MOLECULAR AND *IN SILICO* CHARACTERIZATION OF *TRYPTOPHAN DECARBOXYLASE* (*TDC*) GENE FROM *MITRAGYNA SPECIOSA* (KRATOM)

BY

NURUL IZZAHTUL LIYANA BINTI AZMAN

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

2026

MOLECULAR AND *IN SILICO* CHARACTERIZATION OF *TRYPTOPHAN DECARBOXYLASE* (*TDC*) GENE FROM *MITRAGYNA SPECIOSA* (KRATOM)

BY

NURUL IZZAHTUL LIYANA BINTI AZMAN

A dissertation submitted in fulfillment of the requirement for the degree of Master of Science.

Kulliyah of Science

International Islamic University Malaysia

FEBRUARY 2026

ABSTRACT

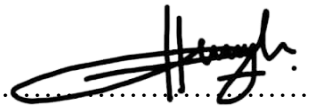
Mitragyna speciosa, commonly known as kratom or "ketum", is a tropical plant belonging to the Rubiaceae (coffee) family, predominantly found in Southeast Asia, particularly in Malaysia, Indonesia and Thailand. It is widely recognized for its medicinal properties, primarily due to mitragynine, a major psychoactive compound with stimulant and analgesic effects. Mitragynine biosynthesis involves a complex metabolic pathway in which the enzyme tryptophan decarboxylase (TDC) plays a critical role by converting tryptophan into tryptamine, a key alkaloid precursor. This study aimed to identify and characterize the *TDC* gene from the *M. speciosa* using molecular techniques and bioinformatics, focusing on its gene structure, three-dimensional modelling and mechanistic interaction, particularly the protein-ligand interaction between the TDC and tryptophan. Functional analysis revealed that the region spanning amino acid positions 57 to 430 represents a conserved TDC protein domain involved in metal-ion interaction and ligand binding within a full-length protein of 506 amino acids. The *TDC* gene was successfully isolated using the conventional Cetyltrimethylammonium bromide (CTAB), followed by Polymerase Chain Reaction (PCR) amplification. The resulting 1,599 bp nucleotide sequence was validated via Needleman-Wunsch pairwise alignment. Phylogenetic analysis was conducted using MEGA-X, and domain prediction with INTERPRO confirmed the presence of the Pyridoxal-Dependent Decarboxylase domain, which binds pyridoxal phosphate (PLP) and is essential for catalyzing the decarboxylation of amino acids, enabling the conversion of L-tryptophan to tryptamine. Physicochemical analysis using PROTPARAM classified TDC as a polar protein with an optimum isoelectric point (pI) of 6.05. The three-dimensional structure modelling identified the AlphaFold model as the most reliable, with ERRAT quality score of 95.7983. Tunnel visualization revealed a ligand-accessible channel measuring 4.7 Å in length with a bottleneck radius of 2.25 Å, suggesting ideal path for substrate transport. Molecular docking analysis using Webina and SeamDock platforms showed that Webina provided a more stable and stronger interaction, with a binding affinity of -9.901 kcal/mol, compared to -5.4 kcal/mol for SeamDock. The protein–ligand complex was further refined using YASARA energy minimization, resulting in a reduced total potential energy of -6.688×10^4 kcal/mol, indicating a stable and optimized complex structure. In conclusion, this study elucidates the functional role of TDC in the biosynthesis of tryptamine-derived alkaloids in *M. speciosa*. The findings provide a molecular basis for future metabolic engineering and synthetic biology studies targeting TDC to enhance the production of specific pharmaceutically important alkaloids, particularly mitragynine.

البحث ملخص

ميتراجينا سبيسيوزا، المعروفة باسم الكراتوم أو "الكيتوم"، نبات استوائي ينتمي إلى فصيلة الفوّفية (القهوة)، وينتشر بشكل رئيسي في جنوب شرق آسيا، خاصةً في ماليزيا وإندونيسيا وتايلاند. يُعرف على نطاق واسع بخصائصه الطبية، ويرجع ذلك أساسًا إلى مركبها النفسي الرئيسي الميتراجينين، الذي يتميز بتأثيرات منشطة ومسكنة. يتضمن تخليق الميتراجينين مسارًا أيضيًا معقدًا، حيث يلعب إنزيم التريبتوفان ديكاربوكسيلاز (TDC) دورًا محوريًا في تحويل التريبتوفان إلى تريبتامين، وهو سلَف أساسي للقلويدات. هدفت هذه الدراسة إلى تحديد وتوصيف جين *TDC* من *M. speciosa* باستخدام التقنيات الجزيئية والمعلوماتية الحيوية، مع التركيز على بنيته الجين، والنمذجة ثلاثية الأبعاد، والتفاعل الآلي، ولا سيما التفاعل بين البروتين والربيطة مع التريبتوفان. كشف التحليل الوظيفي أن المنطقة الممتدة من الموضع 57 إلى الموضع 430 من الأحماض الأمينية تمثل نطاقًا بروتينيًا محفوظًا لإنزيم ديكاربوكسيلاز المعتمد على البيريدوكسال (TDC)، والذي يشارك في تفاعل الأيونات المعدنية وربط الليجاندات ضمن بروتين كامل الطول مكون من 506 أحماض أمينية. تم عزل جين *TDC* بنجاح باستخدام طريقة بروميد سيتيل تريميثيل الأمونيوم (CTAB) التقليدية، متبوعةً بتضخيم تفاعل البوليميراز المتسلسل (PCR). تم التحقق من صحة تسلسل النيوكليوتيدات الناتج، البالغ طوله 1599 زوجًا قاعديًا، عبر محاذاة نيدلمان–وونش الثنائية. أُجري تحليل تطوري باستخدام برنامج MEGA-X، وأكد تنبؤ النطاقات باستخدام برنامج INTERPRO وجود نطاق ديكاربوكسيلاز المعتمد على البيريدوكسال، والذي يرتبط بفوسفات البيريدوكسال (PLP) وهو ضروري لتحفيز نزع الكربوكسيل من الأحماض الأمينية، مما يُمكّن من تحويل التربتوفان إلى تريبتامين. صنّف التحليل الفيزيائي الكيميائي باستخدام برنامج PROTPARAM إنزيم TDC كبروتين قطبي ذي نقطة تعادل كهربائي مثالية (pI) تبلغ 6.05. أظهر نمذجة البنية ثلاثية الأبعاد أن نموذج AlphaFold هو الأكثر موثوقية، حيث بلغت درجة جودة ERRAT 95.7983. وكشف تصوير النفق عن قناة يمكن للرابط الوصول إليها بطول 4.7 أنغستروم ونصف قطر عنق الزجاجة 2.25 أنغستروم، مما يشير إلى مسار مثالي لنقل الركيزة. وأظهر تحليل الإرساء الجزيئي باستخدام منصتي Webina وSeamDock أن Webina وفرت تفاعلًا أكثر استقرارًا وقوة، مع ألفة ارتباط تبلغ –9.901 كيلو كالوري/مول، مقارنةً بـ –5.4 كيلو كالوري/مول لـ SeamDock. وتم تحسين معقد البروتين–الرابط بشكل أكبر باستخدام تقنية YASARA لتقليل الطاقة، مما أدى إلى انخفاض إجمالي طاقة الوضع إلى –6.688 × 10^4 كيلو كالوري/مول، مما يدل على بنية معقدة مستقرة ومُحسَّنة. في الختام، توضح هذه الدراسة الدور الوظيفي لـ TDC في التخليق الحيوي للقلويدات المشتقة من التربتامين في *M. speciosa*. توفر النتائج أساسًا جزيئيًا لدراسات الهندسة الأيضية والبيولوجيا التركيبية المستقبلية التي تستهدف TDC لتعزيز إنتاج قلويدات محددة ذات أهمية صيدلانية، وخاصة الميتراجينين.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science.



