THAI NGUYEN UNIVERSITY
UNIVERSITY OF EDUCATION

Hoang Viet Cuong

STUDY OF CHLOROPLAST GENOME CHARACTERISTICS AND ANALYSIS OF SAPONIN ACTIVITY ISOLATED FROM "LAN TAI CAO" (*HOYA VERTICILLATA* VAR. *VERTICILLATA*)

Speciality: Genetics

Code: 9420121

DISSERTATION SUMMARY

THAI NGUYEN - 2026

This dissertation was completed at the Department of Genetics and Biotechnology, Faculty of Biology, Thai Nguyen University of Education.

Supervisors: **1. Prof. Dr. Chu Hoang Mau**
2. Assoc. Prof. Dr. Tu Quang Tan

Reviewer 1:

Reviewer 2:

Reviewer 3:

The dissertation will be defended before the University-level Dissertation Examination Committee convened at Thai Nguyen University of Education, Thai Nguyen University,
at ……, on …… ……, 2026.

The dissertation may be consulted at:

1. National Library of Vietnam
2. Digital Center - Thai Nguyen University
3. Library of Thai Nguyen University of Education

PUBLICATIONS RELATED TO THE DISSERTATION

Journal Articles

1. Hoang, C. V., Tu, T. Q., Nguyen, L. T. N., Nguyen, H. D., Nguyen, Q. H., & Chu, M. H. (2023). Two New C21 Steroidal Glycosides from the Leaves of *Hoya parasitica*. *Records of Natural Products*, 17(6), pp. 1046-1051. http://doi.org/10.25135/rnp.419.2307.2831 (SCI-E).

2. Cuong Viet Hoang, Tan Quang Tu, Thu Thi Mai Lo, Mau Hoang Chu (2024), "Dataset on intergenic spacer regions of chloroplast genome and potential DNA barcode of *H. verticillata* var. *verticillata* in Vietnam", *Data in Brief*, 54, 110471. https://doi.org/10.1016/j.dib.2024.110471 (Scopus).

3.Hoang Viet Cuong, Pham Thi Thu Hien, Tu Quang Tan, Chu Hoang Mau (2024). "*Phylogenetic analysis based on psbz-rps14 and rbcl-accd sequences to support the identification of Hoya verticillata var. verticillata*". Journal of Science and Technology, Thai Nguyen University, 230(01), 177–184. (ACI)

4. Cuong Viet Hoang, Tan Quang Tu, Hung Duc Nguyen, Mau Hoang Chu (2025), "*In silico* studies of saponins from *Hoya verticillata* var. *verticillata* with important apoptosis potency", *Letters in Organic Chemistry*, 22 (11), pp. 846 - 854. https://doi.org/10.2174/0115701786363779250528013022 (SCI-E).

Chloroplast Genome Sequence Registered in GenBank

1. Jo,S., Mau,C.H., Tan,T.Q., Cuong, H.V., Lee,C., Quan,N.H., Thuong,S.D. and Lan,N.T.N. (2023), "*Hoya verticillata* chloroplast, complete genome", GenBank: OR475244.1; https://www.ncbi.nlm.nih.gov/nuccore/OR475244.1

INTRODUCTION

1. Rationale

Hoya verticillata var. *verticillata*, also referred to as *Hoya parasitica* (Roxb.) Wall. ex Traill is a perennial epiphyte that occurs in tropical moist forests and belongs to the *Hoya* genus. *Hoya* is a genus of flowering plants in the Apocynaceae family, with most species being vines with succulent leaves and attractive inflorescences. *H. verticillata* var. *verticillata* occurs in areas ranging from eastern India to southern China and some countries in Southeast Asia, such as the Philippines, Indonesia, Laos, Myanmar, Cambodia, Thailand, and Vietnam.

H. verticillata var. *verticillata* is highly variable and bears some morphological similarities with some other species within the *Hoya* genus. In particular, when the plant material is fragmented or in powdered form, species identification becomes difficult, and the species may be confused with other species within the same genus based on morphology alone. It is therefore imperative to sequence its chloroplast genome and identify chloroplast genomic regions that can act as molecular markers for differentiating *H. verticillata* var. *verticillata* from other species within the *Hoya* genus.

Research on the chloroplast genome in plants has led to significant advances and is considered to be of major importance in the improvement of the understanding of the mechanisms of photosynthesis, plant taxonomy, and evolution. The analysis of the chloroplast DNA (cpDNA) has been found to be valuable in the understanding of plant evolution, diversity, and phylogeny. The application of next-generation sequencing technologies has accelerated the construction of cpDNA databases. The development of specific databases and the creation of bioinformatics tools have improved the handling and analysis of the large amount of chloroplast DNA sequence data generated.

It was found through various studies that *H. verticillata* var. *verticillata* has been used as a medicinal plant for acute kidney failure and for the

treatment and healing of wounds. Preliminary studies have shown the presence of flavonoids, phenolics, and saponins in the species of the *Hoya* genus, including *H. verticillata* var. *verticillata*. However, the chemical components, the search for new compounds, and the biological activity as the basis for the medical application are still limited. The following questions arise from the above analysis: What bioactive compounds are found in the H. verticillata var. verticillata plant? Are the bioactive compounds applicable in medicine?

For these reasons, we conducted the doctoral research entitled: **Study of chloroplast genome characteristics and analysis of saponin activity isolated from "Lan tai cáo" (*Hoya verticillata* var. *verticillata*).**

2. Research objectives

To analyze the characteristics of the chloroplast genome and establish a phylogenetic tree of *H. verticillata* var. *verticillata* and other species of the genus *Hoya*.

To isolate and elucidate the structures of saponins extracted from *H. verticillata* var. *verticillata* and to analyze the *in silico* biological activities of the isolated saponins.

3. Research contents

3.1. Chloroplast genome sequencing of *H. verticillata* var. *verticillata*

i) Collect plant samples and identify them to prepare materials for molecular biological and phytochemical analyses.

ii) Identify *H. verticillata* var. *verticillata* using comparative morphological methods.

iii) Sequence chloroplast DNA nucleotides and establish the chloroplast genome map of *H. verticillata* var. *verticillata*. Describe its chloroplast genome characteristics.

3.2. Phylogenetic analysis, survey of highly nucleotide-divergent DNA markers from the chloroplast genome, and proposal of potential DNA barcodes for species/genus identification

i) Obtain the sequence data of the chloroplast genome from *Hoya* species and perform the analysis using the chloroplast DNA sequences from the species.

ii) Survey the DNA sequences with high nucleotide sequence divergence and construct phylogenetic trees using the DNA sequences retrieved from GenBank, and propose the potential DNA barcodes for species and genus identification in the *Hoya* plant species.

3.3. Study on the isolation, structural elucidation, and biological activity analysis of several saponins from *H. verticillata* var. *verticillata*

(i) Evaluation of the biological activities of saponin-containing fractions from the leaves of *H. verticillata* var. *verticillata*.

(ii) Isolation of saponins and structural elucidation of the isolated saponins.

(iii) Analysis of the cytotoxic activity of saponins isolated from *H. verticillata* var. *verticillata* using *in silico* molecular interaction and molecular dynamics models.

4. Novel contributions of the dissertation

4.1. This dissertation is a novel study on the characteristics of the complete chloroplast genome of *H. verticillata* var. *verticillata*. It proposes potential DNA barcodes to support species discrimination and genus identification within *Hoya*.

