta_n$ is a torsion phase

$K_\phi,n$ is a torsion amplitude

$q_i$ and $q_j$ are a partial charges

$R_{min,ij}$ is a Lennard-Jones radius

$\varepsilon_{ij}$ is a Lennard-Jones well depth.



© University of Venda

It is common practice to add the van der Waals (VDW) interactions and the Coulombic repulsion (CR) to account for all of the other non-bonded interactions. Van der Waals interactions are almost always expressed as the sum of repulsive and attractive components. This is the most common way to do so. The Coulomb interactions between fixed point charges easily account for the electrostatics of class I force fields $q_i$ and $q_j$ that are centered on the atoms, as shown in equation 3.4; these are also called "partial charges."

Equation 3.2 also might contain some extra non-bonded terms to account for situations where interactions like hydrogen bonding must be explicitly counted. More sophisticated force fields may also include higher-order terms, cross terms (like stretch-bend, bend-bend, and torsion-stretch), and terms that account for environments that are not symmetrical. Many force fields, also known as Class II force fields, have higher order terms to handle bond and valence angle terms and/or cross terms between, for example, bonds and valence angles and dihedrals [22-25]. Additionally, a variety of force fields contain these higher-order terms. Optimizing force fields allows for the performance of MD simulations and more static computations like energy minimization [20]. They can be set up so that they can reproduce the conformational behaviour of single molecules or the conformational behaviour of interactions between molecules. They can also be made more specific for gas phase computations or for simulations in a condensed phase. In gas-phase vs. solution-phase simulations, parameter adjustments are often required to account for the different interaction environments. For solution-phase, parameters like the dielectric constant are altered to represent solvent effects such as screening and hydrogen bonding, while gas-phase settings focus on isolated molecule interactions.

Energy minimization is an important part of the molecular mechanics method, which calculates energy as a function of the nuclear coordinates. Thus, during the energy minimization process, computer visualizations are first used to create a trial 3D-structure of a molecule. Using an energy minimization strategy, the atoms are shifted without breaking bonds until no atoms have net forces, and the molecule's energy is lowest. This energy minimum corresponds to one of the stable conformations of the molecule, although it is not always the most stable conformation. Due to the inability of energy minimization methods to overcome energy barriers, minimizing a trial molecule



© University of Venda

continues until the first local energy minimum is obtained. By doing the calculation again with a different starting geometry or, more efficiently, by using a conformational search technique, you can find several local energy minima, including the one with the least energy, which is called the global energy minima

Molecular mechanics methods can be distinguished from one another in different ways, such as by the number of terms they use, the kind of terms they use, and the way they are parameterized. As a result, various software applications may use distinct functional variations of the force field. Some examples of MM force fields include MM2, MM3, MMFF, Amber, CFF93, and UFF [26]. It is important to remember that the absolute values of the energies obtained using the various force field approaches are meaningless. For programs using the same force field, discrepancies can still arise due to differences in implementation, integration methods, or cutoff distances. Although two programs might theoretically give identical results, slight setup or numerical handling variations can lead to minor differences. They are only helpful for comparing values obtained for various geometries of the same system. Except in closely related systems, not only should figures for energies obtained from one program not be compared to those obtained from another program, but the energy calculated for a molecule in one program should not be compared to the energy calculated for another molecule in the same program [27].

The following computational techniques frequently make use of molecular mechanics: Molecular dynamics (MD), which models the quantum mechanical properties of matter by simulating the dynamic behaviour of molecules (using, among other programs, Amber [28], CHARMM [29], MM3 [30], and Gromos [31]); protein folding, which refers to techniques for predicting the three-dimensional structure of proteins based on their sequences; protein-ligand docking—methods for predicting protein-ligand binding energy (AutoDock [32], DOCK [33], Glide [34], ICM [35], etc.); protein-protein or protein-DNA/RNA docking—methods for predicting protein-protein or protein-protein or protein-DNA/RNA binding energy.

The field of computational biology has made remarkable progress in recent years, owing to the development of various simulation methods. One such method is molecular mechanics, which finds application in many computational techniques, including MD. MD is a powerful simulation



© University of Venda

method that models the dynamic behavior of molecules by projecting their quantum mechanical characteristics. It uses programs such as Amber, CHARMM, MM3, and Gromos to perform simulations.

Protein folding is another critical process in computational biophysics that predicts the three-dimensional configuration of proteins based on their sequences. This process is important as the three-dimensional shape of a protein is responsible for its function. Several methods are available for protein folding, including ab initio, homology modeling, and MD simulations.

Protein-ligand docking is yet another method that estimates protein-ligand binding energy. It involves predicting the binding affinity between a protein and a ligand molecule. This estimation is accomplished by programs like AutoDock, DOCK, Glide, and ICM, which apply different approaches to calculate the binding energy.

Similarly, protein-protein or protein-DNA/RNA docking estimates protein-protein or protein-DNA/RNA binding energy. Various programs like ZDOCK, RosettaDock, HADDOCK, and ClusPro [36] are available to perform these calculations, using different algorithms to estimate the binding energy and predict the orientation of the interacting molecules. Besides, reactive force field (ReaxFF, [37]) MD is a simulation method between quantum mechanics and classical molecular dynamics. It uses a reactive force field to simulate chemical reactions in complex systems, making it a useful tool in computational biology.

#### 3.3.1.1. Limitation of molecular mechanics methods

Molecular mechanics is an efficient method and has broad applicability. However, it does have several important drawbacks, the most important are as follows [19, 27, 38]:

- Molecular mechanics can not provide any information about the electronic structure and, as a result, can not handle processes such as the breaking and formation of covalent bonds, electron excitations, charge transfers, protonation states, etc.
- It can not be used to analyse features dependent on the distribution of electrons, such as NMR shielding constants or spectroscopic data, etc.



© University of Venda

- Force fields generally are unable to describe stacking interactions appropriately.
- The precision of the procedure is relatively limited and significantly less than that of QM calculations.
- The procedure of parameterisation must first be completed to obtain the empirical parameters needed to use the approach.
- Most of the time, the predictions of MM energy are meaningless when given as absolute numbers. They are usually only useful when compared to studies of the same system.

### 3.3.2. Quantum mechanical methods

#### 3.3.2.1. Electronic structure methods

Electronic structure methods are computational techniques used to study the electronic properties of molecules and materials [39]. These methods are based on quantum mechanics and employ mathematical models to calculate the distribution of electrons within a system. The primary goal of these methods is to determine the electronic wavefunction (ψ), which describes the spatial distribution of electrons, their energies, and their probabilities in different orbitals [38].

##### 3.3.2.1.1. The Schrödinger equation

The Schrödinger equation in quantum mechanics defines the wave function. There are two different equations, the time-independent and the time-dependent Schrödinger equation.

###### 3.3.2.1.1.1. Time-dependent Schrödinger equation:

The time-dependent Schrödinger equation governs the evolution of a quantum system's wavefunction over time [20]. It provides a complete description of the system's dynamics by accounting for how the wavefunction changes due to the influence of kinetic and potential energies and can be written as,

$$i\hbar \frac{\partial \psi}{\partial t}(x,t) = H\psi(x,t), \quad (3.5)$$

where, ψ(x,t) is the time-dependent wavefunction,

H is the Hamiltonian operator, representing the system's total energy (H = T + V, where T is the kinetic energy, and V is the potential energy),



© University of Venda

i is the imaginary unit,

$\hbar$ is the reduced Planck's constant,

t is time.

To derive the time-dependent Schrödinger equation, the classical energy of a system is given as:

$$E = T + V = \frac{p^2}{2m} + V(x), \tag{3.6}$$

where p is the momentum, m is the particle's mass, and V(x) is the potential energy.

In quantum mechanics:

$$p \rightarrow -i\hbar \frac{\partial}{\partial x}, \tag{3.7}$$

$$T = \frac{p^2}{2m} \rightarrow -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}. \tag{3.8}$$

Substituting these operators into the Hamiltonian, the Hamiltonian operator becomes:

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x). \tag{3.9}$$

Substituting H into the time-dependent Schrödinger equation, we obtain the general form:

$$I\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + V(x,t)\,\psi(x,t). \tag{3.10}$$

**3.3.2.1.1.2. Time-independent Schrödinger equation:**

For systems with constant energy, the wavefunction can be expressed as the product of a spatial part and a time-dependent phase factor. This leads to the time-independent Schrödinger equation, which describes the spatial properties of stationary states.

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x)}{\partial x^2} + V(x)\psi(x) = E\psi(x). \tag{3.11}$$

Derivation from the time-dependent Schrödinger equation:



© University of Venda

Assume a separable form of the wavefunction:

$$\psi(x,t) = \psi(x)e^{-\frac{iEt}{\hbar}}. \tag{3.12}$$

Substitute $\psi(x,t)$ into the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\left(\psi(x)e^{-\frac{iEt}{\hbar}}\right) = H\left(\psi(x)e^{-\frac{iEt}{\hbar}}\right). \tag{3.13}$$

The time derivative yields:

$$i\hbar\frac{\partial}{\partial t}\left(\psi(x)e^{-\frac{iEt}{\hbar}}\right) = E\psi(x)e^{-\frac{iEt}{\hbar}}. \tag{3.14}$$

Cancelling the time-dependent term $e^{-\frac{iEt}{\hbar}}$, we obtain:

$$E\psi(x) = H\psi(x). \tag{3.15}$$

Substitute the Hamiltonian operator:

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x))\psi(x) = E\psi(x). \tag{3.16}$$

Rearranging gives:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x)}{\partial x^2} + V(x)\psi(x) = E\psi(x). \tag{3.17}$$

The time-dependent Schrödinger equation provides a full description of quantum systems, incorporating changes in both spatial and temporal behavior. The time-independent Schrödinger equation is a particular case describing systems in stationary states, where the energy remains constant.

**3.3.2.1.1.3. Born-Oppenheimer approximation and variational principle**

The Born-Oppenheimer (BO) [40] approximation is a fundamental concept in quantum chemistry and molecular physics that simplifies the complex problem of solving the Schrödinger equation



© University of Venda

for molecules. It simplifies the problem by assuming that the nuclei are much heavier and move more slowly than electrons. This allows the separation of the motions of electrons and nuclei, where the nuclei are treated as fixed points during electronic calculations, and the electronic wavefunction is determined accordingly. This approximation significantly reduces the complexity of quantum mechanical computations, making them feasible for larger molecular systems.

The BO approximation is applied in several widely used electronic structure methods, such as Hartree-Fock (HF), Density Functional Theory (DFT), and various ab initio methods [20]. These methods compute the electronic structure at fixed nuclear positions, simplifying the computational burden. While the nuclear motion is typically considered separately, the BO approximation remains essential for the efficient treatment of molecular systems, allowing for practical calculations of molecular properties like bond energies, reaction dynamics, and electronic excitations. However, the BOA has several limitations, primarily when nuclear and electronic motions become coupled, as in conical intersections where potential energy surfaces of different electronic states intersect. The approximation struggles in systems with strong electron-nuclear correlations, such as in excited states or reactions involving light atoms and fails for heavy nuclei where nuclear-electronic coupling becomes significant.

Another approach to solving the complex problem of the Schrödinger equation for systems with many interacting particles is the variational principle. The variational principle is a method in quantum mechanics used to approximate the ground state energy of a system. It states that the energy calculated with any trial wavefunction, $\Psi_{trial}$, is always an upper bound to the true ground state energy, $E_0$.

Mathematically expressed as:

$$E_{trial} = \frac{\langle \Psi_{trial} | \hat{H} | \Psi_{trial} \rangle}{\langle \Psi_{trial} | \Psi_{trial} \rangle} \geq E\_0, \quad (3.18)$$

where $E_0$ is the true ground-state energy.

- Trial wavefunction ($\Psi_{trial}$):

A guess for the actual wavefunction of the system, which depends on adjustable parameters.



© University of Venda

- Optimization:By minimizing $E_{trial}$ with respect to these parameters, the trial wavefunction approaches the true ground-state wavefunction, and the energy estimate improves.

### 3.3.2.1.1.4. Slater determinant

A Slater determinant is a mathematical construct commonly used in quantum chemistry to represent the wavefunction of multi-electron systems, ensuring the correct antisymmetry required for fermions. Fermions, such as electrons, adhere to the Pauli exclusion principle, which states that no two fermions can simultaneously occupy the same quantum state. To apply this principle, the Slater determinant is designed so that if two electrons are exchanged, the wavefunction changes sign, reflecting the antisymmetry of the system.

In the context of a Slater determinant, the total wavefunction of a system with N electrons is expressed as a combination of single particle wavefunctions, each representing one electron's spatial position and spin. This construction leads to a multivariable wavefunction that includes contributions from each electron's quantum state. Mathematically, the Slater determinant for N electrons is given by:

$$\Psi = \frac{1}{\sqrt{N}} \begin{vmatrix} \chi_1(\mathrm{X}_1) & \chi_2(\mathrm{X}_1) & \cdots & \chi_N(\mathrm{X}_1) \\ \chi_1(\mathrm{X}_2) & \chi_2(\mathrm{X}_2) & \cdots & \chi_N(\mathrm{X}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathrm{X}_N) & \chi_2(\mathrm{X}_N) & \cdots & \chi_N(\mathrm{X}_N) \end{vmatrix}. \tag{3.19}$$

Where $\chi_i(\mathrm{X}_k)$ represents the spatial and spin wavefunction of the *i*-th electron in the system, and the determinant ensures the antisymmetry required by the Pauli exclusion principle. This form is crucial for quantum mechanical methods like Hartree-Fock theory, where the overall wavefunction of the electrons is approximated by a single Slater determinant of molecular orbitals [41, 42].



© University of Venda

### 3.3.2.2. Schrödinger equation-based methods

#### 3.3.2.2.1. The Hartree-Fock (HF) method

The Hartree-Fock (HF) method is a self-consistent field method that is used to approximate the molecular wave function of a system [43]. The method starts with an initial guess of the molecular wave function, which is then used to calculate the molecular electron density. This density is then used to calculate an updated set of molecular orbitals, which are used to update the molecular electron density until the density no longer changes [43].

The Hartree-Fock method starts with the Schrödinger equation, which describes the behavior of a system of N electrons in a molecule. The Schrödinger equation can also be written as [39],

$$H\Psi = E\Psi, \tag{3.20}$$

where H is the Hamiltonian operator of the system, Ψ represents the many-electron wave function, and E is the total energy. Here, Ψ (1, 2,…, N) is the complete wave function that describes all N electrons.

The Hartree-Fock method represents the wave function as a Slater determinant, a linear combination of molecular orbitals. The molecular orbitals are calculated by solving the Fock equation [43],

$$F\Phi_i = \varepsilon_i\Phi_i, \tag{3.21}$$

where F is the Fock matrix, $\Phi_i$ is the i-th molecular orbital, and $\varepsilon_i$ is the orbital energy.

In Hartree-Fock (HF) molecular orbital theory, the Fock matrix element $f_{i,j}$ is expressed as:

$$f_{i,j} = h_{i,j} + \sum_k [(ij \mid kk) - (ik \mid kj)]P_{k,l} \tag{3.22}$$

which is derived from the electron density and the Coulomb and exchange integrals [43];

where, $h_{i,j}$ represents the core Hamiltonian matrix element, which accounts for the kinetic energy of electrons and the nuclear attraction,



© University of Venda

$(ij \mid kk)$ and $(ik \mid kj)$ are two-electron integrals describing electron-electron repulsion, $P_{k,l}$ is the electron density matrix element, representing electron distribution in molecular orbitals.

This equation assumes molecular orbitals are orthonormal, a standard condition in HF theory. Orthonormality simplifies calculations by ensuring that the overlap between different molecular orbitals is zero, facilitating the accurate determination of electronic properties and total molecular energy.

In the HF method, the molecular wavefunction is represented as a product of one-electron wavefunctions, called orbitals. These orbitals are determined by solving the Hartree-Fock equations, which are a set of non-linear integral-differential equations. The solutions to the Hartree-Fock equations are self-consistent in that the electron density, derived from the wavefunction, is used to optimize the potential energy and the orbitals, and the process is repeated until convergence.

The self-consistent field (SCF) process in the Hartree-Fock (HF) method is a central step in determining the wavefunction of a molecule. The goal of the SCF process is to find the molecular orbitals and the electron density that minimize the energy of the system [44].

The SCF process starts with an initial guess for the molecular orbitals, which are used to calculate the electron density. The electron density is then used to update the potential energy of the system, and the updated orbitals are used to calculate a new electron density. This process is repeated until the energy of the system and the electron density converge to a self-consistent solution, meaning that the orbitals and electron density are no longer changing. In each iteration of the SCF process, the electron density is used to calculate the Fock matrix, which represents the potential energy of the electrons due to the electrostatic interactions with the nuclei and the electron-electron interactions. The Fock matrix is used to update the molecular orbitals, which are then used to calculate a new electron density.

The SCF process is converged when the difference between the energy and electron density from the previous iteration and the current iteration is smaller than a specified limit [45]. At this point,



© University of Venda

the self-consistent solution of the molecular orbitals and electron density has been found, and the energy of the system is minimized. The Hartree-Fock method provides a reliable approximation of the molecular wave function and is widely used in quantum chemistry to study the properties of molecules [43, 46, 47]. However, it is limited by its lack of ability to account for electron correlation, an important aspect of electronic structure that affects the chemical reactivity and physical properties of molecules. To overcome this limitation, post-HF methods, such as configuration interaction (CI), Møller-Plesset perturbation theory, and density functional theory, have been developed to correct for electron correlation (the interactions between electrons in a molecule that are not accounted for by the Hartree-Fock method, which treats electrons independently) and improve the accuracy of the predictions [43, 46, 47].

#### 3.3.2.2.2. Post-Hartree-Fock methods

Post-Hartree-Fock (HF) methods are electronic structure calculation methods that build upon the results of a Hartree-Fock calculation by including electron correlation effects that were neglected in the Hartree-Fock calculation. In HF, the electronic wavefunction is represented as a single Slater determinant, a mathematical representation of the distribution of electrons in different orbitals. However, this approach neglects electron-electron correlation effects, which are important for accurately describing the electronic structure of molecules.

Examples of Post-Hartree-Fock (HF) methods include:

- **Configuration interaction (CI):** a wavefunction-based method that includes electron-electron interactions beyond the mean-field approximation of Hartree-Fock.
- **second-order Moller-Plesset perturbation theory (MP2):** a method that considers electron-electron interactions through a perturbation series expansion of the electron-electron interaction energy.
- **Coupled-cluster (CC):** a method that includes electron-electron interactions through a series expansion of the cluster operators. The method can be implemented at various levels of approximation, such as Coupled-cluster with single and double excitations (CCSD) and Coupled-cluster with triple excitations (CC3).
- **Density functional theory (DFT):** a method that calculates the electronic structure of a molecule based on the electron density rather than the wavefunction. The method provides a more computationally efficient alternative to wavefunction-based methods.



© University of Venda

- **Quantum Monte Carlo (QMC):** a statistical sampling method that can be used to calculate the electronic structure of a molecule.

### 3.3.2.2.2.1. Configuration Interaction (CI)

The CI method is one of the most straightforward approaches to addressing electron correlation. It is used to address the limitations of the Hartree-Fock (HF) method by taking into account the interactions between electrons. This is achieved by constructing the wavefunction as a linear combination of multiple Slater determinants, each representing a different configuration of electrons in different orbitals [41, 43].

To obtain an unnormalized CI expansion of an N-electron wave function, we can first utilize the self-consistent field (SCF) wave function and orbitals. This involves restricting the Hartree-Fock solution to the region covered by a designated basis set. Thus, the CI wavefunction can be written as follows;

$$\Psi = \Phi_{SCF} + \sum_{i,a} C_i^a \Phi_i^a + \sum_{i<j,a<b} C_{ij}^{ab} \Phi_{ij}^{ab} + \cdots \ (up\ to\ N\ excitations), \tag{3.23}$$

where, $\Phi_i^a$ is a singly excited configuration where $\phi_i$ an occupied orbital and $\phi_a$ is an unoccupied orbital,

$\Phi_{ij}^{ab}$ is a double-excitation configuration,

$C_i^a$ and $C_{ij}^{ab}$ are the corresponding coefficients.

The molecular orbitals $\phi_i$ and $\phi_a$ are typically found by solving the Hartree-Fock equations, where an eigenfunction of the Fock operator, $\hat{f}\phi_a = \varepsilon_a\phi_a$, can be selected as the orbital unoccupied in the SCF function. The coefficients $C_i^a$, $C_{ij}^{ab}$, are determined by minimizing the total energy of the system using a variational method [42]. This process is non-iterative in configuration interaction (CI) because the molecular orbitals (MOs) are fixed throughout the calculation. Only the coefficients are adjusted to find the optimal wavefunction that best describes the electronic structure.



© University of Venda

The steps in determining these coefficients are straightforward:

1. Start with a wavefunction that is a combination of determinants, as described earlier.
2. The determinants $\Phi_i$ are expressed using molecular orbitals, which are linear combinations of atomic orbitals.
3. Minimize the total energy by solving a linear variational problem to adjust the coefficients while keeping the MOs fixed.
4. Use the optimized coefficients to improve the overall wavefunction for the given MOs.

This process is repeated until convergence is reached, meaning the energy minimum is found and the wavefunction parameters are fixed. By including multiple Slater determinants in the wavefunction, CI accounts for the complex interactions between electrons, resulting in a more accurate description of the electronic structure of a molecule [41]. In this way, CI methods provide a more accurate description of the electronic structure of molecules, including bond strengths and dissociation energies, by taking into account electron-electron correlation effects. However, the method is computationally intensive and becomes impractical for large and complex molecular systems. To overcome these limitations, advanced CI methods, such as multi-reference CI, have been developed to improve the accuracy and efficiency of the calculations [41].

#### 3.3.2.2.2.2. Møller-Plesset perturbation theory

Since the origins of quantum chemistry, perturbation theory has aided in producing descriptions of the electronic structure of atoms and molecules that include electron correlation corrections. Back in 1934 [48], Møller and Plesset1 introduced the use of second-order perturbation theory to fix electron pair correlation within the Hartree–Fock (HF) method [49]. This became known as Møller–Plesset perturbation theory [49], commonly called MPPT or MP. Despite being ignored initially, MP played a major role in shaping the development of ab initio quantum chemical methods over the past years.

In perturbation theory, the zero-order wavefunction serves as the starting point, or unperturbed wavefunction, from which corrections are applied. In this context, the Hartree-Fock (HF) wavefunction is commonly used as the zero-order approximation. To calculate correlation energy, perturbation theory relies on the selection of an unperturbed Hamiltonian operator, which is



© University of Venda

typically constructed by summing Fock operators. These Fock operators account for the average electron-electron repulsion, but they do so twice, leading to a perturbation that equals the exact electron repulsion operator (Vee) minus twice the average electron repulsion ⟨Vee⟩. The resulting difference between these two quantities gives rise to the fluctuation potential, a key operator in perturbation theory.

This fapproximation ensures that solutions to the unperturbed Schrödinger equation are known and leads to the concept of size extensivity. Size extensivity refers to the property whereby the total energy of a system scales appropriately with the size of the system. This is in contrast to size consistency, which is a different concept that ensures the energy of two non-interacting subsystems equals the sum of their individual energies. Thus, while size extensivity ensures proper scaling with system size, size consistency ensures that energy calculations reflect the additive nature of separate subsystems.

The sum of MO energies forms the zeroth-order energy, with the HF determinant being the corresponding wave function. This can be expressed as:

$$W_0 = \langle \Phi_0 | H_0 | \Phi_0 \rangle = \left\langle \Phi_0 \right| \sum_{i=1}^{N_{elec}} F_i \left| \Phi_0 \right\rangle = \sum_{i=1}^{N_{elec}} \varepsilon_i . \quad (3.24)$$

Moving on to the first-order energy correction,

$$W_1 = \langle \Phi_0 | H' | \Phi_0 \rangle = \langle V_{ee} \rangle - 2\langle V_{ee} \rangle = -\langle V_{ee} \rangle, \quad (3.25)$$

it pertains to the perturbation operator averaged over the zeroth-order wave function. It is important to keep in mind that orbital energy takes into account the interaction of electrons with not just all the nuclei but also with each other, thus counting electron–electron repulsion twice.

The overcounting of electron-electron repulsion at zeroth order can be corrected through this method. The first-order energy ($W_0$ plus $W_1$) equals the HF energy. MPn denotes the total energy at order n, while E(MPn) indicates the correction in the same order.

This can be expressed mathematically as follows:



© University of Venda

$$MP0 = E(MP0) = \sum_{i=1}^{N_{elec}} \varepsilon_i \tag{3.26}$$

For the first-order energy, we have:

$$MP1 = E(MP0) + E(MP1) = E(HF). \tag{3.27}$$

The first contribution to the correlation energy is the second-order correction to the energy derived from a sum over doubly excited determinants. Such determinants can be generated by raising electrons from orbitals i and j to orbitals a and b, respectively [20]. However, it is important to note that each excited state can only be counted once, and the summation must be restricted accordingly; thus the second-order wave function can be written as:

$$W2 = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\langle \Phi_0 | \mathrm{H}' | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | \mathrm{H}' | \Phi_0 \rangle}{\mathrm{E}_0 - \mathrm{E}_{ij}^{ab}}. \tag{3.28}$$

The matrix elements between the HF and a doubly excited state can be calculated using two-electron integrals over the molecular orbitals, as described by equation (3.28):

$$\langle \Phi_0 | H_0 | \Phi_{ij}^{ab} \rangle = \langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle. \tag{3.29}$$

Thus, the second order Møller-Plesset correction can be explicitly expressed as:

$$E(MP2) = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\left( \langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_a \phi_b \rangle \right)^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \tag{3.30}$$

After obtaining the two-electron integrals across MOs, a summation of these integrals produces the second-order energy correction. These integrals typically number in the $M_4$ (where M denotes the number of basis functions) basis range, which increases the energy calculation proportionally with the system size [50]. However, the transformation of integrals from atomic orbital (AO) to molecular orbital (MO) basis increases at a rate of $M_5$ basis. The second-order Møller-Plesset perturbation theory (MP2) is a method that utilizes $M_5$ basis but is relatively inexpensive as it does not require all of the two-electron integrals across MOs. Only integrals that correspond to a combination of two occupied and two virtual MOs are needed. Practical calculations show that the



© University of Venda

cost of calculating MP2 energy for systems with several hundred basis functions is similar to or less than that of calculating HF energy. Typically, MP2 accounts for 80-90% of correlation energy and is the most economical method for incorporating electron correlation.

The calculation of fourth- and fifth-order energies is made possible by the second-order wave function. These energies include contributions from determinants that are excited singly, doubly, triply, or quadruply. The computational cost of fourth-order energy, MP4 (fourth-order Møller-Plesset perturbation theory, SDQ), increases as the $M_6$ basis grows. In contrast, the contribution of the triply excited determinants, resulting in the triples contribution, and increases as the $M_7$ basis grows. Despite this, MP4 remains a computationally feasible model for many molecular systems and requires a similar amount of time as Configuration Interaction Simple Singles and Doubles (CISD) [50]. The fifth-order correction to the energy involves contributions from S, D, T, and Q (single, double, triple, and quadruple excitations, respectively), and the sixth-order term introduces quintuple and sextuple excitations. The complexity of the equations required to calculate the fifth-order Møller-Plesset perturbation theory (MP5) and the sixth-order Møller-Plesset perturbation theory (MP6) contributions renders them impractical for use with larger systems. The computational effort for MP5 increases as the M8 basis grows, while the computational effort for MP6 increases as the M9 basis grows. There is little knowledge on the performance of MPn beyond MP4.

To analyze the convergence properties of the perturbation series, one can study the partitioned Hamiltonian as shown below:

$$\mathrm{H}(\lambda) = \mathrm{H}_0 + \lambda \mathrm{H}'. \tag{3.31}$$

This equation utilizes a parameter, $\lambda$, which connects the reference system (where $\lambda$ = 0) to the real physical system (where $\lambda$ = 1) [50, 51]. To analyze the convergence behavior, complex values should also be allowed for $\lambda$.

The energy as a function of λ can be expressed by the series:

$$E(\lambda) = \sum_{i=0}^{\infty} W_I \lambda^i. \tag{3.32}$$



© University of Venda

Ideally, the HF, MP2, MP3, and MP4 results should demonstrate a monotonic convergence towards a limiting value, with the corrections becoming smaller numerically and having the same sign as the order of perturbation increases [50]. Unfortunately, this is not typically the case. Even when the reference is adequately described by a single determinant, oscillations in a given property as a function of perturbation order can be observed. Cremer's analysis indicates that smooth convergence of total energy can only be expected in systems containing well-separated electron pairs; otherwise, oscillations occur. In reality, only lower orders of perturbation theory can be performed, resulting in significant differences between HF and MP2 observations, with the MP3 result closer to the HF and the MP4 moving away from the HF result.

MP2 tends to overestimate correlation effects but still yields better results than MP3, particularly when working with medium-sized basis sets. On the other hand, MP4 often overestimates the effect of singles and triple contributions since they are only included in the series starting at fourth order. If the reference wave function shows a significant multireference character, a perturbation expansion that relies on a single determinant will exhibit poor convergence. When the reference wave function experiences symmetry breaking, the MP approach is likely to yield in inadequate results. Due to its uncertain convergence, the MP method is not as widely used, although the MP2 version remains a cost-effective method for incorporating electron correlation effects. MP2 is the most economical ab initio method for correcting HF findings for correlation effects. This is also true for higher MPn methods, as long as the focus is on a specific correlation effect, such as connected three-electron (MP4) or connected four-electron correlations (MP6). One can quickly analyze correlation effects by partitioning MPn energies into S, D, T, and so on, comparing changes in response densities from order to order, or describing the dependency of other correlation-corrected response properties on *n*. Studying the correlations has made it clear that MPn has limitations, as seen in studies [52-54].

#### 3.3.2.2.2.3 Coupled-cluster (CC) Methods

Coupled-Cluster (CC) theory is a highly accurate and widely used method for incorporating electron-electron interactions in quantum chemistry calculations. It represents a significant advancement over methods like Hartree-Fock and Møller-Plesset perturbation theory by including



© University of Venda

electron correlation through a series expansion of the cluster operators, which account for electron excitation and correlation beyond the mean-field approximation [55].

The CC method is the wavefunction ansatz, written as an exponential operator acting on the Hartree-Fock wavefunction. The coupled-cluster wavefunction can be expressed as:

$$| \Psi_{CC} \rangle = e^{T} | \Phi_0 \rangle, \tag{3.33}$$

where $e^{T}$ is the exponential of the cluster operator T and $| \Phi_0 \rangle$ is the reference Hartree-Fock wavefunction. The cluster operator T is a sum of excitation operators that include single (S) and double (D) excitations and potentially higher-order excitations, depending on the level of approximation chosen [56].

$$T = T_1 + T_2 + T_3 + \cdots \tag{3.34}$$

where $T_1$, $T_2$, and higher $T_n$ correspond to single, double, and higher-order excitations, respectively. By exponentiating this operator, the CC method captures not just single- and double-excitation contributions but also the interdependencies of these excitations that account for electron correlation [55].

The most commonly used form of CC theory is the Coupled-Cluster with Single and Double excitations (CCSD) method. In this approximation, the cluster operator is truncated to include only single and double excitations:

$$| \Psi_{CCSD} \rangle = e^{T_1 + T_2} | \Phi_0 \rangle. \tag{3.35}$$

The CCSD method includes the correlation effects from both single-electron and double-electron excitations, which are essential for a highly accurate treatment of electron correlation. The energy of the system is then calculated by solving for the coupled-cluster amplitudes through a series of equations derived from the CC wavefunction ansatz. The CCSD method is known for its high accuracy and often provides results close to the full configuration interaction (FCI) solution, especially for smaller molecular systems [57].

For systems requiring even greater accuracy, Coupled-Cluster with triple excitations (CC3) includes the contributions of triple excitations. This method involves truncating the cluster operator at the third excitation level:

$$| \Psi_{CC3} \rangle = e^{T_1 + T_2 + T_3} | \Phi_0 \rangle. \tag{3.36}$$



© University of Venda

The CC3 method captures more electron correlation than CCSD by including the effects of triple excitations, making it even more accurate. However, the computational cost of CC3 is significantly higher than CCSD due to the increased complexity of solving for the triple-excitation amplitudes. Despite this, CC3 is often employed for high-accuracy calculations, especially when studying larger systems or phenomena that require precise correlation effects [58].

Coupled-cluster methods are highly praised for their accuracy and reliability in capturing electron correlation. One of the key advantages is that CC methods are size-consistent, meaning that the energy of a system consisting of two non-interacting subsystems is equal to the sum of the energies of the individual subsystems. Additionally, CC theory is size-extensive, ensuring that the total energy scales appropriately with the size of the system [59].

Moreover, CC methods can be systematically improved by including higher-order excitations (such as CCSDT for triple excitations or even higher-order terms). This allows for a precise treatment of electron correlation effects, especially for systems with significant electron-electron interactions [56].

Despite their high accuracy, CC methods have computational challenges, particularly for larger systems. The inclusion of higher-order excitations significantly increases the computational cost, making methods like CCSDT and CC3 computationally expensive and often impractical for very large systems. Furthermore, solving for the coupled-cluster amplitudes typically requires iterative methods and can be time-consuming for systems with many basis functions [60].

**3.3.2.2.2.4. Density functional theory (DFT)**

DFT descriptions have recently made significant progress in handling systems consisting of thousands of electrons, allowing for computational experiments on various materials. The primary variable in Kohn and Sham's method is the electron density, which has been tested and proven effective in solving the n-electron problem. DFT is based on the idea that the properties of a many-electron system can be described by the spatial distribution of electrons, represented by the electronic density, which is effective in solving the n-electron problem. DFT operates on the idea that the properties of a many-electron system can be described by the spatial distribution of



© University of Venda

electrons, represented by the electron density ρ(r) [61]. It can be expressed mathematically as a variational principle.

In the DFT approach, the ρ(r) represents the number of electrons at a specific position r per unit volume. It is related to the orbitals $\varphi_i$ in the Kohn-Sham system through the following equation:

$$\rho(r) = \sum_i |\varphi_i(r)|^2 = 2\sum_i^{occ} |\varphi_i(r)|^2. \quad (3.37)$$

Here, $\varphi_i(r)$ are the single-particle KS orbitals in the non-interacting reference system, and the factor of 2 accounts for spin degeneracy in doubly occupied orbitals. To determine the total number of electrons n, the electron density is integrated over all space:

$$\int \rho(r)dr = n. \quad (3.38)$$

Electron density determines the total number of electrons and serves as the basis for calculating potentials, energies, and other system properties [61].

The Hohenberg-Kohn theorem establishes that the external potential is a function of the electron density, meaning that the density ρ(r), an observable in three-dimensional space, can describe all properties of the system [62]. The total energy functional in Kohn-Sham DFT is expressed as:

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho]. \quad (3.39)$$

The kinetic energy of the non-interacting system in the KS approach is written as:

$$T_s[\rho] = \sum_i \left\langle \varphi_i \left| -\frac{1}{2}\nabla^2 \right| \varphi_i \right\rangle. \quad (3.40)$$

This term is derived from the KS orbitals $\varphi_i$ and approximates the kinetic energy of the interacting system.

The interaction between electrons and nuclei is represented by:



© University of Venda

$$V_{ne}[\rho] = \int \rho(r)v(r)\,dr. \tag{3.41}$$

where v® is the external potential at position r.

The electron-electron Coulomb repulsion is given by:

$$J[\rho] = \frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r-r'|}\,dr\,dr'. \tag{3.42}$$

The exchange-correlation energy approximates the difference between the true interacting system and the non-interacting reference system. It is expressed as:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]). \tag{3.43}$$

The exchange-correlation energy, $E_{xc}[\rho]$, is the most challenging term to calculate and is approximated using different exchange-correlation functionals. Various approximations, such as the local density approximation (LDA), generalized gradient approximation (GGA), and Hybrid functionals, are used to estimate it. The accuracy of DFT is highly dependent on the choice of the functional.

DFT is computationally efficient and capable of describing large systems. It is widely used in computational chemistry, materials science, and condensed matter physics to predict geometries, reaction mechanisms, and spectroscopic properties [61]. However, DFT calculations rely on approximate exchange-correlation functionals, which may introduce errors. The choice of functional significantly affects the results, and the predicted reaction mechanisms often require additional evidence for verification

**3.3.2.2.2.5. Quantum Monte Carlo (QMC)**

The fundamental principle behind the Monte Carlo (MC) method is the use of random numbers in problem-solving. MC is a practical approach for addressing a wide range of problems: physical process simulation, multidimensional integral evaluation, integral and linear operator equation solutions, and statistical mechanics applications. All of these uses of MC can be extended to address issues emerging from quantum mechanics, and the convergence of MC methods and



© University of Venda

quantum mechanics is commonly known as quantum Monte Carlo (QMC). Simulation methods utilizing QMC are highly effective for certain types of interacting many-particle models [63].

There are several established and emerging QMC computational methods. However, the two main techniques used in QMC are the variational quantum Monte Carlo (VMC) and the diffusion quantum Monte Carlo (DMC). VMC is generally considered more efficient and straightforward than DMC but has a lower degree of accuracy. VMC is centered on the Monte Carlo integration of the Rayleigh-Ritz quotient, while DMC originates from the basis that the Schrodinger equation can be expressed in imaginary time $\tau$ (measured in Hartree atomic units) as shown below on equation 3.44:

$$\frac{\delta\psi}{\delta\tau} = -H\psi. \tag{3.44}$$

VMC is a relatively straightforward extension of classical Monte Carlo, first developed by McMillan in 1965 for studying liquid helium [64]. The VMC approach calculates molecular properties using the Monte Carlo integration method using a trial wave function. A trial wave function in QMC approximates the true wavefunction that exactly solves the Schrödinger equation. By performing MC integration, one can calculate an estimated energy for a given trial wave function and then refine the wave function by varying its parameters to minimize the estimated energy. To compute the expectation value of energy, the Rayleigh-Ritz quotient is assessed using Monte Carlo integration with a trial function $\phi_T$, as shown in equation 3.45:

$$E_{VMC} = \frac{\int \phi_T^*(R) H_{\phi T}(R) dR}{\int \phi_T^*(R)_{\phi T}(R) dR}. \tag{3.45}$$

The Boltzmann distribution is used in classical MC to the coordinates R= ($r_1$, $r_2$, …, $r_N$), for an interaction V(R). The trial wavefunction's squared modulus $|\phi_T(R)|^2$ is used for sampling the electron coordinates, ensuring the wavefunction satisfies the antisymmetry requirement for fermions.

Since electrons are fermions, the trial wavefunction $\phi_T(R)$ must be antisymmetric under the exchange of two electrons. This antisymmetry is commonly implemented using a Slater



© University of Venda

determinant. However, a single Slater determinant neglects correlation effects, yielding results similar to Hartree-Fock theory [64]. In DMC, the fixed-node approximation is used to address the antisymmetry constraint. The nodes of the trial wavefunction $\phi_T(R)$—the points where $\phi_T(R)=0$ is fixed, ensuring the antisymmetry property. Errors introduced by incorrect nodal surfaces are known as fixed-node errors and are a critical limitation of DMC.

The expression of VMC energy can be represented by utilizing the local energy [65],

$$E_L(R) = \frac{H\phi_T(R)}{\phi_T}, \tag{3.46}$$

also known as $E_L$.

For an eigenfunction $\psi_i(R)$, the local energy $E_L(R)$ is constant and equals the corresponding eigenvalue $E_i$.

In DMC, the goal is to find the exact ground-state wavefunction and energy of the system by using a random walk algorithm to evolve the trial wavefunction over time. The random walk is performed in a high-dimensional configuration space, where each point in the space represents a possible configuration of the system. The evolution is governed by the imaginary-time Schrödinger equation, which eliminates the excited states from the wavefunction, leaving only the ground state. For time-independent Hamilton operators H, the lowest eigenfunction $\psi_0$ of H is found by solving the time-dependent Schrodinger equation in imaginary time eqn. (3.44). To achieve this, the solution $\psi(\tau)$ can be expressed in terms of the eigenfunctions $\psi_k$ of the Hamiltonian H as follows [66]:

$$H_{\psi_k} = E_k\psi_k. \tag{3.47}$$

As a result, the solution for $\psi_0$ can be determined. The exact ground state energy or DMC can be written as:

$$E_{DMC} = \frac{\int E_L(R)\phi_G(R)\psi_0(R)dR}{\int \phi_G(R)\psi_0(R)dR}. \tag{3.48}$$



© University of Venda

Here, $\phi_G(R)$ represents a guiding function to ensure efficient sampling. Alternatively, the energy can be expressed directly as,

$$E_{DMC} = \frac{\int \phi_G(R) \mathrm{H} \psi_0(R) dR}{\int \phi_G(R) \psi_0(R) dR} = \mathrm{E}_0, \quad (3.49)$$

where the trial function is projected out using a Hamiltonian function.

QMC has several advantages over other computational methods for quantum mechanics. For example, it is well-suited for studying systems with strong correlations and for solving problems that are difficult to treat using other methods, such as first-principles calculations of the properties of materials. Additionally, QMC is not limited by the size of the system, making it possible to study large systems with many degrees of freedom.

### 3.4. Linear combination of atomic orbitals (LCAO)

Theoretically, it is possible to determine the exact structure of molecular orbitals (MO) for a molecular system by solving specific differential equations [67]. The process of molecular orbital theory begins by assigning electrons in pairs to spatial functions labeled as $\psi_1, \psi_2, \psi_3, \ldots \psi_N$. These functions are then used to construct a multi-electron wavefunction represented as a single determinant (eqn. 3.19) [68]. The mathematical function that determines spin orbitals is called a Slater determinant, named after the scientist John Slater. This functional form is interesting because it results in indistinguishable electrons, which aligns with the findings of quantum mechanics.

A molecular orbital wave function, $\psi_i$, can be represented by a combination of atomic functions, $\phi_\mu$, where μ ranges from 1 to the total number of atomic orbitals in the molecule [68, 69]. Each atomic function $\phi_\mu$ corresponds to a distinct atom within the molecule.

This representation can be expressed as;

$$\psi_i = c_{1i}\phi_1 + c_{2i}\phi_2 + c_{3i}\phi_{3\ldots} = \sum_\mu c_{\mu i}\phi_\mu, \quad (3.50)$$

where, $\psi_i$ is the molecular orbital wave function



© University of Venda

$C_{\mu i}$ is the expansion coefficient of the ith atomic orbital

$\phi_{\mu}$ is the ith atomic orbital in the basis set.

The expansion in which the numerical coefficients $C_{\mu i}$, may be either positive or negative and are capable of being either real or complex numbers, referred to as a linear combination of atomic orbitals (LCAO). To find solutions to the wave equation for a molecule, one must specifically search for coefficients that give the charge distribution and energies of the MO's [70]. The method for finding these coefficients is based on the variation theorem, which suggests that the best coefficients for a particular MO are those that minimize the energy value [68, 71]. The coefficients for the expansion can be obtained by solving the Roothaan equations [68],

$$\sum_{v}( F_{\mu v} - \epsilon_i S_{\mu v}) c_{\mu i} = 0, \tag{3.51}$$

where, $S_{\mu v}$ is an overlap integral

$\epsilon_i$ is the one electron energy associated with $\psi_i$

$F_{\mu v}$ (or $\hat{f}$) is the component of the Fock matrix, representing the effective one-electron Hamiltonian in the Hartree-Fock method.

The Schrödinger equation's kinetic and potential terms can be represented in an unspecified manner using a one-electron Hamiltonian, $\hat{f}$. It is important to note that $\hat{f}$ is not the actual Hamiltonian operator for the molecule but rather an approximation to the Hartree-Fock Hamiltonian. The Hartree-Fock approach yields a differential equation to be solved, with the LCAO serving as a numerical method for this solution. However, the problem could also be addressed by other numerical methods, such as using a fourier transform technique with a plane-wave basis set or the finite element method over a grid in physical space (x, y, z).

The molecular orbital $\psi_i$ describes the probability density of finding an electron in the molecule, derived from a combination of the atomic orbitals $\phi_{\mu}$ from each atom. Compared to the valence bond approach, which focuses on localized bonding structures, the MO approach provides a more comprehensive view, especially for conjugated systems. In conjugated materials, electrons are



© University of Venda

delocalized across multiple atomic centers, which the MO approach naturally accommodates by allowing orbitals to spread over the entire molecule. This delocalization leads to a more accurate electronic structure, facilitating a clearer understanding of properties like conductivity and reactivity in conjugated materials [71].

### 3.5. Basis set

In modern methods of describing the electronic structure of both molecular and extended systems, the expansion of molecular orbitals in wave function or Kohn-Sham density functional methods is fundamental. A basis set is a set of mathematical functions used to describe a molecule's electronic structure or other system. These functions are typically chosen so that they can be combined in various ways to approximate the wavefunction of the system to a desired level of accuracy. They can be viewed as a set of basic functional building blocks that can be combined or "stacked" in different ways to approximate the electron density of the system, as shown [72],

$$\psi = a_1\phi_1 + a_2\phi_2 \cdots + a_k\phi_k, \tag{3.52}$$

where, k is the size of the basis set

$\phi_1, \phi_2, \ldots, \phi_k$ are the basis functions

$a_1, a_2, \ldots, a_k$ are the normalization constants.

The representation of an atomic orbital can be done through a singular function of either exponential or Gaussian type. Alternatively, it can be expressed by a linear combination of two or more functions. John C. Slater was the first to introduce the concept of calculating orbitals through the use of basis sets, which are made up of functions that are combined in a linear manner to represent atomic orbitals. These basis sets are made up of functions referred to as contractions or basis functions. Slater-type orbitals (STOs) were preferred as basis functions due to their similarity to the hydrogen atom's atomic orbitals. However, they tend to be less precise near the nucleus. The basis function is defined as,

$$BF = Ne^{(-zr)}, \tag{3.53}$$



© University of Venda

where, N is the normalization constant,

z is the orbital exponent,

r is the radius in Angstroms.

STOs are characterized by the function that is dependent on spherical coordinates, that is,

$$\phi_1(z,n,l,m;r,\theta,\phi) = Nr^{n-1}e^{(-zr)}Y_{l,m}(\theta,\phi), \quad (3.54)$$

where r, θ, and φ are the spherical coordinates

$Y_{l,m}$ are the angular momentum components (functions describing the "shape")

$n, l$ and $m$ are the quantum numbers: principal quantum number, angular momentum, and magnetic quantum number, respectively.

The STO equation simplifies the equations of hydrogen-like systems and has the form,

$$STO = \left[\frac{\alpha^3}{\pi}\right]^{0.5} e^{(-ar)}, \quad (3.55)$$

where α is the Slater orbital exponent.

Equation 3.52 represents the eigenvalue equation, and STOs are an estimation of the solutions to this equation.

During the 1950s, Frank Boys proposed a modification to the wavefunction by incorporating Gaussian type functions, which include an exponential factor of $e^{-a_i(r-R)^2}$ instead of the $e^{-\alpha r}$ present in the STOs [72]. These functions are much simpler to calculate. However, they do not represent the electron density of the real situation or the STOs. Nonetheless, this dilemma can be resolved to a great extent by using more Gaussian-type orbitals (GTOs). In such case, the MO ψ can be expanded into a set of known functions χ (basis set) [73]:

$$\psi = \sum_{i=1}^{M} c_i\chi_i. \quad (3.56)$$



© University of Venda

Nowadays, most ab initio molecular computations use GTO's as their basis functions [73]. The primary advantage of GTOs over STOs in electron-repulsion integrals is computational efficiency. GTOs are mathematically simpler to handle in quantum chemistry calculations because they have a closed-form expression for their integrals, especially when dealing with electron-electron repulsion integrals, which are computationally expensive. Mathematically, a basis set can be expressed as a linear combination of primitive Gaussian functions, which will be taken as cartesian Gaussian-type function (GTF):

$$\chi_i = N(x-X)^k(y-Y)^l(z-Z)^m e^{-a_i(r-R)^2}. \quad (3.57)$$

In a primitive function, the center is R (x,y,z), a nuclear position. The sum of k, l, and m determines the angular momentum, while the exponent $a_i$forms the radial extent of the function, N is the normalization constant [73]. The number of functions M and their angular momentum distribution, as well as the exponents $a_i$ values attributed to the accuracy of the expansion of the equation (3.56). To simplify the computational process, the primitive basis set is usually contracted, making K fixed linear combinations $k_j$ from M primitive functions $\chi_i$ (where K < M):

$$k_j = \sum_{i=1}^{M} d_{ji}\chi_i. \quad (3.58)$$

Due to the fact that the core orbitals are unaffected by the molecular environment, they can therefore be expressed by,

$$\psi = \sum_{j=1}^{K} c_j k_j, \quad (3.59)$$

a fixed linear combination without the loss of accuracy.

Contraction minimizes the number of coefficients $c_j$ from M to K for each molecular orbital, and thus simplifies the variational problem [73].

Two strategies can be used to increase the speed of integral calculations using Gaussian basis sets [74]. First, the one-electron basis functions are typically grouped in sets that share common properties of the Gaussian basis functions, allowing for more efficient calculations. This grouping



© University of Venda

reduces redundancy and makes the process faster. Second, more sophisticated techniques are used to handle the repetitive nature of calculations, specifically those that need to be performed K4 times. These might involve screening to discard unnecessary integrals that do not contribute significantly to the final result or applying algorithms such as density fitting or Cholesky decomposition, which optimize the evaluation of integrals by approximating complex electron repulsion integrals with a smaller, more manageable set of auxiliary functions. While these methods introduce additional computational steps, they significantly reduce the time complexity, making calculations much more efficient, particularly for larger molecular systems.

Different basis sets can represent different aspects of the molecular system, such as electron correlation, polarization, and diffuse electron density [75]. There are many different types of basis sets available, ranging from simple basis sets, such as the minimal basis set, to more complex basis sets, such as the augmented and double-zeta basis sets [73]. The choice of the basis set depends on the level of accuracy required for the calculation, the size of the molecule, and the computational resources available. Basis sets are important components of quantum chemical calculations, and they can be classified into various types based on their properties. Some commonly used basis sets include minimal, Pople, correlation, correlation consistent (cc) basis sets, and other split valence, and plane wave basis sets [72, 73].

### 3.5.1. Minimal or single zeta (SZ) basis set:

A minimal basis set is a basis set that uses only one basis function to describe each atomic orbital. This means that one basis function required to describe the free atom is considered for a given atom. The number of contracted functions in a minimum basis set is equal to the number of occupied orbitals for the isolated atom, with the fact that all components of a given function type are included [72]. This type of basis set is also known as a single-zeta (single z, SZ) basis set. For computational purposes, many minimal basis sets have been proposed, with the most common being the STO-nG basis sets developed by John Pople and his group. These basis sets involve a linear combination of "n" GTOs fitted to each STO. However, minimal basis sets provide poor results for molecules since they are inadequate at describing the deviation from the spherical symmetry of the atoms and the process of bond formation [73]. This is because these basis sets describe each atomic orbital using only one primitive orbital, which is unable to capture the complexity of molecular systems. For methods including electron correlation, where the virtual



© University of Venda

(unoccupied) orbitals are used for expanding the many-electron wave function, SZ basis sets are likewise unsuitable.

Minimal or SZ basis sets contain only one primitive Gaussian function per atom, which represents the atomic orbitals of each atom in the molecule. Polarized and non-polarized minimal basis sets are commonly used in molecular orbital calculations for small molecules. Examples of minimal or SZ basis sets include STO-3G, STO-4G, STO-6G, and STO-3G* [72]. STO-3G includes three basis functions - two p-type Gaussian and one s-type Gaussian function. STO-4G is a basis set consisting of four different functions: a s-type Gaussian function, two p-type Gaussian functions, and a d-type Gaussian function. STO-3G* enlarges the description of electron density around the nucleus using three polarisation functions enhanced by the three basis functions featured in STO-3G. Alternatively, STO-6G contains twice the number of basis functions integrated into STO-3G but no extra polarisation functions.

### 3.5.2. Double- and triple-zeta basis sets and the split-valence basis sets:

John Pople and his colleagues in the 1970s introduced the split-valence basis sets to improve the reliability and precision of basis sets for quantum chemical computations. The split-valence (SV) basis set involves the application of a single function for orbitals that do not belong to the valence shell and two functions for those that do. In contrast, the double zeta (z) (DZ) basis set uses two basis functions, in contrast to the one function in the minimal basis set [72]. The core orbitals' functions are not often replicated, resulting in DZ basis sets that are only double zeta for the valence orbitals. DZ basis sets are much better at describing changes due to molecular bond formation, as functions with different exponents can be used in different ratios to describe that bonding is different in different directions. However, a higher angular momentum function should also be added to describe the charge polarization for HF and DFT methods, and the electron-electron cusp at correlated levels. Such basis sets are double zeta (z) plus polarization (DZP). The structure of the basis set is described for the entire molecule, as opposed to specific atoms or orbitals, which differs from the approach Pople and colleagues took. Split-valence basis sets possess a specific configuration that enables quick and easy determination of their construction. The concept of augmenting the number of s- and p-functions while also incorporating higher



© University of Venda

angular momentum functions can be extended to produce triple zeta plus polarization (TZP), quadruple zeta plus polarization (QZP), and so on [73].

The 3-21G basis set is the smallest split-valence basis set that uses a three-primitive expansion for the 1s orbital and splits the valence orbitals into two basis functions. The first of these functions consists of a contraction of two Gaussian primitives, while the second is an uncontracted Gaussian primitive [72]. For instance, in the case of the hydrogen atom, the lone valence orbital is split up into two Gaussian groups, one carrying two primitives and the other carrying one primitive.

The 4-31G and 6-31G basis sets are variations of the 3-21G basis set that are slightly larger [72]. These basis sets and other split-valence basis sets, such as SV- or DZ-type basis sets, use inner and outer basis functions for valence orbitals. This allows for these orbitals to expand or contract in molecular calculations, thus adjusting flexibly to the bonding requirements of the molecule. The 6-311G basis set is an example of a triply split-valence basis set that is of TZ (triple zeta (z)) quality in the valence part but only minimal in the core [72]. Split-valence basis sets such as the 3-21G basis set allow orbitals to change size, which is the first step in providing a molecular orbital environment. Even the atomic orbital shape can be changed in molecular orbital formulation.

The use of DZ and TZ basis sets accounts for more refined descriptions of electron correlation, allowing calculations to better model phenomena like bond formation, electron density redistribution, and molecular reactivity. The inclusion of these functions enables computational models to more accurately represent how atoms and electrons behave in a molecular environment, thus capturing the physical effects of bonding, orbital overlap, and interaction with the surrounding medium.

### 3.5.3. Polarization basis set:

The introduction of polarization functions to basis sets is required in order to describe the distortion of s or p orbitals during the formation of chemical bonds in molecules. The polarization functions involve the addition of p- or d-type basis functions. These functions allow for charge polarization to occur away from the atomic distribution and enable the molecular wave function to become more flexible in its ability to deviate from spherical symmetry within the vicinity of each atom.



© University of Venda

The decision to include polarization and diffuse functions depends on the molecular properties. Furthermore, incorporating higher basis sets, such as 6-31G, can increase the basis set's flexibility in the molecule's valence region [72]. The introduction of p, d, and f functions in the basis set can replicate the distortion of s, p, and d orbitals. This leads to double-zeta plus polarization (DZP) or split-valence plus polarization (SVP) basis sets. Incorporating polarization functions (such as d or f orbitals) into these basis sets (e.g., DZP or TZP) helps to account for the distortion of electron clouds in response to chemical bonding or external fields, which can affect molecular geometry, electronic distribution, and interaction with other molecules. Thus, the inclusion of polarization functions noticeably enhances the description of molecular geometries, such as bond lengths and angles, as well as relative molecular energies. As more functions are added to the basis set, the computations involve an exponentially larger number of basis functions, integrals, and matrix dimensions. To illustrate, we consider methane, which requires 17 basis functions and 38 primitive Gaussians with a 6-31G basis set. However, a 6-31G(d) basis set requires 27 basis functions and 48 primitive Gaussians. Furthermore, with 6-31G(d, f), 39 basis functions and 60 primitive Gaussians are required.

The choice of basis sets and polarization functions in computational chemistry is important for obtaining accurate results. The use of polarization functions with SZ basis sets is not recommended, as the error in the sp-part of the basis is much larger than the error from the lack of polarization functions [73]. On the other hand, a DZP basis set is defined by adding a single d-function to a DZ-type basis set, and early applications of TZP-type basis sets often added two or three d-functions and one f-function based on performance criteria for a selection of properties and systems. The addition of d- and f-functions to TZ, QZ (quadruple zeta (z)), and 5Z (quintuple zeta (z)) basis sets is necessary for consistent polarization [76]. The number of d-functions required for consistent polarization depends on the level of theory used, and the errors in small basis sets are systematic due to the lack of deep enough molecular energy. Basis sets of polarized triple-zeta quality are the smallest to be recommended, and convergence towards the basis set limit is faster at the polarized triple-zeta level.



© University of Venda

### 3.5.4. Diffuse basis sets

When calculating anions, it is important to enhance the wave function in areas far from the nucleus as the extra electron is typically weakly bound [73]. To do this, diffuse functions with small exponents are included in the basis set, extending far from the nucleus [73]. These functions are well-known for improving the depiction of the extra electron in anions and have been added to various basis sets, resulting in enhanced accuracy in calculating proton affinities and the electronic excited states of molecules. Dunning and his colleagues [73, 74] proposed the term "aug-" ("augmented") to describe the addition of one diffuse basis function to every atom for each angular momentum of basis functions already present on that atom, resulting in fully augmented basis sets. In contrast, the earlier "plus" basis sets originally introduced by Pople and his colleagues [77] added solely diffuse s and p subshells to non-hydrogenic functions, with no diffuse functions added to hydrogen atoms, and are referred to as minimal augmentation. The definition of minimal augmentation is categorized into three types, represented by the symbols “+”, “maug-”, and “ma-” [77]. The "+", "maug-", and "ma-" types of minimal augmentations are used to add diffuse functions to basis sets for various atoms. The "+ " type can be applied to any basis set where exponential parameters have been defined for adding diffuse functions to the 6-31G or 6-311G basis sets. The "maug-" type can be used with any aug-type correlation consistent basis set, while the "ma-" type can be used with any nondiffuse basis set. To create diffuse functions for atoms heavier than helium using the "ma-" type of minimal augmentation, one simply divides the smallest s and p exponential parameters already present by a factor of 3 [77].

Schleyer and colleagues incorporated diffuse s- and p-type basis functions with small exponents into the Pople-style k-*lmn*G basis sets, and the exponents for the diffuse functions were determined through HF optimization on the atomic anions [73]. They discovered that adding diffuse functions to 4-31G, denoted as 4-31 + G, had a tremendous impact on calculated proton affinities, and the results were very similar to those of 6-31 + G*, which is 6-31G* augmented with diffuse functions [78].

### 3.5.5. Correlation consistent basis sets

Correlation-consistent (cc) basis sets, introduced by Dunning and colleagues, are fundamental in post-Hartree-Fock (HF) calculations for effectively converging electronic energies toward the



© University of Venda

complete basis set (CBS) limit [79]. The CBS limit represents an idealized target where the basis set approaches an 'infinite' size, theoretically capturing all electron correlation with total accuracy. While achieving an actual infinite basis is unfeasible, extrapolation techniques allow researchers to approximate the CBS energy. For instance, calculations with progressively larger basis sets, such as cc-pVDZ, cc-pVTZ, and cc-pVQZ, yield a converging trend that can be extrapolated to estimate the CBS limit [80].

Each cc basis set includes polarization functions (e.g., d, f, g), which systematically increase the level of electron correlation captured, with notation such as ‘cc-p’ indicating ‘correlation-consistent polarized’ and ‘V’ signifying valence-only orbitals. This basis set family includes options like cc-pVDZ (double-zeta), cc-pVTZ (triple-zeta), and higher zeta levels to yield smoother convergence toward the CBS limit. Augmented versions, such as aug-cc-pVDZ, incorporate additional diffuse functions to enhance calculations for systems with extended electron distributions [81, 82].

For greater efficiency in achieving near-CBS accuracy, explicitly correlated methods such as CCSD(T)-F12 can closely approximate CBS-level results using moderate-sized basis sets, reducing computational costs significantly [83]. These techniques, alongside traditional extrapolation, provide a practical approach to achieving results that approximate the CBS limit in computationally intensive post-HF applications.

### 3.5.6. Basis set error

One of the significant sources of error in the solution of the electronic Schrödinger equation is the basis set truncation error [84]. This error is the difference between the true solution of the electronic Schrödinger equation and the experimental value corrected for non-Born-Oppenheimer (non-BO) effects and possibly relativistic corrections [72]. It occurs when we use a limited basis set size for the analytical expansion method applied to atoms and molecules, leading to a computational result that depends on the size of the basis set used. As we increase the size of the basis set, the result approaches a certain limiting value, although the approach may or may not be monotonic. The difference between the limiting value and the value obtained by a truncated basis set defines the basis set truncation error [78].



© University of Venda

When calculating the interaction energy of two atoms or molecules, A and B, it is common practice to determine the difference in energy levels between the product complex AB and its constituents, A and B as;

$$\Delta E(R) = E_{AB}(R) - (E_A + E_B) \tag{3.60}$$

where R is the distance between Atom (or molecule) A and B

$E_{AB}(R)$ is the is the total energy of the interacting system

$E_A$ and $E_B$, are the energies of the two atoms in isolation.

If a large saturated basis set is used, the interaction energy will be expressed as,

$$\Delta E_S(R) = E_{AB,S}(R) - \left(E_{A,S} + E_{B,S}\right), \tag{3.61}$$

and if an unsaturated basis set is used, the interaction energy can be expressed as:

$$\Delta E_U(R) = E_{AB,U}(R) - \left(E_{A,U} + E_{B,U}\right). \tag{3.62}$$

Thus, the basis set truncation error can be written as:

$$E(R) = \Delta E_S(R) - \Delta E_U(R). \tag{3.63}$$

This error can be further divided into the basis set error associated with the limited size of the 1-electron particle basis and the n-electron error associated with the incompleteness of the n-electron basis. A complete basis set eliminates the basis set error, leaving only the intrinsic error of the corresponding model [72]. One-electron basis set truncation errors can be significant, especially if inadequate basis sets are chosen for the given problem or the desired accuracy. However, theoretical estimates are expected to be good under the assumption that non-dynamical electron correlation plays no role [72]. When the atoms of interacting molecules or different parts of the same molecule approach each other, their basis functions overlap, leading to an effective increase in the basis set. This improves the calculation of derived properties such as energy. If the total



© University of Venda

energy is minimized as a function of the system geometry, the short-range energies from the mixed basis sets must be compared with the long-range energies from the unmixed sets to avoid a mismatch known as basis set superposition error (BSSE) [72].

Obtaining accurate results greatly depends on the atomic orbital approximation basis set used. A basis set can be classified as either saturated or unsaturated. In the case of a saturated atomic basis set, the atomic basis functions placed on atom B do not improve the 'atomic' energy of atom A because the basis set on A is already saturated [78]. However, if the atomic basis set on A is not saturated, the basis functions placed on B will improve the atomic basis set on A towards saturation. Similarly, the unsaturated basis set on B will also be improved by the basis functions on A. To shape the molecular orbitals, polarization functions can be added to A and B, altering the atomic basis functions. In this scenario, the polarization functions placed on B will partially contribute to enhancing the energy of A and vice versa. Therefore, the interaction energy of the A-B system, $E_{AB}$ (R), contains these 'atomic' energy gains, and $\Delta E_U(R)$ also contains these energy gains in addition to the difference properly accounted for as the interaction energy between A and B [78].

The Basis Set Superposition Error (BSSE) counterpoise correction can be mathematically expressed as:

$$\Delta E'_U(R) = E_{AB,U}(R) - E_{A\,(B),U}(R) - E_{B(A),U}(R). \tag{3.64}$$

This equation incorporates the atomic energy of A, which is calculated using both the basis sets on A and B. It should be noted that the atomic energy is determined without the presence of the atomic nucleus B. The basis functions on B, without the actual atomic nucleus B, are commonly referred to as the "ghost" basis functions on B [78].

There are several computational methods used to minimize the basis set superposition error (BSSE). The counterpoise method, which was proposed by Boys and Bernardi, is one of the methods used to address BSSE in the electronic structure calculations. The method involves calculating the difference between $\Delta E'_U(R)$ and $\Delta E_U(R)$,



© University of Venda

$$E_{CP}(R) = \Delta E'_U(R) - \Delta E_U(R), \quad (3.65)$$

and this difference can be used as an indicator of the completeness or degree of saturation of the basis set used. If the difference is positive, it suggests that the basis set used is incomplete or fully saturated. The counterpoise method is currently the most practical approach for correcting BSSE [78].

Another method is the chemical Hamiltonian approach (CHA), which prevents basis set mixing by replacing the conventional Hamiltonian with a new one. The CHA removes the terms of the Hamiltonian responsible for BSSE by removing all projector-containing terms, which would allow basis set extension [72].

### 3.6. Geometry optimization

Geometry optimization, also known as energy minimization, is an iterative process that starts with an initial guess of the molecular or material structure and adjusts the positions of the atoms to find a more stable configuration. This process uses mathematical optimization algorithms to alter the positions of atoms in a molecule or solid to reduce the net forces acting on them until they become negligible [85]. The stable states of molecular systems correspond to global and local minima on their PES.

In recent years, several optimization algorithms based on natural, redundant, or delocalized internal coordinates have been developed. Optimizations using these coordinates are typically more efficient than cartesian coordinates, requiring fewer optimization cycles. The use of an appropriate coordinate system is also one of the best ways to improve the quality of geometry optimization [86], ensuring speed and reliability in both optimization and transition state searches [87, 88].

Several different algorithms can be used for geometry optimization, including steepest descent, conjugate gradient, and Newton-Raphson methods. Here we discuss a few of them.



© University of Venda

### 3.6.1. Newton and Quasi-Newton methods

Quasi-Newton and Newton methods have become the popular means for optimizing equilibrium geometries. These techniques are also good when searching for TSs [89]. The Newton-Raphson method involves the Taylor series expansion of the PES at the current geometry and is known to be the most computationally expensive per step among all the methods used for this purpose [85]. The method updates the geometry iteratively using an equation that takes into account the gradient and Hessian of the PES,

$$x^{k+1} = x^k - V_K^{-1}F(x^k), \quad V_k \epsilon \delta_b F(x^k), \quad k = 0, 1, \ldots\ldots, \tag{3.66}$$

for solving a nonlinear algebraic equation f(x) = 0. Despite its high computational cost, the Newton-Raphson method remains a valuable tool for accurately determining the equilibrium configuration of molecules and solids. And is one of the most important advantages of this method is its faster convergence rate compared to many other iterative methods for solving nonlinear equations [90]. However, it should be noted that this method requires the derivative of the function f(x) to be available, which can be a disadvantage in certain cases where the derivative is not easily accessible.

Quasi-Newton methods are an effective class of optimization algorithms that accelerate the steepest-descent technique for function minimization by using computational history to generate a sequence of approximations to the inverse of the Hessian matrix. These methods converge faster than the steepest-descent method and are commonly used in large-scale optimization problems. Quasi-Newton methods use,

$$x^{k+1} = x^k - V_K^{-1}F(x^k), \quad V_k \epsilon R^{n\times n}), \quad k = 0, 1, \ldots\ldots, \tag{3.67}$$

for solving nonlinear equations.

The most commonly used Quasi-Newton methods are the Broyden-Fletcher-Goldfarb-Shanno (BFGS) and Davidon-Fletcher-Powell (DFP) algorithms [91]. Newton-Raphson methods offer fast convergence and high accuracy, making them ideal for refining the equilibrium geometry of molecules and solids. They are widely applicable across various molecular systems, including



© University of Venda

effective transition state searches. However, they are computationally expensive due to the need for calculating both the gradient and Hessian matrix, which can be costly for large systems.

### 3.6.2. Steepest descent method

The steepest descent method is well-known for minimizing multivariable unconstrained optimization problems [92]. It is the simplest gradient descent method [93, 94]. Unlike other methods that require calculating multiple-second derivatives, this method uses an approximation. In this method, it assumes that the second derivative remains constant. The formula for this iterative method is given by:

$$x^{k+1} = x^k - a_K^{-1}\nabla f(x^k). \tag{3.68}$$

The Steepest Descent method, unlike Newton-Raphson, prioritizes rapid progress by calculating gradients at each point without the need for second derivative calculations. However, since it is only an approximation, it typically requires more iteration to reach the minimum [85]. This approach is named after its strategy of minimizing geometry in the opposite direction of the maximum (i.e., steepest) gradient initially [85]. After reaching the minimum in the first direction, the next steepest direction is followed for further minimization. This process continues until a tolerance level is met and a minimum is achieved in all directions [85].

The steepest descent method is known for being simple and computationally inexpensive, as it only requires the first-order gradient and does not involve second derivatives, unlike methods such as Newton-Raphson. This makes it an attractive option for systems that are far from equilibrium, as it can quickly guide the system towards a local minimum. However, its main drawback is slower convergence due to the reliance on first-order derivatives, often requiring more iterations to reach the optimal configuration compared to second-order methods. Additionally, the method's efficiency is highly dependent on the initial geometry; poor starting points can lead to slow or failed convergence. Another challenge is the risk of stagnation in shallow local minima if the optimization path is not properly adjusted.



© University of Venda

### 3.6.3. Conjugate gradient method

The conjugate gradient method is a conjugate direction method in which the selected continuous direction vector is treated as a conjugate version of the continuous gradient obtained during the method [95]. It has an intermediary role that lies between the steepest descent method and Newton's method. The conjugate gradient method is often implemented as an iterative algorithm. It can be considered as being between Newton's method, a second-order optimization algorithm, and the steepest descent method, a first-order optimization algorithm [95]. In the Conjugate Gradient method, the initial phase of the search occurs in the opposite direction of the largest gradient, similar to the Steepest Descent method. These methods offer a practical solution for dealing with extensive systems, thanks to their ability to minimize storage requirements in comparison to limited memory quasi-Newton methods. The underlying concept of these methods involves selecting a fresh search direction that effectively reduces energy while remaining in close proximity to the previous search direction's minimum. In cases where there is interdependence among the coordinates in the Hessian, the optimal search directions are not orthogonal but rather conjugate. Two widely recognized conjugate gradient methods are Fletcher-Reeves90 and Polak-Ribiere91, respectively expressed as,

$$s_i = -\mathbb{g}_i + \frac{\mathbb{g}_i^T \mathbb{g}_i}{\mathbb{g}_{i-1}^T \mathbb{g}_{i-1}} s_{i-1}, \tag{3.69}$$

and

$$s_i = -\mathbb{g}_i + \frac{(\mathbb{g}_i - \mathbb{g}_{i-1})^T \mathbb{g}_i}{\mathbb{g}_{i-1}^T \mathbb{g}_{i-1}} s_{i-1}. \tag{3.70}$$

To achieve convergence, line searches are needed at each step in conjugate gradient methods, unlike quasi-Newton methods. These methods only require the storage of three vectors: the current and previous search directions ($s_i$ and $s_{i-1}$) and the current and prior gradients ($g_i$ and $g_{i-1}$)

The conjugate gradient method is a memory-efficient optimization technique making it advantageous for large-scale systems where memory usage is critical. It offers a good balance between the steepest descent method's simplicity and the advanced optimization path handling of Newton-Raphson methods, resulting in better performance than the steepest descent in many cases. This method is versatile and well-suited for large systems, as it reduces storage requirements



© University of Venda

compared to more computationally expensive methods like Quasi-Newton. However, one of its main drawbacks is that it does not guarantee convergence to a global minimum, especially in complex systems with many degrees of freedom. Additionally, the requirement for line searches at each iteration can slow down convergence compared to methods that do not need this step.

The efficiency of optimization can be measured by the number of energy and gradient evaluations needed to achieve convergence. Three choices influence the efficiency of geometry optimization: (1) the coordinates used to describe the system, (2) the mathematical technique used for optimization, and (3) the quality of the quadratic approximation (the force constants or Hessian) [96].

Optimization methods can be categorized into three groups based on the information used. The first group uses only the objective function, while the second group incorporates both the function and its gradient. The third group goes a step further by considering the function, gradient, and Hessian [97].

The choice of optimization algorithm will depend on the specific problem and the goals of the optimization. For example, steepest descent algorithms move the molecular geometry in the direction of steepest decrease in energy, while conjugate gradient algorithms use a combination of the gradient and previous optimization steps to update the geometry. Newton-Raphson methods use second-derivative information to optimize the energy and are generally faster than other optimization algorithms. Once the optimization process has converged, the final molecular or material structure represents the most stable configuration with the lowest energy. This structure can then be used to compute different molecular properties, such as molecular geometries, bond lengths, and reaction energetics. Therefore, choosing the right optimization algorithm is important for obtaining accurate results and saving computational time.

### 3.7. Calculation of molecular properties

The calculation of molecular properties is an essential step in molecular modeling and computational chemistry. It relies on computational techniques methods to obtain important information about molecular structure, stability, and reactivity, which is significant for the study



© University of Venda

and development of molecular systems and materials. Molecular Properties can be categorized based on their calculation methods. The following is an outline of the different calculation methods used to categorize properties [98];

- From the potential hyper surface (PHS), several properties can be derived that take into account the electronic energy with respect to nuclear coordinates. These properties include equilibrium geometries, the geometries and energies of potential isomers or transition structures, harmonic force constants (which determine harmonic vibration frequencies and, subsequently, IR and Raman spectra), barriers for rotational, inversion, and isomerization processes, as well as bond dissociation and atomization energies. Such properties are important in understanding the behavior and reactivity of molecules in different chemical environments.
- First-order properties, such as dipole moment and higher multipole moments, density and spin density at the nucleus, and the field gradient at the nucleus, can be derived from the wave function or reduced density matrices.
- After performing a conventional ab-initio calculation on the state in question, some properties require the application of perturbation theory. These properties are known as the second-order properties and third-order property. Second-order properties include electric polarizabilities, magnetic susceptibilities, NMR chemical shifts, and indirect spin-spin coupling. An example of a third-order property is hyperpolarizability.
- Properties involving multiple electronic states consist of optical transition energies and intensities (observed in ultraviolet (UV) spectra), ionization potentials (evident in photoelectron (PE) spectra), properties associated with circular dichroism, and others.
- Non-adiabatic effects give rise to properties that fall outside the scope of the Born-Oppenheimer (BO) approximation. These include non-adiabatic transition probabilities and Jahn-Teller distortions.
- Unbound states possess distinct properties, including rates of autoionization, which can be classified as a separate category.

### 3.8. Hydrogen bonding

#### 3.8.1. Hydrogen bond and its significance

Hydrogen bond (HB) is a concept (i.e., the theoretical framework that explains the nature of hydrogen bonding and its role in various scientific contexts) of utmost significance, as it holds a



© University of Venda

fundamental role in several scientific disciplines [99]. The reason for such interest in HB arises from its important role in enzymatic catalysis [100], proton transfer reactions [101], molecular arrangement in crystals [102, 103], and its vital involvement in various biological processes. [104, 105]. Mineralogy, material science, inorganic and organic chemistry, supramolecular chemistry, biochemistry, molecular medicine, and pharmacy can be influenced by this fundamental area of study. The HB is the key to understanding molecular conformation, molecular aggregation, and the function of countless chemical systems, from inorganic to biological. Due to its site-specificity, hydrogen bonding is central to understanding microscopic structures and functions in numerous molecular and supramolecular systems, such as hydrogen-bonded water or alcohol networks, organic compounds in solution, hydrogen-bond crystal engineering, polymers, self-assembled supramolecular architectures, proteins, and DNA [106].

There are several definitions of the HB [107, 108]. One of the most recommended definitions of the HB is the one that was proposed by the IUPAC, which states that ‘’ the HB is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation’’ [109]. The definition of HB can be illustrated as X–H···Y–Z, with X-H representing the donor and Y-Z representing the acceptor. The acceptor can take the form of either an atom or an anion Y or a fragment or molecule Y–Z, where Y is bonded to Z. In certain instances, X and Y may be identical, resulting in what is known as symmetric hydrogen bonds. The acceptor is a region abundant in electrons, such as a lone pair of Y or a π-bonded pair of Y–Z [110]. N, O, and F are typically the atoms represented by X and Y [111, 112]. HBs that form between a hydrogen atom that is covalently bonded to a more electronegative atom (O, N, F) are known as conventional HBs, and the ones that involve hydrogen atoms covalently bonded to less electronegative atoms (such as C) or hydrogen atoms that are not covalently bonded to an electronegative atom are known as unconventional HB s.

HBs can occur within a molecule or between neighboring molecules, as shown in Figure 3.4. The one occurring within a molecule is called intramolecular HB, while the one occurring between.is known as the intermolecular HB.



© University of Venda

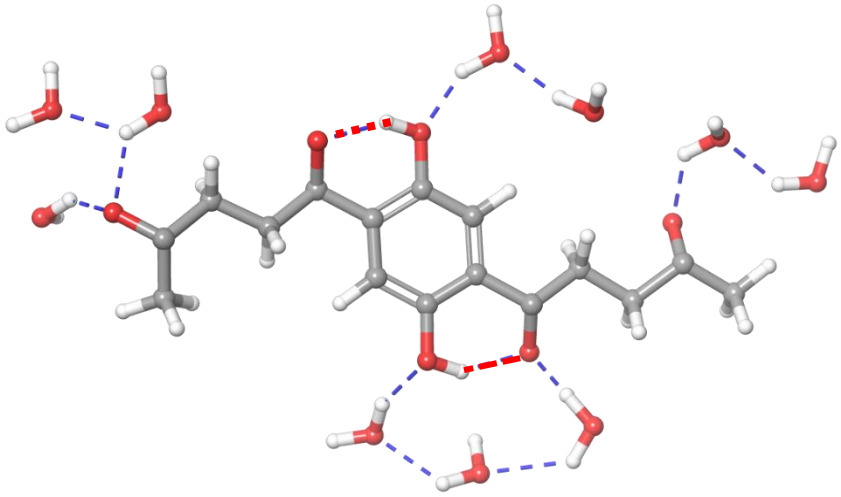


**Figure 3.4: Illustration of intramolecular and intermolecular HBs.** Red-segment lines show intramolecular lines and blue-segment lines show intermolecular lines.

### 3.8.2. Classification of hydrogen bonds

The classification of HBs can be approached in different ways, and there is no universally accepted standard method for categorizing them. Among the numerous parameters available, such as interaction energy, HB distance, HB radii, atoms in molecules (AIM) topological parameters (r(bond critical points, (BCP)) and r2 r(BCP)), orbital energies (from NBO approach), and more, one must select one or multiple criteria to classify the HB [113]. This work categorizes HBs as



© University of Venda

strong, moderate, or weak based on their strength and energy. The strength of an HB is determined by factors such as bond length (the distance between the H atom and the acceptor), bond angle, and bond energy.

**Bond angle (θ):** The formation of a bond angle (θ) between X–H···Y atoms reflects the directional nature of HBs. This directionality, characterized by linearity, helps differentiate HBs from VDW interactions [114]. The linearity of HBs is directly related to the distance between the atoms; as the angle θ approaches 180°, the distance decreases. The extent of directionality is influenced by the polarity of the donor atom, with greater polarity resulting in a more linear interaction. Additionally, HBs exhibit directionality on the acceptor side as well.

**Bond energy**: The interaction energy of stable aggregates can be determined through calorimetry or mass-spectrometry experiments [108]. However, these techniques are complex, and experimental values are limited to a small number of hydrogen-bonded dimers. Another approach is to measure the sublimation heat [115] of crystals, which reflects the overall interaction energy of all intermolecular interactions within the crystal [116].

It is also important to note that the direct determination of the energy of HBs in the solid state is not possible through experimentation alone. However, with the help of a sound structural model obtained from experiments, computational techniques can be used to calculate the energy of the HBs within the structure.