…….
Nurul Hidayah Samsulrizal
Supervisor

…….
Zarina Zainuddin
Co-Supervisor

…….
Mohd Hafiz Arzmi
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Science

…….
Tamil Chelvan Meenakshi Sundram
Internal Examiner

…….
Amirul Adli Abd Aziz
External Examiner

This dissertation was submitted to the Department of Plant Science and is accepted as a fulfilment of the requirement for the degree of Master of Science

…….
Che Nurul Aini Che Amri
Head, Department of Plant Science

This dissertation was submitted to the Kulliyyah of Science and is accepted as a fulfillment of the requirement for the degree of Master of Science

…….
Kamaruzzaman Yunus
Dean, Kulliyyah of Science

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Nurul Izzahtul Liyana binti Azman

Signature…… *Izzah* Date…… 23/02/2026

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

MOLECULAR CHARACTERIZATION OF *TRYPTOPHAN DECARBOXYLASE* (TDC) GENE IN *MITRAGYNA SPECIOSA* (KRATOM)

I declare that the copyright holder of this dissertation are jointly owned by the student and IIUM.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below:

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.

3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Nurul Izzahtul Liyana binti Azman

Izzah	23/02/2026
……	……..
Signature	Date

This thesis is dedicated to me and those who support me for laying the foundation of what I turned out to be in life.

ACKNOWLEDGEMENTS

All praise is due to Allah, the Almighty, whose grace and mercy have guided and sustained me throughout the course of my study. Despite the challenges faced, His blessings have made the completion of this thesis possible. Alhamdulillah, for His endless guidance and strength.

I wish to express my deepest gratitude to my supervisor, Asst. Prof. Dr. Nurul Hidayah binti Samsulrizal, for her continuous guidance, patience, and invaluable support. Her insightful comments, constructive suggestions, and detailed supervision have been instrumental in shaping the quality of this work. Despite her many commitments, she always found time to provide advice, encouragement, and moral support, for which I am sincerely grateful.

My sincere appreciation also goes to my co-supervisors, Assoc. Prof. Dr. Zarina binti Zainuddin and Assoc. Prof. Ts. Dr. Mohd Hafiz bin Arzmi, for their guidance, feedback, and cooperation throughout this research.

I would like to extend my heartfelt thanks to Muhammad Amirul Husni bin Samsulrizal, my senior, for his kind assistance, technical guidance, and support during the laboratory experiments. His help and encouragement were invaluable to the successful completion of this research.

Not to forget, my appreciation goes to the technical and administrative staff for their assistance in managing laboratory equipment, materials, and documentation, which facilitated the smooth progress of my research. I am also thankful to several lecturers who have shared their knowledge, insights, and research experiences during our conversations, which inspired and deepened my understanding throughout my academic journey.

Finally, I am deeply grateful to my beloved family for their prayers, patience, and unwavering support. Their love and encouragement have been my greatest motivation to stay strong and complete this study.

Alhamdulillah, all praise be to Allah for granting me the perseverance and ability to complete this thesis successfully.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Å	Angstrom
%	Percentage
bp	base pair
kDA	kilodalton
kJ/mol	kilo Joule per mole
BS	bootstrap value
ϕ	phi angles
ψ	psi angles
kcal/mol	kilocalorie per mol

LIST OF ABBREVIATIONS

M. speciosa	*Mitragyna speciosa*
TDC	Tryptophan Decarboxylase
HPLC-PDA	High-Performance Liquid Chromatography - Photodiode Array
DART-MS	Direct Analysis in Real Time - Mass Spectrometry
qNMR	Quantitative Nuclear Magnetic Resonance
MIA	monoterpenoid indole alkaloids
STR	strictosidine synthase
SGD	strictosidine β-D-glucosidase
TIA	terpenoid indole alkaloid
MEP pathway	2-C-methyl-D-erythritol 4-phosphate pathway
3D	three-dimensional
CASTp	Computed Atlas of Surface Topography of Proteins
PCR	Polymerase Chain Reaction
AGE	Agarose Gel Electrophoresis
CTAB	cetyltrimethylammonium bromide
HCl	hydrochloric acid
EDTA	ethylenediamine tetra-acetic
NaCl	Sodium Chloride
CIA	chloroform: isoamyl-alcohol
TE	Tris-EDTA
CDS	complete coding sequence
Tm	annealing temperature
BLAST	Basic Local Alignment Search Tool
nr	non-redundant
ORF	Open Reading Frame

Pfam	Protein families
GO	Gene Ontology
MSA	Multiple Sequence Alignment
MEGA X	Molecular Evolutionary Genetic Analysis
JTT	Jones-Taylor-Thornton
pI	isoelectric point
GRAVY	grand average of hydropathicity
PSIPERED	Protein Structure Prediction Informatics of Secondary Structure
pLDDT	predicted Local Distance Difference Test
PDB	Protein Data Bank
GMQE	Global Model Quality Estimation
SAVES	Structural Analysis and Verification Server
CAVER	Analysis of Channels, CAVities, and EneRgies
.sdf	Structure Data File
.pdbqt	Protein Data Bank, Partial Charge & Atom Type File
NW	Needleman-Wunsch alignment
PLP	pyridoxal 5'-phosphate
BR	bottleneck radius
gDNA	genomic DNA

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF STUDY

Mitragyna speciosa (*M. speciosa*), commonly known as kratom, is a tropical evergreen tree belonging to the Rubiaceae (coffee) family and is native to Southeast Asia, particularly Malaysia, Indonesia, and Thailand. It has been used for its medicinal and stimulant properties (Brose et al., 2021). Traditionally, the leaves of kratom are harvested, dried, and then brewed into teas or chewed. These preparations are used for their stimulant and analgesic effects, particularly among labourers seeking energy-boosting properties to endure the hot and humid conditions (Veltri & Grundmann, 2019). Phytochemical profiling has identified more than 40 unique alkaloids in kratom. Mitragynine and 7-hydroxymitragynine are the most common, accounting for around 66% and 2% of the alkaloid composition, respectively (Ameline et al., 2023; Adkins et al., 2011). Other alkaloids such as speciogynine, speciociliatine, and paynantheine have been reported with lower psychoactive effects (Ameline et al., 2024). These compounds contribute to kratom's wide range of pharmacological activities, such as analgesic, stimulant, and opioid-like effects (Bowe & Kerr, 2024; Smith et al., 2023b). Plant-derived natural products have gained global interest, and kratom is notable for its complex phytochemistry and diverse therapeutic potential. Given this growing attention, it is crucial to comprehend the status, benefits and impact of kratom at the national level, particularly in Malaysia.

Despite these concerns, there are certain communities still using kratom for self-medication and sociocultural reasons, which emphasizes the need to provide scientific understanding of its bioactive compounds. A deeper understanding of *M. speciosa*'s pharmacological activities requires studying the genes involved in alkaloid biosynthesis,

particularly the *tryptophan decarboxylase* (*TDC*) gene, which encodes a key enzyme in the shikimate pathway responsible for the biosynthesis of numerous alkaloids (Facchini et al., 2000)

Tryptophan decarboxylase (TDC) is a crucial enzyme in the biosynthesis of indole alkaloids in *M. speciosa*, which catalyzes the decarboxylation of the amino acid tryptophan to yield tryptamine, a precursor in the synthesis of mitragynine and other alkaloid compounds. TDC activity is an essential step in metabolic advancement for the production of these pharmacologically active alkaloids (You et al., 2021). Therefore, it is important to isolate and characterize the molecular structure of the *TDC* gene in kratom to understand its regulatory functions in alkaloid biosynthesis and to explore its potential modification for targeted compound production. Insights into the structural and functional features of TDC can facilitate metabolic engineering approaches aimed at enhancing alkaloid yield or producing specific tryptamine-derived compounds in heterologous systems. Such knowledge could also support biotechnological applications in pharmaceutical development, particularly for the biosynthesis of analgesic and psychoactive alkaloids.