4.2. Two new saponins were isolated and structurally characterized from the leaves of *H. verticillata* var. *verticillata*: 3β,20-dihydroxy-pregn-5-ene-3-*O*-β-D-fucopyranoside 20-*O*-β-D-glucopyranoside and 3β,20-dihydroxy-pregn-5-ene-3-*O*-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranoside-20-*O*-β-D-glucopyranoside.

4.3. The interaction capacity and *in silico* molecular dynamics of saponins isolated from *H. verticillata* var. *verticillata* were evaluated, and parasiticoside B (3β,20-dihydroxy-pregn-5-ene-3-*O*-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranoside-20-*O*-β-D-glucopyranoside) was identified as showing the highest cytotoxic potential against the cancer-related protein

Bcl-2. Parasiticoside B is therefore a promising candidate for further experimental investigation targeting Bcl-2 in cancer.

5. Scientific and practical significance of the dissertation

The findings of the dissertation possess both scientific and practical significance in the approach to molecular evolutionary studies based on chloroplast genome sequences, as well as in the isolation and activity determination of saponins for pharmaceutical research and applications from *Hoya verticillata* var. *verticillata*.

The research results contribute to elucidating the complete chloroplast genome structure of *H. verticillata* var. *verticillata* and the characteristic chloroplast genome architecture of species within the genus *Hoya*. The findings have preliminarily enabled the construction of phylogenetic schemes based on whole chloroplast genome sequences or chloroplast gene regions, thereby providing a foundation for evaluating phylogenetic affinity and the degree of molecular evolution among species of the genus *Hoya*.

The two newly isolated saponins from *H. verticillata* var. *verticillata* provide supplementary data for the screening of biologically active compounds applicable to pharmacognostic and biomedical research. The complete chloroplast DNA sequence of *H. verticillata* var. *verticillata* has enriched the available data resources for molecular evolutionary analyses within the genus *Hoya*. Potential DNA barcodes identified in this study may be used to support the taxonomic identification of the genus *Hoya* and of *H. verticillata* var. *verticillata*.

The papers published in international and domestic journals of science and technology, together with the complete chloroplast genome structure of *H. verticillata* var. *verticillata* deposited in GenBank, constitute valuable reference materials for both research practice and academic training.

6. Structure of the dissertation

This dissertation consists of 165 pages (including appendices). The structure of the dissertation is as follows: Introduction (4 pages); Chapter 1: Literature Review (37 pages); Chapter 2: Materials and Methods (16 pages); Chapter 3: Results and Discussion (48 pages); Conclusions and

Recommendations (2 pages); Publications Related to the Dissertation (1 page); References (18 pages); and Appendices (39 pages). This dissertation contains 11 tables, 20 figures, 10 appendices, and 175 references.

CHAPTER 1. LITERATURE REVIEW

The dissertation reviewed and synthesized 131 English-language references on three main topics:

1. Characteristics of Chloroplast Genomes in Plants and in the Genus *Hoya*;
2. Phylogenetic analysis based on chloroplast DNA sequences;
3. Chemical constituents and biological activities of selected *Hoya* species.

1.1. Characteristics of Chloroplast Genomes in Plants and in the Genus *Hoya*

The chloroplast genomes of angiosperms in general, and of the genus *Hoya* in particular, typically have a quadripartite structure consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeat (IR) regions. In many *Hoya* species, abnormal expansion of the IR regions results in a marked contraction of the SSC region, making *Hoya* one of the angiosperm genera with the shortest SSC regions recorded. This feature is not only structurally distinctive but also provides important information for studying the evolutionary mechanisms of chloroplast genomes in this group.

In recent years, the number of sequenced *Hoya* chloroplast genomes has increased rapidly, with more than 50 species published and deposited in public databases. These data have contributed to clarifying chloroplast genome characteristics, supporting infrageneric classification, and providing evidence for phylogenetic studies in the family Apocynaceae. However, compared with the rich species diversity of *Hoya*, with more than 500 species described, the number of complete chloroplast genomes

currently available remains limited. This indicates a continuing need to expand the genetic database for comparative and taxonomic studies.

To date, there has been no full report on the complete chloroplast genome of *H. verticillata* var. *verticillata*, a species naturally distributed in Vietnam. Sequencing and analyzing the chloroplast genome of this species would not only supplement data for the genus *Hoya* but also help clarify its molecular evolutionary characteristics, while providing useful genetic information for biodiversity research and conservation.

1.2. Phylogenetic analysis based on chloroplast DNA sequences

The emergence and development of biotechnology based on molecular biology has provided new and more objective approaches for phylogenetic research and analysis. Studies based on chloroplast DNA sequences, including coding genes, spacers, and introns, and even complete chloroplast genomes, have been proven to be effective in phylogenetic research and analysis for many plant groups, including *Hoya*, a plant genus belonging to family Apocynaceae. Specifically, identifying potential DNA regions for use as a barcode is a critical task and plays a vital role in species identification and taxonomy.

However, the chloroplast genome resources published for *Hoya*, a plant genus, are few compared to its rich species diversity, leading to insufficient representative data and reducing the reliability of phylogenetic analyses. Although bioinformatics tools play a vital role in improving the quality of phylogenetic research and analysis, insufficient data remains a critical problem for phylogenetic research and analysis, especially for *Hoya*, a plant genus, and *Hoya* species in Vietnam, such as *H. verticillata var. verticillata*, for which no published phylogenetic analysis based on chloroplast genomes has been reported, and no data are available.

Thus, exploiting the chloroplast genome to construct phylogenetic trees and, at the same time, identifying potential DNA regions for use as a barcode in taxonomy and species identification are critical and urgent tasks for this study.

1.3. Chemical constituents and biological activities of some *Hoya* species

The number of plant species under investigation for their chemical constituents and biological activities is increasing, and natural compounds of the *Hoya* genus have been of particular interest. Many scientific reports have confirmed that species of this genus have various secondary metabolites such as flavonoids, alkaloids, and especially saponins, which have high pharmaceutical potential. However, detailed scientific studies on saponin compounds in *Hoya* species are still meager in terms of the number of species studied and depth of research, leading to gaps in assessing biological value and potential use.

H. verticillata var. verticillata, a species naturally occurring in Vietnam, has not yet been systematically reported with respect to its saponin constituents and related biological activities. Elucidating new saponin compounds from this species would not only expand the chemical profile of this particular genus but would also offer opportunities for discovering compounds with high pharmaceutical potential.

It is interesting to note that no study using *in silico* molecular docking and molecular dynamics simulations was conducted to investigate interactions between saponin compounds of *H. verticillata var. verticillata* and proteins related to apoptosis, an essential process in carcinogenesis.

CHAPTER 2. MATERIALS AND METHODS

2.1. Research materials

Plant materials

Samples of *H. verticillata* var. *verticillata* (voucher specimen HOYP202302TQ1) were collected in Sơn Dương Commune, Tuyên Quang Province, Vietnam, in February 2023 at coordinates 21°41'04" N, 105°29'03" E. Another voucher specimen (HOYP202305TN) was collected in Hang Trai area, Quang Sơn Commune, Thái Nguyên Province, Vietnam, at coordinates 21°43'34" N, 105°50'52" E.