Computational methods have become indispensable for detailed and accurate assessments of HB interactions. Advanced quantum chemical calculations, paired with robust basis sets, enable the precise decomposition of bond energy. This computational approach provides a quantitative measure and reveals the distinct energetic contributions of HBs, which are primarily calculated within isolated systems, often in the gas phase [117, 118].

The theoretical breakdown of hydrogen bond energy generally follows an equation where bond energy (BE) is given by:



© University of Venda

$$\text{bond energy (BE)} = E_{elst} + E_{exch} + E_{ind} + E_{dis}. \quad (3.71)$$

This equation allows us to determine the characteristics of a specific HB interaction. The term $E_{elst}$ primarily represents the classical electrostatic interaction between the individual monomers. On the other hand, the term $E_{exch}$ represents the exchange-repulsion contribution resulting from the overlap of wave functions of each unit, along with the need for antisymmetry due to fermionic behavior exhibited by electrons in the dimer. The $E_{ind}$ term depicts the induction contribution, encompassing the polarization that arises from the response of each monomer to the electric field exerted by the other monomer, as well as the charge transfer between them. The $E_{dis}$ term represents the dispersion contribution arising from the Coulomb correlation among electrons in one monomer and those in another.

The decomposition of HB energy into distinct contributions such as electrostatics, exchange-repulsion, induction, and dispersion is a widely used approach in theoretical chemistry, although it has nuanced interpretations. Research using advanced computational methods like symmetry-adapted perturbation theory (SAPT) and local energy decomposition (LED) has shown that while electrostatic interactions are often the dominant force in HB formation, the exact partitioning of HB energy can vary depending on the model and method applied. SAPT, for instance, effectively divides interaction energies into electrostatic, exchange-repulsion, induction, and dispersion terms, with electrostatics typically contributing the largest stabilizing effect in HB systems, especially in gas-phase dimers. This method is known for its consistency across various hydrogen-bonded systems, even though dispersion and induction contributions may differ in solid-state environments [118].

The assertion that HB energy could be derived from electrostatics alone (including higher-order multipole moments) aligns with earlier theories suggesting electrostatics as the sole necessary component for HB formation in simple systems. However, modern studies clarify that a purely electrostatic view does not capture the entirety of HB interactions, especially for complex or biological systems where dispersion and induction become more pronounced, especially under non-idealized conditions like those in the solid state. Using SAPT or LED [119, 120], interaction energy components for accurate descriptions of HBs can be calculated across different states of



© University of Venda

matter, making these methods valuable for validating HB models and capturing the nuanced nature of HB energy.

### 3.8.2.1. Strong hydrogen bonds

The HBs that exhibit strong strength are characterized by their short length, symmetry, proton-centered nature, and linear X–H···Y structure bonds that can be understood as a result of three-centre–four-electron interactions with a covalent character [121]. Their distances typically fall within the range of 1.2–1.5 ˚ A [112], while the bond angle is approximately between 170 and 180 degrees.

Isabel Rozas's group considers strong HB to be those with BE greater than -41.84 kJ $mol^{-1}$ [122]. However, Hibbert and Emsley propose that the bond energy falls within the range of -50.21 to -100.42 kJ $mol^{-1}$, considering them moderately strong HBs [100]. Desiraju, on the other hand, defines strong HBs as those with interaction energies ranging from -16.74 to -62.76 kJ $mol^{-1}$, while very strong HBs have energies greater than -62.76 kJ $mol^{-1}$ [123].

The formation of a strong HB requires the close proximity of donor and acceptor heteroatoms and similar or identical proton affinities in the surrounding environment. It is important to note that there are various types of strong HBs, including low-barrier HBs (LBHBs), charge-assisted HBs (CAHBs), and resonance-assisted HBs (RAHBs).

**The low-barrier hydrogen bonds (LBHBs):** Cleland and Kreevoy initially described them as "short and strong" HBs [124]. These particular HBs have been observed in complexes formed by ylides that contain nitrogen, oxygen, and carbon. The distances between the atoms involved in these HBs typically fall within the 1.39–1.46 angstroms range. Moreover, their interaction energies span from -88.70 to -102.93 kJ $mol^{-1}$ [125].

**Charge-assisted hydrogen bonds (CAHB):** These HBs are formed between charged species and are influenced by electrostatic forces. These HBs, as described by Gilli, exhibit a significant covalent character and are considered very strong interactions [126]. Meot-Ner further categorizes



© University of Venda

CAHBs as a distinctive class of HBs between ions and molecules, displaying strengths ranging from -20.92 to -146.44 kJ $mol^{-1}$ [127].

**Resonance-assisted hydrogen bonds (RAHBs):** They were first introduced by Gilli et al. in 1989 [128]. These bonds are based on the concept that the interaction between an HB and hetero-conjugated systems can greatly enhance the strength of the bond itself [128]. A cooperative effect has been observed in molecules where hydrogen bonding occurs along with conjugated systems (multiple π-bonds). This effect was qualitatively recognized by Huggins in 1936 [129] and later quantitatively studied by Coulson [129], who understood that a purely electrostatic interpretation of hydrogen bonding was insufficient. Contemporary authors describe RAHB as a positive synergism between the strengthening of hydrogen bonding and the π-delocalization of the interleaving resonant fragment [130].

### 3.8.2.2. Moderate hydrogen bonds

Moderate HBs, in their nature, can be recognized by their moderate bond energy and the bond length and angle that lie between strong and weak HBs. The length of these bonds usually falls within the range of 2.2-2.3 Å, while the angle they form is typically found to be around 130-170°. Alkorta and co-workers [122] considered moderate HB as those with BE between -20.92 and -41.84 kJ $mol^{-1}$.

### 3.8.2.3. Weak hydrogen bonds

Weak HBs primarily exhibit electrostatic properties. They are characterized by their elongated and asymmetrical structure, with the proton positioned slightly off-center and often bent X–H···Y bonds [131]. These HBs exhibit a bond length ranging from 2.0 to 3.0, while the bond angle falls within the range of 90-130° [113].

The classification given by Desiraju identifies these interactions as having BE below -16.74 kJ $mol^{-1}$ [123]. Hibbert and Emsley, on the other hand, suggest that their bond energies vary between



© University of Venda

-10.04 and -50.21 kJ $mol^{-1}$ [132]. Isabel Rozas's group states weak HBs have BE of up to -20.92 kJ $mol^{-1}$ [122].

### 3.8.3. Hydrogen bond pattern

Two molecular systems have the ability to engage in multiple simultaneous interactions. These interactions can manifest in various patterns, either occurring in a parallel or bifurcated manner. To illustrate, parallel hydrogen bonding (HB) complexes can arise between fused rings having three HB acceptors and fused-ring systems with three HB donors (referred to as dddaaa systems), or between fused rings having two HB acceptors and fused-ring systems with two HB donors (known as a dd-aa system). Additionally, parallel HB complexes can form between a system with one donor and two acceptors (known as a d-aa system), or between a system with two donors and one acceptor (known as a dd-a system). On the other hand, bifurcated HB systems can occur in a d-aa system and a dd-system.

The strength of the interaction between systems is determined by the amount of electron density present, which differs depending on the number of centers involved and the sides of the rings formed. Experimental studies have shown that "parallel" HB interactions are more stable and preferred compared to "bifurcated" interactions [112]. Furthermore, "bifurcated" interactions with two donors are more stable than those with only one donor. For example, the ddaa complex, which has a "parallel" HB system, exhibits a higher interaction energy compared to the corresponding "bifurcated" HB systems (dda and aad) and other "parallel" HB complexes. When considering only the "bifurcated" complexes, the one with two HB donors (dda) and a six-sided "ring" is more stable than the one with only one donor (aad) and a four-sided "ring" [112].

The electron density exhibited a high degree of hydrogen bonding for ddaa, a moderate to strong HB for dda, and a moderate to weak HB for aad. The complexes of ddaa, dda, and aad displayed symmetric BCPs (bond critical points) and a single RCP (ring critical point), indicating the distance and properties of the HB.



© University of Venda

### 3.9. The computational study of molecules in solution

#### 3.9.1. The study of solvent effects

Many chemical reactions take place within solutions, particularly within living organisms. It is essential to employ effective computational methods to depict the properties of molecules within a solution accurately and the barriers and energies involved in chemical reactions [133, 134]. There are two primary approaches to account for solvent effects: implicit and explicit [135, 136]. Implicit solvation models view the solvent as a continuous medium and analyze how a solute molecule interacts within this medium [137]. On the other hand, explicit solvation models examine the interactions between a solute molecule and individual solvent molecules [136]. While an overview of studying molecules and processes in solutions has been previously published, this paper will delve into two specific methods.

When a molecule transitions from the gas phase to a solvent, its energy changes due to interactions with the solvent. This energy alteration is the free energy of solvation, represented as $\Delta G_{solv}$.

##### 3.9.1.1. Implicit solvation models

The first implicit approach, proposed by Born and Onsager, entails substituting a specific solvent-solute interaction with an average field, denoted as $V_{int}$, incorporated into the Hamiltonian within the Hartree-Fock approximation [135]. However, this method has limitations as it does not offer comprehensive insights into phenomena such as hydrophilic effects and HBs.

One of the most widely utilized implicit solvation models is the Polarizable Continuum Model (PCM) [138-142], which employs the Apparent Surface Charge (ASC) methodology [142, 143]. Within PCM, the solvent is viewed as a polarizable continuum characterized by a dielectric constant (ε) [142, 143]. The solute is depicted as embedded within a cavity surrounded by the solvent. The boundary of this cavity is determined through the superimposition of interlocking atomic spheres with radii approximating the VDW radii of the relevant atoms [142, 143]. The points of intersection are smoothed by 'rolling' a probe sphere, sized appropriately for the specific solvent, along the surface of these interlocking spheres [142, 143]. The free energy of solvation holds considerable importance in the thermodynamic characterization of solutions, playing a



© University of Venda

pivotal role in various fields, including chemistry, biology, and pharmaceutical science. Within the PCM approach, the free energy of solvation is expressed as;

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{dr} + \Delta G_{cav}, \quad (3.72)$$

where, $\Delta G_{el}$ is the free energy change related to electrostatic interactions

$\Delta G_{dr}$ is the free energy change related to dispersion interactions

$\Delta G_{cav}$ is the free energy change corresponding to the formation of the cavity within the solvent system to host the solute molecule (cavitation energy).

$\Delta G_{dr}$ and $\Delta G_{cav}$ together form the non-electrostatic component of the free energy of solvation, $G_{non-el}$. Thus, the total free energy of solvation can also be written as:

$$\Delta G_{solv} = G_{el} + G_{non-el}. \quad (3.73)$$

Within the PCM family, several methods exist for calculating the free energy of solvation [138-142], including dielectric polarizable continuum model (DPCM) [144], conductor-like polarizable continuum model (CPCM) [144-146], isodensity polarizable continuum model (IPCM) [147], self-consistent isodensity polarizable continuum model (SCIPCM) [147], and integral equation formalism polarizable continuum model (IEFPCM) [142, 148]. However, among these methods, CPCM stands out as one of the most commonly used approaches due to its computational efficiency and widespread use in calculating the free energy of solvation.

### 3.9.1.2. Explicit solvation models

The explicit solvation approach uses molecular dynamics or Monte Carlo methods, where a solute molecule is placed within a cubic box containing a finite number of solvent molecules [135, 136]. By applying Periodic Boundary Conditions, this finite set of particles simulates behavior as part of a larger bulk system [135, 136]. Consequently, this approach considers the combined effects of the bulk solvent alongside interactions between solvent-solute and solvent-solvent molecules [135, 136]. However, a drawback of this method is its computational expense.



© University of Venda

### 3.10. Selection of the solvents for calculations in solution in the current study

Solvents are substances that are capable of dissolving other substances to form a homogenous solution. Solvents have different properties, such as polarity, dielectric constant, hydrogen bonding ability, and viscosity, among others. These properties determine how well a solvent can interact with a solute and influence the physical and chemical properties of the solute-solvent system.

When studying the behavior of molecules in solution, the choice of solvent is critical. The polarity of the solvent molecules and their ability to interact with the solute, such as through intermolecular HBs, have a significant impact on determining the solute-solvent interactions and the conformational preferences of the solute. Polarity is an important property of solvents that determines their ability to dissolve polar or non-polar substances. Polar solvents have a high dielectric constant and can dissolve polar substances, such as salts, and polar organic molecules, like alcohols, because they can form strong electrostatic interactions with them. On the other hand, non-polar solvents like chloroform or benzene have a low dielectric constant and can not dissolve polar substances. They are only able to dissolve non-polar substances such as alkanes, alkenes, and aromatics because they share the same non-polar nature.

In the case of acylphloroglucinols, which have several HB donor or acceptor groups, the possibility of solute-solvent intermolecular hydrogen bonds needs to be considered. Additionally, biologically active molecules can be present in non-polar media within living organisms. As a result, in this study, the selection of solvents included both polar and non-polar solvents to mimic the situation in living organisms where these molecules may act.

The selected solvents for this study were water ($\varepsilon = 78.39$), acetonitrile ($\varepsilon = 36.64$), and chloroform ($\varepsilon = 4.90$), each with a different dielectric constant, which enables informative comparisons of the situation of the solute molecules in regions with different polarities. The different polarities of these solvents entirely cover the polarity range of possible media in which acylphloroglucinols molecules may be present within a living organism, according to individual compounds' preferences for more or less polar solvents.



© University of Venda

The solvents selected also have different hydrogen bonding abilities. Water was selected because it has unique properties, including its ability to form hydrogen bonds with both the hydrogen bond donor and acceptor centers of the acylphloroglucinol molecules, as well as among themselves. Furthermore, water is abundant and low-cost, making it a practical choice for many research applications. Acetonitrile can only form intermolecular hydrogen bonds with the H-bond donors of acylphloroglucinols, and it has no solvent-solvent intermolecular hydrogen bonds. Acetonitrile is widely used in various chemical reactions and analytical techniques, such as high-pressure liquid chromatography (HPLC) and NMR spectroscopy. Lastly, chloroform was chosen because it can not form intermolecular hydrogen bonds and is useful for studying the behavior of hydrophobic molecules in non-polar environments.

### 3.11. Computational approaches in the study of biologically active molecules

Computational approaches in the study of biologically active molecules refer to the use of computer simulations and mathematical models to understand the structure, function, and interactions of biological molecules. Computational approaches, also known as *in silico* experiments, have become a valuable tool in many areas of scientific research, particularly in the field of biology. These methods are often complex and require specialized knowledge in computational chemistry, molecular biology, and other fields. *In silico* studies can provide valuable insights into molecular behavior and interactions and can help to reduce the time and cost required for physical experimentation. However, they are not a substitute for physical experimentation and must be validated using experimental data. One of the key advantages of these approaches is that they enable researchers to study complex biological systems in a controlled and systematic way. This is particularly important when it comes to understanding biological processes that occur over long time scales or that involve many different components and interactions.

Computational approaches, on the other hand, use computer simulations to model the interactions between drugs and their targets, allowing for virtual screening of large numbers of compounds in a fraction of the time and cost required for traditional laboratory testing. This enables researchers to rapidly identify promising candidates for further development and testing. In summary, the use of computational approaches has transformed the study of biologically active molecules and drug design. These approaches provide valuable insights into biological processes, support the



© University of Venda

optimization of drug candidates, and greatly enhance the efficiency of the drug development process.

### 3.12. Molecular docking

Molecular docking has been utilized since the early 1980s to gain insights into and make predictions about molecular recognition. Its main goal is identifying potential binding modes and estimating the binding affinity [149, 150]. This method has gained significant attention in drug design, as it has proven valuable. A study by the Tufts Center for the Study of Drug Development in 2016 shows that the expenses involved in developing and introducing a new drug to the market have risen by nearly 145% over the past decade [151]. Additionally, although the average duration for bringing a drug to clinical trials has decreased, the success rate of obtaining approval from the United States (US) Food and Drug Administration (FDA) has dropped to 12% [151]. However, computer-aided drug design has played an important role in reducing costs and accelerating drug discovery by guiding experimental research toward the most promising compounds more efficiently [152]. Computer-aided drug design has greatly benefited from the implementation of molecular docking and other related techniques. These methods are a valuable alternative to the time-consuming and costly process of high-throughput screening (HTS) in experiments. These methods have proven successful in generating informed predictions, leading to the discovery of numerous promising lead compounds. Several approved drugs, including imatinib, zanamivir, nelfinavir, erdafitinib, and various clinical candidates, have been identified or optimized using computational methods

In principle, molecular docking involves two key stages. The first stage focuses on predicting how a ligand will conform and identifying its position within the protein binding spot; this position is often called 'the pose'. The next stage involves the use of the scoring function to assess the quality of the binding mode. The sampling algorithms involved in the initial stage and the scoring functions will be briefly discussed in the next sections.

#### 3.12.1. Sampling algorithms

Many sampling algorithms have been developed to date. Algorithms searching for the conformational flexibility of small molecules are divided into three main categories: systematic,



© University of Venda

stochastic, and deterministic methods. Each category includes different approaches that can be used to identify the most energetically favorable conformation of a molecule. The following sub-sections briefly discuss some of these approaches to highlight their main concepts.

#### 3.12.1.1. Geometry-based methods

Geometry-based methods are widely used in molecular shape mapping and ligand docking in protein active sites. One of the popular techniques under this category is Matching Algorithms (MA). These algorithms use shape features and chemical information to match a ligand into a protein active site, representing the protein and ligand as pharmacophores. The distance between each pharmacophore within the protein and ligand is calculated for a match.

Moreover, new ligand conformations are ruled by the distance matrix between the pharmacophore and corresponding ligand atoms. The chemical properties, such as hydrogen-bond donors and acceptors, can be considered during the match. MAs have significant advantages in terms of speed; thus, they are commonly used to enrich active compounds from large libraries. Several programs like DOCK, FLOG [153], LibDock [154], and SANDOCK [155] are available for matching algorithms for ligand docking.

#### 3.12.1.2. Systematic methods

Systematic search methods for ligand docking can be classified into three categories: exhaustive, fragmentation, and conformational ensemble methods. Each method handles ligand flexibility differently.

##### 3.12.1.2.1. Exhaustive method

The exhaustive method involves systematically rotating all possible rotatable bonds in the ligand at a given interval but is limited to small, relatively flexible ligands due to the huge number of possible combinations. To make this approach practical, geometrical or chemical constraints are often applied to the initial screening of the ligand poses. An example of a popular program using the exhaustive sampling method is Glide [156].



© University of Venda

### 3.12.1.2.2. Fragment-based methods

Fragment-based methods are a popular approach for the *de novo* design of ligands and modifications of known ligands that may enhance their binding to a target protein [152]. Fragmentation methods divide the ligand into modular pieces and incrementally grow the ligand binding conformation by adding fragments one at a time [150]. One of the fragments is anchored to the protein binding site, and the anchor is typically chosen to be the fragment that shows maximum interaction complementarity with the receptor surface and has the minimum number of alternate conformations. However, these methods are also limited to medium and smaller-sized ligands, similar to exhaustive methods [150].

There are several methods included in this approach, such as incremental construction (IC) [157], Multiple Copy Simultaneous Search (MCSS) [158, 159], and ligand unbiased dynamic search (LUDI) [159]. In the process of IC methods, a ligand is fragmented by breaking its rotatable bonds, resulting in multiple fragments. From these fragments, one is selected to be docked into the active site as an anchor. Typically, the chosen fragment is the largest or one that plays an important functional role or interacts significantly with the protein. The remaining fragments are then systematically incorporated. To fit-in in the active site, different orientations are generated, allowing for the ligand to exhibit flexibility. This method has been used in DOCK 4.0 [33], FlexX [157], Hammerhead [160], SLIDE [161], and eHiTS [162].

MCSS makes thousands of multiple copies of a functional group, which are randomly placed in the considered binding site [150]. Through energy minimization and or quenched MD in the protein's forcefield, the copies undergo optimization. It is important to note that the copies only interact with the protein and not each other. This allows for the identification of a collection of favorable binding sites and orientations for the functional group, determined by their interaction energies. The binding site can be effectively mapped by the use of different functional groups. Through the linkage of these different functional groups, it becomes possible to design new molecules that perfectly match the binding site.

On the other hand, LUDI methods are based on the formation of hydrogen bonds and hydrophobic contacts between the ligand and protein [150]. LUDI’s approach is based on rules about energetically favorable non-bonded contact geometries between the protein and ligand functional



© University of Venda

groups. This approach helps to optimize hydrogen bonding and hydrophobic interactions, which both stabilize the ligand at the target site and help to increase its affinity for the protein [163]. After generating the interaction sites, LUDI fits the fragment onto the interaction sites and evaluates the fit according to distance criteria [164]. The final step is the connection of some or all the fitted fragments to a single molecule [150].

#### 3.12.1.2.3. Conformational ensemble methods

Conformational ensemble methods represent ligand flexibility by rigidly docking an ensemble of pre-generated ligand conformations [150]. This approach removes the computational cost of exploring ligand conformational space, but it involves additional tools to generate the required ensemble of conformations of the ligand. A potential drawback of this method is the possibility of the inclusion of the bioactive conformation from the collection of generated conformers.

### 3.12.1.3. Stochastic methods

Stochastic algorithms sample ligand binding orientations and conformations by introducing ligand modifications based on randomly generated values at each step [150, 165]. Algorithm-dependent criteria determine whether to accept or reject changes. Stochastic algorithms have the advantage of generating many molecular conformations and navigating various energy possibilities, thereby increasing the probability of finding a global energy minimum [150]. The drawback of this approach is that it is computationally expensive. Different types of stochastic algorithms are used in molecular docking, including Monte Carlo, ant colony optimization (ACO), genetic algorithm, and tabu search methods. Each of these algorithms differs in the way they generate given moves and the probability criteria of acceptance [150].

#### 3.12.1.3.1. Monte Carlo methods

Monte Carlo (MC) methods have proven to be valuable in determining the ligand pose in the binding pocket. In these methods, different conformations are generated by bond rotation or rigid-body movements (translation or rotation) [150, 152]. Each conformation then goes through an energy-based selection test. If the conformation passes the criterion, it will be saved and further modified to generate the next conformation, and the process will proceed until the predefined quantity of conformations is obtained. So, the newly generated conformation is retained if the new



© University of Venda

internal energy is lower than previously. Otherwise, it is rejected with a Boltzmann probability based on the effective temperature [166]. One of the advantages of this method is that the ligands are described using internal coordinates, which is less time-consuming than using 3D cartesian coordinates, and pre-calculated grid potentials are often used to speed up the process [148]. Another advantage is that the Monte Carlo method is superior to molecular dynamics-based simulation methods as it allows the ligand to cross the energy barriers on the PES [152]. Programs that can use the Monte Carlo method include an earlier version of AutoDock, ICM, QXP, and Affinity [167].

#### 3.12.1.3.2. Ant colony optimization method

Ant colony optimization (ACO) is a method that replicates the behavior of real ants, finding the shortest direct path between a nest and a food source [150]. In the context of protein-ligand docking, an artificial ant colony is utilized to search for the lowest energy pose of the ligand in the binding site. The ants imitate the behavior of real ants by marking low-energy ligand conformations with pheromone trails. These pheromone trail signals are then adjusted in subsequent iterations to produce low-energy pose with higher chances [168]. PLANTS [168], a docking program that uses ACO, is an excellent example of this approach.

#### 3.12.1.3.3. Tabu search method

Tabu search (TS) is a variant of the Monte Carlo approach. The fundamental concept behind this approach is to prevent the search algorithm from revisiting already explored space by either prohibiting or imposing penalties on moves that lead to solution space that has already been visited [165]. Spatial constraints are organized into a "taboo list" of previously identified solutions. Then, a "tabu list" is used to arrange spatial constraints based on previously obtained solutions. The search starts from a single solution and uses Gaussian or Cauchy random variables to generate different conformations through a mutation-like process. These newly generated conformations are then evaluated and assigned scores, with the highest-scoring conformation ultimately being selected as the final solution [165]. PRO_LEADS [169] and PSI-DOCK [170] are one of the most well-known TS-based methods.



© University of Venda

#### 3.12.1.3.4. Genetic algorithm method

In a genetic algorithm method, the ligand fitting points are mapped onto the complementary protein's fitting points, and each member of the population is encoded as a 'chromosome' that contains information about this mapping [150, 152, 165]. The chromosomes are represented as binary strings called genes, which encode for different degrees of freedom, such as ligand rotation and translation or torsional angles. The initial generation of solutions evolved using crossover and mutation operators, which respectively act on two-parent solutions or one single solution. The resulting 'children' elements are subjected to random mutations [150, 152, 165]. In this process, the genetic operators affect the genes, and the result is a new ligand structure. New structures are assessed by a scoring function, and the ones that survived and exceeded a threshold can be used for the next generation. The fitness score of each chromosome is based on its predicted binding affinity, and the chromosomes within the population are ranked according to the fitness score. AutoDock [171], DIVALI [172], GOLD [173], and DARWIN [174] are just a few examples of software applications that have successfully incorporated genetic algorithms into their frameworks.

### 3.12.1.4. Deterministic methods

Deterministic methods rely primarily on the architecture of the first conformation. This dependence governs the moves that can be made to generate the next state. However, these methods face the problem of being trapped in local minima due to the high energetic barriers that must be overcome [175]. To address this problem, deterministic MD-based docking algorithms have explored two divergent mathematical approaches.

The first approach involves implementing simulated annealing with various integrated tricks to jump over barriers. This approach has been successful in several docking programs. The second approach, heuristic TS, forces sampling unexplored space during docking simulations by maintaining lists of already explored conformations. The second approach is heuristic-TS, which forces the sampling of unexplored space during docking simulations by maintaining a list of explored conformations [165]. Thus, searches that approach these regions are penalized.



© University of Venda

### 3.12.2. Scoring functions

Scoring functions are efficient mathematical methods to predict the degree of interaction, or binding affinity, between multiple molecules. The scoring function serves the purpose of distinguishing between correct and incorrect poses, as well as identifying binders from inactive compounds, all within a reasonable amount of computation time [152]. There are four points to take into account when evaluating the dependability of a scoring function. These factors include [176];

- scoring power, which refers to the capacity to generate scores that are highly correlated with experimental binding affinity data
- ranking power, which involves correctly ranking a group of ligands that bind to a common target protein based on their binding affinities when their binding poses are known
- docking power, which involves identifying the native binding pose of a ligand as the one with the best score, and when screening a large set of generated decoy poses
- screening power, which involves identifying the true binders to a given target protein among a library of random molecules.

While an ideal scoring function would perform equally well on all four tasks, current scoring functions tend to excel at only one or two of these tasks at a time [150]. When it comes to scoring functions, the focus lies on estimating the binding affinity between the protein and ligand rather than directly calculating it. These functions rely on different assumptions and simplifications to achieve their purpose. Scoring functions can be grouped into classical scoring functions and machine learning-based scoring functions [177].

#### 3.12.2.1. Classical scoring functions

Classical scoring functions (CSFs) are based on the idea that change in free energy can break down when a ligand binds to its target into several separate energy contributions, which are linearly combined. However, it is not always the case that such a linear correlation will exist [178]. The general expression of CSFs can be written as:

$$E_{bind} = \sum_{m=1}^{M} W_m x_m. \quad (3.74)$$



© University of Venda

Classical scoring functions have two significant drawbacks: limited depiction of protein flexibility and the implicit handling of solvent [150]. These methods can be subdivided into three classes: physics-based, empirical, and knowledge-based.

### 3.12.2.2. Force-field-based scoring function

Force-field-based scoring functions calculate the binding energy of the protein-ligand complex by adding up the effect of bonded interactions (such as bond stretching, angle bending, and torsion angles) and non-bonded interactions (such as van der Waals and electrostatic interactions) as shown in equation 3.4 [150, 152]. The extended force-field-based scoring function can take the form:

$$V = W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}} + W_{sol}\sum_{i,j}\left(S_i V_{j+} S_j V_i\right)e^{\left(-\frac{r_g^2}{2\sigma^2}\right)}. \quad (3.75)$$

To determine the energy between two atoms, i and j, the combined effects of van der Waals forces, hydrogen bonding, coulombic interactions, and desolvation are considered. The empirical free energy is calibrated using weight factors, denoted as W. The initial term within the equation accounts for the VDW component and is characterized by a Lennard-Jones potential function. The second term represents the hydrogen bond contribution and is usually explained by adding an additional term to the bond energy. The electrostatic component, determined using a Coulombic formula, is represented by the third term. However, the fourth term poses challenges when considering solvation and how the solvent interacts with the binding of the ligand. To address this problem, the use of implicit solvent methods like Poisson-Boltzmann (PB) or Generalised-Born (GB) continuum solvation models is a common practice [179]. Alternatively, computationally intensive approaches that explicitly account for water molecules, such as free energy perturbation (FEP) [180, 181] and thermodynamic integration (TI) [182], can also be used. In addition, the extensions can also consider the contributions of entropy. Several software programs that consider these functions include DOCK [183], GOLD [184], and AutoDock [171].



© University of Venda

### 3.12.2.3. Empirical scoring functions

Empirical scoring functions sum up energetic components such as hydrogen bonds, hydrophobic effects, ionic interactions, protein-ligand clashes, and binding entropy to estimate the binding energy of a protein-ligand complex [150, 152]. When calculating a final score, the coefficients derived from multiple linear regression analyses are used to multiply each component and then get added together. These coefficients are obtained by fitting the regression model into a training set of protein-ligand complexes with known binding affinities [150]. The empirical scoring function can be expressed as;

$$\Delta G = \Delta G_0 + \Delta G_{rot} + N_{rot} + \Delta G_{hb} \sum_{neutral\ H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{io} \sum_{ion\ init.} f(\Delta R, \Delta\alpha) + \Delta G_{aro} \sum_{aro\ int.} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \sum_{lipo\ cont.} f^*(\Delta R), \quad (3.76)$$

where, $\Delta G$ refers to the estimated free energy of binding,

$\Delta G_0$ refers to the regression constant

$\Delta G_{rot}$, $\Delta G_{hb}$, $\Delta G_{io}$, $\Delta G_{aro}$, $\Delta G_{lipo}$ represents regression coefficients for each corresponding free energy term

$f(\Delta R, \Delta\alpha)$ refers to the scaling function penalizing deviations from the ideal geometry

$N_{rot}$ refers to the number of rotate bonds that are immobilized in the complex.

When comparing force-field scoring functions to empirical scoring functions, it is evident that the latter is significantly faster in calculating binding scores [150]. Hence, this is primarily due to the simplified treatment of energy terms in empirical scoring functions, making the evaluation process relatively straightforward. The accuracy and coverage of the protein-ligand training set used to develop the model directly influence the accuracy of the empirical scoring functions, as seen in the correlation between these factors. LUDI [184], ChemScore [185], GlideScore [34], and ChemPLP [186, 187] are among the examples of empirical scoring functions.

### 3.12.2.4. Knowledge-based scoring functions

In knowledge-based scoring functions [153], statistical analysis is used to derive the interatomic contact frequencies and distance between protein and their docked ligand. Through the use of the



© University of Venda

Boltzmann law, these functions convert atom pair preferences into distance-dependent pairwise potentials, as illustrated in equation (3.77);

$$PMFF_{score} = \sum_{\substack{kI \\ r<r_{cut-off}^{ij}}} A_{ij}(r) \tag{3.77}$$

whereby,

$$A_{ij}(r) = -k_B TIn\left[f_{Vol_{corr}}^{j}(r)\frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}}\right]; \tag{3.78}$$

where, kB, represents the Boltzmann constant,

T, represents the absolute temperature of the system,

$\rho_{ij}(r)$, denotes the number density of the protein-ligand atom pair at distance r in the training set,

$\rho^*(r)$, represents the pair density in a reference state where interatomic interactions are absent.

The protein-ligand density functions $\rho_{ij}(r)$ are not determined by averaging different states of the same protein but by summing the static densities observed in different proteins. The scoring mechanism considers the preferred contacts and repulsive interactions between every atom in the ligand and protein within a specified cutoff. This is because knowledge-based scoring functions operate under the assumption that favorable interactions are more frequent than unfavorable ones. Because of this, the final score gets determined by prioritizing preferred contacts and penalizing repulsive interactions within the designated cutoff.

Knowledge-based functions have the advantage of computational simplicity, which makes them ideal for screening large compound databases. Additionally, these functions can effectively model interactions such as sulphur-aromatic or cation-, which are often neglected in empirical approaches. In comparison to physics-based and empirical scoring functions, knowledge-based scoring functions strike a favorable balance between accuracy and speed by avoiding the need for ab initio calculations or reproducing binding affinities. Furthermore, their insensitivity to the training set is attributed to the large and diverse protein-ligand dataset database [188].



© University of Venda

Despite the advancements in deriving knowledge-based scoring functions, these functions still face challenges. One of these challenges is the underrepresentation of certain interactions in the limited training sets of crystal structures and the inherent bias in selecting proteins for successful structure determination. This can lead to parameters that may not be suitable for widespread use, especially when dealing with interactions involving metals or halogens.

Another challenge is the calculation of the reference state ($\rho^*(r)$) [189]. Currently, two classical strategies are used to determine the reference state: the traditional atom-randomized reference state and the corrected reference state. The former approximates the reference state by randomly distributing atomic pairs in the training set. However, this approach neglects the effects of excluded volume and interatomic connectivity [189].

Later approaches introduced correction terms such as the volume factor correction term for the reference state to address these limitations [190]. Despite these efforts, accurately calculating the reference state remains a challenge for knowledge-based scoring functions. This is particularly relevant regarding binding mode predictions and virtual screening, as the pairwise potentials derived from bound structures may not be sensitive enough to different ligand positions, potentially yielding good scores even for bad binding modes.

A third approach to address this issue is to use iterative methods to bypass the reference state's accurate calculation. The basic idea behind this method is to adjust the pair potentials through iteration until the interaction potentials replicate the experimentally determined pair distribution function in the training set. This results in a set of potentials that can be used effectively.

Several knowledge-based functions, such as GOLD/ASP [191], PMF [153], DrugScore [192], SMoG [193], and Bleep [194], differ in their training set sizes, energy function formats, atom type definitions, distance cutoffs, and other parameters.

[188].

### 3.12.2.5. Machine learning-based scoring functions

Machine learning scoring functions (MLSFs) have been developed using a different set of descriptors derived from protein-ligand complexes, and a subset of these models have



© University of Venda

demonstrated better performance compared to CSFs. In contrast to (CSFs), MLSFs use advanced algorithms, including gradient boosting decision tree (GBDT), random forests (RF), support vector machines (SVM), artificial neural network (ANN), and deep learning (DL), to effectively model and solve non-linear problems [150]. The binding energy on MLSF can be expressed as;

$$E_{bind} = f_{ML}(x_m), \qquad ML = FR, GBDT, SVM, DL, ANN. \tag{3.79}$$

In order to develop a dependable machine learning-based scoring function (MLSF), it is necessary to incorporate three essential components [195]:

- a suitable machine learning algorithm capable of performing classification and/or regression analysis
- an accurate representation of protein-ligand interactions
- a reliable benchmark dataset that can be used for constructing and evaluating the performance of the models.



© University of Venda

### 3.13. References

1. Ebbin, D. D., & Gammon, S. (2009). General chemistry (9th ed.). Houghton Mifflin Co.
2. Chang, R. (2010). Chemistry (10th ed.). McGraw Hill.
3. Brown, T. L., LeMay, H. E., Bursten, B. E., Murphy, C. J., & Woodward, P. M. (2017). Chemistry: The central science (14th ed.). Pearson.
4. Wüthrich, K. (1992). The use of NMR spectroscopy in determining the three-dimensional structures of proteins. *Annual Review of Biophysics and Biomolecular Structure*, 21, 593–626. https://doi.org/10.1146/annurev.bb.21.060192.001051
5. Kornberg, R. D. (1999). Structural basis of RNA polymerase II transcription initiation. Cell, 99(3), 259–262. https://doi.org/10.1016/S0092-8674(00)81779-9
6. Pauling, L. (1939). The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structural chemistry. Cornell University Press.
7. Pauling, L. (1940). The nature of the chemical bond: and the structure of molecules and crystals (No. 04; RMD, QD469 P3.).
8. Dove, M. T. (2003). Structure and dynamics: An atomic view of materials (Vol. 1). Oxford University Press.
9. Schlegel, H. B. (2003). Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *Journal of Computational Chemistry*, 24(12), 1514–1527. https://doi.org/10.1002/jcc.10231
10. Billeter, S. R., Turner, A. J., & Thiel, W. (2000). Linear scaling geometry optimization and transition state search in hybrid delocalized internal coordinates. *Physical Chemistry Chemical Physics*, 2(10), 2177–2186. https://doi.org/10.1039/a909486e
11. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3, 1–14.
12. Weser, O., Hein-Janke, B., & Mata, R. (2022). Automated handling of complex chemical structures in z-matrix coordinates—the chemcoord library. *Journal of Computational Chemistry*, 44(5), 710–726. https://doi.org/10.1002/jcc.27029
13. Bristol University. (n.d.). Molecular modelling: Energy and forces. http://www.chm.bris.ac.uk/pt/ajm/mmhtm/MM_L2p16.htm (Accessed: July 23, 2023)
14. Stereoelectronics. (n.d.). Computational chemistry software guide. https://www.stereoelectronics.org/webSC/SC102.html (Accessed: June 23, 2023)



© University of Venda

15. Cotton, F. A. (1991). Chemical applications of group theory. John Wiley & Sons.
16. Vendrell, O., Gatti, F., Lauvergnat, D., & Meyer, H. D. (2007). Full-dimensional (15-dimensional) quantum-dynamical simulation of the protonated water dimer. I. Hamiltonian setup and analysis of the ground vibrational state. *The Journal of Chemical Physics*, 127(18).
17. Krogsgaard-Larsen, P., Liljefors, T., & Madsen, U. (2002). Textbook of drug design and discovery (3rd ed.). Taylor & Francis.
18. Roy, K., Kar, S., & Das, R. N. (2015). Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Academic Press.
19. González, M. A. (2011). Force fields and molecular dynamics simulations. École Thématique de la Société Française de la Neutronique, 12, 169–200.
20. Hehre, W. J. (2003). A guide to molecular mechanics and quantum chemical calculations. Wavefunction Inc.
21. MacKerell Jr, A. D. (2004). Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry*, 25(13), 1584–1604.
22. Lii, J. H., & Allinger, N. L. (1991). Molecular mechanics analysis. *Journal of Computational Chemistry*, 12, 186–199.
23. Sun, H. (1998). J. Phys. Chem. B, 102, 7338–7346.
24. Derreumaux, P., & Vergoten, G. (1995). J. Chem. Phys, 102, 8586.
25. Halgren, T. A. (1996). Merck molecular force field. *Journal of Computational Chemistry*, 17, 490–516.
26. Vanommeslaeghe, K., & Guvench, O. (2014). Molecular mechanics. *Current Pharmaceutical Design*, 20(20), 3281–3292.
27. Martin, P. (n.d.). Structural bioinformatics. http://ncbr.muni.cz/~martinp/C3210/StructBioinf7.pdf (Accessed: May 14, 2023)
28. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry, 26*(16), 1668-1688.
29. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., ... Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry, 30*(10), 1545-1614.



© University of Venda

30. Allinger, N. L. (1977). Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *Journal of the American Chemical Society*, *99*(25), 8127–8134. *(MM3 is an extension of the MM2 force field).*
31. van Gunsteren, W. F., & Berendsen, H. J. C. (1987). GROMOS: GROningen MOlecular Simulation package. *Biomolecular Chemistry*. *(Later developments can be found in studies like Oostenbrink et al., 2004.)*
32. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry, 30*(16), 2785-2791.
33. Ewing, T. J. A., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design, 15*(5), 411–428.
34. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry*, 47(7), 1750–1759.
35. Abagyan, R., & Totrov, M. (1994). Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *Journal of Molecular Biology, 235*(3), 983–1002. *(ICM is part of this development in docking studies.)*
36. Rodrigues, J. P., & Bonvin, A. M. (2014). Integrative computational modeling of protein interactions. *The FEBS journal*, *281*(8), 1988-2003.
37. Senftle, T. P., Hong, S., Islam, M. M., Kylasa, S. B., Zheng, Y., Shin, Y. K., ... & Van Duin, A. C. (2016). The ReaxFF reactive force-field: development, applications and future directions. *npj Computational Materials*, *2*(1), 1-14.
38. Dorsett, H., & White, A. (2000). Overview of molecular modelling and ab initio molecular orbital methods suitable for use with energetic materials. *DSTO Aeronautical and Maritime Research Laboratory*, 1–39.
39. Zettili, N. (2009). Quantum mechanics: Concepts and applications. Wiley.
40. Sutcliffe, B. T. (1992). The Born-Oppenheimer Approximation. In *Methods in Computational Molecular Physics* (pp. 19-46). Boston, MA: Springer US.
41. Szabo, A., & Ostlund, N. S. (1989). Modern quantum chemistry: Introduction to advanced electronic structure theory. Dover Publications.



© University of Venda

42. Szabo, A., & Ostlund, N. S. (2012). Modern quantum chemistry: Introduction to advanced electronic structure theory. Courier Corporation.
43. Helgaker, T., Jorgensen, P., & Olsen, J. (2000). Molecular electronic-structure theory. Wiley.
44. Moldenhauer, P. A., & Pitzer, R. M. (1977). Hartree-Fock theory. *The Journal of Chemical Physics*, 66(3), 851–861. https://doi.org/10.1063/1.434813
45. Helgaker, T., Klopper, W., & Koch, H. (2000). The Hartree-Fock method. *The Journal of Chemical Theory and Computation*, 6(1), 1–24.
46. Levine, I. N. (2005). Quantum chemistry (6th ed.). Prentice Hall.
47. Harrison, W. A. (2013). Electronic structure and the properties of solids. Dover Publications.
48. Cremer, D. (2011). Møller–Plesset perturbation theory: From small molecule methods to methods for thousands of atoms. Wiley Interdisciplinary Reviews: Computational Molecular Science, 1(4), 509-530.
49. Møller, C., & Plesset, M. S. (1934). Note on an approximation treatment for many-electron systems. *Physical Review*, 46(7), 618.
50. Jensen, F. (2017). Introduction to computational chemistry. John Wiley & Sons.
51. Olsen, J., Jørgensen, P., Helgaker, T., & Christiansen, O. (2000). Divergence in Møller–Plesset theory: A simple explanation based on a two-state model. *The Journal of Chemical Physics*, 112(22), 9736–9748.
52. Kraka, E., Gauss, J., & Cremer, D. (1991). Determination and use of response densities in "40 years in quantum chemistry". *Journal of Molecular Structure* (Theochem), 234, 95–126.
53. Gauss, J., & Cremer, D. (1992). Analytical energy gradients in Møller–Plesset perturbation and quadratic configuration interaction methods: Theory and application. *Advances in Quantum Chemistry*, 23, 205–299.
54. He, Z., & Cremer, D. (1996). Sixth-order many-body perturbation theory. IV. Improvement of the Møller–Plesset correlation energy series by using Pade, Feenberg, and other approximations up to sixth order. *International Journal of Quantum Chemistry*, 59, 71–95.
55. Purvis, G. D., & Bartlett, R. J. (1982). A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics*, *76*(4), 1910-1918.
56. Crawford, T. D., Schaefer, H. F., Lipkowitz, K. B., & Boyd, D. B. (2000). Reviews in Computational Chemistry.



© University of Venda

57. Raghavachari, K., & Trucks, G. W. (1989). Highly correlated systems. Excitation energies of first row transition metals Sc–Cu. *The Journal of chemical physics*, *91*(2), 1062-1065.
58. Knizia, G., & Chan, G. K. L. (2012). Density matrix embedding: A simple alternative to dynamical mean-field theory. *Physical review letters*, *109*(18), 186404.
59. Bartlett, R. J., & Musiał, M. (2007). Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, *79*(1), 291-352.
60. Bartlett, R. J., & Shavitt, I. (1977). Comparison of high-order many-body perturbation theory and configuration interaction for $H_2O$. *Chemical Physics Letters*, *50*(2), 190-198.
61. Kohn, W., & Sham, L. J. (1996). Density functional theory. In Proceedings of the Italian Physical Society (Vol. 49, pp. 561–572). Editrice Compositori.
62. Cohen, A. J., Mori-Sánchez, P., & Yang, W. (2012). Challenges for density functional theory. *Chemical Reviews*, 112(1), 289–320.
63. Sandvik, A. W. (2007). An introduction to quantum monte carlo methods. In Strongly Correlated Magnetic and Superconducting Systems: Proceedings of the El Escorial Summer School Held in Madrid, Spain, 15–19 July 1996 (pp. 109-135). Berlin, Heidelberg: Springer Berlin Heidelberg.
64. Ceperley, D. M. (2018). An overview of quantum Monte Carlo methods. In Theoretical and Computational Methods in Mineral Physics: Geophysical Applications (pp. 129-136). Walter de Gruyter GmbH. https://doi.org/10.2138/rmg.2010.71.6
65. Acioli, P. H. (1997). Review of quantum Monte Carlo methods and their applications. *Journal of Molecular Structure: THEOCHEM*, 394(2–3), 75–85.
66. Lüchow, A. (2011). Quantum Monte Carlo methods. Wiley Interdisciplinary Reviews: *Computational Molecular Science*, 1(3), 388–402.
67. Pople, J. A., & Beveridge, D. L. (1970). Molecular orbital theory. McGraw-Hill.
68. Hehre, W. J. (1976). Ab initio molecular orbital theory. *Accounts of Chemical Research*, 9(11), 399–406.
69. Sánchez-Portal, D., Ordejon, P., Artacho, E., & Soler, J. M. (1997). Density-functional method for very large systems with LCAO basis sets. *International Journal of Quantum Chemistry*, 65(5), 453–461.
70. Kier, L. (2012). Molecular orbital theory in drug research (Vol. 10). Elsevier.



© University of Venda

71. Alias, A. N. (2010). Organic semiconductor characterization using Linear Combination of Atomic Orbital (LCAO). Esteem Academic Journal, 6(1), 1–13.
72. Ramachandran, K. I., Gopakumar, D., & Namboori, K. (2008). Basis sets. Computational Chemistry and Molecular Modeling: Principles and Applications, 115-138.
73. Jensen, F. (2013). Atomic orbital basis sets. Wiley Interdisciplinary Reviews: Computational Molecular Science, 3(3), 273–295.
74. Pople, J. A., & Hehre, W. J. (1978). Computation of electron repulsion integrals involving contracted Gaussian basis functions. *Journal of Computational Physics*, 27(2), 161–168.
75. Wilson, S. (1987). Basis sets. Advances in Chemical Physics: Ab initio Methods in Quantum Chemistry Part I, 67, 439-500.
76. Frisch, M. J., Pople, J. A., & Binkley, J. S. (1984). Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *The Journal of Chemical Physics*, 80(7), 3265–3269.
77. Papajak, E., Zheng, J., Xu, X., Leverentz, H. R., & Truhlar, D. G. (2011). Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions. *Journal of Chemical Theory and Computation*, 7(10), 3027–3034.
78. Huzinaga, S. (1985). Basis sets for molecular calculations. *Computer Physics Reports*, 2(6), 281–339.
79. Dunning Jr, T. H. (1989). Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2), 1007–1023.
80. Holm, S., Unzueta, P. A., Thompson, K., & Martínez, T. J. (2023). Single-Point Extrapolation to the Complete Basis Set Limit through Deep Learning. Journal of Chemical Theory and Computation, 19(14), 4474-4483.
81. Jensen, F. (2001). Polarization consistent basis sets: Principles. *The Journal of Chemical Physics*, 115(20), 9113–9125.
82. Hill, A. N., Meijer, A. J., & Hill, J. G. (2022). Correlation consistent basis sets and core polarization potentials for Al–Ar with ccECP pseudopotentials. *The Journal of Physical Chemistry A*, 126(34), 5853–5863.
83. Drabik, G., & Radoń, M. (2024). Approaching the Complete Basis Set Limit for Transition Metal Spin–State Energetics.



© University of Venda

84. Boys, S. F., & Bernardi, F. (1970). The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics*, 19(4), 553–566.
85. UTEP. (n.d.). Molecular dynamics simulation tutorial. https://www.utep.edu/science/upbit/curriculum/mdsim.html (Accessed: August 13, 2023)
86. Reveles, J. U., & Köster, A. M. (2004). Geometry optimization in density functional methods. *Journal of Computational Chemistry*, 25(9), 1109–1116.
87. von Arnim, M., & Ahlrichs, R. (1999). Geometry optimization in generalized natural internal coordinates. *The Journal of Chemical Physics*, 111(20), 9183–9190.
88. Baker, J., Kessi, A., & Delley, B. (1996). The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of Chemical Physics*, 105(1), 192–212.
89. Schlegel, H. B. (2011). Geometry optimization. Wiley Interdisciplinary Reviews: *Computational Molecular Science*, 1(5), 790–809.
90. Verbeke, J., & Cools, R. (1995). The Newton-Raphson method. *International Journal of Mathematical Education in Science and Technology*, 26(2), 177–193.
91. Hennig, P., & Kiefel, M. (2013). Quasi-Newton methods: A new direction. *The Journal of Machine Learning Research*, 14(1), 843–865.
92. Mishra, S. K., Ram, B., Mishra, S. K., & Ram, B. (2019). Conjugate gradient methods. Introduction to Unconstrained Optimization with R, 211-244.
93. Meza, J. C. (2010). Steepest descent. Wiley Interdisciplinary Reviews: Computational Statistics, 2(6), 719–722.
94. Yuan, Y. X. (2006). A new stepsize for the steepest descent method. *Journal of Computational Mathematics*, 149–156.
95. Cornell University. (n.d.). Conjugate gradient methods. https://optimization.cbe.cornell.edu/index.php?title=Conjugate_gradient_methods (Accessed: July 23, 2023)
96. Pulay, P., & Fogarasi, G. (1992). Geometry optimization in redundant internal coordinates. *The Journal of Chemical Physics*, 96(4), 2856–2860.
97. Baker, J., Kessi, A., & Delley, B. (1996). The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of Chemical Physics*, 105(1), 192–212.



© University of Venda

98. Kutzelnigg, W. (1989). Ab initio calculation of molecular properties. *Journal of Molecular Structure: THEOCHEM*, 202, 11–61.

99. Gerlt, J. A., Kreevoy, M. M., Cleland, W. W., & Frey, P. A. (1997). Chemistry & Biology, 4, 259.

100. Perrin, C. L., & Nielson, J. B. (1997). Annual Review of Physical Chemistry, 48, 511.

101. Zundel, G. (2007). Hydrogen bonds with large proton polarizability and proton transfer processes in electrochemistry and biology. *Advances in chemical physics*, 3, 1.

102. Leiserowitz, L. (1976). Acta Crystallographica Section B, 32, 775.

103. Bernstein, J., Etter, M. C., & Leiserowitz, L. (1994). The role of hydrogen bonding in molecular assemblies. *Structure correlation*, 431-507.

104. Lehn, J. M. (1990). Angewandte Chemie International Edition in English, 29, 1304.

105. Pokorná, P., Krepl, M., Kruse, H., & Šponer, J. (2017). *Journal of Chemical Theory and Computation*, 13, 5658.

106. Zhao, G. J., & Han, K. L. (2012). Hydrogen bonding in the electronic excited state. *Accounts of Chemical Research*, 45(3), 404–413.

107. Kojić-Prodić, B., & Molčanov, K. (2008). The nature of hydrogen bonds: New insights into old theories. *Acta Chimica Slovenica*, 55(4).

108. Buckingham, A. D., Del Bene, J. E., & McDowell, S. A. C. (2008). The hydrogen bond. *Chemical Physics Letters*, 463(1–3), 1–10.

109. Arunan, E., Desiraju, G. R., Klein, R. A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D. C., Crabtree, R. H., Dannenberg, J. J., Hobza, P., & Kjaergaard, H. G. (2011). Definition of the hydrogen bond (IUPAC recommendations 2011). *Pure and Applied Chemistry*, 83(8), 1637–1641.

110. Purdue University. (n.d.). Hydrogen bonds. https://www.chem.purdue.edu/gchelp/liquids/hbond.html (Accessed: August 24, 2023)

111. Paoloni, L. (1959). Nature of the hydrogen bond. *The Journal of Chemical Physics*, 30(4), 1045–1058.

112. Rozas, I. (2007). On the nature of hydrogen bonds: An overview on computational studies and a word about patterns. *Physical Chemistry Chemical Physics*, 9(22), 2782–2790.

113. Lozano-Casal, P. (2006). Structural and computational studies of small organic and biological molecules.



© University of Venda

114. Chickos, J. S., & Acree Jr, W. E. (2002). Enthalpies of sublimation of organic and organometallic compounds. 1910–2001. *Journal of Physical and Chemical Reference Data*, 31(2), 537–698.

115. Gavezzotti, A. (1998). The crystal packing of organic molecules: Challenge and fascination below 1000 Da. *Crystallography Reviews*, 7(1), 5–121.

116. Del Bene, J. E., & Jordan, M. J. T. (2001). Journal of Molecular Structure, 573, 11–23.

117. Tsuzuki, S., & Luthi, H. P. (2001). The Journal of Chemical Physics, 114, 3949–3957.

118. Tarek Ibrahim, M., Wait, E., & Ren, P. (2024). Quantum Mechanics Characterization of Non-Covalent Interaction in Nucleotide Fragments. *Molecules*, *29*(14), 3258.

119. Altun, A., Neese, F., & Bistoni, G. (2018). Local energy decomposition analysis of hydrogen-bonded dimers within a domain-based pair natural orbital coupled cluster study. *Beilstein journal of organic chemistry*, *14*(1), 919-929.

120. Rackers, J. A., & Ponder, J. W. (2019). Classical Pauli repulsion: An anisotropic, atomic multipole model. The Journal of chemical physics, 150(8).

121. Grabowski, S. J. (2006). Theoretical studies of strong hydrogen bonds. Annual Reports Section C: *Physical Chemistry*, 102, 131–165.

122. Alkorta, I., Rozas, I., & Elguero, J. (1998). Non-conventional hydrogen bonds. *Chemical Society Reviews*, 27(2), 163–170.

123. Desiraju, G. R. (2002). Hydrogen bridges in crystal engineering: Interactions without borders. *Accounts of Chemical Research*, 35(7), 565–573.

124. Cleland, W. W., & Kreevoy, M. M. (1994). Low-barrier hydrogen bonds and enzymic catalysis. *Science*, 264(5167), 1887–1890.

125. Rozas, I., Alkorta, I., & Elguero, J. (2000). Behavior of ylides containing N, O, and C atoms as hydrogen bond acceptors. *Journal of the American Chemical Society*, 122(45), 11154–11161.

126. Gilli, G., & Gilli, P. (2000). Towards an unified hydrogen-bond theory. *Journal of Molecular Structure*, 552(1-3), 1-15.

127. Meot-Ner, M. (2005). The ionic hydrogen bond. *Chemical Reviews*, 105(1), 213–284.

128. Gilli, G., Bellucci, F., Ferretti, V., & Bertolasi, V. (1989). Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the beta-diketone fragment. *Journal of the American Chemical Society*, 111(3), 1023.



© University of Venda

129. Huggins, M. L. (1936). Hydrogen bridges in organic compounds. *The Journal of Organic Chemistry*, 1(5), 407-456.

130. Bertolasi, V., Pretto, L., Gilli, G., & Gilli, P. (2006). π-Bond cooperativity and anticooperativity effects in resonance-assisted hydrogen bonds (RAHBs). Acta Crystallographica Section B: *Structural Science*, 62(5), 850–863.

131. Gilli, P., & Gilli, G. (2010). Hydrogen bond models and theories: The dual hydrogen bond model and its consequences. *Journal of Molecular Structure*, 972, 270.

132. Hibbert, F., & Emsley, J. (1990). Hydrogen bonding and chemical reactivity. *Advances in Physical Organic Chemistry*, 26, 255–379.

133. Yakupov, M. Z., Lyapina, N. K., Shereshovets, V. V., & Imashev, U. B. (2001). The solvent effect on the rate of reaction between propanethiol and chlorine dioxide. *Kinetics and catalysis*, *42*, 609-612.

134. Chattaraj, P. K. (2009). Chemical reactivity theory: A density functional view. CRC Press.

135. Charlotte, F. F. (1987). General Hartree-Fock program. *Computer Physics Communications*, 43, 365.

136. Wiberg, K. B. (1997). Properties of some condensed aromatic systems. *Journal of Organic Chemistry*, 62, 5727.

137. Miertus, S., & Tomasi, J. (1982). Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chemical Physics*, 65, 245.

138. Barone, V., & Cossi, M. (1997). A new definition of cavities for the computation of solvation free energy by the polarizable continuum model. *The Journal of Chemical Physics*, 107, 3221.

139. Tomas, J., Mennucci, B., & Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chemical Reviews*, 105, 3093.

140. Barone, V., Cossi, M., & Tomas, J. (1998). Geometry optimization of molecular structures in solution by the polarizable continuum model. *Journal of Computational Chemistry*, 19, 417.

141. Cossi, M., Scalmani, G., Rega, N., & Barone, V. (1997). New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *The Journal of Chemical Physics*, 107, 3041.



© University of Venda

142. Cancès, E., Mennucci, B., & Tomas, J. (1997). A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *The Journal of Chemical Physics*, 107, 3041.

143. Cossi, M. (2004). Continuum solvation model for infinite systems. *Chemical Physics Letters*, 384, 179.

144. Foresman, J. B., & Frisch, A. (1995). Exploring chemistry with electronic structure methods. Gaussian, Inc.

145. Frediani, L., Cammi, R., Corni, S., & Tomas, J. (2004). A polarizable continuum model for molecules at diffuse interfaces. *The Journal of Chemical Physics*, 120, 3893.

146. Cossi, M., Rega, N., Scalmani, G., & Barone, V. (2003). Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *Journal of Computational Chemistry*, 24, 681.

147. Okeke, I. N., Aboderin, O. A., Byarugaba, D. K., Ojo, K. K., & Opintan, J. A. (2007). Growing problem of multidrug-resistant enteric pathogens in Africa. Emerging Infectious Diseases, 13, 1645.

148. Tomas, J., Mennucci, B., & Cancès, E. (1999). The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *Journal of Molecular Structure (Theochem),* 464, 211.

149. Ripphausen, P., Nisius, B., Peltason, L., & Bajorath, J. (2010). Quo vadis, virtual screening? A comprehensive survey of prospective applications. *Journal of medicinal chemistry*, *53*(24), 8461-8467.

150. Stanzione, F., Giangreco, I., & Cole, J. C. (2021). Use of molecular docking computational tools in drug discovery. *Progress in Medicinal Chemistry*, 60, 273–343.

151. DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of health economics*, *47*, 20-33.

152. 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.

153. Muegge, I., & Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *Journal of Medicinal Chemistry*, *42*(5), 791–



© University of Venda

804. *(FLOG was developed as part of docking methods, sometimes referenced alongside scoring functions.)*

154. Diller, D. J., & Merz, K. M. (2001). High throughput docking for library design and library prioritization. *Proteins: Structure, Function, and Bioinformatics*, *43*(2), 113–124.

155. Good, A. C., & Whiteside, G. W. (1999). SANDOCK: A new screening technique for automated docking. *Proteins: Structure, Function, and Bioinformatics, 35*(4), 512–525.

156. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., & Shaw, D. E. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7), 1739–1749.

157. Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261, 470–489.

158. Miranker, A., & Karplus, M. (1991). Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins: Structure, Function, and Bioinformatics*, 11(1), 29–34.

159. Eisen, M. B., Wiley, D. C., Karplus, M., & Hubbard, R. E. (1994). HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins: Structure, Function, and Bioinformatics*, 19(3), 199–222.

160. Welch, W., Ruppert, J., & Jain, A. N. (1996). Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chemistry & Biology*, 3(6), 449–462.

161. Schnecke, V., & Kuhn, L. A. (2000). Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design*, 20, 171–190.

162. Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S., & Johnson, A. P. (2006). eHiTS: An innovative approach to the docking and scoring function problems. *Current Protein and Peptide Science*, 7, 421–435.

163. Patil, R., Das, S., Stanley, A., Yadav, L., Sudhakar, A., & Varma, A. K. (2010). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface lead the pathways of drug-designing. PLoS One, 5(8), e12029.

164. Böhm, H. J. (1992). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*, 6, 593–606.



© University of Venda

165. Abraham, D. J. (2003). Burger's medicinal chemistry and drug discovery: Docking and scoring in drug discovery. https://doi.org/10.1002/0471266949.bmc140

166. Sotriffer, C., Klebe, G., Stahl, M., & Böhm, H. J. (2003). Docking and scoring functions/virtual screening. Burger's Medicinal Chemistry and Drug Discovery, 281–331.

167. Squires, M., Ward, G., Saxty, G., Berdini, V., Cleasby, A., King, P., Angibaud, P., Perera, T., Fazal, L., Ross, D., & Jones, C. G. (2011). Potent, selective inhibitors of fibroblast growth factor receptor define fibroblast growth factor dependence in preclinical cancer models. *Molecular Cancer Therapeutics*, 10(9), 1542–1552.

168. Korb, O., Stützle, T., & Exner, T. E. (2006). PLANTS: Application of ant colony optimization to structure-based drug design. In Ant Colony Optimization and Swarm Intelligence (pp. 247–258). Springer.

169. Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R., & Eldridge, M. D. (1998). Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins: Structure, Function, and Bioinformatics, 33(3), 367–382.

170. Pei, J., Wang, Q., Liu, Z., Li, Q., Yang, K., & Lai, L. (2006). PSI-DOCK: Towards highly efficient and accurate flexible ligand docking. *Proteins: Structure, Function, and Bioinformatics*, 62(4), 934–946.

171. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14), 1639–1662.

172. Clark, K. P., & Ajay. (1995). Flexible ligand docking without parameter adjustment across four ligand–receptor complexes. *Journal of Computational Chemistry*, 16(10), 1210–1226.

173. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52(4), 609–623.

174. Taylor, J. S., & Burnett, R. M. (2000). DARWIN: A program for docking flexible molecules. *Proteins: Structure, Function, and Bioinformatics*, 41(2), 173–191.

175. Brooijmans, N., & Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*, 32(1), 335–373.



© University of Venda

176. Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der Deutschen Chemischen Gesellschaft, 27(3), 2985–2993.

177. Li, J., Fu, A., & Zhang, L. (2019). An overview of scoring functions used for protein–ligand interactions in molecular docking. Interdisciplinary Sciences*: Computational Life Sciences*, 11, 320–328.

178. Ain, Q. U., Aleksandrova, A., Roessler, F. D., & Ballester, P. J. (2015). Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdisciplinary Reviews: *Computational Molecular Science*, 5(6), 405–424.

179. Zou, X., Yaxiong, K., & Kuntz, I. D. (1999). Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *Journal of the American Chemical Society*, 121(35), 8033–8043.

180. Kollman, P. (1993). Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, 93(7), 2395–2417.

181. Briggs, J. M., Marrone, T. J., & McCammon, J. A. (1996). Computational science: New horizons and relevance to pharmaceutical design. *Trends in Cardiovascular Medicine*, 6(6), 198–203.

182. Rotstein, S. H., & Murcko, M. A. (1993). GenStar: A method for de novo drug design. *Journal of Computer-Aided Molecular Design*, 7, 23–43.

183. Åqvist, J., Luzhkov, V. B., & Brandsdal, B. O. (2002). Ligand binding affinities from MD simulations. *Accounts of Chemical Research*, 35(6), 358–365.

184. Khamis, M. A., Gomaa, W., & Ahmed, W. F. (2015). Machine learning in computational docking. *Artificial Intelligence in Medicine*, 63(3), 135–152.

185. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, 11, 425–445.

186. Korb, O., Stützle, T., & Exner, T. E. (2006). PLANTS: Application of ant colony optimization to structure-based drug design. In International Workshop on Ant Colony Optimization and Swarm Intelligence (pp. 247–258). Springer.



© University of Venda

187. Li, J., Fu, A., & Zhang, L. (2019). An overview of scoring functions used for protein–ligand interactions in molecular docking. Interdisciplinary Sciences: *Computational Life Sciences*, 11, 320–328.

188. Huang, S. Y., & Zou, X. (2006). An iterative knowledge-based scoring function to predict protein–ligand interactions: II. Validation of the scoring function. *Journal of Computational Chemistry*, 27(15), 1876–1882.

189. Thomas, P. D., & Dill, K. A. (1996). Statistical potentials extracted from protein structures: How accurate are they? *Journal of Molecular Biology*, 257(2), 457–469.

190. Muegge, I. (2006). PMF scoring revisited. *Journal of Medicinal Chemistry*, 49(20), 5895–5902.

191. Mooij, W. T., & Verdonk, M. L. (2005). General and targeted statistical potentials for protein–ligand interactions. *Proteins: Structure, Function, and Bioinformatics*, 61(2), 272–287.

192. Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, 295(2), 337–356.

193. DeWitte, R. S., & Shakhnovich, E. I. (1996). SMoG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *Journal of the American Chemical Society*, 118(47), 11733–11744.

194. Mitchell, J. B., Laskowski, R. A., Alex, A., & Thornton, J. M. (1999). BLEEP—potential of mean force describing protein–ligand interactions: I. Generating potential. *Journal of Computational Chemistry*, 20(11), 1165–1176.

195. Zhang, X., Shen, C., Jiang, D., Zhang, J., Ye, Q., Xu, L., Hou, T., Pan, P., & Kang, Y. (2023). TB-IECS: An accurate machine learning-based scoring function for virtual screening. *Journal of Cheminformatics*, 15(1), 63.



© University of Venda

# CHAPTER 4

# COMPUTATIONAL STUDIES OF EUGLOBALS

## 4.1. Introduction to the chapter

Euglobals (EGs) with anticancer or antimalarial properties have been introduced in chapter 2, section 2.4.2.4. This chapter presents a computational study of selected EGs *in-vacuo*, in-solution, and through docking studies. Detailed results are respectively presented in Sections 4.4.1, 4.5.1 and 4.6.1. The results presented in Section 4.4.1 have already been published [1].

The primary objective of the studies *in-vacuo* and in solution was those of providing extensive understanding of the molecular properties of EG molecules in different media. The primary objective of the docking studies is to analyze the protein-ligand interactions and to prioritize candidates with favorable pharmacokinetic and safety profiles.

The acylphloroglucinol moiety of EGs contains two acyl groups (Fig. 2.19). The current study focuses on two types of EG molecules: those in which the R of one acyl group is R = H (so that the acyl group is an aldehyde group) and the R in the other acyl group is R = isobutyl, and those where both acyl groups are aldehydes.

## 4.2. Computational details

### 4.2.1. Methodology for in-vacuo and in-solution studies.

The selection of computational methods was guided by two primary criteria: achieving a balance between result quality and computational cost and ensuring meaningful comparisons with previous studies on ACPLs. The first criterion necessitated considering the considerable size of EG molecules, the number of molecules under consideration, and the numerous potential conformers for each molecule. The second criterion emphasized the importance of employing methods consistent with prior research. This facilitates comparisons of obtained values and trends with those established as general—or nearly general—for ACPLs or ACPL subclasses with specific characteristics.



© University of Venda

The use of multiple methods to verify obtained results and inferred trends is considered advantageous, as different methods may yield better evaluations of certain properties while providing weaker evaluations of others. The selected levels of theory encompass two ab initio methods: Hartree-Fock (HF) and second-order Møller-Plesset Perturbation Theory (MP2, [2]), as well as Density Functional Theory (DFT).

HF is the least expensive method, while MP2 is the most costly. HF accounts for electron correlation only to a limited extent, primarily through Pauli's exclusion principle. In contrast, DFT incorporates some electron correlation effects, and MP2 considers both electron correlation and dispersion effects, which are important for accurately describing hydrogen bonding.

HF and MP2 calculations were performed using the 6-31G(d,p) basis set chosen for its lower computational cost. Previous studies on ACPLs (e.g., [3-8]), have utilized HF/6-31G(d,p) calculations, demonstrating reasonable geometry parameters and trend identification for ACPLs. It was considered valuable to verify these findings for EGs. Using the 6-31G(d,p) basis set for MP2 is necessary for a meaningful comparison with HF results due to the high computational expense of MP2 calculations for molecules of this size.

DFT calculations employed the widely used B3LYP functional [9, 10] and the 6-31+G(d,p) basis set. Previous studies have demonstrated the importance of diffuse functions on heavy atoms for accurate DFT results for ACPLs studies [3-8]. Comparative studies have shown that including diffuse functions on heavy atoms improves agreement with experimental data, particularly for properties related to molecular polarization [11]. Given the capabilities of the three methods, MP2 results are considered the most suitable benchmarks.

Vibrational frequencies (harmonic approximations) were computed at the HF/6-31G(d,p) and DFT/B3LYP/6-31G+(d,p) levels on the respective optimized geometries, with the computed values scaled by factors of 0.9632 for DFT results and 0.8992 for HF results [12]. Frequency calculations are crucial for confirming whether the obtained stationary points represent true minima and for providing zero-point energy (ZPE) corrections, which are related to vibrational motions, as well as energy values corrected for ZPE. Additionally, changes in the vibrational



© University of Venda

frequency of the OH groups acting as hydrogen bond donors enable comparisons of hydrogen bond strengths.

All calculations were conducted with fully relaxed geometry (full optimization) in a vacuum environment. The HF-optimized geometries served as input geometries for DFT and MP2 calculations. The initial step involved preparing the 3D structures using Gauss View 4.1.2 software. Afterward, minimization and optimization were performed using Gaussian-09, Revision E.01 [13], following the previously mentioned methods. Each molecule produced multiple conformers as part of this procedure. Visualization of the input and output geometries utilised Gauss View 4.1.2 [14], Chem3D Ultra and ChemDraw Ultra [15].

In solution calculations utilized the conductor-like polarizable continuum model (CPCM). These calculations were carried out at the HF and DFT levels with fully relaxed geometry. PCM re-optimization is important for identifying conformational changes induced by the solvent.

All energy values presented in this study are given in kcal $mol^{-1}$, and distances are provided in Ångströms (Å). For simplicity, the calculation methods are abbreviated as follows: DFT for DFT/B3LYP/6-31G + (d,p), HF for HF/6-31G(d,p), and MP2 for MP2/6-31G(d,p).

### 4.2.2. Methodology for docking Studies

**Ligands preparation:** The docking studies involved selecting the lowest energy conformers from the DFT results as ligands for docking. The relevant properties predicted by the DFT calculations were retained without any changes. The ionization states of the ligands remained unchanged, and no alternative tautomers or chiral configurations were generated different from those used in the initial DFT calculations.

**Preparation of protein molecules**: The following X-ray crystal structures of anticancer proteins were obtained from the Protein Data Bank (PDB) [16-18]: EGFR (PDB; 6LUD), P13K (PDB; 5JHB), BRAF V600b (PDB; 6V34), JAK3 (PDB; 7C3N), CDK-2 (PDB; 1DI8), Topo I (PDB; IT8I), H5P90 (PDB; 3TUH), HER2 (PDB; 3RCD), and antimalarial proteins PFPMT (PDB; 3UJ9), PFLDH (PDB; 1U5A), PFMDH (PDB; 6R8G). The Maestro 9.3 protein preparation wizard



© University of Venda

application was utilized to refine the raw PDB structures. This involved tasks such as adding hydrogen atoms, assigning bond orders, creating zero-order bonds to metals, forming disulfide bonds, adjusting charges, and orienting groups.

**Receptor grid generation:** GLIDE molecular docking requires a ligand to bind with the X-ray crystal structure of the protein in order to establish the active site receptor grid. This receptor grid-based docking enables ligands to bind in multiple possible conformations. To set up the receptor grid, a grid file was generated from the prepared protein using the protein preparation wizard. In the Define Receptor section, the option for Pick to Identify the Ligand (Molecule) and Show Markers was checked. Subsequently, an atom was selected in the ligand to define the active site. The scaling factor and partial charge cutoff for van der Waals radius scaling were set to 0.25 and 1 Å, respectively. Other parameters such as sites, constraints, rotatable groups, and excluded volume were maintained at their default settings as per Maestro 9.3 [19, 20].

.

**GLIDE molecular docking:** GLIDE molecular docking procedures were performed following ligand and protein preparation, as well as the definition of the active site grid on the protein. GLIDE utilizes a systematic computational simulation approach to assess specific ligand poses and flexibility. The GLIDE systematic method is characterized by its rapid and accurate molecular docking process, with the GScore serving as its output, which is an empirical scoring function derived from various parameters. The GLIDE XP visualizer module was employed to analyze specific ligand-protein interactions. The ligands selected for docking were evaluated against the X-ray crystal structures of anticancer proteins, EGFR (PDB; 6LUD), P13K (PDB; 5JHB), BRAF V600b (PDB; 6V34), JAK3 (PDB; 7C3N), CDK-2 (PDB; 1DI8), Topo I (PDB; IT8I), H5P90 (PDB; 3TUH), HER2 (PDB; 3RCD), and antimalarial proteins PFPMT (PDB; 3UJ9), PFLDH (PDB; 1U5A), PFMDH (PDB; 6R8G) using GLIDE. The most suitable compounds for each target were determined based on thermodynamic optimal energy values, types of interactions, bonding potential, and conformations [21, 22].

### 4.2.3. Methodology for physicochemical properties studies

Absorption, Distribution, Metabolism, and Excretion (ADME) properties studies were conducted using the QikProp tool from Schrodinger 2012 to predict the physicochemical properties of the



© University of Venda

lowest energy conformers of the ligands molecules obtained from the DFT results. This tool provides predictions for several molecular properties. The following properties related to molecular permeability and absorption [23, 24] were computed for the ACPLs in this study: predicted apparent permeability (QPPCaco, nm/sec), using Caco-2 cells, a human epithelial colorectal adenocarcinoma cell line often utilised as a model for the gut-blood barrier; predicted apparent permeability across MDCK (Madin-Darby Canine Kidney) cells (QPPMDCK, nm/sec), with MDCK cells serving as a suitable model for the blood-brain barrier; predicted blood-brain partition coefficient (QPlogBB); predicted binding affinity to HAS (a human serum albumin, QPlogKhsa); and the percentage of predicted human oral absorption (%PHOA)

### 4.3. Selection of structures

#### 4.3.1. Considered EG molecules in which one acyl group is the aldehyde group and the other has R = isobutyl

The selected EG molecules exhibit two structural isomers distinguished by the placement of the acyl groups-one as an aldehyde and the other with R = isobutyl. The study aims to identify differences in molecular properties resulting from reversing the positions of these two groups while keeping the terpene moiety the same. As a result, two isomers are identified: ε-isomers, where the aldehyde is attached at C1 and the R = isobutyl group at C5, and ζ-isomers, where their positions are reversed. When only one type is reported in literature, the other is modeled to broaden the comparison; the modelled one is identified by adding an asterix (*) in the acronym denoting it.

Significant effects are anticipated due to the non-equivalence of the C1 and C5 positions concerning the terpene moiety. C1 follows the first atom (C2) shared by both moieties, with the subsequent atom (O8) in the terpene moiety serving as an H-bond acceptor. In contrast, C5 is not sequential to the other common atom (C3), and the substituent at the intermediate atom (C4) includes an OH group capable of donating and accepting H-bonds. The grandinol molecule (G) is chosen as a convenient benchmark to explore the terpene moiety's influence on the acylphloroglucinol moiety of EGs, and how this influence varies between the two isomer types. This molecule shares the same acyl groups as the ε/ζ EGs but has a methyl at C3, making the C1 and C5 positions equivalent concerning the substituent.



© University of Venda

EG molecules with two acyl groups can potentially have up to two simultaneous Intramolecular Hydrogen Bonds (IHBs) or just one IHB in different positions, depending on the orientation of the acyl groups and neighboring hydroxyl (OH) groups. The study takes into account all possible IHB patterns.

The significance of this study lies in understanding how the positioning of substituents within a molecule can profoundly affect its activity, a phenomenon well-established by quantitative structure-activity relationship (QSAR) studies on several compounds [25-27]. The sophisticated relationship between a molecule's structure and biological activity highlights the importance of investigating how the arrangement of its constituent parts influences its overall function [28, 29]. For instance, previous QSAR analyses have showed the hydrophobic effect's strong dependence on substituents' positioning [26, 30].

In this study, ten pairs of ε-isomers and ζ-isomers were considered, for a total of 20 molecules, each pair sharing the same terpene moiety. Table 4.1 lists these molecules and their corresponding acronyms in this text.

**Table 4.1.** List of calculated euglobal molecules in which the R chain of one of the acyl groups is R = isobutyl, and acronyms with which they are denoted in this work [1].
The table highlights the pairs of corresponding ε-type / ζ-type molecules.
In the cases in which the second molecule of a pair is not reported in literature, the corresponding molecule was also calculated, and is denoted only by its acronym. Under the column heading 'molecule', the notation a) represents a model molecule with reversed positions of the two acyl groups with respect to the corresponding molecule that has been reported in the literature.

| Terpene moiety | | ε-type molecule | | ζ-type molecule | |
|---|---|---|---|---|---|
| | | molecule | acronym | molecule | Acronym |
| monoterpene | α-phellandrene | Euglobal IIC | M1-ε | Euglobal T1 | M1-ζ |
| | α-pinane | Euglobal G2 | M2-ε | Euglobal G1 | M2-ζ |
| | β-pinane | Euglobal G3 | M3-ε | Euglobal G4 | M3-ζ |
| | pinane | Euglobal G5 | M4-ε | _a) | M4*-ζ |
| | γ- terpinene | Euglobal G6 | M5-ε | Euglobal G7 | M5-ζ |
| | γ- terpinene | Euglobal G8 | M6-ε | _a) | M6*-ζ |
| | α-terpinene | Euglobal G9 | M7-ε | Euglobal G10 | M7-ζ |



© University of Venda

| | | | | | |
|---|---|---|---|---|---|
| | α-terpinene | Euglobal G11 | M8-ε | _a) | M8*-ζ |
| | terpinolene | a) | M9*-ε | Euglobal G12 | M9-ζ |
| sesquiterpene | bicyclogermacrane | Euglobal VII | M10-ε | _a) | M10*-ζ |

Figures 4.1 and 4.2 depict the calculated ζ-isomers and ε-isomers, respectively, providing visual representations of the terpene moieties across different pairs. The molecular formula of M1–M9 molecules is $C_{23}H_{30}O_5$, despite differences in their terpene moieties, with a molecular mass of 386.48 atomic mass units (a.m.u.). M10, on the other hand, has a molecular formula of $C_{28}H_{38}O_5$ and a molecular mass of 454.60 a.m.u.