Bioinformatics provides an effective *in silico* approach to explore molecular biology and genetics by integrating computational, mathematical, and statistical methods (Can, 2014). It facilitates researchers carrying out sequence analysis, consisting of the identification of homologous genes (similar genes), alignment of multiple sequences, discovery of important sequence patterns, and analysis of gene evolution (Patil & Gupta, 2022). Widely used tools include FASTA and BLAST for sequence alignment, Clustal Omega for multiple sequence alignment, and Autodock for protein structure prediction and molecular docking (Saran et al., 2023). These tools also help build evolutionary trees and determine the interactions of proteins with other molecules. Using these tools, researchers can determine and analyze complex biological data. Therefore, this study aims to characterize the *TDC* gene in *M. speciosa* using computational tools to establish the structure and function of the gene, which would prove useful in future studies of metabolic engineering and synthetic biology for alkaloid biosynthesis.

1.2 PROBLEM STATEMENT

Research on *M. speciosa* has been overshadowed by studies on the pharmacological activity of mitragynine, the major indole alkaloid, rather than on the upstream enzymes responsible for its biosynthesis. Tryptophan decarboxylase (TDC) is one of the key enzymes that catalyze the conversion of tryptophan to tryptamine, a key precursor in mitragynine biosynthesis. Previous studies reported the *TDC* gene in *M. speciosa* for its cloning, expression, and functional role. Furthermore, the *TDC* gene from *M. speciosa* had been successfully cloned and characterised, showing the sequence homology of TDC with other plant species (Charoonratana et al., 2013a). Another study found that manipulating TDC expression can directly influence the production of mitragynine, while precursor feeding, such as tryptamine, enhances alkaloid content (Charoonratana et al., 2013b; Mohamad Zuldin et al., 2013; Wungsintaweekul et al., 2012). Moreover, expression of TDC in *Escherichia coli* yielded a soluble protein, highlighting its potential for biochemical and metabolic engineering applications (Charoonratana et al., 2013a). However, the molecular characterisation, especially in sequence analysis, 3D structural prediction, and substrate-binding interaction, remains unexplored.

Despite these studies, there are still major gaps in the molecular characterization of TDC in *M. speciosa,* namely, (1) limited bioinformatics approaches to extensively explorethe *TDC* gene and its encoded protein, particularly regarding sequence analysis, evolutionary relationships, and functional annotation. (2) Absence of 3D structural models limits understanding of domains, active sites, and stability, and (3) Substrate-binding interactions with tryptophan remain unclear, raising questions about enzyme efficiency and specificity. Addressing these gaps will advance fundamental understanding of the biosynthetic pathway and support applications in synthetic biology and drug development.

To address these gaps, this study employs bioinformatics and computational methods to improve the molecular characterization of TDC from *M. speciosa*, thereby addressing existing limitations. Sequence analysis will be performed to identify the gene and compare it with homologs from similar plant species. Then, 3D protein structure modelling will be conducted to predict the structural characteristics and functional domains of the enzyme. Molecular docking and energy minimization analyses will be used to investigate the interaction between TDC and tryptophan, identifying potential active sites and assessing the stability of enzyme-ligand binding. This integrative approach will yield new insights into the structural and functional properties of *TDC*, establishing foundation for metabolic engineering of alkaloid pathways and potential therapeutic applications.

1.3 RESEARCH OBJECTIVES

1.3.1 To isolate the *tryptophan decarboxylase* (*TDC*) gene responsible for mitragynine biosynthesis in *M. speciosa.*

1.3.2 To perform *in silico* characterization and structural analysis of the isolated *TDC* gene, encompassing 3D protein structure construction.

1.3.3 To determine and validate the stability of binding interactions between the TDC protein and tryptophan through molecular docking analysis.

1.4 RESEARCH HYPOTHESIS

1.4.1 The *TDC* gene from *M. speciosa* that is essential for the biosynthesis of mitragynine can be successfully isolated.

1.4.2 *In silico* characterization and structural analysis of the isolated *TDC* gene will be successfully predicted and validated using computational modelling.

1.4.3 The TDC protein exhibits stable and favourable binding interactions with tryptophan, which can be confirmed through molecular docking and energy minimization analysis.

CHAPTER TWO

LITERATURE REVIEW

2.1 *MITRAGYNA SPECIOSA* (KRATOM)

Mitragyna speciosa, commonly known as kratom, is native to Southeast Asia, particularly in countries such as Thailand, Malaysia, and Indonesia, where it typically grows in marshy forests and along riverbanks (Grundmann et al., 2023; Veltri & Grundmann, 2019). This species belongs to the Rubiaceae (coffee) family and has been classified taxonomically based on its morphological and phytochemical traits (Brown et al., 2017). The taxonomic classification of *M. speciosa* is shown in Table 2.1.

The leaves of kratom have been traditionally consumed in various forms, such as teas, capsules, and powders, to relieve pain and combat fatigue (Alghalith et al., 2024; Prozialeck, 2016). In hot and humid climates, such as those found in Southeast Asia, local labourers have historically used kratom leaves to boost energy and physical endurance. In folk medicine, kratom leaves are commonly chewed or brewed into decoctions prepared using either hot or cold water, producing effects ranging from mild stimulation to relaxation (Grundmann et al., 2023). Although less common, the leaves may also be smoked to achieve a stimulating effect (Hassan et al., 2013).

Table 2.1 Taxonomic classification of *Mitragyna speciosa*

Rank	**Classification**
Kingdom	Plantae
Division	Magnoliophyta (flowering plants)
Class	Magnoliopsida (Dicotyledons)
Order	Gentianales
Family	Rubiaceae (coffee family)
Genus	*Mitragyna*
Species	*Mitragyna speciosa* (Korth.) Havil

Morphologically, *M. speciosa* is a medium- to large-sized evergreen tree that can grow between 4–16 m tall with a broad canopy adapted to humid tropical environments in Malaysia and Thailand. Its leaves are oval to elliptical, measuring 7–15 cm in length and 4–10 cm in width, with distinct vein colouration (green, red, or white) which reflects differences in maturity and alkaloid profiles (Leksungnoen et al., 2022). Leaf structure, including thickness and venation, has been linked to photosynthetic efficiency and variations in mitragynine content. At the same time, environmental factors such as water availability and light intensity significantly influence growth and alkaloid production (Leksungnoen et al., 2025). The tree produces spherical clusters of small white to pale-yellow flowers with five lobes, which represent the largest flowering heads of the genus and support pollination by insects, particularly bees (Pootakham et al., 2022; Fluyau & Revadigar, 2017). These morphological traits, as illustrated in Figure 2.1, demonstrate the ecological adaptability of *M. speciosa* and its importance, both as a culturally significant and pharmacologically valuable plant.