Test bacterial strains

The bacterial strains used for antibacterial assays were *Escherichia coli* (ATCC25922), *Staphylococcus aureus* (ATCC13709), *Pseudomonas aeruginosa* (ATCC15442), and *Bacillus subtilis* (ATCC6633). These strains were supplied by the Microbial Strain Museum, Thai Nguyen Central General Hospital, and the Department of Microbiology, University of Medicine and Pharmacy, Thai Nguyen University. The strains were cultured and maintained at the Faculty of Biology, University of Education, Thai Nguyen University, for experimentation. All strains used in this study are human pathogens; *S. aureus* and *B. subtilis* are Gram-positive, whereas the remaining strains are Gram-negative. All experiments were conducted in the laboratory under full compliance with laboratory safety regulations.

Cancer cell lines

The cancer cell lines used included human lung cancer cells (TA549), human breast cancer cells (MCF-7), and human liver cancer cells (HepG2). These cell lines were provided by Prof. Dr. J. M. Pezzuto, Long Island University, USA, and Prof. Jeanette Maier, University of Milan, Italy, and were preserved at the Biological Testing Laboratory, Institute of Biology, Vietnam Academy of Science and Technology.

Data

Data used in the study included chloroplast genomes and chloroplast DNA region sequences of species published on GenBank, as well as protein information and structures from the Protein Data Bank.

2.2. Chemicals and equipment

Chemicals

DPPH, ascorbic acid, DMSO, SRB, acetic acid, α-glucosidase from yeast, pNPG, 4-nitrophenol (Sigma, USA); DMEM, MEME, L-glutamine, sodium pyruvate, $NaHCO_3$, penicillin/streptomycin, 10% FBS, trypsin-EDTA; normal-phase silica gel (230–400 mesh), RP-18 silica gel (150 μm), Sephadex LH-20, Diaion HP-20 resin (Japan); ellipticine, sulforhodamine B, trichloroacetic acid, Tris base, phosphate buffer, and other chemicals (Sigma, Merck).

Equipment

All equipment used in the experiments was modern, newly available, and highly accurate, and belonged to the laboratories of the Faculty of Biology, University of Education - Thai Nguyen University; the Institute of Biology, Vietnam Academy of Science and Technology; and the International Biological Material Research Center (IBMRC), Korea Research Institute of Bioscience and Biotechnology (KRIBB).

2.3. Research methods

The dissertation employed four major groups of methods:

1. Chloroplast genome research methods;
2. Phylogenetic analysis methods;
3. Methods for biological activity analysis, isolation, and compound identification;
4. Data analysis methods.

2.3.1. Chloroplast genome research methods

These included DNA extraction and genome sequencing, genome assembly and annotation, and genome analysis.

2.3.2. Phylogenetic analysis methods

These comprised phylogenetic analyses based on whole chloroplast genome sequences and phylogenetic analyses based on selected chloroplast DNA regions.

2.3.3. Methods for biological activity analysis, isolation, and compound identification

These included qualitative analysis of organic compounds in *H. verticillata* var. *verticillata*, isolation of saponins from the plant, and *in silico* molecular docking and molecular dynamics simulation.

2.3.4. Data analysis

Bioinformatics software used for genome analyses included MEGA11, BioEdit, and BLAST in NCBI. Statistical data were analyzed using SPSS version 29.0.2.0 for Windows. Differences were compared using the LSD test (Least Significant Difference Test) and Duncan's test at $\alpha = 0.05$.

2.4. Research site and completion of the dissertation

The dissertation was completed at the Faculty of Biology, University of Education, Thai Nguyen University.

CHAPTER 3. RESULTS AND DISCUSSION

3.1. Characteristics of the chloroplast genome of *H. verticillata* var. *verticillata*

3.1.1. Genome structure

The complete chloroplast genome of *Hoya verticillata* var. *verticillata* is 176,429 bp in length (Figure 3.1). It comprises the four typical regions of a circular DNA molecule, including the large single-copy (LSC) region (91,020 bp), the small single-copy (SSC) region (2,297 bp), and two inverted repeat (IR) regions, each 41,556 bp in length, with a GC content of 37.1% (Table 3.1).

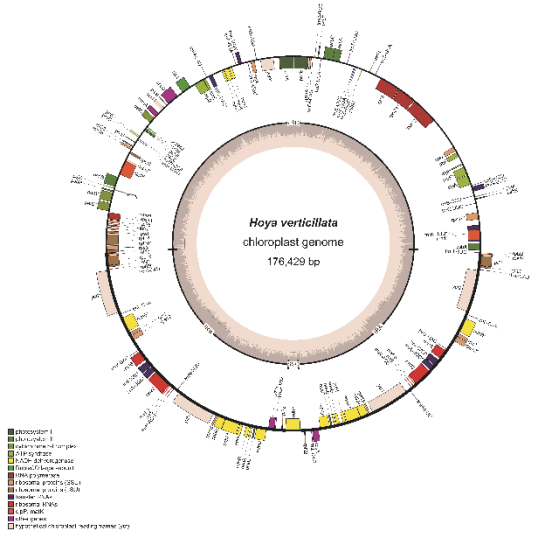


Figure 3.1. Map of the chloroplast genome of Hoya verticillata var. *verticillata collected in Vietnam. Genes located inside the outer circle are transcribed clockwise, whereas those located outside are transcribed*

counterclockwise. Genes belonging to different functional groups are color-coded. The inner concentric circle indicates the quadripartite structure of the chloroplast genome, including the LSC, SSC, and IR regions. The light gray region represents AT content and the dark gray region represents GC content.

The output information on the chloroplast genome from GeSeq and tRNAscan-SE showed that the chloroplast genome of *H. verticillata* var. *verticillata* contains a total of 144 genes, including 98 protein-coding genes, 38 tRNA genes, and 8 rRNA genes (Table 3.1). These genes were classified into three main groups based on their functions. The group of genes related to photosynthesis comprises 44 genes, including gene groups encoding the subunits of ATP synthase, the cytochrome complex, photosystem I and II, NADPH dehydrogenase, and the large subunits of enzymes involved in photosynthesis and light-dependent processes; the remaining 60 genes belong to functional groups related to transcription and translation. The remaining nine genes were classified into other gene groups, including 5 genes with functions related to RNA processing (matK - encoding a protein involved in the splicing of introns from messenger RNAs), 4 conserved open reading frame genes, cytochrome c synthesis (ccsA), fatty acid synthesis (accD), carbon metabolism (cemA), and protein degradation. The annotated chloroplast genome sequence of *H. verticillata* var. *verticillata* was registered in GenBank under accession number OR475244.1.

Table 3.1. Summary of the chloroplast genome of H. verticillata var. *verticillata*

Component	Number
Genome size (bp)	176,429
LSC region size (bp)	91,020
SSC region size (bp)	2,297
IR region size (bp)	41,556
GC content (%)	37.1
Total number of genes	144
Total number of protein-coding genes	98
Total number of tRNA genes	38

Component	Number
Total number of rRNA genes	8

The codon usage frequency of the 98 protein-coding genes found in the coding regions was 29,223, in which codons ending in U and A were found more frequently than those ending in G and C. AUU (encoding isoleucine) was the most frequently found codon among the 64 codons, with an RSCU value of 4.27, followed by GAA (encoding glutamic acid) with an RSCU of 4.24, and AAA (encoding lysine) with an RSCU of 4.19. Meanwhile, the rarest codons were AUC, CUG, and GUG, with an RSCU value of 0.01. In total, 37 codons were used more frequently, with RSCU > 1. Codon AUA was used once.