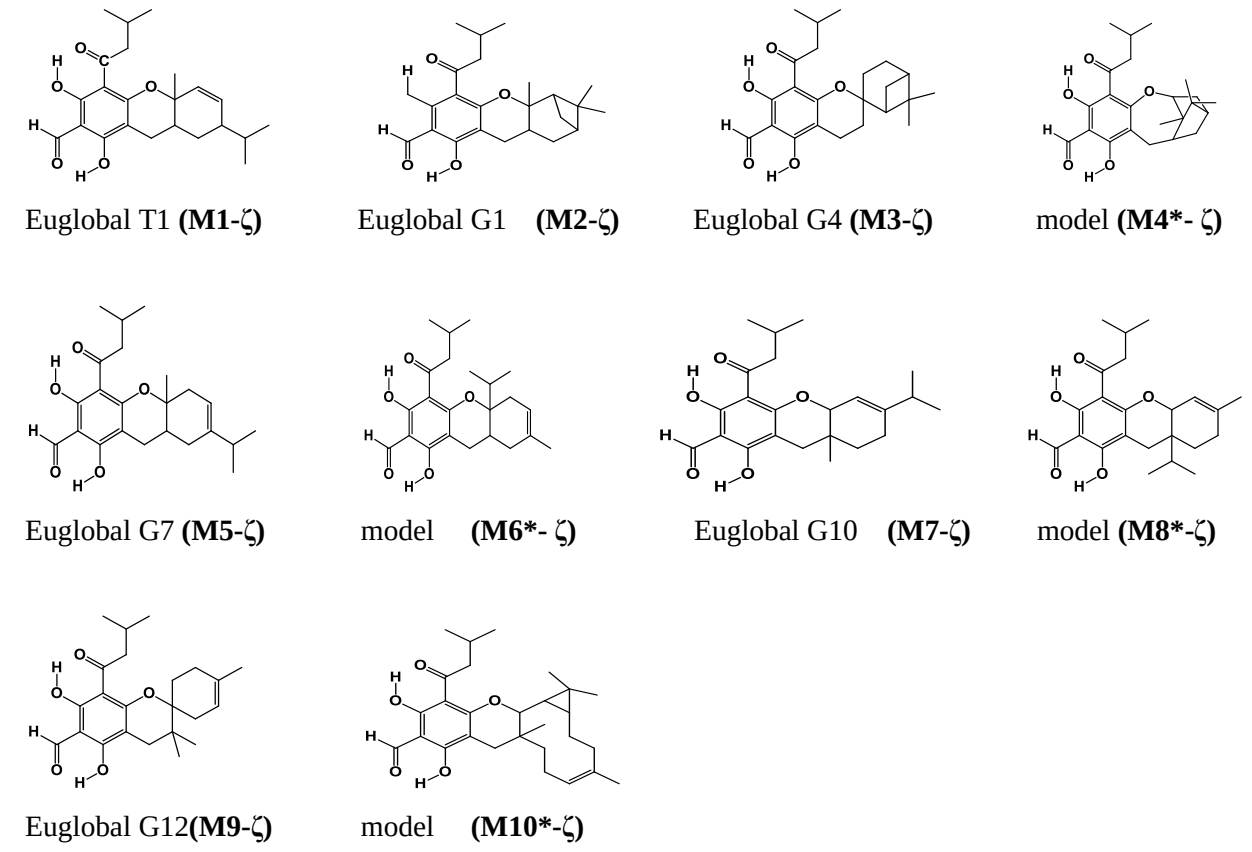
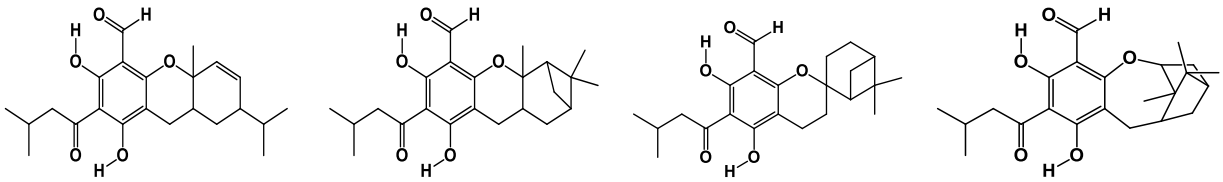


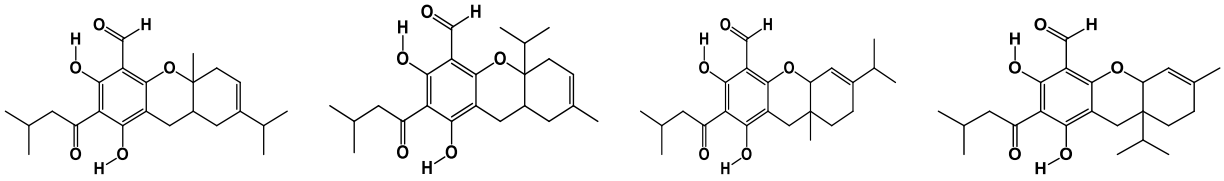

**Figure 4.1. The molecular structures of the considered ζ-isomers EG molecules.**
The asterisk indicates that the considered molecule is a model structure. H atoms attached to C atoms are not displayed to make it easier to visualize the molecular structure.



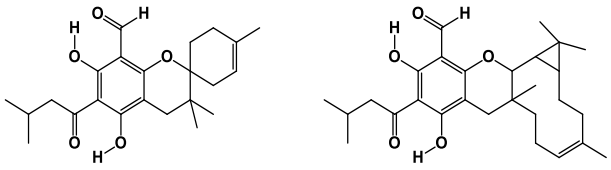
© University of Venda



Euglobal IIC **(M1-ε)** Euglobal G2 **(M2-ε)** Euglobal G3 **(M3-ε)** Euglobal G5 **(M4-ε)**



Euglobal G6 **(M5-ε)** Euglobal G8 **(M6-ε)** Euglobal G9 **(M7-ε)** Euglobal G11 **(M8-ε)**



model **(M9*-ε)** Euglobal VII **(M10-ε)**

**Figure 4.2. The molecular structures of the considered ε-isomers EG molecules.**
The asterisk indicates that the considered molecule is a model structure. H atoms attached to C atoms are not displayed to make it easier to visualize the molecular structure.

### 4.3.2. Considered EG molecules in which both acyl groups are aldehyde groups

In cases where both acyl groups in EG molecules are aldehyde groups (denoted with the symbol τ), there is no effect from position reversal.

This study also considers 18 τ-type EG molecules, divided into two groups: ten models with the same terpene moiety as the calculated ε-isomer/ζ-isomer pairs, and eight selected molecules reported in existing literature. The latter set was chosen to provide ɐ a broader set of results for τ molecules. Figure 4.3 illustrates the calculated τ molecules.



© University of Venda

Molecules M1-τ to M9-τ have the same molecular formula $C_{19}H_{22}O_5$ (despite varying terpene moieties), with a molecular mass of 330.37 a.m.u. Molecule M10-τ has a molecular formula of $C_{24}H_{30}O_5$ and a molecular mass of 398.49 a.m.u.

Molecule N1-τ has a molecular formula of $C_{27}H_{36}O_5$ and a molecular mass of 440.57 a.m.u. Molecules N2-τ and N3 have a molecular formula of $C_{22}H_{28}O_5$ with a molecular mass of 372.45 a.m.u. Molecules N4 and N5-τ have a molecular formula of $C_{23}H_{30}O_5$ with a molecular mass of 386.48 a.m.u. Molecule N6-τ has a molecular formula of $C_{27}H_{34}O_5$ with a molecular mass of 438.56 a.m.u. Molecules N7-τ and N8-τ have a molecular formula of $C_{28}H_{38}O_5$ with a molecular mass of 454.60 a.m.u.

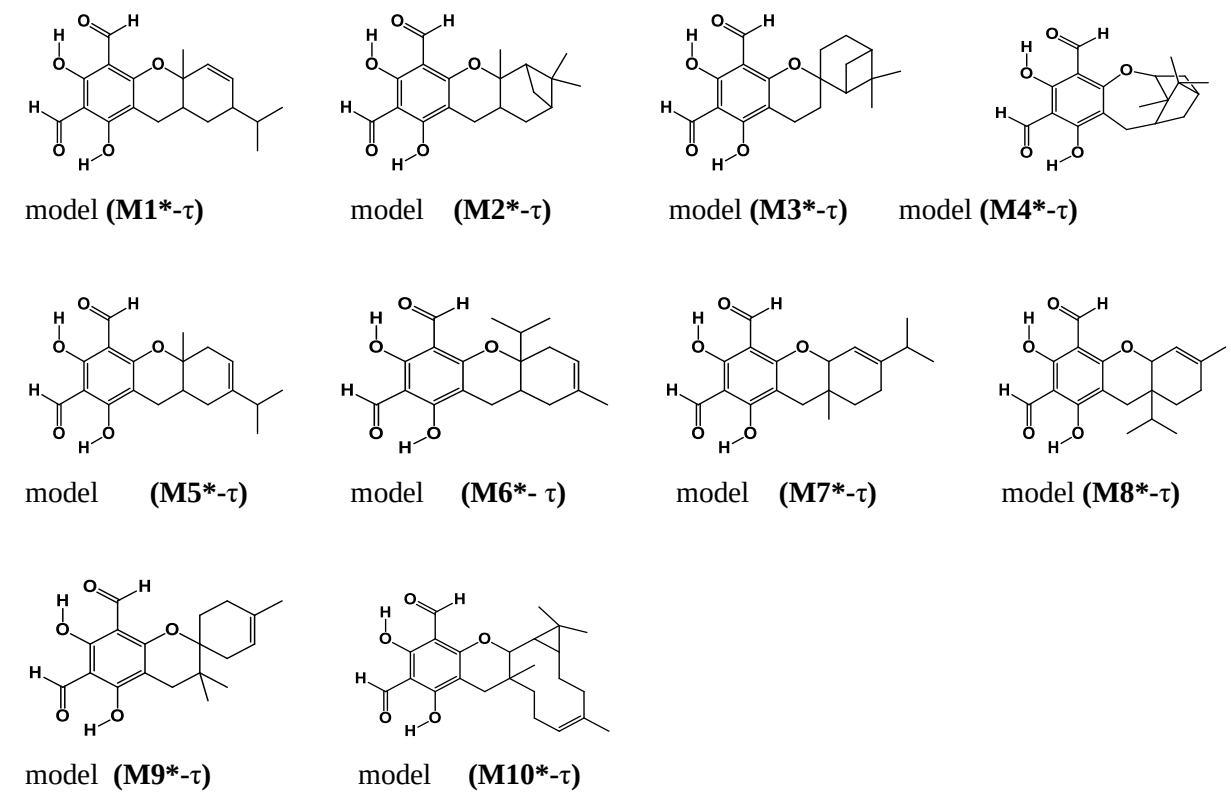



**Figure 4.3. The molecular structures of the considered τ -type EG molecules (model structures).** H atoms attached to C atoms are not displayed to make it easier to visualize the molecular structure.



© University of Venda

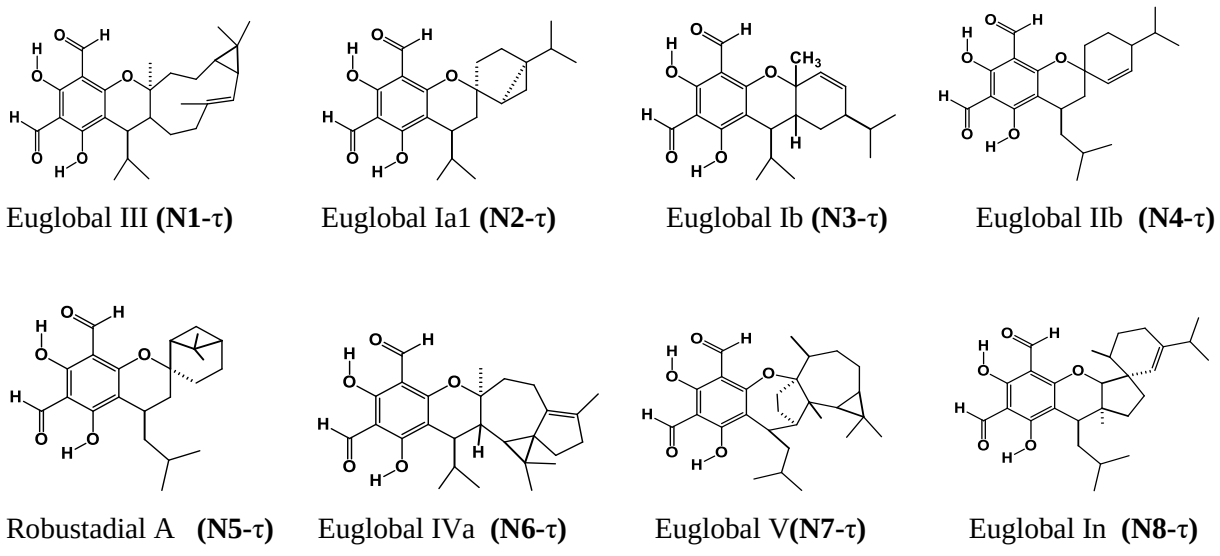



**Figure 4.4. The molecular structures of the considered τ-type EG molecules (reported from the literature).**
H atoms attached to C atoms are not displayed to make it easier to visualize the molecular structure.

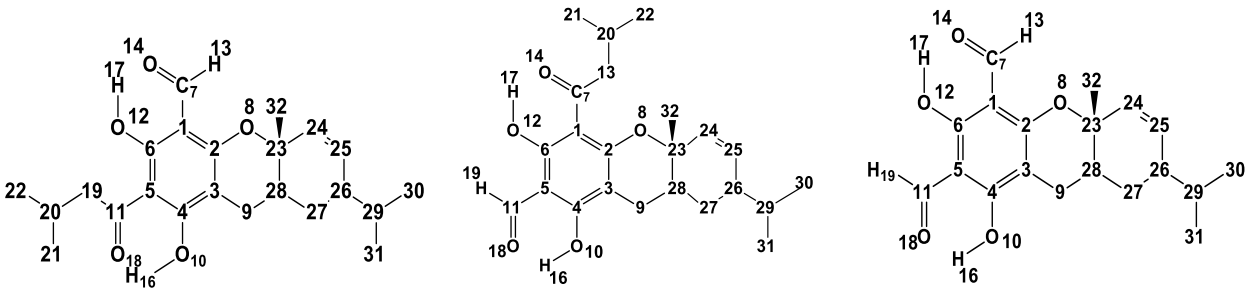
## 4.4. *In-vacuo* results

### 4.4.1. Naming of structures and conformers

In Figure 4.5, the atom numbering scheme is designed to maintain consistency with the numbering conventions utilized in previous studies of ACPLs for the acylphloroglucinol moiety ensuring consistency [4, 5]. The O atom within the terpene moiety corresponds to O8 in the general structure of ACPLs (Fig. 2.4.1), with the remaining atoms of the terpene moiety numbered sequentially after the atoms of the acylphloroglucinol moiety. While the terpene moieties differ across different pairs, necessitating distinct atom numbering, this specific analysis primarily focuses on the atoms within the acylphloroglucinol moiety. Therefore, individualized numbering for the other terpene moieties is not introduced.



© University of Venda



ε-type isomer ζ-type isomer τ-type structure

**Figure 4.5. Illustration of ε-isomer/ζ-isomer pairs and a τ-type structure with an identical monoterpene or sesquiterpene moiety.** The atom numbering utilised for each structure in this study is also shown. In this study, the atom numbering of the ACPL moiety is consistent across all the molecules; however, for different isomer pairs, the atom numbering of the terpene moiety depended on its nature. In order to provide a clear representation of the molecular structure, the carbon atoms are assigned numbers, and the hydrogen atoms bonded to them are not displayed; instead, they are assigned the same number as their corresponding carbon atom. When there is a need to differentiate between two hydrogen atoms attached to a single carbon atom, subscripts 'a' and 'b' can be used (e.g., H13a and H13b).

Most of the EGs studied here are being analyzed for the first time. EG1 (M2-ζ), EG2 (M2-ε), and EG4 (M3-ζ), previously analyzed using HF and DFT methods [31,32], have been reanalyzed with Gaussian-09. This allows for fair comparisons with the results of other EGs and adds the MP2 level, which was not used before.

To keep things organized, structures and conformers are labeled with acronyms, as a practice found useful for all other ACPLs (e.g., [4, 5]). These acronyms use symbols to show the type of molecule and its characteristics. For instance, the first part of the acronym identifies the molecule, and the second part identifies the conformer. For example, an acronym starts with "M" for "molecule," followed by a number indicating the terpene moiety and a Greek letter (ε or ζ) for the isomer. If a structure is  in nature but is introduced as a model for comparison, an asterisk is added after the number.

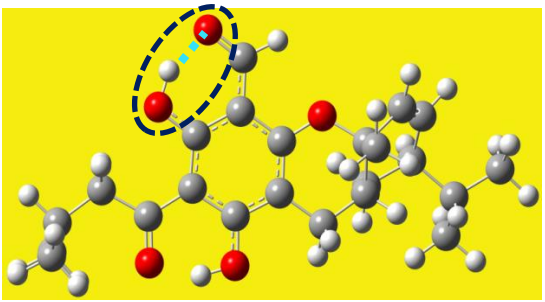
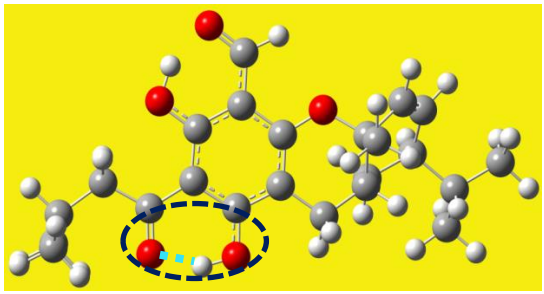
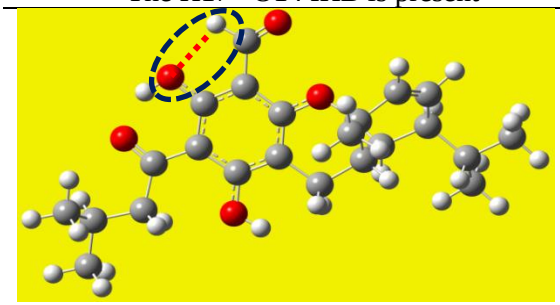
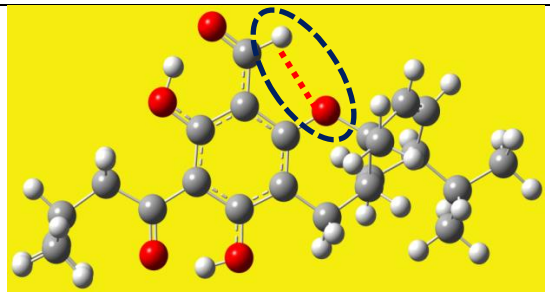
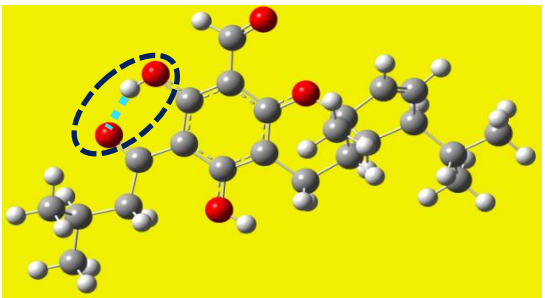
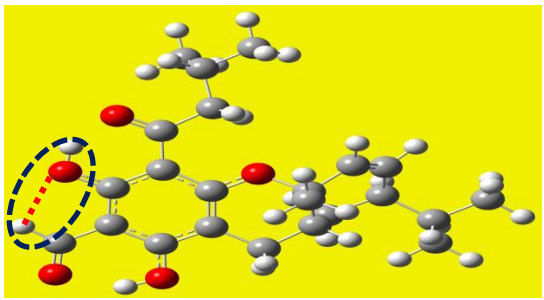


© University of Venda

Conformer specifications are indicated by lowercase letters, each representing specific geometric characteristics. Table 4.2 categorizes these letters, and Figure 4.6 illustrates their meanings. More detailed explanations of each characteristic are provided in the following section.

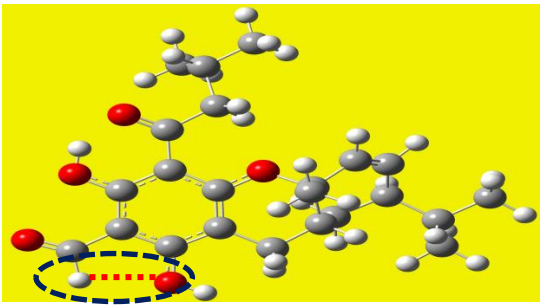
**Figure 4.6. Illustrations of the characteristics that are represented by the letters in the acronym [1].** These characteristics include the presence and types of IHBs, the orientation of the OH groups, the orientation of the $sp^2$ O of an acyl group when not engaged in IHBs and the orientation of the isobutyl chain. For better visualization, the dark-blue segmented circles highlight the positions of the considered atoms, the light blue segment lines represent the O–H···O IHB, and the red segment lines represent the C–H···O IHB.

**the presence and types of IHBs**

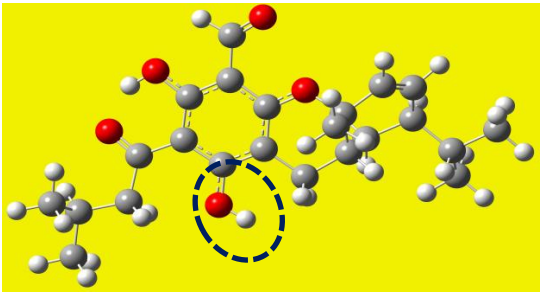
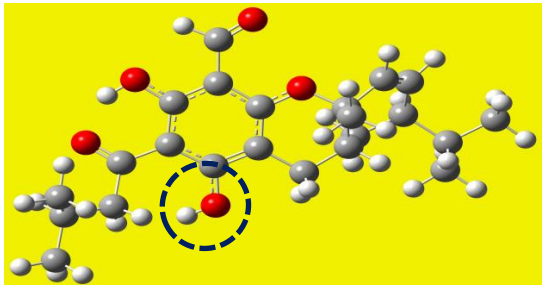
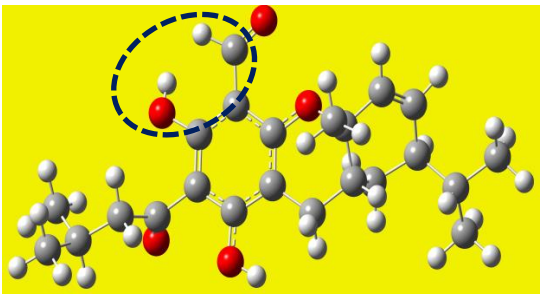
| | |
|---|---|
|  <br>a<br>The H17···O14 IHB is present |  <br>b<br>The H16···O18 IHB is present |
|  <br>n<br>the C7–H13···O12 IHB is present |  <br>m<br>The C7–H13···O8 IHB is present |
|  <br>c<br>the H17···O18 IHB is present |  <br>q<br>The C11–H19···O12 IHB is present |



© University of Venda

| | |
|---|---|
|  <br>p<br>the C11–H19···O10 IHB is present | |

**the orientation of the OH groups**

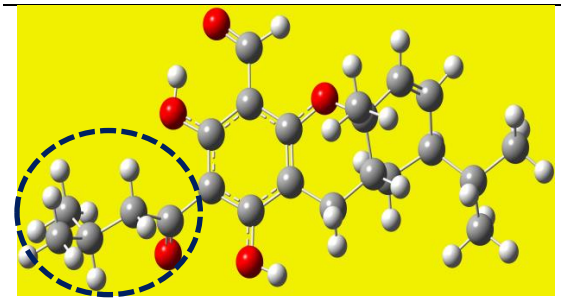
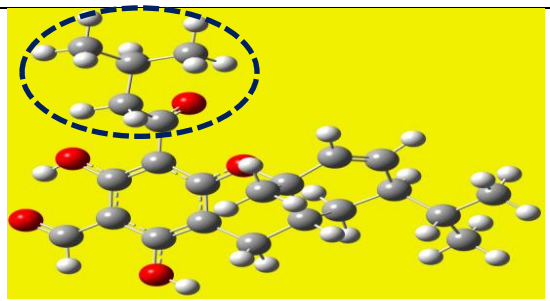
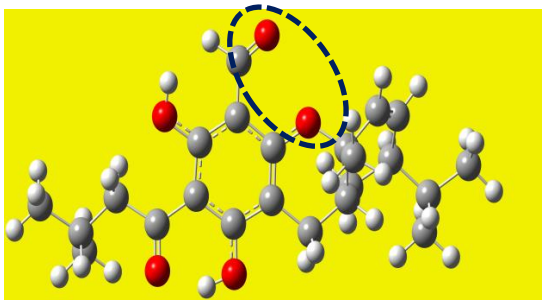
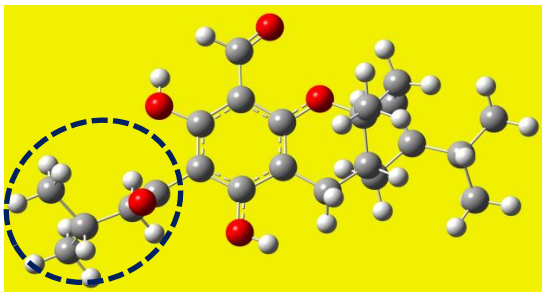
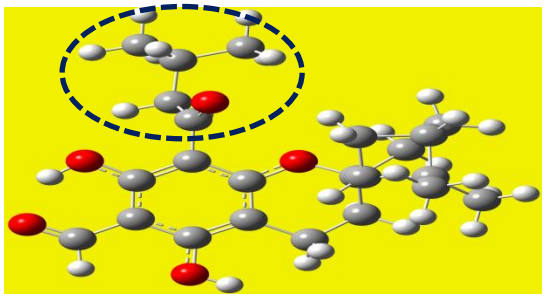
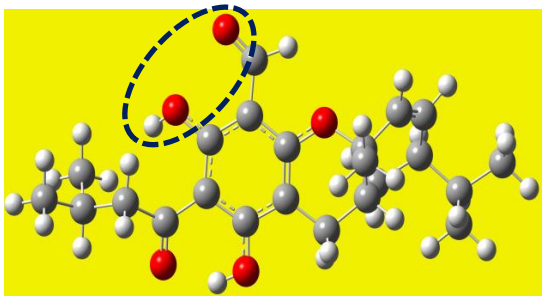
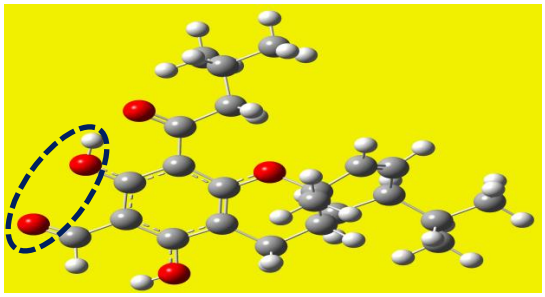
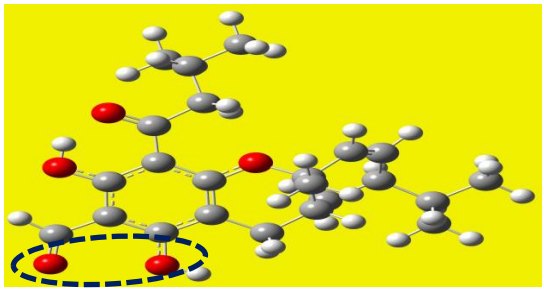
| | |
|---|---|
|  <br>r<br>the C3–C4–O10–H16 torsion angle is close to 0° |  <br>w<br>the C3–C4–O10–H16 torsion angle is close to 180° |
|  <br>u<br>O12–H17 is not engaged in an IHB and is oriented toward the acyl group attached at C1 | |

the orientation of the $sp^2$ O of an acyl group when not engaged in IHBs

| | |
|---|---|
| | |



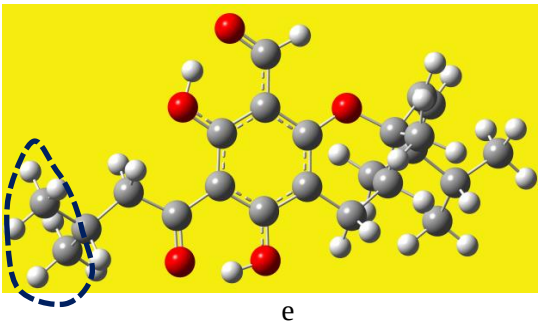
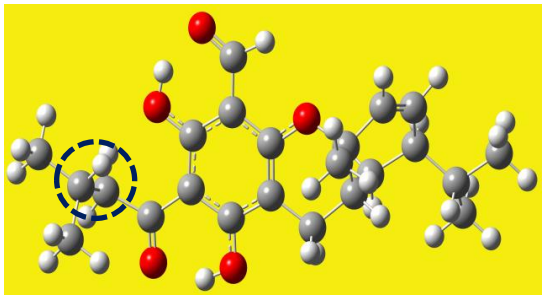
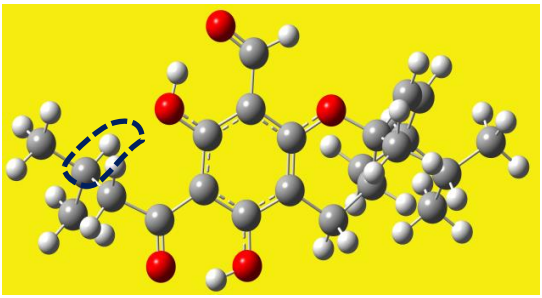
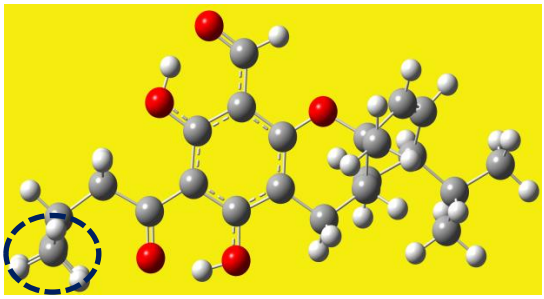
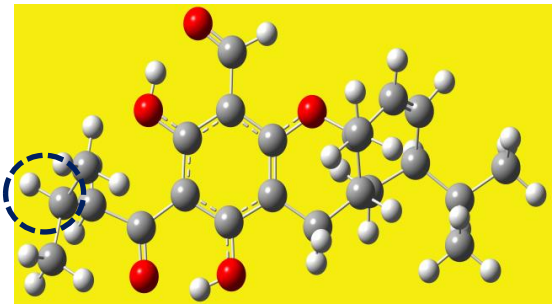
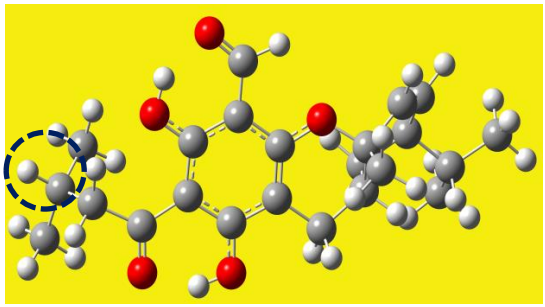
© University of Venda

| | |
|---|---|
|  |  |
| t<br>O18 is not engaged in an IHB and is off-plane, away from us | z<br>O14 is not engaged in an IHB and is off-plane, away from us |
|  |  |
| x<br>O14 is not engaged in an IHB and is oriented towards O8 | s<br>O18 is not engaged in an IHB and is off-plane, towards us |
|  |  |
| v<br>O14 is not engaged in an IHB and is off-plane, towards us | y<br>O14 is not engaged in an IHB and is oriented towards O12 |
|  |  |
| k<br>O18 is not engaged in an IHB and is oriented towards O12 | j<br>O18 is not engaged in an IHB and is oriented towards O10 |

the orientation of the isobutyl chain



© University of Venda

| | |
|---|---|
|  <br>e<br>the two methyls of the isobutyl group are oriented on the other side with respect to the $sp^2$ O of that group |  <br>d<br>the isobutyl group is oriented 'towards us', with the H atom attached to C20 facing towards the O12 or towards the O8 (in the ζ-type) |
|  <br>g<br>the isobutyl group is oriented away from us, with the H atom attached to C20 facing towards O12 or towards the O8 (in the ζ-type) |  <br>f<br>the two methyls of the isobutyl group are oriented towards the $sp^2$ O of that group |
|  <br>i<br>the isobutyl group is oriented 'towards us', with the H atom attached to C20 facing away the O12 or away the O8 (in the ζ-type) |  <br>h<br>the isobutyl group is oriented away from us, with the H atom attached to C20 facing away the O12 or away the O8 (in the ζ-type) |



© University of Venda

Since this study focuses on numerous geometry characteristics unique to EGs, a different lettering system is used compared to previous works on ACPLs. The molecular geometries depicted in figures consistently adhere to the patterns shown in Figure 4.5 for easier comparisons. Hence, terms describing orientations not based on specific atoms (like "towards us," "away from us," "upwards," "downwards") refer to the images as presented according to this convention.

### 4.4.2. Conformational preferences and energetics

#### 4.4.2.1. Conformers' geometrical characteristics

The following characteristics play a significant role in influencing the conformer's energy :

- the presence and types of intramolecular hydrogen bonds (IHBs)
- the orientation of the OH groups
- the orientation of the $sp^2$ O of an acyl group when not involved in an IHB
- the orientation of the isobutyl chain

These characteristics are represented by letters in the acronyms in the same order (also listed in Table 4.2 and depicted individually in Figure 4.6). To simplify, the acyl group with R = H will be referred to as AC-ald, and the acyl group with R = isobutyl as AC-isb throughout the text.

**Table 4.2.** Symbols utilised in the acronyms denoting the molecules and their conformers [1].

| Feature category | Symbol | Meaning of symbol |
|---|---|---|
| Molecule type | ε | Type-1 molecule |
| | ζ | Type-2 molecule |
| | * | Molecule in literature, but used as model to complete a pair |
| IHBs present | a | The H17···O14 IHB is present |
| | b | The H16···O18 IHB is present |
| | c | The H17···O18 IHB is present |
| | m | The C7–H13···O8 IHB is present |
| | n | The C7–H13···O12 IHB is present |
| | p | The C11–H19···O10 IHB is present |
| | q | The C11–H19···O12 IHB is present |
| Orientation of OHs | r | the C3–C4–O10–H16 torsion angle is close to 0° |
| | w | the C3–C4–O10–H16 torsion angle is close to 180° |
| | u | O12–H17 is not engaged in an IHB and is oriented toward the acyl group attached at C1 |
| Orientation of $sp^2$ O not | x | O14 is not engaged in an IHB and is oriented towards O8 |
| | y | O14 is not engaged in an IHB and is oriented towards O12 |
| | v | O14 is not engaged in an IHB and is off-plane, towards us |



© University of Venda

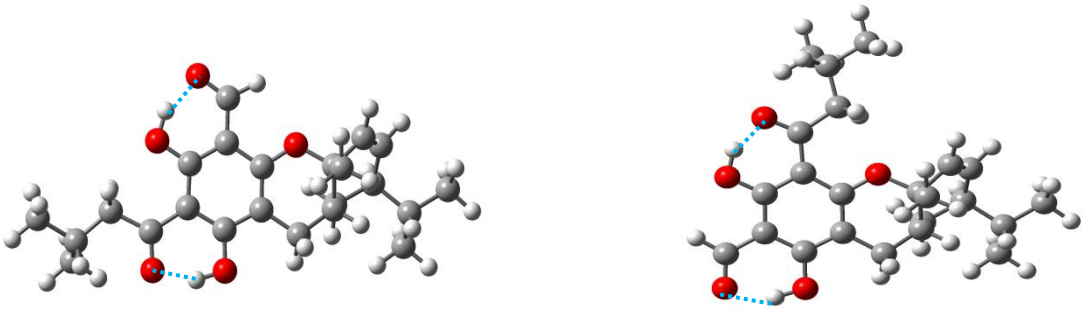
| | | |
|---|---|---|
| engaged in an IHB | z | O14 is not engaged in an IHB and is off-plane, away from us |
| | j | O18 is not engaged in an IHB and is oriented towards O10 |
| | k | O18 is not engaged in an IHB and is oriented towards O12 |
| | s | O18 is not engaged in an IHB and is off-plane, towards us |
| | t | O18 is not engaged in an IHB and is off-plane, away from us |
| Orientation of the isobutyl group | e | the isobutyl group is oriented on the other side with respect to the $sp^2$ O of the acyl group |
| | f | the isobutyl group is oriented towards the $sp^2$ O of the acyl group |
| | d | the isobutyl group is oriented towards us, with the H atom attached to C20 facing towards the $sp^2$ O of the acyl group |
| | i | the isobutyl group is oriented towards us, with the H atom attached to C20 facing away the $sp^2$ O of the acyl group |
| | g | the isobutyl group is oriented away from us, with the H atom attached to C20 facing towards the $sp^2$ O of the acyl group |
| | h | the isobutyl group is oriented away from us, with the H atom attached to C20 facing away the $sp^2$ O of the acyl group |

O–H···O IHBs are the primary stabilizing factors for ACPLs. Three types of O–H···O IHBs can form in EG molecules: H17···O14, H16···O18, and H17···O18, labeled as a, b, and c in the acronyms (Table 4.2, Figure 4.6). The acceptor atom is an $sp^2$ O in all cases. While H17···O14 and H16···O18 can exist simultaneously (Figure 4.7) or individually (Figure 4.6), H17···O18 can only exist alone (Figure 4.6). Additionally, C–H···O IHBs, although weaker, also contribute to stabilization. These are considered due to their known influence on the orientation of substituent chains [44]. Stronger C–H···O IHBs include C7–H13···O8, C7–H13···O12, C11–H19···O10, and C11–H19···O12, denoted by individual letters in the acronyms (Table 4.2, Figure 4.8).

Additional weaker C–H···O IHBs may also be present, including C19–H19a···O12, C19–H19b···O12, C19–H19a···O10, C19–H19b···O10, C13–H13a···O12, C13–H13b···O12, C13–H13a···O8, C13–H13b···O8, C20–H20···O18, C20–H20···O14, C20–H20···O12, C20–H20···O8, C9–H9···O10, C21–H21a···O12, C21–H21a···O10, C22–H22a···O12, C22–H22a···O18, C22–H22a···O10. Figure 4.8 displays representative combinations in selected conformers. Up to six C–H···O IHBs can coexist simultaneously in a given conformer, with their patterns depending on the positions of AC-ald and AC-isb (isomer type), the orientation of the acyl groups, and the geometry of the isobutyl chain. Table S4.1 outlines the different combinations for various conformer types.



© University of Venda



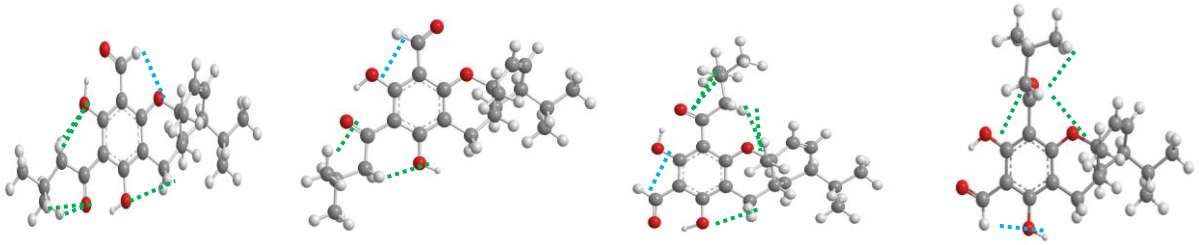
M1–ζ-a-b-m-w-e

ε-type

M1–ζ-a- b-q-w -e

ζ-type

**Figure 4.7. Illustration of instances of both H17··O14 and H16··O18 being simultaneously present in an ε-type conformer (left) and in an ζ-type conformer (right).** IHBs are represented by light-blue dashed segments, and the acronyms representing the conformers are provided under the corresponding images.



M1-ε-a-b-m-w-e M1-ε-c-n-r-x-e

ε-type isomer

M1–ζ-a- b-q-w-e M1–ζ-c-p-r-z-e

ζ-type isomer

**Figure 4.8. Illustrative patterns of C–H···O IHBs) in chosen conformers of euglobal molecules are presented.** The optimized geometries of M1-ε and M1–ζ are utilized for this purpose. Dashed segments denote the presence of IHBs, while the acronyms representing the conformers are provided below each respective image.

The weaker C–H···O intramolecular hydrogen bonds (IHBs) are distinguishable by comparing their lengths. In the M1–ε-a-b-m-w-e conformer, for example, the H···O lengths (in Å) for all C–H···O IHBs are as follows: 2.388 for C7–H13···O8, 2.504 for C19–H19b···O12, 2.602 for C22–H22a···O18, 2.517 for C20–H20···O18, and 2.511 for C9–H9···O10. Similarly, in the M1–ζ-a-b-q-w-e conformer, the H···O lengths are: 2.417 for C11–H19···O12, 2.518 for C13–H13b···O8,



© University of Venda

2.606 for C22–H22a···O14, 2.491 for C20–H20···O14, and 2.514 for C9–H9···O10. In both cases, the first-listed C–H···O IHB is shorter than the others, and an individual symbol denotes it. However, all the others can still be considered C–H···O IHBs because their lengths, although close to the sum of the VDW radii of O and H (2.7 Å), are shorter than this sum, meeting the basic criteria for identifying hydrogen bonds.

The orientations of the OH groups also influence the energy of the conformers [3, 5]. The two possible orientations of O10–H16 are labeled with the same letters (r and w, Table 4.2) used in other ACPLs. The orientation of O12–H17 only needs specification when it does not form the H17···O14 IHB and is oriented upwards (towards the acyl group attached to C1); in such cases, it is denoted as "u" (Table 4.2).

When the $sp^2$ O of the acyl group is not involved in an IHB, the acyl group remains coplanar with the benzene ring when R = H. However, when R ≠ H, it rotates upon optimization, causing the $sp^2$ O to become significantly off-plane, although not perpendicular to the plane [4, 5, 31, 33]. Symbols are used to indicate when O14 or O16 is off-plane (v, z, s, t, Table 4.2) and specify the side for which its dihedral angle with respect to the plane of the benzene ring is <90°. For example, 'v-x' indicates that O14 is off-plane on the 'towards us' side of the ring (v), oriented towards the side of the terpene moiety (x).

The orientation of the isobutyl chain can also influence the energy of a conformer, so specific symbols are used to denote the relevant (most common) orientations it can adopt (Table 4.2). The meanings of these symbols are illustrated in the last section of Figure 4.6, using conformers of the G molecule to focus on the geometry of the isobutyl chain. As mentioned earlier, the G molecule shares the same acyl groups as the ε/ζ EGs, along with a methyl at C3, making it a useful reference for the acylphloroglucinol moiety of EGs. The atom numbering assigned to G corresponds to that of the acylphloroglucinol moiety of the ζ-type isomer in Figure 4.5, ensuring clear correspondence with the numbering of EGs; for example, H20 refers to the hydrogen atom attached to carbon 20. By comparing the geometries of G conformers with the corresponding geometries of the acylphloroglucinol moiety in EG conformers, we can identify how the terpene moiety affects the acylphloroglucinol moiety.



© University of Venda

Figure S4.1 shows the optimized geometries of all calculated conformers of the EGs under consideration, comparing pairs of isomers. On the other hand, Figure S4.2 showcases the geometries of the calculated conformers of G (only those corresponding to EGs' conformers have been calculated due to G's reference role) and τ-type EGs.

In most instances, the optimized geometries closely resemble the input ones. However, there are cases where significant changes occur upon optimization. For most ζ-isomers and the G molecule, inputs containing only the H16···O18 IHB optimize to outputs containing both H17···O14 and H16···O18. This suggests not only the rotation of the AC-isb group (consistent with previous findings [3]), but also the 'upwards' rotation of O12–H17, likely favored by the presence of the second acyl group. As noted previously [3-5], inputs where an $sp^2$ O not engaged in an IHB is coplanar to the benzene ring do not change significantly when R = H. However, they optimize to a geometry where the acyl group rotates when R ≠ H, smoothing the O↔O repulsion with the neighboring O atom. This indicates that u-type conformers are commonly observed for ε-isomers, where O14 remains 'on the right' if initially placed in that position. However, they are rare for ζ-isomers because O14 rotates off-plane from the 'on the right' input position, then further rotates 'to the left' to form the H17···O14 IHB.

Some u-type inputs of M7–ζ and M10*–ζ do not change upon optimization and yield u-type conformers. This difference may be attributed to the presence of the terpene moiety. In ACPLs with the 'canonical' structure, u-conformers are possible only if O14 is engaged in an IHB with the other OH ortho to the acyl group at C1. In the conformers of ζ-isomers where no O–H···O IHB is present, O12–H17 tends to be oriented 'downwards,' and O10–H16 'to the right,' consistent with findings for other ACPLs [3-5].

The orientation of the isobutyl chain can change during optimization for the conformers of the ζ-isomers, particularly when the AC-isb is closer to the bulky terpene moiety. For example, conformers with g-type inputs for M2–ζ, i-type inputs for M3–ζ and M4*–ζ, and h, i, and d-type inputs for M5–ζ, M9–ζ, M10*–ζ, and M6–ζ, all optimize to the f-geometry of the respective molecule. It is notable that in the f-geometry, the two H atoms attached to C13 form C–H···O8 IHBs, whereas in the g, d, h, i geometries, only one of them forms a C–H···O8 IHB.



© University of Venda

#### 4.4.2.2. Conformers' relative energies and factors influencing them

The calculations enabled the consideration of different energy-related values like relative energy, corrected relative energy (which includes both electronic and zero-point energy corrections), zero-point energy correction, relative Gibbs free energy (which includes both electronic and thermal free energy), and its thermal correction. The corrected values and Gibbs free energies were obtained from the frequency calculation results.

Table 4.3 presents the overall ranges of relative energies based on the types of O–H···O hydrogen bonds present. Additional tables (Tables S4.1, S4.2, S4.3, S4.4, and S5.4) provide more detailed comparisons and relative energies values for different conformers and molecules, including grandinol.

**Table 4.3.** Relative energy ranges of the calculated conformers of the ε and ζ isomers and of the grandinol molecule, according to the O–H···O IHBs present [1].

DFT/B3LYP/6-31+G(d,p), HF/6-31G(d,p) and MP2/6-31G(d,p) results *in vacuo* from full optimization calculations, respectively denoted as DFT, HF and MP2 in the columns' headings.

| O–H···O IHBs present | isomer | Relative energy ranges (kcal mol$^{-1}$) | | |
|---|---|---|---|---|
| | | DFT | HF | MP2 |
| H17···O14, H16···O18 | ε | 0.000–2.931 | 0.000–3.680 | 0.000–2.094 |
| | ζ | 0.000–3.925 | 0.000–4.375 | 0.000–3.785 |
| | G | 0.000–2.975 | 0.000–3.711 | 0.000–1.958 |
| | τ* | 0.000 | 0.000 | 0.000 |
| | | | | |
| H17···O18 | ε | 16.025–22.475 | 14.430–21.986 | 14.445–20.489 |
| | ζ | 12.197–17.849 | 9.549–17.353 | 9.606–17.279 |
| | G | 16.665–16.825 | 14.151–14.497 | 14.596–14.969 |
| | τ* | 14.897–18.964 | 13.729–19.397 | |



© University of Venda

| | | | | |
|---|---|---|---|---|
| H17···O14 | ε | 16.014–19.741 | 13.071–19.277 | 13.709–16.934 |
| | ζ | 16.330–20.519 | 14.650–19.446 | 14.992–19.457 |
| | G | 17.615–19.309 | 16.652–18.413 | 16.961–18.285 |
| | τ* | 15.970–17.837 | 14.836–17.492 | |
| | | | | |
| H16···O18 | ε | 16.434 – 21.074 | 16.156 – 20.399 | 13.467– 19.078 |
| | ζ | 14.400–15.458 | 11.408–13.848 | 11.058–13.124 |
| | G | - | - | - |
| | τ* | 15.224–18.097 | 15.406–18.787 | |
| | | | | |
| none | ε | 30.701–34.728 | 27.554–31.126 | 26.762–29.017 |
| | ζ | 28.802–31.789 | 27.164–29.544 | 26.766–29.236 |
| | G | 31.660 | 29.404 | 29.045 |
| | τ* | 30.319–32.461 | 29.601–33.293 | |

Table 4.5 displays the relative energy ranges of the calculated conformers of the considered euglobals (comparing ε and ζ isomers) and the grandinol molecule, with respect to the presence of O–H···O hydrogen bonds. The results from the DFT/B3LYP/6-31 + G(d,p), HF/6-31G(d,p), and MP2/6-31G(d,p) calculations *in vacuo* are listed under the headings DFT, HF, and MP2, respectively. These ranges consider all conformers of the same isomer type where the specified hydrogen bonds are present, providing a comprehensive view of the energy variations across different conformers and molecules.

Earlier studies [8] indicated that for ACPLs with multiple possible hydrogen bonds, conformers with the lowest energy typically have the most simultaneous hydrogen bonds. This pattern also holds true for EGs: the lowest energy conformers typically feature two hydrogen bonds. Table 4.5



© University of Venda

illustrates the energy ranges based on the presence of O–H···O hydrogen bonds. It is evident that only conformers with two hydrogen bonds are common and likely contribute to biological activities, while other conformers are primarily included for comparison purposes to analyze factors affecting energy.

Comparing the lowest energy ranges shows that transitioning from two hydrogen bonds to one results in similar energy increases for ε-isomers, but the increase varies based on the type of remaining hydrogen bond for ζ-isomers (larger when only H17···O14 is present, somewhat smaller with only H16···O18, and even smaller with only H17···O18). Removing both hydrogen bonds leads to nearly double the energy increase compared to removing just one. For the G molecule, the lower ends of the energy ranges show slightly greater increases when removing H16···O18 from conformers with both H17···O14 and H16···O18 and when removing both hydrogen bonds simultaneously.

The positioning of the isobutyl chain's methyl groups can significantly affect the energy of conformers. The e-conformers, where both methyl groups are oriented away from the $sp^2$ O of AC-isb, are generally favored in the HF and DFT results. On the other hand, when the methyl groups face the $sp^2$ O of AC-isb, there is an energy increase ranging from 1.746 to 1.780 kcal $mol^{-1}$ for ε-isomers and 1.457 to 2.631 kcal $mol^{-1}$ for ζ-isomers in HF calculations, and 1.823 to 1.940 kcal $mol^{-1}$ for ε-isomers and 1.693 to 2.019 kcal. $mol^{-1}$ for ζ-isomers in DFT calculations. In MP2 results, the lowest energy conformer tends to be either d- or g-type.

When the isobutyl chain is oriented such that C20 is perpendicular to the plane of the benzene ring, two conformers arise, with the chain on either side of the plane. Although these conformers are mirror images with respect to the plane, their energies may slightly differ. For example, the g and d conformers form such a pair, with the relative energy of the g conformer usually slightly higher than that of the d conformer. Similarly, the h and i conformers are symmetric, with the relative energy of the h conformer usually slightly higher than that of the i conformer.
Figure S4.3 illustrates the trends of relative energies—both uncorrected and corrected for ZPE—for each calculated molecule across the utilized computational methods. In all the diagrams, there's a gradual decrease in energy corresponding to conformers with two O–H···O IHBs, followed by a



© University of Venda

sharp increase to a higher energy level for conformers with one O–H···O IHB. Furthermore, there is another sharp increase to the highest-energy region for conformers with no IHBs.

The ZPE-corrected and uncorrected relative energies are quite similar for the lowest energy conformers (those with two O–H···O IHBs). However, there may be a slightly greater difference between ZPE-corrected and uncorrected energies for conformers with one O–H···O IHB, observed across both DFT and HF results.

Figure S4.4 compares how the ZPE-corrected and uncorrected relative energies differ for the lowest energy conformers (a-b-conformers) of both ε and ζ isomers within the same pair. Additionally, Figure S4.5 illustrates that the Gibbs relative free energies are quite similar for these lowest energy conformers, although HF values are sometimes slightly smaller than DFT ones for conformers with only one O–H···O IHB.

Table S4.6 provides details about the ZPE correction to the electronic energy and the thermal correction to the Gibbs free energy. These corrections are very similar for isomers within the same pair and also for pairs M1–M9, which have the same number of atoms but different terpene moieties. However, they are notably larger for the M10*–ζ and M10–ε pair, which have bulkier terpene moieties. It is notable that HF corrections tend to be larger than DFT corrections, likely due to HF's underestimation of the IHB strength and DFT's overestimation, inluencing the computed vibrational frequencies of the donor OHs (red shifts). Table S4.7 presents the same corrections for the G molecule, providing comprehensive information about it.

### 4.4.3. Characteristics of the IHBs

#### 4.4.3.1. Characteristics of the O–H···O IHBs

##### 4.4.3.1.1. Parameters of the O–H···O IHBs

H-bonds are characterized by their parameters: bond length (the H···O distance), O···O distance, and bond angle (OĤO). Tables S4.8, S4.9, and S4.10 present these parameters for the calculated conformers of the ε-isomers, ζ-isomers, and grandinol, respectively. The data suggest that these H-bonds fall into the moderate to strong range, similar to other ACPLs [3, 4], due to factors like



© University of Venda

the acceptor being an $sp^2$ O atom, closing a 6-member ring, and exhibiting characteristics of resonance-assisted H-bonds [33, 34].

The bond length is often used to compare the strengths of H-bonds, with shorter lengths indicating stronger bonds. DFT, HF, and MP2 results show similar trends, with MP2 values being intermediate between the other two. MP2 results are typically more accurate, as DFT/B3LYP tends to overestimate H-bond strengths and HF tends to underestimate them. Previous studies [3, 4], have shown that experimental results on H-bond lengths align with the range defined by DFT and HF results. Figure S4.6 provides visual comparisons of the lengths of H17···O14 and H16···O18 in the lower energy conformers, where both H-bonds are present simultaneously, with separate diagrams for each H-bond in each isomer.

Considering the aim of this study, it is important to compare the lengths of the three IHBs when the acceptor atom belongs to AC-ald or AC-isb, and when each acyl group is attached to C1 or C5. Figure S4.6 illustrates that H16···O18 is shorter than H17···O14 in ε-isomer conformers but longer in ζ-isomer conformers. Table S4.11 organizes the length values based on the acyl group of the acceptor atom (AC-ald or AC-isb), and Figure 4.9 visually compares H17···O14 and H16···O18 lengths when they are present together in lower energy conformers for representative molecule (with Figure S4.7 shows for the conformers of all considered molecules).

**Figure 4.9. Representative diagram showing the comparison of the length (Å) of the H17···O14 and H16···O18 intramolecular hydrogen bonds in the lower energy conformers of the calculated euglobal IIC (M1-ε), in the results of the selected calculation methods.**

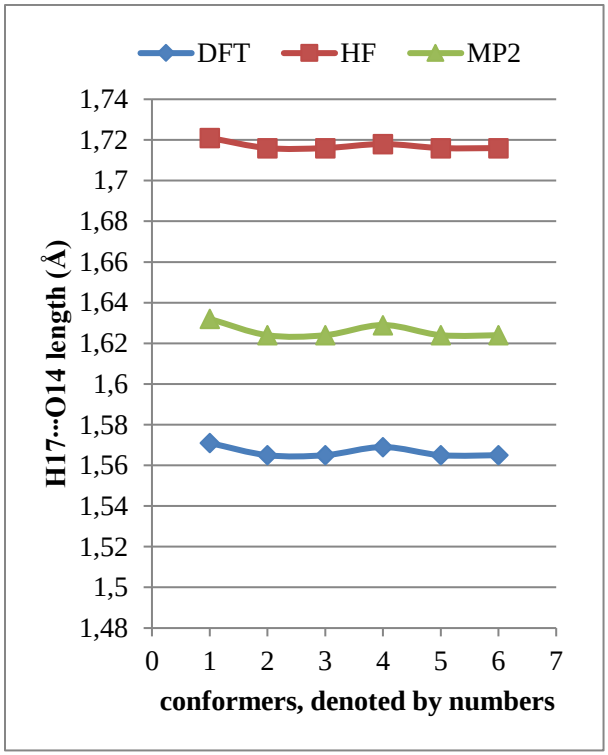
Results from HF/6-31G(d,p), DFT/B3LYP/6-31+G(d,p) and MP2/6-31G(d,p) full optimization calculations *in vacuo*, respectively denoted as HF, DFT and MP2 in the figures' legends. The H17···O14 and H16···O18 IHBs are selected because they are simultaneously present in all the low energy conformers. In each diagram, the conformers are denoted by numbers on the x axis; the correspondence between numbers and conformers is shown in tables preceding the diagrams.

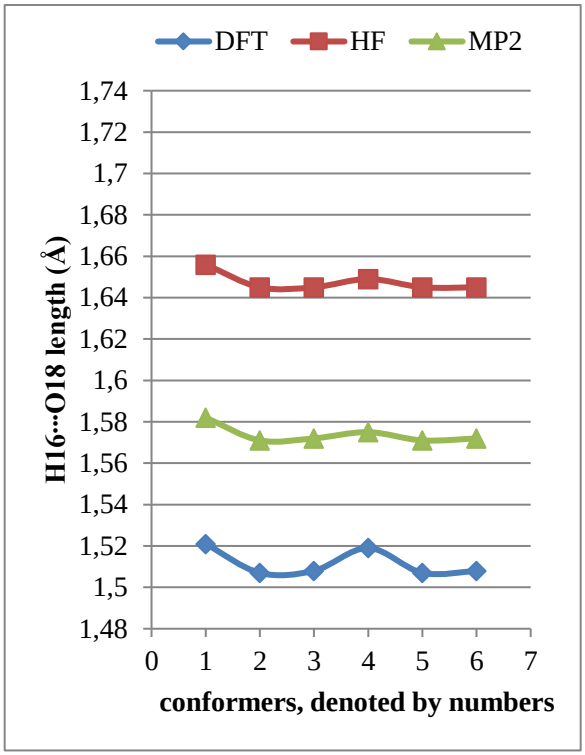
Numbers denoting the conformers of euglobal IIC on the x axis.



© University of Venda

| # | Conformer | # | Conformer | # | Conformer |
|---|---|---|---|---|---|
| 1 | M1–ε-a-b-m-w-e | 3 | M1–ε-a-b-m-w-g | 5 | M1–ε-a-b-m-w-i |
| 2 | M1–ε-a-b-m-w-d | 4 | M1–ε-a-b-m-w-f | 6 | M1–ε-a-b-m-w-h |




Consistent with findings for other ACPLs, the IHB with the acceptor O from AC-ald (an acyl group with R = H) tends to be longer than the IHB with the acceptor O from AC-isb (an acyl group with R ≠ H). Regarding position, the IHB involving AC-ald is shorter when attached to C1 compared to C5. The IHB involving AC-isb is often slightly shorter when attached to C1 than when attached to C5. This suggests that the bulky terpene moiety near C1 has a greater effect on the IHB involving AC-ald than on the IHB involving AC-isb.

Comparisons with the conformers of grandinol (G) further support this conclusion. When considering the IHB involving AC-ald, its length in the ε isomers (H17···O14 IHB) is notably shorter than in G (H16···O18 IHB). In contrast, in the ζ isomers (H16···O18 IHB), it is slightly longer than in G. For the IHB involving AC-isb, its length in the ζ isomers (H17···O14 IHB) is similar to or shorter than in G. In contrast, in the ε isomers (H16···O18 IHB), it is comparable to



© University of Venda

or slightly longer than in G. This indicates that the proximity of the bulky terpene moiety to C1 shortens the IHB involving the acyl group at C1, with a greater effect observed for AC-ald than for AC-isb. This effect difference could be because the bulkier R in AC-isb already shortens the IHB involving it [4].

To further validate the influence of the terpene moiety's proximity on the IHB involving AC-ald, ten τ-type EGs (where both acyl groups are AC-ald) were also computed, sharing the same terpene moieties as the ε and ζ isomers. They are identified with the same molecular labels (M1, M2, etc.) as the ε and ζ isomers, followed by an asterisk to denote model structures, and by τ. Their relative energies are reported in Table S4.12, and the IHB parameters in Table S4.13. Since the two acyl groups are identical, any differences in their IHB characteristics are due to the surrounding molecular context. In conformers with both H17···O14 and H16···O18 present, the former is consistently shorter than the latter. In conformers with only one present, H17···O14 is almost always slightly shorter than H16···O18. This confirms the terpene moiety's influence on the IHB involving the acyl group at C1.

To further validate the influence of the terpene moiety's proximity on the H17···O18 IHB is consistently longer than H16···O18 in both the ε and ζ isomers as well as in grandinol. However, in the conformers of the τ models, it may vary, either longer or shorter. This indicates that its length being longer in the ε and ζ isomers and in grandinol might be associated with an acyl group with a bulkier R (AC-isb).

#### 4.4.3.1.2. Energy increase accompanying the removal of O–H···O IHBs

When an IHB is removed, there is a notable increase in energy. This increase is not solely due to the energy of the IHB itself but also because of changes in geometry and the emergence of repulsion between the oxygen atoms previously connected by the hydrogen atom. Comparing these energy increases for each IHB in EGs can give insights into their relative strengths or the stabilizing effects they bring.

Table 4.5 and Figure S4.3 highlight the significant energy differences linked to the presence of two, one, or no IHBs. Table S4.14 provides detailed comparisons. Since removing an IHB involves geometry changes, there is no pair of conformers differing only by the presence or absence of one



© University of Venda

IHB. Thus, we consider the average energy of conformers with a specific IHB and compare it to those without that IHB and no other changes in the IHB patterns. This is crucial due to the dominant role of O–H···O IHBs in energy contribution.

For instance, the average energy of specific conformers in M1–ε serves as a reference for understanding the energy changes associated with the removal of certain IHBs. The conformers a-b-m-w-e, a-b-m-w-d, a-b-m-w-g, a-b-m-w-f, a-b-m-w-i, and a-b-m-w-h are used as the reference energy for removing both the H17···O14 and H16···O18 IHBs.

Next, the energy of conformers resulting from the removal of H17···O14 is analyzed. These are conformers b-w-x-e, b-m-r-y-e, and b-n-r-x-e, as they maintain the H16···O18 IHB without introducing any other O–H···O IHBs.

Similarly, conformer M1–ε-a-m-r-t-e, resulting from the removal of H16···O18, is compared with the average energy of conformers containing both IHBs. Comparing the conformers a-m-r-t-e, and r-t-x-u-e allows further evaluation of the energy increase upon removing H17···O14.

Moreover, the average energy of conformers b-w-x-e, b-m-r-y-e, and b-n-r-x-e is compared with the energy of conformer r-t-x-u-e to assess the energy increase associated with the removal of H16···O18. However, these can not be compared with the average energy of conformers c-n-r-x-e, c-n-w-x-e, and c-m-w-y-e because they contain another O–H···O IHB (H17···O18), even though H16···O18 is absent.

#### 4.4.3.1.3. Vibrational frequency decreases associated with O–H···O IHBs

When hydrogen bonds form, they typically cause an elongation of the bond length of the donor, leading to a decrease in its infrared (IR) vibrational frequency [35-38], known as a redshift. Redshifts are more common for highly polar donors such as O–H. Comparing these redshifts can offer further insights into the relative strengths of hydrogen bonds, as larger redshifts generally indicate stronger hydrogen bonds.



© University of Venda

The calculated vibrational frequencies, specifically the stretching mode of the two O–H groups, in the ε- and ζ-isomers' conformers are listed in Table S4.15, while those for grandinol are in Table S4.16. Figure S4.8 compares the DFT and HF results for the frequencies of these OH groups in the respective isomers. The diagrams depict an initial plateau of lower frequencies corresponding to conformers with both O10–H16 and O12–H17 engaged in hydrogen bonds. Subsequently, fluctuations between higher and lower frequency values occur, depending on whether the OH group is involved in a hydrogen bond.

In the diagrams, an initial lower-frequency plateau corresponds to the conformers with lower energy, where both O10–H16 and O12–H17 are engaged in hydrogen bonds. Subsequently, fluctuations occur between higher frequency values when a given OH group is not involved in a hydrogen bond, and lower values when it is engaged in one. Notably, the difference between the plateau and the higher frequency values is more pronounced in the DFT results compared to the HF results. The detailed reasons for this discrepancy are discussed further in the following paragraphs.

The redshift in the vibrational frequency of a hydrogen bond (H-bond) donor is determined by comparing its frequency when it is not engaged in an H-bond (i.e., free) with its frequency when it forms an H-bond. This difference represents the redshift, calculated as the difference between the frequencies in these two scenarios. To evaluate the redshifts of either O10–H16 or O12–H17, each case where one acts as an H-bond donor is considered, using the average of its frequencies in the conformers where it is free as a reference. Table S4.17 provides the redshifts observed in the calculated conformers of the ε and ζ isomers and in the G molecule. At the same time, Table S4.18 compares these redshifts using the same criteria as Table S4.11. Figure 4.10 further illustrates this comparison for the representative lowest energy conformers of euglobal IIC (M1-ε) and euglobal T1 (M1-ζ)' pairs (Figure S4.9 shows the comparison of each isomer pair for other molecules).

**Figure 4.10. Representative comparison of the redshifts in the vibrational frequency of O12–H17, when it forms the H17···O14 intramolecular hydrogen bond (IHB), and of O10–H16, when it forms the H16···O18 IHB, in the lowest energy conformers of euglobal IIC (M1-ε) and euglobal T1 (M1-ζ)' pairs.**



© University of Venda

Results from HF/6-31G(d,p), DFT/B3LYP/6-31+G(d,p) calculations *in vacuo*, respectively denoted as HF and DFT in the diagrams' legends.

The H17···O14 and H16···O18 IHBs are selected because they are simultaneously present in all the low energy conformers. H17···O14 involves the acyl group with R = H (AC-ald) in the ε-isomers and the acyl group with R = isobutyl (AC-isb) in the ζ-isomers. Conversely, H16···O18 involves AC-isb in the ε-isomers and AC-ald in the ζ-isomers.

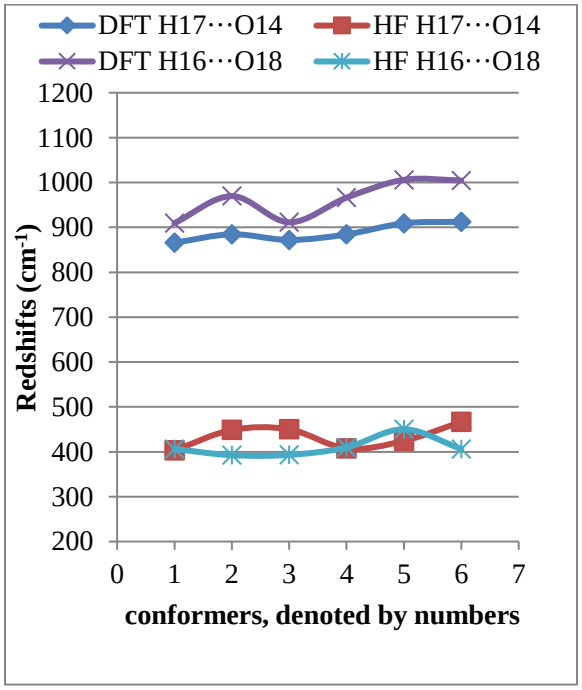
In each diagram, the conformers are denoted by numbers on the x axis; the correspondence between numbers and conformers is shown in tables preceding the diagrams.

Numbers denoting the conformers of M1-ε on the x axis.

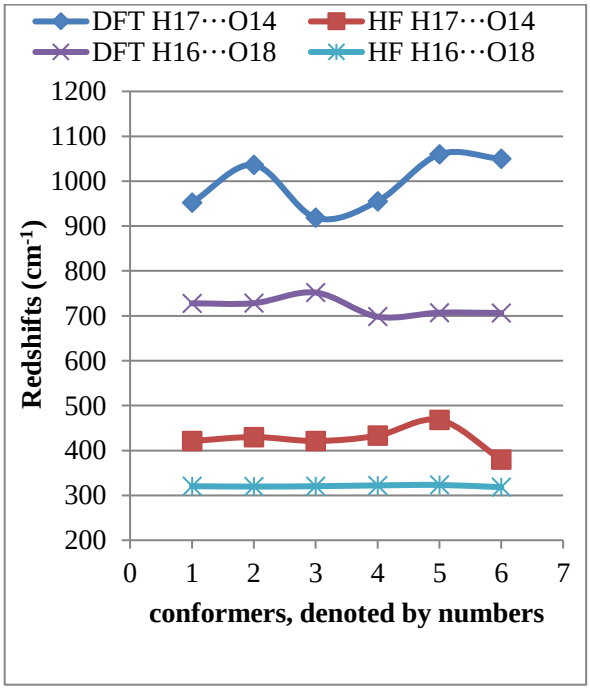
| # | Conformer | # | Conformer | # | Conformer |
|---|---|---|---|---|---|
| 1 | M1–ε-a-b-m-w-e | 3 | M1–ε-a-b-m-w-g | 5 | M1–ε-a-b-m-w-i |
| 2 | M1–ε-a-b-m-w-d | 4 | M1–ε-a-b-m-w-f | 6 | M1–ε-a-b-m-w-h |

Numbers denoting the conformers of M1-ζ on the x axis.

| # | Conformer | # | Conformer | # | Conformer |
|---|---|---|---|---|---|
| 1 | M1–ζ-a-b-q-w-e | 3 | M1–ζ-a-b-q-w-d | 5 | M1–ζ-a-b-q-w-h |
| 2 | M1–ζ-a-b-q-w-g | 4 | M1–ζ-a-b-q-w-f | 6 | M1–ζ-a-b-q-w-i |




M1-ε




M1-ζ



© University of Venda

The comparison facilitated by Table S4.17 and Table S4.18 categorizes the redshifts according to the types of H-bonds formed. By following the approximate rule that greater redshifts correspond to stronger H-bonds, one can roughly compare the strengths of the H-bonds present in the calculated molecules. The redshift is found to be most significant for the H-bonds involving AC-isb, specifically O10–H16 in the ε isomers and O12–H17 in the ζ isomers, when the other H-bond is also present (a-b conformers). Redshifts for H-bonds involving AC-ald are slightly smaller, consistent with the observation that the first H-bond in ACPLs with R = H is weaker than in those with R ≠ H. Moreover, these redshifts are greater when the other H-bond is also present (a-b conformers). However, the redshifts are considerably smaller when either H17···O14 or H16···O18 is present alone, suggesting a potential coupling between the two H-bonds in a-b conformers despite their positions not indicating cooperation. The indications regarding the relative strengths of the H-bonds provided by the redshifts align with those obtained from the analysis of the H-bonds' lengths.

Tables S17 and S18, along with Figure S4.9, highlight a notable difference between the redshift estimations obtained from HF calculations compared to those from DFT calculations. While the trends observed in both methods are comparable, the redshifts estimated by HF are considerably smaller than those by DFT. This discrepancy can be attributed to the previously mentioned phenomenon wherein DFT/B3LYP tends to overestimate the strength of H-bonds, leading to larger redshift estimations, whereas HF underestimates it, resulting in smaller redshift estimations.

In general, the actual redshifts are likely to fall somewhere between these two extremes, representing an intermediate value. However, despite potentially overestimating the actual redshifts, the DFT calculations offer the advantage of highlighting the differences, thereby facilitating comparisons between different scenarios.

Table S4.19 presents the vibrational frequencies of the OH groups in the τ-type model structures, while Table S4.20 reports the redshifts observed when these OH groups form hydrogen bonds (IHBs). Comparing these frequencies with those of the OH groups in the lowest energy conformers



© University of Venda

of the ε- and ζ-isomers, when they form IHBs with AC-ald, shows a notable difference depending on the position of AC-ald.

In the ζ-isomers, when O10–H16 forms an IHB with AC-ald, its frequencies are similar to those observed in the τ-type model structures (2952.03–3019.05 $cm^{-1}$ and 2975.03–2984.91 $cm^{-1}$ respectively in the DFT results). However, in the ε-isomers, when O12–H17 forms the H17···O14 IHB, its vibrational frequencies are lower compared to the corresponding τ-type conformers (2929.37–2941.85 $cm^{-1}$/DFT and 2811.53–2856.16 $cm^{-1}$/DFT respectively). Additionally, when both IHBs are present in the τ-type structures, the redshifts of O12–H17 are greater than those of O10–H16. These findings suggest that the proximity of the bulky terpene moiety to C1 strengthens the IHB-engaging AC-ald at C1.

Furthermore, the redshifts of either O12–H17 or O10–H16 are slightly smaller in the τ-type conformers, with only one IHB confirming that each IHB strengthens the other when both are present.

#### 4.4.3.2. Characteristics of the C–H···O IHBs

##### 4.4.3.2.1. Parameters of the C–H···O IHBs

As reported in previous studies, C–H···O hydrogen bonds have been acknowledged as true hydrogen bonds for over three decades [39, 40]. These bonds play a significant role in determining the conformational preferences of ACPLs, including influencing the orientation of chains containing the donor groups in the substituents. In this study, particular emphasis is placed on understanding their effects.

As described in Section 4.4.1.2.1, the presence of one or more of the stronger C–H···O hydrogen bonds (such as C7–H13···O8, C7–H13···O12, C11–H19···O10, and C11–H19···O12) is indicated in the acronyms used to label the conformers (refer to Table 4.2). Representative combinations of these bonds in selected conformers are illustrated in Figure 4.8, while Table 4.3 provides a list of their combinations in different types of conformers.

The parameters characterizing the C–H···O hydrogen bonds in the calculated conformers of the ε-isomers, ζ-isomers, and grandinol are listed in Tables S4.21, S4.22, and S4.23, respectively. These



© University of Venda

bonds consistently exhibit bond lengths well below the sum of the van der Waals radii of the hydrogen and oxygen atoms, confirming their status as true hydrogen bonds.

Interestingly, the length of these bonds is influenced by the presence of other intramolecular hydrogen bonds. For instance, in the ε-isomers, the distance between H13 and O8 (which represents the bond length of the C7–H13···O8 hydrogen bond) tends to be longer when H17···O14 is also present and shorter when it is absent. Similarly, in the ζ-isomers, the length of the C11–H19···O12 hydrogen bond tends to be longer when the H17···O14 hydrogen bond is present.

Furthermore, the length of the C–H···O hydrogen bonds formed by the hydrogen atom of the AC-ald group is often shorter in the ε-isomers compared to the ζ-isomers. The lengths of the C–H···O hydrogen bonds formed by the carbon atoms of the isobutyl chain do not exhibit significant differences when the AC-isb group is attached to either C1 or C5. For instance, in M1-ζ (where AC-isb is attached to C1) and M1-ε (where it is attached to C5), the lengths (in angstroms) of these hydrogen bonds show similar values: 2.340 for the C13–H13a···O8 bond and 2.518 for the C13–H13b···O8 bond in M1-ζ; 2.367 for the C19–H19a···O12 bond and 2.504 for the C19–H19b···O12 bond in M1-ε.

Table S4.24 provides data on the lengths of the C–H···O hydrogen bonds in the conformers of the calculated τ-type model structures. Similarly to the ε and ζ isomers, the presence of O–H···O hydrogen bonds affects these lengths. For example, when H17···O14 is present, the bond length of H13···O8 becomes longer, similar to what happens in the ε-isomers. Similarly, the bond length of H13···O12 increases when H17···O18 is present. Moreover, the length of H19···O12 becomes longer when H16···O18 is present, and when H17···O18 IHB exists, the bond length of H19···O10 IHB increases.

Comparing the lowest energy conformers of τ models with the ε-isomers shows that H13···O8 is longer in the former, suggesting that the presence of AC-isb in the ε-isomers affects the length of H13···O8. Similarly, H19···O12 is longer in the τ models compared to the ζ-isomers, indicating that AC-isb also influences the length of this bond.



© University of Venda

#### 4.4.3.2.2. Vibrational frequency increases associated with C–H···O IHBs

The vibrational frequencies of the C–H bonds in AC-ald, acting as donors in C–H···O hydrogen bonds (IHBs), often exhibit a phenomenon known as a 'blue shift.' This occurs when the bond length of the donor C–H decreases, leading to an increase in its stretching mode frequency [41-43]. Blue shifts are commonly observed in C–H bonds where the carbon has $sp^2$ hybridization [44]. Specifically, an increase in the C–H stretching mode frequency is frequently noted for aldehyde groups [44-46] and formic acid, where the bonded carbon is also part of a C=O group [47].

A blue shift in the C–H stretching frequency indicates the presence of C–H···O hydrogen bonds. This shift would not occur without forming such bonds [48]. Consequently, detecting a blue shift in the C–H stretching frequency can be a diagnostic tool for identifying C–H···O hydrogen bonds, even within complex molecules like proteins [49].

Table S4.25 provides the vibrational frequencies of the stretching mode of the C–H bonds of AC-ald in both the ε and ζ isomers, while Table S4.26 presents the corresponding blue shifts, evident in the DFT results. The magnitude of the blue shift depends not only on the type of C–H···O hydrogen bond formed but also on the presence of O–H···O hydrogen bonds in the conformer.

In the ε isomers, the C7–H13 bond can form a C–H···O hydrogen bond with either O8 or O12 as the acceptor. The blue shift is more pronounced when O8 is the acceptor, with the highest values observed in conformers where both H17···O14 and H16···O18 are present. Slightly smaller blue shifts are seen in conformers where only H17···O14 is present, and considerably smaller shifts occur in conformers where H17···O14 is absent, but either H16···O18 or H17···O18 is present. When C7–H13 forms the C7–H13···O12 hydrogen bond, the blue shifts are smaller compared to when it forms C7–H13···O8. Additionally, the shifts are slightly greater in conformers where H16···O18 is present compared to those where H17···O18 is present.

In ζ isomers, the C11–H19 bond can form a C–H···O hydrogen bond with either O10 or O12. Its blue shift is more pronounced when O12 is the acceptor. The highest shifts are observed in conformers where both H17···O14 and H16···O18 are present, followed by slightly smaller shifts in conformers where only H16···O18 is present and considerably smaller shifts in conformers



© University of Venda

where only H17···O14 is present. When C11–H19 forms C11–H19···O10, the blue shift is slightly smaller compared to when it forms C11–H19···O12, particularly when H17···O18 is present, and significantly smaller when H17···O14 is present.

Table S4.27 provides the vibrational frequencies of C11–H19 in the calculated conformers of the grandinol molecule, along with the corresponding blue shifts when it forms C–H···O hydrogen bonds. The patterns of blue shifts are similar to those observed for C11–H19 in the calculated ζ isomers. Specifically, the blue shift for C11–H19···O12 is slightly greater than that for C11–H19···O10. Moreover, the shift is greater when both H17···O14 and H16···O18 are present but considerably smaller when only one of them is present. Similarly, the blue shift for C11–H19···O10 is greater when H17···O18 is present but significantly smaller when no O–H···O hydrogen bond is present.

The discrepancy between the DFT and HF results for the blue shift can be attributed to the tendency of DFT to overestimate the strength of hydrogen bonds, resulting in an exaggeration of their effects. This likely leads to overemphasizing the blue-shift effect in the DFT results. While this overestimation may make it easier to identify patterns due to the slightly higher values, the reliability of pattern identification remains unaffected. This is because comparisons are made based on values that have been similarly overemphasized, ensuring consistency in the analysis.

Table S4.28 presents the vibrational frequencies of the C–H bonds in the acyl groups within the conformers of the calculated τ-type model structures. At the same time, Table S4.29 reports the corresponding blue shifts. Similar to the ε and ζ isomers, the blue shift observed here is influenced by the presence of O–H···O hydrogen bonds. Specifically, the blue shift of the C7–H13 bond is amplified in the presence of H17···O14, while the blue shift of C11–H19 is heightened when either H16···O18 or H17···O18 IHBs are present.

### 4.4.4. Other molecular properties

#### 4.4.4.1. Energies of the frontier molecular orbitals and selected derived quantities

The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) can be used to determine the properties of a molecule. The HOMO dictates the molecule's ability to donate electrons, while the LUMO governs its ability to accept electrons. Analyzing their



© University of Venda

distribution enables predictions about the molecule's reactivity and the reactivity of specific regions, facilitating the identification of its more active sites. The molecule's reactivity, spectroscopic properties, and electric conductivity, among other properties, are actively influenced by the energy gap between the HOMO and LUMO. A smaller energy gap promotes intramolecular charge transfer, resulting in higher reactivity that somewhat influences the molecule's biological activities [50-52].

The individual energies of the HOMO and LUMO, along with their energy gap, can also be used as descriptors in studying a molecule's activity [53]. These descriptors are commonly employed in QSAR [54] and other investigations into molecules' biological activities [55, 56], such as anticancer effects [57-59]. The applications of these descriptors are diverse; for example, the HOMO-LUMO energy gap serves as a criterion for predicting toxicity in aquatic environments. Therefore, it is relevant in designing new environmentally benign substances. It is recommended that the energy gap be greater than 138 kcal $mol^{-1}$ to avoid reactive toxicity, which is often associated with high reactivity.