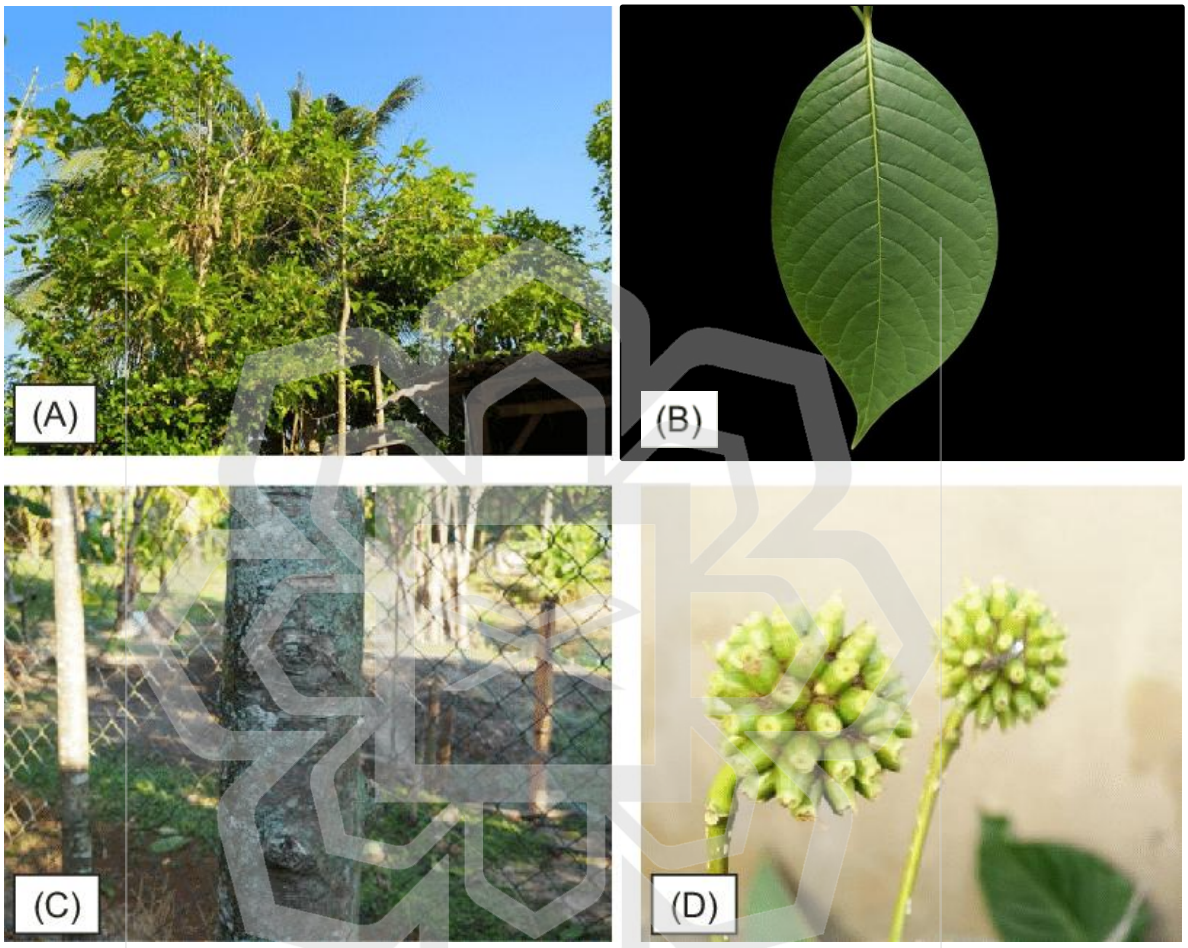


Figure 2.1 Morphological features of *Mitragyna speciosa*. (A) Mature kratom tree; (B) leaf exhibiting green venation; (C) straight, greyish trunk; and (D) flower clusters retrieved from Ghazalli et al., 2021.

2.1.1 *M. speciosa* Usage in Malaysia

M. speciosa has gained global attention for its diverse preparations and the wide range of reported pharmacological and adverse effects (Veltri & Grundmann, 2019). Although several studies have highlighted its potential role in managing pain and opioid dependence, concerns remain regarding its safety and potential for addiction (Khalid et al., 2021). In Malaysia, where kratom has a long history of traditional use, its legal status was reclassified in 2003 as a restricted substance under the Poisons Act 1952 (Halim et al., 2021). This regulation has prompted increased local research into kratom's pharmacological effects, risks, and therapeutic potential to inform evidence-based policies (Hassan et al., 2023; Halim et al., 2021; Khalid et al., 2021). Malaysia also remains one of the key producers of kratom, underscoring its cultural heritage and economic relevance (Hassan et al., 2023).

There are dual perceptions on kratom in Malaysia; on the positive side, it is important for its cultural medicinal use in treating pain and opioid withdrawal, while on the negative side, its misuse has given rise to legal problems and health issues (Fluyau & Revadigar, 2017). The Malaysian government has regulated kratom consumption under the Poisons Act 1952 due to research showing that uncontrolled and excessive use of the drug might lead to addiction, cognitive impairments, and adverse physiological effects (Phoong et al., 2025).

Studies in Malaysia have provided valuable insights into kratom's health impacts. For example, Mohammad Farris and Darshan (2021) reported that chronic and frequent kratom use could lead to elevated blood pressure and cardiovascular fluctuations associated with circulating mitragynine. Similarly, Hassan et al. (2023) demonstrated that mitragynine may induce kidney and liver damage and cause neural impairments such as memory loss, although no significant effects were observed on pulmonary or cardiac functions. Overall, these studies highlight the complex pharmacological profile of kratom in Malaysia, reflecting both its therapeutic promise and associated health risks.

2.1.2 Phytochemistry of *M. speciosa*

M. speciosa exhibits a high phytochemical diversity, predominantly characterized by the presence of alkaloids as its main secondary metabolites. Most of these compounds belong to the indole alkaloid class, including mitragynine, 7-hydroxymitragynine, paynantheine, and speciogynine, which are unique to this species and contribute to its pharmacological properties (Figure 2.2) (Cinosi et al., 2015). Mitragynine is the predominant alkaloid, exhibiting stimulant effects at low doses and sedative effects at elevated doses, whereas 7-hydroxymitragynine acts as a potent agonist at opioid receptors (Kruegel et al., 2016). The alkaloid profile is affected by environmental conditions such as light intensity, temperature, soil nutrient availability, water stress, and extraction methods, with methanolic extracts typically producing higher concentrations and enhanced antioxidant activities (Zailan et al., 2022).