In *H. verticillata* var. *verticillata*, mono-nucleotide repeat sequences were identified with a frequency of 34 (53.12%), di-nucleotide repeat sequences with a frequency of 14 (21.88%), tri-nucleotide repeat sequences with a frequency of 6 (9.38%), tetra-nucleotide repeat sequences with a frequency of 9 (14.06%), and penta-nucleotide repeat sequences with a frequency of 1 (1.56%). Therefore, the most common mono-nucleotide repeat sequences were A/T, with a frequency of 34 (53.12%) (Figure 3.2).

The frequency of simple repeats in different regions of the genome also showed clear differentiation. The LSC region had a frequency of 49 repeat sequences, including 27 mono-nucleotide, 11 di-nucleotide, 4 tri-nucleotide, 6 tetra-nucleotide, and 1 penta-nucleotide repeat; the IR region had a repeat frequency of 15, including 7 mono-nucleotide, 3 di-nucleotide, 2 tri-nucleotide, and 3 tetra-nucleotide repeats, and no penta-nucleotide repeat sequences (Figure 3.2A); the SSC region contained only one gene and did not contain simple repeat sequences (Figure 3.2B). In the coding sequence (CDS) regions, intergenic spacer (IGS) regions, and non-coding regions (introns), the frequency of simple repeat sequences was mainly present in the spacer regions, with a frequency of 43 (CDS: 11; introns: 10).

3.1.2. Comparison of the complete chloroplast genome of *Hoya verticillata* var. *verticillata* collected in Vietnam with other *Hoya* species on GenBank

The genomes were aligned using the Mauve algorithm, in which colored blocks represent homologous regions among species (Figure 3.4). Results indicate conservation amongst *Hoya* species in terms of their respective chloroplast genomes due to similar arrangements of homologous blocks and all similarities being present; thus the block patterns are presented in a clear collinear fashion. Also represented graphically are the nucleotide positions by number and the colored blocks represent particular features unique to each of the six species studied. An additional observation is the inverted repeat regions (IR) of each genome expand into both large single copy region (LSC) and small single copy region (SSC).

Structural analysis also indicated that these six chloroplast genomes maintained a very high degree of similarity; the genes were arranged continuously along the entire genome and there were almost no large-scale rearrangements such as inversions or translocations.

Nucleotide diversity (Pi) was calculated by sliding-window analysis performed in DnaSP v.6.12.03, using a window length of 600 bp and a step size of 200 bp. The results showed that both single-copy regions were identified as having significantly greater sequence divergence than the inverted repeat region IR. With a Pi threshold value of 0.03, eight highly variable regions were identified, including six intergenic regions (*trnK-rps16, rps16-trnQ, psbI-atpA, psbZ-rps14, ndhC-trnV, rbcL-accD*) and two intron regions belonging to the genes *rpl16* and *ndhF* (Figure 3.5). Six highly variable regions belonged to the large single-copy region (LSC), one region was located between the large single-copy region (LSC) and the inverted repeat region (IR), and the ndhF gene belonged to the small single-copy region (SSC).

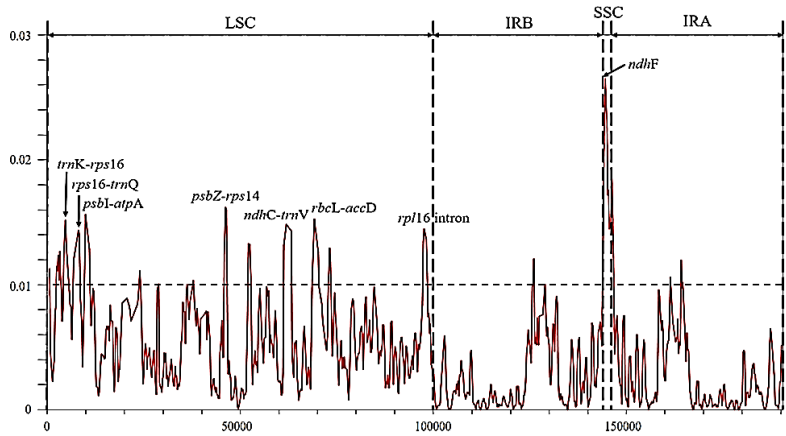


Figure 3.5. Graph comparing nucleotide diversity (Pi) values by sliding-window analysis (window length 600 bp; step size 200 bp) across the whole chloroplast genome of 30 Hoya species. The Y-axis indicates Pi values at the midpoint of the sliding window and the X-axis indicates nucleotide position. The blue dotted line indicates the threshold Pi value of 0.01. Genes exceeding this threshold are labeled above the peaks of the graph.

The LSC/IRB boundary was always located below the *rpl22* gene at the 3' end of the *rpl22* gene; the LSC/IRB junction was located at positions 29-31 bp of the *rpl22* gene, and the IRb/SSC junction was located 26-35 bp away from the ndhF gene. The SSC/IRA junction was located 89-97 bp from the ndhF gene, and the IRA/LSC junction was located at the *rps19-trnH* interval. The difference observed in the *rps19-trnH* interval of *H. verticillata* var. *verticillata* was that it contained the *rpl22* gene for about 30 bp and the junction overlapped by 1 bp within the *rpl22* gene. The contraction of the SSC region and expansion of the IR region in the *Hoya* genome led to changes in the relative lengths of the four structural regions and the whole chloroplast genome sequence; however, this variation was very small (Figure 3.6).

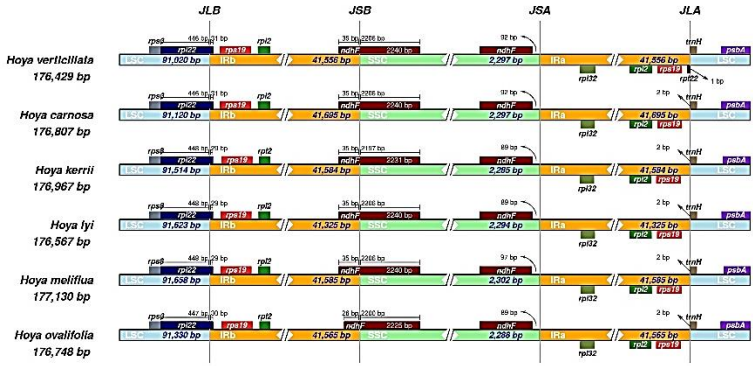


Figure 3.6. Comparison of the junctions of the LSC, SSC, and two IR regions in six complete chloroplast genomes of Hoya. The abbreviations JLB, JSB, JSA, and JLA indicate the junctions LSC/IRB, SSC/IRB, SSC/IRA, and LSC/IRA, respectively.

3.2. PHYLOGENETIC ANALYSIS BASED ON CHLOROPLAST GENOME SEQUENCES AND SOME DNA REGIONS OF THE CHLOROPLAST GENOME

3.2.1. Phylogenetic analysis based on complete chloroplast genomes

Data for phylogenetic analysis based on chloroplast DNA sequences of 50 genomes belonging to the genus *Hoya*, the family Apocynaceae, and the outgroup were searched and downloaded from GenBank (Appendix PL1). The phylogenetic tree established in Figure 3.7 was divided into two branches; the first branch included three species of the subfamily Asclepiadeae, and the second branch included species of the genus *Hoya* and species of other genera in the same tribe Marsdenieae. Notably, species of the genus *Hoya* were distributed into two groups: one large group with 26 species and one small group with 3 species. In the large group, *H. verticillata* var. *verticillata* was distributed in a monophyletic clade, had a very close relationship with other *Hoya* species, and was placed in the same group as *H. carnosa* (bootstrap value 100%).