The shapes of the frontier orbitals for all the calculated conformers of all considered EG molecules are shown in Figure S4.10. Both orbitals are mainly distributed on the benzene ring and on all or nearly all the O atoms (including O8). The $sp^2$ C atom of the acyl groups may also be involved, with a higher frequency observed for the LUMO; in some cases, they also extend to the C atom of the terpene moiety bonded to C3. In the conformers of M5-ε and M5-ζ, the HOMO is distributed on the terpene moiety. In the conformers of M7-ε, the HOMO is distributed nearly completely or to a large extent on the terpene moiety.

Considering the significance of the values associated with the frontier orbitals in studying biological activities, all the computed values are documented in the ESI. The energies of the HOMO and LUMO orbitals for the conformers of the calculated ε- and ζ-isomers are presented in Table S4.30, while Table S4.31 presents the HOMO-LUMO energy gap values. Figure S4.11 provides a comparison of the gap values across the results obtained from three selected computational methods, and Figure 4.11 and figure S4.12 illustrates their trends for the low energy conformers in ε- and ζ-isomers' pairs. The observed values demonstrate the well-known



© University of Venda

phenomenon where DFT tends to underestimate the energy of the frontier orbitals gap (it assigns negative values to the LUMO energies, while ab initio methods assign positive values).

Numbers denoting the conformers of euglobal IIC on the x axis.

**Figure 4.11. Representative comparison of the HOMO-LUMO energy gap of the lowest energy conformers of euglobal IIC (M1-ε) and euglobal T1 (M1-ζ)' pairs.**

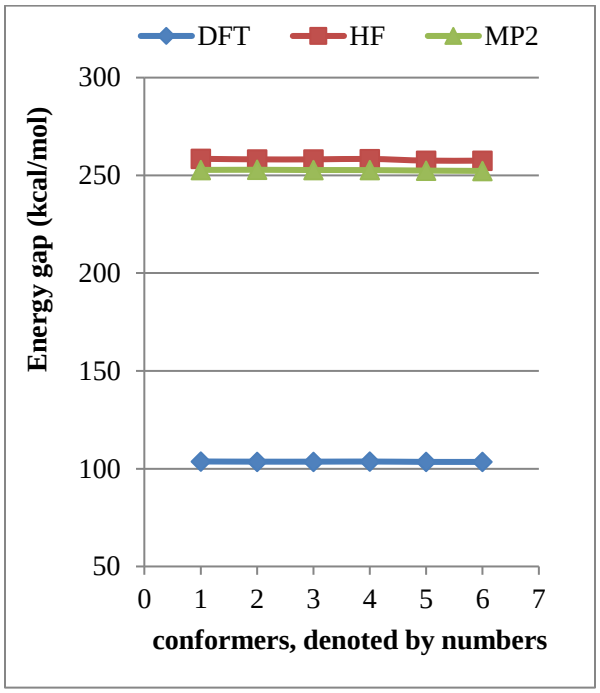
Results from HF/6-31G(d,p), DFT/B3LYP/6-31+G(d,p) and MP2/6-31G(d,p) full optimization calculations *in vacuo*, respectively denoted as HF, DFT and MP2 in the figures' legends.

For each pair, the conformers' pairs are denoted by numbers on the x axis; the correspondence between numbers and pairs is shown in tables before each diagram.

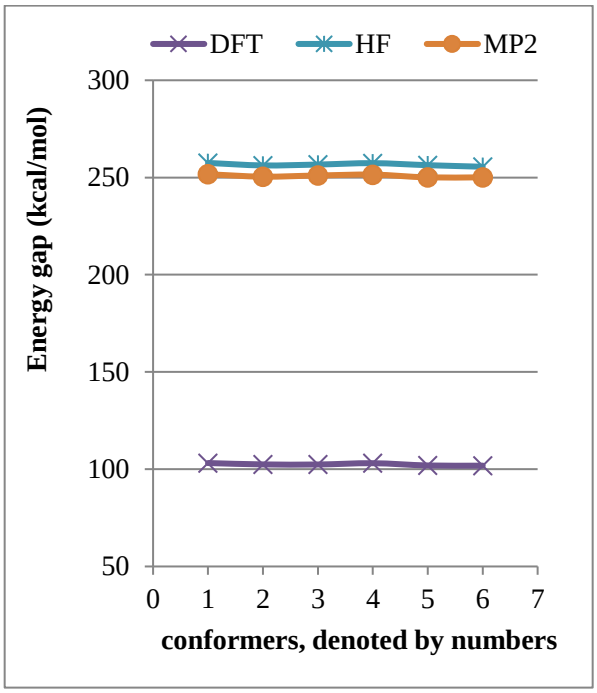
| # | Conformer | # | Conformer | # | Conformer |
|---|---|---|---|---|---|
| 1 | M1–ε-a-b-m-w-e | 3 | M1–ε-a-b-m-w-g | 5 | M1–ε-a-b-m-w-i |
| 2 | M1–ε-a-b-m-w-d | 4 | M1–ε-a-b-m-w-f | 6 | M1–ε-a-b-m-w-h |

Numbers denoting the conformers of euglobal T1 on the x axis.

| # | Conformer | # | Conformer | # | Conformer |
|---|---|---|---|---|---|
| 1 | M1–ζ-a-b-q-w-e | 3 | M1–ζ-a-b-q-w-d | 5 | M1–ζ-a-b-q-w-h |
| 2 | M1–ζ-a-b-q-w-g | 4 | M1–ζ-a-b-q-w-f | 6 | M1–ζ-a-b-q-w-i |




M1-ε




M1-ζ



© University of Venda

Ab initio results are generally regarded as more reliable and realistic since they stem from wavefunction-based approaches, and the HOMO and LUMO—being molecular orbitals—are wavefunctions themselves. Among the obtained results, the MP2 results are considered the most accurate as MP2 incorporates both correlation and dispersion effects. The difference between the MP2 and DFT estimations is typically around 147–151 kcal mol−1. Comparatively, the MP2 and HF results exhibit very similar trends; MP2 results are generally slightly smaller than HF ones, with a difference of 5–8 kcal mol−1, which diminishes significantly for conformers lacking any O–H···O IHB. Other than these observations, no distinct patterns relating the HOMO-LUMO gap to the presence of IHBs are noted. The smallest energy gap magnitude among the considered molecules corresponds to the M5-ε and M5-ζ isomers.

The ranges of the HOMO-LUMO energy gap for the conformers of the ε- and ζ-isomers of the same pair are comparable, indicating that the reversal of the positions of AC-ald and AC-isb does not influence the gap. For completeness, Table S4.32 reports the HOMO-LUMO energy gaps for the calculated conformers of grandinol.

The energies of the frontier orbitals are utilized to calculate global chemical reactivity descriptors [60-64], which are useful for evaluating and comparing various aspects of a molecule's reactivity. The term "global" indicates that these quantities consider the overall molecule; however, local estimations are also possible, allowing for comparisons of reactivity in different areas of the molecule.

Since the HOMO-LUMO energy gap remains unaffected by the reversal of the positions of the two acyl groups in the ε- and ζ-isomers, these quantities are not expected to be influenced. Therefore, their calculation does not strictly align with the objective of the present work. Instead, descriptors that can be derived using simple arithmetic operations without adding additional computational tasks to the study are preferred. This approach enables the calculation of examples and facilitates comparisons among different conformers of the same molecule. The names and operational definitions of these descriptors are as follows:

chemical hardness

$$\eta = (E_{LUMO} - E_{HOMO})/2$$



© University of Venda

electronic chemical potential

$$\mu = -(ELUMO + EHOMO)/2$$

electrophilicity index

$$\omega = \mu 2/2\eta$$

maximum electronic charge

$$\Delta N_{max} = -\mu/\eta$$

absolute electronegativity

$$\chi = -\mu$$

softness

$$S = 1/\eta$$

The relationship between the chemical hardness (η) and the HOMO-LUMO energy gap is direct, indicating the molecule's resistance to electron cloud polarization or deformation [65]. Smaller reactivity is associated with greater hardness. Chemical softness, inversely related to chemical hardness, signifies a higher tendency for electron cloud polarization and increased reactivity with higher softness values.

The electronic chemical potential (μ) pertains to the ability of a molecular system "to exchange electron density with the environment at the ground state" [62], indicating the direction of electron density flow in polar reactions. In such reactions, electron density flows from the molecule with higher μ to the one with lower μ, where the former acts as an electron donor and the latter as an electron acceptor. This concept is related to electronegativity, where the absolute electronegativity (χ) is the opposite of μ.

The electrophilicity index (ω) quantifies the energy stabilization of a molecule when it gains additional electron density from the environment [62]. The extent of stabilization depends on the



© University of Venda

molecule's tendency to acquire electron density (expressed by μ) and its resistance to exchanging electron density with the environment (expressed by η). However, the number of electrons a molecule can acquire is limited, with the maximum number denoted as $\Delta N_{max}$, a reactivity index.

Table 4.4 presents these six descriptors for the calculated conformers of the representative ε- and ζ-isomers in the DFT results. Tables S4.33, S4.34, and S4.35 present these six descriptors for the calculated conformers of the ε- and ζ-isomers, in the DFT, HF, and MP2 results, respectively. Tables S4.36, S4.37, and S4.38 combine the comparison for corresponding isomers and the comparisons across methods: η and S (Table S4.36), μ and χ (Table S4.37), and ω and $\Delta N_{max}$ (Table S4.38). Including the results from all the calculation methods allows for a broader comparison and choice possibilities, considering that several studies focusing on biologically active molecules analyze and report ab initio results [66–71].

**Table 4.4. Global reactivity descriptors for the calculated conformers of the representative ε- and ζ-isomer euglobal molecules. DFT/B3LYP/6-31+G(d,p) results.**

Results *in vacuo* from full optimization calculations.
The following descriptors are considered: chemical hardness (η), electronic chemical potential (μ), electrophilicity (ω), maximum electronic charge ($\Delta N_{max}$), absolute electronegativity (χ), and softness (S).
The molecules are denoted with the symbols listed in table 1, and the conformers with the symbols listed in table 2. For each molecule, the conformers are listed in order of increasing relative energies in the DFT results.

| Molecules and conformers | Global reactivity descriptors | | | | | |
|---|---|---|---|---|---|---|
| | η (eV) | μ (eV) | ω (eV) | $\Delta N_{max}$ | χ (eV) | S (eV) |
| **ε-isomers** | | | | | | |
| M1–ε-a-b-m-w-e | 2.25 | -4.17 | 3.87 | 1.86 | 4.17 | 0.44 |
| M1–ε-a-b-m-w-d | 2.25 | -4.18 | 3.89 | 1.86 | 4.18 | 0.45 |
| M1–ε-a-b-m-w-g | 2.25 | -4.18 | 3.90 | 1.86 | 4.18 | 0.45 |
| M1–ε-a-b-m-w-f | 2.24 | -4.17 | 3.88 | 1.86 | 4.17 | 0.45 |
| M1–ε-a-b-m-w-i | 2.24 | -4.17 | 3.88 | 1.86 | 4.17 | 0.45 |
| M1–ε-a-b-m-w-h | 2.24 | -4.17 | 3.88 | 1.86 | 4.17 | 0.45 |
| M1–ε-a-m-r-t-e | 2.22 | -4.13 | 3.84 | 1.86 | 4.13 | 0.45 |
| M1–ε-b-w-x-u-e | 2.33 | -4.10 | 3.62 | 1.76 | 4.10 | 0.43 |
| M1–ε-c-n-r-x-e | 2.25 | -4.07 | 3.69 | 1.81 | 4.07 | 0.45 |
| M1–ε-c-n-w-x-e | 2.15 | -4.16 | 4.02 | 1.93 | 4.16 | 0.46 |
| M1–ε-b-m-w-y-e | 2.17 | -4.20 | 4.06 | 1.94 | 4.20 | 0.46 |



© University of Venda

| | | | | | | |
|---|---|---|---|---|---|---|
| M1–ε-b-n-w-x-e | 2.17 | -4.22 | 4.10 | 1.94 | 4.22 | 0.46 |
| M1–ε-c-m-w-y-e | 2.30 | -4.03 | 3.54 | 1.76 | 4.03 | 0.44 |
| M1–ε-r-t-x-u-e | 2.30 | -4.02 | 3.53 | 1.75 | 4.02 | 0.44 |
| | | | | | | |
| **ζ-isomers** | | | | | | |
| M1–ζ-a-b-q-w-e | 2.24 | -4.15 | 3.84 | 1.85 | 4.15 | 0.45 |
| M1–ζ-a-b-q-w-g | 2.22 | -4.17 | 3.92 | 1.88 | 4.17 | 0.45 |
| M1–ζ-a-b-q-w-d | 2.22 | -4.17 | 3.91 | 1.88 | 4.17 | 0.45 |
| M1–ζ-a-b-q-w-f | 2.24 | -4.15 | 3.84 | 1.85 | 4.15 | 0.45 |
| M1–ζ-a-b-q-w-h | 2.21 | -4.18 | 3.96 | 1.89 | 4.18 | 0.45 |
| M1–ζ-a-b-q-w-i | 2.21 | -4.18 | 3.95 | 1.89 | 4.18 | 0.45 |
| M1–ζ-c-p-r-z-e | 2.24 | -4.13 | 3.81 | 1.85 | 4.13 | 0.45 |
| M1–ζ-a-q-r-e | 2.27 | -4.10 | 3.69 | 1.80 | 4.10 | 0.44 |
| M1–ζ-a-w-k-e | 2.24 | -4.10 | 3.74 | 1.83 | 4.10 | 0.45 |
| M1–ζ-c-r-x-e | 2.12 | -4.26 | 4.28 | 2.01 | 4.26 | 0.47 |
| M1–ζ-a-p-r-e | 2.26 | -4.05 | 3.62 | 1.79 | 4.05 | 0.44 |
| M1–ζ-a-p-r-k-e | 2.23 | -4.09 | 3.75 | 1.84 | 4.09 | 0.45 |
| M1–ζ-r-j-z-e | 2.34 | -4.05 | 3.50 | 1.73 | 4.05 | 0.43 |

As anticipated, the values of these quantities for the ε- and ζ-isomers are comparable within the same calculation method. Discrepancies in the results of different methods reflect differences in evaluating the individual energies of HOMO and LUMO and their energy gap, as discussed previously. Thus, DFT values of η are notably smaller than HF and MP2 values; correspondingly, DFT values of S, χ, ω, $\Delta N_{max}$, and the magnitude of μ are substantially greater than the ab initio ones. While values of the same quantity are generally comparable across molecules within the same calculation methods, DFT results for M5–ε, M5–ζ, and M7–ε show slightly smaller η, μ, and χ values compared to other isomers, with somewhat greater values for S. For M4–ε, M4*–ζ, M5–ε, and M5–ζ, DFT results for ω and $\Delta N_{max}$ are slightly greater than those of other isomers. The results of the two ab initio methods are very close, with MP2 values slightly smaller than HF ones for η and slightly greater for μ, χ, ω, and $\Delta N_{max}$.

Tables S4.39, 4.40, and S4.41 present the values of the six descriptors for the calculated τ-type model structures across DFT, HF, and MP2 results, respectively. The values of η and S exhibit comparability to those of the ε and ζ isomers, with observations similar to those mentioned earlier, such as the smaller η values of M5 and M7-ε and the greater S values of M5 and M7-ε. Meanwhile, the values of μ, χ, ω, and $\Delta N_{max}$ are slightly elevated for the τ structures, with a particularly noticeable difference for ω. This indicates that substituting one of the AC-ald groups in the τ-



© University of Venda

structures with AC-isb (as in the EGs) has a non-negligible influence on chemical reactivity. The increased bulk of the isobutyl group may partially account for the decrease in reactivity by hindering access to the $sp^2$ O of AC-isb.

### 4.4.4.2. Dipole moments of the conformers

The dipole moment is a property often associated with the biological activity of molecules [72] and is commonly listed among QSAR descriptors. Early investigations into anticancer agents, as noted by [73], highlighted its significance in the activity of anthracyclines, a major class of anticancer drugs, where activity is observed only within a specific range of dipole moment values. Table S4.42 presents the dipole moments of the calculated conformers of ε- and ζ-isomers. At the same time, Figure S4.13 illustrates the results from the three calculation methods for each isomer, and Figure S4.14 compares the values for the lower energy conformers in ε-isomer/ζ-isomer pairs. Figure 4.12 compares the values for the lower energy conformers in representative ε and ζ isomer pairs (i.e. euglobal IIC (M1-ε) and euglobal T1 (M1-ζ) pair), of the calculated euglobal molecules. The patterns observed in the results of the three methods are similar.

**Figure 4.12. Comparison of the dipole moments of the lower energy conformers in representative ε and ζ isomer pairs (i.e. euglobal IIC (M1-ε) and euglobal T1 (M1-ζ)), of the calculated euglobal molecules.**
Results from HF/6-31G(d,p), DFT/B3LYP/6-31+G(d,p) and MP2/6-31G(d,p) full optimization calculations *in vacuo*, respectively denoted as HF, DFT and MP2 in the figures' legends.
For each pair, the conformers' pairs are denoted by numbers on the x axis; the correspondence between numbers and pairs is shown in tables before each diagram.

Numbers denoting the conformers of M1–ε on the x axis.

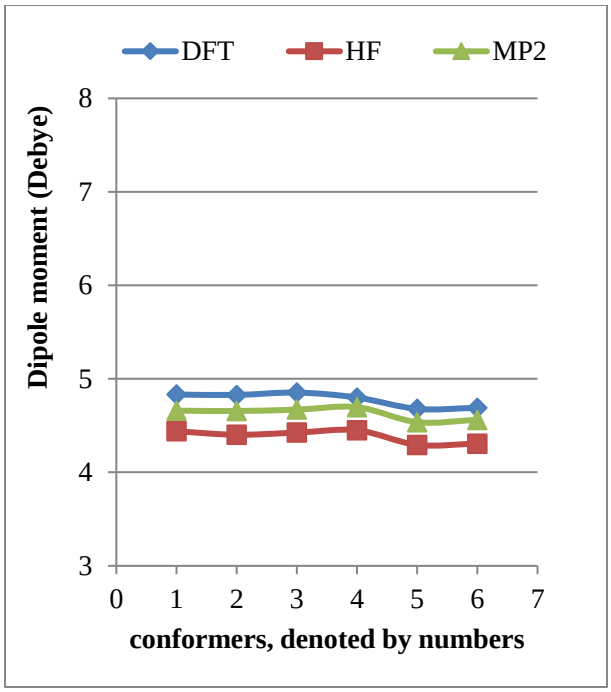
| # | Conformer | # | Conformer | # | Conformer |
|---|---|---|---|---|---|
| 1 | M1–ε-a-b-m-w-e | 3 | M1–ε-a-b-m-w-g | 5 | M1–ε-a-b-m-w-i |
| 2 | M1–ε-a-b-m-w-d | 4 | M1–ε-a-b-m-w-f | 6 | M1–ε-a-b-m-w-h |

Numbers denoting the conformers of M1–ζ on the x axis.

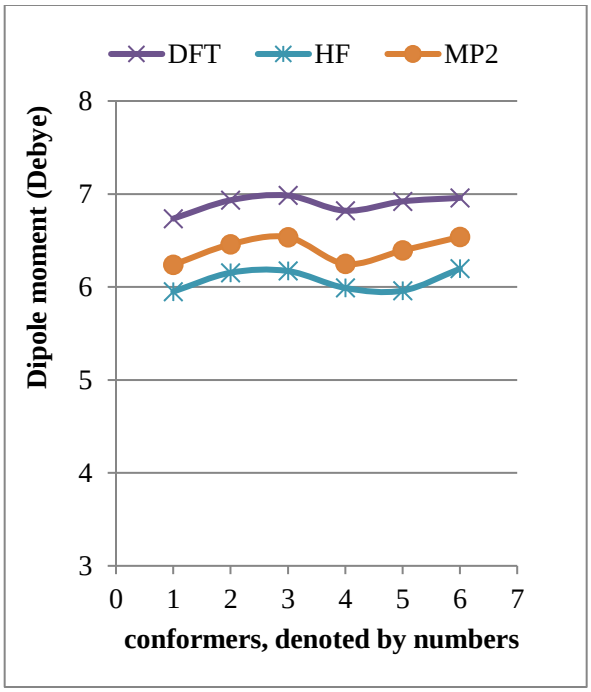
| # | Conformer | # | Conformer | # | Conformer |
|---|---|---|---|---|---|
| 1 | M1–ζ-a-b-q-w-e | 3 | M1–ζ-a-b-q-w-d | 5 | M1–ζ-a-b-q-w-h |
| 2 | M1–ζ-a-b-q-w-g | 4 | M1–ζ-a-b-q-w-f | 6 | M1–ζ-a-b-q-w-i |



© University of Venda




M1-ε




M1-ζ

The dipole moments of the ε-isomers are consistently smaller than those of the ζ-isomers. Among conformers with both H17···O14 and H16···O18, the dipole moments are comparable within the same isomer. The smallest dipole moments are found in conformers with only H16···O18. The highest values are observed in conformers with only H17···O14 (a-conformers) or only H17···O18 (c-conformers), where the orientation of O12–H17 is 'to the left' and O14 is 'to the right' (with respect to the orientation used in this study). Within each category (a or c), the value is greater when H10···O16 is oriented 'to the right' (r-conformers) compared to when it is oriented 'to the left' (w-conformers).

In the case of ε-isomers, the highest dipole moments for a given molecule may belong to either a-conformers or c-conformers, depending on the molecule. On the other hand, in the case of ζ-isomers, the highest values consistently belong to a-conformers, with c-conformers following at a certain distance. These patterns arise from the dependency of the dipole moment on the relative orientations of the OH groups, O14 and O18. Additionally, there is some reliance on the terpene moiety, as evidenced by variations such as the conformers of M4-ε having the lowest dipole moments and those of M7-ε having the highest among the ε-isomers. Similarly, the conformers of



© University of Venda

M4*-ζ exhibit the lowest dipole moments, while those of M8*-ζ have the highest among the ζ-isomers.

#### 4.4.4.3. The Mulliken charges on relevant atoms

Tables S4.43, S4.44, and S4.45 provide the Mulliken charges on the O atoms of the calculated conformers of the ε- and ζ-isomers, based on the DFT, HF, and MP2 results, respectively. Tables S4.46 and S4.47 delineate the ranges of the magnitudes of these charges concerning the O–H···O IHBs present, focusing on the DFT and MP2 results, respectively. The MP2 findings show more consistent patterns. Specifically, the magnitude of the negative charge on the O atoms increases when engaged in an IHB. This discrepancy is more pronounced for acceptor atoms (O14 or O18) and less for donor atoms (O10 or O12). Moreover, the difference is more pronounced when both O12–H17···O14 and O10–H16···O18 are present, compared to when either is singularly present, indicating a stronger electron-withdrawing effect in the former scenario. This observation supports the notion of some coordination between the two IHBs, even if not consecutive. Additionally, the charge's magnitude on acceptor atoms is higher on O18 for ε-isomers and on O14 for ζ-isomers, emphasizing a greater effect on AC-isb.

Table S4.48 presents the charges on the H atoms of the OH groups in the calculated conformers of the ε- and ζ-isomers. These charges notably increase when the H atom participates in an IHB. The patterns align with those observed for the O atoms, highlighting consistent trends in IHB strengths (with stronger IHBs exhibiting greater charge differences between the O and H atoms). The highest charges are observed on H16 when engaged in H16···O18 interactions.

In addition, Tables S4.49, S4.50, and S4.51 detail the Mulliken charges on the O atoms for the conformers of the calculated τ-type model structures. Table S4.52 reports the charges on the H atoms. Notably, the charges on O14 and O18, acting as IHB acceptors, demonstrate comparability, indicating that differences observed for the ε- and ζ-isomers are influenced by the terpene moiety's presence. Furthermore, the charges exhibit some influence from O–H···O IHBs, with slightly higher charges on O8 or O12 in their presence. Similarly, the charges on the H atoms are comparable, with H17 displaying only a slightly greater charge than H16.



© University of Venda

The charges obtained from the three calculation methods are generally consistent across all considered atoms except O8. MP2 yields the highest charge values, while DFT yields the smallest. Moreover, MP2 offers more closely aligned atom values in analogous scenarios (such as similar IHBs), thus elucidating clearer patterns. Notably, DFT produces significantly smaller charges than the other methods for O8. While delving into the reasons behind this discrepancy would necessitate further computations, it is not a focal point of the current investigation and may be explored in a separate study.

### 4.5. In-solution results

The polarity of a solvent plays a crucial role in determining its behavior in various chemical processes. One way to measure the polarity of a solvent is by looking at its dielectric constant, also known as permittivity. The Dielectric Constant (ε) is a dimensionless constant that indicates how easily a material can be polarized by the imposition of an electric field [74]. The results show a clear trend in the polarity of *in-vacuo*, chloroform, acetonitrile, and water. *In-vacuo*, being a vacuum, has a dielectric constant of 1, the lowest value on the scale. Chloroform, a nonpolar solvent, has a dielectric constant of 4.1. Acetonitrile, a polar aprotic solvent, has a dielectric constant of 36.6. Finally, water, a highly polar solvent, has a dielectric constant of 80 at room temperature [75]. These values indicate that the polarity of a solvent increases as its dielectric constant increases. In other words, more polar solvents have higher dielectric constants than nonpolar solvents. Therefore, we can conclude that the polarity of the solvents in the given sequence is *in-vacuo* < chloroform < acetonitrile < water. The calculations consider the following: relative energies and energy-related quantities (HOMO-LUMO energy gap, solvation energy, dispersion energy, total electrostatic energy) and dipole moment.

#### 4.5.1. Relative energies of the conformers across different media

Calculations in solution were performed using both HF and DFT levels for all conformers of the considered EG molecules. Table S4.53 and S.54 present the relative energy of these conformers in various solvents (vacuum, chloroform, acetonitrile, and water). The values indicate that the conformers have lower energy in solution than in vacuum across all methods. Moreover, the relative energy of the conformers slightly decreases with increasing solvent polarity for conformers where both the O–H···O IHBs are present. It drastically decreases in those where the



© University of Venda

IHBs are absent. This suggests that solvent molecules can interact with the hydroxyl groups and an $sp^2$ O atom not engaged in the IHBs, leading to the stabilization of the solute.

The solvent effect, represented by the free energy of solvation ($\Delta G_{solv}$), indicates the stabilization of a solute molecule in solution due to solute-solvent interactions. This property helps to understand a system's thermodynamic stability [76]. Solvation free energies, as described in [77], provide insight into the free energy change associated with the transfer of a molecule between an ideal gas and a solvent at a specific temperature and pressure. Tables S4.55 and S4.56 report the free energy of solvation (solvent effect, $\Delta G_{solv}$) and its electrostatic component ($G_{el}$) for the conformers of all considered EG molecules.

The calculated $\Delta G_{solv}$ is mainly influenced by the type of solvent considered, the type of molecule, and the type of conformer considered. The magnitude value of $\Delta G_{solv}$ differs significantly across different media. The extent of the magnitude of solvent effect for different media is in the following order: $\Delta G_{solv}$ in acetonitrile > $\Delta G_{solv}$ in chloroform > $\Delta G_{solv}$ in water, as shown in Table S4.55 and S4.56; and Table 4.5 indicates the range across conformers of different molecules. It is important to note that a negative sign of the solvent effect indicates that the solvent brings no stabilization, while the positive sign shows stabilization.



© University of Venda

**Table 4.5. Ranges of the solvation free energy ($\Delta G_{solv}$, kcal mol$^{-1}$) for the calculated conformers of EG molecules considered in this work.** DFT/B3LYP/6-31+G(d,p) and HF/6-31G(d,p) result from full optimization calculations, respectively denoted as DFT and HF in the columns' headings.

| Molecule | DFT | | | HF | | |
|---|---|---|---|---|---|---|
| | chlrf | actn | aq | chlrf | actn | aq |
| M1–ε | 1.71–8.29 | 6.40–14.73 | -10.14–4.57 | -0.43–7.68 | 4.60–14.43 | -10.50–1.94 |
| M2–ε | 1.77–8.41 | 6.18–14.96 | -8.76–5.06 | -1.28–7.56 | 3.55–10.38 | -8.31–2.73 |
| M3–ε | 1.4–8.12 | 5.99–14.78 | -10.14–4.73 | -2.09–2.70 | 2.95–9.57 | -9.72–1.62 |
| M4–ε | 1.18–8.43 | 6.17–15.09 | -10.22–5.30 | -0.75–3.50 | 4.03–10.22 | -9.31–2.88 |
| M5–ε | 0.92–7.75 | 4.38–13.52 | -10.8–3.48 | -2.15–2.84 | 2.46–9.60 | -12.98– (-0.58) |
| M6–ε | 2.21–8.47 | 6.37–14.82 | -9.64–4.87 | -0.47–4.08 | 4.43–10.99 | -10.09–1.93 |
| M7–ε | 0.80–7.32 | 4.27–13.13 | -11.92–2.95 | -1.47–2.96 | 3.04–9.68 | -11.84–0.76 |
| M8–ε | 1.40–7.99 | 5.84–15.08 | -10.77–4.20 | -1.52–3.12 | 3.43–9.95 | -11.02–1.37 |
| M9*–ε | 2.02–8.26 | 6.44–14.79 | -9.40–4.86 | -1.59–2.91 | 3.43–9.75 | -10.73–1.09 |
| M10–ε | 6.69–12.57 | 12.99–20.34 | -3.73–9.81 | 1.95–4.59 | 9.22–12.74 | -2.04–4.15 |
| | | | | | | |
| M1–ζ | 2.28–8.85 | 6.79–15.23 | -9.16–5.25 | 2.28–8.97 | 4.76–11.42 | -6.82–3.08 |
| M2–ζ | 2.60–8.62 | 6.83–14.96 | -8.17–5.14 | 2.60–8.76 | 4.46–11.21 | -7.74–3.76 |
| M3–ζ | 2.27–8.68 | 6.65–15.04 | -8.67–5.13 | 2.27–8.68 | 4.54–11.01 | -8.64–3.85 |
| M4*–ζ | 1.94–8.48 | 6.48–14.84 | -9.36–4.86 | 1.20–8.68 | 3.88–11.19 | -9.43–4.17 |
| M5–ζ | 1.26–7.54 | 5.01–13.08 | -11.10–3.01 | 1.26–8.31 | 3.63–10.30 | -12.54–0.37 |
| M6*–ζ | 3.08–9.23 | 7.29–15.51 | -8.09–5.79 | 3.08–9.23 | 5.76–12.38 | -6.76–4.00 |
| M7–ζ | 2.55–9.50 | 7.19–15.80 | -8.73–6.03 | 2.55–9.50 | 3.33–15.55 | -11.43–3.76 |
| M8*–ζ | 1.48–8.73 | 5.89–15.13 | -10.79–5.20 | 1.48–8.73 | 4.55–11.11 | -7.87–3.13 |
| M9–ζ | 2.34–8.34 | 6.80–14.52 | -9.03–4.47 | 2.34–8.34 | 3.80–15.26 | -10.33–3.36 |
| M10*–ζ | 7.47–13.33 | 13.10–20.85 | -2.11–10.58 | 7.47–13.33 | 6.89–21.05 | -5.56–5.34 |
| | | | | | | |
| M1*–τ | -4.01–3.22 | -0.92–8.21 | -17.89– (-2.08 | -4.35–3.30 | -0.92–8.64 | -20.48– (-3.75) |
| M2*–τ | -2.61–3.28 | 0.72–8.32 | -15.50– (-1.86) | -3.51–3.38 | 0.52–8.80 | -19.01– (-3.42) |
| M3*–τ | -4.91–2.82 | -1.73–7.95 | -18.56– (-2.25) | -4.95–3.00 | -1.27–8.51 | -20.75– (-3.76) |
| M4*–τ | -4.11–3.37 | -0.69–8.55 | -17.37– (-1.45) | -4.62–3.36 | -0.93–8.82 | -20.26– (-0.46) |
| M5*–τ | -5.22–2.56 | -2.96–6.85 | -19.82– (-3.23) | -5.24–2.87 | -2.46–7.68 | -21.85– (-4.43) |
| M6*–τ | -3.21–3.40 | -0.25–8.23 | -16.97– (-1.98) | -3.64–3.44 | -0.35–8.62 | -19.98– (-1.12) |
| M7*–τ | -3.62–2.14 | -1.47–6.48 | -15.27– (-3.86) | -3.52–2.29 | -0.81–6.98 | -16.89– (-5.44) |
| M8*–τ | -4.02–3.28 | -1.06–8.29 | -18.05– (-2.11) | -4.48–3.38 | -1.14–8.73 | -20.89– (-3.84) |
| M9*–τ | -0.78–8.37 | -0.98–8.29 | -18.05– (-2.11) | -4.48–3.38 | -1.14–8.73 | -20.89– (-3.85) |
| M10*–τ | 1.45–7.51 | 6.12–13.82 | -10.89–3.21 | 0.74–7.48 | 5.76–14.13 | -14.19–1.42 |
| | | | | | | |
| N1-τ | 5.52–13.05 | 11.04–20.35 | -6.41–9.76 | 3.48–13.03 | 9.64–20.69 | -4.61–6.89 |
| N2-τ | 0.87–8.66 | 5.06–14.92 | -11.67–4.78 | -0.23–8.68 | 4.15–15.29 | -10.25–2.75 |
| N3-τ | 1.27–9.07 | 5.79–15.61 | -11.25–5.23 | -0.19–9.08 | 4.62–15.96 | -9.07–3.40 |
| N4-τ | 0.20–8.35 | 4.81–14.87 | -12.41–4.22 | -1.14–8.22 | 3.54–15.02 | -10.64–1.31 |
| N5-τ | 0.75–8.56 | 5.28–15.06 | -11.62–4.73 | -0.83–8.52 | 3.72–15.34 | -9.53–2.65 |



© University of Venda

| | | | | | | |
|---|---|---|---|---|---|---|
| N6-τ | 7.19–12.49 | 13.19–20.11 | -2.96–20.11 | 3.13–12.67 | 9.23–20.61 | -4.32–5.70 |
| N7-τ | 4.58–11.91 | 10.15–19.37 | -6.94–19.37 | 1.81–11.77 | 7.42–19.50 | -5.48–6.13 |
| N8-τ | -6.03–14.97 | -8.74–22.74 | -19.10–22.74 | 3.59–14.65 | 9.63–22.67 | -3.32–8.01 |

In acetonitrile, the solvent effect is always positive for all calculated conformers of different molecules (except for most conformers of M*–τ molecules), with the highest value reaching up to 21.05/ HF and 20.85/ DFT kcal $mol^{-1}$ and the lowest value at 2.46/HF and 4.15/HF kcal $mol^{-1}$. These positive values indicate that the acetonitrile solvent brings no stabilization. On the other hand, in chloroform, the solvent effect is mostly positive for all calculated conformers of different molecules, with a few exceptions depending on the type of conformers, the type of molecule considered, and the method utilized. For example, the highest value on the upper ends of the range is 13.33/ HF and 13.33/ DFT kcal $mol^{-1}$, and the lowest value on the lower ends of the range is -2.15/ HF and 0.34/ DFT kcal $mol^{-1}$. The positive values indicate that the solvent brings no stabilization, and the negative values show stabilization by the solvent. In the case of water, the solvent effect is mostly negative for all calculated conformers of different molecules, with a few exceptions depending on the type of conformers and the type of molecule considered. For example, the highest value on the upper ends of the range is 5.34/ HF and 10.58/ DFT kcal $mol^{-1}$, and the lowest value on the lower ends of the range is -12.98/ HF and -12.45/ DFT kcal $mol^{-1}$. Similarly to chloroform, the negative values indicate that the water brings stabilization, and the positive show no stabilization.

The position of AC-isb plays a significant role in the solvent effect. The solvent effect is slightly greater when AC-isb is attached to C5 (ε-isomers) than when it is attached to C1 (ζ-isomers) in acetonitrile and water and greatest in chloroform. This suggests that the bulky terpene moiety near C1 influences solute-solvent interactions. All the a-b conformers have positive values of the solvent effect, which are greater in magnitude compared to other types of conformers. a-b conformers are those in which the two IHBs are present. The results suggest a clear preference for forming intramolecular hydrogen bonds within solute atoms rather than intermolecular solute-solvent bonds, especially when a specific group within the solute molecule acts as the hydrogen bond donor. The challenge of solvent molecules acting as hydrogen bond acceptors for solute atoms can be attributed to the presence of stronger hydrogen bond acceptors or the scarcity of



© University of Venda

stronger hydrogen bond donors in the solute than the solvent. Conformers of higher energy, or those in which only one IHB or none are present, have small values of $\Delta G_{solv}$. This indicates that solvent molecules can form intermolecular hydrogen bonds with the central solute atoms acting as hydrogen bond acceptors or acceptors, stabilizing the solute molecules.

The extent of the solvent effect also shows some dependence on the terpene moiety of the molecule considered. $\Delta G_{solv}$ is different for different terpene moieties. Molecules with a less bulky substituent on the terpene moiety show smaller values of $\Delta G_{solv}$. Amongst the molecules with AC-is (i.e., ε-type and ζ-type isomers) present, the solvent effect is more pronounced in M5 and least pronounced in M10. Furthermore, for those in which both R are AC-ald, the solvent effect is more pronounced in M5 and least pronounced in M10. The decrease in solute-solvent interactions for polar atoms with high substitution levels can be attributed to the expansion of excluded volume, which hinders the approach of solvent molecules to the central solute atoms [76].
The electrostatic component ($G_{el}$) of $\Delta G_{solv}$ exhibits negative values in all three solvents, with its magnitude greater in water than in the other two solvents.

#### 4.3.3.1. Energies of the frontier molecular orbitals in different media

It has been observed that when the polarity of a solvent increases, the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) gradually becomes narrower [78]. This narrowing is primarily caused by the presence of the solvent, which can either decrease or increase the energy levels of HOMO and LUMO, ultimately affecting the HOMO-LUMO energy gap [79]. The polarity of the solvent plays a crucial role in modifying this energy gap. When the polarity of the solvent increases, it has an influence on the overall stability of the molecular orbitals. This, in turn, brings about a shift in the energy levels of both the HOMO and LUMO orbitals [78]. As a result, the energy needed for an electron to move between these orbitals decreases, making it easier for this transition to occur at lower energy levels.

In addition, the presence of a solvent can also influence the way molecules interact with each other. The polarity of the solvent can either stabilize or destabilize the molecular orbitals, depending on the characteristics of the solvent and the molecule being dissolved [79]. As a result, this can lead to either a rise or a fall in the energy difference between the highest occupied molecular orbital



© University of Venda

(HOMO) and lowest unoccupied molecular orbital (LUMO), highlighting just how significant the polarity of a dissolving substance can be in the realm of molecular chemistry.

For the considered molecules in this study, in most cases (especially for the lowest energy conformers), the energy gap decreases as the medium polarity increases. In some cases, it slightly increases, as shown in Table S4.57 and S4.58. Furthermore, there is no consistency in how it decreases or increases with the solvent's polarity.

**4.5.2. Parameters of the O–H···O and C–H···O IHBs in different medium**

The results from Table S4.59 and S4.60 show that when there is a stabilizing or destabilizing of the hydrogen bond of the solute, chloroform shows minimal effect, acetonitrile shows moderate, and water shows a significant stabilization/ destabilizing effect.

Chloroform molecules do not have significant dipole moments, so hydrogen bonding interactions with polar atoms from the solutes are weak. As a result, when a hydrogen bond between a hydrogen donor and an oxygen acceptor in a solute encounters chloroform molecules, it experiences little destabilization due to weak solvent-solute interactions. Therefore, the hydrogen bond is relatively unaffected and maintains its strength. Acetonitrile molecules have a permanent dipole moment due to the electronegativity difference between carbon and nitrogen atoms. So, when an intramolecular hydrogen bond from a solute encounters acetonitrile molecules, the polar solvent molecules can interact more strongly with the hydrogen bond than chloroform molecules. The polar nature of acetonitrile leads to increased solvation of the hydrogen and oxygen atoms involved in the hydrogen bond, resulting in some destabilization of the hydrogen bond compared to chloroform.

The effect observed in water is because water's strong polarity and hydrogen bonding capability lead to significant solvation of the hydrogen and oxygen atoms involved in the intramolecular hydrogen bond. The extensive solvation in water results in considerable stabilization of the intramolecular hydrogen bond compared to chloroform and acetonitrile.

Solvent molecules have the role of stabilizing or destabilizing the hydrogen bond of the solute. Considering this work, this role is limited due to the type of the conformer and the type of molecule. In ε-type, solvent molecules seem to have an effect of increasing the bond length of



© University of Venda

H17···O18 and H17···O14 and decreasing H16···O18 as the solvent's polarity increases. On the other hand, a decrease in the bond length of H17···O18 and H17···O14 and an increase of H16···O18 as the polarity of the solvent increases is shown in ζ-type isomers. Furthermore, this analysis suggests that the position of AC-isb influences the effect of solvent molecules on the intramolecular hydrogen bond formed within the considered solute molecule in this study. When AC-isb is attached at the C5 position, it leads to an increase of the O12–H17···O18 and O12–H17···O14 as the medium's polarity increases, and when it is attached to the C1 position, the observed trend reverse. Similarly, when AC-isb is attached at the C5 position, the H16···O18 bond length decreases, while when it is at C1, the trend reverses.

Similarly to the ζ-type isomers, O12–H17···O18 and O12–H17···O14 bond length decreases as the solvent polarity increases in τ-type molecules. However, when considering the bond length of O10–H16···O18, it increases with the solvent's polarity when only O12–H17···O14 IHB is absent and decreases when it is present. Thus, this did not occur in the ζ-type isomers. So, this suggests the effect of the presence of the AC-isb on how solvent effect affects the O10–H16···O18 bond length.

The bond length of C7–H13···O8 and C11–H19···O10 shortens as the solvent polarity increases in all the considered molecules. Solvent molecules from different media have a stabilizing effect on the intramolecular C–H···O hydrogen bond.

### 4.5.3. Conformers' dipole moment in different media

Regarding the dipole moment behavior of solutes in different solvents, the results show that the dipole moment of calculated conformers of considered molecules increases as the polarity of the surrounding medium increases. This indicates that the polarity of the medium can significantly influence the dipole moment of the molecule. Specifically, it has been observed that the dipole moment increases in the following order: *in-vacuo* (vc) < in chloroform (chlm) < in acetonitrile (actn) < in water (wtr) as shown in Table S4.61 and S4.62. However, the magnitude of the increase in the dipole moment across different media is not only influenced by the polarity of the solvent but also by other factors such as the type of conformer, the presence and position of AC-isb, and



© University of Venda

the type of terpene moieties for this class of molecules. The results of the two methods used in the study show similar patterns.

To compare the increase in the dipole moment in different isomers across different media, τ-type is used a reference molecule to analyze the differences when AC-isb is attached to C1 or C5. The increase in the dipole moment shows significant differences when AC-isb is attached to C1 or C5. For an illustrative example, Table S4.61 and S4.62 compares the increase in the dipole moment across different solvents for the lowest energy conformers (in which both the IHB are present) of the corresponding molecules. The results show that the increase in the dipole moment is significant in both cases, but more pronounced when AC-isb is attached to C1. This suggests that the position of the AC-isb group in the molecule can have a significant effect on the dipole moment and its behavior in different solvents.

The results also show that the increase in the dipole moment is more pronounced in the a- and c-conformers in all molecules and least pronounced in b-conformers. This shows that the type of IHB presence in a molecule not only influences the value of dipole moment *in vacuo* but also influences how it changes across different media. There is also some dependence on the terpene moiety, as highlighted by the fact that, for example, the conformers of M4 in ε-type and ζ-type isomers and τ-type have the lowest increase dipole moments across different media and the conformers of M7 have the highest increase.

**4.6. Molecular docking results**

The docking protocols and parameters were initially validated using a re-docking approach [80]. This method involved extracting a known inhibitor (ligand) from a co-crystal structure of the protein of interest obtained from the Protein Data Bank (PDB), then re-docking it into the same protein using the chosen docking method. Successful re-docking, where the ligand aligns in the same orientation as in the original co-crystal structure, confirms the validity of the docking method. The re-docking results showed that the ligands matched their original binding modes within the proteins, affirming the reliability of the docking protocols and the accuracy of the obtained results.



© University of Venda

For these docking studies, twenty-three anticancer plant-derived compounds, fifteen modeled structures, and eleven molecular targets (proteins) were examined. Eight targets are associated with anticancer activity and three with antimalarial activity; details, including target names, PDB IDs, and brief descriptions, are provided in Table 4.6. Structural representations are shown in Figure S4.15, with binding sites highlighted by co-crystallized ligands. Each ACPL was docked with its relevant targets, and their binding energies and intermolecular interactions were assessed and are reported in Table S4.63, alongside those of the co-crystal ligands, serving as comparative references.

**Table 4.6.** List of the considered anticancer and antimalarial targets with their PDB ID and short descriptions.

| **Considered receptors** | **PDB ID** | **Short description** |
|---|---|---|
| **Anticancer target** | | |
| Epidermal Growth Factor Receptor (EGFR) | 6LUD | • trans-membrane glycoprotein that regulates cellular proliferation<br>• drives metastasis through paracrine loops involving tumor-stromal interactions, enabling invasion, survival of circulating tumor cells, and colonization of distant organs. |
| Janus kinase (JAK3) | JAK3 | • a tyrosine kinase that plays a central role in lymphopoiesis, forming lymphocytes or white blood cells.<br>• mutations in JAK3 can lead to severe combined immunodeficiency or various types of cancers. |
| Topoisomerase I (Topo I) | IT8I | • enzyme that alters DNA supercoiling, which is crucial for replication and transcription<br>• the emergence and spread of cancer have been intimately associated with topoisomerase dysregulation. |
| Phosphatidylinositol 3-kinase (PI3K) | 5JHB | • lipid kinase that regulates diverse cellular processes, including proliferation, adhesion, survival, and motility.<br>• dysregulated PI3K pathway signaling occurs in one-third of human tumors. |
| a mutant of BRAF gene (BRAFV600E) | 6V34 | • a constitutively active onco-kinase<br>• is the most common genetic alteration in papillary thyroid carcinoma (PTC), and anaplastic thyroid carcinoma as well, albeit |



© University of Venda

| | | |
|---|---|---|
| heat shock protein 90 (HSP90) | 3TUH | stabilizes oncogenic substrate "client" proteins in cellular processes promoting tumorigenesis. |
| human epidermal growth factor receptor 2 (HER2) | 3RCD | • a transmembrane tyrosine kinase receptor protein<br>• promotes the growth of cancer cells |
| Cyclin-dependent kinase 2 (CDK2) | 1DI8 | • drives the progression of cells into the S- and M-phases of the cell cycle<br>• associated with tumor growth in multiple cancer types |
| **Antimalarial targets** | | |
| Plasmodium falciparum lactate dehydrogenase (PfLDH | 1U5A | enzyme for energy generation of malarial parasites |
| Plasmodium falciparum malate dehydrogenase (PfMDH) | 6R8G | • enzyme that plays a crucial role in the energy production of the malaria parasite<br>• might complement PfLDH function in Plasmodium falciparum<br>• presence of PfMDH in other malaria parasites and its absence in mammals make it an ideal target for the development of selective antimalarial |
| Plasmodium falciparum Phosphoethanolamine-methyltransferase (PfPMT) | 3UJ9 | • catalyzes the three-step methylation of phosphoethanolamine to form phosphocholine, a critical step in the synthesis of phosphatidylcholine in a select number of eukaryotes including human malaria parasites<br>• essential for parasite gametocytogenesis and malaria transmission<br>• similarly to PfMDH, the presence of PfPMT orthologs in other malaria parasites that infect humans and their absence in mammals make them ideal targets |

The interactions between ligands and proteins, such as those examined in this study and listed in Maestro software, include hydrogen bonds (H-bonds), aromatic H-bonds, hydrophobic interactions, π···π stacking interactions, polar interactions, glycine interactions, and cation···π interactions. H-bonds, which are well understood and recognized as the strongest non-covalent interactions, play a critical role in biological complexes by determining molecular recognition specificity. Aromatic H-bonds are a specific type of H-bond, where an aromatic ring acts as the acceptor in an O–H···π interaction.



© University of Venda

Hydrophobic interactions describe the tendency of non-polar molecules or regions of molecules to cluster together in aqueous environments. In ligand-protein complexes, these interactions occur between the non-polar regions of the ligand and the protein's binding site, contributing significantly to the overall binding. They can also be influenced or adjusted, making them relevant in drug design.

π···π stacking interactions involve the interaction between aromatic planes, typically in edge-to-face or parallel face-to-face configurations. In these interactions, an aromatic ring from the protein's amino acid (commonly phenylalanine, followed by tyrosine, tryptophan, or histidine) interacts with an aromatic ring from the ligand. These interactions are crucial for biological recognition and the organization of biomolecular structures.

Polar interactions refer to the non-bonded interactions between polar donor and acceptor atoms in the ligand-protein complex. The interaction is considered significant within a certain atom-atom distance range, such as 2.5–3.5 Å. Polar interactions are frequently observed in bound fragments, particularly when the solvent is excluded from the model.

Glycine interactions are particularly notable since glycine often acts as an acceptor in C-H···O H-bonds and as a donor in amide···π interactions.

Cation···π interactions involve a positively charged nitrogen atom interacting with the electron-rich region of an aromatic ring. In the case of the ACPLs studied here, which lack nitrogen atoms, these interactions typically occur between the cationic amino acid residue in the protein and the aromatic ring of the ligand. These interactions are important for small-molecule recognition at protein binding sites.

### 4.6.1. Molecular docking results of selected anticancer targets against the considered compounds.

#### 4.6.1.1. Molecular interaction of considered compounds with EGFR

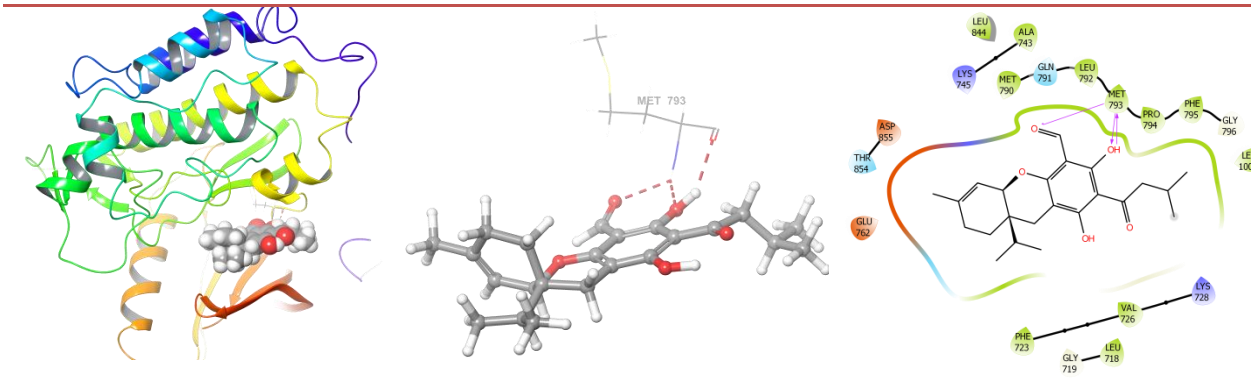
The molecular docking analysis of EGFR with the selected compounds shows that all the molecules can interact effectively at the active site. M8-ε is the best docked compound, with a binding energy of -9.201 kcal $mol^{-1}$. Although its interaction with EGFR is not significantly



© University of Venda

stronger than that of the co-crystal ligand, it forms three HBs with MET793. It also engages in hydrophobic interactions with LEU1001, PHE795, PRO794, MET793, LEU792, MET790, and ALA743 (Figure 4.6.1).

Additionally, among the compounds analyzed, those with better energy scores include N5-τ, M5-ζ, M9*-τ, N3-τ, and M8*-τ, with XP Gscores -9.077, -8.915, -8.365, -8.345, and -8.229 kcal mol$^{-1}$ respectively. While the presence and position of the AC-isb do influence the energy score, no clear pattern or trend is observed. Furthermore, the nature of the terpene moiety appears to significantly influence the energy scores, particularly noted in molecules where both R groups are AC-ald. Among the τ-type compounds, M9*-τ, M8*-τ, and M3*-τ exhibit better energy scores, while among the τ*-type, N5-τ and N3-τ stand out.




a b c

**Figure 4.6.1. Molecular insight into compound M8-ε docked on the anticancer target (EGFR, PDB: 6LUD), highlighting the residues within the binding site pocket.** Panel a) showcases the docking pose of the protein-ligand complex. Panels b and c offer 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.

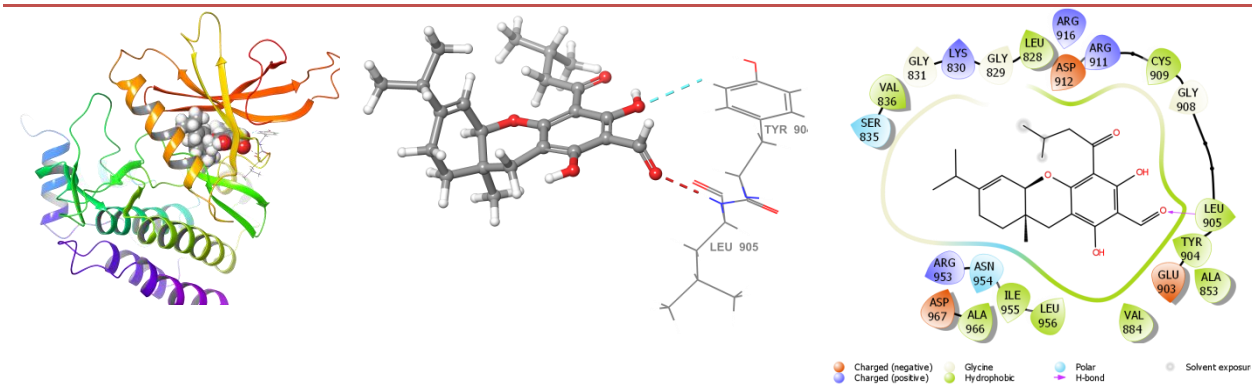
### 4.6.1.2. Molecular interaction of considered compounds with JAK3

The results of the molecular docking analysis of JAK3 with the selected compounds indicate that all the molecules can interact at the active site. Among them, M8-ε emerges as the top-docked compound, with a binding energy of -8.929 kcal mol$^{-1}$. Its interaction with JAK3 is not comparatively stronger than that of the co-crystal ligand; it forms one HB with LEU905. It also



© University of Venda

engages in hydrophobic interactions with ILE955, LEU956, VAL884, TYR904, LEU905, CYS, and LEU828 (Figure 4.6.2).




a b c

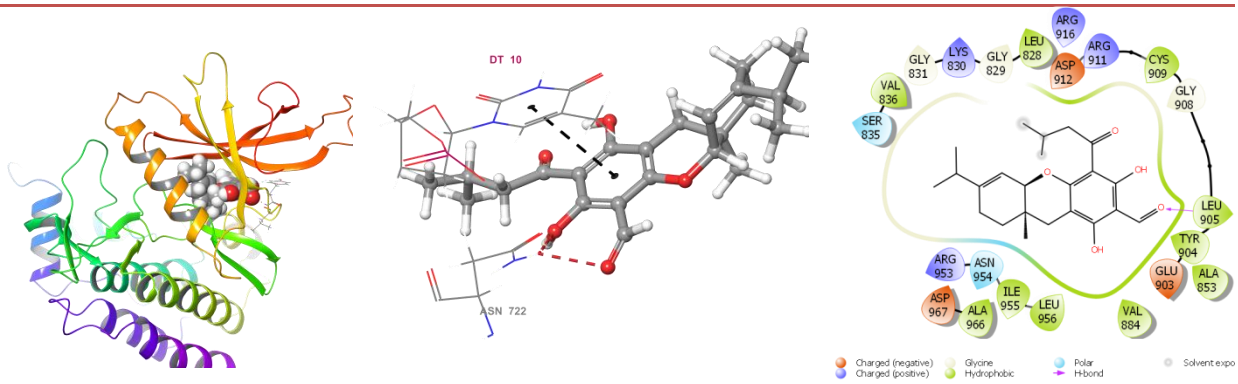
**Figure 4.6.2. Molecular insight into compound M8-ε docked on the anticancer target (JAK3, PDB: 7C3N), highlighting the residues within the binding site pocket.** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs, and light blue-dashed segments denote aromatic HBs.

### 4.6.1.3. Molecular interaction of considered compounds with Topo I

The molecular docking analysis of Topo I with the selected compounds demonstrates that all the molecules can effectively interact at the active site. M7-ε stands out as the best-docked compound, exhibiting a binding energy of -8.515 kcal $mol^{-1}$. Its interaction with Topo I is notably stronger than that of the co-crystal ligand, forming two HBs with ASN A:722, engaging in π-π stacking with DT B:9, and establishing hydrophobic interactions with PRO A:431 and LEU A:429, along with a polar interaction with ASN A:352 (Figure 4.6.3).



© University of Venda




a b c

**Figure 4.6.3. Molecular insight into compound M7-ε docked on the anticancer target (Topo I, PDB: IT8I), highlighting the residues within the binding site pocket.** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs, and black-dashed segments denote π-π stacking interactions.

### 4.6.1.4. Molecular interaction of considered compounds with P13K

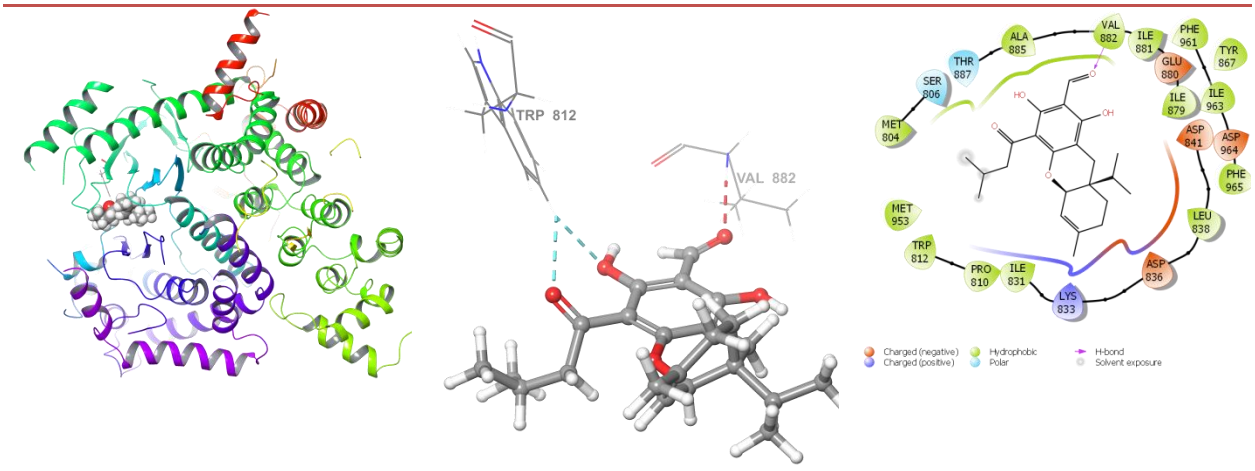
The majority of the considered molecules bind to site B1, while only a few interact with site B2. Among them, M8*-ζ is identified as the best-docked compound against P13K at site B1, with a binding score energy of -9.527 kcal $mol^{-1}$. Its interaction with P13K is notably stronger than that of the co-crystal ligand. M8*-ζ forms one HB with VAL882, two aromatic HBs with TRP 812, and engages in hydrophobic interactions with ILE881, VAL882, ALA885, and MET 953 (Figure 4.6.4).

Considering site B2, the molecules bound to this site did not exhibit better binding score energy. Most of the molecules binding to this site belong to the ζ-type and τ-type categories—specifically, only one molecule from each of the ζ-type and τ-type bound to this site.

The molecule identified as the best-docked compound at site B2 is M6*-ζ, which belongs to the ζ-type molecules. It achieved a binding score of -5.453 kcal $mol^{-1}$. This interaction is relatively comparable to that of the co-crystal ligand. It involves one HB with TYR608, two HBs with LYS425, and hydrophobic interactions with LEU423 and PRO424 (Figure 4.6.5).

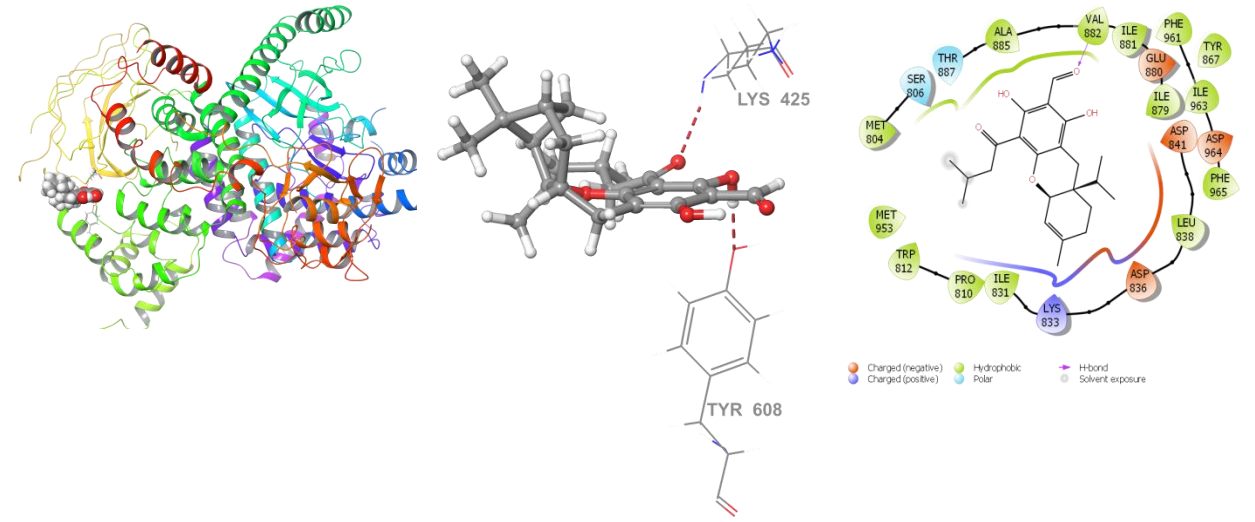


© University of Venda




a b c

**Figure 4.6.4. Molecular insight into compound M8*-ζ docked on the anticancer target (P13K, PDB: 5JHB), highlighting the residues within the binding site pocket (site B1).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b and c show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs, and light blue-dashed segments denote aromatic HBs.




a b c

**Figure 4.6.5. Molecular insight into compound M6*-ζ docked on the anticancer target (P13K, PDB: 5JHB), highlighting the residues within the binding site pocket (site B2).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b and c show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.



© University of Venda

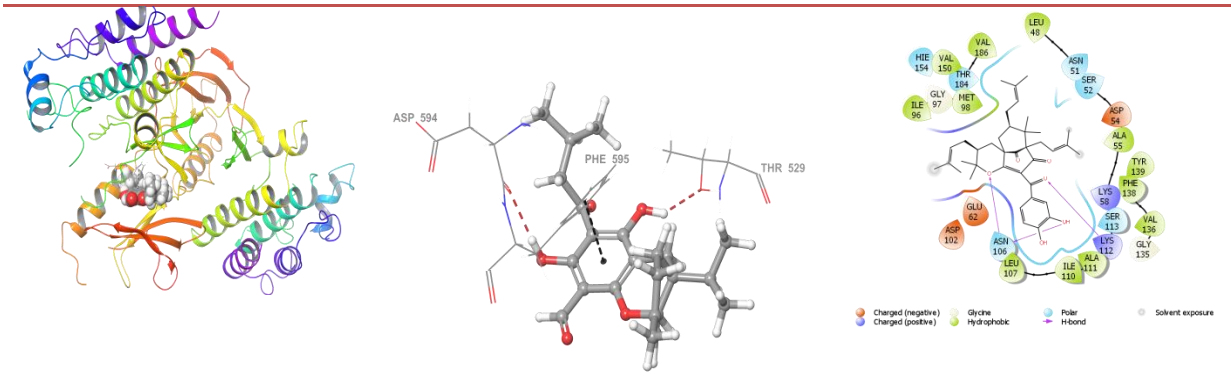
#### 4.6.1.5. Molecular interaction of considered compounds with BRAF V600B

In the ligand-BRAF V600B complex, the ligands show better binding scores at site C1 and C2 than at site C3.

At site C1, M8-ε achieves the highest docking score with a binding score energy of -10.266 kcal $mol^{-1}$. This interaction is not significantly stronger than that of the co-crystal ligand, it involves two HBs with ASP594 and THR529, π-π stacking with PHE595, and hydrophobic interactions with PHE583, CYS532, TRP531, ILE527, LEU514, LEU505, PHE595, LEU597, VAL471, and ILE463 (Figure 4.6.6).

For site C2, M7-ε emerges as the best-docked ligand with a binding score energy of -10.270 kcal $mol^{-1}$. Similarly, this interaction is not notably stronger than the co-crystal ligand. However, it encompasses a HB with CYS532, an aromatic HB with PHE595, and hydrophobic interactions with PHE583, PHE595, VAL471, VAL482, ALA481, ILE527, VAL528, TRP531, CYS532, and ILE463 (Figure 4.6.7).

At site C3, N8-τ exhibits the best docking score with a binding score energy of -10.270 kcal $mol^{-1}$. This interaction is comparatively stronger than the co-crystal ligand and involves three HBs with LEU577 and SER637, along with hydrophobic interactions with ILE666, TYR656, PRO655, and ILE644 (Figure 4.6.8).

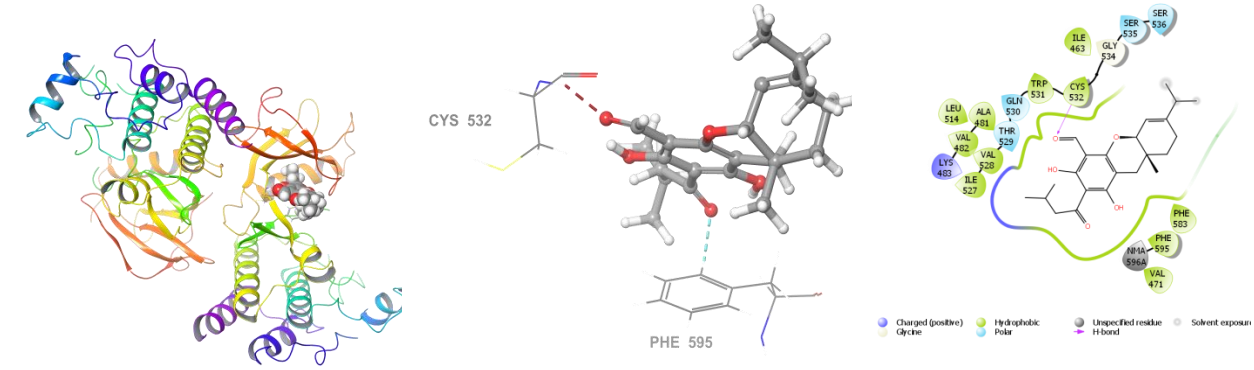



a b c



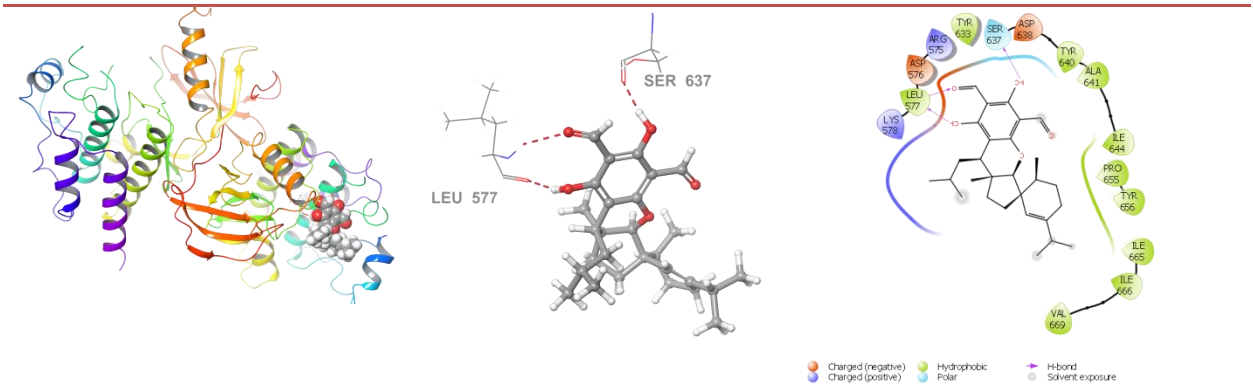
© University of Venda

**Figure 4.6.6. Molecular insight into compound M8-ε docked on the anticancer target (BRAF V600b, PDB; 6V34), highlighting the residues within the binding site pocket (site C1).** Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, and black-dashed segments denote π-π stacking interactions.




a b c

**Figure 4.6.7. Molecular insight into compound M7-ε docked on the anticancer target (BRAF V600b, PDB; 6V34), highlighting the residues within the binding site pocket (site C2).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs, and light blue-dashed segments denote aromatic HBs.




a b c

**Figure 4.6.8. Molecular insight into compound N8-τ docked on the anticancer target (BRAF V600b, PDB; 6V34), highlighting the residues within the binding site pocket (site C3).** Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

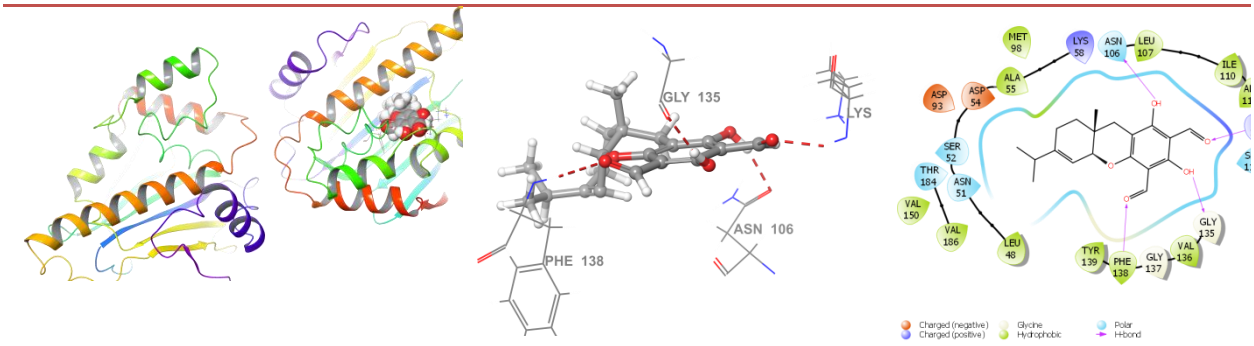


© University of Venda

### 4.6.1.6. Molecular interaction of considered compounds with H5P90

The molecular docking results show that most of the considered molecules bind to both identified binding sites. M7*-τ demonstrates the strongest docking scores in both active sites, with binding score energies of -8.134 kcal $mol^{-1}$ and -8.055 kcal $mol^{-1}$ for the G1 and G2 sites, respectively.

At active site G1, the interaction between M7*-τ and H5P90 is closely similar to that of the co-crystal ligand. M7*-τ forms four HBs with PHE138, GLY135, LY112, and ASN106, and engages in polar interactions with SER113, ASN106, SER52, and ASN51 (Figure 4.6.9). Similarly, at active site G2, M7*-τ displays similar interaction patterns observed in active site G1.




a b c

**Figure 4.6.9. Molecular insight into compound M7*-τ docked on the breast cancer target (H5P90, PDB: 3TUH), highlighting the residues within the binding site pocket (site G1).** Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

### 4.6.1.7. Molecular interaction of considered compounds with HER2

A considerable number of the considered molecules exhibit binding affinity with the H1 and H5 active sites compared to the remaining sites, while none bind to the H6 site.

The best-docked complexes from each active site are as follows: M8*-τ–HER2 (H1), M7-ζ–HER2 (H2), N4-τ–HER2 (H3), M10*-τ–HER2 (H4), and N4-τ–HER2 (H5), with binding score energies of -8.021, -4.724, -5.580, -6.812, and -7.889, respectively.



© University of Venda

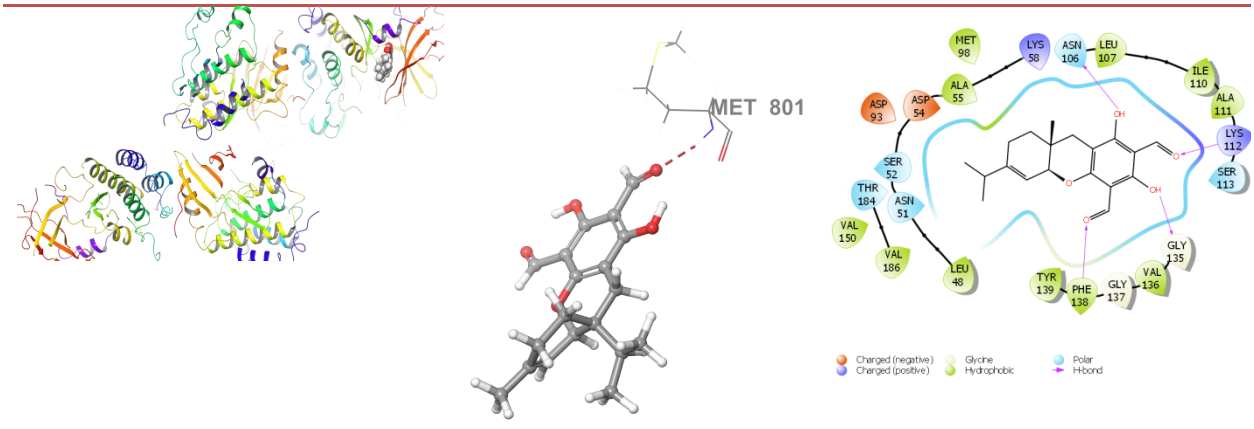
At active site H1, the interaction between M8*-τ and HER2 is not notably stronger than that of the co-crystal ligand. M8*-τ forms one HB with MET801 and engages in hydrophobic interactions with CYS805, MET801, LEU800, and ALA751 (Figure 4.6.10).

At active site H2, the interaction between M7-ζ and HER2 is also not significantly stronger than the co-crystal ligand. M7-ζ forms one HB with TYR1005 and two aromatic HBs with TYR1005 and TYR803 (Figure 4.6.11).

At active site H3, the interaction between N4-τ and HER2 is comparatively stronger than the co-crystal ligand. N4-τ forms one HB with MET801 and engages in hydrophobic interactions with CYS805, MET801, LEU800, and ALA751 (Figure 4.6.12).

At active site H4, the interaction between M10*-τ and HER2 is relatively similar to that of the co-crystal ligand. M10*-τ forms only one HB with MET801 and engages in hydrophobic interactions with CYS805, MET801, LEU800, ALA751, and LEU852 (Figure 4.6.13).

At active site H5, the interaction between N4-τ and HER2 is not significantly stronger than the co-crystal ligand. N4-τ forms only one hydrogen bond with MET801 and engages in hydrophobic interactions with CYS805, MET801, LEU800, and ALA751 (Figure 4.6.14).



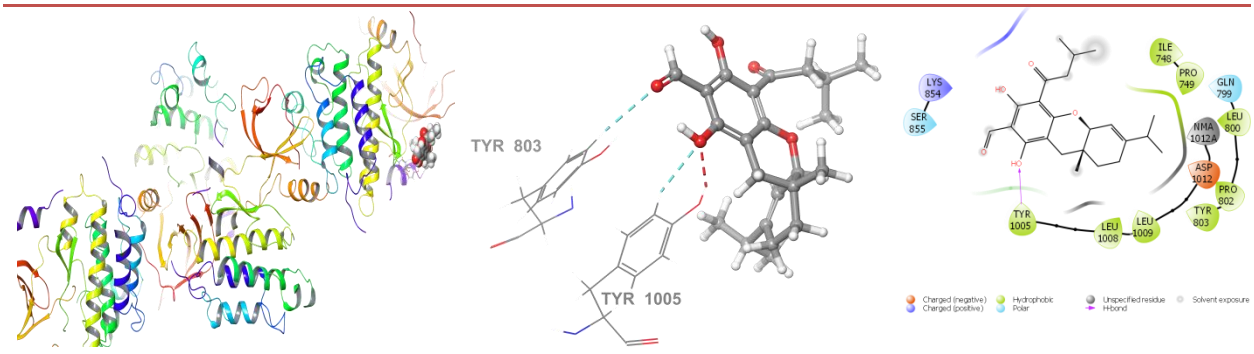

a b c

**Figure 4.6.10. Molecular insight into compound M8*-τ docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H1).** Panel a) showcases the



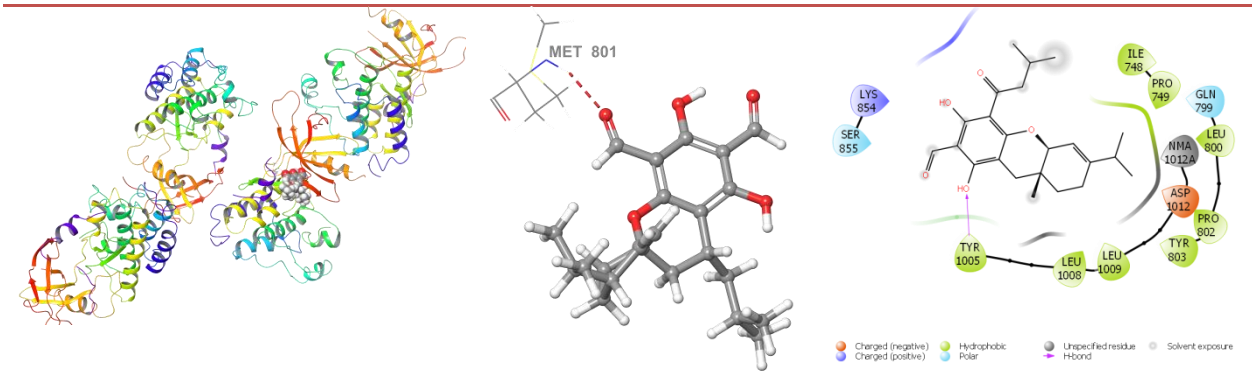
© University of Venda

docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.




a b c

**Figure 4.6.11. Molecular insight into compound M7-ζ docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H2).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs, and light blue-dashed segments denote aromatic HBs.

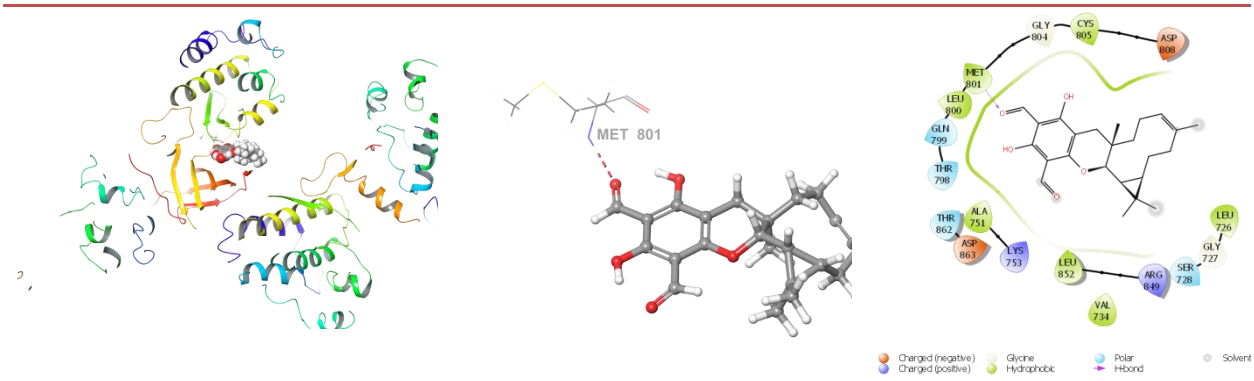



a b c

**Figure 4.6.12. Molecular insight into compound N4-τ docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H3).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.