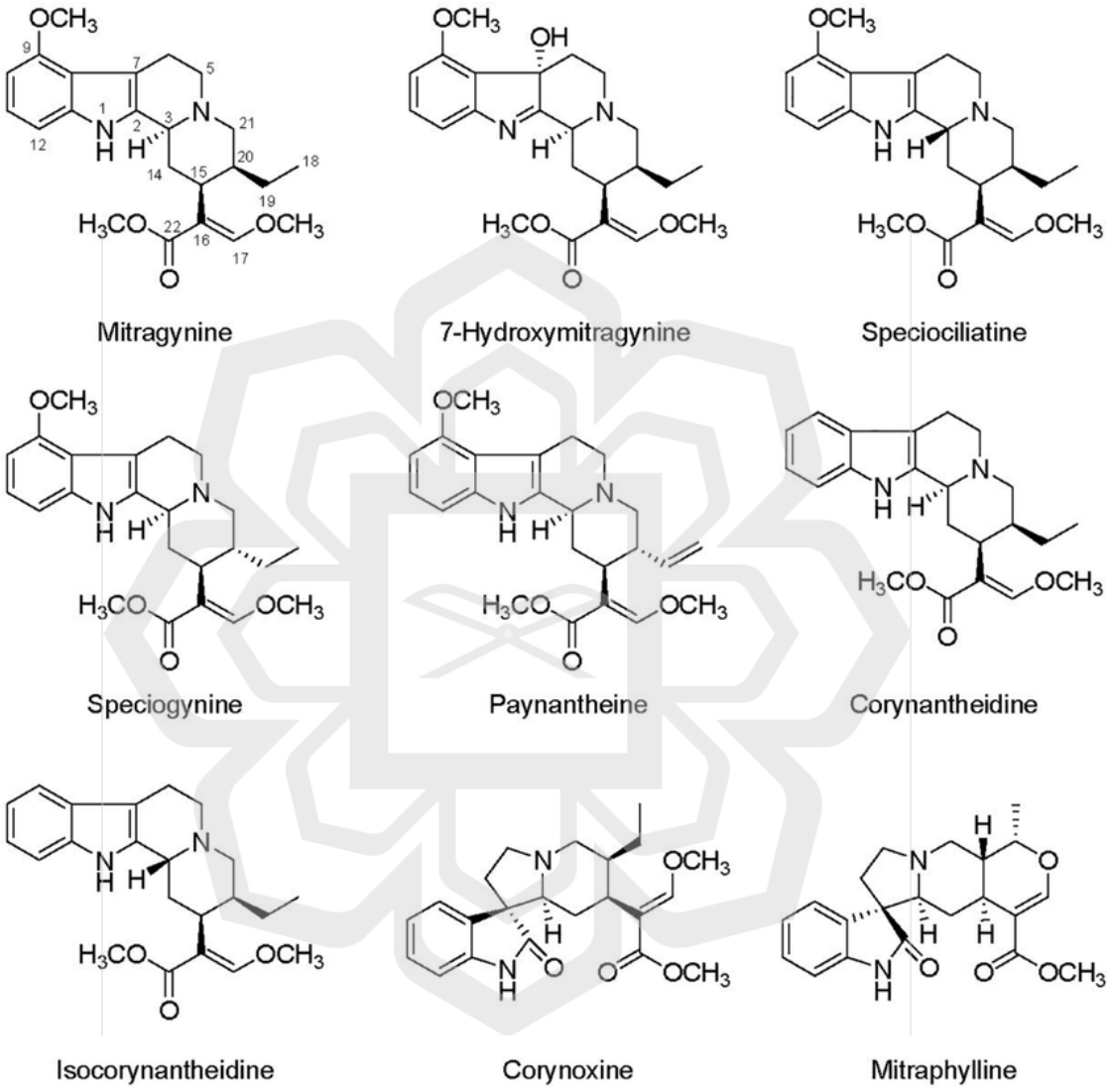


Figure 2.2 Chemical structures of major indole alkaloids identified in *M. speciosa* retrieved from Zhang et al. (2020).

2.1.3 Importance of *M. speciosa*

M. speciosa has been utilized for a long time since it offers both therapeutic and hallucinogenic effects. Begum et al. (2025) discussed several traditional applications of kratom leaves, including the alleviation of pain, fever, cough, anxiety, depression, obesity, diarrhea, wound healing, diabetes, hypertension, cancer prevention, and enhancement of sexual function. The leaves also consist of more than forty alkaloids, with mitragynine and 7-hydroxymitragynine being the most important. Singh et al. (2020) discovered that long-term kratom users among males had improvements in reported sexual performance, including improved stamina, delayed ejaculation, and overall pleasure, supporting its traditional use as a traditional aphrodisiac.

Researchers have extensively examined the molecular and pharmacological properties of kratom alkaloids. Kitajima and Takayama (2014) characterized mitragynine and 7-hydroxymitragynine as novel opioid agonists with significant analgesic properties and unique structures compared to morphine, highlighting their therapeutic promise in pain treatment. Begum et al. (2025) elaborated on the dose-dependent effects of kratom, indicating that lower dosages function as stimulants while larger doses serve as sedatives, and emphasized its role in alleviating opiate withdrawal symptoms. Advanced analytical techniques such as HPLC-PDA, DART-MS, and qNMR have been employed to ensure reliable quantification of these compounds (Garba et al., 2024; Norliana-Izzati et al., 2024; Lesiak et al., 2014). These techniques facilitate the measurement of mitragynine and other alkaloids in various plant parts and products, proving beneficial for pharmacological research and forensic applications.

In summary, *M. speciosa* is important for both cultural and pharmacological purposes. The substance contains a diverse array of bioactive alkaloids that help with its traditional and medicinal applications. The acquired understanding regarding its pharmacological effects, metabolism, and safety issues forms a critical foundation for

investigating the production of mitragynine and associated chemicals, which will be examined in the next section on the biosynthesis pathway.

2.1.4 Health Risk of Using Kratom

Although kratom offers several health benefits, its use is accompanied by potential health risks. Halpenny (2017) noted that kratom provides analgesic effects with a lower risk of addiction compared to conventional opioids; however, concerns about abuse and dependency persist. Manda et al. (2014) reported that mitragynine and 7-hydroxymitragynine exhibit moderate permeability and significant plasma protein binding, which may result in interactions with medications that are substrates of P-glycoprotein, highlighting the need for cautious therapeutic application. Adverse effects, including organ damage and dependency, have been documented, particularly with concentrated or processed kratom formulations (Saingam et al., 2023; Papadi et al., 2022). In contrast, traditional consumption of fresh leaves appears to be less harmful. Grundmann et al. (2023) emphasized that variability among commercially available products complicates safety assessments and underscored the urgent need for further clinical studies to evaluate their efficacy, safety, and associated risks.

While these bioactivities indicate therapeutic promise, the phytochemical profile of *Mitragyna speciosa* also raises important safety concerns. Mitragynine and related alkaloids have been linked to hepatotoxicity and cardiotoxicity, especially at high doses or with prolonged use (Lu et al., 2014; Harizal et al., 2010). Additionally, geographical variations in alkaloid content pose challenges for standardization and reliable evaluation of efficacy and safety (Meireles et al., 2019). The unique alkaloid composition of *M. speciosa* underpins both its therapeutic potential and the controversies surrounding its use, making phytochemical characterization a critical factor in understanding the plant's role in traditional medicine and modern pharmacological research.

2.2 INDOLE ALKALOID BIOSYNTHETIC PATHWAY

Section 2.1 emphasizes that *M. speciosa* synthesizes pharmacologically significant indole alkaloids, including mitragynine, which are classified under monoterpenoid indole alkaloids (MIAs). MIAs represent a diverse group of plant secondary metabolites primarily located within the Apocynaceae, Loganiaceae, and Rubiaceae families, including *M. speciosa* (Cázares-Flores et al., 2016; Araújo-Junior et al., 2007). This group of compounds is widely recognized for its medicinal value, particularly due to its anticancer, antimicrobial, and anti-inflammatory properties (Parammal et al., 2023; Qu et al., 2018; Byeon et al., 2014). The biosynthesis proceeds through two primary metabolic pathways: the shikimate pathway, producing tryptophan as the indole precursor, and the plastidial MEP pathway (2-C-methyl-D-erythritol 4-phosphate pathway), which supplies secologanin as the monoterpenoid component (Facchini & De Luca, 2008).

The standard MIA pathway involves the decarboxylation of tryptophan by the enzyme tryptophan decarboxylase (TDC), resulting in the production of tryptamine. Tryptamine acts as the precursor metabolite for indole alkaloid biosynthesis (Byeon et al., 2014; Gerasimenko et al., 2002). Tryptamine subsequently condenses with secologanin, derived from the MEP pathway, to produce strictosidine, a universal precursor for various MIAs (Qu et al., 2018). This process is an important stage in alkaloid diversification, eventually leading to the synthesis of complex alkaloids like mitragynine in *M. speciosa*. Clarifying this pathway is crucial, as it improves our comprehension of plant metabolic evolution and establishes a basis for metabolic engineering strategies focused on the sustainable production of valuable alkaloids. Figure 2.3 presents a summary of the proposed biosynthetic framework for MIAs in *M. speciosa*.