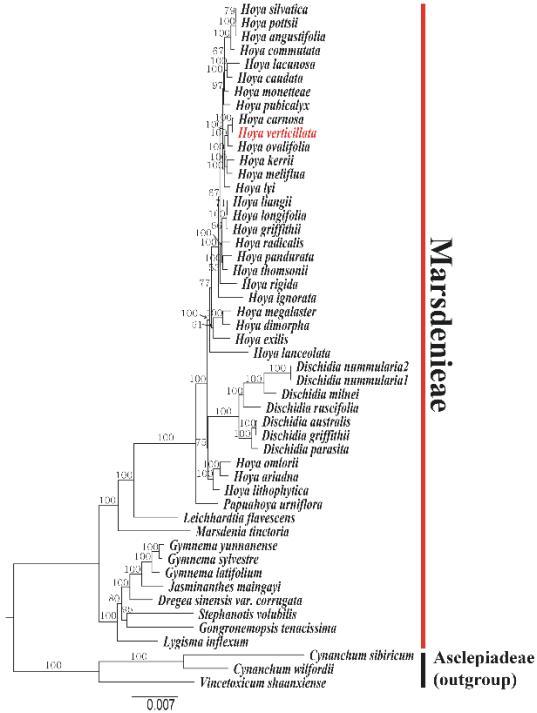


Figure 3.7. Phylogenetic tree of 50 species of the family Apocynaceae based on 79 protein-coding gene sequences. The aligned sequence had a length of 75,252 bp. The Maximum Likelihood (ML) tree was generated by RaxML with -ln L = 196716.416607. Values at the nodes represent bootstrap values. Cynanchum sibiricum, Cynanchum wilfordii, and Vincetoxicum shaanxiense, belonging to Asclepiadeae, were used as the outgroup (not belonging to the tribe Marsdenieae).

To compare, distinguish, and accurately determine the origin of species in the genus *Hoya*, the research group added three other species, namely *Vincetoxicum shaanxiense, Cynanchum wilfordii*, and *Cynanchum sibiricum*; the earliest diverging branch was *Lygisma inflexum*; the second earliest diverging branch included 7 species (*Gymnema yunnanense, Gymnema sylvestre, Gymnema latifolium; Jasminathes maingayi, Dregea sinensis* var. *corrugata, Stephanotis volubilis, Gongronemopsis*

tenacissima), among which *Gymnema yunnanense*, *Gymnema sylvestre*, and *Gymnema latifolium* showed close relationships with bootstrap = 100; the third branch consisted of 3 species: *Marsdenia tinctoria, Leichhardtia flavescens,* and *Papuahoya urniflora*; the fourth branch consisted of 10 species and was divided into two clades, *Dischidia* with 7 species and *Hoya* with 3 species; the fifth and largest branch comprised 26 *Hoya* species. *H. verticillata var. verticillata* shared similar characteristics and was placed in the same branch as *H. carnosa*, had a close relationship with *H. ovalifolia*, with bootstrap value 100. The phylogenetic tree also showed that the three species *H. omlorii*, *H. ariadna*, and *H. lithophytica* were distributed in a paraphyletic group, suggesting ongoing evolutionary divergence in these lineages.

3.2.2. Phylogenetic analysis based on some DNA regions of the chloroplast genome

The results of phylogenetic analysis based on 3 gene regions and 6 spacer regions with high nucleotide diversity in the chloroplast genome of *H. verticillata var. verticillata* identified the gene region *ndhF* and the spacer regions *trnK-rps16, rps16-trnQ, psbI-atpA, ndhC-trnV, and rbcL-accD* as potential DNA barcode candidates that can be used to support the differentiation of species within the genus *Hoya* and the differentiation of the genus *Hoya* from other genera in the family Apocynaceae. Thus, in addition to the genes that have been demonstrated to be potential candidate markers for supporting species differentiation, namely *matK* and *rbcL*, this study additionally identified the gene region *ndhF* (Figure 3.8) and the spacer regions *trnK-rps16, rps16-trnQ, psbI-atpA, ndhC-trnV, and rbcL-accD*. However, further experiments are needed to confirm these sequences as DNA barcodes for practical use.

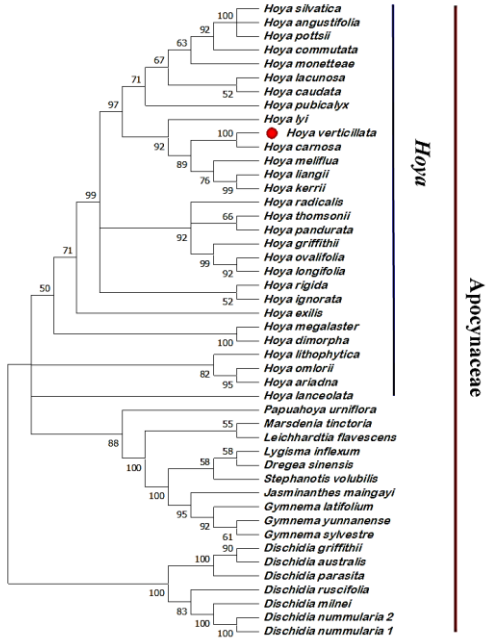


Figure 3.8. Phylogenetic relationship inferred by the Maximum Likelihood method based on the sequence of the ndhF gene region. H. verticillata var. *verticillata and H. carnosa were grouped in the same clade with strong bootstrap support (100%). Species of the genera Marsdenia, Lygisma, Leichhardtia, Dregea, Stephanotis, Jasminanthes, Gymnema, and Dischidia in the family Apocynaceae were used as the outgroup. Numbers at each node are bootstrap support values.*

3.3. ANALYSIS OF THE BIOLOGICAL ACTIVITIES OF SAPONINS ISOLATED FROM *H. VERTICILLATA* VAR. *VERTICILLATA*

3.3.1. Chemical constituents of *H. verticillata* var. *verticillata*

The crude extract was used in the experiment to investigate the chemical constituents of the leaves using chemical reagents. From the results obtained, there were 5 chemical constituents identified: flavonoids, steroids, coumarins, saponins, and tannins.

3.3.2. Antibacterial activity

The crude extract showed a wide range of antibacterial activity against both Gram-negative bacteria (*E. coli, P. aeruginosa*) and Gram-positive bacteria (*S. aureus, B. subtilis*). The extract was more effective at a concentration of 100 mg/mL compared to the antibiotic amoxicillin against *E. coli*. The extract was equally effective compared to the antibiotic amoxicillin against *P. aeruginosa and B. subtilis*. The antibacterial effect of the extract was highly significant against *E. coli* at $P < 0.05$.

3.3.3. Biological activities of saponin-containing fractions from the extract of *H. verticillata* var. *verticillata*

Two fractions (among five fractions: B1, B2, B3, B4, B5) containing saponins from the extract of *H. verticillata* var. *verticillata* were further evaluated for antioxidant activity, cytotoxicity against cancer cell lines, and α-glucosidase inhibitory activity. The results identified fraction B3 as having stronger activity than fraction B2, and these fractions were selected for subsequent analyses.