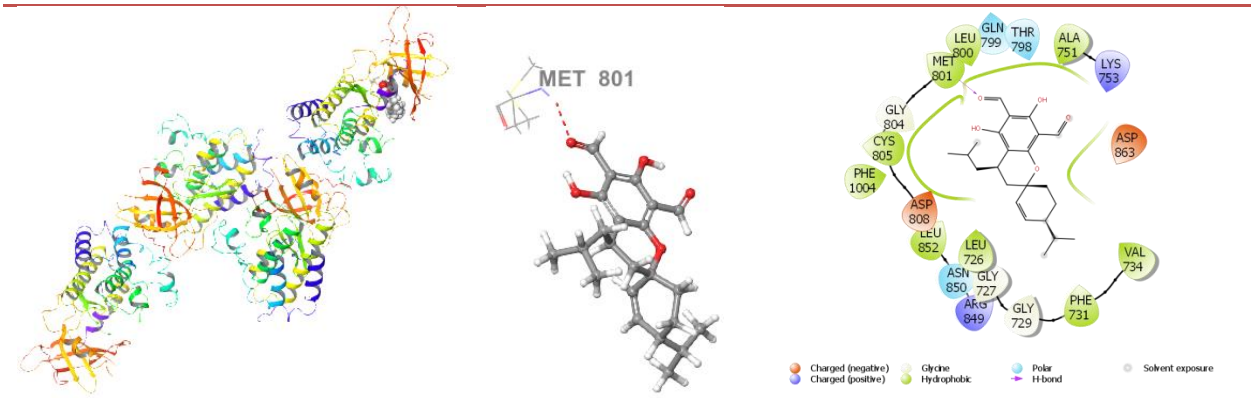


© University of Venda




a b c

**Figure 4.6.13. Molecular insight into compound M10*-τ docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H4).** Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.




a b c

**Figure 4.6.14. Molecular insight into compound N4-τ docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H5).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.

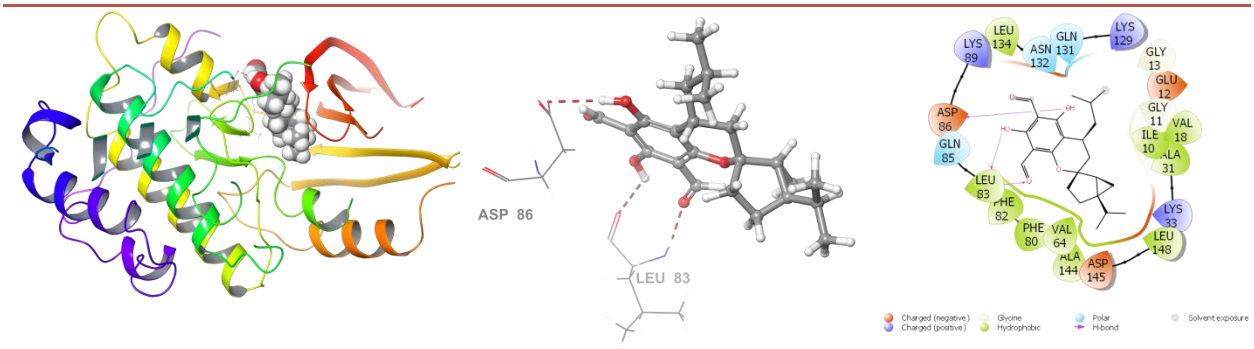
### 4.6.1.8. Molecular interaction of considered compounds with CDK-2

The molecular docking analysis of CDK-2 with the selected compounds shows that most of the considered molecules are capable of interacting at the active site. Among them, N3-τ emerges as the best-docked compound against CDK-2, showing a binding score energy of -9.687 kcal $mol^{-1}$. Its interaction with CDK-2 is relatively similar to that of the co-crystal ligand, forming three HBs



© University of Venda

with LEU83 and ASP86, and engaging in hydrophobic interactions with LEU83, PHE82, PHE80, and ALA144 (Figure 4.6.15).




a b c

**Figure 4.6.15. Molecular insight into compound N3-τ docked on the anticancer target (CDK-2, PDB: 1DI8), highlighting the residues within the binding site pocket.** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.

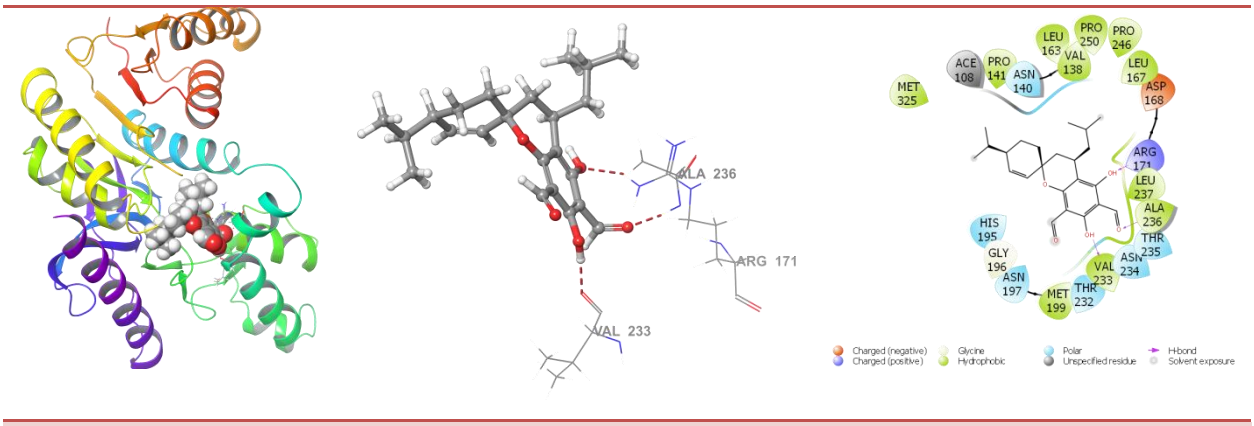
### 4.6.2. Molecular docking result of selected antimalarial-targets against the considered compounds.

#### 4.6.2.1. Molecular interaction of considered compounds with PFLDH

The molecular docking analysis of PFLDH with the selected compounds shows that nearly all the considered molecules can interact at the active site. N5-τ emerges as the top-performing compound against PFLDH, showing a binding score energy of -7.919 kcal $mol^{-1}$. Its interaction with PFLDH is notably stronger than that of the co-crystal ligand. Specifically, N5-τ forms three HBs with VAL233, ALA236, and ARG171 and engages in hydrophobic interactions with ALA236, LEU237, and VAL138 (Figure 4.6.16).



© University of Venda




a b c

**Figure 4.6.16. Molecular insight into compound N5-τ docked on the antimalarial target (PFLDH, PDB: 1U5A), highlighting the residues within the binding site pocket.** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.

### 4.6.2.2. Molecular interaction of considered compounds with PFMDH

The molecular docking results show that only a few molecules bind to the M1, M2, M4, and M5 active sites, while most bind to the M3 active site. The best-docked complexes from each active site are as follows: M5-ε/PFMDH (M1), M1*-τ/PFMDH (M2), M4*-τ/PFMDH (M3), M5-ζ/PFMDH (M4), and M7-ε/PFMDH (M5), with binding score energies of -4.519, -6.521, -5.737, -5.284, and -5.140, respectively.

At active site M1, the interaction between M5-ε and PFMDH is not significantly stronger than that of the co-crystal ligand. M5-ε forms three HBs with VAL233, ALA236, and ARG171 and engages in hydrophobic interactions with ALA236, LEU237, and VAL138 (Figure 4.6.17).

At active site M2, the interaction between M1*-τ and PFMDH is also not notably stronger than the co-crystal ligand. M1*-τ forms only one HB with PRO D:281 and engages in hydrophobic interactions with VAL D:282, PHE D:284, LEU D:250, PHE B:195, VAL B:187, and VAL B:161 (Figure 4.6.18).

At active site M3, the interaction between M4*-τ and PFMDH is not significantly stronger than the co-crystal ligand. M4*-τ forms three HBs with GLY13, ILE12, and GLY78 and engages in

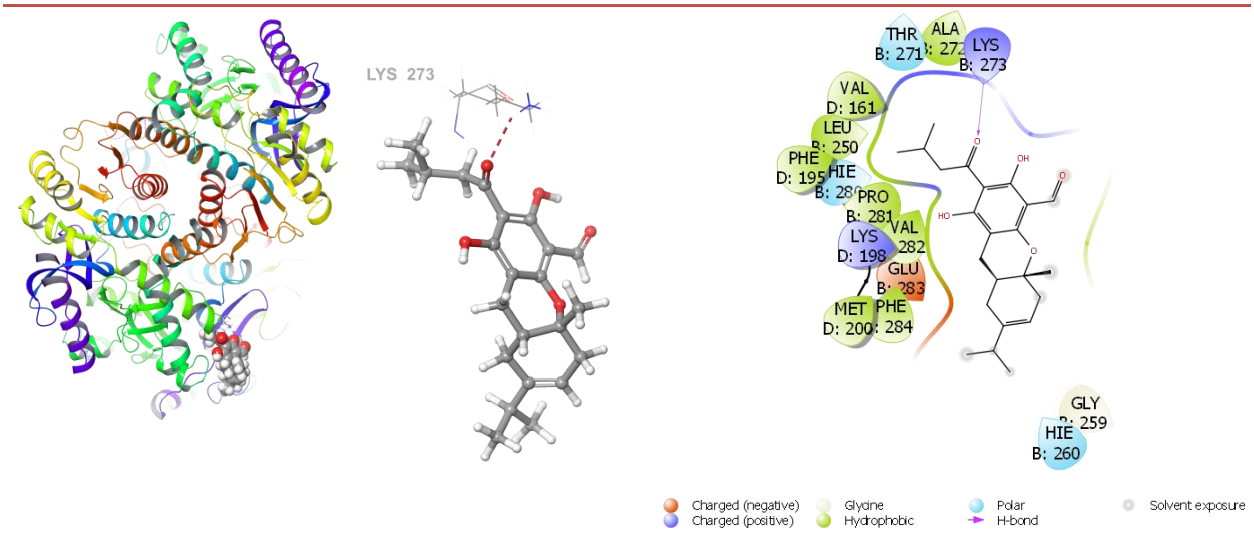


© University of Venda

hydrophobic interactions with ILE12, along with a polar interaction with THR76 and GLN80 (Figure 4.6.19).

At active site M4, the interaction between M5-ζ and PFMDH is also not significantly stronger than the co-crystal ligand. M5-ζ forms only one HB with LEU A:159 and engages in hydrophobic interactions with ALA C:272 and PRO C:281 (Figure 4.6.20).

At active site M5, the interaction between M7-ε and PFMDH is also not significantly stronger than the co-crystal ligand. M7-ε forms two HBs with LYS A:273 and PRO A:281 and engages in hydrophobic interactions with LEU C:159, VAL C:161, ALA A:272, VAL C:187, and PRO 281 (Figure 4.6.21).

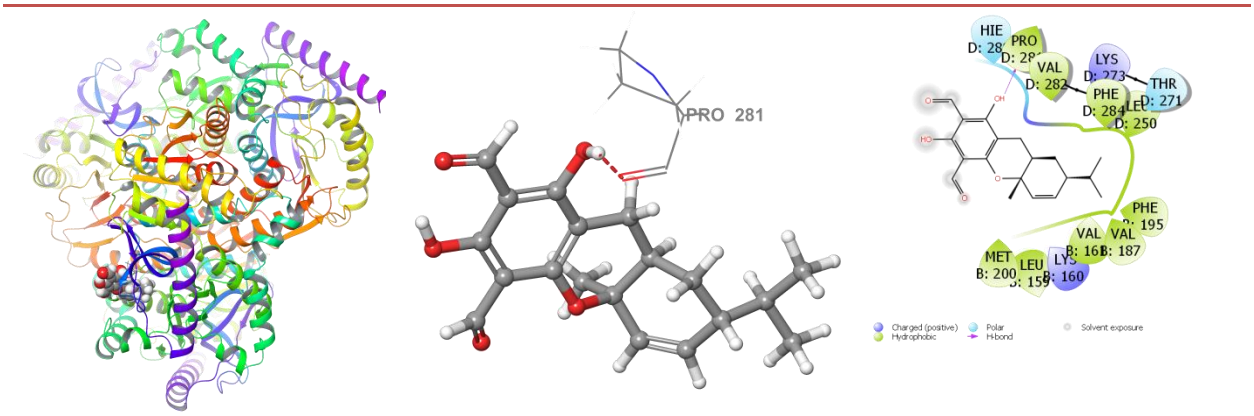



a b c

**Figure 4.6.17. Molecular insight into compound M5-ε docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M1).** Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

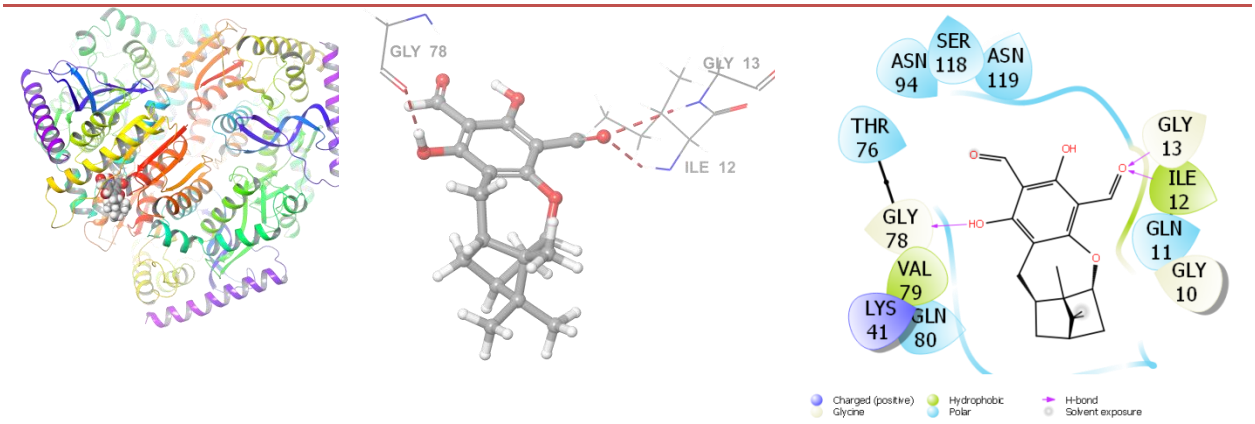


© University of Venda




a b c

**Figure 4.6.18. Molecular insight into compound M1*-τ docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M2).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.

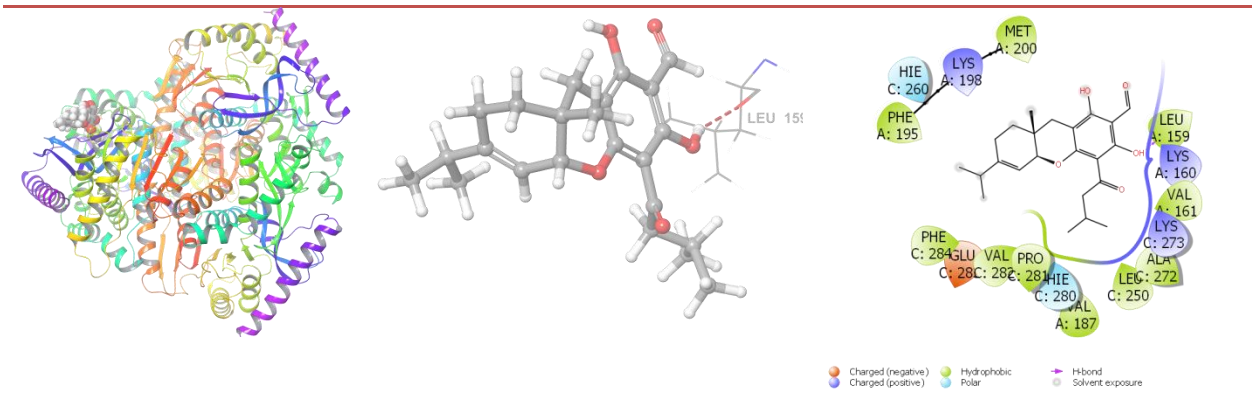



a b c

**Figure 4.6.19. Molecular insight into compound M4*-τ docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M3).** Panel a) showcases the docking pose of the protein-ligand complex. Panels b) and c) show 3D and 2D diagrams illustrating the interactions between the ligand and the active site of the target. In panel b, red-dashed segments denote HBs.

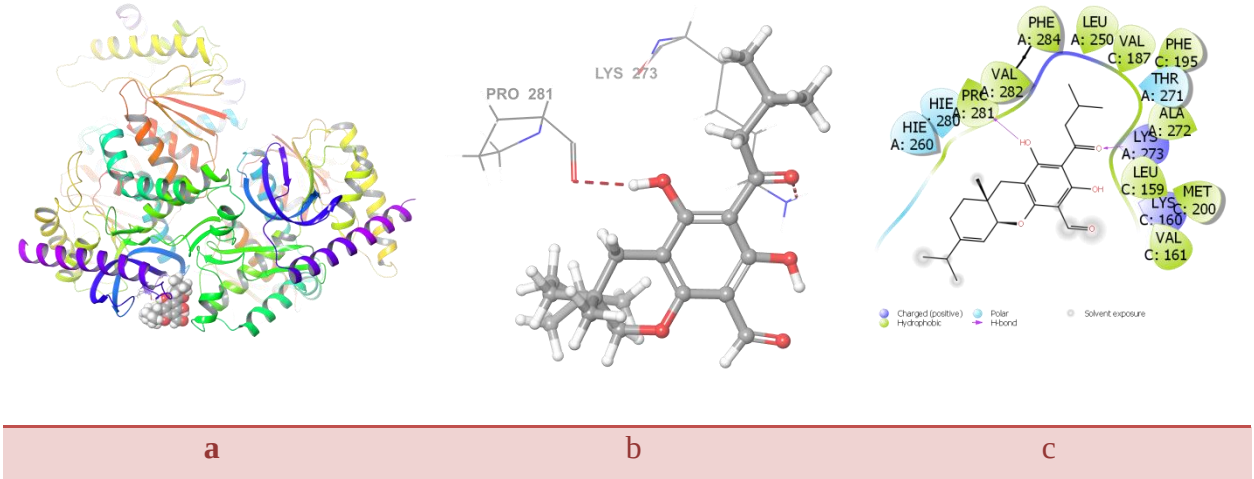


© University of Venda




a b c

**Figure 4.6.20. Molecular insight into compound M5-ζ docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M4).** Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.




a b c

**Figure 4.6.21. Molecular insight into compound M7-ε docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M5).** Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

## 4.7. Analysis of physicochemical properties

In this study, an analysis of the physicochemical properties and ADMET properties of 38 compounds was performed to assess the rate of drug attrition in later stages of drug development. The results show that out of the 38 compounds examined, five did not meet Lipinski's rule of five (Table S4.64). Lipinski's rule of five is a widely used guideline in pharmaceutical science that



© University of Venda

evaluates the drug-likeness and oral bioavailability of compounds in humans based on specific criteria. These criteria include a molecular mass (denoted as MW from the old-type molecular weight term) of less than 500 a.m.u., a partition coefficient (logP) of less than 5, a maximum of 5 hydrogen bond donor (HBD), and a maximum of 10 hydrogen bond acceptor (HBA). Among the compounds studied, M10-ε, M10*-ζ, N1-τ, N6-τ, and N8-τ violated Lipinski's rule of five by having higher logP values, indicating high lipophilicity and potentially poor absorption rates. While this may be desirable in certain cases, it also increases the risk of toxicity and other adverse effects.

In order to minimize the risk of failures during clinical phases, it is crucial to thoroughly analyze the pharmacokinetic properties of drug-like compounds, including Absorption, Distribution, Metabolism, and Excretion (ADME), in humans. This can be achieved by examining certain physicochemical properties such as Caco-2, MDCK cell permeability, log Khsa for human serum albumin binding, QP log P o/w, QP log BB, and the percentage of human oral absorption.

When assessing the permeability of drugs across cell membranes, higher values for Caco-2 and MDCK cell permeability are generally preferred as they indicate better absorption into the bloodstream. In this study, the considered compounds displayed a range of Caco-2 and MDCK cell permeability values, ranging from 119.560 to 498.516 and 49.812 to 233.119, respectively. These values were considered satisfactory. Notably, compounds containing an AC-isb exhibited higher Caco-2 and MDCK cell permeability values compared to those without. Furthermore, compounds with AC-isb at C5 demonstrated higher Caco-2 and MDCK cell permeability values in comparison to compounds with AC-isb at C1.

The blood-brain barrier (BBB) is a crucial barrier that separates the brain from the rest of the body. When it comes to drugs that target the central nervous system (CNS), it is important for them to have good penetration through the BBB. On the other hand, drugs that target peripheral organs should have low BBB penetration in order to minimize any potential side effects on the CNS. In this study, the QP log BB results showed that all of the selected compounds had low absorption into the CNS, with BBB values ranging from -1.067 to -1.748. The compound with the lowest absorption, specifically with R=AC-isb, demonstrated the least ability to penetrate the BBB.



© University of Venda

In order to anticipate the binding capacity of compounds to human serum albumin (HSA), the Qikprop parameter QP log Khsa (-1.5 to 1.5) is used. A higher log Khsa value signifies a greater affinity for HSA, which can significantly influence the drug's pharmacokinetics and distribution within the body. The calculated values for these compounds were considered acceptable, falling within the range of 0.125 to 1.379.

When it comes to the percentage of human oral absorption, higher values are typically preferred because they indicate that a larger amount of the drug is entering the bloodstream after being taken orally. This is particularly desirable for oral medications. However, it's important to note that the acceptable values for this parameter can vary significantly depending on factors such as the specific drug being used, the disease being treated, and the intended dosing schedule. In this case, the compounds in question demonstrate a remarkable percentage of human oral absorption, exceeding 78.301%.

The compound under consideration exhibits favorable physicochemical properties as well as ADMET properties. It is possible to enhance these physicochemical properties through structural modifications using the structure-activity relationship approach. It is evident that the presence and placement of the AC-isb greatly influence both the physicochemical properties and ADMET properties.

### 4.8. Concluding remarks

The study’s results show that the position of two acyl groups (that is, AC-isb and AC-ald) in euglobals; either one at C1 and the other one at C5, or vice versa, has a significant effects on several molecular properties, including their magnitudes and patterns. Energy profiles of low-energy conformers that had both H17⋯O14 and H16⋯O18 are quite closely matched. The characteristics of the O–H⋯O IHBs, and other properties that have these bonds are strongly influenced by the length of R, like all other ACPs with longer length and smaller red shift when R = H. The QSAR descriptors also include other properties that may be dependent on the relative position of both acyl groups, e.g., dipole moment. The effects could originate from the proximity of the terpene moiety to the C1-attached acyl group.



© University of Venda

The presence of different IHBs has an influence on the properties of an IHB; the presence of other O–H···O IHBs for any one of the O–H···O IHBs, the presence of one or another O–H···O IHB for a C–H···O IHB character. This implies that there are interrelations between the HBs which could be explained only after further detailed investigation.

When assessing the cooperativity of C7–H13···O12 (ε-isomers) due to H17···O18 or C11–H19···O12 (ζ-isomers) by H17···O14, the pattern of cooperativity can be similar to the well-established cooperativity between O–H···O and C–H···O H-bonds; it is known for consecutive O–H···O H-bonds (e.g., [98] among others). Hence, extension from O–H···O and C–H···O HBs must be acceptable. Other interactions are observed for the non-consecutive IHBs, which have an influence on a much greater number of things that should also be studied separately at some general level not being restricted to a specific class.

When AC-ald and AC-isb have their positions swapped, there is no significant influence on the HOMO-LUMO gap. Thus, the global reactivity descriptors also show that the ε- and ζ-isomers are similar in terms of reactivity. Three levels of theory used for calculations outline similar patterns concerning effects on investigated molecular characteristics when changing the positions of two acyl groups with respect to each other. The reliability of conclusions drawn from these results is increased by the different natures of the methods used, ab initio and density functional theory. It is also well known that the estimation of the HOMO-LUMO energy gap differs between DFT and ab initio methods without the need for any specific analysis. On one hand, HF does not exhibit blue shifts for the ε- and ζ-isomers while recognizing them for grandinol and τ model structures in spite of strong steric repulsion; this might best suit a particular research effort focused on how different calculation methods discriminate among blue shifts.

We noted the exceptions in the Mulliken charge of the O8 atom, which should be verified using an alternative approach to determine charges, such as natural bond orbitals (NBO, [80]). Conversely, in this study, specific atomic charges do not fall under a high priority category. It gives more prominence to features that can help correlate with the biological activity of molecules.



© University of Venda

Reporting and analysis of them were mainly for their indirect contribution to investigating IHBs in these molecules.

The data obtained show that all the EGs are structural isomers differing by terpene moiety, except M10. One of the descriptors for QSAR is the molecular mass. Therefore, it could be interesting to know whether the phenomenon that they have a common molecular mass may play a role in some aspects of their anticancer activity. However, compound M10 with a considerably higher molecular mass exhibits analogous patterns of H-bonding abilities and HOMO and LUMO distributions as other compounds considered, which implies that fundamental structural characteristics (presence and nature of two moieties) have a dominant role for their activities.

The presence of hydroxyl groups and an $sp^2$ oxygen atom plays a crucial role in stabilizing ligand-receptor complexes through intermolecular HBs. Additionally, the presence of aromatic rings contributes significantly to interactions via aromatic H-bonds, pi-cation, and π-π stacking interactions. These findings indicate that molecules demonstrating strong interactions with the target protein are promising candidates for further analysis of their action mechanisms.

Moreover, the presence of the AC-isb group and the reversal of the positions of AC-ald and AC-isb exert notable influences on physicochemical properties and ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties. The inclusion of the AC-isb group leads to increased values of QPlogPo/w, QPlogKhsa, Caco-2, and MDCK cell permeability, Percent-Human-Oral-Absorption, and QPlogBB. Furthermore, these properties exhibit greater values when the AC-isb group is positioned at C5 compared to C1.

This knowledge is intended to serve as a foundation for further studies, particularly in investigating the potential roles of these natural compounds within the medicinal field.



© University of Venda

## 4.9. References

1. Tshilande, N., & Mammino, L. (2023). Comparison of the molecular properties of euglobals differing by the mutual positions of the two R–C=O groups (R=H and $CH_2CH(CH_3)_2$): A computational study. *Chemistry*, *5*(4), 2120-2154.
2. Dean, J. A., & Liu, D. (2007). *Mechanisms of molecular activity*. John Wiley & Sons.
3. Mammino, L., & Kabanda, M. M. (2007). Model structures for the study of acylated phloroglucinols and computational study of the caespitate molecule. *Journal of Molecular Structure: THEOCHEM*, *805*(1-3), 39-52.
4. Mammino, L., & Kabanda, M. M. (2009). A study of the intramolecular hydrogen bond in acylphloroglucinols. *Journal of Molecular Structure: THEOCHEM*, *901*(1-3), 210-219.
5. Kabanda, M. M., & Mammino, L. (2012). The conformational preferences of acylphloroglucinols—a promising class of biologically active compounds. *International Journal of Quantum Chemistry*, *112*(23), 3691-3702.
6. Oprea'i, T. I., & Waller'r, C. L. (2009). Structure-activity relationships. *Reviews in Computational Chemistry*, *11*, 127.
7. Patle, S., & Saini, S. (2016). Hydrophobic effect on drug stability and oral bioavailability. *Journal of Pharmaceutical Sciences*, *105*, 4037-4049.
8. Cremer, D. (2011). Møller–Plesset perturbation theory: From small molecule methods to methods for thousands of atoms. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *1*(4), 509-530.
9. Mammino, L., & Kabanda, M. M. (2009). A computational study of the effects of different solvents on the characteristics of the intramolecular hydrogen bond in acylphloroglucinols. *The Journal of Physical Chemistry A*, *113*(52), 15064-15077.
10. Mammino, L., & Kabanda, M. M. (2012). Computational study of the patterns of weaker intramolecular hydrogen bonds stabilizing acylphloroglucinols. *International Journal of Quantum Chemistry*, *112*(14), 2650-2658.
11. Mammino, L., & Kabanda, M. M. (2013). The role of additional O–H… O intramolecular hydrogen bonds for acylphloroglucinols' conformational preferences *in vacuo* and in solution. *Molecular Simulation*, *39*(1), 1-13.



© University of Venda

12. Polo, V., Kraka, E., & Cremer, D. (2002). Some thoughts about the stability and reliability of commonly used exchange–correlation functionals–coverage of dynamic and nondynamic correlation effects. *Theoretical Chemistry Accounts, 107*, 291-303.
13. Chiodo, S. G., Leopoldini, M., Russo, N., & Toscano, M. (2010). The inactivation of lipid peroxide radical by quercetin: A theoretical insight. *Physical Chemistry Chemical Physics, 12*(27), 7662-7670.
14. Baker, J., Autschbach, J., & Zheng, J. (2010). Performance of DFT methods for molecular properties of polarizable molecules. *Journal of Chemical Theory and Computation, 7*(1), 26-35.
15. Merrick, J. P., Moran, D., & Radom, L. (2007). An evaluation of harmonic vibrational frequency scale factors. *The Journal of Physical Chemistry A, 111*(45), 11683-11700.
16. Gajiwala, K. S., Feng, J., Ferre, R., Ryan, K., Brodsky, O., Weinrich, S., Kath, J. C., & Stewart, A. (2013). Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. *Structure, 21*(2), 209-219.
17. Sogabe, S., Kawakita, Y., Igaki, S., Iwata, H., Miki, H., Cary, D. R., Takagi, T., Takagi, S., Ohta, Y., & Ishikawa, T. (2013). Structure-based approach for the discovery of pyrrolo [3, 2-d] pyrimidine-based EGFR T790M/L858R mutant inhibitors. *ACS Medicinal Chemistry Letters, 4*(2), 201-205.
18. Stamos, J., Sliwkowski, M. X., & Eigenbrot, C. (2002). Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *Journal of Biological Chemistry, 277*(48), 46265-46272.
19. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein−ligand complexes. *Journal of Medicinal Chemistry, 49*(21), 6177-6196.
20. Repasky, M. P., Shelley, M., & Friesner, R. A. (2007). Flexible ligand docking with Glide. *Current Protocols in Bioinformatics, 18*(1), 8-12.
21. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., & Shaw, D. E. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry, 47*(7), 1739-1749.



© University of Venda

22. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry, 47*(7), 1750-1759.
23. Jorgensen, W. L., & Duffy, E. M. (2002). Prediction of drug solubility from structure. *Advanced Drug Delivery Reviews, 54*(3), 355-366.
24. Lu, J. J., Crimin, K., Goodwin, J. T., Crivori, P., Orrenius, C., Xing, L., Tandler, P. J., Vidmar, T. J., Amore, B. M., Wilson, A. G., & Stouten, P. F. (2004). Influence of molecular flexibility and polar surface area metrics on oral bioavailability in the rat. *Journal of Medicinal Chemistry, 47*(24), 6104-6107.
25. Kozuka, M., Sawada, T., Kasahara, F., Mizuta, E., Amano, T., Komiya, T., & Goto, M. (1982). The granulation-inhibiting principles from *Eucalyptus globulus* Labill. II. The structures of euglobal-Ia1, -Ia2, -Ib, -Ic, -IIa, -IIb, and -IIc. *Chemical and Pharmaceutical Bulletin, 30*(6), 1952-1963.
26. Singh, I. P., Sidana, J., Bharate, S. B., & Foley, W. J. (2010). Phloroglucinol compounds of natural origin: Synthetic aspects. *Natural Product Reports, 27*(3), 393-416.
27. Coppen, J. J. W. (Ed.). (n.d.). *Eucalyptus: The genus Eucalyptus*. Available online: https://books.google.co.za/books?id=0dRlDMvlhQ0C&pg=PA286 (Accessed; November 19, 2021).
28. Begleiter, A., Lin, D., Larson, K. K., Lang, J., Wu, X., Cabral, T., Taylor, H., Guziec, L. J., Kerr, P. D., Hasinoff, B. B., & Guziec Jr, F. S. (2006). Structure-activity studies with cytotoxic anthrapyrazoles. *Oncology Reports, 15*(6), 1575-1580.
29. Perlovich, G. L., Volkova, T. V., Manin, A. N., & Bauer-Brandl, A. (2008). Influence of position and size of substituents on the mechanism of partitioning: A thermodynamic study on acetaminophens, hydroxybenzoic acids, and parabens. *AAPS PharmSciTech, 9*(1), 205-216.
30. Wiley, J. L., Marusich, J. A., & Huffman, J. W. (2014). Moving around the molecule: Relationship between chemical structure and in vivo activity of synthetic cannabinoids. *Life Sciences, 97*(1), 55-63.
31. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery Jr, J. A., Vreven, T. K. K. N., Kudin, K. N., Burant, J. C., & Millam, J. M. (2013). *Gaussian 03, revision C.02*. Gaussian, Inc.
32. Dennington, R. D. I. I., Keith, T., & Millam, J. G. (2007). *Version 4.1*. Semichem, Inc.



© University of Venda

33. Mammino, L. (2014). A computational study of euglobal G1—an acylphloroglucinol with anticancer activity. *Current Bioactive Compounds, 10*(3), 163-180.
34. Mammino, L. (2014). A computational study of euglobal G4. In *Proceedings of the 10th Theoretical Chemistry Conference in Africa (TCCA) and the Eastern Southern Africa Environmental Chemistry Conference (ESAECC)* (pp. 29-55).
35. Spoliti, M., Bencivenni, L., Quirante, J. J., & Ramondo, F. (1997). Molecular conformations and harmonic force field of 1,3,5-benzenetriol molecule from ab initio and density functional theory investigations. *Journal of Molecular Structure: THEOCHEM, 390*(1-3), 139-148.
36. Gilli, G., Bellucci, F., Ferretti, V., & Bertolasi, V. (1989). Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the β-diketone fragment. *Journal of the American Chemical Society, 111*(3), 1023-1028.
37. Bertolasi, V., Gilli, P., Ferretti, V., & Gilli, G. (1991). Evidence for resonance-assisted hydrogen bonding. 2. Intercorrelation between crystal structure and spectroscopic parameters in eight intramolecularly hydrogen bonded 1,3-diaryl-1,3-propanedione enols. *Journal of the American Chemical Society, 113*(13), 4917-4925.
38. Guevara-Vela, J. M., Gallegos, M., Valentín-Rodríguez, M. A., Costales, A., Rocha-Rinza, T., & Pendás, Á. M. (2021). On the relationship between hydrogen bond strength and the formation energy in resonance-assisted hydrogen bonds. *Molecules, 26*(14), 4196.
39. Grosch, A. A., van der Lubbe, S. C., & Fonseca Guerra, C. (2018). Nature of intramolecular resonance-assisted hydrogen bonding in malonaldehyde and its saturated analogue. *The Journal of Physical Chemistry A, 122*(6), 1813-1820.
40. Barnes, A. J. (2004). Blue-shifting hydrogen bonds—are they improper or proper? *Journal of Molecular Structure, 704*(1-3), 3-9.
41. Spoliti, M., Bencivenni, L., Quirante, J. J., & Ramondo, F. (1997). Molecular conformations and harmonic force field of 1,3,5-benzenetriol molecule from ab initio and density functional theory investigations. *Journal of Molecular Structure: THEOCHEM, 390*(1-3), 139-148.
42. Gilli, G., Bellucci, F., Ferretti, V., & Bertolasi, V. (1989). Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the β-diketone fragment. *Journal of the American Chemical Society, 111*(3), 1023-1028.
43. Bertolasi, V., Gilli, P., Ferretti, V., & Gilli, G. (1991). Evidence for resonance-assisted hydrogen bonding. 2. Intercorrelation between crystal structure and spectroscopic parameters



© University of Venda

in eight intramolecularly hydrogen bonded 1,3-diaryl-1,3-propanedione enols. *Journal of the American Chemical Society*, *113*(13), 4917-4925.

44. Joseph, J., & Jemmis, E. D. (2007). Red-, blue-, or no-shift in hydrogen bonds: A unified explanation. *Journal of the American Chemical Society*, *129*(15), 4620-4632.
45. Struble, M. D., Kelly, C., Siegler, M. A., & Lectka, T. (2014). Search for a strong, virtually "no-shift" hydrogen bond: A cage molecule with an exceptional OH···F interaction. *Angewandte Chemie*, *126*(34), 9070-9074.
46. Desiraju, G. R. (1996). The C−H···O hydrogen bond: Structural implications and supramolecular design. *Accounts of Chemical Research*, *29*(9), 441-449.
47. Gu, Y., Kar, T., & Scheiner, S. (1999). Fundamental properties of the CH···O interaction: Is it a true hydrogen bond? *Journal of the American Chemical Society*, *121*(40), 9411-9422.
48. An, N. T., Duong, N. T., Tri, N. N., & Trung, N. T. (2022). Role of OH···O/S conventional hydrogen bonds in considerable C-sp²-H blue-shift in the binary systems of acetaldehyde and thioacetaldehyde with substituted carboxylic and thiocarboxylic acids. *RSC Advances*, *12*(54), 35309-35319.
49. Cuc, N. T. T., Phan, C. T. D., Nhung, N. T. A., Nguyen, M. T., Trung, N. T., & Ngan, V. T. (2021). Theoretical aspects of nonconventional hydrogen bonds in the complexes of aldehydes and hydrogen chalcogenides. *The Journal of Physical Chemistry A*, *125*(48), 10291-10302.
50. Tabayashi, K., Chohda, M., Yamanaka, T., Tsutsumi, Y., Takahashi, O., Yoshida, H., & Taniguchi, M. (2010, June). Hydrogen bonding interaction of small acetaldehyde clusters studied with core-electron excitation spectroscopy in the oxygen K-edge region. In *Journal of Physics: Conference Series* (Vol. 235, No. 1, p. 012017). IOP Publishing.
51. Trung, N. T., Hung, N. P., Hue, T. T., & Nguyen, M. T. (2011). Existence of both blue-shifting hydrogen bond and Lewis acid–base interaction in the complexes of carbonyls and thiocarbonyls with carbon dioxide. *Physical Chemistry Chemical Physics*, *13*(31), 14033-14042.
52. Tang, Q., Huang, T., Huang, R., Cao, H., Wang, L., & Zheng, X. (2021). Theoretical researches on binding modes and stability of hydrogen bonds between uracil and formic acid.
53. Scheiner, S., Kar, T., & Gu, Y. (2001). Strength of the CαH··O hydrogen bond of amino acid residues. *Journal of Biological Chemistry*, *276*(13), 9832-9837.



© University of Venda

54. Scheiner, S. (2020). The hydrogen bond: A hundred years and counting. *Journal of the Indian Institute of Science, 100*(1), 61-76.
55. Chen, J. C., Li, J., Qian, L., & Zheng, K. C. (2005). Electronic structures and SARs of the isomeric complexes α-, β-, γ-[Ru(mazpy)$_2$Cl$_2$] with different antitumor activities. *Journal of Molecular Structure: THEOCHEM, 728*(1-3), 93-101.
56. Hossan, A., Alrefaei, A. F., Katouah, H. A., Bayazeed, A., Asghar, B. H., Shaaban, F., & El-Metwaly, N. M. (2023). Synthesis, anticancer activity, and molecular docking of new pyrazolo [1,5-a] pyrimidine derivatives. *Journal of Saudi Chemical Society, 27*(2), 101599.
57. Karrouchi, K., Brandán, S. A., Sert, Y., El-Marzouqi, H., Radi, S., Ferbinteanu, M., Faouzi, M. E. A., & Garcia, Y. (2020). Synthesis, X-ray structure, vibrational spectroscopy, DFT, biological evaluation, and molecular docking studies of (E)-N'-(4-(dimethylamino)benzylidene)-5-methyl-1H-pyrazole-3-carbohydrazide. *Journal of Molecular Structure, 1219*, 128541.
58. Rahmouni, N. T., el Houda Bensiradj, N., Megatli, S. A., Djebbar, S., & Baitich, O. B. (2019). New mixed amino acids complexes of iron (III) and zinc (II) with isonitrosoacetophenone: Synthesis, spectral characterization, DFT study, and anticancer activity. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *213*, 235-248.
59. Kumer, A., Sarker, M. N., & Sunanda, P. A. U. L. (2019). The theoretical investigation of HOMO, LUMO, thermophysical properties and QSAR study of some aromatic carboxylic acids using HyperChem programming. *International Journal of Chemistry and Technology*, *3*(1), 26-37.
60. Vijayaraj, R., Subramanian, V., & Chattaraj, P. K. (2009). Comparison of global reactivity descriptors calculated using various density functionals: A QSAR perspective. *Journal of Chemical Theory and Computation, 5*(10), 2744-2753.
61. Mebi, C. A. (2011). DFT study on structure, electronic properties, and reactivity of cis-isomers of [(NC$_5$H$_4$-S)$_2$Fe(CO)$_2$]. *Journal of Chemical Sciences, 123*, 727-731.
62. Domingo, L. R., Ríos-Gutiérrez, M., & Pérez, P. (2016). Applications of the conceptual density functional theory indices to organic chemistry reactivity. *Molecules, 21*(6), 748.
63. Tanış, E., Cankaya, N., & Yalçın, S. (2019). Synthesis, characterization, computation of global reactivity descriptors and antiproliferative activity of N-(4-nitrophenyl) acrylamide. *Russian Journal of Physical Chemistry B, 13*, 49-61.



© University of Venda

64. Miar, M., Shiroudi, A., Pourshamsian, K., Oliaey, A. R., & Hatamjafari, F. (2021). Theoretical investigations on the HOMO–LUMO gap and global reactivity descriptor studies, natural bond orbital, and nucleus-independent chemical shifts analyses of 3-phenylbenzo[d]thiazole-2(3H)-imine and its para-substituted derivatives: Solvent and substituent effects. *Journal of Chemical Research*, *45*(1-2), 147-158.
65. Kaya, S., & Kaya, C. (2015). A new method for calculation of molecular hardness: A theoretical study. *Computational and Theoretical Chemistry*, *1060*, 66-70.
66. Oláh, J., Van Alsenoy, C., & Sannigrahi, A. B. (2002). Condensed Fukui functions derived from stockholder charges: Assessment of their performance as local reactivity descriptors. *The Journal of Physical Chemistry A*, *106*(15), 3885-3890.
67. Vektariene, A., Vektaris, G., & Svoboda, J. (2009). A theoretical approach to the nucleophilic behavior of benzofused thieno[3,2-b]furans using DFT and HF based reactivity descriptors. *Arkivoc: Online Journal of Organic Chemistry*.
68. Said, M. S., & Najim, Z. A. (2012). Theoretical approach to relate the reactivity descriptors and Mulliken charges with carcinogenicity of some methylated benzo[a]anthracene. *Pakistan Journal of Analytical & Environmental Chemistry*, *13*(1), 8.
69. Babu, N. S., & Jayaprakash, D. (2019). Studies for reactivity indexes of cyanuric acid tautomers in different solvents by ab initio Hartree–Fock (HF) theory. *Journal of Advances in Chemistry*, *11*(8).
70. Mendoza Huizar, L. H., Rios-Reyes, C. H., Olvera-Maturano, N. J., Robles, J., & Rodriguez, J. A. (2014). Chemical reactivity of quinclorac employing the HSAB local principle-Fukui function. *Open Chemistry*, *13*(1), 000010151520150008.
71. Prasad, S., & Ojha, D. P. (2018). Geometric structures, vibrational spectroscopic and global reactivity descriptors of nematogens containing strong polar group—A comparative analysis using DFT, HF and MP2 methods. *Molecular Crystals and Liquid Crystals*, *666*(1), 12-28.
72. Das, A., Das, A., & Banik, B. K. (2021). Influence of dipole moments on the medicinal activities of diverse organic compounds. *Journal of the Indian Chemical Society*, *98*(2), 100005.
73. Bushelyev, S. N., & Stepanov, N. F. (1989). Elektronnaya Struktura y Biologhicheskaya Aktivnost Molecul. *Khimiya, Snanye*.



© University of Venda

74. Engineering Toolbox. (n.d.). *Liquid dielectric constants*. Retrieved from https://www.engineeringtoolbox.com/liquid-dielectric-constants-d_1263.html
75. Engineering Toolbox. (n.d.). *Relative permittivity*. Retrieved from https://www.engineeringtoolbox.com/relative-permittivity-d_1660.html
76. Goncalves, P. F., & Stassen, H. (2005). Free energy of solvation from molecular dynamics simulation applying Voronoi-Delaunay triangulation to the cavity creation. *The Journal of Chemical Physics, 123*(21).
77. Duarte Ramos Matos, G., Kyu, D. Y., Loeffler, H. H., Chodera, J. D., Shirts, M. R., & Mobley, D. L. (2017). Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database. *Journal of Chemical & Engineering Data, 62*(5), 1559-1569.
78. Matczak, P., & Domagała, M. (2017). Heteroatom and solvent effects on molecular properties of formaldehyde and thioformaldehyde symmetrically disubstituted with heterocyclic groups $C_4H_3Y$ (where Y = O–Po). *Journal of Molecular Modeling, 23*, 1-11.
79. Yoosefian, M., & Etminan, N. (2016). The role of solvent polarity in the electronic properties, stability and reactivity trend of a tryptophan/Pd doped SWCNT novel nanobiosensor from polar protic to non-polar solvents. *RSC Advances, 6*(69), 64818-64825.
80. Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. Molecular modeling of proteins, 365-382.



© University of Venda

University of Venda
Creating Future Leaders

# CHAPTER 5

# COMPUTATIONAL STUDIES OF SELECTED SIMPLE ACYLPHLOROGLUCINOLS

## 5.1. Introduction to the chapter

The selected compounds with anticancer or antimalarial properties have been introduced in Chapter 2, section 2.4.2. This chapter presents a computational study of selected simple ACPLs (with the acylphloroglucinol moiety containing only one acyl group) *in-vacuo*, in-solution, and through docking studies. Detailed results are presented in sections 5.4, 5.5, and 5.6, respectively. The primary objective of the studies *in vacuo* and in-solution was to provide an extensive understanding of the molecular properties of the selected molecules in different media. The primary objective of the docking studies is to analyze the protein-ligand interactions and to prioritize candidates with favorable pharmacokinetic and safety profiles. The computational methods selected for this chapter are the same as those outlined in Chapter 4, section 4.2.

## 5.2. Selection of structures

This study focuses on eight monomeric ACPLs (Fig.1), which are ACPLs containing a single acylphloroglucinol moiety, often referred to as "simple" ACPLs, with established antimalarial and/or anticancer activities. Table 5.1 lists these compounds, providing their common names, acronyms used throughout the study, and their proven biological activities. Their chemical structures are shown in Fig. 1. The atoms within the acylphloroglucinol moiety are numbered as depicted in Fig. 2, while the relevant atoms of individual substituents are numbered in Fig. 3. Only the atoms from OH groups and $sp^2$ O atoms are numbered, as they are involved in IHBs and other interactions. These atoms are numbered independently, with sufficiently high numbers to avoid overlap with the numbering in the acylphloroglucinol moieties. Table 5.2 provides a summary of the substituents, detailing the presence of OH groups and $sp^2$ O atoms, offering a clear overview of their number and positions, complementing the visual representation in Fig. 2. It is also noteworthy that Mono2 and Mono3 are structural isomers.



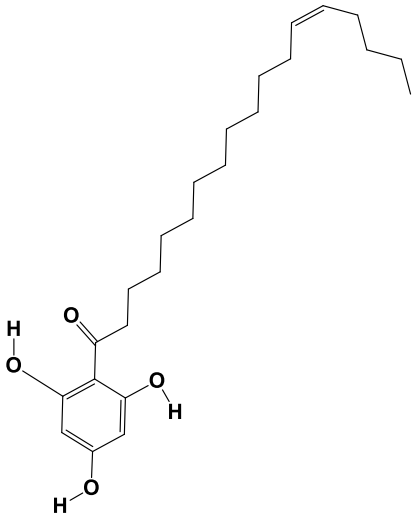
© University of Venda

**Table 5.1. List of calculated selected ACPL molecules and acronym denoting the name them.**

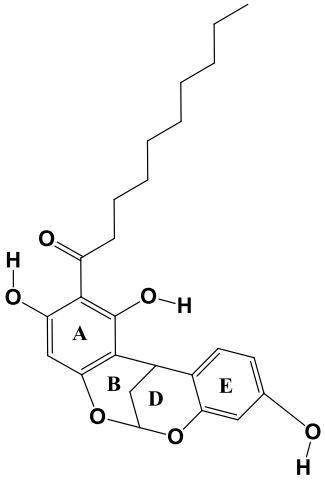
| Name of molecule (as reported in the literature) | Acronym denoting the name of the molecule |
|---|---|
| Thouvenol A | Mono1 |
| Myristicyclins A | Mono2 |
| Myristicyclins B | Mono3 |
| Knipholone | Mono4 |
| knipholone anthrone | Mono5 |
| 1-(2,6-dihydroxy-3-methyl-4-((3-methylbut-2-en-1-yl)oxy)phenyl)- 3-methylbutan-1-one | Mono6 |
| Antiarone J | Mono7 |
| Iriflophenone 4-glucoside | Mono8 |



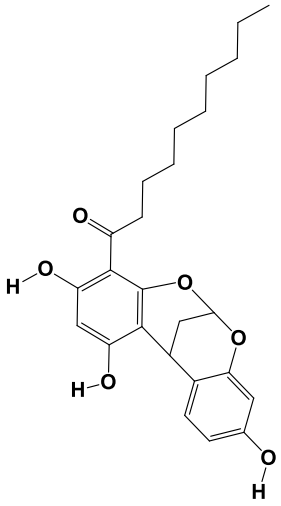
© University of Venda



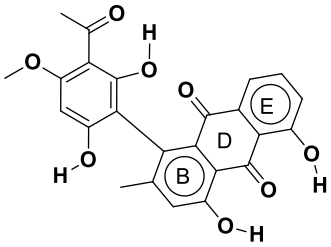
Mono1



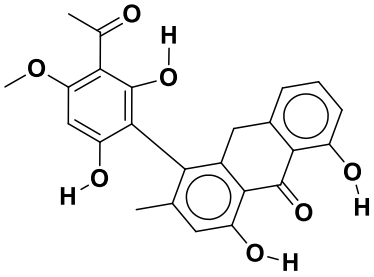

Mono2



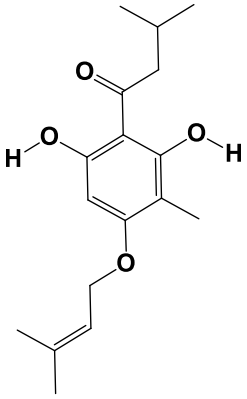
Mono3




Mono4



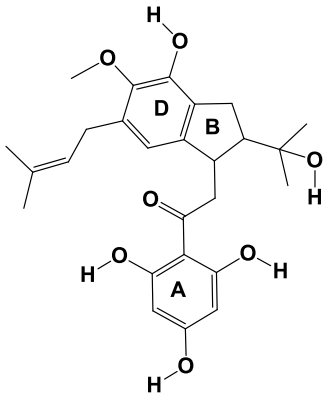
Mono5



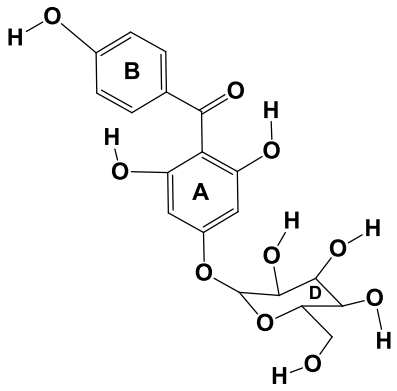
Mono6



© University of Venda




Mono7




Mono8

---

**Figure 5.1.** Molecular structures of the eight monomeric ACPL molecules considered in this study.
The atoms of the phloroglucinol moiety are numbered as shown in fig. 5.1. Only the atoms that are relevant for the analysis of the results are numbered individually in fig.3.
The rings are denoted with uppercase letters (A, B, D, and E), with the letter A denoting the ring of the phloroglucinol moiety.

---

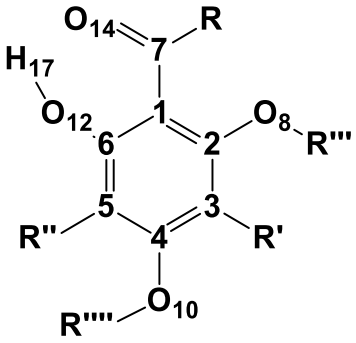



**Figure 5.2.** The general structure of acylphloroglucinols and atom numbering utilized in this work. The structure considers the possibility of substituents also at O8 and O10.



© University of Venda

**Group** **Substituents with numbering of relevant atoms**

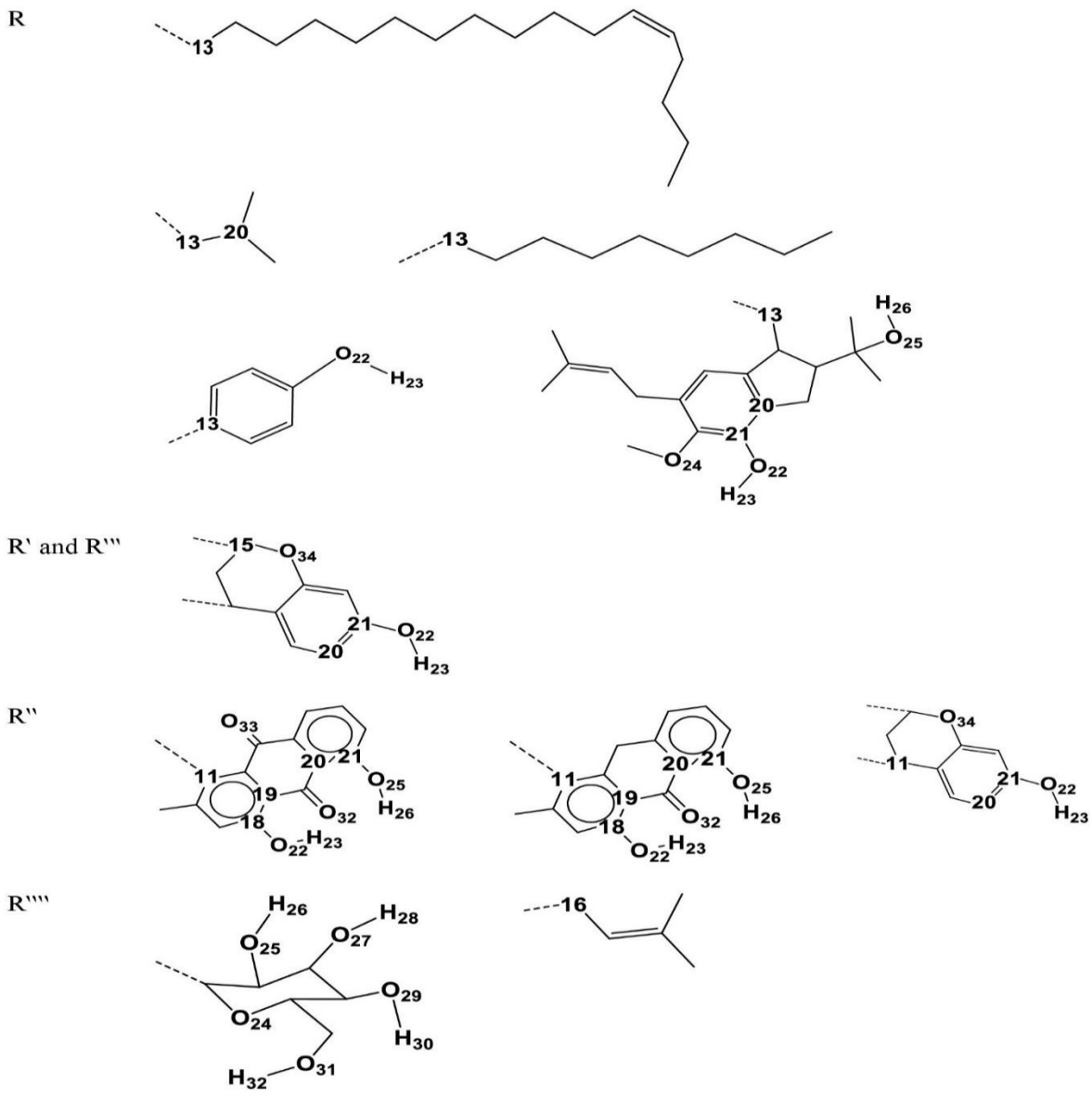



**Figure 5.3. Atom numbering of the different R, R', R'', and R'''' substituents appearing in the ACPL molecules shown in Fig. 1.**
The figure shows only the structures of the substituents, organised according to their positions (R, R', R'', R''''). The first carbon atom in each substituent is numbered, as explained in the caption of Fig. 2. The other atoms that are relevant for the analysis of the results are numbered as indicated in the structures above. When there is only one OH group in the substituent, it is conventionally numbered as O22–H23. Additional OH groups are assigned different and conventional numbers.



© University of Venda

**Table 5.2. Nature of the R, R', R", R"', R"" substituents and presence of OH groups and $sp^2$ O in the molecules of the acylphloroglucinols considered in this study.**

| Acronym | R, R', R", R"', R"" substituents | OH groups | $sp^2$ O |
|---|---|---|---|
| Mono1 | R = $(CH_2)_{11}C=C(CH_2)_3CH_3$ | O8H15, O10H16, O12H17 | O14 |
| Mono2 | R = $(CH_2)_8CH_3$<br>Terpene moiety attached at C4-C5 | O8H15, O12H17, O22H23, | O14 |
| Mono3 | R = $(CH_2)_8CH_3$<br>Terpene moiety attached at C2-C3 | O10H16, O12H17, O22H23 | O14 |
| Mono4 | R = R"' = $CH_3$<br>R" = an anthraquinone moiety | O10H16, O12H17, O22H23, O25H26 | O14, O32, O33 |
| Mono5 | R = R"' = $CH_3$<br>R" = an anthraquinone moiety | O10H16, O12H17, O22H23, O25H26 | O14, O32 |
| Mono6 | R = isobutyl, R' = $CH_3$<br>R"" = prenyl | O8H15, O12H17 | O14 |
| Mono7 | R = dihydrobenzofuran | O8H15, O10H16, O12H17, O22H23, O25H26 | O14 |
| Mono8 | R = phenol<br>R"" = glucoside | O10H16, O12H17, O22H23, O25H26, O27H28, O29H30, O31H32 | O14 |

### 5.3. Naming of conformers

Based on a method developed in earlier studies of ACPLs [1-9], conformers are represented by acronyms, with each letter signifying a particular geometric feature. These features include O–H···O IHBs like H15···O14, H17···O14, H23···O32, H26···O32, H26···O14, and H16···O33, as well as O–H···π interactions. Other important characteristics involve the orientation of OH groups, the arrangement of substituents (including ring systems) in relation to moiety A, and the location of C13 concerning the plane of ring A. The symbols that represent these geometric features are listed in Table 5.3, with further explanations provided in Figure 5.4.



© University of Venda

**Table 5.3. Symbols utilised in the acronyms denoting the molecules and their conformers.**

| Moiety or molecule considered | Feature category | Symbol | Meaning of symbol |
|---|---|---|---|
| acylphloroglucinol moiety | Presence and position of the first IHB | d | the H15···O14 first IHB is present |
| | | s | the H17···O14 first IHB is present |
| | Orientation of OHs *ortho* to acyl group | - | if engaged in the first IHB, indicated by *d* or *s* |
| | | - | if not engaged in the first IHB and oriented away from the acyl group, no indication is given |
| | | u | if not engaged in the first IHB and oriented toward the acyl group |
| | C3–C4–O10–H16 (or C16 if R"" is present) torsion angle | r | the angle is close to 0° |
| | | w | the angle is close to 180° |
| | | m | the angle is close to +90° (‘’towards us’’) |
| | | y | the angle is close to -90° (‘’away from us’’) |
| | Orientation of O14 when not engaged in the first IHB | z | oriented off-plane with respect to the plane of ring A, and away from us |
| general | absence of IHB | - | no IHB present |
| Mono1, Mono2 and Mono3 | Orientation of R with respect to ring A | a | R (including C13) is coplanar to ring A |
| | | b | R (including C13) is off-plane with respect to ring A and “towards us” |
| Mono2, Mono3, Mono4 and Mono5 | C19-C18-O22-H23 torsion angle | v | the angle is close to 180° |
| | C20-C21-O25-H26 torsion angle | k | the angle is close to 180° |
| Mono4 and Mono5 | Presence and position of other IHBs | ε | the H16··O33 IHB is present (Mono4) |
| | | x | the H23···O32 IHB is present (Mono4, Mono5) |
| | | j | the H25···O32 IHB is present (Mono4, Mono5) |
| Mono6 | Orientation of the isobutyl group | e | oriented to the other side with respect to O14 |
| | | f | oriented towards O14 |
| | | c | oriented “towards us”, with the H atom attached to C20 oriented towards O14 |
| | | i | oriented “towards us”, with the H atom attached to C20 oriented away from O14 |
| | | g | oriented “away from us”, with the H atom attached to C20 oriented towards O14 |
| | | h | oriented “away from us”, with the H atom attached to C20 oriented away from O14 |
| Mono7 | Presence and position of other IHBs | ʌ | the H26···O14 IHB is present |
| | | χ | the H23···O24 IHB is present |

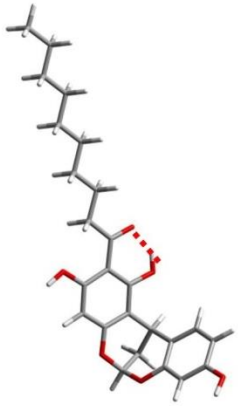
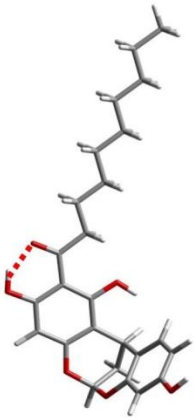
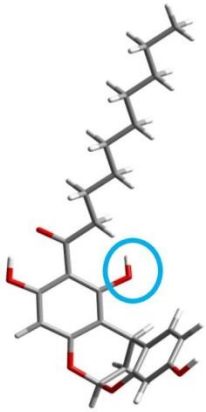
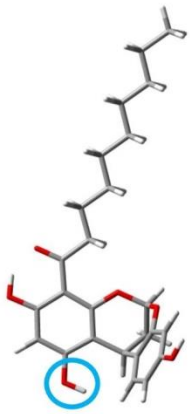
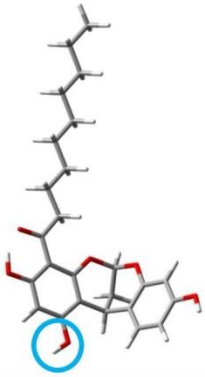
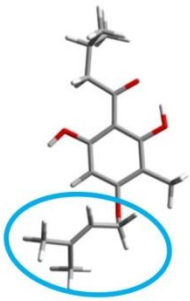
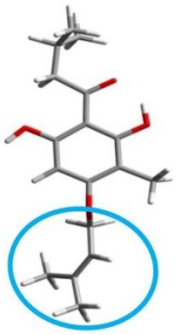
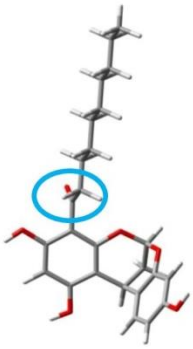
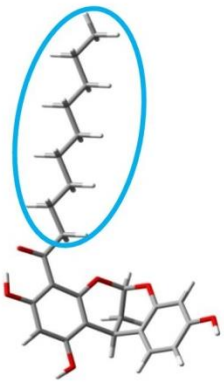
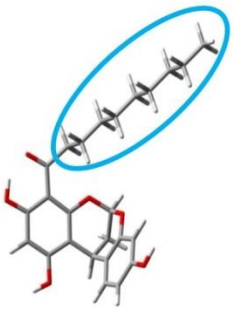
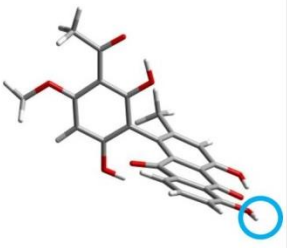


© University of Venda

| | | | |
|---|---|---|---|
| | Orientation of O22–H23 | λ | oriented away with respect to O24 |
| | C19-C18-O22-H23 torsion angle | p | the angle is close to +90° |
| | | q | the angle is close to -90° |
| | Orientation of the B-D ring system and the prenyl chain | α | ring system oriented away with respect to O14; prenyl chain oriented “towards us” |
| | | β | ring system oriented away with respect to O14; prenyl chain oriented “away from us” |
| | | γ | ring system oriented towards O14 |
| Mono8 | IHBs present in ring D | η | H17···π (C13) |
| | | κ | O27–H28···O29, O29–H30···O31, O31–H32···O24 |
| | | μ | O25–H26···O10, O27–H28···O25, O29–H30···O27, O31–H32···O29 |
| | | ξ | O25–H26···O10, O27–H28···O29, O29–H30···O31, O31–H32···O24 |
| | | ς | O25–H26···O10, O27–H28···O25, O29–H30···O31, O31–H32···O24 |
| | | δ | O25–H26···O27, O27–H28···O29, O29–H30···O31 |
| | | τ | O27–H328···O25, O29–H30···O27, O31–H32···O29 |
| | Orientation of ring B and O22–H23 | ω | ring B oriented “towards us”; O22–H23 oriented “downwards” |
| | | t | ring oriented “towards us”; O22–H23 oriented “upwards” |
| | | n | ring oriented “away from us”; O22–H23 oriented “downwards” |

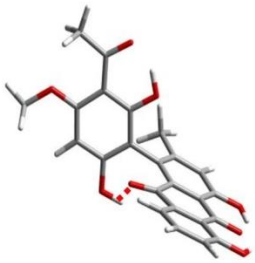
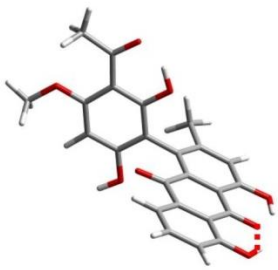
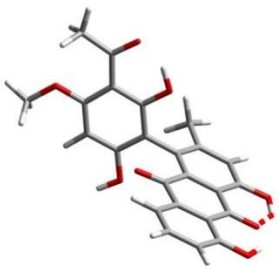
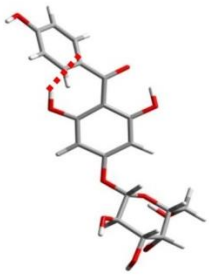
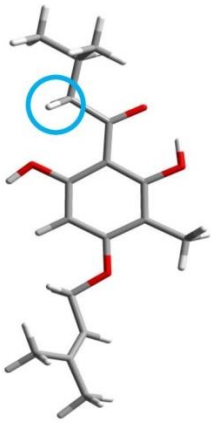
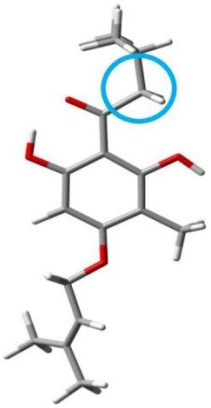
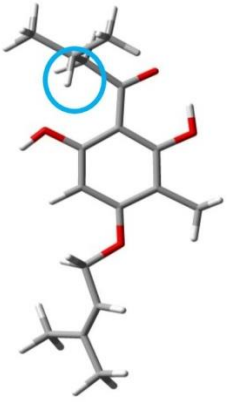
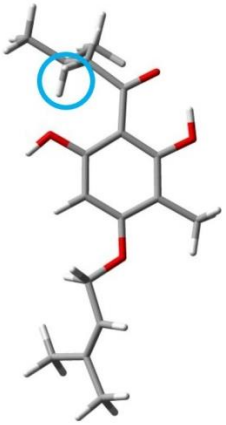
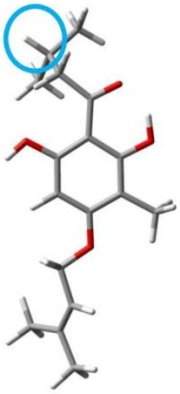
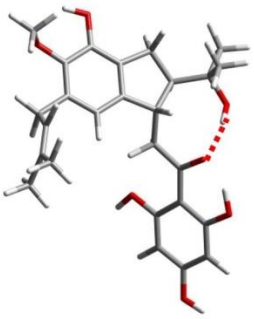
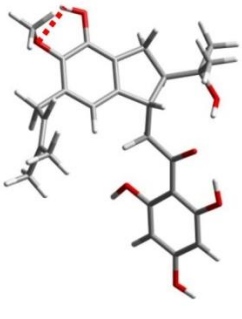
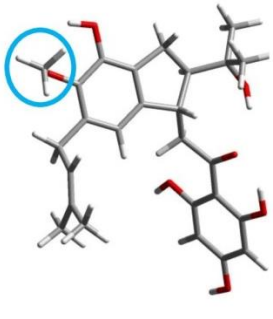


© University of Venda

| | | | |
|---|---|---|---|
|  |  |  |  |
| d | s | u | r |
|  |  |  |  |
| w | m | y | z |
|  |  |  |  |
| a | b | v | K |

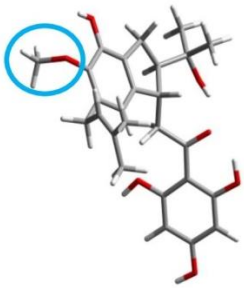
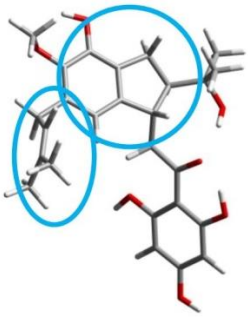
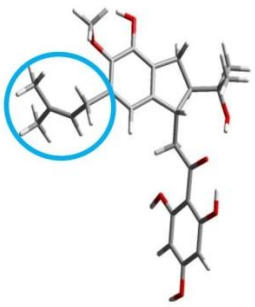
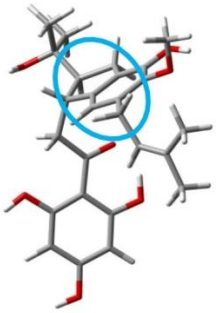
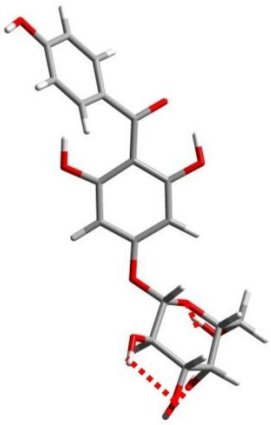
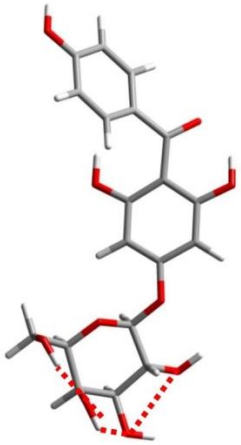
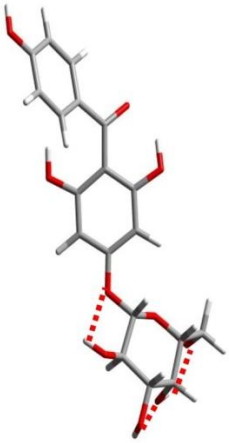
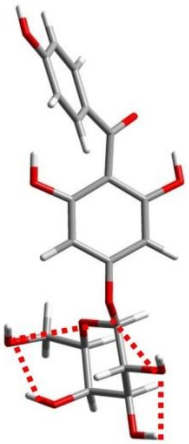
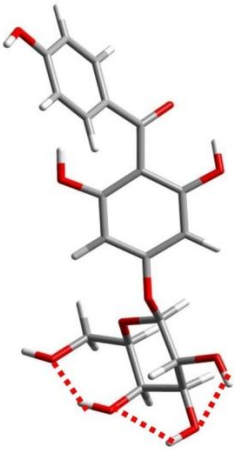
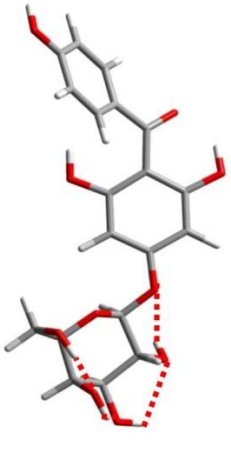
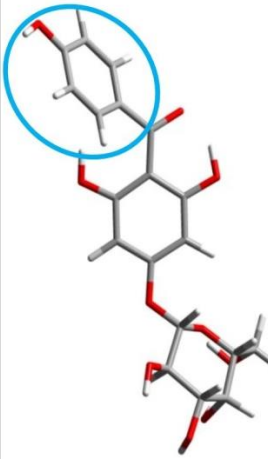
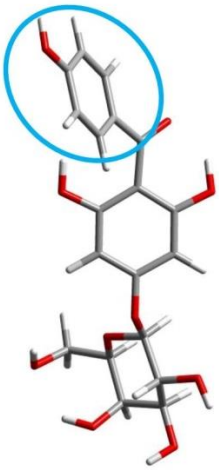


© University of Venda

| | | | |
|---|---|---|---|
|  |  |  |  |
| ε | x | j | η |
|  |  |  |  |
| e | f | c | g |
|  |  |  |  |
| h | ʌ | χ | p |



© University of Venda

| | | | |
|---|---|---|---|
|  |  |  |  |
| q | α | β | γ |
|  |  |  |  |
| κ | μ | ξ | ς |
|  |  |  |  |
| δ | τ | ω | t |



© University of Venda

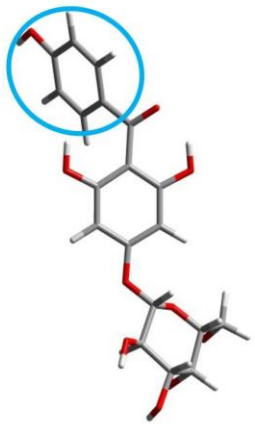
| | | | |
|---|---|---|---|
|  | | | |
| n | | | |

**Figure 5.4. Illustrations depict the features represented by letters in the acronyms for the calculated conformers of the ACPL molecules considered in this study.** The characteristics and their corresponding letters are detailed in Table 5.3. To clear visualization, blue circles emphasize the areas related to each characteristic, while red dashed lines indicate the IHBs under consideration.

The symbols used to represent the characteristics of the acylphloroglucinol moiety are retained from previous studies on ACPLs for consistency. It is worth revisiting the structure-related definitions of some of these symbols. When different substituents are present in the meta position relative to the acyl group, the carbon atoms of the ring are numbered so that the larger substituent is attached at C3. In this context, “d” denotes the first intramolecular hydrogen bond (IHB) forming on the same side as the substituent at C3, while “s” indicates the first IHB forming on the opposite side. The symbol “r” signifies that O10–H16 is oriented toward the side of the substituent at C3, whereas “w” indicates that it is oriented toward the opposite side. Thus, the distinction between “d” and “s” pertains to the presence of different substituents at C3 and C5, becoming more pronounced when a substituent other than hydrogen is at C3 and hydrogen is at C5 [1, 2]. If there are no substituents other than hydrogen at C3 and C5, the resulting symmetry negates the difference; in these cases, the letter “d” is conventionally used for the first IHB, with no “s” IHBs.



© University of Venda

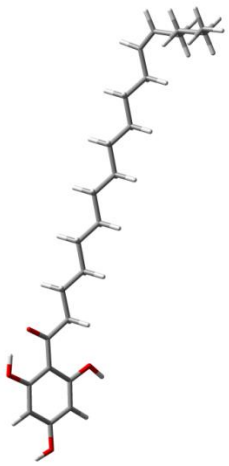
Here, “r” indicates that O10–H16 is oriented toward the side of the first IHB, while “w” indicates it is oriented towards the other side.

### 5.4. Results *in vacuo*

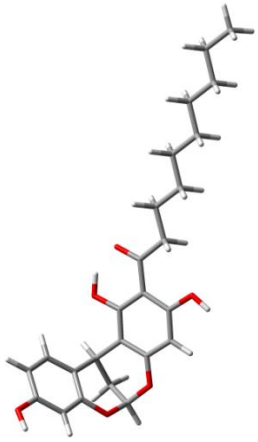
#### 5.4.1. Conformers’ geometrical characteristics

This study investigates the geometrical characteristics of conformers of the considered monomeric selected ACPLs. The optimized geometries of these conformers are depicted in Figure S.5.1. Figure 5.5 presents the optimized geometries of the lowest energy conformer for each of the ACPL molecules considered.

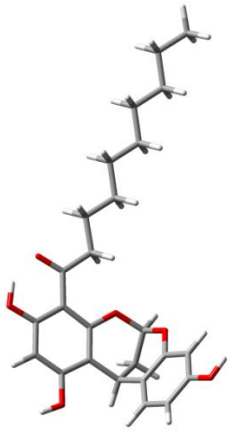
---



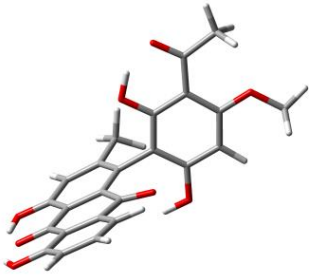
Mono1-d-r-a



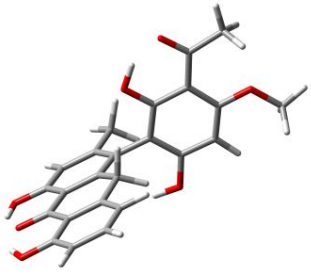
Mono2-d-v-a



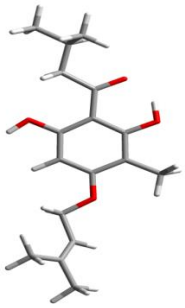
Mono3-s-x-w-a



Mono4-d-ε-r-x-j



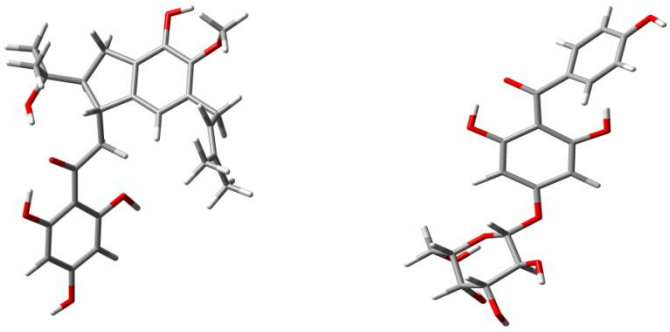
Mono5-d-r-x-j



Mono6-d-w-e



© University of Venda



Mono7-d-r-ʌ-χ-α-p Mono8-η-d-u-y-κ-ω

**Figure 5.5.** The lowest energy conformer of each of the considered ACPL molecules. DFT results *in vacuo*.

Similar to earlier studies of ACPLs, the primary factor influencing conformational preferences is the formation of IHBs. The $sp^2$ oxygen of the acyl group (O14) can form an IHB with either of the *ortho* hydroxyl groups, leading to H15···O14 or H17···O14; both are referred to as the "first IHB" and are denoted by the letters "d" or "s" in the acronyms for the conformers. Previous studies have indicated that these bonds range from moderate to strong in strength [2]. They cannot form if the hydrogen atom is replaced by other functional groups (Mono3, Mono4, Mono5 for H15).

Additional O–H···O IHBs can form in some of these molecules. Both Mono4 and Mono5 contain an $sp^2$ oxygen (O32) within the intermediate D ring of the B-D-E ring system, while Mono4 also features an additional oxygen (O33). O32 can engage in IHBs with H23 and H26 simultaneously (cooperative interactions involving H23···O32 and H26···O30, with bifurcation at O32) or with just one of them at a time. In Mono4, O33 can form IHBs with H16, but it is too far (3.9 Å) from H17, even when the H17···O14 IHB is absent and H17 is oriented toward it. In Mono7, H25 can form an IHB with O14, which may cooperate with H17···O14 if both are present.