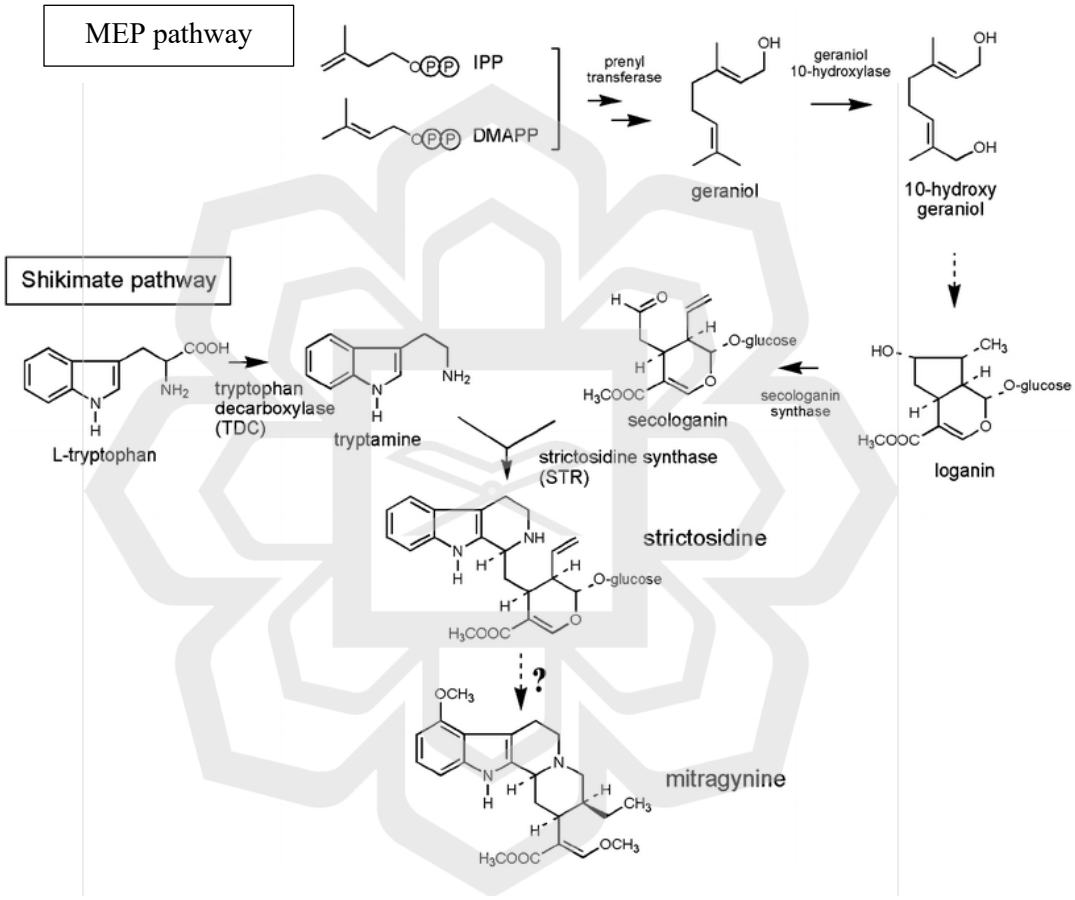


Figure 2.3 Biosynthesis pathway of *M. speciosa* retrieved from Charoonratana et al. (2013a).

2.2.1 Tryptophan-Derived Pathway

The biosynthesis of indole alkaloids in *M. speciosa* begins with the synthesis of tryptophan, an aromatic amino acid produced via the shikimate pathway. The pathway initiates with the condensation of phosphoenolpyruvate and erythrose 4-phosphate, followed by a sequence of enzymatic reactions that culminate in the production of chorismate, which is subsequently converted into anthranilic acid by anthranilate synthase (Dubouzet et al., 2013). An additional enzymatic step at this intermediate stage leads to the formation of tryptophan, serving as the essential precursor for subsequent alkaloid biosynthesis (You et al., 2021; Dubouzet et al., 2013). Figure 2.4 schematically illustrates the enzymatic sequence, highlighting the key enzymes and intermediates involved in tryptophan biosynthesis.

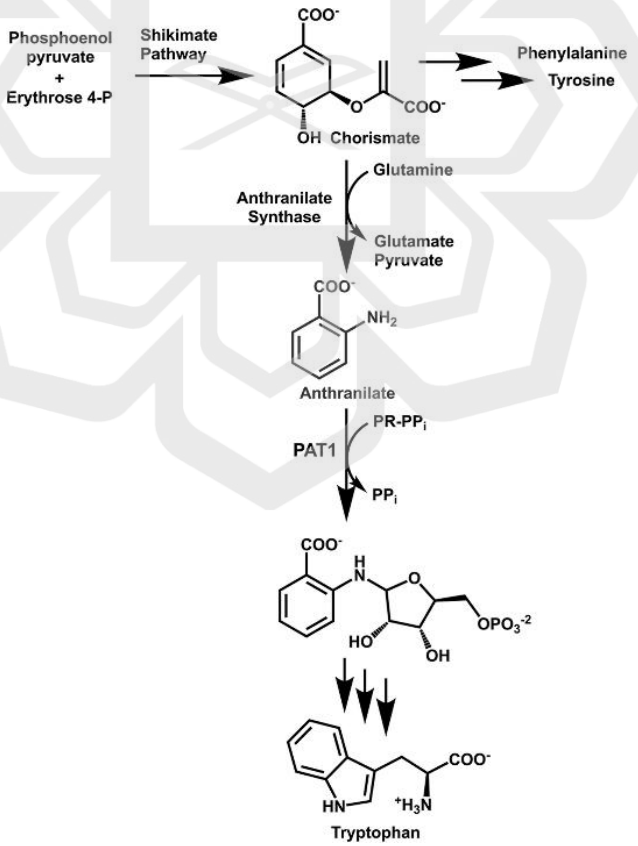


Figure 2.4 Tryptophan biosynthetic enzyme retrieved from Li et al. (2023).

Tryptophan, upon synthesis, is transformed by tryptophan decarboxylase (TDC) into tryptamine, an essential entry metabolite in the MIA pathway. TDC is a cytosolic enzyme that depends on pyridoxal phosphate and is classified within the family of aromatic L-amino acid decarboxylases (You et al., 2021). This enzyme is recognized as a rate-limiting factor that governs the transfer of carbon from primary metabolism to secondary metabolite pathways, thereby regulating the availability of tryptamine for indole alkaloid biosynthesis (Dubouzet et al., 2013). Functional studies in related species, including *Catharanthus roseus*, indicate that the suppression of TDC leads to a reduction in MIA production, while its overexpression increases alkaloid yields (Pan et al., 2012).

Moreover, TDC activity is modulated by developmental and environmental factors, including light conditions, temperature, water and nutrient availability, and biotic stress, underscoring its central role in maintaining the balance between primary amino acid metabolism and secondary metabolite biosynthesis (Pan et al., 2012; Miranda-Ham et al., 2007). In *M. speciosa*, tryptamine produced by TDC acts as the substrate for strictosidine synthase (STR), which catalyzes the condensation of tryptamine with secologanin to yield strictosidine. This metabolite serves as the primary intermediate from which the structural diversity of MIAs, such as mitragynine, originates.

2.2.2 Strictosidine as Central Intermediate

After tryptophan is converted into tryptamine, the indole alkaloid pathway joins the terpenoid pathway through the enzyme strictosidine synthase (STR). STR catalyzes the condensation of tryptamine and secologanin to form (S)-strictosidine, a key intermediate in the biosynthesis of monoterpenoid indole alkaloids (MIAs) (Sheng & Himo, 2020).