3.3.4. Isolation of saponins

3.3.4.1. Extraction and isolation of saponins

The dried leaf powder (53.6 g) of *H. verticillata* var. *verticillata* was extracted by ultrasound-assisted extraction using EtOH/H_2O (75:25) under conditions of 200 W, 60°C, for 45 min, yielding 500 mL per extraction (2 extractions). The crude extract was further fractionated by VLC on RP-18 silica gel and eluted successively with 100% H_2O, EtOH/H_2O 50:50, and 100% EtOH, yielding 3 fractions (A-C) based on TLC analysis. Fraction B (5.60 g) was fractionated by MPLC on silica gel 60 ($CHCl_3$/MeOH/H_2O 80:20:2, 75:25:3, v/v/v) to obtain 5 fractions (B1-B5).

Fraction B2 (356.5 mg) was chromatographed by MPLC on silica gel 60 ($CHCl_3$/MeOH/H_2O 80:20:2, v/v/v) to obtain 4 fractions (B2.1-B2.4). Fraction B2.2 (42.3 mg) was further fractionated by MPLC on silica gel 60 ($CHCl_3$/MeOH/H_2O 75:25:3, v/v/v) to obtain 5 fractions (B2.2.1-B2.2.5). Fraction B2.2.3 (15.3 mg) was further fractionated by MPLC on silica gel 60 ($CHCl_3$/MeOH/H_2O 75:25:3, v/v/v) to obtain Compound 1 (3.3 mg).

Fraction B3 (255.6 mg) was further fractionated by MPLC on silica gel 60 ($CHCl_3$/MeOH/H_2O 75:25:3, v/v/v) to obtain 4 fractions (B3.1-B3.4). Fraction B3.3 (85.2 mg) was fractionated by MPLC on silica gel 60 ($CHCl_3$/MeOH/H_2O 75:25:3, v/v/v) to obtain 5 fractions (B3.3.1-B3.3.5). Fraction B3.3.3 (29.8 mg) was separated by MPLC on silica gel 60 ($CHCl_3$/MeOH/H_2O 70:30:5, v/v/v) to obtain Compound 2 (3.1 mg).

3.3.4.2. Acid hydrolysis and gas chromatography

Each glycoside (3 mg) was hydrolyzed with 5 mL of 2N CF_3COOH solution for 3 h at 95°C. The product was extracted with CH_2Cl_2, neutralized with MeOH, and then analyzed by TLC ($CHCl_3$/MeOH/H_2O 8:5:1). The monosaccharide mixture was derivatized with anhydrous pyridine and $HSCH_2CH(NH2)COOCH_3$·HCl at 60°C, followed by HMDS-TMCS (3:1). The final product was extracted with n-hexane and analyzed by GC (1 µL), compared with Sigma-Aldrich standards. Retention times (min): rhamnose 11.07; fucose 13.82; glucose 17.51.

Based on the biological activity of the extract of *H. verticillata* var. *verticillata*, fractionation was conducted to isolate new compounds. From the chromatographic results of fractions B2 and B3, further fractionation and nuclear magnetic resonance spectroscopy yielded two new compounds:

Parasiticoside A (Compound 1): amorphous white powder, [α]25 = -380 (MeOH, c 0.30). HR-ESI-MS: m/z 649.3563 $[M+Na]^+$ (assigned as $C_{33}H_{54}NaO_{11}$, 649.3558). ^{13}C and ^{1}H NMR data are given in Table 3.10.

Parasiticoside B (Compound 2): amorphous white powder, [α]25 = -450 (MeOH, c 0.22). HR-ESI-MS: m/z 795.4142 $[M+Na]^+$ (assigned as $C_{39}H_{64}NaO_{15}$, 795.4137). 1^H and 13^C NMR data are given in Table 3.10.

Table 3.10. ^{13}C and ^{1}H NMR spectral data of Compounds 1 and 2 (solvent: pyridine-d5, δ in ppm, J in Hz)

1			**2**		
Position	**δC**	**δH**	**Position**	**δC**	**δH**
1	37.5	1.06 m, 1.62 m	1	37.4	1.08 m, 1.66 m
2	31.3	1.70 m, 2.14 m	2	32.3	1.75 m, 2.18 m
3	78.1	3.93 ov	3	77.9	3.97 ov
4	39.3	2.28 m, 2.72 m	4	39.2	2.40 m. 2.82 m
5	140.9	-	5	140.9	-

6	125.0	5.67 br s	6	125.0	5.72 br s
7	32.1	1.40 m, 1.95 m	7	32.7	1.43 m, 1.97 m
8	31.8	1.30 ov	8	32.2	1.35 ov
9	50.3	0.80 m	9	50.3	0.87 m
10	36.9	-	10	36.3	-
11	21.0	1.27 m, 1.45 m	11	23.9	1.18 m, 1.48 m
12	39.1	1.00 m, 1.82 m	12	39.3	1.11 m, 1.83 m
13	40.5	-	13	40.6	-
14	56.7	1.08 ov	14	56.7	1.09 ov
15	28.0	1.91 m, 2.40 m	15	28.0	1.92 m, 2.48 m
16	26.4	1.05 m, 1.53 m	16	26.8	1.09 m, 1.54 m
17	62.4	1.60 ov	17	62.3	1.64 ov
18	16.9	0.85 s	18	16.9	0.91 s
19	16.0	1.28 s	19	15.9	1.26 s
20	81.3	3.98 ov	20	81.2	4.01 ov
21	23.5	1.12 d (6.0)	21	23.6	1.28 d (6.0)
Fuc-1	99.8	4.78 d (7.6)	Fuc-1	99.9	4.80 d (7.6)
2	74.2	4.52 dd (9.4, 7.6)	2	74.4	4.53 dd (9.4, 7.6)
3	76.6	4.08 dd (9.4, 2.9)	3	76.4	4.12 dd (9.4, 2.9)
4	72.9	3.93 br d (2.9)	4	73.1	3.96 br d (2.9)
5	70.8	3.62 br q (6.4)	5	70.9	3.68 br q (6.4)
6	17.0	1.46 d (6.4)	6	16.8	1.49 d (6.4)
Glc-1	103.4	4.90 d (7.6)	Glc-1	103.5	4.88 d (7.6)
2	74.7	4.05 ov	2	74.6	4.06 ov
3	78.0	4.26 ov	3	77.9	4.28 ov
4	72.2	4.18 ov	4	72.3	4.18 ov
5	78.1	3.96 m	5	78.1	3.96 m
6	62.4	4.33 m, 4.54 m	6	62.3	4.30 m, 4.52 m
			Rha-1	101.2	6.35 br s
			2	72.1	4.79 br s
			3	72.2	4.61 dd (9.4, 3.5)
			4	73.8	4.35 dd (9.9, 9.4)
			5	69.0	4.83 dq (9.9, 5.9)
			6	18.7	1.78 d (5.9)

3.3.4.3. Structural elucidation of saponins

The process of acid hydrolysis and 2D NMR analysis of Compounds 1 and 2 identified the sugar moieties as follows: D-fucopyranosyl (Fuc), D-glucopyranosyl (Glc) for Compound 1; D-fucopyranosyl (Fuc), D-glucopyranosyl (Glc), and α-rhamnopyranosyl (Rha) for Compound 2. GC

analysis showed the absolute configurations: Fuc and Glc in the D form, Rha in the L form (retention times: Rha 11.07; Fuc 13.82; Glc 17.51 min). The β orientation of Fuc and Glc was demonstrated by the value 3JH-1,H-2 ~7.6 Hz, whereas α-pyranoid Rha was identified with 3JH-1,C-1 ~166 Hz.