While O–H···O IHBs involving $sp^3$ oxygen acceptors are generally weaker than those involving $sp^2$ oxygen, they still play significant stabilizing roles. The glucoside moiety of Mono8 exhibits the most extensive variety of possible IHBs, including O25–H26···O10, O25–H26···O27, O27–H28···O29, O27–H28···O25, O29–H30···O31, O29–H30···O27, O31–H32···O24, and O31–



© University of Venda

H32···O29. Up to four of these IHBs can be present at the same time, exhibiting consecutive and cooperative interactions.

O–H···π interactions have long been recognized as hydrogen bonds, primarily due to their significance in biomolecules [10, 11] and continue to attract interest [12]. Their contributions to the conformational preferences of ACPLs have been explored in previous studies [4], with further emphasis placed on specific ACPL molecules. In Mono8, O8–H15 and O12–H17 can participate in this type of IHB with ring B.

The orientations of the hydroxyl groups also affect the conformers' energy and are specified in their acronyms (Table 5.3). When involved in the first IHB, the orientation of O8–H15 or O12–H17 is automatically determined (toward O14). If they are not engaged in the first IHB, only the higher-energy orientation (toward the acyl group) is indicated (denoted by the letter "u"). The hydroxyl groups of the acylphloroglucinol moiety typically remain coplanar with the benzene ring [1, 13]; from a rotational symmetry perspective, they prefer a uniform orientation ($C_{3v}$, see Fig. 5.6).

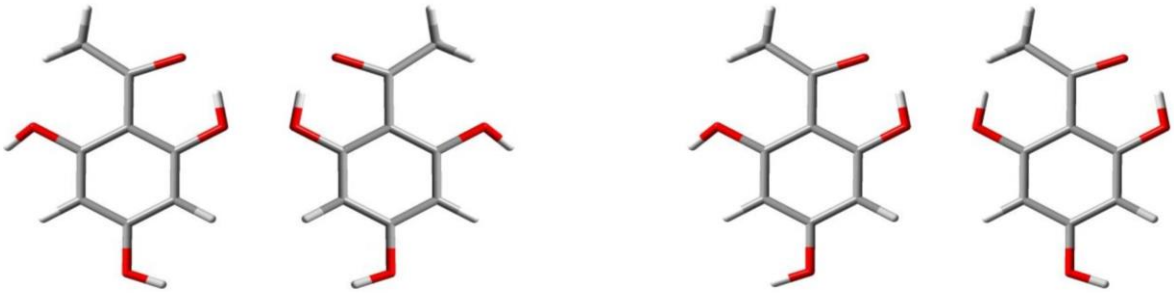



**Figure 5.6. Illustrations of uniform and non-uniform orientations of the OH groups in the acylphloroglucinol moiety.**



© University of Venda

The orientation of substituents relative to the acylphloroglucinol moiety can also influence conformational preferences and energy, especially in cases where certain orientations promote interactions, such as IHBs, between them.

The optimized geometries largely mirror the input geometries, consistent with the phenomena previously observed for ACPLs. For example, since R is not hydrogen in any of the compounds considered when O14 is not involved in the first IHB and remains coplanar with the ring in the input, the acyl group rotates during optimization. This causes O14 to shift off-plane (but not perpendicular) to minimize repulsion with O8 and O12 [1, 2].

### 5.4.2. Conformers' relative energies and factors influencing them

The relative energies of the four lowest energy conformers for each molecule are presented in Table 5.4, while the energies of all calculated conformers can be found in Table S5.1.

**Table 5.4.** Relative energies of the four lowest energy conformers of the calculated ACPL molecules *in vacuo*, from full optimisation calculations.

HF/6-31G(d,p), DFT/B3LYP/6-31+G(d,p) and MP2/ 6-31G(d,p) results, respectively denoted as DFT, HF and MP2 in the columns' headings. For each molecule, the conformers are listed in order of increasing relative energies in the DFT results.

| Molecules and conformers | Relative energy (kcal $mol^{-1}$) | | |
|---|---|---|---|
| | DFT | HF | MP2 |
| Mono1-d-r-a | 0.000 | 0.000 | 0.000 |
| Mono1-d-w-a | 1.380 | 1.516 | 1.402 |
| Mono1-d-u-r-a | 2.879 | 4.373 | 2.820 |
| Mono1-d-u-w-a | 3.229 | 4.785 | 3.189 |
| | | | |
| Mono2-d-v-a | 0.000 | 0.041 | 0.051 |
| Mono2-s-v-a | 0.080 | 0.000 | 0.000 |
| Mono2-s-v-u-a | 4.178 | 5.019 | 3.165 |
| Mono2-d-x-a | 4.327 | 5.822 | 4.096 |
| | | | |
| Mono3-s-x-w-a | 0.000 | 0.000 | 3.334 |
| Mono3-s-v-w-a | 0.030 | 0.135 | 3.491 |
| Mono3-s-x-w-b | 0.534 | 1.024 | 0.000 |
| Mono3-s-x-r-a | 3.135 | 3.994 | 5.772 |
| | | | |
| Mono4-d-ε-r-x-j | 0.000 | 0.000 | 0.000 |
| Mono4-d-w-x-j | 2.091 | 2.669 | 3.410 |



© University of Venda

| | | | |
|---|---|---|---|
| Mono4-d-ε-r-v-j | 12.259 | 10.941 | 11.408 |
| Mono4-d-ε-r-x-k | 12.923 | 11.510 | 12.342 |
| | | | |
| Mono5-d-r-x-j | 0.000 | 0.000 | 0.000 |
| Mono5-d-w-x-j | 3.873 | 4.058 | 4.286 |
| Mono5-d-r-v-j | 12.880 | 11.948 | 11.683 |
| Mono5-d-r-x-k | 13.489 | 12.498 | 12.124 |
| | | | |
| Mono6-d-w-e | 0.000 | 0.000 | 0.461 |
| Mono6-d-w-g | 0.418 | 0.889 | 0.000 |
| Mono6-d-w-c | 0.443 | 0.887 | 0.000 |
| Mono6-s-w-f | 1.133 | 1.064 | 1.320 |
| | | | |
| Mono7-d-r-ʌ-χ-α-p | 0.000 | 0.000 | 0.000 |
| Mono7-d-w-ʌ-χ-α-p | 1.325 | 1.466 | 1.321 |
| Mono7-d-w-ʌ-χ-α-q | 1.561 | 1.589 | 2.252 |
| Mono7-d-w-ʌ-χ-β-p | 1.845 | 1.815 | 0.150 |
| | | | |
| Mono8-η-d-u-y-κ-ω | 0.000 | 0.951 | 0.019 |
| Mono8-η-d-u-y-κ-t | 0.000 | 1.038 | 0.000 |
| Mono8-η-d-u-w-μ-t | 1.213 | 0.000 | 1.416 |
| Mono8-d-y-κ-ω | 1.378 | 1.607 | 3.948 |

Comparing conformers that differ by only one characteristic offers insights into the stabilizing effects of that characteristic. The acronyms used help in this comparison. For example, the acronyms Mono2-s-v-a and Mono2-d-v-a indicate that these two conformers are appropriate for evaluating the influence of the two first intramolecular hydrogen bonds (H15···O14 and H17···O14), as they differ solely by the position of the first IHB (represented by the letters s and d). Similarly, Mono2-s-v-a and Mono2-s-v-u-a can be compared to assess the effect of the different orientations of O8–H15 when it is not involved in the first IHB (pointing away from the acyl group in Mono2-s-v-a and towards the acyl group in Mono2-s-v-u-a). A comparison between Mono2-d-v-a and Mono2-x-a illustrates the energy increase that occurs when the H15···O14 IHB is removed.

These comparisons also highlight the consistency of the findings with previous studies on ACPLs. For instance, the comparison of Mono2-s-v-a and Mono2-s-v-u-a (and similar pairs across all studied compounds) shows that the observed trends align with the general conclusions established for ACPLs, specifically that orientations towards the acyl group lead to higher energy states [1, 2].



© University of Venda

Compounds Mono2 and Mono6 are the only ones for which a comparison between the effects of H15···O14 and H17···O14 interactions can be made, as both types of conformers are possible. The relative energy values show that the energy of the d conformers is slightly lower than that of the corresponding s conformers, which is consistent with the results obtained for all other ACPLs.

Tables S2 and S3 present the relative energies, zero-point energy (ZPE) corrections, corrected relative Gibbs free energies ($\Delta G_{corrected}$, the sum of electronic and thermal free energies), and the free energy correction ($G_{corr}$) for all calculated conformers. Table S5.2 summarizes the DFT results, while Table S5.3 contains the HF results. Table 5.5 outlines the ranges of ZPE corrections and free energy corrections. The relative energies, both corrected for ZPE and uncorrected, generally follow similar patterns, with a few exceptions (e.g., Mono7-d-w-ʌ-λ-α-p at the DFT level and Mono8-η-d-u-y-κ-t, Mono8-η-d-u-w-μ-t at the HF level). In most cases, conformers with higher relative energy also show higher relative Gibbs free energy. The ZPE and Gibbs free energy corrections are relatively consistent for conformers of the same molecule but vary between different molecules. The largest corrections are found for Mono1 conformers, while the smallest are observed for Mono4.

**Table 5.5. Ranges of the ZPE correction to the electronic energy ($ZPE_{corr}$) and of the thermal correction to the Gibbs free energy ($G_{corr}$) for the calculated conformers of the molecules considered in this work.** DFT/B3LYP/6-31+G(d,p) and HF/6-31G(d,p) results *in vacuo* from full optimization calculations, respectively denoted as DFT and HF in the columns' headings.

| Molecule | $ZPE_{corr}$ (kcal $mol^{-1}$) | | $G_{corr}$ (kcal $mol^{-1}$) | |
|---|---|---|---|---|
| | DFT | HF | DFT | HF |
| Mono1 | 364.29–365.30 | 390.82–391.52 | 319.07–323.08 | 347.40–347.87 |
| Mono2 | 319.69–320.77 | 343.41–344.79 | 278.46–281.20 | 303.12–305.93 |
| Mono3 | 319.82–321.16 | 343.55–345.14 | 279.04–281.47 | 303.55–305.82 |
| Mono4 | 235.02–236.20 | 252.57–255.20 | 195.93–199.65 | 215.37–219.64 |
| Mono5 | 246.75–247.93 | 266.03–267.48 | 209.70–212.11 | 229.82–232.10 |
| Mono6 | 238.09–239.21 | 255.56–256.95 | 203.17–205.23 | 221.04–223.41 |
| Mono7 | 337.66–338.40 | 363.49–364.33 | 294.51–296.26 | 321.50–323.35 |
| Mono8 | 238.86–239.73 | 259.03–260.07 | 201.38–203.68 | 221.53–225.11 |



© University of Venda

Figure S5.2 illustrates the trends in relative energies—both uncorrected and corrected for ZPE—across the calculated conformers for each molecule, based on the computational methods used. All the diagrams show a slow-gradient low-energy region, corresponding to conformers with the initial O–H···O IHBs. A sharp rise in energy corresponds to the few conformers that lack IHBs.

### 5.4.3. Characteristics of the intramolecular hydrogen bonds

The strength of intramolecular hydrogen bonds (IHBs) can be evaluated by comparing several parameters: the H···O bond length, O···O distance, OĤO angle, the energy increase when an IHB is removed without a new one forming, and the reduction in IR vibrational frequencies of the donor OH groups. Table S5.4 details the IHB parameters for the calculated conformers of each molecule, while Table 5.6 shows the ranges of their lengths. The parameters for the first IHB suggest it falls within the moderate-to-strong range, similar to other ACPLs [2]. Its strength is attributed to its closure of a six-membered ring and its status as a resonance-assisted hydrogen bond [14-17].

**Table 5.6.** Ranges of the lengths of the intramolecular hydrogen bonds in the calculated conformers of the considered ACPL molecules *in vacuo*.

DFT/B3LYP/6-31+G(d,p), HF/6-31G(d,p), and MP2/6-31G(d,p) result from full optimization calculations, respectively denoted as DFT, HF, and MP2 in the columns' headings. The molecules are denoted with the symbols listed in Table 5.1.

| Molecules | Considered IHB | Range of the length of the IHB (Å) | | |
|---|---|---|---|---|
| | | DFT | HF | MP2 |
| Mono1 | H15···O14 | 1.563–1.584 | 1.692–1.719 | 1.615–1.652 |
| | | | | |
| Mono2 | H15···O14 | 1.545–1.546 | 1.672–1.673 | 1.599–1.600 |
| | H17···O14 | 1.570–1.588 | 1.696–1.733 | 1.630–1.661 |
| | | | | |
| Mono3 | H17···O14 | 1.545–1.565 | 1.681–1.697 | 1.615–1.667 |
| | | | | |
| Mono4 | H15···O14 | 1.539–1.547 | 1.665–1.671 | 1.607–1.649 |
| | H23···O32 | 1.615–1.654 | 1.766–1.785 | 1.682–1.707 |
| | H25···O32 | 1.684–1.694 | 1.800–1.814 | 1.712–1.727 |
| | | | | |
| Mono5 | H15···O14 | 1.536–1.541 | 1.666–1.670 | 1.590–1.595 |
| | H23···O32 | 1.609–1.660 | 1.747–1.760 | 1.687–1.697 |
| | H25···O32 | 1.627–1.682 | 1.767–1.783 | 1.699–1.714 |



© University of Venda

| | | | | |
|---|---|---|---|---|
| Mono6 | H15···O14 | 1.538–1.585 | 1.663–1.722 | 1.607–1.649 |
| | H17···O14 | 1.563 | 1.691 | 1.614 |
| | | | | |
| Mono7 | H15···O14 | 1.565–1.601 | 1.703–1.746 | 1.612–1.683 |
| | H26···O14 | 1.923–1.979 | 2.057–2.128 | 1.963–2.838 |
| | | | | |
| Mono8 | H15···O14 | 1.627–1.656 | 1.808–1.856 | 1.712–1.751 |
| | H32···O10 | 2.535–2.632 | 2.535–2.632 | 2.415–2.521 |
| | H32···O28 | 2.416–2.430 | 2.441–2.452 | 2.335–2.351 |
| | H33···O27 | 2.353–2.402 | 2.371–2.408 | 2.306–2.340 |
| | H33···O29 | 2.183–2.402 | 2.209–2.408 | 2.134–2.340 |
| | H34···O28 | 2.202–2.219 | 2.227–2.233 | 2.135–2.151 |
| | H34···O31 | 1.898–2.219 | 1.947–2.233 | 1.861–2.151 |
| | H35···O29 | 1.992–2.617 | 2.036–2.429 | 1.946–2.269 |
| | H17···π (C13) | 2.172–2.193 | 2.301–2.638 | 2.144–2.161 |

The first IHB, H15···O14, is slightly shorter than H17···O14, with both bonds being longer in u-conformers compared to those where the other *ortho* OH is oriented away from the acyl group. This is consistent with those for ACPLs in general.

The nature of the R group also affects the first IHB's length, influenced by steric factors or its potential to interact with O8–H15 or O12–H17. For example, in Mono8 u-conformers, the first IHB is longer due to the interaction between the ortho OH not involved in the IHB and the R group (a phenol) through an O–H···π interaction. This interaction pulls the acyl group toward the non-involved OH, lengthening the first IHB.

In Mono7, the R group contains an OH group (O25–H26) that can form an IHB with O14. In ʌ-conformers, this H26···O14 IHB bifurcates from the first IHB (on O14), slightly extending the first IHB. However, the first IHB remains stronger because it forms a six-membered ring and is resonance-assisted, whereas the H26···O14 bond forms an eight-membered ring and is not resonance-assisted.

For the IHBs in the glucoside moiety of Mono8, O25–H26···O27 is slightly shorter than O25–H26···O10, and O27–H28···O29 is shorter than O27–H28···O25. Similarly, O29–H30···O31 is



© University of Venda

shorter than O29–H30···O27, while O31–H32···O29 is shorter than O31–H32···O24. Among these, O29–H30···O31 is the shortest, and O25–H26···O10 is the longest, influenced by O10's position.

In Mono4 and Mono5, two bifurcated intramolecular hydrogen bonds (IHBs), H23···O32 and H26···O32, are present in their ring systems. These IHBs are shorter in Mono5 compared to Mono4, likely due to the presence of an $sp^2$ oxygen atom (O33) in Mono4, which may influence the bond lengths.

The energy increase observed upon removing a hydrogen bond (through a 180° rotation of the donor group) does not necessarily reflect the bond's strength, as this process often involves additional structural changes, such as an off-plane shift of O14 in ACPLs. However, if removing an IHB does not lead to the formation of another IHB, comparing the energy increases between conformers with and without the IHB provides insight into the relative stabilizing effects of the IHBs. For instance, this is exemplified by the comparison between Mono2-d-x-a and Mono2-x-a. Table 5.7 presents the energy increases associated with the removal of specific IHBs, using appropriately selected conformers for comparison.



© University of Venda

**Table 5.7.** Energy increase on removal of the first intramolecular hydrogen bond (IHB) from the lowest energy conformer of the ACPL molecules considered in this work.

Results from full optimisation calculations HF/6-31G(d,p), DFT/B3LYP/6-31+G(d,p) and MP2/ 6-31G(d,p) calculations, respectively denoted as DFT, HF and MP2 in the columns' headings.

The energy difference is taken as « energy of the lowest energy conformer resulting from the removal of the first IHB minus energy of the lowest energy conformer with the IHB».

| Molecules and conformers | | Energy difference (kcal mol$^{-1}$) | | |
|---|---|---|---|---|
| With the IHB | Without the IHB | DFT | HF | MP2 |
| **Mono1** | | | | |
| Mono1-d-r-a | Mono1-r-a | 14.717 | 15.476 | 12.423 |
| | | | | |
| **Mono2** | | | | |
| Mono2-d-v-a | Mono2-x-a | 16.643 | 14.187 | 13.407 |
| | | | | |
| **Mono3** | | | | |
| Mono3-s-x-w-a | Mono3-v-w-a | 13.739 | 11.128 | 11.869 |
| | | | | |
| **Mono4** | | | | |
| Mono4-d-w-v-k | Mono4-w-v-k | 13.044 | 10.080 | 9.901 |
| | | | | |
| **Mono5** | | | | |
| Mono5-d-r-x-j | Mono5-r-x-j | 13.528 | 10.372 | 10.671 |
| | | | | |
| **Mono6** | | | | |
| Mono6-d-w-e | Mono6-w-f | 15.142 | 11.821 | 12.158 |
| | | | | |
| **Mono7** | | | | |
| Mono7-d-r-ʌ-χ-α-p | Mono7-w-ʌ-χ-α-p | 14.314 | 11.687 | 9.434 |
| | | | | |
| **Mono8** | | | | |
| Mono8-η-d-u-y-κ-ω | Mono8-y-κ-ω | 11.949 | 6.542 | 10.077 |

Infrared (IR) vibrational frequencies were calculated for all compounds using the harmonic approximation. The frequencies of O–H groups are particularly relevant in the context of IHBs. When an O–H group acts as a hydrogen bond donor, its vibrational frequency decreases (redshift), while its bond length increases. The magnitude of the redshift serves as an indicator of the H-bond's strength, with stronger bonds showing a greater redshift.



© University of Venda

The calculated frequencies for all OH groups in each molecule are listed in Table S5.5 for DFT results and Table S5.6 for HF results. Some OH vibrations are coupled, displaying both symmetric and asymmetric modes, as confirmed through vibrational visualization. In Mono4 and Mono5, the vibrations of O22–H23 and O25–H26 are coupled, showing identical frequencies when these groups form simultaneous, cooperative IHBs with O32.

The redshift of a hydrogen bond donor is measured as the difference between its frequency in a free state (not participating in a hydrogen bond) and when it forms a hydrogen bond. In this study, the reference frequency for each OH group was determined as the average of its frequencies in conformers where it is not engaged in an IHB. The redshifts for IHBs in the calculated conformers of ACPLs are provided in Table S5.7 for DFT results and Table S5.8 for HF results, with their ranges presented in Table 5.8. The computed redshifts align with the IHB strength comparisons inferred from bond lengths. The largest redshift is observed for IHBs, where O8–H15 is the donor. However, this redshift is slightly smaller in the conformers of Mono7 and Mono8 compared to the other molecules, likely due to the presence of cyclic substituents in Mono7 and Mono8.

**Table 5.8.** Range of the red-shifts in the calculated vibrational frequencies (harmonic approximation) of the O–H bonds that act as IHB donors in the calculated conformers of the considered ACPL molecules.

DFT/B3LYP/6-31+G(d,p) and HF/6-31G(d,p) results from full optimisation, respectively denoted as DFT and HF in the columns' headings. The molecules are denoted with the symbols listed in table 5.1.

| Molecules | Considered IHB | Range of the red-shifts ($cm^{-1}$) | |
|---|---|---|---|
| | | DFT | HF |
| Mono1 | H15···O14 | 674.92–755.20 | 240.38–306.78 |
| | | | |
| Mono2 | H15···O14 | 799.89–806.20 | 236.34–295.69 |
| | H17···O14 | 701.82–802.49 | 340.49–342.33 |
| | | | |
| Mono3 | H17···O14 | 721.40–804.83 | 286.62–326.76 |
| | | | |
| Mono4 | H15···O14 | 47.69–786.17 | 316.36–336.80 |
| | H23···O32 | 482.45–603.90 | 179.91–191.87 |
| | H25···O32 | 415.40–541.73 | 164.91–184.79 |
| | | | |
| Mono5 | H15···O14 | 730.52–779.67 | 305.61–325.88 |
| | H23···O32 | 449.35–635.64 | 208.61–243.60 |



© University of Venda

| | | | |
|---|---|---|---|
| | H25···O32 | 438.51–589.70 | 178.15–221.17 |
| | | | |
| Mono6 | H15···O14 | 629.13–789.35 | 250.00–341.51 |
| | | | |
| Mono7 | H15···O14 | 575.69–724.94 | 203.26–272.35 |
| | H26···O14 | 86.41–147.25 | 37.49–82.90 |
| | H23···O32 | 79.46–86.26 | 42.77–45.67 |
| | | | |
| Mono8 | H15···O14 | 515.89–598.79 | 146.45–177.59 |

The redshifts that occur when both O22–H23 and O25–H29 form H23···O32 and H26···O32 at the same time are significantly smaller than when either hydrogen bond is present on its own. This is due to the longer length of these hydrogen bonds when they coexist, which is linked to their bifurcated structure on O32.

Similarly, the redshifts observed when O22–H23 and O25–H26 form H23···O32 and H26···O32 together are much smaller than when only one of the hydrogen bonds is present. This decrease in redshift is due to the increased length of the hydrogen bonds when they are both present, associated with their bifurcation on O32.

### 5.4.4. HOMO-LUMO energy gap of the conformers.

Table S5.9 presents the HOMO-LUMO energy gap values for the conformers of the studied molecules, as determined by three different calculation methods. Figure S5.3 provides a comparison of these values across the computational methods used. The results reveal a well-known pattern where the DFT values are notably lower than those obtained from the two ab initio methods. However, the trends among conformers of the same molecule remain consistent across all three methods. Table 5.9 presents the ranges of the HOMO-LUMO energy gaps, illustrating their dependence on the molecular structure: the smallest energy gap is found in Mono4, while the largest is associated with Mono1.



© University of Venda

**Table 5.9. Ranges of the HOMO-LUMO energy gap of the calculated conformers of the considered ACPL molecules.**
DFT/B3LYP/6-31+G(d,p), HF/6-31G(d,p) and MP2/6-31G(d,p) results *in vacuo* from full optimization calculations, respectively denoted as DFT, HF and MP2 in the columns' headings.

| Molecules | HOMO-LUMO energy gap range (kcal $mol^{-1}$) | | |
|---|---|---|---|
| | DFT | HF | MP2 |
| Mono1 | 103.30–126.82 | 260.06–277.56 | 253.92–287.81 |
| Mono2 | 95.79–123.32 | 245.39–271.57 | 237.10–263.66 |
| Mono3 | 99.12–121.23 | 249.94–272.86 | 243.27–267.02 |
| Mono4 | 63.96–79.84 | 213.01–228.71 | 198.24–215.05 |
| Mono5 | 84.22–96.30 | 232.99–247.21 | 225.21–236.25 |
| Mono6 | 96.99–122.23 | 251.45–288.63 | 244.26–281.30 |
| Mono7 | 92.04–102.01 | 245.77–261.28 | 235.02–258.62 |
| Mono8 | 92.53–111.09 | 239.68–265.48 | 237.42–258.52 |

In addition to the molecular structure, the HOMO-LUMO energy gap is also affected by conformational characteristics such as intramolecular hydrogen bonds (IHBs) and the orientation of OH groups. For OH groups capable of forming IHBs, the combined effects of both the IHB and the orientation of the donor OH group are considered. The impact of the first IHB's position depends on the type of R' and the overall molecule. For example, in Mono2, the d-conformers (with O14···H15 IHB) have a larger HOMO-LUMO energy gap than the s-conformers (with O14···H17 IHB), whereas in Mono5, the opposite is true. Additionally, the presence of the first IHB ensures that O14 and C13 remain coplanar with ring A. When this IHB is absent, O14 and C13 shift off-plane, resulting in a smaller HOMO-LUMO energy gap. For Mono4 and Mono5, the gap is smaller when both OH groups attached to R" are oriented towards O32, forming two simultaneous IHBs. The gap increases when one OH group is oriented away from O32, removing one IHB, and increases further when both OH groups are oriented away, removing both IHBs.

The orientation of OH groups that do not or cannot form IHBs also significantly influences the energy gap. For example, the energy gap is larger when the ortho phenol OH group, which is not involved in the first IHB, is oriented away from the acyl group compared to when it is oriented towards it (as seen in the u-conformers). Additionally, considering the orientation of the O10–H16 group, the energy gap is greater in the w-conformer than in the r-conformer.



© University of Venda

### 5.4.5. Dipole moments of the conformers.

Table S5.10 presents the dipole moment values for the calculated conformers of the studied ACPLs, while Figure S5.4 compares these values across the three computational methods, showing similar patterns. Figure S5.5 provides a comparison based on the types of conformers.

The dipole moment is primarily affected by the relative orientation of the OH groups [1]. The phenol OH groups in the phloroglucinol moiety are easier to interpret because they lie in the same plane: their overall contribution is zero when they have a uniform orientation (as shown in Fig. 5.6) and significant when their orientations differ. Therefore, in molecules where the three phenol OH groups are aligned uniformly, the dipole moment contributions mainly come from other parts of the molecule.

In molecules with additional OH groups, analysing individual contributions becomes more complex (e.g, [18]), particularly when OH groups from different moieties are on separate planes. Some trends can still be identified. For example, in Mono2 conformers, the dipole moment is larger when O22–H23 (in the E ring) and O8–H15 are aligned in the same direction and smaller when they are aligned oppositely. The conformer with the smallest dipole moment in Mono2 is Mono2-d-x-a, where O22–H23 and O12–H17 oppose the orientation of O8–H15. Similarly, in Mono3 conformers, the dipole moment is larger when O22–H23 and O10–H16 are oriented the same way, and smaller when oriented oppositely, with the smallest value found in Mono4-d-ε-r-x-k.

In Mono4, the dipole moments are generally smaller than those of the Mono5 conformers. This difference likely results from the opposing effect of the C=O33 bond relative to the combined influence of two OH groups on the opposite side of the ring system, present in Mono4 but absent in Mono5. The impact of OH group orientations is further illustrated in Mono8, where different orientations of the four OH groups in the glucoside moiety lead to a wide range of dipole moments, from as high as 6.050, 8.507, and 6.398 debye to as low as 1.191, 1.454, and 1.610 debye.



© University of Venda

## 5.5. Results in solution

### 5.5.1. Relative energies of the conformers.

Table S5.11 outlines the relative energies of the calculated conformers from the DFT results across four different media: vacuum, chloroform, acetonitrile, and water. Meanwhile, Table S12 presents the HF results. Figure S5.6 illustrates these comparisons for each of the examined ACPL molecules.

According to the DFT results, the lowest energy conformer remains consistent across most media, except for Mono3 in acetonitrile, where the two lowest energy conformers switch positions. Generally, as solvent polarity increases, the relative energy of a conformer tends to decrease, although Mono1, Mono6, and Mono7 are notable exceptions. The energy reduction is more pronounced for conformers lacking any IHBs. This phenomenon can be attributed to the PCM models, which, while not explicitly accounting for strong solute-solvent interactions like hydrogen bonds, implicitly consider them through polarization charges. Additionally, OH groups that are not involved in IHBs can engage in stronger solute-solvent interactions, such as solute-solvent hydrogen bonds.

The free energy of solvation ($\Delta G_{solv}$) represents the change in Gibbs free energy when a solute dissolves in a solvent, reflecting the stabilization of solute molecules due to interactions with the solvent. A more negative $\Delta G_{solv}$ indicates a stronger tendency for the molecule to dissolve. Table S13 details the solvation free energy ($\Delta G_{solv}$), along with its electrostatic ($G_{el}$) [19] and dispersion components [20-24], for the calculated conformers of the ACPL molecules in chloroform, acetonitrile, and water based on DFT results. Table S14 presents similar data from HF results, and Figure S5.7 provides a graphical comparison.

In acetonitrile, $\Delta G_{solv}$ values are highest and mostly positive for the conformers of compounds Mono1-Mono7, except for Mono4-d-w-v-k, Mono4-w-v-k, and Mono5-d-v-k-w, while all conformers of Mono8 show negative values. Chloroform yields lower $\Delta G_{solv}$ values compared to acetonitrile, remaining positive for Mono1, Mono2 (except Mono2-x-a), Mono3, Mono6, and Mono7, with only one insignificant exception; for Mono4 and Mono5, values are mostly negative, and all conformers of Mono8 are negative and with larger magnitudes. In water, Mono1 displays



© University of Venda

small $\Delta G_{solv}$ values (both positive and negative), while Mono2-Mono7 have more negative values, and Mono8 shows the most negative values with the greatest magnitude. Notably, conformers lacking the first intramolecular hydrogen bond (IHB) exhibit lower $\Delta G_{solv}$ than other conformers across all three solvents. Compounds with additional OH groups (Mono4, Mono5, Mono8) have more negative values, particularly Mono8, which stands out with significantly lower values. Table 5.10 summarizes the ranges of $\Delta G_{solv}$ for each molecule's conformers in the three solvents, excluding those without any IHB due to their significantly different values.

**Table 5.10. Ranges of the solvation free energy ($\Delta G_{solv}$, kcal mol$^{-1}$) for the calculated conformers of the molecules considered in this work (excluding the conformers without any IHB because of the significant difference of their values with respect to the others).** DFT/B3LYP/6-31+G(d,p) and HF/6-31G(d,p) results from full optimization calculations, respectively denoted as DFT and HF in the columns' headings.

| Molecule | DFT | | | HF | | |
|---|---|---|---|---|---|---|
| | chlrf | actn | aq | chlrf | actn | aq |
| Mono1 | 6.39–7.10 | 13.62–14.78 | (-0.26)–0.95 | (-4.00)–(-2.47) | 2.57–4.64 | (-12.47_–(-7.67) |
| Mono2 | 1.08–2.49 | 6.53–9.04 | (-7.51)–(-6.05) | (-4.82)–1.83 | 0.56–8.39 | (-17.67_–(-8.80) |
| Mono3 | 0.61–3.58 | 6.11–9.68 | (-11.35)–(-5.15) | (-3.80)–1.96 | (1.68)–8.19 | (-16.40)–(-8.92) |
| Mono4 | (-8.28)–(-1.06) | (-5.69)–3.80 | (-28.25)–(-12.36) | (-0.82)–1.68 | 3.79–7.39 | (-12.08)–(-5.50) |
| Mono5 | (-2.87)–0.41 | (1.17)–5.09 | (-16.19)–(-9.39) | (-1.34)–1.32 | 3.30–7.10 | (-12.76)–(-6.20) |
| Mono6 | 2.20–3.62 | 6.35–8.36 | (-4.08)–(-1.73) | (-0.72)–2.10 | 3.36–6.56 | (-6.70)–(-3.51) |
| Mono7 | (-0.03)–2.38 | 4.59–7.46 | (-14.67)–(-10.86) | (-1.96)–3.19 | (-2.04)–8.99 | (-17.54)–(-11.76) |
| Mono8 | (-9.72)–(-8.04) | (-7.30)–(-4.12) | (-31.96)–(-27.64) | (-8.49)–(-6.10) | (-6.46)–(-2.63) | (-33.68)–(-28.60) |

Smaller or more negative $\Delta G_{solv}$ values are associated with molecules that have a higher number of OH groups, indicating more sites available for strong solute-solvent interactions. For instance, Mono8, with seven OH groups, exhibits the most favorable $\Delta G_{solv}$ for dissolution. Among the molecules with more positive $\Delta G_{solv}$ values, Mono6 has only two OH groups, while Mono1 features a long alkyl chain that is less conducive to interactions with polar solvents. Similarly, conformers lacking IHBs or with fewer IHBs have $\Delta G_{solv}$ values more favorable for dissolution, as the OH groups not involved in IHBs can participate more readily in solute-solvent interactions.

The electrostatic ($G_{el}$) and dispersion components of $\Delta G_{solv}$ are negative across all three solvents, with their magnitudes being greater in water compared to chloroform and acetonitrile. For each molecule, $G_{el}$ is more pronounced for conformers without the first IHB, and it is especially large



© University of Venda

for Mono8, which has the most OH groups. The dispersion component tends to be larger in compounds with long R chains, with Mono1 showing the highest values.

There are notable differences between the DFT and HF results. It's important to remember that all potential solute-solvent interactions are non-covalent and heavily influenced by dispersion components, which DFT accounts for more effectively than HF.

### 5.5.2. Characteristics of intramolecular hydrogen bonds

A solvent can influence the IHBs by interacting with either the donor or the acceptor. Table S5.15 provides details on the various IHB parameters present in the calculated conformers across the three selected solvents, with DFT results reported. Table S5.16 presents the corresponding HF results. Table S5.17 compares the lengths of the IHBs in the four different media (vacuum and the three solvents). Figure S5.8 visually represents the comparison of IHB lengths across these media. Finally, Table 5.11 summarizes the ranges of IHB lengths in all four environments.

**Table 5.11.** Range of the lengths of the intramolecular hydrogen bonds in the calculated conformers of the considered ACPL molecules in chloroform, acetonitrile and water (respectively denoted as chlrf, actn, and aq in the column headings).
DFT/B3LYP/6-31+G(d,p) and HF/6-31G(d,p) results from full optimisation PCM calculations, respectively denoted as DFT and HF in the columns' headings. The molecules are denoted with the symbols listed in table 5.1.

| Molecules | Considered IHB | Method | Range of the length of the IHB (Å) | | | |
|---|---|---|---|---|---|---|
| | | | vac | chlrf | actn | aq |
| Mono1 | H15···O14 | DFT | 1.563–1.585 | 1.552–1.719 | 1.548–1.719 | 1.547–1.719 |
| | | HF | 1.692–1.719 | 1.679–1.708 | 1.677–1.703 | 1.677–1.702 |
| | | | | | | |
| Mono2 | H15···O14 | DFT | 1.545–1.546 | 1.538–1.539 | 1.536–1.536 | 1.535–1.536 |
| | | HF | 1.672–1.673 | 1.669–1.669 | 1.667–1.667 | 1.667–1.667 |
| | H17···O14 | DFT | 1.570–1.588 | 1.554–1.567 | 1.548–1.559 | 1.559–1.696 |
| | | HF | 1.696–1.733 | 1.686–1.718 | 1.682–1.712 | 1.682–1.711 |
| | | | | | | |
| Mono3 | H17···O14 | DFT | 1.545–1.565 | 1.536–1.550 | 1.533–1.545 | 1.532–1.544 |
| | | HF | 1.681–1.697 | 1.677–1.686 | 1.678–1.683 | 1.678–1.682 |
| | | | | | | |



© University of Venda

| | | | | | | |
|---|---|---|---|---|---|---|
| Mono4 | H15···O14 | DFT | 1.539–1.547 | 1.533–1.537 | 1.530–1.533 | 1.529–1.533 |
| | | HF | 1.665–1.671 | 1.660–1.664 | 1.658–1.661 | 1.665–1.671 |
| | H23···O32 | DFT | 1.615–1.654 | 1.602–1.644 | 1.596–1.640 | 1.595–1.640 |
| | | HF | 1.766–1.785 | 1.767–1.775 | 1.767–1.772 | 1.766–1.785 |
| | H25···O32 | DFT | 1.684–1.694 | 1.679–1.686 | 1.677–1.683 | 1.677–1.682 |
| | | HF | 1.800–1.814 | 1.803–1.812 | 1.803–1.811 | 1.800–1.814 |
| | | | | | | |
| Mono5 | H15···O14 | DFT | 1.536–1.541 | 1.529–1.535 | 1.527–1.530 | 1.527–1.530 |
| | | HF | 1.666–1.670 | 1.660–1.662 | 1.657–1.658 | 1.658–1.671 |
| | H23···O32 | DFT | 1.609–1.660 | 1.591–1.651 | 1.583–1.647 | 1.582–1.647 |
| | | HF | 1.747–1.760 | 1.743–1.754 | 1.740–1.752 | 1.747–1.758 |
| | H25···O32 | DFT | 1.627–1.682 | 1.606–1.669 | 1.598–1.663 | 1.597–1.663 |
| | | HF | 1.767–1.783 | 1.758–1.774 | 1.753–1.770 | 1.762–1.776 |
| | | | | | | |
| Mono6 | H15···O14 | DFT | 1.538–1.585 | 1.525–1.573 | 1.520–1.557 | 1.519–1.574 |
| | | HF | 1.663–1.722 | 1.663–1.713 | 1.661–1.708 | 1.661–1.707 |
| | H17···O14 | DFT | 1.563 | 1.548 | 1.543 | 1.542 |
| | | HF | 1.691 | 1.681 | 1.678 | 1.678 |
| | | | | | | |
| Mono7 | H15···O14 | DFT | 1.565–1.601 | 1.536–1.580 | 1.550–1.575 | 1.538–1.574 |
| | | HF | 1.703–1.746 | 1.689–1.740 | 1.684–1.736 | 1.683–1.735 |
| | H26···O14 | DFT | 1.923–1.979 | 1.863–1.885 | 1.850–1.860 | 1.849–1.858 |
| | | HF | 2.057–2.128 | 2.014–2.060 | 1.990–2.036 | 1.988–2.034 |
| | | | | | | |
| Mono8 | H15···O14 | DFT | 1.627–1.656 | 1.619–1.639 | 1.615–1.633 | 1.615–1.632 |
| | | HF | 1.808–1.856 | 1.809–1.837 | 1.809–1.861 | 1.809–1.861 |
| | H32···O10 | DFT | 2.535–2.632 | 2.562–2.681 | 2.559–2.691 | 2.558–2.692 |
| | | HF | 2.535–2.632 | 2.540–2.633 | 2.538–2.641 | 2.537–2.642 |
| | H32···O28 | DFT | 2.416–2.430 | 2.434–2.444 | 2.435–2.445 | 2.435–2.445 |
| | | HF | 2.441–2.452 | 2.445–2.455 | 2.447–2.452 | 2.447–2.452 |
| | H33···O27 | DFT | 2.353–2.402 | 2.383–2.422 | 2.396–2.441 | 2.397–2.428 |
| | | HF | 2.371–2.408 | 2.395–2.428 | 2.404–2.448 | 2.405–2.450 |
| | H33···O29 | DFT | 2.183–2.402 | 2.192–2.239 | 2.192–2.248 | 2.192–2.249 |
| | | HF | 2.209–2.408 | 2.215–2.241 | 2.215–2.249 | 2.215–2.249 |
| | H34···O28 | DFT | 2.202–2.219 | 2.239–2.249 | 2.255–2.261 | 2.256–2.262 |
| | | HF | 2.227–2.233 | 2.258–2.261 | 2.269–2.272 | 2.270–2.273 |
| | H34···O31 | DFT | 1.898–2.219 | 1.871–1.926 | 1.866–1.907 | 1.865–1.906 |
| | | HF | 1.947–2.233 | 1.941–2.015 | 1.943–2.009 | 1.944–2.009 |
| | H35···O29 | DFT | 1.992–2.617 | 2.018–2.501 | 2.013–2.522 | 2.013–2.524 |
| | | HF | 2.036–2.429 | 1.941–2.261 | 1.943–2.272 | 1.944–2.273 |
| | H17···π (C13) | DFT | 2.172–2.193 | 2.170–2.315 | 2.168–2.322 | 2.300–2.323 |
| | | HF | 2.301–2.638 | 2.300–2.323 | 2.295–2.315 | 2.300–2.323 |



© University of Venda

In many of these compounds, the length of the IHBs tends to decrease as the polarity of the medium increases. This is observed in bonds such as H17···O14, H15···O14, H23···O32, H26···O32, H26···O14, and H30···O31. Except for O31, all these bonds involve an $sp^2$ oxygen as the acceptor, making them stronger and less susceptible to solvent weakening. Additionally, the area around these bonds is hydrophobic [5], reducing the likelihood of disruption by water molecules. In contrast, the lengths of H26···O10, H26···O27, H28···O25, H28···O29, H30···O27, and H32···O29 tend to increase with greater medium polarity. Another observation is that IHB bond lengths are more similar in acetonitrile and water than in chloroform or *in vacuo*, although HF results present some exceptions to these trends.

Moreover, Tables S5.15, S5.16, and S5.17 analyse O–H···π interactions with the B aromatic ring, where the H17···C13 distance serves as an indicator of how the OH group approaches the π cloud. This distance increases with medium polarity.

### 5.5.3. HOMO-LUMO energy gap and dipole moment of the conformers.

The presence of a solvent can influence the energies of the HOMO and LUMO in different ways, which in turn impacts the HOMO–LUMO energy gap [25, 26]. The extent of this influence differs based on the molecular system and the nature of the solvent.

Table S5.18 provides the HOMO–LUMO energy gap values for the calculated conformers of the selected ACPLs in the four media examined, as obtained from DFT calculations, while Table S5.19 shows the corresponding values from HF results. Figure S5.9 graphically compares these values across the four media, with separate plots for DFT and HF due to the differences in scales.

The trends across the four media are quite similar for the conformers of the same molecule. The conformers with the minimum and maximum gap values remain consistent across all media. Some specific patterns are evident for certain molecules. For example, the HOMO–LUMO energy gap values *in vacuo* for most conformers of Mono7 and Mono8 are much smaller than in solution and tend to increase slightly with higher medium polarity. Conversely, for all conformers of Mono3,



© University of Venda

the gap values *in vacuo* are larger than in solution, decreasing as the medium's polarity increases. The trends for other molecules' conformers are less consistent.

As already known, solvents with differing polarities influence the dipole moment of solute molecules differently. Table S5.20 presents the dipole moment values for the calculated conformers of selected ACPLs in the four media based on DFT results, while Table S5.21 shows the corresponding HF results. Figure S10 provides a graphical representation of these trends.

The findings align with typical behaviours, where the dipole moment of a conformer generally increases with the medium's polarity. However, there are some exceptions, particularly in the HF results for compound Mono8, where some values *in vacuo* are higher than those in solution. The most significant increases in dipole moment are observed for molecule Mono5.

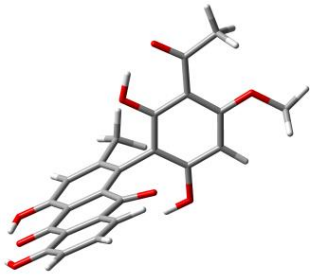
**5.6. Molecular docking results.**

In this study, eight ACPL molecules and eleven selected molecular targets (proteins) were used for docking analyses. Among these targets, eight are associated with anticancer activity, while three are linked to antimalarial properties. Their names, PDB IDs, and brief descriptions are provided in section 4.6 Each ACPL was docked with its corresponding target, and the resulting binding energies and intermolecular interactions were evaluated and listed in Table 5.12, along with those of the co-crystal ligands for reference comparison. The docking studies shows that several ACPL molecules, particularly Mono4, Mono5, Mono7, and Mono8 fig 5.7, show promising binding affinities across various cancer and antimalarial targets. Mono4 stands out for its strong interactions with critical cancer-related proteins such as EGFR, JAK3, HER2, CDK-2, and Topo I, indicating its potential to disrupt multiple cancer pathways, particularly in HER2-positive cancers. Mono8 also shows a dual application, displaying high binding affinities not only for cancer targets like EGFR, HER2, and CDK-2 but also for antimalarial targets such as PFLDH and PFPMT, suggesting it could be effective against Plasmodium parasites. Mono5 demonstrates selective efficacy with high affinities toward P13K and CDK-2, making it a promising candidate for cancers involving these pathways, such as certain leukemias and solid tumors. Mono7, though showing more moderate binding scores, reveals potential for antimalarial applications, especially when combined with other agents to enhance efficacy against drug-resistant strains. The co-

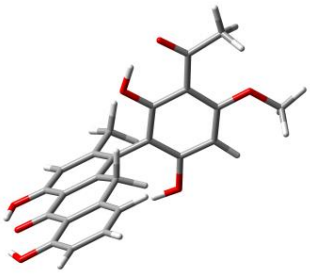


© University of Venda

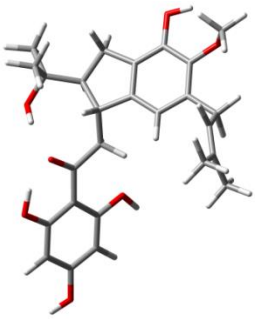
crystallized ligands, serving as benchmarks, confirm that these ACPL molecules approach or even surpass binding scores on targets like HER2 and P13K, particularly Mono4 and Mono5. Overall, the results suggest that Mono4 and Mono8 could serve as versatile candidates across cancer and antimalarial therapies, while Mono5 and Mono7 offer targeted opportunities in oncology and malaria, respectively, meriting further studies in pharmacokinetics and efficacy.



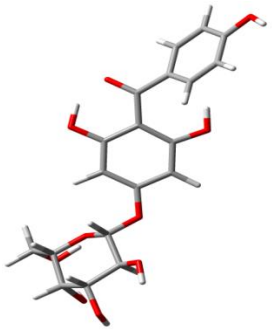
Mono4



Mono5



Mono7



Mono8

**Figure 5.7.** Optimized ACPL molecules with promising biochemical interactions: DFT results *in vacuo*.

**Table 5.12**. Binding score energy (kcal $mol^{-1}$) for the docking of the ACPL molecules considered in this work against the selected biochemical targets.

Results from the QickProp tool (Schrödinger suite).
When the target biomolecule contains more than one active site, they are indicated under the name of the protein.



© University of Venda

(-) indicates the absence of detectable interactions between the biological target and the considered ACPL molecule.

| Molecule considered | Anticancer targets considered | | | | | |
|---|---|---|---|---|---|---|
| | EGFR | P13K | | JAK3 | CDK-2 | Topo I |
| | | (B1) | (B2) | | | |
| | | | | | | |
| Mono1 | -4.142 | -5.183 | – | -4.799 | -6.870 | -8.116 |
| Mono2 | -7.405 | -6.509 | -5.189 | -6.494 | -7.549 | -7.903 |
| Mono3 | -6.762 | -5.768 | -5.233 | -6.393 | -6.779 | -8.497 |
| Mono4 | -8.013 | -6.790 | -6.649 | -7.619 | -6.621 | -9.008 |
| Mono5 | -7.502 | -10.250 | -6.691 | -8.544 | -7.866 | -9.928 |
| Mono6 | -6.529 | -5.545 | -4.445 | -6.024 | – | -7.367 |
| Mono7 | -6.675 | -7.385 | -5.276 | -7.275 | -7.426 | -8.782 |
| Mono8 | -7.290 | -7.994 | -7.224 | -9.814 | -8.491 | -8.412 |
| Co-crystilised ligands | -11.256 | -8.282 | -5.810 | -9.261 | -9.142 | -7.916 |

| Anticancer targets considered | | | | | |
|---|---|---|---|---|---|
| | H5P90 | | BRAF V600b | | |
| | (G1) | (G2) | (C1) | (C2) | (C3) |
| | | | | | |
| Mono1 | -5.500 | -5.582 | -8.085 | -8.215 | -4.626 |
| Mono2 | -5.898 | -6.116 | -8.744 | -8.433 | -6.853 |
| Mono3 | -5.924 | -5.614 | -9.167 | -8.676 | -4.772 |
| Mono4 | -6.570 | -7.038 | -11.013 | -10.770 | -5.683 |
| Mono5 | -7.365 | -7.269 | -10.832 | -9.133 | -5.836 |
| Mono6 | -5.886 | -5.917 | -8.583 | -7.896 | -5.410 |
| Mono7 | -7.209 | -6.903 | -10.863 | -10.862 | -6.038 |
| Mono8 | -9.458 | -8.944 | -9.585 | -9.087 | -6.596 |
| Co-crystilised ligands | -8.997 | -8.821 | -13.710 | -13.682 | -3.355 |

| | Anticancer targets considered | | | | | |
|---|---|---|---|---|---|---|
| | HER2 | | | | | |
| | (H1) | (H2) | (H3) | (H4) | (H5) | (H6) |
| | | | | | | |
| Mono1 | -5.918 | – | – | -4.398 | -4.984 | – |
| Mono2 | -6.602 | – | – | – | -6.801 | – |
| Mono3 | -6.338 | – | – | – | -6.558 | – |
| Mono4 | -10.779 | -5.133 | -4.978 | -4.763 | -10.611 | -5.756 |
| Mono5 | -9.927 | – | -5.837 | -5.147 | -10.500 | -5.514 |



© University of Venda

| | | | | | | |
|---|---|---|---|---|---|---|
| Mono6 | -8.061 | – | – | -4.823 | -6.801 | – |
| Mono7 | -7.355 | -4.902 | -5.955 | -5.412 | -7.618 | – |
| Mono8 | -10.350 | -6.200 | -5.004 | – | -10.579 | -5.521 |
| Co-crystilised ligands | -12.033 | -7.582 | -3.699 | -6.974 | -12.083 | -7.443 |

| | Antimalarial targets considered | | | | | | |
|---|---|---|---|---|---|---|---|
| | PFLDH | PFMDH | | | | | PFPMT |
| | | (M1) | (M2) | (M3) | (M4) | (M5) | |
| | | | | | | | |
| Mono1 | -4.680 | -4.561 | -6.491 | -6.604 | – | -5.838 | – |
| Mono2 | -4.621 | -4.616 | -6.495 | -4.958 | -5.200 | -5.419 | – |
| Mono3 | -4.705 | -6.842 | -7.027 | -6.483 | -5.657 | -5.737 | – |
| Mono4 | -6.333 | – | – | -5.821 | – | – | |
| Mono5 | -6.639 | – | – | -5.748 | – | – | |
| Mono6 | -4.750 | -6.822 | -6.128 | -5.465 | -5.713 | -5.686 | -6.168 |
| Mono7 | -6.891 | -5.193 | -6.143 | -7.264 | -5.891 | – | – |
| Mono8 | -8.209 | -5.840 | -8.030 | -7.533 | -5.670 | -5.851 | – |
| Co-crystilised ligands | -4.724 | -7.312 | -7.184 | -9.633 | -7.312 | -7.312 | -10.834 |

The following sections provide a detailed discussion of the types of molecular interactions exhibited by ACPL molecules with the highest docking scores for each biochemical target. These interactions include hydrogen bonding, hydrophobic contacts, π-π stacking, and electrostatic interactions, each playing a crucial role in the stability and specificity of ligand-target binding. By examining the top-scoring ACPL-target complexes, we gain insights into the key structural and chemical features responsible for binding affinity, helping to identify the most promising candidates for therapeutic applications. Each section will address these interactions in the context of specific protein targets, including EGFR, HER2, CDK-2, P13K, Topo I, and the antimalarial enzymes PFLDH and PFPMT.

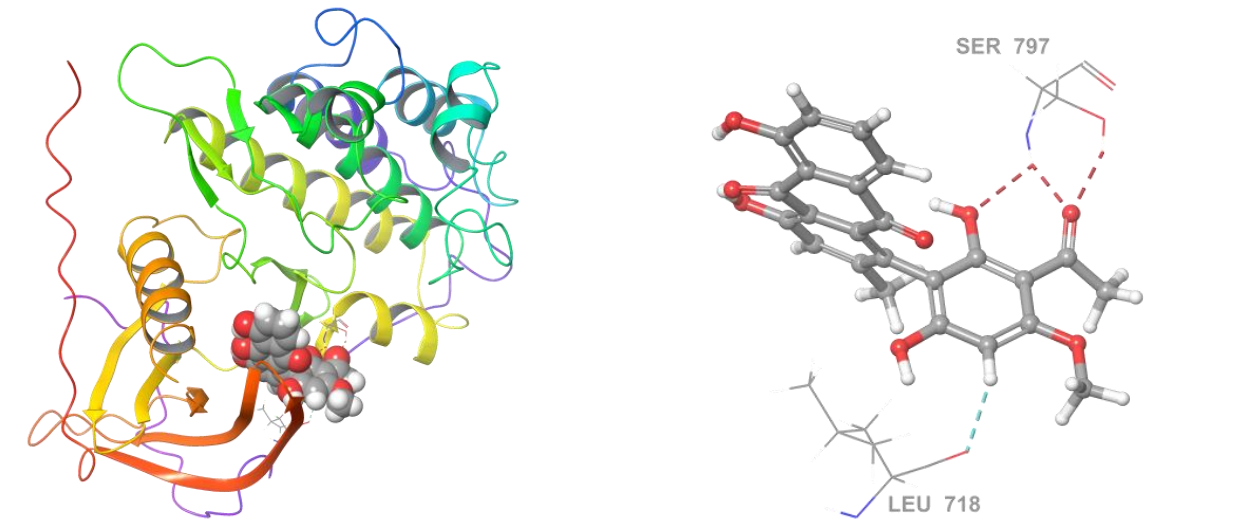
### 5.6.1. Molecular interaction of considered compounds with EGFR

The molecular docking results of EGFR against the selected compounds show that all considered molecules can interact at the active site (Table S5.17). Mono4 emerges as the top-docked compound, showing a binding energy of -8.013 kcal $mol^{-1}$. Despite not showing a stronger interaction with EGFR than the co-crystal ligand, Mono4 engages in two HBs with SER797 (Fig.



© University of Venda

**5.8.1**) and shows hydrophobic interactions with LEU718 and ALA743. This suggests a promising potential for Mono4 as a potential inhibitor of EGFR.




A b

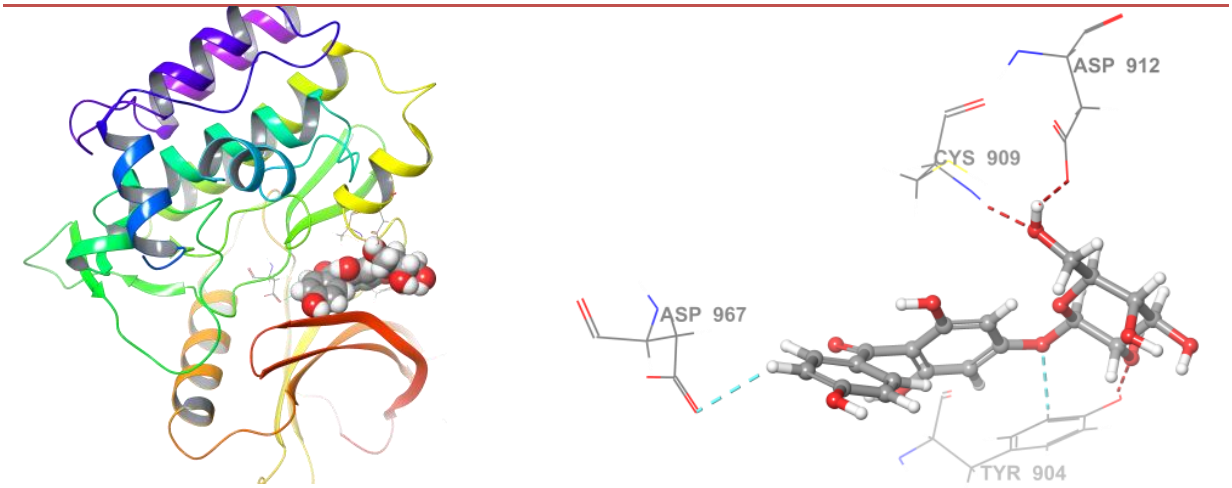
**Figure 5.8.1.** Molecular insight into compound Mono4 docked on the lung cancer target (EGFR, PDB: 6LUD), highlighting the residues within the binding site pocket. Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, and light blue-dashed segments denote aromatic HBs.

### 5.6.2. Molecular interaction of considered compounds with JAK3

The molecular docking results of JAK3 against the selected compounds show that all considered molecules can interact at the active site. Mono8 emerges as the best-docked compound, showing a binding energy of -9.814 kcal $mol^{-1}$. Remarkably, Mono8's interaction with JAK3 is comparable to that of the co-crystal ligand. Mono8 engages in three HBs with ASP912, DYS909, and TYR904, along with two aromatic HBs with ASP967 and TYR904 (Fig. **5.8.2**). Additionally, it shows hydrophobic interactions with ALA853, LEU828, VAL836, and CYS909. These findings suggest that Mono8 holds promising potential as a JAK3 inhibitor.



© University of Venda




a b

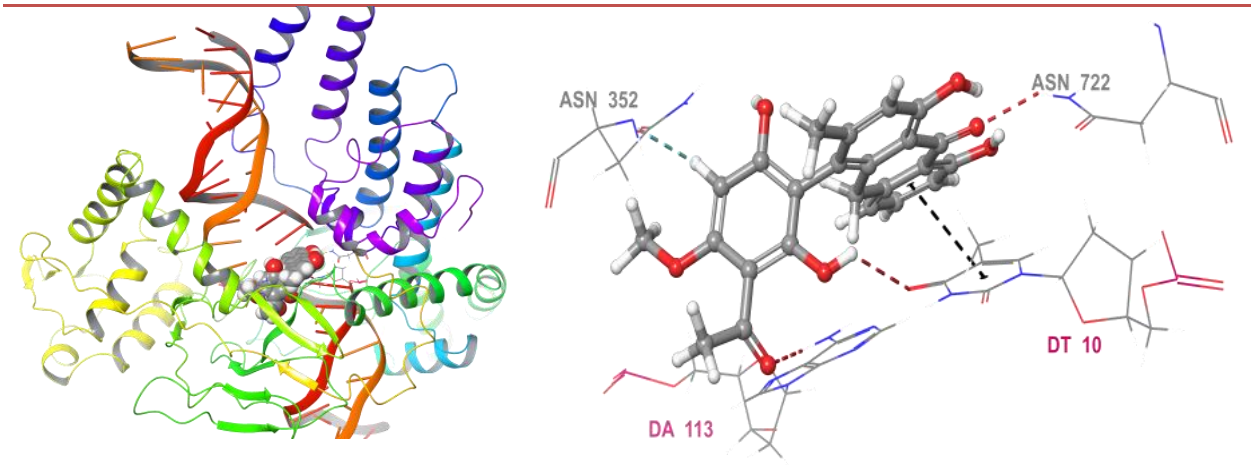
**Figure 5.8.2.** Molecular insight into compound Mono8 docked on the lung cancer target (JAK3, PDB: 7C3N), highlighting the residues within the binding site pocket. Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, and light blue-dashed segments denote aromatic HBs.

### 5.6.3. Molecular interaction of considered compounds with Topo I

The molecular docking results of Topo I against the selected compounds show that all considered molecules can interact at the active site. Mono5 emerges as the best-docked compound, showing a binding energy of -9.928 kcal $mol^{-1}$. Mono5's interaction with Topo I is notably stronger than that of the co-crystal ligand. It engages in three HBs with ASN A:722, DT B:10, and DA D:112, along with aromatic HBing with ASN352, π-π stacking with DT B:10 (Fig. **5.8.3**)., and polar interactions with ASN A:352, THR A:718, and ASN A:352 These findings highlight Mono5's potential as a promising Topo I inhibitor.



© University of Venda




a b

**Figure 5.8.3.** Molecular insight into compound Mono5 docked on the lung cancer target (Topo I, PDB: IT8I), highlighting the residues within the binding site pocket. Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; the red-dashed segments denote HBs, light blue-dashed segments denote aromatic HBs, and black-dashed segments denote π-π stacking interactions.

### 5.6.4. Molecular interaction of considered compounds with P13K

The molecular docking results of PI3K against the selected compounds indicate that most of the considered molecules bind to site B1, with a few binding to site B2. Mono5 emerges as the best-docked compound against PI3K on-site B1, showing a binding score energy of -10.250 kcal $mol^{-1}$. Mono5's interaction with PI3K is comparatively stronger than that of the co-crystal ligand. It engages in three HBs with VAL882, ASP950, and TYR867 and an aromatic HB with TRP 812. Additionally, Mono5 shows hydrophobic interactions with PRO810, TYP812, ILE879, ILE881, VAL882, and ALA885 (Fig. **5.8.4**). These findings suggest Mono5's potential as a promising PI3K inhibitor targeting site B1.



© University of Venda

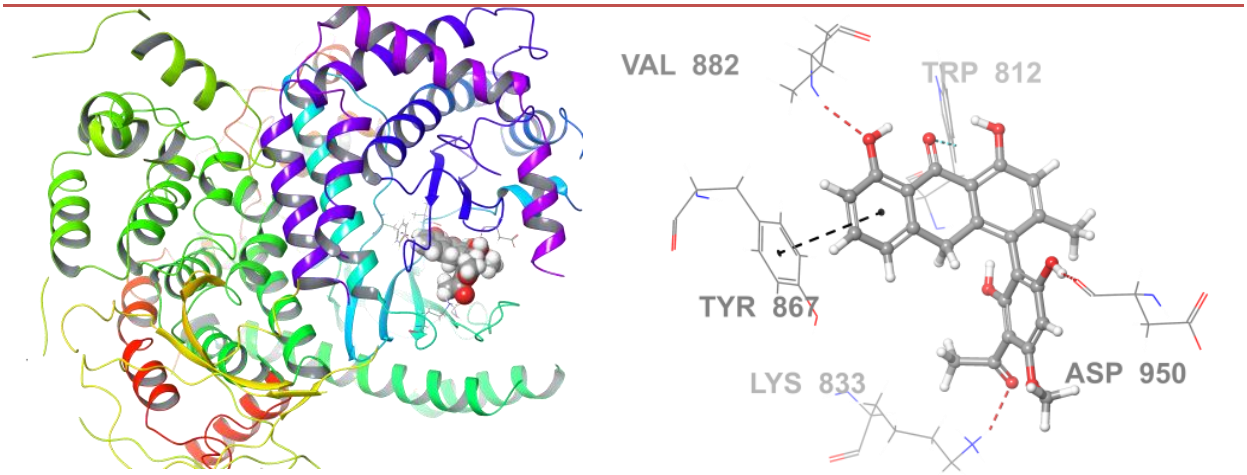
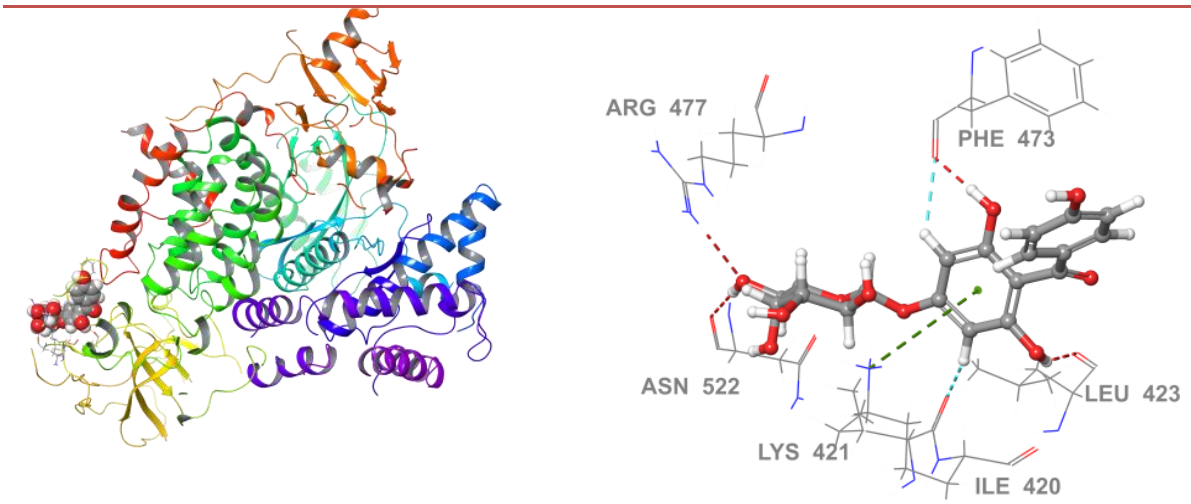



**Figure 5.8.4.** Molecular insight into compound Mono5 docked on the lung cancer target (P13K, PDB: 5JHB), highlighting the residues within the binding site pocket (site B1). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; the red-dashed segments denote HBs, light blue-dashed segments denote aromatic HBs, and black-dashed segments denote π-π stacking interactions.

The molecular docking results of PI3K against the selected compounds indicate that Mono8 is the best-docked compound against PI3K on-site B2, showing a binding score energy of -7.224 kcal $mol^{-1}$. Mono8's interaction with PI3K is also comparatively stronger than that of the co-crystal ligand. Mono8 engages in three HBs with VAL882, LYS833, and ASP950 and an aromatic HB with TRP 812. Additionally, it shows π-π stacking interaction with TYR867 and hydrophobic interactions with VAL882, ILE881, ILE879, and TYR867 (Fig. **5.8.5**). These findings suggest that Mono8 holds promise as a potential PI3K inhibitor targeting site B2.



© University of Venda




a b

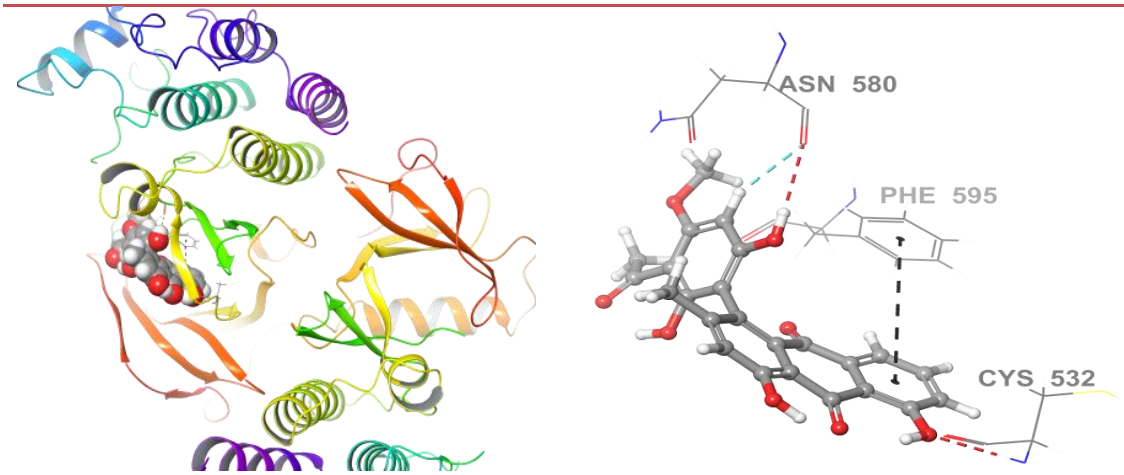
**Figure 5.8.5.** Molecular insight into compound Mono8 docked on the lung cancer target (P13K, PDB: 5JHB), highlighting the residues within the binding site pocket (site B2). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, light blue-dashed segments denote aromatic HBs and green-dashed segments denote pi-cation interactions.

### 5.6.5. Molecular interaction of considered compounds with BRAF V600B

The docking analysis of BRAF V600B shows three identified docking sites denoted as C1, C2, and C3. The ligand-protein complexes show better binding scores on site C1 than on sites C2 and C3. C1 shows the best docking performance among the sites with Mono4, showing a binding score energy of -11.013 kcal $mol^{-1}$. While this interaction is not comparatively stronger than the co-crystal ligand, it involves two HBs with ASN580 and CYS532 and an aromatic HB with ASN580. Additionally, there is π-π stacking with PHE595 and hydrophobic interactions via ILE463 (Fig. **5.8.6**). These findings suggest that Mono4 has potential as a BRAF V600B inhibitor targeting site C1.



© University of Venda



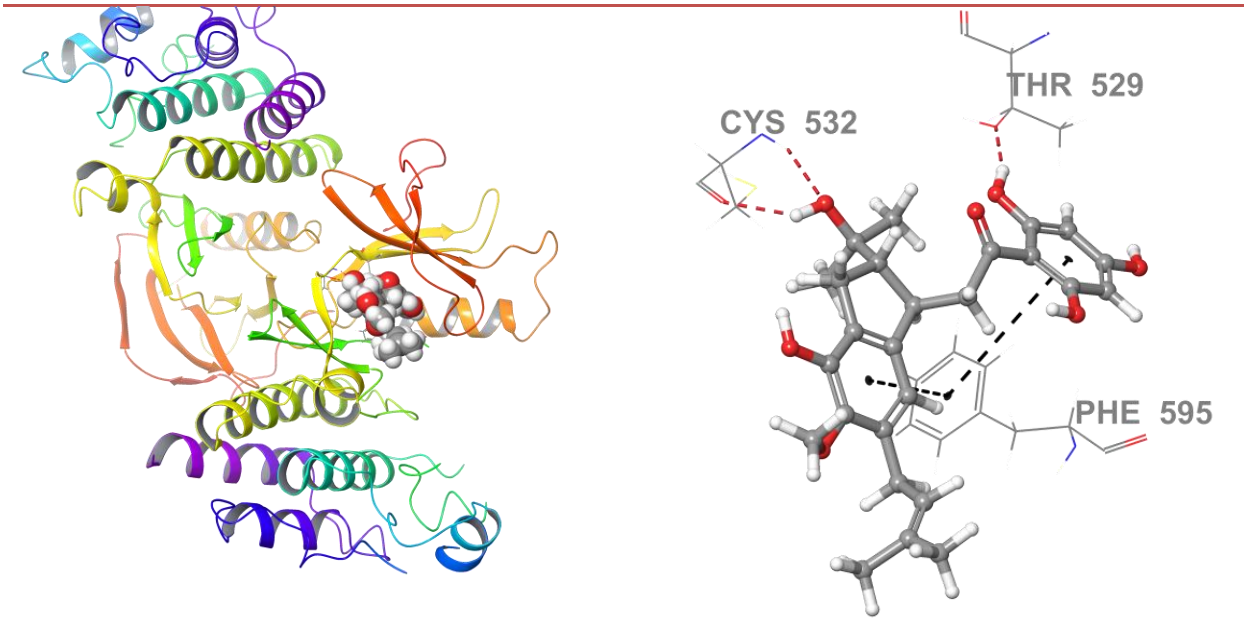

a b

**Figure 5.8.6.** Molecular insight into compound Mono4 docked on the lung cancer target (BRAF V600b, PDB; 6V34), highlighting the residues within the binding site pocket (site C1). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, light, blue-dashed segments denote aromatic HBs and black-dashed segments denote π-π stacking interactions.

The C2 site shows the best docking complex with Mono7, demonstrating a binding score energy of -10.862 kcal $mol^{-1}$. Although this interaction is not comparatively stronger than the co-crystal ligand, it involves a HB with CYS532 and THR529. Additionally, there is π-π stacking with PHE595 and hydrophobic interactions via ILE527, TRP531, CYS532, PHE595, and ILE463 (Fig. **5.8.7**). These findings suggest that Mono7 has potential as a BRAF V600B inhibitor targeting site C2.



© University of Venda



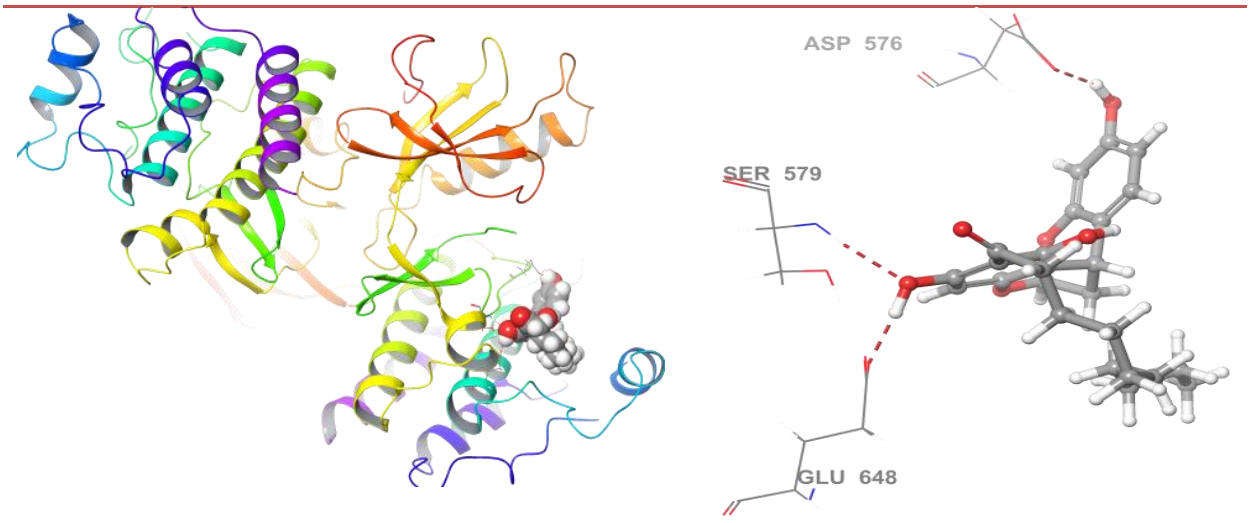

a b

**Figure 5.8.7.** Molecular insight into compound Mono7 docked on the lung cancer target (BRAF V600b, PDB; 6V34), highlighting the residues within the binding site pocket (site C2). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, and black-dashed segments denote π-π stacking interactions.

The C3 site shows the best docking performance with Mono6, showing a binding score energy of -6.853 kcal $mol^{-1}$. Notably, this interaction is comparatively stronger than the co-crystal ligand and involves three HBs with ASP576, SER579, and GLU648. Additionally, there are hydrophobic interactions via VAL654, ILE644, ALA641, and TYR640 (Fig. **5.8.8**). These findings suggest that Mono6 has significant potential as a BRAF V600B inhibitor targeting site C3.



© University of Venda




a b

**Figure 5.8.8.** Molecular insight into compound Mono6 docked on the lung cancer target (BRAF V600b, PDB; 6V34), highlighting the residues within the binding site pocket (site C3). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

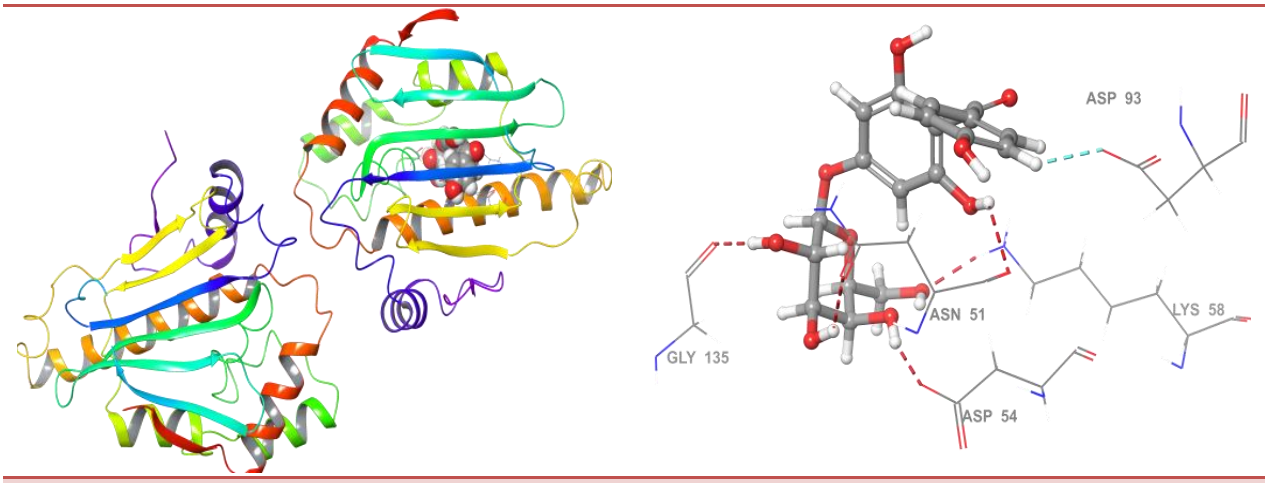
### 5.6.5. Molecular interaction of considered compounds with H5P90

Most of the considered molecules bind to both active sites, but Mono8 shows the best docking performance in both sites, with binding score energies of -9.458 kcal $mol^{-1}$ and -8.944 kcal $mol^{-1}$ for sites G1 and G2, respectively.

At active site G1, the interaction of Mono8 with H5P90 is relatively stronger than the co-crystal ligand. Mono8 engages in five HBs with LYS58, ASP54, ASN51, and GLY135, along with an aromatic HB with ASP93 (Fig. **5.8.9**). Additionally, there are polar interactions with ASN106 and ASN51 and hydrophobic interactions with ILE91 and LEU107. These findings suggest that Mono8 has significant potential as a ligand for both active sites, particularly at site G1.



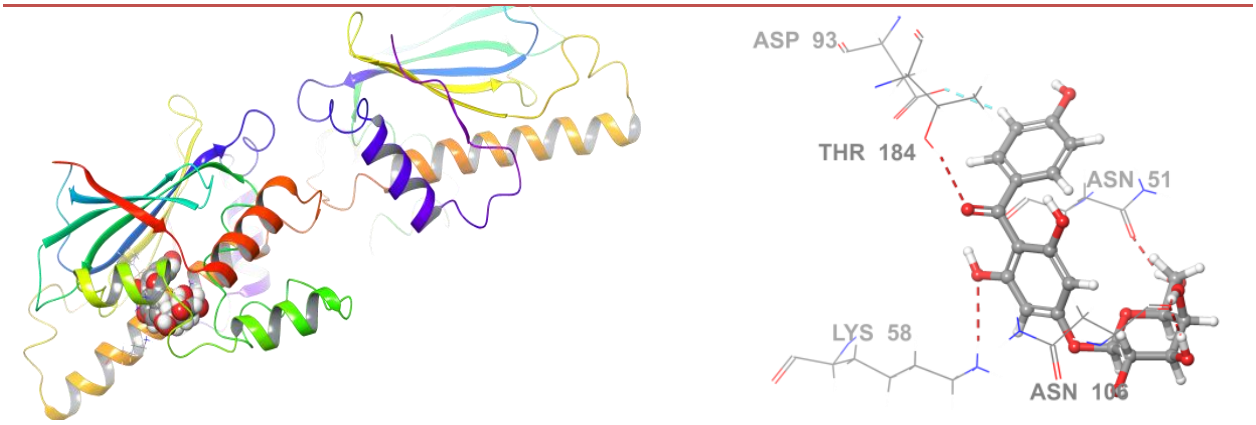
© University of Venda




a b

**Figure 5.8.9.** Molecular insight into compound Mono8 docked on the breast cancer target (H5P90, PDB: 3TUH), highlighting the residues within the binding site pocket (site G1). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

At active site G2, the interaction of Mono8 with H5P90 is relatively closer to one of the co-crystal ligand. It interacts through five HBs with ASN106, ASN51, LYS58, THR184, one aromatic H-bond with ASP93 (Fig. **5.8.10**), polar interactions with SER113, LYS112, ALA111, ILE110, LEU107, ASN106, TYR139, PHE138, LEU48 and hydrophobic interactions with ASN51, SER52, ALA55, LYS58, MET98.




a b

**Figure 5.8.10.** Molecular insight into compound Mono8 docked on the breast cancer target (H5P90, PDB: 3TUH), highlighting the residues within the binding site pocket (site G2). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the



© University of Venda

ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

### 5.6.6. Molecular interaction of considered compounds with HER2

All the selected compounds successfully bind to active sites H1 and H5 with favorable binding score energies. However, for active sites H2, H3, H4, and H6, only a few compounds interact with these sites. The best-docked complexes from each active site are as follows: Mono4/HER2 (H1) with a binding score energy of -10.779, Mono8/HER2 (H2) with -6.200, Mono7/HER2 (H3) with -5.955, Mono7/HER2 (H4) with -5.412, Mono4/HER2 (H5) with -10.611, and Mono4/HER2 (H6) with -5.756. These results suggest that while all compounds show strong binding affinity to active sites H1 and H5, there is inconsistency in their interactions with sites H2, H3, H4, and H6. This inconsistency may affect the design of specific inhibitors targeting these different active sites.