Strictosidine is the precursor to over 2,000 to 3,000 diverse alkaloids, that significant in pharmaceutical uses (Dudley et al., 2021; Yang et al., 2010; Barleben et al., 2007).

Strictosidine β-D-glucosidase (SGD) hydrolyzes strictosidine into reactive intermediates,which leads to diverse alkaloid (Arafa et al., 2019; Gerasimenko et al., 2002). Strictosidine is tightly regulated in plants, often stored in vacuoles and released under stress to help plant defence (Guirimand et al., 2011; Guirimand et al., 2010). Environmental stresses can increase STR and SGD gene expression, boosting alkaloids production (Zhu et al., 2015; Burlat et al., 2004).

While well studied in *Catharanthus roseus*, *M. speciosa* uses a similar pathway. In *M. speciosa*, strictosidine leads to unique alkaloids like mitragynine, which important for its effects (Schotte et al., 2023; Brose et al., 2021). Transcriptomic data show STR and related genes are co-expressed in clusters, enhancing pathway regulation (Nützmann et al., 2016). Strictosidine biosynthesis occurs mainly in roots, while mitragynine accumulates in leaves and stems, showing tissue-specific metabolism (Schotte et al., 2023). Overall, Strictosidine links primary pathways to alkaloid metabolism and highlights the role of TDC, which makes tryptamine for strictosidine synthesis. Studying TDC is thus important for understanding alkaloid production in *M. speciosa* (Ahmad et al., 2022).

2.3 TRYPTOPHAN DECARBOXYLASE (TDC)

Tryptophan decarboxylase (TDC) is an enzyme that requires pyridoxal 5′-phosphate (PLP) for its activity and is classified within the aromatic L-amino acid decarboxylase (AADC) family (You et al., 2021; Moreno-Arribas & Lonvaud-Funel, 2001). The enzyme facilitates the decarboxylation of tryptophan, resulting in the production of tryptamine, a crucial precursor in the biosynthesis of indole alkaloids, including serotonin, melatonin, and

pharmacologically important MIAs (Bhowal et al., 2021). TDC exhibits substrate specificity for tryptophan, often serving as a rate-limiting step that regulates the metabolic flow toward tryptamine-derived compounds in various plant systems (Kang et al., 2009).

By all established biochemical and regulatory criteria, TDC represents the rate-limiting step. It catalyzes the first committed and irreversible step in the tryptamine biosynthetic pathway with high substrate specificity for tryptophan, thereby directing metabolic flux away from primary metabolism toward tryptamine-derived compounds (Facchini et al., 2000). Experimental studies have shown a good positive correlation between the expression or enzymatic activity of TDC with the accumulation of downstream tryptamine-derived metabolites (Lehmann et al., 2010). Moreover, TDC is tightly regulated at the transcriptional level in response to developmental and environmental cues, and genetic manipulation of TDC expression significantly alters metabolic flux through the pathway, suggesting that TDC exerts substantial control over pathway throughput and operates as a key rate-limiting enzyme (Facchini, 2001; De Luca et al., 1988).

The enzymatic properties of TDC differ across plant species; however, its conserved function as the initial step in tryptamine biosynthesis highlights its critical role in secondary metabolism (Kang et al., 2008). TDC serves as a regulatory checkpoint that significantly impacts the accumulation and diversity of alkaloids in plants. TDC is crucial in *M. speciosa*, as tryptamine serves as a precursor for mitragynine and various other alkaloids. Enhancing TDC activity has been proposed as a strategy to improve alkaloid yields, presenting opportunities for metabolic engineering and biotechnological applications (Salim et al., 2021). Understanding the biochemical characteristics and regulatory mechanisms of TDC is crucial for elucidating its role in plant metabolism and for leveraging its potential in the production of pharmacologically significant compounds.

2.3.1 Role of TDC in Alkaloid Biosynthesis

Tryptophan decarboxylase (TDC) plays a pivotal role in the terpenoid indole alkaloid (TIA) biosynthetic pathway, facilitating the conversion of tryptophan to tryptamine, which is a crucial precursor for strictosidine synthesis (Zhu et al., 2015; Zhu et al., 2014). Strictosidine is produced via the condensation of tryptamine and secologanin, facilitated by strictosidine synthase (STR), and functions as a crucial intermediate in TIA biosynthesis (Bahieldin et al., 2018). This underscores the critical role of TDC in initiating the indole branch of the pathway, thereby enabling the synthesis of various TIAs with notable pharmaceutical significance (Liu et al., 2016).

TDC enzymatic activity in *C. roseus*, a model plant for TIA research, is strongly influenced by both metabolic and environmental variables such as biotic and abiotic stresses, hormonal signaling, and light conditions. TDC activity is boosted by increased substrate availability, especially higher tryptophan levels, which in turn increases flow along the TIA pathway (Fouad et al., 2021; Liu et al., 2014). This regulation has a direct effect on the accumulation of alkaloids, which means that TDC expression is a critical factor in the alkaloid yield. Additionally, transcriptional analyses indicate that TDC is significantly affected by signaling molecules, particularly jasmonates, which are recognized for their role in upregulating TIA pathway genes in response to stress or developmental signals (Hedhili et al., 2007). Regulation of TDC is interconnected with other metabolic pathways, especially the shikimate pathway, which produces tryptophan as the starting material for TDC.

In addition to its established function in alkaloid biosynthesis, TDC is also involved in other metabolic pathways that come from indole. In certain plants, TDC activity plays a role in the biosynthesis of serotonin and melatonin, highlighting its multifunctional contribution to secondary metabolism (Bhowal et al., 2021; Kang et al., 2008). These

parallel functions expand the biosynthetic capabilities of plants and enhance adaptive responses to biotic and abiotic stresses (Zhu et al., 2015; Liu et al., 2014).

In summary, TDC's biochemical importance and regulatory complexity have prompted extensive molecular studies aimed at elucidating its structural and functional characteristics. This research offers a comprehensive understanding of the enzyme's biochemical characteristics and evolutionary connections, which will be examined in the molecular and *in silico* characterization of TDC.

2.3.2 Molecular Characterization of the *TDC* Gene

Tryptophan decarboxylase (TDC) is essential in the biosynthesis of TIA, as it catalyzes the decarboxylation of tryptophan to form tryptamine. Molecular characteristic analyses of the *TDC* gene elucidate its biochemical and functional roles, especially in the regulation of secondary metabolite production.

TDC genes exhibit significant conservation among alkaloid-producing plants, encoding proteins with specific motifs crucial for catalytic function. The pyridoxal phosphate (PLP)-binding domain is a crucial element that supports catalytic efficiency and functional stability across various species (Rahmdel et al., 2024; Li et al., 2015). The conserved domains facilitate the biosynthesis of various secondary metabolites originating from tryptophan, highlighting the enzyme's adaptive function in plant defence and stress responses. Structural analyses (e.g., X-ray crystallography, NMR) have revealed the enzyme's conformation, which enables specific substrate binding and conversion, highlighting its significance in alkaloid biosynthetic pathways (Abu-Zaitoon, 2014).

Previous studies have identified several regulatory mechanisms that govern TDC expression, alongside insights into its structural organization. In *C. roseus*, TDC transcription is significantly induced by jasmonate and other stress-related signals, establishing its connection to both secondary metabolism and plant defence responses (van der Fits et al., 2001). Functional analyses in heterologous systems have shown conserved biochemical properties such as substrate specificity, pH dependence, and kinetic parameters, with notably high Km values for tryptophan, indicating that TDC may serve as a potential rate-limiting step in alkaloid biosynthesis (Kang et al., 2008). However, while the regulatory and biochemical features of TDC have been well documented in other plant species such as *Catharanthus roseus*, *Rauvolfia serpentina*, and *Nicotiana tabacum*, they remain largely unexplored in *M. speciosa*. A limited number of studies have documented the relationship between TDC expression and alkaloid accumulation in response to elicitor treatments (Pootakham et al., 2022; Charoonratana et al., 2013b), highlighting the necessity for thorough sequence, structural, and functional analyses to elucidate its role in mitragynine biosynthesis. In this context, computational approaches offer a complementary strategy, facilitating the prediction of sequence features, structural conformations, and enzyme–substrate interactions, thus providing valuable insights in areas with limited experimental data.