The results of chromatographic and nuclear magnetic resonance spectroscopic analyses identified Compound 1 as 3β,20-dihydroxy-pregn-5-ene-3-*O*-β-D-fucopyranoside 20-*O*-β-D-glucopyranoside and it was named parasiticoside A; Compound 2 as 3β,20-dihydroxy-pregn-5 ene-3-*O*-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranoside-20-*O*-β-D-glucopyranoside, and it was named parasiticoside B. Searches in the data library showed that these are two new compounds. The determination of the molecular structures of these two new compounds is the basis for constructing molecular structural models for use in experimental models.

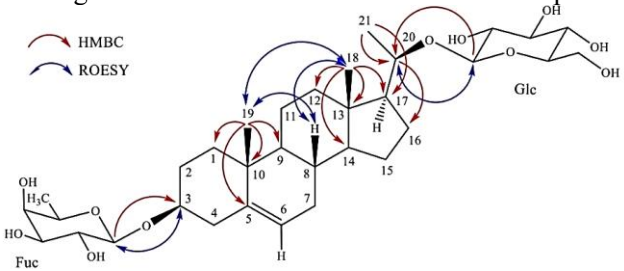


Figure 3.14. Bond connectivity (HMBC) and spatial configuration (ROESY) of Compound 1.

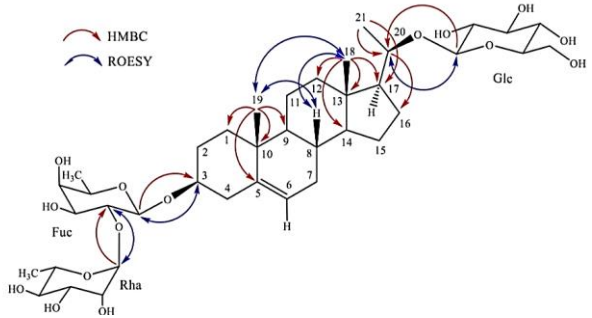


Figure 3.15. Bond connectivity (HMBC) and spatial configuration (ROESY) of Compound 2.

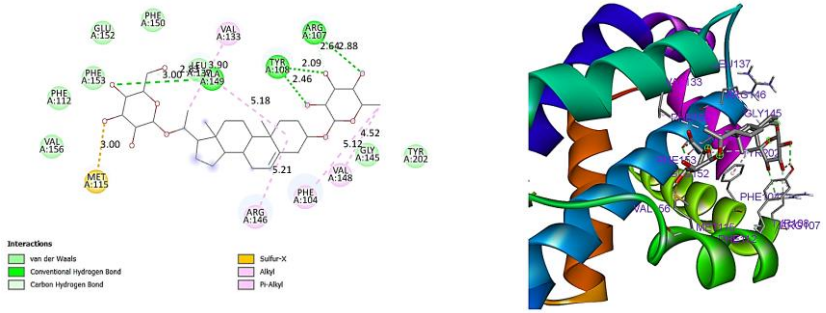
3.3.5. *In silico* molecular interaction and molecular dynamics study of saponins

3.3.5.1. *In silico* molecular interaction study

The *in silico* study was conducted on steroidal saponins isolated from *H. verticillata* as inhibitors of the anti-apoptotic protein Bcl-2, with navitoclax as the control.

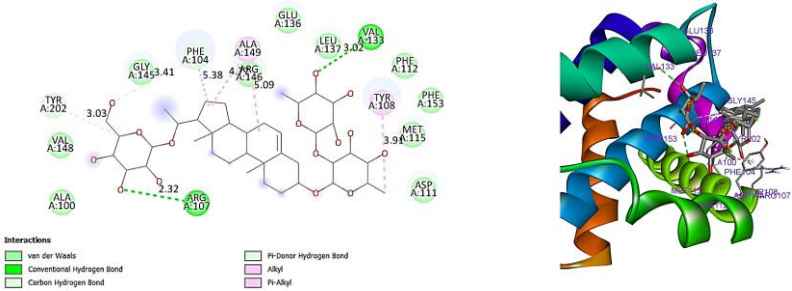
The results showed that Compound 1 (parasiticoside A), Compound 2 (parasiticoside B), Compound 3, Compound 4, Compound 5, and navitoclax all formed Van der Waals interactions with amino acid residues of Bcl-2 (6GL8). Compound 1 interacted with Glu152, Gly145, Phe112, Phe150, Phe153, Tyr202, Val156; Compound 2 with Ala100, Arg146, Asp111, Glu136, Gly145, Leu137, Met115, Phe115, Phe153, Val148. Compound 3 bound to Arg129, Gln118, Glu152, Leu119, Leu137, Phe104, Phe112; Compound 4 with Arg129, Arg146, Asp111, Glu136, Glu152, Phe104, Phe112, Phe153; Compound 5 with Ala149, Arg139, Asp111, Glu152, Met115, Phe104, Phe112, Phe152, Val133. Navitoclax interacted with Ala100, Arg107, Arg146, Asn143, Gln118, Glu136, Gly145, Leu201, Phe153, Tyr108. Each specific molecular interaction with the amino acid residues of 6GL8 is shown in Figure 3.17.

A-1 **A-2**



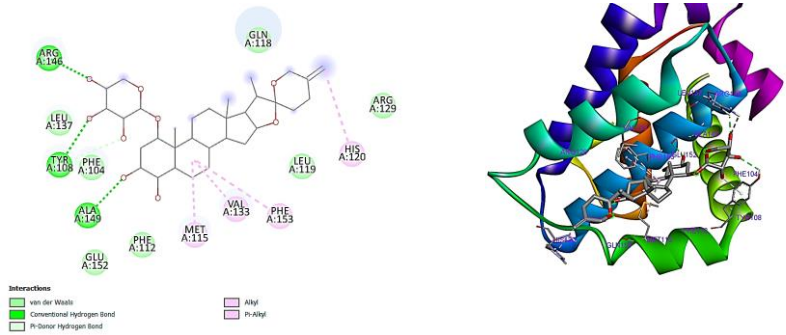
A: 2D and 3D molecular interaction models of Compound **1** with the 6CL8 receptor

B-1 **B-2**



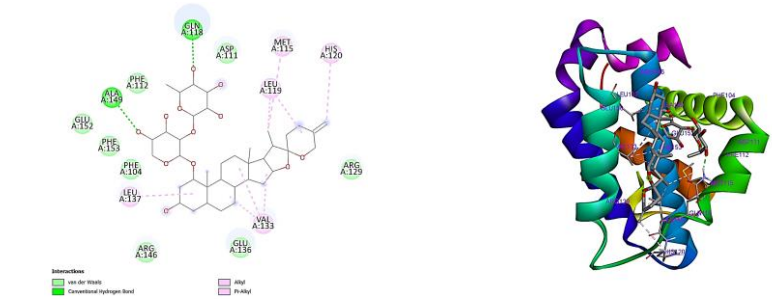
B: 2D and 3D molecular interaction models of Compound **2** with the 6CL8

C-1 **C-2**



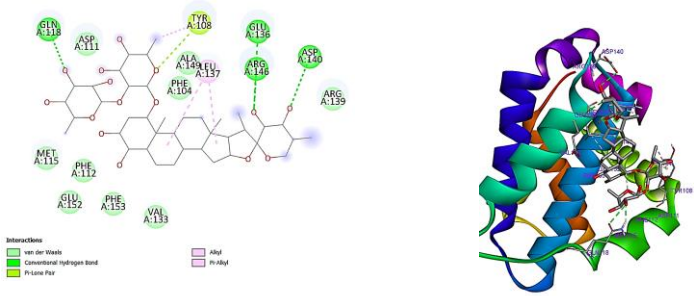
C: 2D and 3D molecular interaction models of Compound **3** with the 6CL8