At active site H1, the interaction between Mono4 and HER2 is not stronger than that of the co-crystal ligand. Mono4 forms two HBs with MET801 and ARG849 and two aromatic HBs with GLN799 and ASN85 (Fig. **5.8.11**). Additionally, there are glycine interactions with GLY727, SER728, GLY729, PHE731, GLY732, THR733, VAL734, ARG849, GLY804, and CYS805. Furthermore, Mono4 shows hydrophobic interactions with ASN850, LEU852, SER783, THR798, GLN799, LEU800, MET801, and ALA751.


**Figure 5.8.11.** Molecular insight into compound Mono4 docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H1). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.



© University of Venda

At active site H2, the interaction between Mono8 and HER2 is not stronger than that of the co-crystal ligand. Mono8 forms five HBs with ARG713, GLU744, GLN799, LYS860, and GLY778, as well as one aromatic HB with GLU744 (Fig. **5.8.12**). Additionally, there are glycine interactions with GLY778, polar interactions with GLN799, and hydrophobic interactions with PRO780, LYS860, and HIE858.

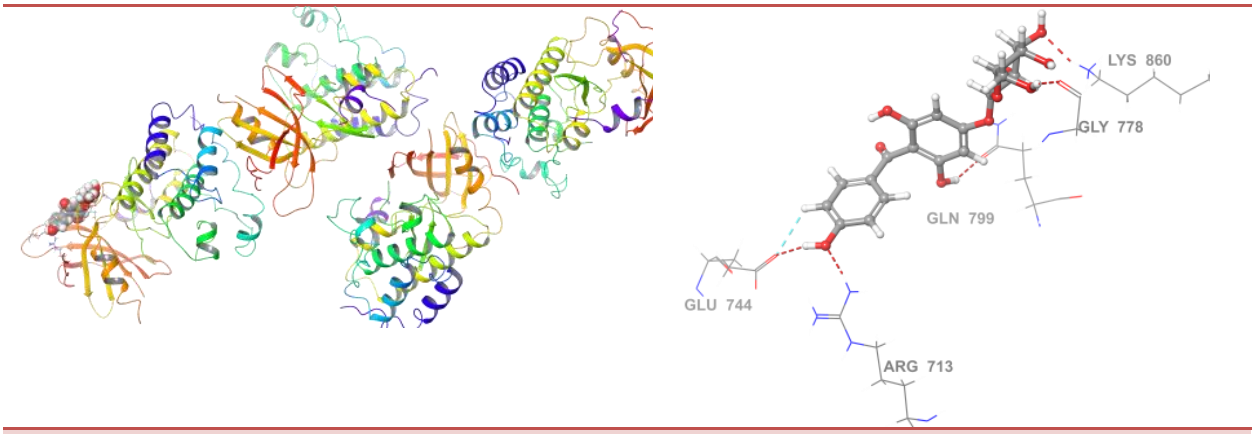
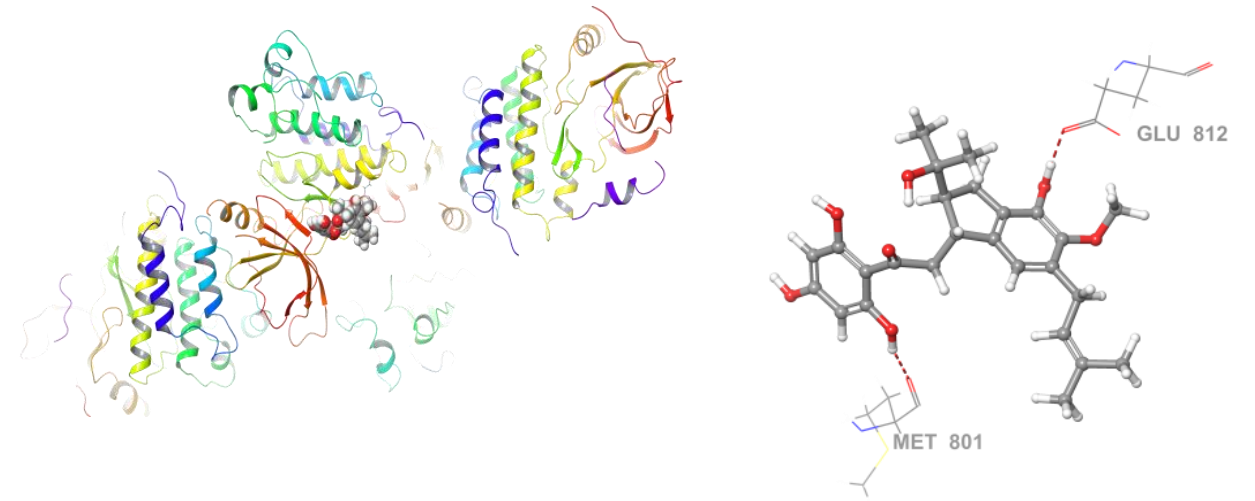



**Figure 5.8.12.** Molecular insight into compound Mono8 docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H2). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

At active site H3, the interaction between Mono7 and HER2 is comparatively stronger than that of the co-crystal ligand. Mono7 forms two HBs with MET801 and GLU812 (Fig. **5.8.13**) and glycine interactions with MET801 and LEU800 residues.



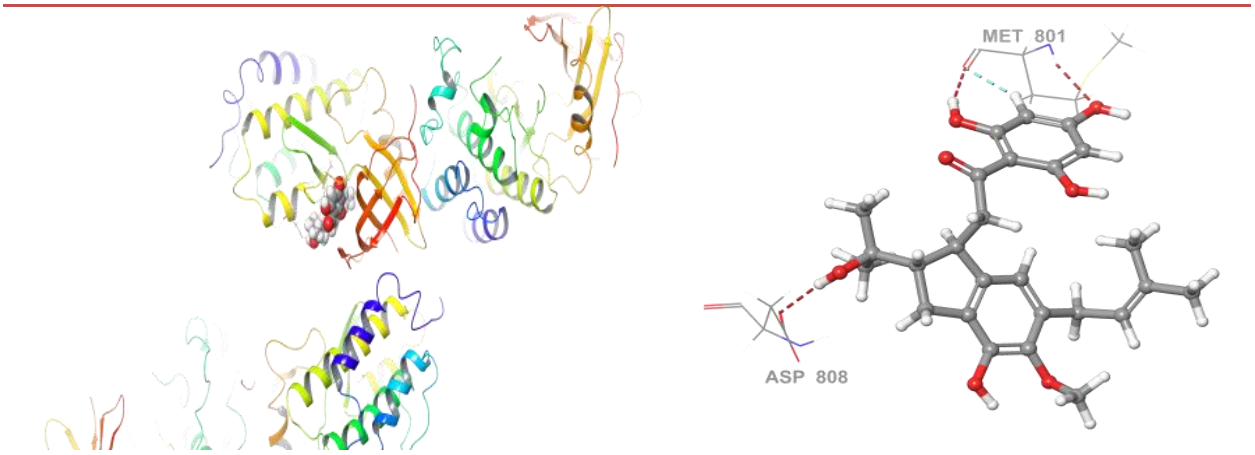
© University of Venda




a b

**Figure 5.8.13.** Molecular insight into compound Mono7 docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H3). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

At active site H4, the interaction between Mono7 and HER2 is not stronger than that of the co-crystal ligand. Mono7 forms three HBs with MET801 and ASP808 and one aromatic HB with MET801 (Fig. **5.8.14**). Additionally, there are polar interactions with GLN799, THR798, and THR862 and hydrophobic interactions with ASP808, CYS805, MET801, and LEU800.




a b

**Figure 5.8.14.** Molecular insight into compound Mono7 docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H4). Panel a) showcases the docking



© University of Venda

pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

At active site H5, the interaction between Mono4 and HER2 is not stronger than that of the co-crystal ligand. Mono4 forms two HBs with MET801 and ARG849 and one aromatic HB with ASN850 (Fig. **5.8.15**). Additionally, there are hydrophobic interactions with MET801, LEU800, GLN799, THR798, LEU726, and glycine interactions with CYS805, GLY804, GLY727, SER728, GLY729, and ALA730.

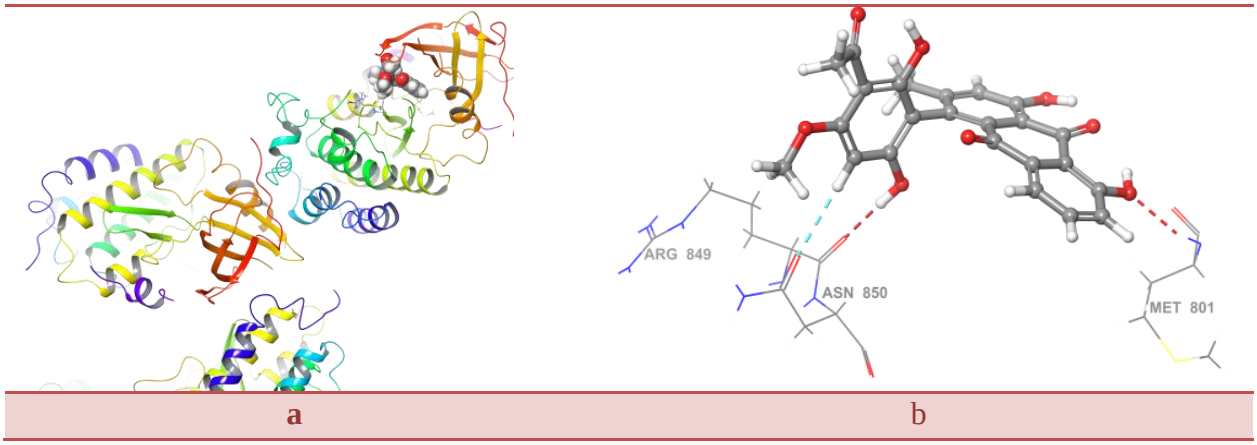
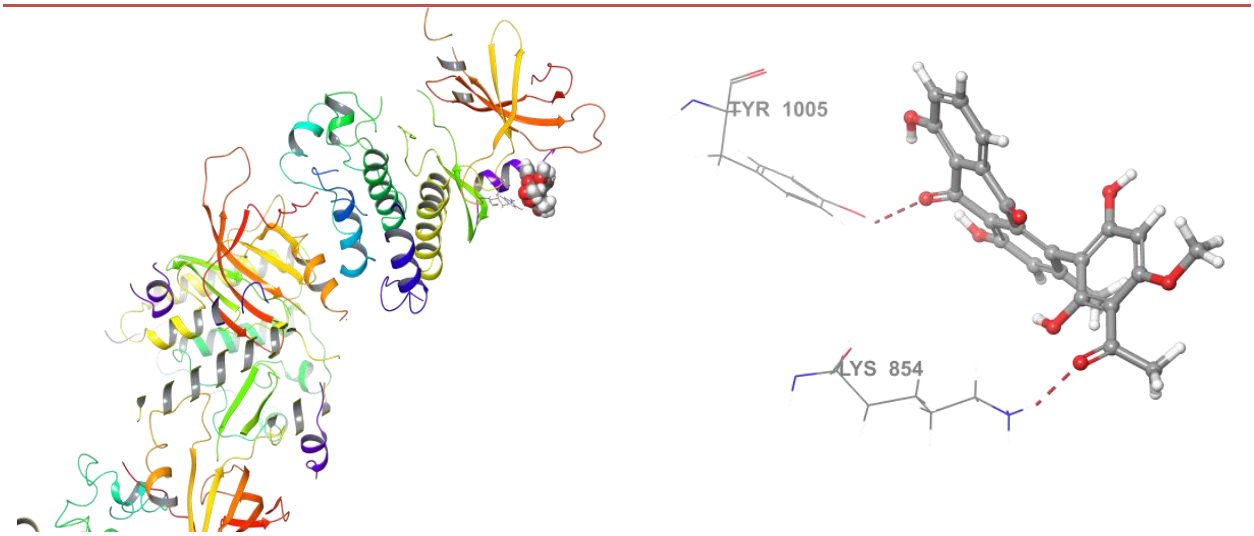



**Figure 5.8.15.** Molecular insight into compound Mono4 docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H5). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

At active site H6, the interaction between Mono4 and HER2 is not stronger than that of the co-crystal ligand. It forms two HBs with TYR1005 and LYS854 (Fig. **5.8.16**). Additionally, there are hydrophobic interactions with TYR1005, LEU1009, and LYS854.



© University of Venda




a b

**Figure 5.8.16.** Molecular insight into compound Mono4 docked on the breast cancer target (HER2, PDB: 3RCD), highlighting the residues within the binding site pocket (site H6). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

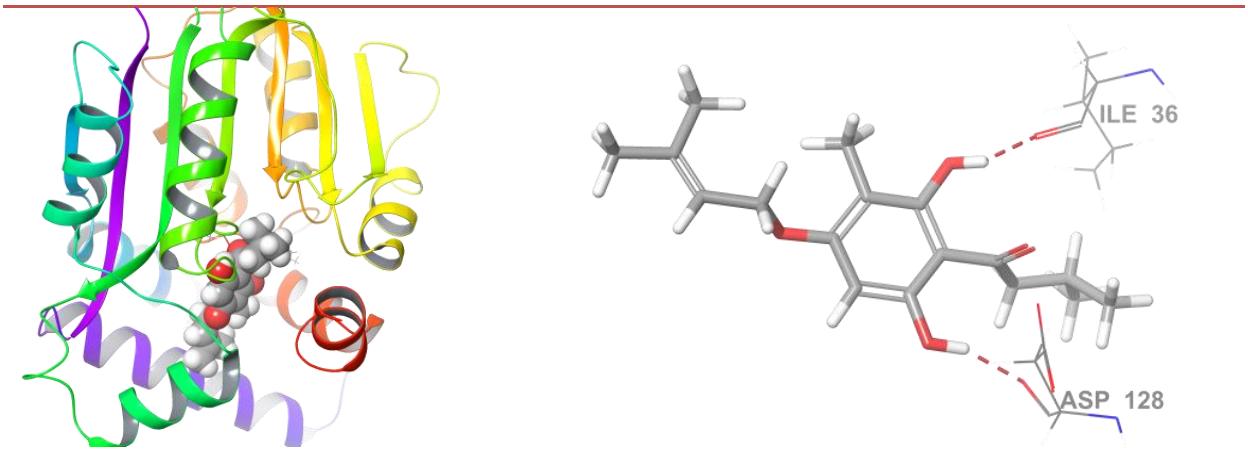
### 5.6.7. Molecular interaction of considered compounds with PFPMT

The molecular docking results of PFPMT against the selected compounds indicate that, apart from Mono3, the considered molecules do not show interactions with PFPMT. However, Mono3 shows binding with PFPMT, with a binding score energy of -6.168 kcal $mol^{-1}$. Nevertheless, its interaction with PFPMT is not notably stronger than that of the co-crystal ligand.

Mono3 forms two HBs, one with ILE36 and the other with ASP128 (Fig. **5.8.17**). Additionally, hydrophobic interactions occur with GLY63, GLY65, ILE90, ASP85, ILE36, TYR27, and LEU240. Furthermore, glycine interactions occur with GLY243 and TRP244.



© University of Venda




a b

**Figure 5.8.17.** Molecular insight into compound Mono3 docked on the antimalarial target (PFPMT, PDB: 3UJ9), highlighting the residues within the binding site pocket. Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

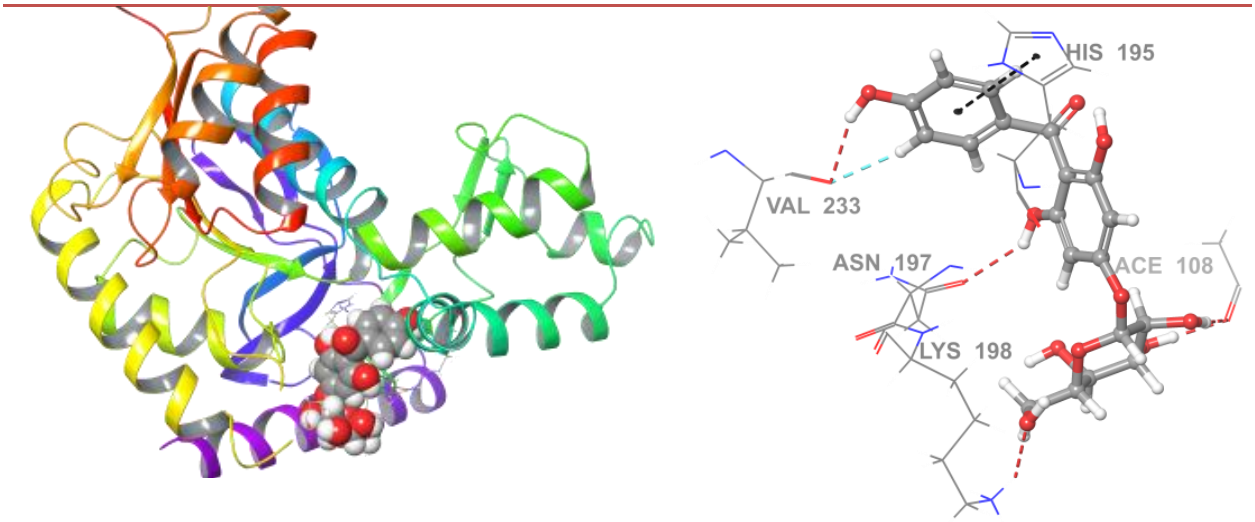
### 5.6.8. Molecular interaction of considered compounds with PFLDH

The molecular docking results of PFLDH against the selected compounds show that all the considered molecules are capable of interacting at the active site. Mono8 emerges as the best-docked compound against PFLDH, showing a binding energy of -8.209 kcal $mol^{-1}$. Notably, its interaction with PFLDH is comparatively stronger than that of the co-crystal ligand.

Mono8 forms five HBs with LYS198, ASN197, ACE108, and VAL233 (Fig. **5.8.18**). Additionally, there are hydrophobic interactions with THR232, VAL233, ASN234, ASN235, and MET325.



© University of Venda




a b

**Figure 5.8.18.** Molecular insight into compound Mono8 docked on the antimalarial target (PFLDH, PDB: 1U5A), highlighting the residues within the binding site pocket. Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, light blue-dashed segments denote aromatic HBs and black-dashed segments denote π-π stacking interactions.

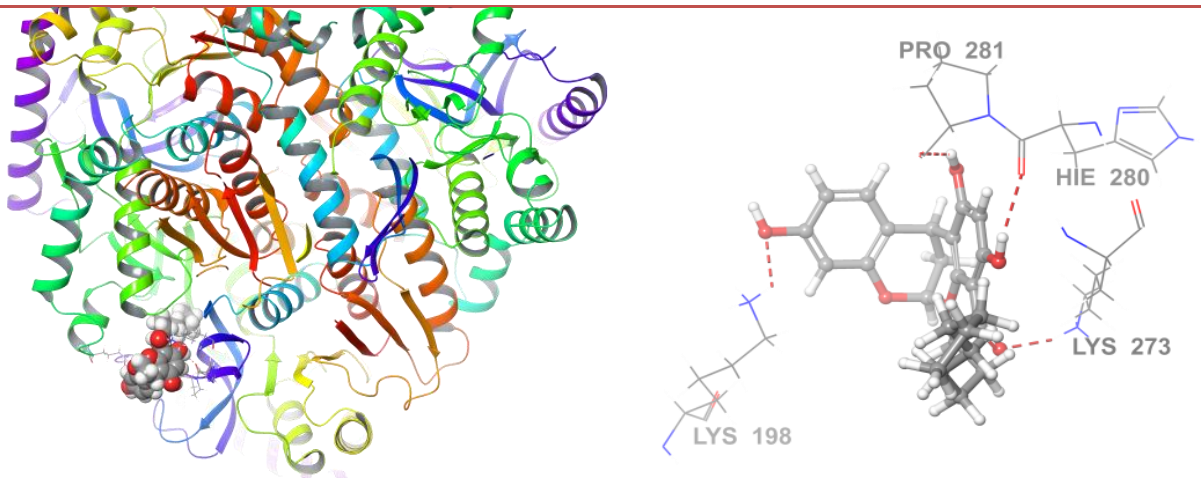
### 5.6.9. Molecular interaction of considered compounds with PFMDH

Among the active sites M1, M2, M4, and M5, only a limited number of the considered molecules show binding, highlighting the selectivity of these sites. Active site M3 interacts with all considered molecules, indicating its broader binding capability. Notably, the best-docked complexes for each active site are Mono2/PFMDH at M1, Mono8/PFMDH at both M2 and M3, and Mono9/PFMDH at both M4 and M5, with binding score energies of -6.842, -8.030, -7.533, -6.080, and -6.247, respectively.

At active site M1, the interaction between Mono2 and PFMDH does not show greater strength than that of the co-crystal ligand. Mono2 engages in four HBs with residues: LYS B:273, HIE B:280, PRO B:281, and LYS D:198 (Fig. **5.8.19**). Additionally, polar interactions are observed with residues VAL B:282, GLU B:283, PHE B:284, THR B:252, and PHE B:251. Furthermore, hydrophobic interactions occur with residues ALA B:272, LEU B:250, and PHE D:195.



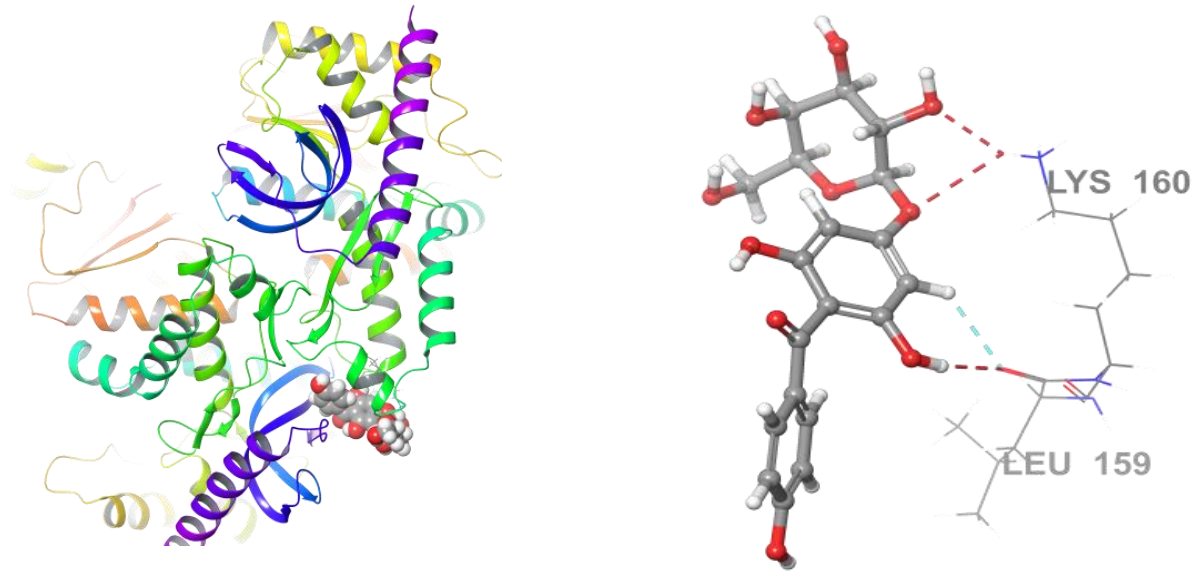
© University of Venda




a b

**Figure 5.8.19.** Molecular insight into compound Mono2 docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M1). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.

At active site M2, the interaction between Mono8 and PFMDH shows relatively greater strength than the co-crystal ligand. Mono8 forms three HBs with residues LYS B:160 and LEU B:159 and an additional aromatic HB with LEU B:159 (Fig. **5.8.20**). Furthermore, hydrophobic interactions are shown with residues PHE B:195, VAL B:187, LEU D:250, and PHE D:284.




a b

**Figure 5.8.20.** Molecular insight into compound Mono8 docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M2). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the



© University of Venda

ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

At active site M3, the interaction between Mono8 and PFMDH does not show significantly greater strength compared to the co-crystal ligand. Mono8 forms four HBs with residues ASN119, THR76, and ASP32, as well as three aromatic HBs with GLN80, GLY78, and VAL117 PRO23 (Fig. **5.8.21**). Additionally, there are glycine interactions involving residues GLY10, SER9, GLY8, THR76, ALA77, GLY78, VAL79, and GLN80. Hydrophobic interactions are observed with GLN11, while polar interactions occur with ASN94, ASN119, SER118, VAL117, MET142, and.

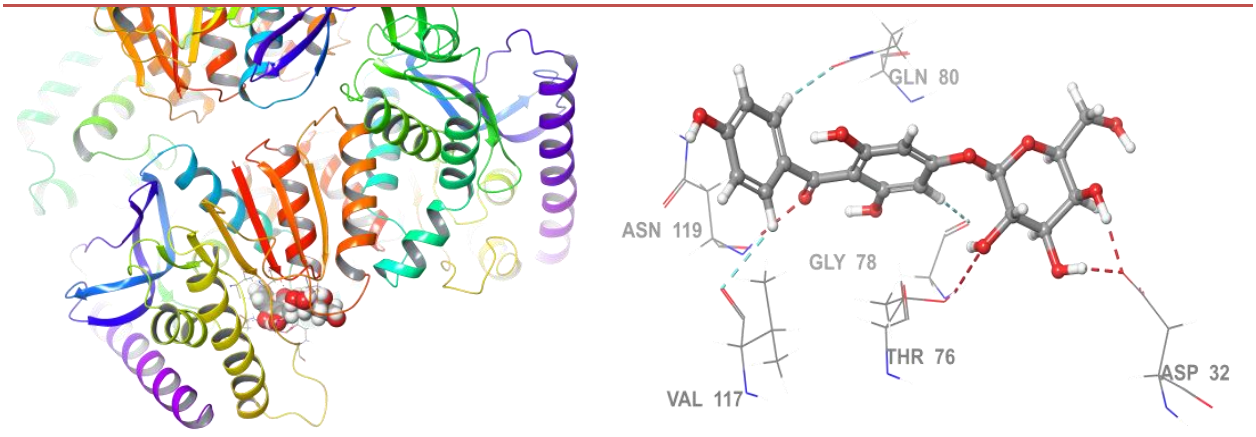
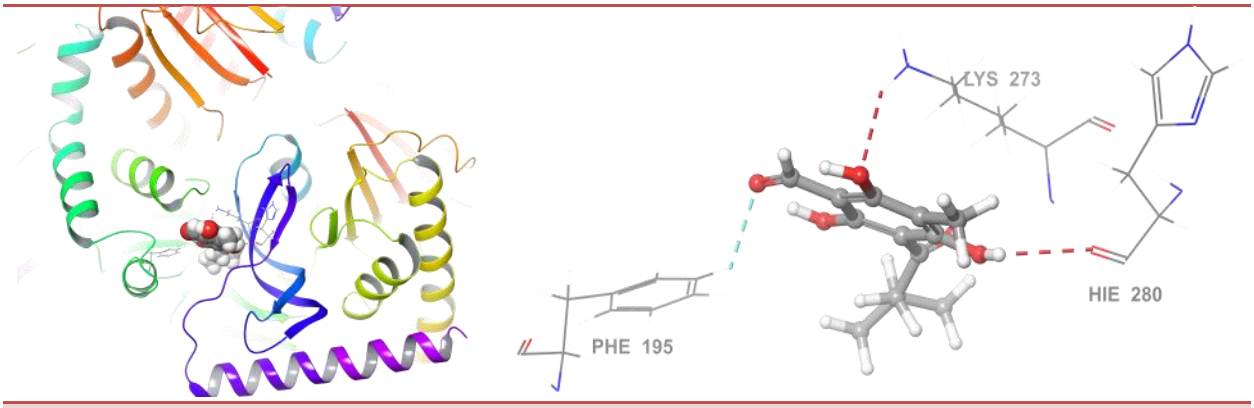



**Figure 5.8.21.** Molecular insight into compound Mono8 docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M3). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

At active site M4, the interaction between Mono9 and PFMDH does not show significantly greater strength than the co-crystal ligand. Mono9 forms two HBs with residues LYS C:273 and HIE C:280 and one aromatic HB with PHE 195 (Fig. **5.8.22**). Additionally, polar interactions involve residues ASN A:188, VAL A:190, and PHE C:284. Hydrophobic interactions are observed with residues PRO C:281, VAL A:187, LEU A:159, VAL A:161, and MET A:200.



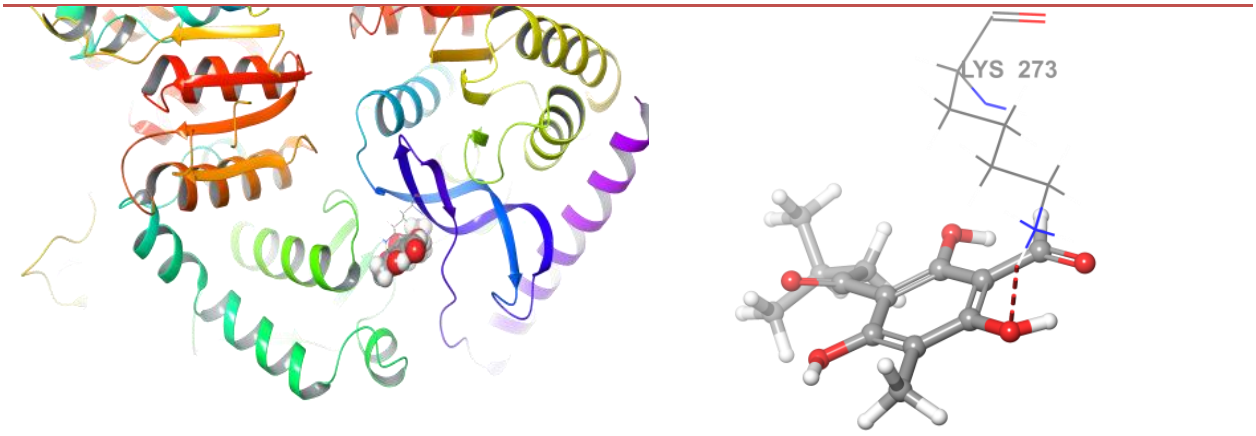
© University of Venda




a b

**Figure 5.8.22.** Molecular insight into compound Mono9 docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M4). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs and light blue-dashed segments denote aromatic HBs.

At active site M5, the interaction between Mono9 and PFMDH does not show significantly greater strength than the co-crystal ligand. Mono9 forms one HB with residue LYS A:273 (Fig. **5.8.23**) and polar interactions involving residues THR A:252 and PHE A:251. Additionally, hydrophobic interactions are shown with residues VAL C:161, LEU C:159, HIE A:280, VAL A:282, and PHE A:284.




a b

**Figure 5.8.23.** Molecular insight into compound Mono9 docked on the antimalarial target (PFMDH, PDB: 6R8G), highlighting the residues within the binding site pocket (site M5). Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs.



© University of Venda

### 5.6.10. Molecular interaction of considered other ACPLS compounds with CDK-2

In the molecular docking study of CDK-2 against the selected compounds, most of the considered molecules, except for Mono3, were found to interact at the active site. Among them, Mono8 showed the strongest binding affinity with CDK-2. Mono3 displayed a binding energy of -8.491 kcal $mol^{-1}$. However, it is worth noting that the interaction of Mono3 with CDK-2 was not stronger than that of the co-crystal ligand. Mono3 formed four HBs with residues LEU83, ASP86, GLU8, and ASN132 (Fig. **5.8.4)**. Additionally, an aromatic HB was observed with GLU8, and a pi-cation interaction occurred with LYS89. Furthermore, hydrophobic interactions with residues LEU134, LYS33, ALA31, VAL64, and GLY13 were noted**.**

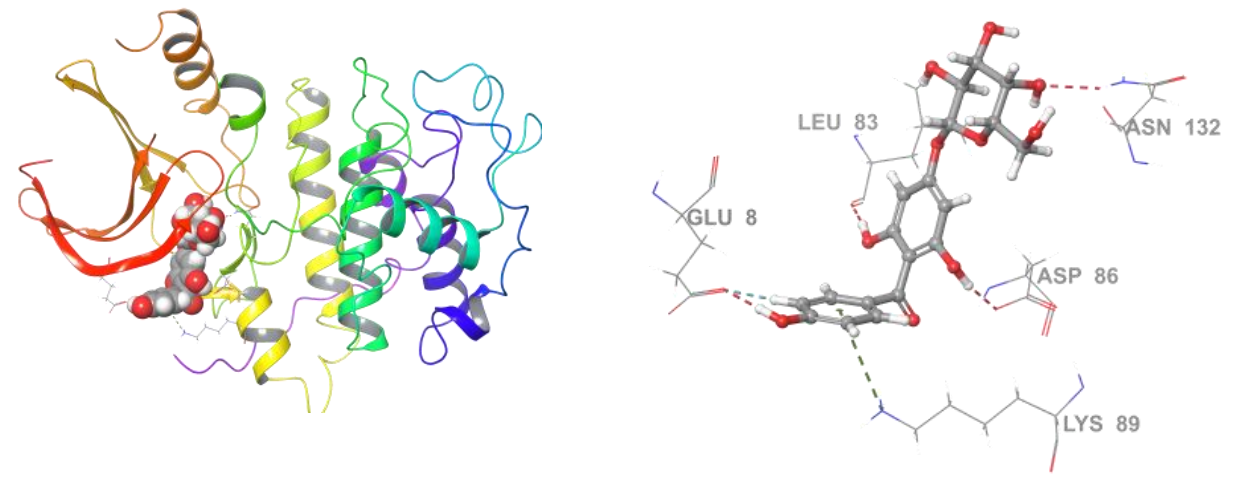



**Figure 5.8.24.** Molecular insight into compound Mono3 docked on the lung cancer target (CDK-2, PDB: 1DI8), highlighting the residues within the binding site pocket. Panel a) showcases the docking pose of the protein-ligand complex. Panel b) shows 3D diagrams illustrating the interactions between the ligand and the active site of the target; and the red-dashed segments denote HBs, light blue-dashed segments denote aromatic HBs and green-dashed segments denote pi-cation interactions.



© University of Venda

### 5.7. Analysis of physicochemical properties

In this study, Quantities relevant to assessing the potential of a compound as a drug were also examined, including properties covered by Lipinski's rule of five and ADME-T properties.

Lipinski's rule of five is a guideline that uses four criteria to make an initial prediction of a compound's drug-likeness. These criteria include a molecular weight (MW) under 500 a.m.u., an octanol/water partition coefficient (ClogPo/w) below 5, a maximum of 5 hydrogen bond donors (HBD), and no more than 10 hydrogen bond acceptors (HBA). Table 5.13 lists the values for the compounds studied. The Pfizer registration system flags compounds that do not meet two or more of these criteria [27]. Among the ACPLs examined, only Mono1 and Mono2 have $ClogP_{o/w}$ values above 5, indicating higher lipophilicity, while all other criteria are met by all compounds. As a result, based on Lipinski's rule, all compounds are considered suitable candidates for further evaluation.

**Table 5.13**. Values of the properties considered by Lipinski's rule-of-five for the molecules considered in this work.

The properties are the molecular mass (MW), the maximum number of H-bond donors (HBD), the maximum number of H-bond acceptors (HBA) and the predicted octanol/water partition coefficient ($QPlogP_{o/w}$). The values were calculated using QikProp [29]. The row under the names of the properties shows the criteria set by Lipinsky rule.

| Molecule considered | MW (a.m.u.) | HBD | HBA | $QPlogP_{o/w}$ | Number of violations of Lipinsky rule |
|---|---|---|---|---|---|
| | < 500 | < 5 | < 10 | < 5 | |
| | | | | | |
| Mono1 | 390.6 | 1.0 | 2.3 | 6.7 | 1.0 |
| Mono2 | 426.5 | 1.0 | 3.8 | 5.5 | 1.0 |
| Mono3 | 426.5 | 2.0 | 4.8 | 4.8 | 0.0 |
| Mono4 | 434.4 | 1.0 | 6.8 | 2.5 | 0.0 |
| Mono5 | 420.4 | 1.0 | 4.8 | 3.4 | 0.0 |
| Mono6 | 292.4 | 0.0 | 2.3 | 4.5 | 0.0 |
| Mono7 | 456.5 | 3.0 | 4.5 | 4.7 | 0.0 |
| Mono8 | 252.3 | 0.0 | 3.3 | 1.9 | 0.0 |



© University of Venda

ADME (Absorption, Distribution, Metabolism, and Excretion) properties are essential for predicting how a drug behaves within the body. The calculated ADME properties are presented in Table 5.14 with the recommended value ranges listed in the second row. The physiological and pharmacological significance of these parameters can be summarized as follows [28]. Drug absorption is indicated by the %PHOA and QPPCaco descriptors, the latter being linked to intestinal absorption. Two descriptors, QPlogBB and QPPMDCK, are associated with the ability to cross the blood-brain barrier (BBB). The BBB is a vital boundary between the brain and the rest of the body, and good BBB penetration is important for drugs aimed at the central nervous system (CNS). Conversely, drugs targeting other organs should have low BBB penetration to reduce potential CNS side effects. Once in the bloodstream, a drug can bind to plasma proteins or remain free to circulate, with human serum albumin being a key protein involved in drug binding. The QPlogKhsa descriptor is used to indicate how much of the drug remains available to exert its intended effect, based on its binding to human serum albumin. The values of these descriptors for the ACPLs under study are reported in Table 5.14. A comparison with the recommended ranges shows that all these compounds meet the criteria for potential drug candidates.

**Table 5.14. ADMET related properties for the eight ACPL compounds considered in this study.**
The properties are the following: percent predicted human oral absorption (%PHOA); predicted apparent Caco-2 cell permeability (QPPCaco, nm/sec); predicted apparent MDCK cell permeability (QPPMDCK, nm/sec); predicted brain/blood partition coefficient (QPlogBB); and predicted binding capacity to human serum albumin (QPlogKhsa). The values were calculated using QikProp [29]. The row under the names of the properties shows the recommended values-ranges.

| Molecule considered | Percent PHOA | QPPCaco | QPPMDCK | QPlogBB | QPlogKhsa |
|---|---|---|---|---|---|
| | <25 % poor, >80 % high | <25-poor, >500-great | <25-poor, >500-great | -3 to 1.2 | -1.5 to 1.5 |
| Mono1 | 100.0 | 370.3 | 169.0 | -2.4 | 1.4 |
| Mono2 | 94.8 | 514.9 | 241.4 | -1.6 | 1.1 |
| Mono3 | 100.0 | 321.7 | 145.2 | -1.9 | 0.8 |
| Mono4 | 72.8 | 56.0 | 21.9 | -2.1 | 0.2 |
| Mono5 | 81.6 | 85.2 | 34.5 | -1.9 | 0.6 |
| Mono6 | 100.0 | 1470.5 | 750.5 | -0.8 | 0.7 |
| Mono7 | 92.0 | 129.4 | 54.3 | -2.3 | 0.9 |
| Mono8 | 79.2 | 196.9 | 85.4 | -1.4 | -0.1 |



© University of Venda

### 5.8. Concluding remarks

The computational analysis of selected ACPLs has identified consistent patterns in their molecular properties that can guide the development of new structures with enhanced therapeutic potential. Key findings show that IHBs stabilize conformers and influence their preferences. This supports earlier studies on ACPLs and increases confidence in predicting similar compounds. The effect of solvents on these molecular properties also offers valuable insights, as they suggest how ACPLs may behave in various environments, including biological systems.

Docking studies further emphasize the therapeutic promise of specific ACPLs like Mono4, Mono5, Mono7, and Mono8, which exhibit strong binding affinities for targets associated with cancer and malaria. For example, Mono4 demonstrates robust interactions with cancer-associated proteins, including HER2 and CDK-2, hinting at its potential for use in HER2-positive cancer therapies. Similarly, Mono8 shows effective binding to both cancer and malaria-related proteins, such as EGFR and PFLDH, suggesting its versatility as a multi-target drug candidate. Mono5 and Mono7 also display notable selectivity: Mono5 aligns with cancer-related pathways involving PI3K, while Mono7 shows moderate efficacy against malaria proteins, potentially aiding in the treatment of resistant strains. Moreover, ACPLs with anticancer activities such as Mono1, Mono7, and Mono8 also interact effectively with malaria targets like PFMDH, while those with antimalarial activity, including Mono2, Mono3, and Mono6, display promising interactions with cancer targets like BRAF V600E, Topo I, and EGFR. ADME-T predictions further support the favourable drug-like properties of these compounds.

These results prompt the need for experimental studies to confirm the pharmacokinetics and bioactivity of ACPLs within biological settings. The dual-action potential of compounds like Mono4 and Mono8 for both cancer and malaria treatments highlight a promising path for developing multifunctional drugs. Future research could focus on evaluating these ACPLs under diverse physiological conditions and optimizing their bioavailability and therapeutic efficacy through additional solvent systems or formulations.



© University of Venda

## 5.9. References

1. Kabanda, M. M., & Mammino, L. (2012). The conformational preferences of acylphloroglucinols—A promising class of biologically active compounds. *International Journal of Quantum Chemistry*, 112(23), 3691-3702.
2. Mammino, L., & Kabanda, M. M. (2009). A study of the intramolecular hydrogen bond in acylphloroglucinols. *Journal of Molecular Structure: THEOCHEM*, 901(1-3), 210-219.
3. Mammino, L., & Kabanda, M. M. (2013). The role of additional O–H···O intramolecular hydrogen bonds for acylphloroglucinols' conformational preferences *in vacuo* and in solution. *Molecular Simulation*, 39(1), 1-3.
4. Mammino, L., & Kabanda, M. M. (2012). Computational study of the patterns of weaker intramolecular hydrogen bonds stabilizing acylphloroglucinols. *International Journal of Quantum Chemistry*, 112(14), 2650-2658.
5. Mammino, L., & Kabanda, M. M. (2009). A computational study of the effects of different solvents on the characteristics of the intramolecular hydrogen bond in acylphloroglucinols. *The Journal of Physical Chemistry A*, 113(52), 15064-15077.
6. Mammino, L. (2019). Intramolecular hydrogen bonding patterns, conformational preferences, and molecular properties of dimeric acylphloroglucinols: An ab initio and DFT study. *Journal of Molecular Structure*, 1176, 488-500.
7. Mammino, L. (2021). Correlation effects in trimeric acylphloroglucinols. *Computation*, 9(11), 121. https://doi.org/10.3390/computation9110121
8. Mammino, L. (2017). Intramolecular hydrogen bonding and conformational preferences of arzanol—An antioxidant acylphloroglucinol. *Molecules*, 22(8), 1294. https://doi.org/10.3390/molecules22081294
9. Mammino, L. (2023). Conformational preferences and intramolecular hydrogen bonding patterns of tetraflavaspidic acid BBBB–a tetrameric acylphloroglucinol. *Physical Sciences Reviews*, 8(11), 4179-4200. https://doi.org/10.1515/psr-2021-0239
10. Desiraju, G. R., & Steiner, T. (1999). The Weak Hydrogen Bond in Structural Chemistry and Biology. Oxford University Press Inc.
11. Steiner, T., & Koellner, G. (2001). Hydrogen bonds with π-acceptors in proteins: Frequencies and role in stabilizing local 3D structures. *Journal of Molecular Biology*, 305(3), 535-557. https://doi.org/10.1006/jmbi.2000.4301



© University of Venda

12. Hayashi, S., Nishide, T., & Nakanishi, W. (2019). Nature of intramolecular O-H···π interactions as elucidated by QTAIM dual functional analysis with QC calculations. *RSC Advances*, 9(27), 15521-15530.
13. Spoliti, M., Bencivenni, L., Quirante, J. J., & Ramondo, F. (1997). Molecular conformations and harmonic force field of 1,3,5-benzenetriol molecule from ab initio and density functional theory investigations. *Journal of Molecular Structure (Theochem)*, 390, 139-148.
14. Gilli, G., Bellucci, F., Ferretti, V., & Bertolasi, V. (1989). Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the beta-diketone fragment. *Journal of the American Chemical Society*, 111, 1023-1028.
15. Bertolasi, V., Gilli, P., Ferretti, V., & Gilli, G. (1991). Evidence for resonance-assisted hydrogen bonding. 2. Intercorrelation between crystal structure and spectroscopic parameters in eight intramolecularly hydrogen-bonded 1,3-diaryl-1,3-propanedione enols. *Journal of the American Chemical Society*, 113, 4017-4025.
16. Gilli, P., Bertolasi, V., Ferretti, V., & Gilli, G. (1994). Covalent nature of the strong homonuclear hydrogen bond. Study of the O-H—O system by crystal structure correlation methods. *Journal of the American Chemical Society*, 116, 909-915.
17. Guevara-Vela, J. M., Romero-Montalvo, E., Costales, A., Pendás, Á. M., & Rocha-Rinza, T. (2016). The nature of resonance-assisted hydrogen bonds: A quantum chemical topology perspective. *Physical Chemistry Chemical Physics*, 18, 26383-26390. https://doi.org/10.1039/C6CP04951D
18. Tshilande, N., & Mammino, L. (2023). Comparison of the Molecular Properties of Euglobals Differing by the Mutual Positions of the Two R–C= O Groups (R= H and CH2CH (CH3) 2): A Computational Study. *Chemistry*, 5(4), 2120-2154.
19. Marenich, A. V., Cramer, C. J., & Truhlar, D. G. (2008). Perspective on foundations of solvation modeling: The electrostatic contribution to the free energy of solvation. *Journal of Chemical Theory and Computation*, 4, 877-887.
20. Amovilli, C. (1994). Calculation of the dispersion energy contribution to the solvation free energy. *Chemical Physics Letters*, 229, 244-249.
21. Curutchet, C., Orozco, M., Luque, F. J., Mennucci, B., & Tomasi, J. (2006). Dispersion and repulsion contributions to the solvation free energy: Comparison of quantum



© University of Venda

mechanical and classical approaches in the polarizable continuum model. *Journal of Computational Chemistry*, 27, 1769-1780.

22. Remsing, R. C., Liu, S., & Weeks, J. D. (2016). Long-ranged contributions to solvation free energies from theory and short-ranged models. *Proceedings of the National Academy of Sciences*, 113(11), 2819-2826.
23. Amovilli, C., & Floris, F. M. (2022). Method to compute the solute–solvent dispersion contribution to the electronic excitation energy in solution. *Journal of Chemical Theory and Computation*, 18, 6816-6825.
24. Liu, S. C., Zhu, X. R., Liu, D. Y., & Fang, D. C. (2023). DFT calculations in solution systems: Solvation energy, dispersion energy, and entropy. *Physical Chemistry Chemical Physics*, 25, 913-931.
25. Matczak, P., & Domagała, M. (2017). Heteroatom and solvent effects on molecular properties of formaldehyde and thioformaldehyde symmetrically disubstituted with heterocyclic groups $C_4H_3Y$ (where Y = O–Po). *Journal of Molecular Modeling*, 23, 282-295.
26. Yoosefian, M., & Etminan, N. (2016). The role of solvent polarity in the electronic properties, stability, and reactivity trend of a tryptophan/Pd doped SWCNT novel nanobiosensor from polar protic to non-polar solvents. *RSC Advances*, 6, 64818-64825.
27. Manthena, V. S., Varma, M., Perumal, O. P., & Panchagnula, R. (2006). Functional role of P-glycoprotein in limiting peroral drug absorption: Optimizing drug delivery. *Current Opinion in Chemical Biology*, 10, 367–373. https://doi.org/10.1016/j.cbpa.2006.06.015
28. Gürdere, M. B., Budak, Y., Kocyigit, U. M., Taslimi, P., Tüzün, B., & Ceylan, M. (2021). ADME properties, bioactivity, and molecular docking studies of 4-amino-chalcone derivatives: New analogues for the treatment of Alzheimer's, glaucoma, and epileptic diseases. *In Silico Pharmacology*, 9(1), 34. https://doi.org/10.1007/s40203-021-00094-x
29. QikProp. Schrödinger. Available online: https://www.schrodinger.com/platform/products/qikprop/ (Accessed: March 13, 2022).



© University of Venda

# CHAPTER 6
# DOCKING STUDIES OF SELECTED DIMERIC AND TRIMERIC ACYLPHLOROGLUCINOLS

## 6.1. Introduction to the chapter

This chapter discusses docking studies performed on a series of selected dimeric and trimeric ACPLs (described in section 2.4.2.3 and 2.2.4.2) with the objective of assessing their binding to relevant anticancer and antimalarial targets. The computational methods used in this chapter are the same as those outlined in Chapter 4, sections 4.2.2.–4.2.3.

## 6.2. Background

ACPLs have demonstrated significant bioactivity [1, 2], but their interactions with key proteins involved in cancer and malaria are not yet fully understood. Table 6.1 lists the selected ACPLs, and their acronyms used in this chapter and their structures are shown in figures in section 2.5.2.3.

This study aims to identify promising drug candidates by systematically evaluating molecules against specific criteria for effectiveness, stability, and safety. The objectives are as follows:

- To determine the binding affinity for the target protein, aiming for a binding energy threshold of -6 kcal/mol or lower to ensure strong and favorable interactions.
- To assess interaction stability through key binding interactions, such as 2–5 hydrogen bonds and stable hydrophobic contacts, confirming each molecule's retention within the binding site.
- To evaluate drug-likeness according to Lipinski's Rule of Five, focusing on optimal molecular weight (≤ 500 Da), hydrogen bond donor/acceptor balance, and a logP ≤ 5, to predict oral bioavailability.
- To analyze ADME (Absorption, Distribution, Metabolism, Excretion) properties, emphasizing high intestinal permeability, moderate plasma protein binding, and favorable pharmacokinetics.



© University of Venda

**Table 6.1.** List of considered acylphloroglucinol molecules and acronyms with which they are denoted in this chapter.

| Class of ACPL | Name of molecules considered | Acronym denoting the name |
|---|---|---|
| DIMERIC ACPL | Acrovestone | dim1 |
| | Bis2-4-diacetylphloroglucyl-methane | dim2 |
| | Butyrylmallatochromanol | dim3 |
| | Butyrylmallatochromene | dim4 |
| | Butyrylmallotojaponin | dim5 |
| | Butyrylmallotolerin | dim6 |
| | Demethylacrovestone | dim7 |
| | Compound 502 | dim8 |
| | Drummondins E | dim9 |
| | Drummondins F | dim10 |
| | Isobutyrylmallatochromene | dim11 |
| | Isobutyrylmallotochromanol | dim12 |
| | Isomallatochromanol | dim13 |
| | Isomallotochromene | dim14 |
| | Isomallotolerin | dim15 |
| | Kosin | dim16 |
| | Mallotochromanol | dim17 |
| | Mallotochromene | dim18 |
| | Mallotojaponin | dim19 |
| | Mallotojaponol | dim20 |
| | Mallotophenone | dim21 |
| | Prenylatedcompound 504 | dim22 |
| | Robustaol A | dim23 |
| | | |
| TRIMERIC ACPL | Agrimol A | trim1 |
| | Agrimol C | trim2 |
| | Agrimol D | trim3 |
| | Agrimol E | trim4 |
| | Agrimol F | trim5 |
| | Agrimol G | trim6 |

To investigate these interactions, molecular docking was conducted to simulate the binding of each molecule to targeted proteins. Docking studies allow us to estimate binding strength and interaction characteristics, providing insights into the suitability of each molecule as a drug candidate. Schrödinger's GLIDE algorithm was used for this work, allowing us to rank compounds based on calculated binding energies and interaction profiles.



© University of Venda

### 6.3. Molecular docking results.

Twenty-three dimeric and six trimeric ACPLs were selected for docking against a range of cancer and malaria targets. Each docking result is detailed for all active sites considered, with emphasis on binding energies, hydrogen bonds, and key hydrophobic interactions. Dimeric and trimeric ACPL results are presented separately for clarity.

The twenty-three dimeric and six trimeric ACPLs considered in this work and eleven suitably selected molecular targets (proteins, described in section 4.6) were utilized for the docking studies. Each of the considered ACPLs was docked against the relevant targets and their binding energies (Table S6.1) and intermolecular interactions were determined, together with those of the co-crystal ligands which serve as comparison references.

#### 6.3.1 Docking results for dimeric ACPLs against anticancer targets

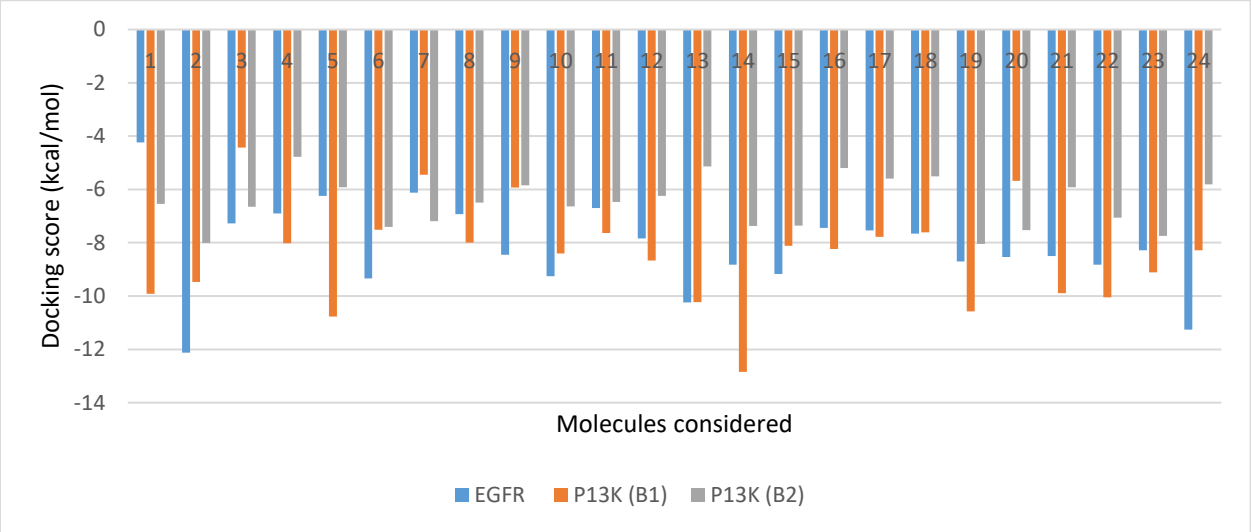
This section presents the molecular docking results for the selected dimeric ACPLs against each anticancer target. Table S6.1 compares the interaction energies between the ACPLs and the co-crystal ligand. Figure 6.1. shows docking scores of dimeric ACPLs against each anticancer target. Figure 6.2. shows molecular insights into the ligand-anticancer target interactions, focusing on the specific residues within the binding site pocket that contribute to ligand binding, and only cases where the ACPLs exhibit interaction energies better than or comparable to those of the co-crystal ligand are highlighted here.

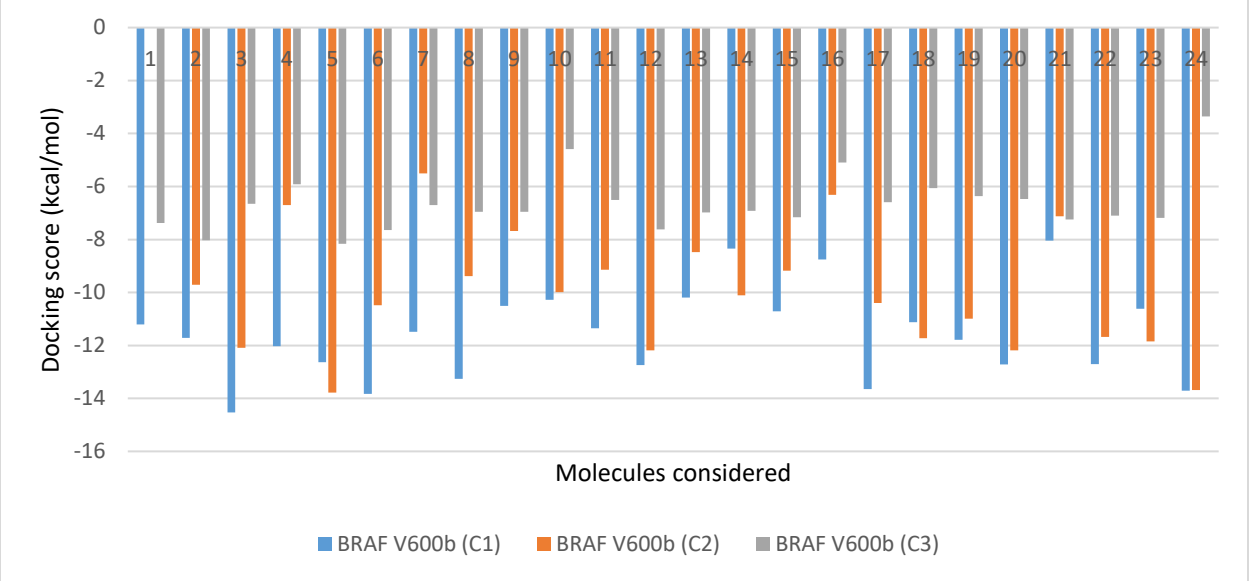
**Figure 6.1.** Docking scores of dimeric ACPLs against each anticancer target.
Numbers denoting the considered molecules on the x-axis.

| # | Conformers | # | Conformers | # | Conformers |
|---|---|---|---|---|---|
| 1 | dim1 | 9 | dim9 | 17 | dim17 |
| 2 | dim2 | 10 | dim10 | 18 | dim18 |
| 3 | dim3 | 11 | dim11 | 19 | dim19 |
| 4 | dim4 | 12 | dim12 | 20 | dim20 |
| 5 | dim5 | 13 | dim13 | 21 | dim21 |
| 6 | dim6 | 14 | dim14 | 22 | dim22 |
| 7 | dim7 | 15 | dim15 | 23 | dim23 |
| 8 | dim8 | 16 | dim16 | 24 | Co-crystalised ligand |



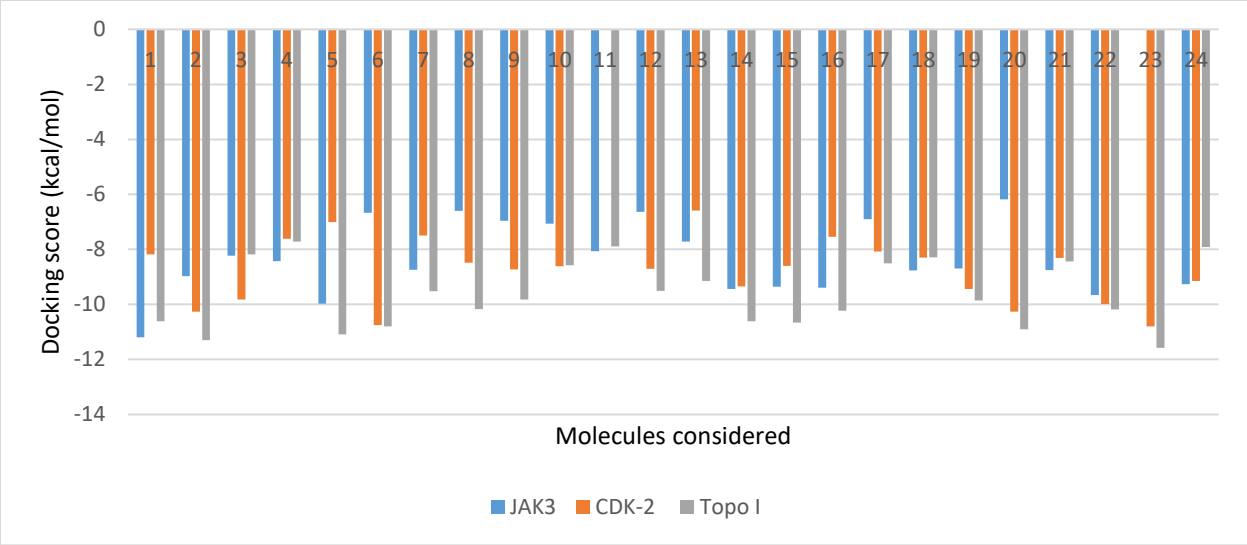
© University of Venda

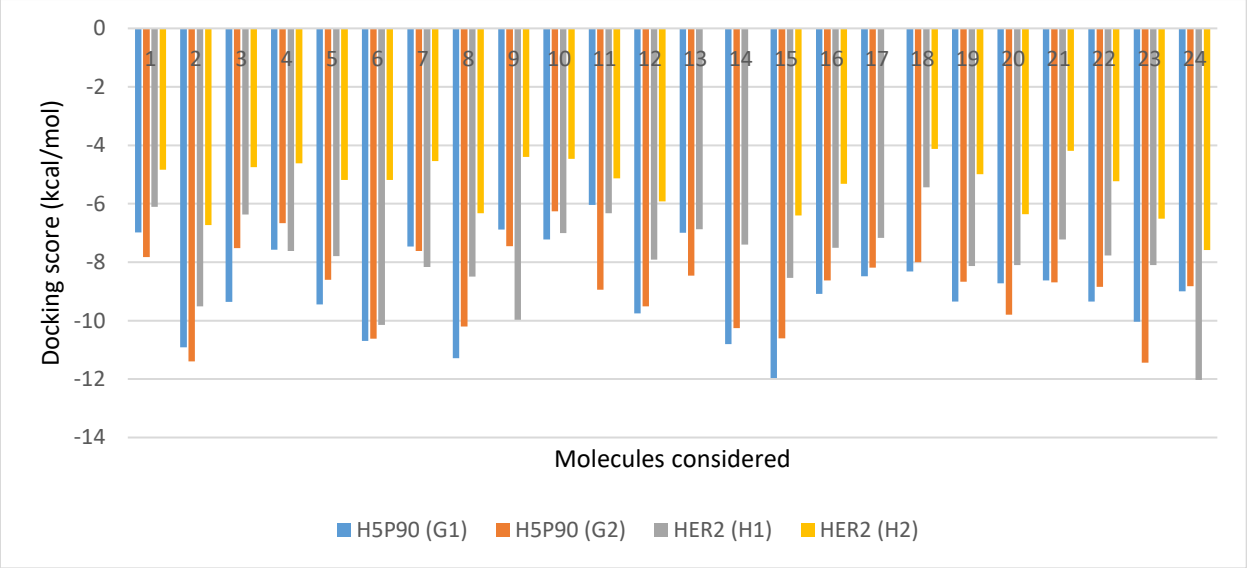





© University of Venda






© University of Venda

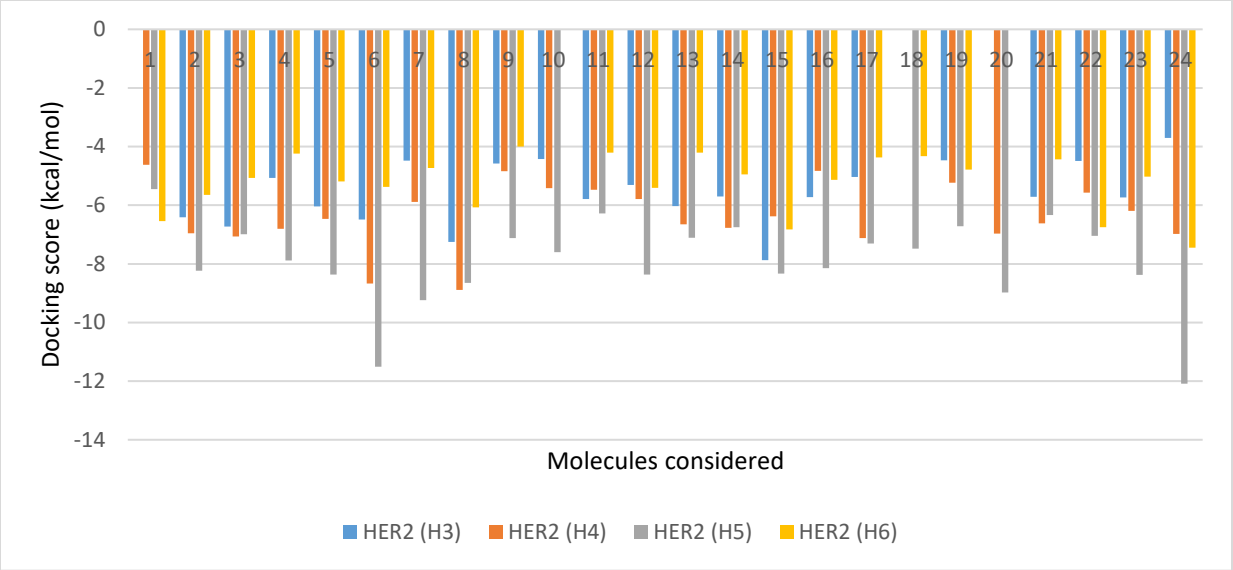
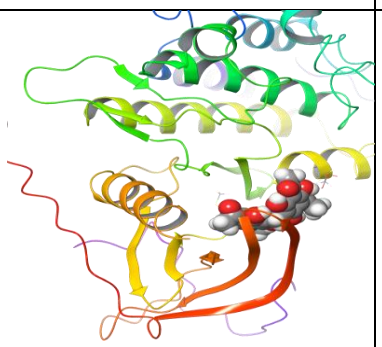
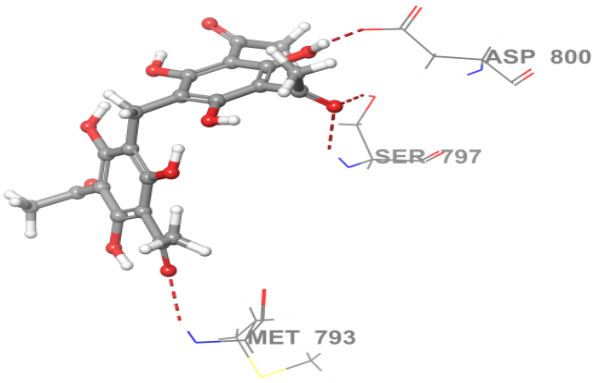
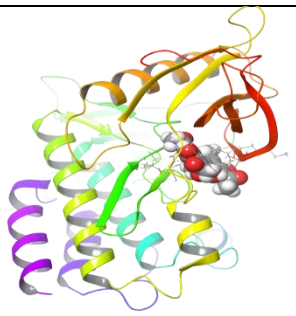
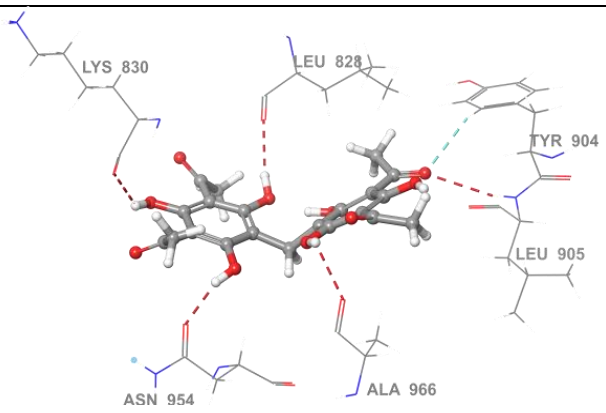
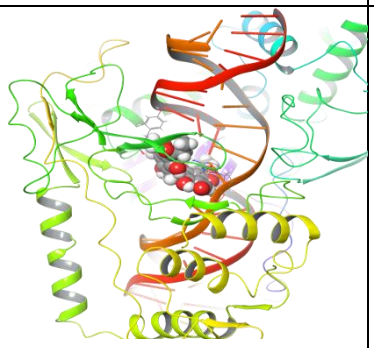
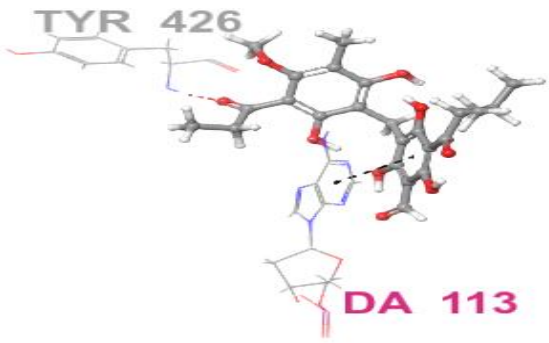
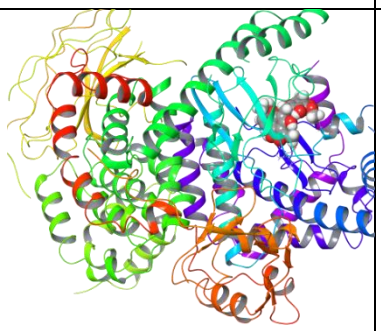
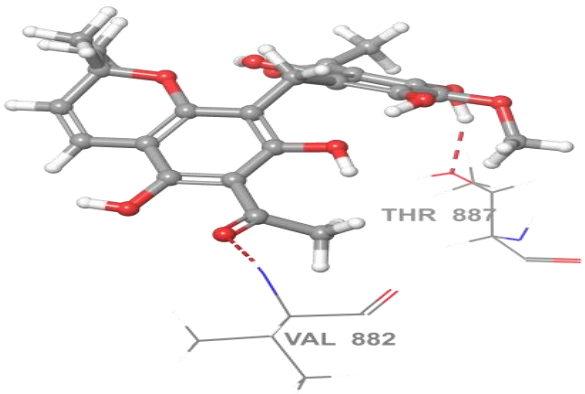
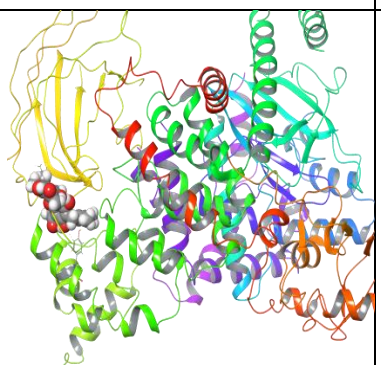
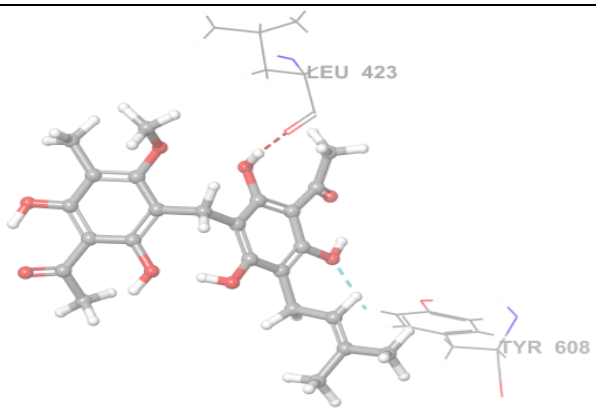



**Figure 6.2.** Molecular insight into dimeric-anticancer target highlighting the residues within the binding site pocket. The diagrams illustrate the interactions between the ligand and the active site of the target, with red-dashed segments denote HBs and black-dashed segments denote π-π stacking interactions, and light blue-dashed segments denote aromatic HBs.

| Docked complex considered | Docking pose of the ligand-target complex | Ligand-target interactions |
|---|---|---|
| dim2 docked on the lung cancer target (EGFR, PDB: 6LUD). |  |   |

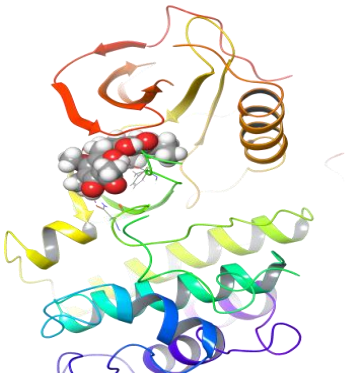
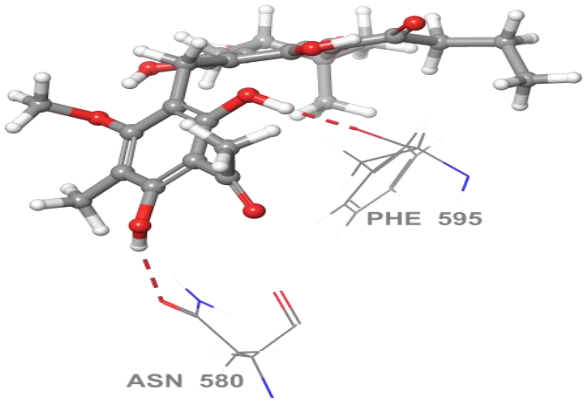
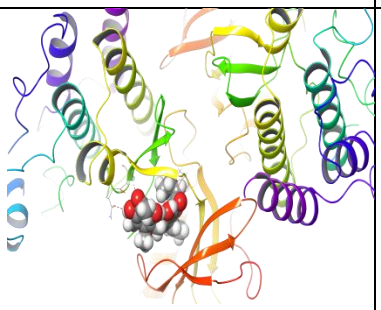
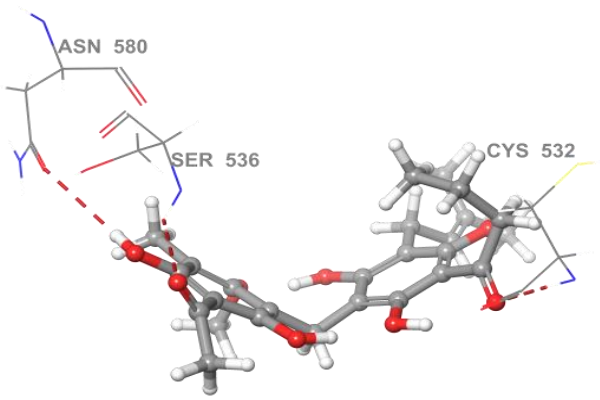
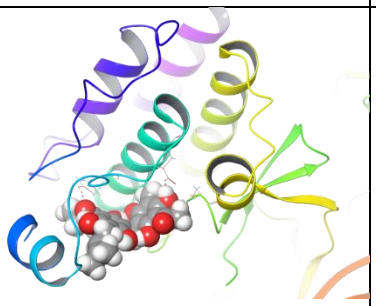
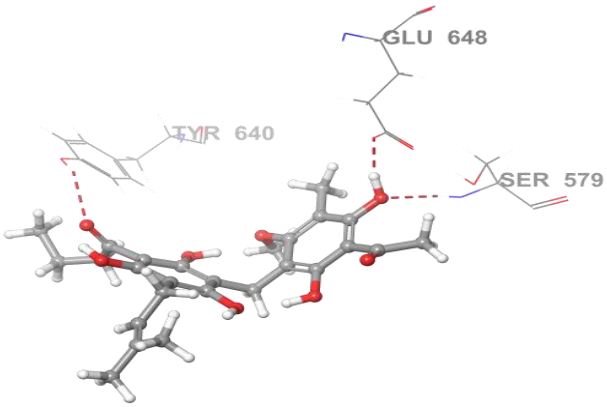
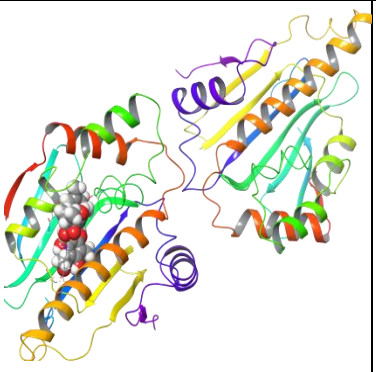
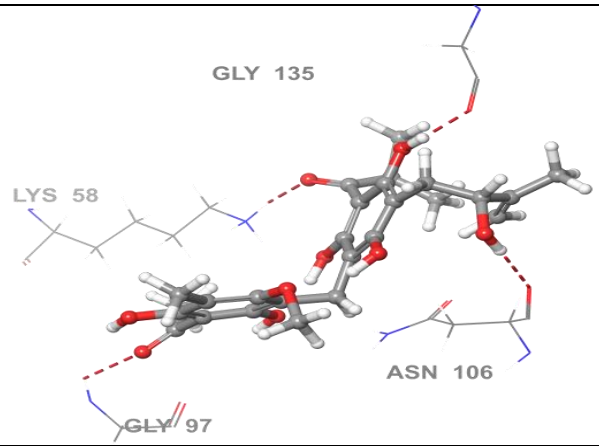


© University of Venda

| | | |
|---|---|---|
| dim2 docked on the lung cancer target (JAK3, PDB: 7C3N). |  |  |
| dim23 docked on the lung cancer target (Topo I, PDB: IT8I). |  |  |
| dim14 docked on the lung cancer target (P13K, PDB: 5JHB), site B1. |  |  |
| dim19 docked on the lung cancer target (P13K, PDB: 5JHB, site B2. |  |  |

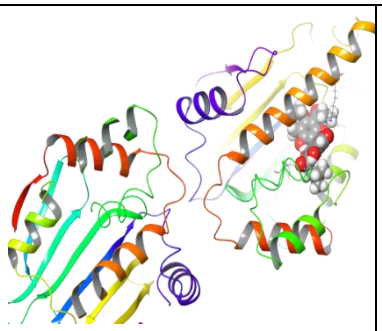
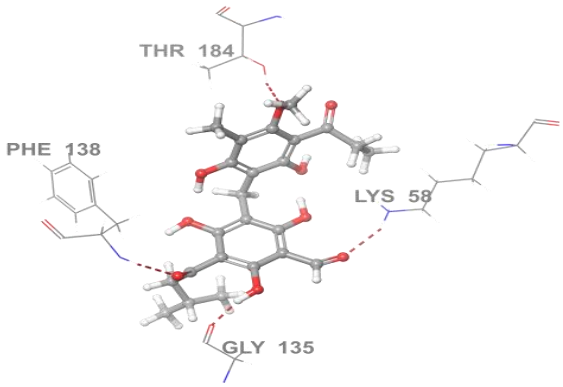
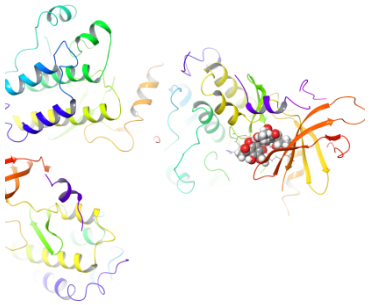
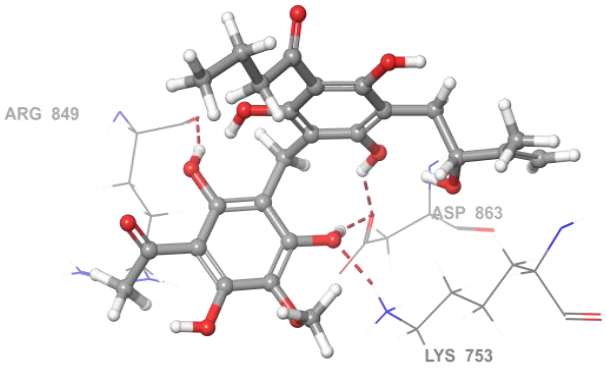
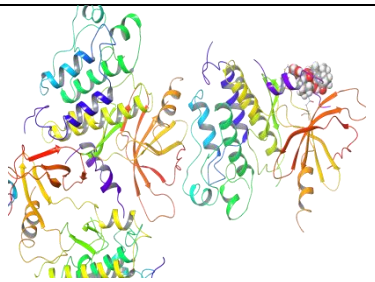
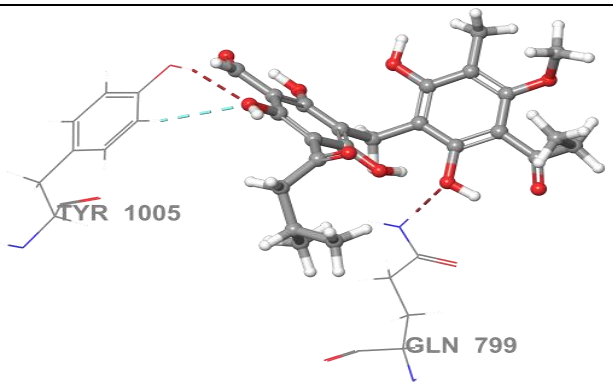
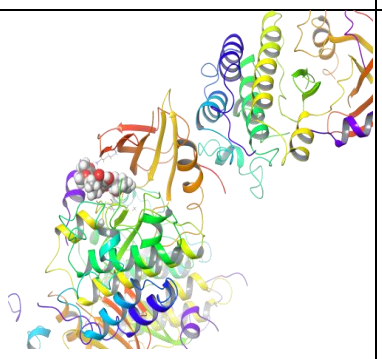
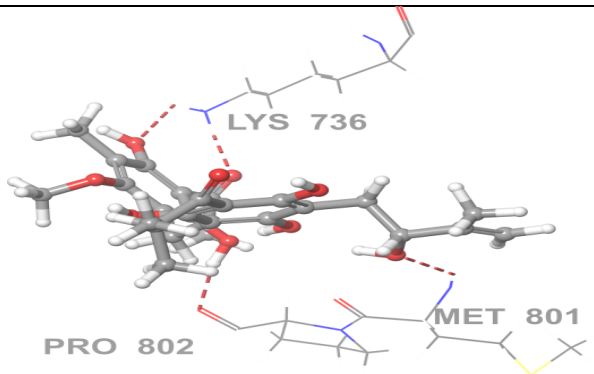


© University of Venda

| | | |
|---|---|---|
| dim3 docked on the lung cancer target (BRAF V600b, PDB; 6V34), site C1. |  |  |
| dim5 docked on the lung cancer target (BRAF V600b, PDB; 6V34), site C2. |  |  |
| dim5 docked on the lung cancer target (BRAF V600b, PDB; 6V34), site C3. |  |  |
| dim15 docked on the breast cancer target (H5P90, PDB: 3TUH), site G1. |  |  |

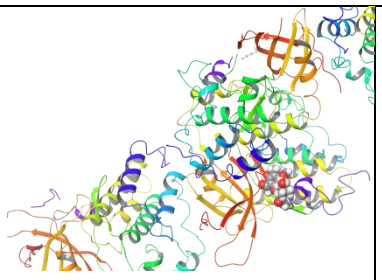
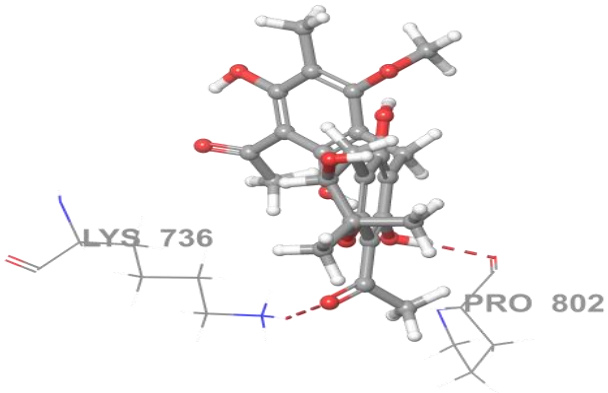
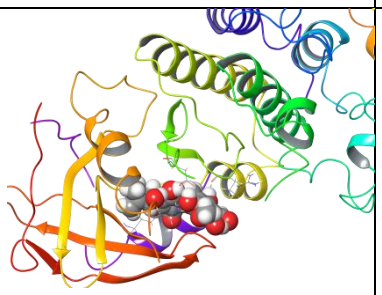
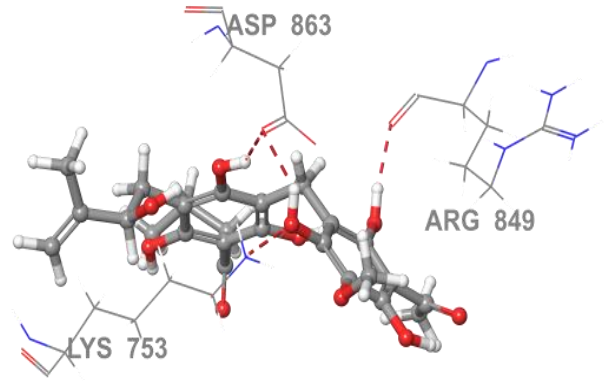
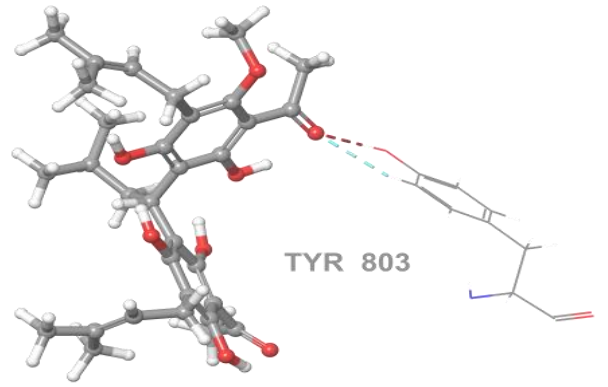


© University of Venda

| | | |
|---|---|---|
| dim23 docked on the breast cancer target (H5P90, PDB: 3TUH), site G2. |  |  |
| dim6 docked on the breast cancer target (HER2, PDB: 3RCD), site H1. |  |  |
| dim2 docked on the breast cancer target (HER2, PDB: 3RCD), site H2. |  |  |
| dim15 docked on the breast cancer target (HER2, PDB: 3RCD), site H3. |  |  |



© University of Venda

| | | |
|---|---|---|
| dim8 docked on the breast cancer target (HER2, PDB: 3RCD), site H4. |  |  |
| dim6 docked on the breast cancer target (HER2, PDB: 3RCD), site H5. |  |  |
| dim15 docked on the breast cancer target (HER2, PDB: 3RCD), site H6. |  |  |

**Docking with EGFR.** Dim2 exhibits the strongest binding, with a binding energy of -12.131 kcal $mol^{-1}$. It forms four H-bonds with MET793, SER797, and ASP800, and shows hydrophobic interactions with residues LEU792, GLN791, and ALA743. This interaction profile suggests that dim2 could be an effective EGFR inhibitor.