D-1 **D-2**



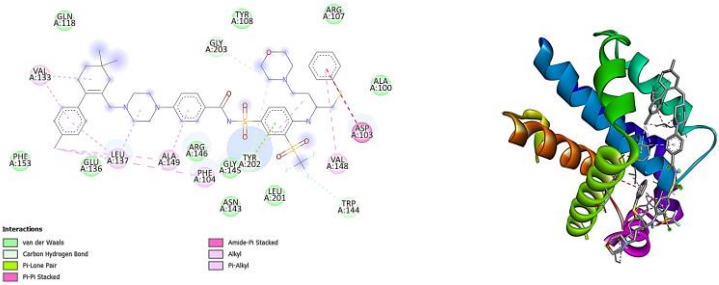
D: 2D and 3D molecular interaction models of Compound **4** with the 6CL8

E-1 **E-2**



E: 2D and 3D molecular interaction models of Compound **5** with the 6CL8

F-1 **F-2**



F: 2D and 3D molecular interaction models of navitoclax with the 6CL8

Figure 3.17. 2D interaction and molecular docking (3D) models of parasiticoside A (A-1, A-2), parasiticoside B (B-1, B-2), 1 (C-1, C-2), 2 (D-1, D-2), 3 (E-1, E-2), and navitoclax (F-1, F-2) with the protein receptor 6GL8.

3.3.5.2. *In silico* molecular dynamics study

In the molecular dynamics study, the main interactions formed between the Bcl-2 protein (6GL8) and the compounds were hydrogen bonds (Figures 3.18 and 3.19).

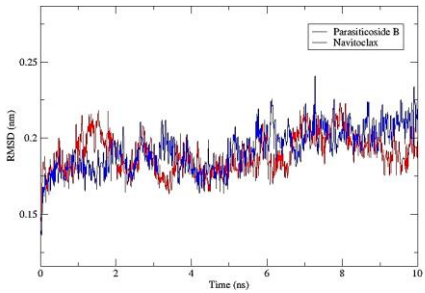


Figure 3.18. Comparative plot of structural stability (RMSD) of the complex of Compound 2 (Parasiticoside B) and Navitoclax during 10 ns of molecular dynamics simulation. The Y-axis shows root mean square deviation (RMSD) in nanometers (nm), ranging from 0 to 0.25 nm; the X-axis shows time in nanoseconds (ns), ranging from 0 to 10 ns.

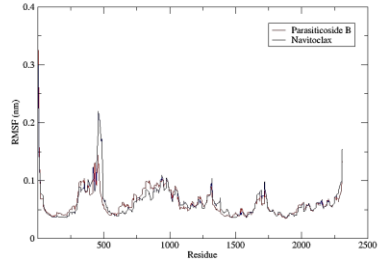


Figure 3.19. Comparative plot of structural fluctuation (RMSF) of the complex of Compound 2 (Parasiticoside B) and Navitoclax during molecular dynamics simulation. The Y-axis shows root mean square fluctuation in nanometers (nm), and the X-axis shows the number of bond units. Parasiticoside B (red line), Navitoclax (blue line).

Compound 2 (Parasiticoside B) formed four hydrogen bonds with Arg107, Phe104, Tyr202, and Val133, resulting in good stability (-8.87 kcal/mol). Compound 1 (Parasiticoside A) also formed four hydrogen bonds with Ala149, Tyr108, Arg107, and Leu137, contributing to its stability, with a binding energy of -8.28 kcal/mol.

Compounds 3 and 4 formed hydrogen bonds with Ala149, showing binding energies of -9.29 and -11.09 kcal/mol, respectively.

Compound 5 formed five hydrogen bonds with Arg146, Asp140, Gln118, Glu136, and Tyr108, together with a weak π-alkyl interaction with Leu137, with a binding energy of -11.16 kcal/mol.

Navitoclax, used as the control, formed three hydrogen bonds with Gly203, Trp144, and Tyr202 and six hydrophobic interactions with Ala149, Asp103, Leu137, Phe104, Val133, and Val148, with a binding energy of -8.34 kcal/mol. The binding energies (values) for all of the above mentioned compounds are as follows Compound 1 (Parasiticoside A): -8.28 kcal/mol, Compound 2 (Parasiticoside B): -8.87 kcals/mol, Compound 3: -9.29 kcal/mol,

Compound 4: -11.09 kcal/mol, Compound 5: -11.16 kcal/mol, and Navitoclax: -8.34 kcal/mol. Especially, the simulation of Compound 2 (Parasiticoside B) and its interaction with protein 6GL8 revealed that Compound 2 had a stable binding affinity throughout the duration of the assay and therefore represents a potential candidate for further experimental investigation of Bcl-2 inhibition in cancer.

CONCLUSIONS AND RECOMMENDATIONS

1. Conclusions

1.1. The complete chloroplast genome of *H. verticillata* var. *verticillata* was sequenced with a size of 176,429 bp and has the typical four-region structure of a circular DNA molecule, including the large single-copy region LSC (91,020 bp), the small single-copy region SSC (2,297 bp), and two inverted repeat regions IR, each IR region being 41,556 bp in size, with a GC content of 37.1%. The chloroplast genome of *H. verticillata* var. *verticillata* contains 144 genes, including 98 protein-coding genes, 38 tRNA genes, and 8 rRNA genes.

1.2. Phylogenetic analysis showed that *H. verticillata* var. *verticillata* has the closest genetic relationship with *H. carnosa* in the genus *Hoya* and has a close genetic relationship with species of other genera within the same tribe Marsdenieae. The gene region ndhF and the spacer regions *trnK-rps16, rps16-trnQ, psbI-atpA, ndhC-trnV*, and *rbcL-accD* were proposed as potential DNA barcode candidates that may be may be used to support the identification of species within the genus *Hoya* and the discrimination of *Hoya* from related genera.

1.3. The saponin-containing fractions separated from the crude extract of the leaves of *H. verticillata* var. *verticillata* exhibited antioxidant activity, cancer cell cytotoxicity, and α-glucosidase inhibitory activity. Two new saponins isolated from fractions B2 and B3 were 3β,20-dihydroxy-pregn-5-ene-3-*O*-β-D-fucopyranoside 20-*O*-β-D-glucopyranoside and 3β,20-dihydroxy-pregn-5 ene-3-*O*-α-L-rhamnopyranosyl-(1→2)-β-D-fucopyranoside-20-*O*-β-D-glucopyranoside. Analysis of the interaction model and *in silico* molecular dynamics of the saponins isolated from *H. verticillata* var. *verticillata* identified Compound 2 as a potential candidate against the Bcl-2 protein in cancer.

2. Recommendations

2.1. Continue molecular evolutionary analysis based on chloroplast DNA sequences of *Hoya* species from GenBank data to strengthen or supplement the system of species in the family Apocynaceae (Asclepiadaceae).

2.2. Continue experimental studies to verify the gene region *ndhF* and the spacer regions *trnK-rps16, rps16-trnQ, psbI-atpA, ndhC-trnV*, and *rbcL-accD* as chloroplast DNA barcodes supporting identification of the genus *Hoya* and differentiation of species within this genus.

2.3. Continue studies to develop an optimal saponin structural model with activity against cancer cell lines based on the results of *in silico* interaction model and molecular dynamics analyses.