© University of Venda

**Docking with JAK3.** Dim2 is again the best-performing compound, with a binding energy of -11.189 kcal $mol^{-1}$. It forms five H-bonds with residues LEU905, ALA966, ASN954, LYS830, and LEU828, and includes an aromatic H-bond with TYR904. These findings indicate that dim2 is a promising JAK3 inhibitor.

**Docking with Topo I.** Dim23 achieves the strongest binding, with a binding energy of -11.577 kcal $mol^{-1}$. It forms H-bonds with TYR A:426 and π-π stacking with DA D:113, plus hydrophobic interactions with THR A:718. This profile highlights dim23 as a potential Topo I inhibitor.

**Docking with PI3K.** PI3K has two active sites, B1 and B2. At site B1, dim14 demonstrates the highest binding affinity, with a binding energy of -12.848 kcal $mol^{-1}$. Dim14 forms three H-bonds with VAL882, ASP950, and TYR867, along with hydrophobic interactions with ILE879, GLU880, and PRO810. At site B2, dim3 shows moderate binding, with a binding energy of -9.634 kcal $mol^{-1}$, forming two H-bonds with residues LEU861 and ASN849. These findings highlight dim14's potential as a PI3K inhibitor, particularly at site B1.

**Docking with BRAF V600B.** BRAF V600B has three active sites, C1, C2, and C3. Dim3 achieves the best binding at site C1 with a binding energy of -14.534 kcal $mol^{-1}$, forming two H-bonds with ASN580 and CYS532, and showing π-π stacking interactions with PHE595. At site C2, dim7 displays a moderate binding energy of -10.862 kcal $mol^{-1}$, forming H-bonds with CYS532 and THR529. At site C3, dim5 shows the highest binding energy of -8.453 kcal $mol^{-1}$, forming three H-bonds with ASP576 and GLU648. These findings support dim3 as the most effective BRAF V600B inhibitor at site C1.

**Docking with HSP90.** HSP90 has two active sites, G1 and G2. Dim15 is the best-docked compound at site G1, with a binding energy of -11.966 kcal $mol^{-1}$. It forms four H-bonds with LYS58, ASP54, ASN51, and GLY135, along with polar interactions with ASN106. At site G2, dim8 shows moderate binding with a binding energy of -10.134 kcal $mol^{-1}$, forming three H-bonds with residues GLY135 and ASP93. These findings suggest dim15's potential as a strong HSP90 inhibitor, particularly at site G1.



© University of Venda

**Docking with HER2.** HER2 has six active sites, denoted H1 through H6. At site H1, dim6 shows the best binding affinity, with a binding energy of -10.15 kcal $mol^{-1}$, forming four H-bonds with MET801 and ARG849. Dim8 performs moderately at site H2, achieving a binding energy of -8.611 kcal $mol^{-1}$ and forming H-bonds with GLY778 and LYS860. At site H3, dim4 binds with a score of -9.327 kcal $mol^{-1}$, forming H-bonds with MET801 and ASP808. At site H5, dim2 demonstrates strong binding with a score of -10.611 kcal $mol^{-1}$, forming H-bonds with GLY727 and LEU852. These results suggest dim6 as a leading candidate for HER2 inhibition, especially at site H1.

### 6.3.2 Docking results for dimeric ACPLs against antimalarial targets

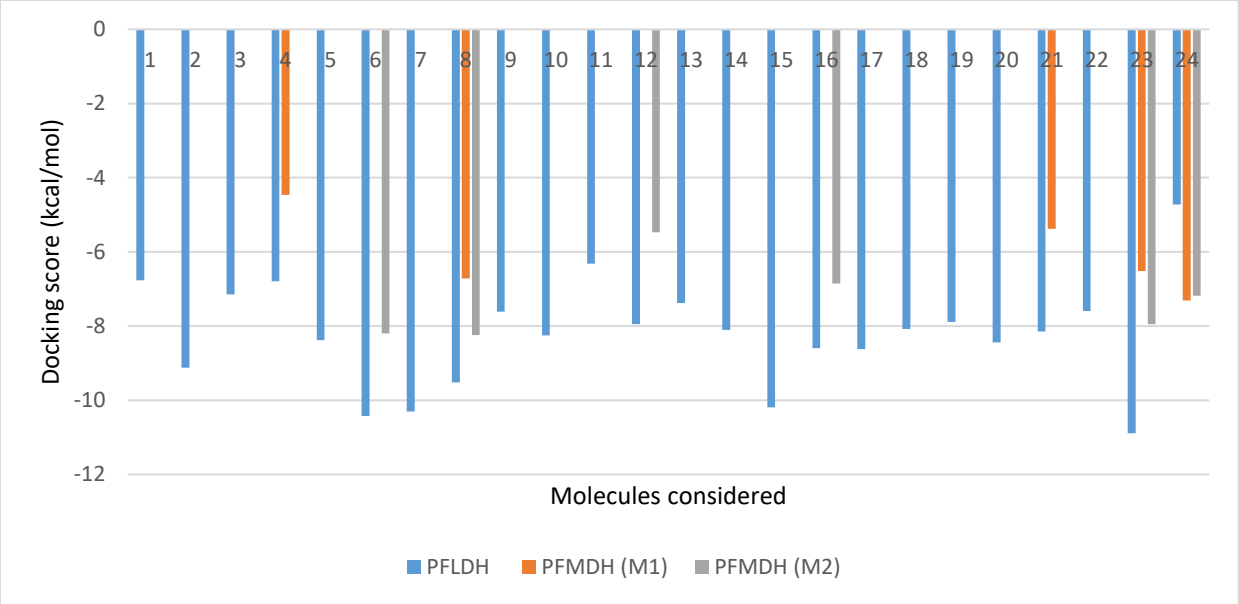
This section presents the results of the molecular docking of the dimeric considered ACPLs against each of the targets selected for the antimalarial activity (all related to *Plasmodium Falciparum*). Figure 6.3. shows docking scores of dimeric ACPLs against each antimalarial target. Figure 6.4. shows molecular insights into the ligand-anticancer target interactions, focusing on the specific residues within the binding site pocket that contribute to ligand binding.

**Figure 6.3**. Docking scores of dimeric ACPLs against each anticancer target.
Numbers denoting the considered molecules on the x-axis.

| # | Conformers | # | Conformers | # | Conformers |
|---|---|---|---|---|---|
| 1 | dim1 | 9 | dim9 | 17 | dim17 |
| 2 | dim2 | 10 | dim10 | 18 | dim18 |
| 3 | dim3 | 11 | dim11 | 19 | dim19 |
| 4 | dim4 | 12 | dim12 | 20 | dim20 |
| 5 | dim5 | 13 | dim13 | 21 | dim21 |
| 6 | dim6 | 14 | dim14 | 22 | dim22 |
| 7 | dim7 | 15 | dim15 | 23 | dim23 |
| 8 | dim8 | 16 | dim16 | 24 | Co-crystalised ligand |



© University of Venda



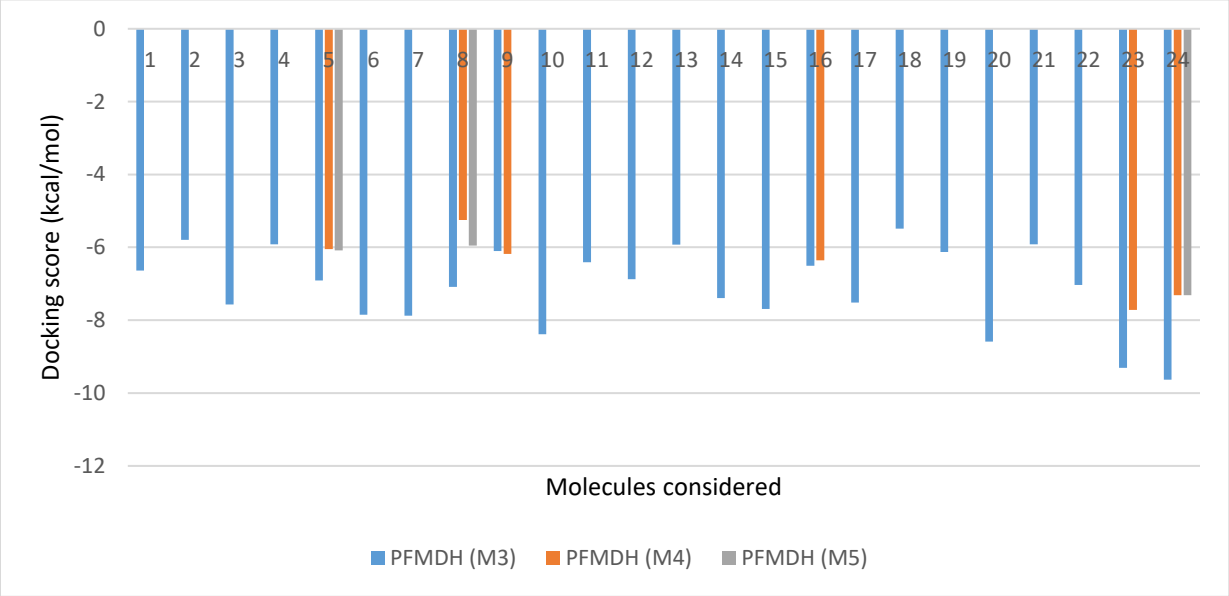
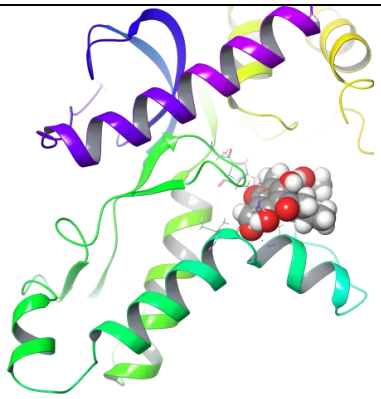
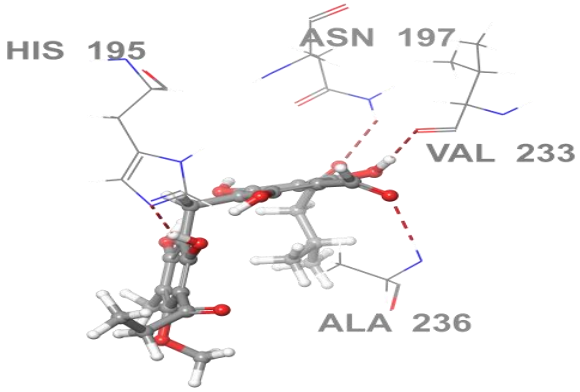
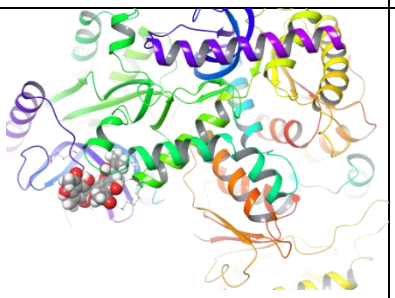
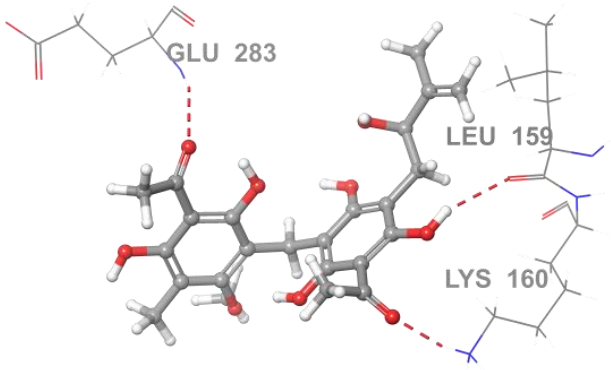
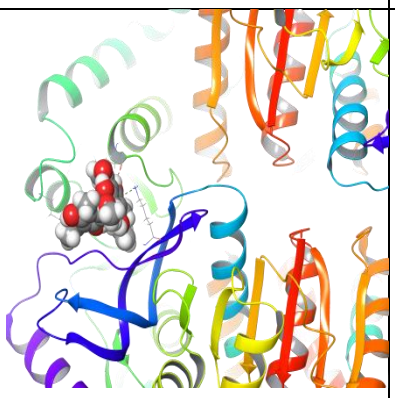
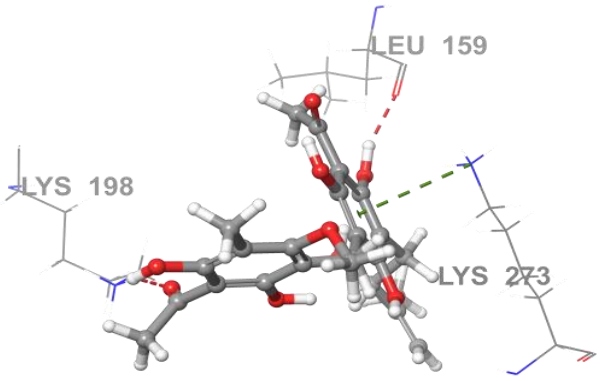




**Figure 6.4**. Molecular insight into dimeric-antimalarial target highlighting the residues within the binding site pocket. The diagrams illustrate the interactions between the ligand and the active site of the target, with



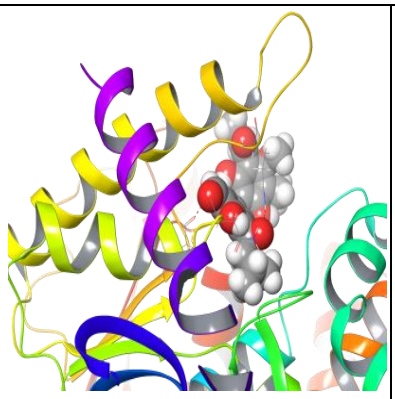
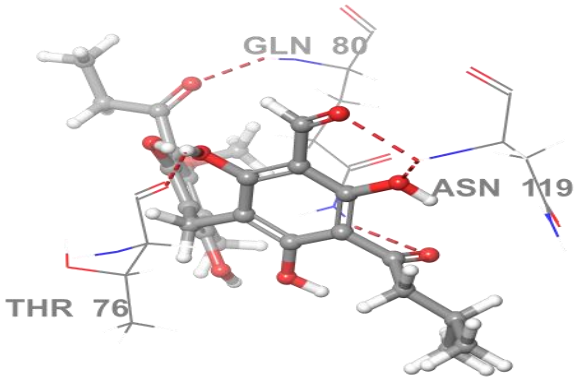
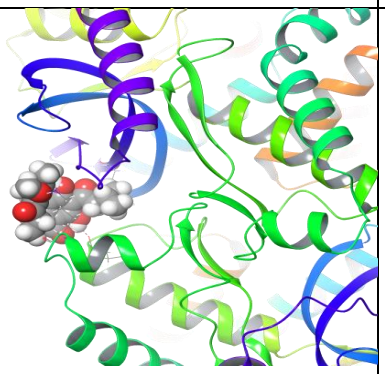
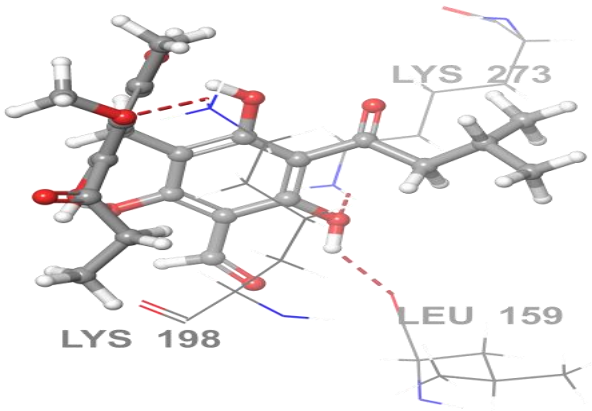
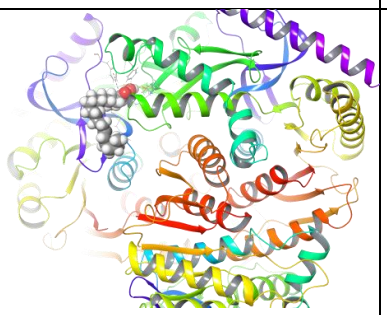
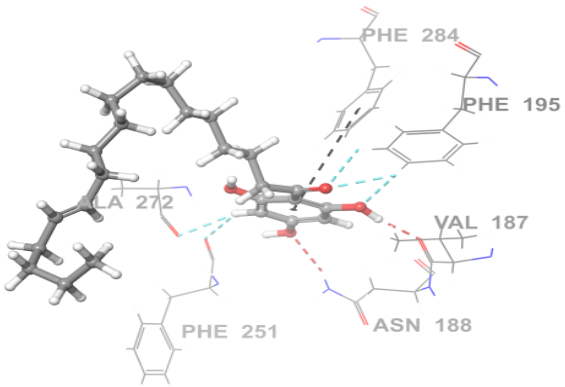
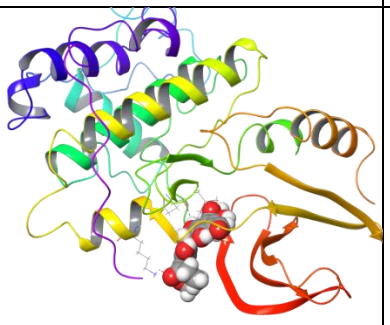
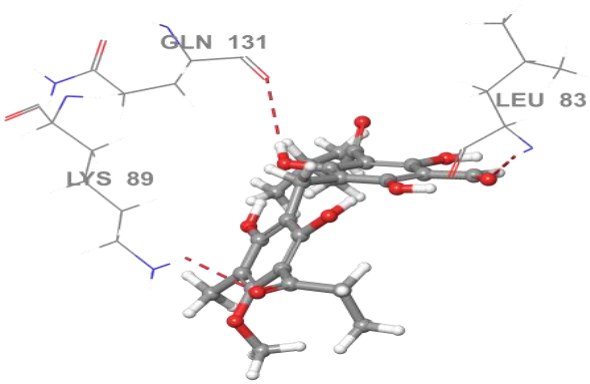
© University of Venda

red-dashed segments denote HBs and black-dashed segments denote π-π stacking interactions, and light blue-dashed segments denote aromatic HBs.

| Docked complex considered | Docking pose of the ligand-target complex | Ligand-target interactions |
|---|---|---|
| dim23 docked on the antimalarial target (PFLDH, PDB: 1U5A) |  |  |
| dim8 docked on the antimalarial target (PFMDH, PDB: 6R8G), site M1. |  |  |
| dim8 docked on the antimalarial target (PFMDH, PDB: 6R8G), site M2. |  |  |



© University of Venda

| | | |
|---|---|---|
| dim23 docked on the antimalarial target (PFMDH, PDB: 6R8G), site M3. |  |  |
| dim23 docked on the antimalarial target (PFMDH, PDB: 6R8G), site M4. |  |  |
| dim24 docked on the antimalarial target (PFMDH, PDB: 6R8G), site M5. |  |  |
| dim24 docked on the lung cancer target (CDK-2, PDB: 1DI8) |  |  |



© University of Venda

**Docking with PFLDH**. PFLDH has one active site where dim23 shows the highest binding affinity, with a binding energy of -10.887 kcal $mol^{-1}$. It forms four H-bonds with VAL233, ALA236, ASN197, and HIS195, along with hydrophobic interactions with LEU237 and SER245. Dim23's binding strength highlights its potential as a PFLDH inhibitor.

**Docking with PFMDH**. PFMDH has five active sites: M1, M2, M3, M4, and M5. Dim8 shows strong binding at site M2, with a binding energy of -8.243 kcal $mol^{-1}$, forming three H-bonds with LYS B:198 and LEU B:159, as well as pi-cation interactions with LYS B:273. At site M3, dim4 demonstrates moderate binding with a binding energy of -7.553 kcal $mol^{-1}$, forming H-bonds with residues ASP C:32 and GLN C:80. Dim8 is thus highl ighted as a viable PFMDH inhibitor, especially at site M2.

**Docking with CDK-2.** CDK-2 has a single active site. Dim23 achieves the highest binding energy, with a score of -10.794 kcal $mol^{-1}$, forming H-bonds with LEU83, LYS89, and GLN131, as well as hydrophobic interactions with LEU134. Dim23's binding profile suggests that it is a strong CDK-2 inhibitor.

### 6.3.3 Docking results for trimeric ACPLs against anticancer targets

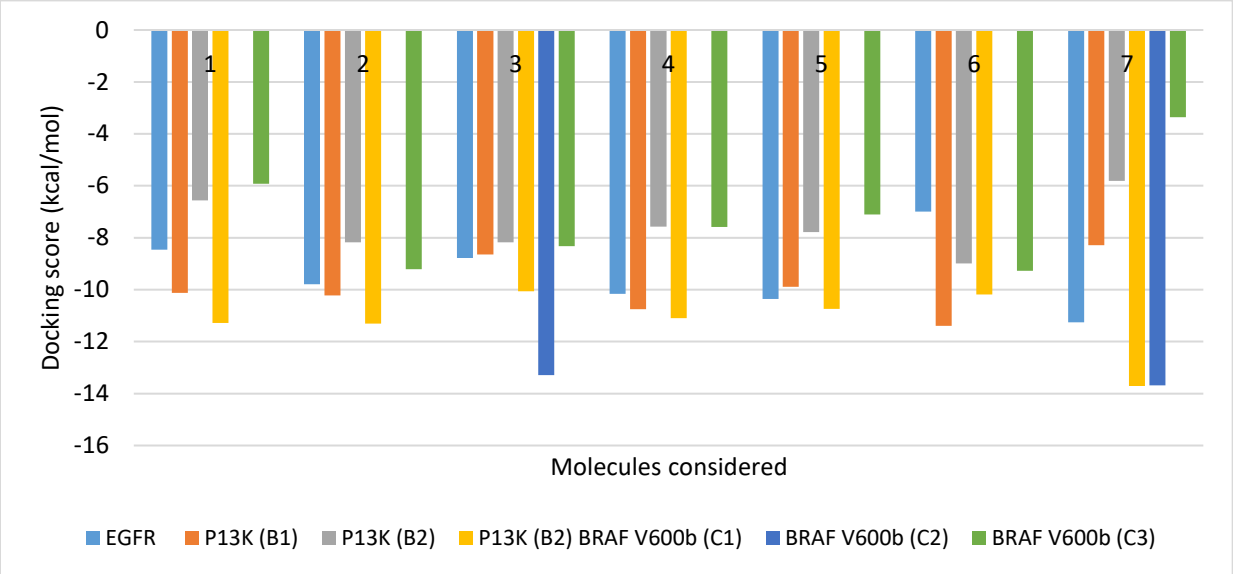
This section presents the molecular docking results for the selected trimeric ACPLs against each anticancer target. Table S6.2 provides a comparison of the interaction energies between the ACPLs and the co-crystal ligand. Figure 6.5. shows docking scores of trimeric ACPLs against each anticancer target. Figure 6.6 shows molecular insights into the ligand-anticancer target interactions, focusing on the specific residues within the binding site pocket that contribute to ligand binding; thus, only cases where the ACPLs exhibit interaction energies better than or comparable to those of the co-crystal ligand are highlighted here.

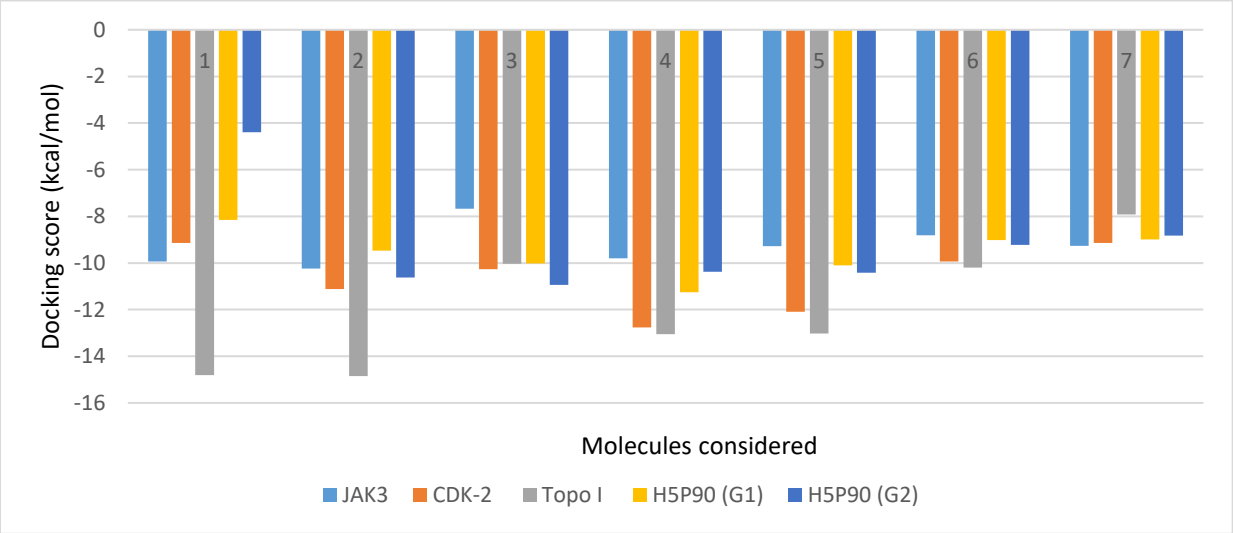
**Figure 6.5.** Docking scores of trimeric ACPLs against each anticancer target.
Numbers denoting the considered molecules on the x-axis.

| # | Conformers | # | Conformers | # | Conformers |
|---|---|---|---|---|---|
| 1 | trim1 | 4 | trim4 | 7 | Co-crystalised ligand |
| 2 | trim2 | 5 | trim5 | | |
| 3 | trim3 | 6 | trim6 | | |



© University of Venda






© University of Venda

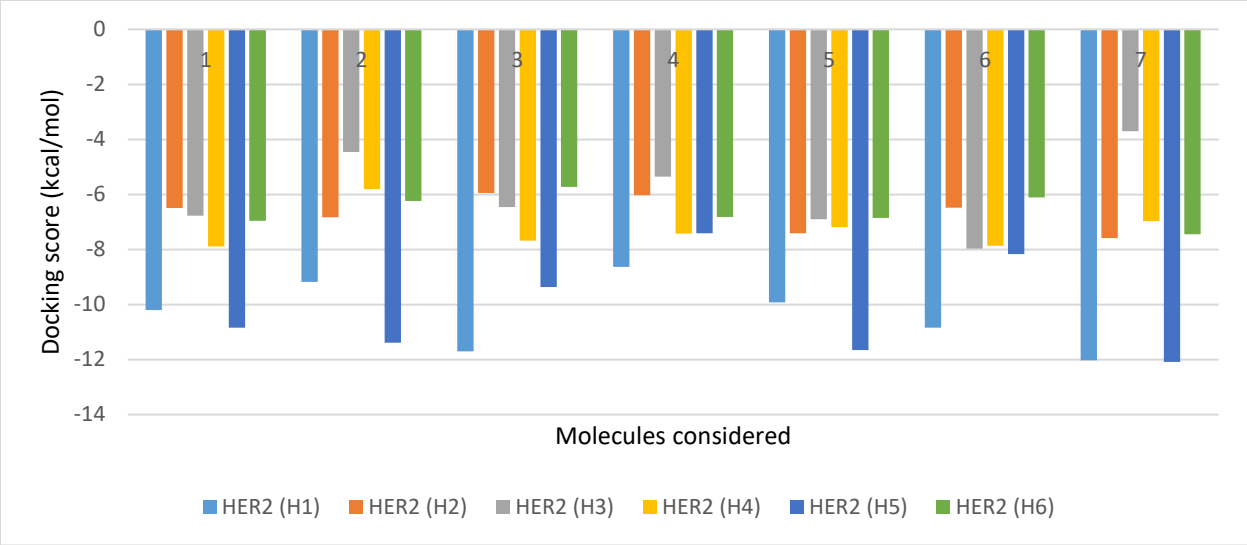
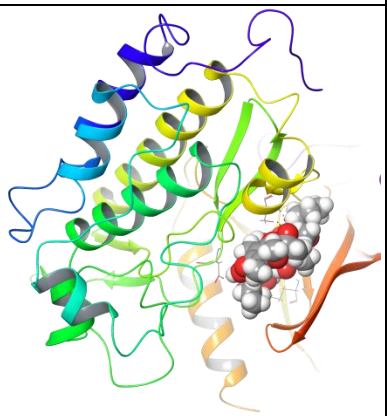
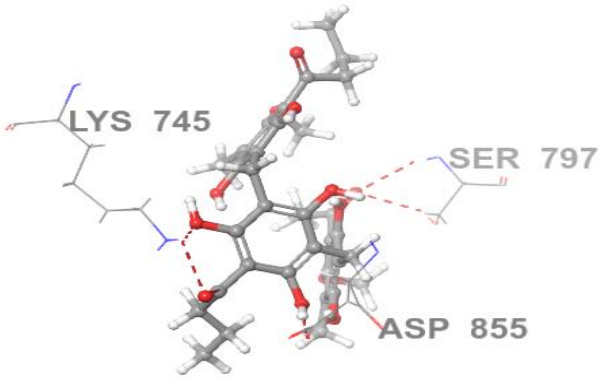
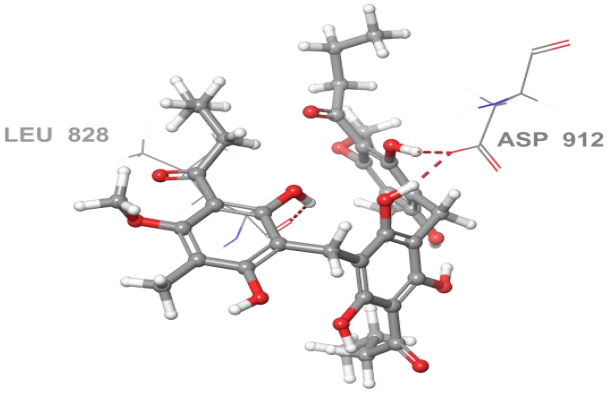
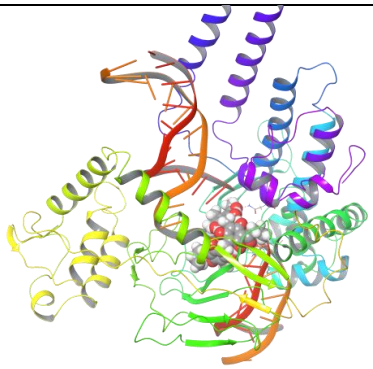
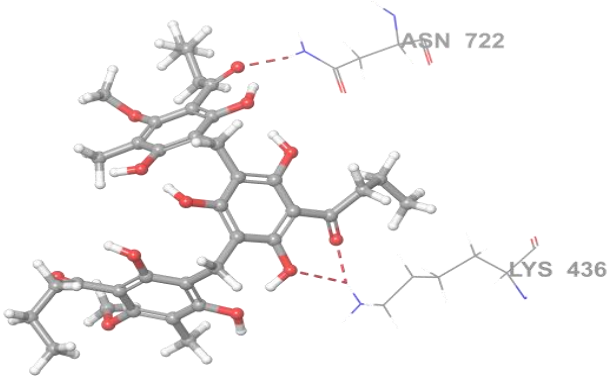
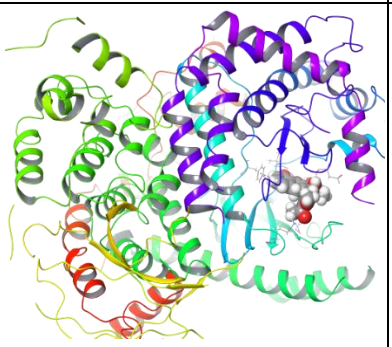
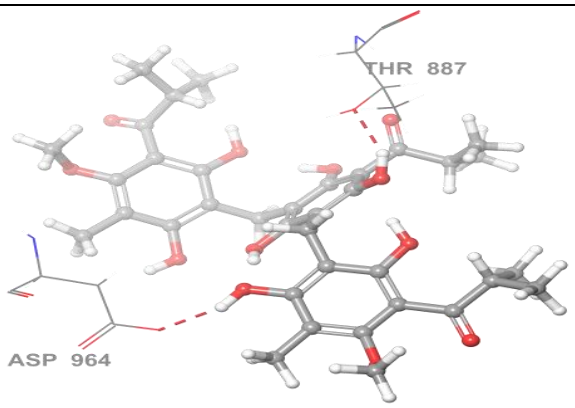



**Figure 6.6**. Molecular insight into trimeric-anticancer target highlighting the residues within the binding site pocket. The diagrams illustrate the interactions between the ligand and the active site of the target, with red-dashed segments denote HBs and black-dashed segments denote π-π stacking interactions, and light blue-dashed segments denote aromatic HBs.

| Docked complex considered | Docking pose of the ligand-target complex | Ligand-target interactions |
|---|---|---|
| trim6 docked on the lung cancer target (EGFR, PDB: 6LUD) |  |   |

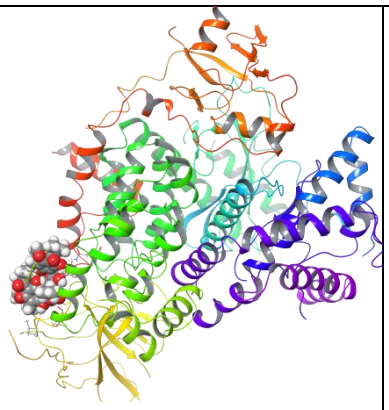
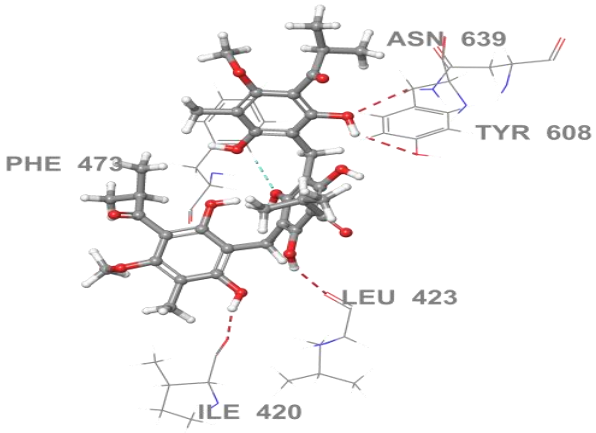
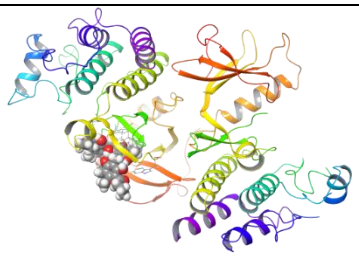
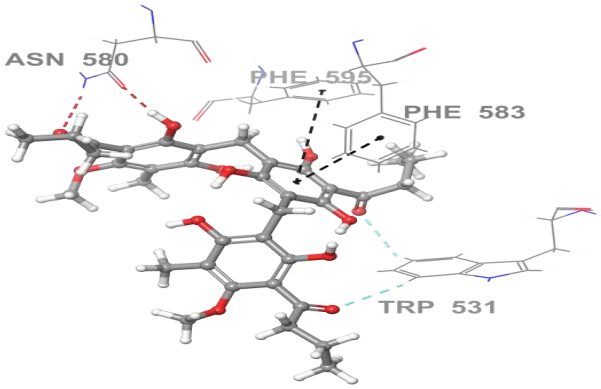
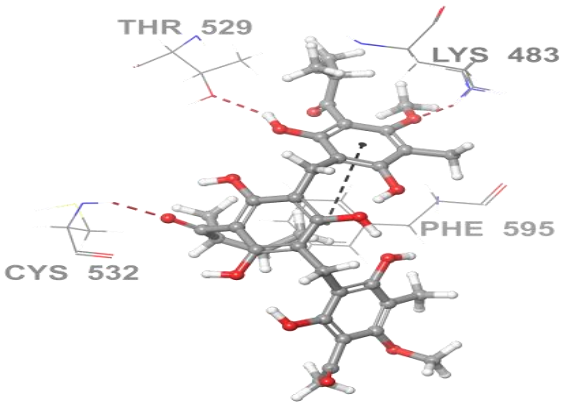


© University of Venda

| | | |
|---|---|---|
| Trim3 docked on the lung cancer target (JAK3, PDB: 7C3N) |  |  |
| trim3 docked on the lung cancer target (Topo I, PDB: IT8I) |  |  |
| Trim7 docked on the lung cancer target (P13K, PDB: 5JHB), site B1. |  |  |

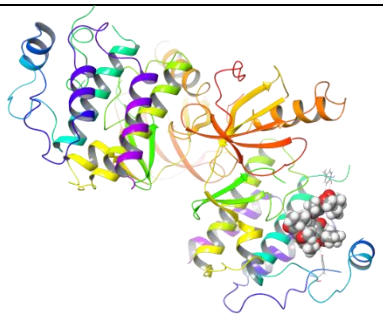
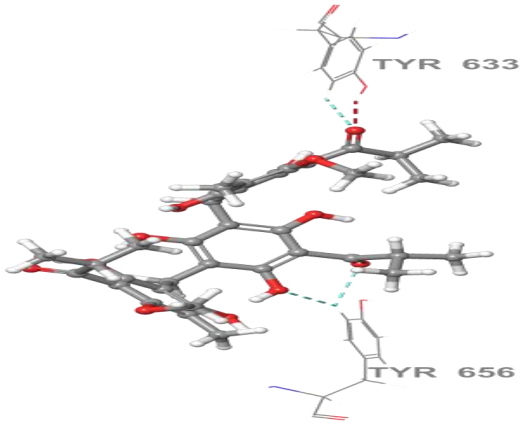
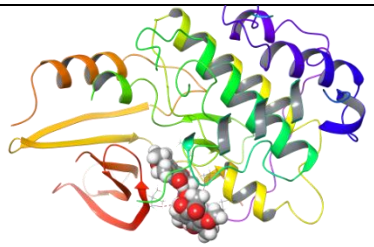
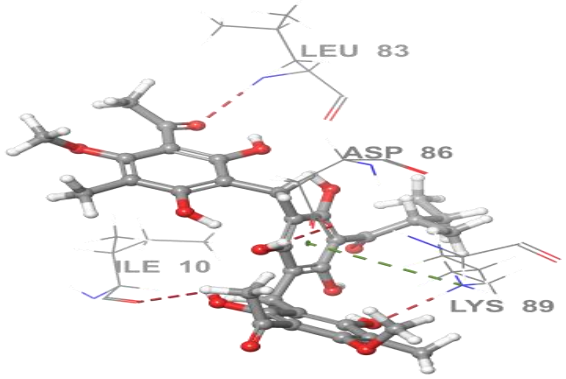
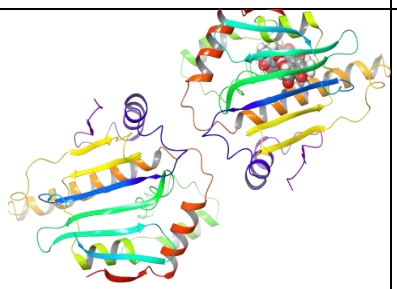
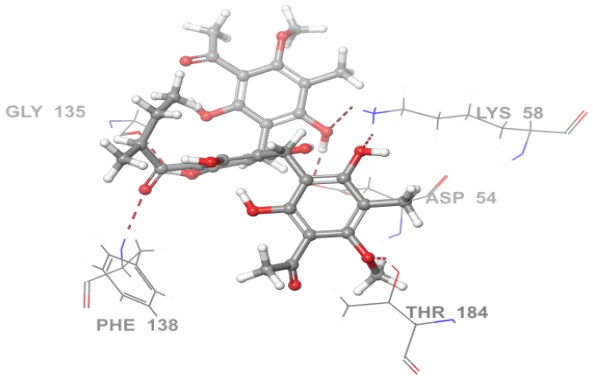
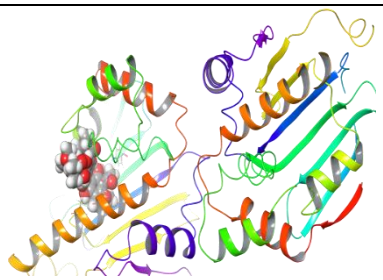
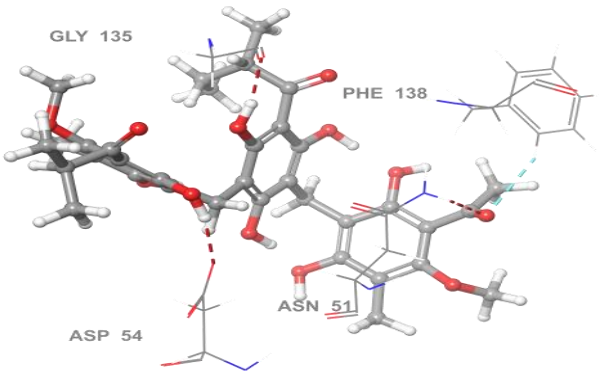


© University of Venda

| | | |
|---|---|---|
| Trim7 docked on the lung cancer target (P13K, PDB: 5JHB), site B2. |  |  |
| trim3 docked on the lung cancer target (BRAF V600b, PDB; 6V34), site C1. |  |  |
| trim4 docked on the lung cancer target (BRAF V600b, PDB; 6V34), site C2. |  |  |

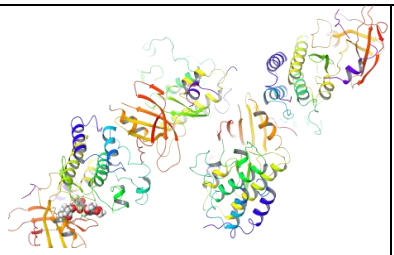
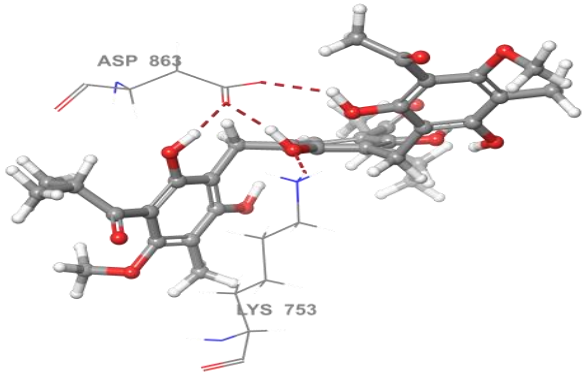
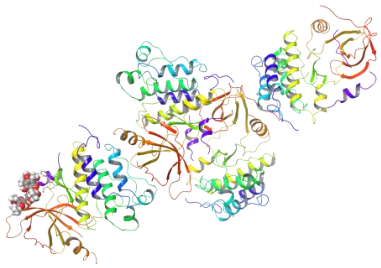
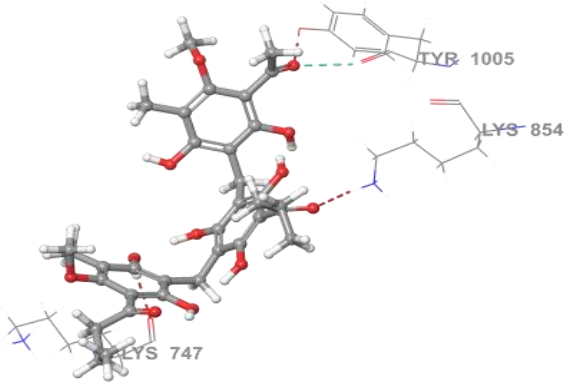
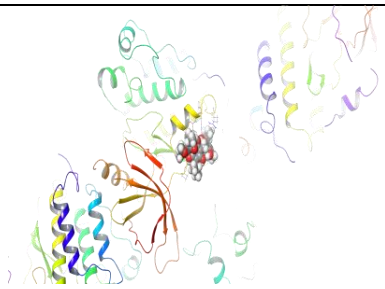
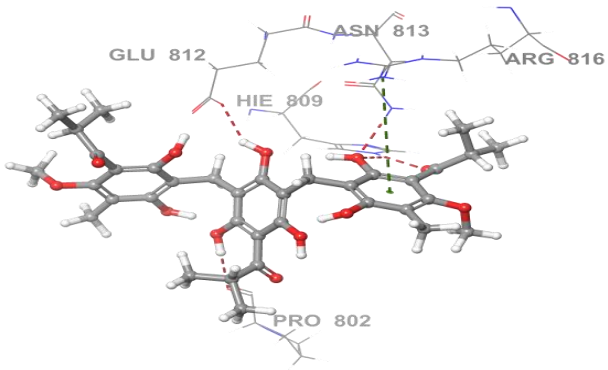
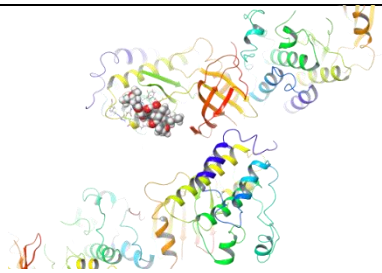
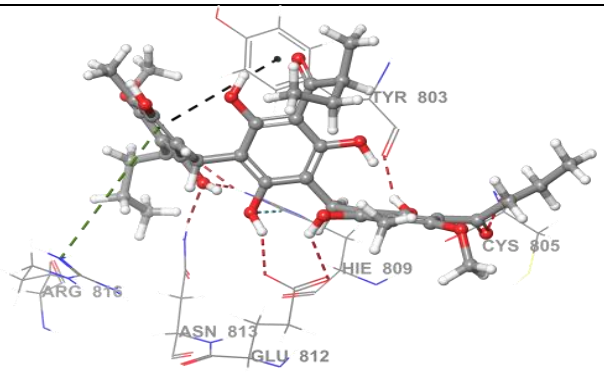


© University of Venda

| | | |
|---|---|---|
| trim7 docked on the lung cancer target (BRAF V600b, PDB; 6V34, site C3. |  |  |
| trim5 docked on the lung cancer target (CDK2, PDB: 1DI8) |  |  |
| trim7 docked on the breast cancer target (H5P90, PDB: 3TUH), site G1. |  |  |
| trim7 docked on the breast cancer target (H5P90, PDB: 3TUH), site G2. |  |  |

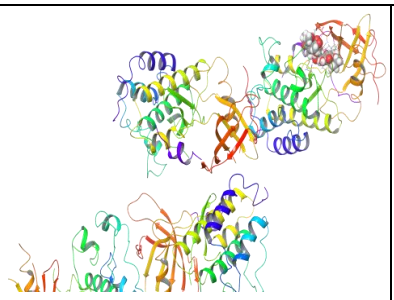
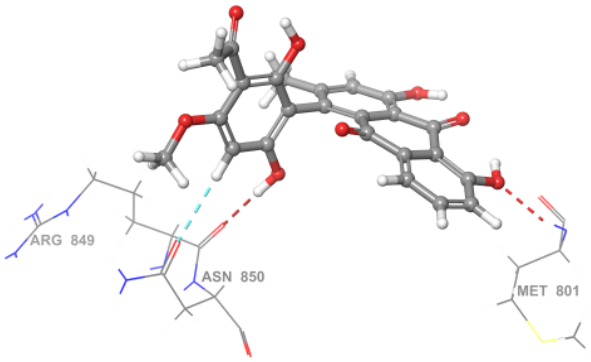
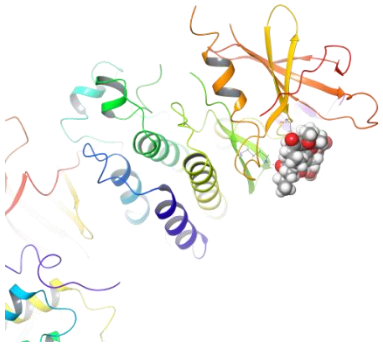
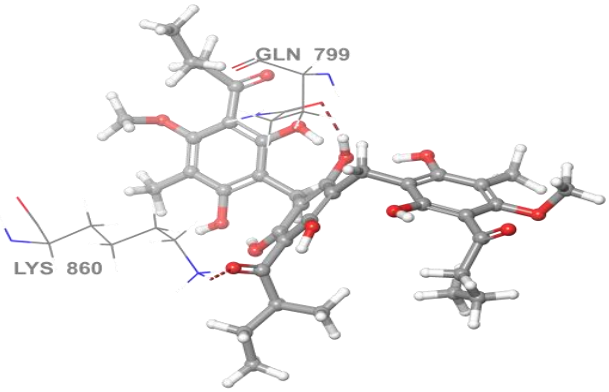


© University of Venda

| | | |
|---|---|---|
| trim4 docked on the breast cancer target (HER2, PDB: 3RCD), site H1. |  |  |
| trim6 docked on the breast cancer target (HER2, PDB: 3RCD), site H2. |  |  |
| trim7 docked on the breast cancer target (HER2, PDB: 3RCD), site H3. |  |  |
| Trim1 docked on the breast cancer target (HER2, PDB: 3RCD), site H4. |  |  |



© University of Venda

| | | |
|---|---|---|
| trim6 docked on the breast cancer target (HER2, PDB: 3RCD), site H5. |  |  |
| trim1 docked on the breast cancer target (HER2, PDB: 3RCD), site H6. |  |  |

**Docking with EGFR.** EGFR has a single active site where trim6 shows the highest binding affinity, achieving a binding energy of -10.359 kcal $mol^{-1}$. It forms four H-bonds with SER797, ASP855, and LYS745, along with hydrophobic interactions with MET790 and LEU718. These results suggest that trim6 could be an effective EGFR inhibitor.

**Docking with JAK3.** JAK3 also has a single active site. Trim3 demonstrates the strongest binding, with a binding energy of -10.24 kcal $mol^{-1}$. It forms three H-bonds with LEU828 and ASP912, along with hydrophobic interactions with ALA966, TYR904, and LEU905. Trim3's binding profile highlights it as a strong JAK3 inhibitor candidate.

**Docking with Topo I.** Trim3 also demonstrates strong binding with Topo I, achieving a binding energy of -14.849 kcal $mol^{-1}$. It forms four H-bonds with LEU83, ASP86, and LYS89, along with pi-cation interactions with LYS89. Trim3's high binding energy suggests that it could be a potent Topo I inhibitor.



© University of Venda

**Docking with PI3K.** PI3K has two active sites, B1 and B2. Trim7 shows the highest binding affinity at site B1, with a binding energy of -11.39 kcal $mol^{-1}$. It forms two H-bonds with THR887 and ASP964, plus hydrophobic interactions with ALA885 and VAL882. At site B2, trim7 achieves a binding energy of -8.985 kcal $mol^{-1}$, forming four H-bonds with ILE420, LEU423, and TYR608. These findings suggest trim7 as a potential PI3K inhibitor across both sites.

**Docking with BRAF V600B.** BRAF V600B has three active sites: C1, C2, and C3. Trim3 shows the strongest binding at site C1, achieving a binding energy of -11.304 kcal $mol^{-1}$. It forms two H-bonds with ASN580 and TRP531 and includes π-π stacking with PHE595. At site C2, trim4 achieves the best score of -13.288 kcal $mol^{-1}$, forming H-bonds with CYS532 and THR529. At site C3, trim7 shows the highest affinity, with a binding energy of -9.277 kcal $mol^{-1}$, forming H-bonds with TYR633. These results highlight trim3, trim4, and trim7 as strong BRAF V600B inhibitors.

**Docking with CDK-2.** CDK-2 has a single active site. Trim5 demonstrates the highest binding energy of -12.768 kcal/mol, forming three H-bonds with ILE10, LEU83, and ASP86, along with pi-cation interactions with LYS89. Trim5 is therefore identified as a strong CDK-2 inhibitor.

**Docking with HSP90.** All the considered trimeric compounds demonstrated binding to both binding sites on HSP90. At the G1 site, trim4 displayed the best docking with a binding score energy of -11.257 kcal/mol, which is comparatively stronger than that of the co-crystal ligand. Trim4 forms six hydrogen bonds (HBs) with residues LYS58, THR184, ASP54, PHE138, and GLY135. In addition, hydrophobic interactions occur with residues VAL150, TYR139, ALA55, SER52, and ALA111, as well as ILE110, stabilizing the binding. At the G2 site, trim7 showed the strongest docking affinity, with a binding score energy of -10.94 kcal/mol, also stronger than that of the co-crystal ligand. This interaction involves three HBs through GLY135, ASN51, and ASP54, along with one aromatic HB with PHE138. Additionally, hydrophobic interactions occur with ALA55 and ILE91, complemented by polar interactions involving SER50, ASN51, SER52, and LYS58.



© University of Venda

**Docking with HER2.** All the considered trimeric compounds are putative binders of HER2. Notably, active sites H1, H4, and H5 showed better binding score energies with these compounds compared to H2 and H3. The best-docked complexes from each active site were identified as follows: trim4/HER2 for H1, trim6/HER2 for H2, trim7/HER2 for H3, trim1/HER2 for H4, trim6/HER2 for H5, and trim1/HER2 for H6, with binding score energies of -11.707, -7.416, -7.969, -7.887, -11.656, and -6.961, respectively.

At active site H1, the interaction between trim4 and HER2 was observed. However, this interaction was not significantly stronger than that of the co-crystal ligand. Trim4 interacts through four H-bonds (HBs) with ASP863 and LYS753 and also engages in hydrophobic interactions with residues VAL797, LEU796, PHE731, ALA730, PHE1004, CYS805, LYS753, ILE752, and ALA751, as illustrated in Figure 6.3.34. At active site H2, the interaction of trim6 with HER2 showed a binding affinity relatively close to that of the co-crystal ligand. Trim6 interacts through three HBs with residues TYR1005, LYS854, and LYS747 and forms one aromatic HB with TYR1005, as shown in Figure 6.3.35. At active site H3, the interaction of trim7 with HER2 was notably stronger than that of the co-crystal ligand. Trim7 forms three HBs with HEI809, ASN813, and PRO802 and a pi-cation interaction with ARG816. Additionally, hydrophobic interactions occur with PRO802. At active site H4, the interaction of trim1 with HER2 did not exceed the binding strength of the co-crystal ligand. Trim1 interacts through five HBs with ASN813, GLU812, HIE809, CYS805, and TYR803. Additionally, there is one aromatic HB with HIE809, a pi-cation interaction with ARG816, and $\pi$-$\pi$ stacking with TYR803. Furthermore, hydrophobic interactions occur with ASN813, and polar interactions with HIE809 and ASP808, as well as glycine interactions with GLY804 and CYS805. At active site H5, the interaction of trim6 with HER2 showed binding energy comparable to the co-crystal ligand. Trim6 interacts through six HBs with ALA730, PHE731, ASN850, LYS753, and ASP863. Additionally, there is one aromatic HB with PHE731 and a pi-cation interaction with LYS753. Hydrophobic interactions include PHE731, LEU796, LEU785, and CYS805. At active site H6, the interaction of trim1 with HER2 was not significantly stronger than that of the co-crystal ligand. Trim1 interacts through two HBs with GLN799 and LYS860, and engages in hydrophobic interactions with LEU800 and PRO802, as well as polar interactions with ILE748, PRO749, and GLN799.



© University of Venda

### 6.3.4 Docking results for trimeric ACPLs against antimalarial targets

This section presents the results of the molecular docking of the considered trimeric ACPLs against each target selected for the antimalarial activity (all related to *Plasmodium Falciparum*). Figure 6.7. shows docking scores of dimeric ACPLs against each antimalarial target. Figure 6.8. shows molecular insights into the ligand-anticancer target interactions, focusing on the specific residues that contribute to ligand binding within the binding site pocket.

**Figure 6.7.** Docking scores of trimeric ACPLs against each antimalarial target.

Numbers denoting the considered molecules on the x-axis.

| # | Conformers | # | Conformers | # | Conformers |
|---|---|---|---|---|---|
| 1 | trim1 | 4 | trim4 | 7 | Co-crystalised ligand |
| 2 | trim2 | 5 | trim5 | | |
| 3 | trim3 | 6 | trim6 | | |

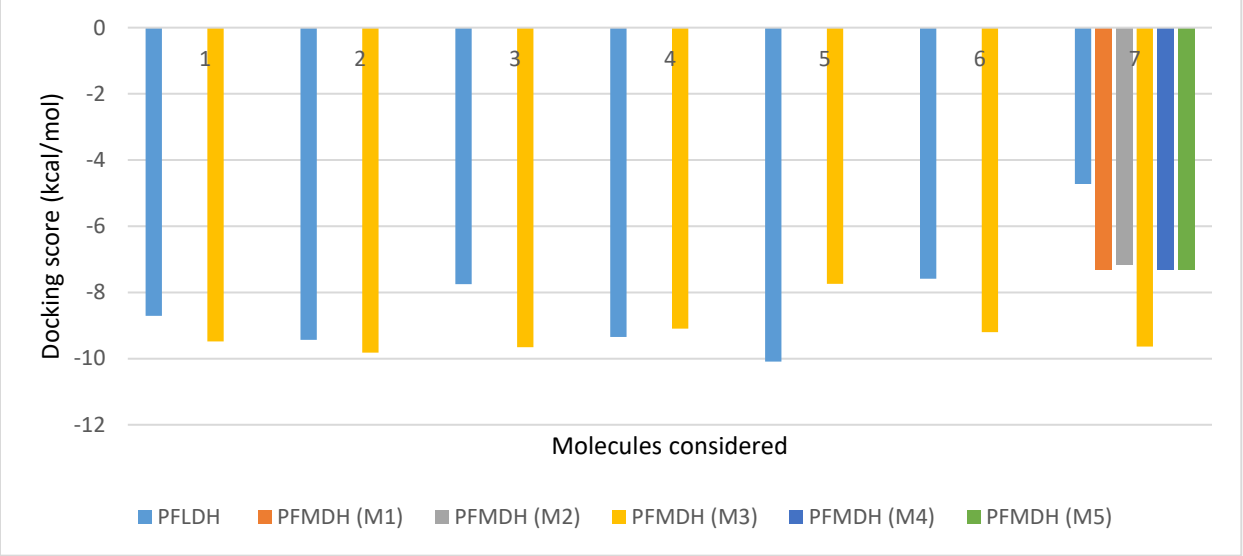
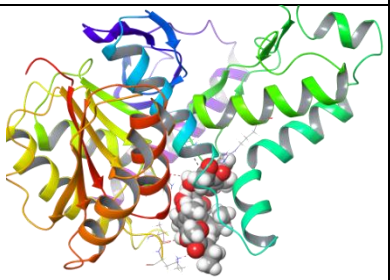
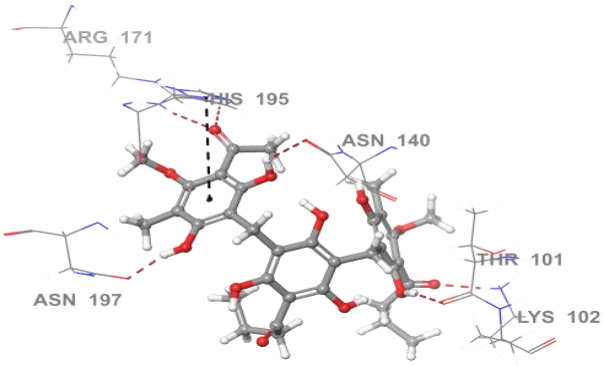
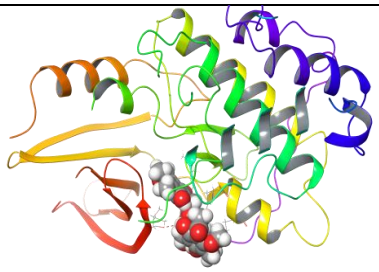
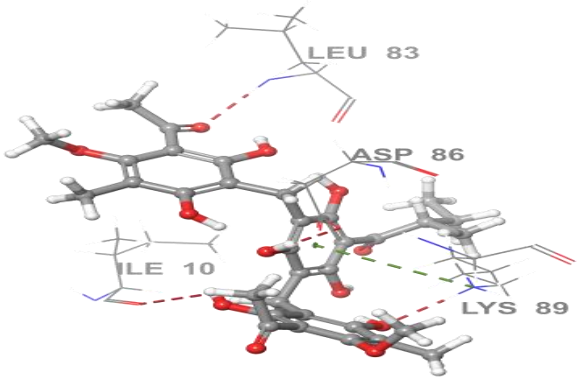



**Figure 6.8.** Molecular insight into trimeric- antimalarial target highlighting the residues within the binding site pocket. The diagrams illustrate the interactions between the ligand and the active site of the target, with red-dashed segments denote HBs and black-dashed segments denote π-π stacking interactions, and light blue-dashed segments denote aromatic HBs.



© University of Venda

| Docked complex considered | Docking pose of the ligand-target complex | Ligand-target interactions |
|---|---|---|
| trim6 docked on the antimalarial target (PFLDH, PDB: 1U5A) |  |  |
| trim3 docked on the antimalarial target (PFMDH, PDB: 6R8G), site M3. |  |  |

**Docking with PFLDH.** PFLDH has one active site where trim6 achieves the strongest binding, with a binding energy of -10.092 kcal/mol. Trim6 forms five H-bonds with ARG171, ASN140, and THR101, as well as π-π stacking with HIS195. Trim6 shows promise as a PFLDH inhibitor.

**Docking with PFMDH.** PFMDH has five active sites: M1, M2, M3, M4, and M5. Trim3 shows the highest affinity at site M3, with a binding energy of -9.824 kcal/mol, forming two H-bonds with ASP C:32 and GLN C:80. These findings suggest trim3 as a viable PFMDH inhibitor.

### 6.4. Analysis of physicochemical properties

The results suggest that eight of the 23 dimeric compounds studied violate Lipinski's rule of five (Table S6.3), indicating potential issues with oral activity. Among these, six compounds (dim2, dim5, dim8, dim12, dim15, and dim20) have one violation each, while dim1 and dim7 show multiple violations. This implies that only two of the 23 dimeric ACPLs may be inactive orally.



© University of Venda

On the other hand, all considered trimeric ACPLs show multiple violations, indicating that these compound classes may not be orally active. The reported violations mainly involve higher molecular weight (mol MW) or higher QPlogPo/w values, suggesting increased lipophilicity and potentially poor absorption rates for these compounds.

The calculated Caco-2 and MDCK cell permeability values (in nm/sec) indicate that all considered compounds fall within the acceptable range (<25 considered poor, >500 considered great). However, the permeability values vary significantly depending on the type or nature of the molecule. For dimeric ACPLs, dim1 and dim11 show better permeability compared to others, with dim1 demonstrating a Caco-2 cell permeability greater than 500 nm/sec. Among trimeric ACPLs, trim1 and trim6 demonstrate better permeability, although still below 500 nm/sec.

The calculated brain/blood partition coefficient (QPlogBB) values for the considered molecules are generally within the acceptable range (-3 to 1.2), except trim2, which has a QPlogBB value of -3.029. The type of the subclass of ACPLs significantly influences the QPlogBB values, with each subclass having its range. For dimeric ACPLs, the calculated QPlogBB values range from -2.469 to -1.284. Trimeric ACPLs have QPlogBB values ranging from -3.029 to -2.476. It is worth mentioning that all considered trimeric ACPLs have lower blood-brain barrier (BBB) penetration than the other class of ACPLs considered.

The calculated QPlogKhsa values, which assess the binding capacity of compounds to HSA, fall within acceptable ranges for most of the considered molecules. Specifically:

- For dimeric ACPLs, the QPlogKhsa values range from -0.469 to -1.481.
- Trimeric ACPLs show QPlogKhsa values ranging from 0.679 to -1.208.

The calculated QPlogKhsa values, which assess the binding capacity of compounds to HSA, fall within acceptable ranges for most of the considered molecules. Specifically:

- For dimeric ACPLs, the QPlogKhsa values range from -0.469 to -1.481.
- Trimeric ACPLs show QPlogKhsa values ranging from 0.679 to -1.208.



© University of Venda

The calculated percentage of human oral absorption falls within acceptable ranges for all subclasses of ACPLs:

- For dimeric ACPLs, the percentage of human oral absorption ranges from 71.420% to 100.000%, with an outlier value of 51.214% for dim2.
- Trimeric ACPLs demonstrate percentages ranging from 44.177% to 57.673%.

These values suggest that the considered ACPLs have favorable potential for oral absorption in humans, with trimeric ACPLs showing a moderate range of absorption percentages compared to dimeric and PCP ACPLs.

### 6.5. Concluding remarks

The aim of the study was to identify potential anticancer and antimalarial compounds using computational docking methods. Drug discovery can be likened to a multi-stage filtering process, where molecules are progressively evaluated, and only the most promising ones advance through each stage. An ideal drug candidate must meet several key criteria: strong binding affinity to biological targets, stable molecular interactions such as hydrogen bonding and hydrophobic interactions, and adherence to drug-likeness rules like Lipinski's rule of five. In addition, compounds exhibiting the ability to interact with multiple targets are particularly valuable, as they may reduce the likelihood of drug resistance by acting on several proteins at once.

Through computational analysis, 23 dimeric and 6 trimeric ACPLs were docked against various cancer and malaria targets. Among these, ten compounds stood out to be promising: dim2 showed the strongest binding with both EGFR and JAK3, making it a strong candidate for cancer therapy; dim23 was particularly effective against Topo I, a target involved in cancer cell proliferation; trim3 exhibited excellent polypharmacological properties, interacting strongly with JAK3, Topo I, and PFLDH, making it a versatile candidate for both anticancer and antimalarial therapies; dim14 and trim7 showed high binding scores with PI3K at site B1, while dim19 and trim7 excelled at site B2; dim3 and trim3 showed the best docking results with BRAF V600B, specifically at site C1; dim15 and trim4 demonstrated strong interactions with H5P90 at site G1, while dim23 and trim7 excelled



© University of Venda

at site G2. These molecules demonstrated stable hydrogen bonding and hydrophobic interactions, which are key for effective drug-receptor binding. Furthermore, the most promising candidates which showed polypharmacological includes dim2, dim23, trim3, and trim7. These findings provide a solid starting point for further experimental work, with the potential for significant contributions to drug discovery and development.

Dimeric ACPLs show a better physicochemical profile than trimeric ACPLs regarding Lipinski's rule of five. ACPLs that showed violations of the rule of five cannot be limited to further studies because studies have also shown that some natural products violate the chemical rules. Furthermore, trimeric also shows better physicochemical properties. However, it’s important to acknowledge that these results are based on computational predictions. While the identified compounds show great potential, further experimental and clinical validation is necessary to confirm their therapeutic effectiveness. Nevertheless, this research provides valuable insights into the molecular interactions of ACPLs with their targets, laying a strong foundation for future drug development efforts.



© University of Venda

## 6.6. References

1. Mammino, L. (2019). Intramolecular hydrogen bonding patterns, conformational preferences and molecular properties of dimeric acylphloroglucinols: An ab initio and DFT study. *Journal of Molecular Structure, 1176*, 488-500.
2. Mammino, L. (2017). Bowl-shaped structures from acylphloroglucinols: An ab initio and DFT study. *Molecular Physics, 115*(17-18), 2254-2266.



© University of Venda

# CHAPTER 7

## 7.0. Introduction to the chapter

This chapter reflects on the key findings of the computational study on selected antimalarial and anticancer ACPLs. These ACPLs were chosen to evaluate their potential as drug candidates against malaria and cancer, addressing the critical need for new therapies to combat these diseases. The study used computational methods to analyze different ACPLs and identify which compounds show the most promise for further development. While the findings have provided valuable insights, several areas require further research to strengthen and expand the scope of the study.

## 7.1. Summary of main findings

The results of this study offer important insights into the molecular properties, stability, and potential therapeutic applications of the selected ACPLs. One of the key findings is the significant role of IHBs in stabilizing the conformers. Specifically, the hydrogen bonds between hydroxyl groups and oxygen atoms—such as H17···O14 or H15···O14 across all ACPLs, and H23···O32 and H25···O32 in compounds with additional $sp^2$ oxygen atoms—were identified as key factors in stabilizing these conformations. These IHBs contribute to the structural stability of the ACPLs and influence their reactivity and ability to interact with biological targets.

In addition to the role of IHBs, the positioning of acyl groups in certain ACPLs, particularly euglobals, was shown to influence molecular properties, including dipole moments and energy profiles. The acyl groups, located at C1 and C5, impact the reactivity of the molecules, with different positioning leading to distinct behavior. This variability highlight the importance of the spatial arrangement of functional groups in determining how these molecules behave in various environments. Overall, the conformational preferences of the ACPLs, shaped by both IHBs and acyl group positioning, are crucial for their interactions with biological systems.

The study also highlighted the effects of solvent polarity on molecular behavior. As the polarity of the medium increased, the HOMO-LUMO energy gap either decreased or increased, depending on the specific conformer or molecule. This relationship between solvent polarity and the energy gap is important because the HOMO-LUMO gap is a key descriptor of molecular stability and



© University of Venda

reactivity. Furthermore, the slight increase in dipole moment in more polar solvents suggests that the molecular conformation adjusts to the surrounding environment, enhancing its stability.

The docking studies provided important insights into the intermolecular interactions between ACPLs and various disease-related proteins. Hydrogen bonding, facilitated by the hydroxyl groups and $sp^2$ oxygen atoms present in ACPLs, was a critical factor in stabilizing ligand-receptor complexes. These interactions are fundamental for the therapeutic potential of the compounds, as they help ensure that the molecules bind effectively to their target proteins. Aromatic rings in the ACPLs also played a significant role, contributing to interactions through aromatic hydrogen bonds, π-cation interactions, and π-π stacking. These additional forces further stabilized the molecular complexes, enhancing the molecules' binding affinity and selectivity.

The study identified several promising ACPL compounds (dim23, trim3, trim6, and trim7) that demonstrated strong interactions across multiple disease-related protein targets, including those associated with cancer (e.g., EGFR, JAK3, Topoisomerase I, HER2) and malaria (e.g., PFLDH, PFMDH). This broad range of interactions suggests that these compounds possess polypharmacological properties. This is particularly advantageous for diseases like cancer and malaria, where drug resistance often develops due to the ability of pathogens or cancer cells to bypass single-target therapies. The ability of these ACPLs to target multiple proteins simultaneously could provide synergistic therapeutic effects, potentially reducing the risk of resistance.

Despite some violations of Lipinski's rule of five, which assesses drug-likeness based on molecular properties, the dimeric and trimeric ACPLs still hold potential. Natural products often defy conventional chemical rules, yet they exhibit therapeutic efficacy. The compounds identified in this study, despite their deviations from Lipinski's rule, may still be practical drug candidates, especially when considering their strong binding interactions and favourable physicochemical properties.



© University of Venda

## 7.2. Reflections and future directions of the study

This study has provided valuable insights into the molecular behavior of selected ACPLs, but several limitations point to opportunities for further research. One limitation is the narrow range of molecules studied. Expanding the research to include a wider variety of ACPL derivatives could offer a better understanding of how structural modifications impact biological activity. Introducing different substituents, such as halogens or hydroxyl groups, might reveal trends that help guide the design of more effective compounds.

The computational methods used in this study, such as HF, DFT, and MP2, provided reliable results. However, more accurate methods like coupled-cluster calculations (e.g., CCSD(T)) could be used in future studies to improve energy estimates, particularly for systems with significant electron correlation.

In addition, incorporating density functional theory (DFT) with Grimme's dispersion correction (DFT-D) could improve the accuracy of the calculations. This correction accounts for weak intermolecular forces like van der Waals interactions and hydrogen bonding, which are important in biological systems. It would provide more accurate predictions of binding energies and conformational preferences for ACPLs.

Hybrid methods such as Quantum Mechanics/Molecular Mechanics (QM/MM) simulations could also be explored. These methods treat the active site of a system quantum mechanically while modeling the surrounding environment using classical mechanics. This approach could offer a more detailed understanding of how ACPLs interact with larger biomolecular systems.

Future research could also benefit from using explicit solvent models instead of implicit ones. Explicit solvent models better replicate biological conditions by representing individual solvent molecules, leading to more accurate predictions of solvation effects in protein-ligand interactions.

Docking studies could be improved by incorporating flexible docking protocols. These protocols account for protein flexibility during binding, providing a more realistic representation of protein-



© University of Venda

ligand interactions. Including explicit water molecules in docking simulations could also help understand water-mediated interactions that influence ligand binding.

Lastly, future studies could investigate how structural changes in ACPL derivatives affect docking outcomes, such as binding affinities and interaction modes. By combining structural insights with advanced computational techniques like free-energy perturbation or molecular dynamics, predictions could be refined further to simulate more realistic physiological environments.

These considerations for future studies would build on the current findings and contribute to a more comprehensive understanding of ACPLs. This approach could lead to their broader therapeutic applications in treating a variety of diseases.

### 7.3. Machine learning applications

Given the volume of data generated in this study, there is significant potential for incorporating machine learning techniques to improve the efficiency of data analysis. Machine learning models can help predict the biological activity of new ACPL derivatives based on their structural properties, reducing the need for labour-intensive computational simulations.

For instance, a machine learning algorithm could be trained to predict binding affinities based on molecular descriptors derived from the current dataset. This would allow for faster screening of new compounds, enabling researchers to focus on the most promising candidates. Machine learning could also assist in identifying patterns in the data that might not be immediately obvious, providing a deeper understanding of the molecular features that contribute to strong binding interactions. Moreover, machine learning could be used to optimize the molecular docking process by predicting the most likely binding conformations for ACPLs. This would streamline the drug discovery process by reducing the need for exhaustive conformational searches.

### 7.4. Conclusion

This ongoing study at the University of Venda has made significant contributions to the understanding of ACPLs and their potential as therapeutic agents. The results so far have provided



© University of Venda

valuable insights into the molecular properties of ACPLs, but there is much more to explore. Future research should focus on expanding the range of ACPL derivatives studied, integrating machine learning techniques for data analysis, and using more advanced computational methods such as dispersion-corrected DFT and QM/MM hybrid simulations. These approaches will help build on the current findings and move closer to the development of ACPL-based therapies with real clinical potential.



© University of Venda