2.4 BIOINFORMATICS TOOLS FOR MOLECULAR CHARACTERIZATION

2.4.1 Sequence and Phylogenetic Analysis

Based on the molecular characterisation of the *TDC* gene outlined in the preceding section, computational methods enhance our understanding of the sequence characteristics, evolutionary relationships, and its functional properties. Multiple sequence alignment

(MSA), phylogenetic analysis, and conserved motif discovery are crucial for investigating the functional role of TDC in various plant species and its involvement in alkaloid production.

Analysis of conserved motifs facilitates the identification of sequence patterns maintained due to selection pressures on critical metabolic genes. These motifs are especially significant in *M. speciosa*, where the routes for making alkaloids are very strictly controlled. Sharma et al. (2019) highlighted the importance of conserved gene sequences in functional regulation, whereas Pootakham et al. (2022) demonstrated that conserved genomic patterns in *M. speciosa* may directly facilitate alkaloid synthesis. This subsection emphasizes the significance of motif recognition in defining functional regions within biosynthetic enzymes like TDC.

In addition, MSA is essential for matching homologous sequences so that you can see areas of similarity and conservation. The MAFFT software is widely used for this purpose due to its ability to efficiently handle large and complex datasets with high accuracy (Katoh & Standley, 2013). Chloroplast genome alignments for *M. speciosa* indicate high sequence similarity to those of other Rubiaceae species, which supports their tight evolutionary connection, and this plant is also placed within this family species in phylogenetic analysis, indicating its distinct evolutionary path and unique genetic characteristics (Chen et al., 2022). These findings highlight the utility of alignment tools in establishing the evolutionary position of TDC among related plant species, while phylogenetic tree construction elucidates the phylogenetic placement of *M. speciosa* and its biosynthetic genes within a broader evolutionary context. Tools like IQ-TREE that use maximum-likelihood methods make it possible to create strong phylogenetic trees (Nguyen et al., 2014). Furthermore, Graham et al. (2024) employed DNA barcoding in conjunction with phylogenetic reconstruction to distinguish between *Mitragyna* species, uncovering sequence changes that correspond to variations in alkaloid content. These results provide essential insights into the genetic diversity and functional adaptations of *M. speciosa.*

Sequence research lays the groundwork for characterizing the *TDC* gene in *M. speciosa* by uncovering conserved motifs, evolutionary relationships, and genetic variation among related species. These findings will be very useful in future computer studies. After sequence characterization level, it is essential to employ structural prediction methods further to examine the 3D conformation of the TDC protein, identify functional domains, and investigate potential substrate-binding regions.

2.4.2 Physicochemical and Structural Prediction

Physicochemical and structural prediction are essential for understanding the functional characteristics of enzymes such as TDC in *M. speciosa*, as outlined in the preceding sequence analysis. Protein parameters, including molecular weight, isoelectric point, hydrophobicity, and stability, significantly influence solubility, folding, and molecular interactions (Handayani et al., 2021). The precise prediction of these parameters yields significant insights into enzyme activity and protein dynamics, facilitating applications in metabolic engineering and drug design. Beyond proteins, physicochemical analyses have also been applied to the alkaloids of *M. speciosa*, particularly mitragynine, the major indole alkaloid comprising up to 66% of the total alkaloid content (Kruegel et al., 2016). Distinct physicochemical traits such as solubility, lipophilicity, and blood–brain barrier permeability influence its receptor interactions, bioavailability, and pharmacokinetics, which underpin its analgesic, mood-enhancing, stimulant, and antidepressant-like properties (Smith et al., 2023a; Farah Idayu et al., 2011).

Secondary structure prediction enhances the functional interpretation of proteins by determining motifs like α-helices, β-sheets, and random coils that determine folding and conformation stability (Singh, 2005). Hybrid computational approaches, such as neural

networks and genetic algorithms, have improved the accuracy of such predictions through the integration of evolutionary and structural data (Armano et al., 2005). TDC initiates the biosynthetic pathway leading to indole alkaloids such as mitragynine, which interact with human receptors, inducing conformational changes that modulate signalling pathways and produce analgesic, anxiolytic, and mood-regulating effects (Cinosi et al., 2015; Yusoff et al., 2014). These findings highlight the significance of secondary structure modelling in elucidating the impact of structural modifications on pharmacological reactions. Moreover, tertiary structure prediction integrates secondary motifs into a complete 3D conformation, essential for defining active sites, substrate binding, and catalytic efficiency. Computational techniques such as molecular dynamics simulations and machine learning algorithms enhance model reliability, especially when sequence similarity with known homologs is low (Hoffmann et al., 2014; Kurgan et al., 2008). For TDC, tertiary modelling provides valuable insights into how the enzyme accommodates L-tryptophan in its active site, thereby supporting its role in tryptamine biosynthesis.

In summary, integrating physicochemical properties with secondary and tertiary structure prediction provides a comprehensive understanding of the structural and functional behaviour of TDC in *M. speciosa*. These predictions establish a strong foundation for subsequent structural analyses, particularly active site and tunnel prediction, which are crucial for investigating substrate binding and catalytic activity of the enzyme.

2.4.3 Active Site and Tunnel Prediction

Active site and tunnel prediction is a highly important computational technique in structural biology that enables researchers to examine the mechanisms by which enzymes locate substrates, accelerate reactions, and facilitate molecular movement. Several tools have been developed for this purpose. Tools such as the Computed Atlas of Surface Topography of Proteins (CASTp) and Depth are used for accurate mapping of surface cavities and ligand-

binding pockets (Mayanja et al., 2022; Tan et al., 2013), whereas DrosteP employs evolutionary conservation to identify biologically significant catalytic regions (Cammisa et al., 2013). Analysis of Channels, CAVities, and EneRgies (CAVER) remains a widely utilized platform for the analysis of internal transport pathways. Protein dynamics are employed to accurately identify tunnels and channels (Štourač et al., 2019). Additional tools for tunnel detection include MOLE (Petřek et al., 2007) and MolAxis (Yaffe et al., 2008), which are used on a large scale. Collectively, these methods provide a comprehensive framework for examining protein function, ligand accessibility, and potential applications in drug discovery and enzyme engineering.

In this study, the prediction of active sites and tunnels is essential for understanding the biosynthetic enzymes that participate in indole alkaloid production. InterProScan integrates multiple protein signature databases, including Pfam, SMART, and PROSITE, to generate robust and biologically relevant predictions regarding conserved domains and catalytic residues (Burge et al., 2012). CAVER is particularly effective for tunnel prediction, as it incorporates protein flexibility and provides real-time visualization along with quantitative parameters such as tunnel length, curvature, and bottleneck radius (Štourač et al., 2019). The predictive analyses yield important insights into the structural characteristics of TDC that govern substrate recognition and catalysis, clarifying its role in the conversion of L-tryptophan to tryptamine within the indole alkaloid biosynthetic pathway of *M. speciosa*. The results offer a structural foundation for forthcoming molecular docking studies examining the interactions and binding affinities between ligands and enzymes.

2.5 MOLECULAR DOCKING STUDIES

Molecular docking serves as an effective method for predicting the interactions between ligands and proteins at the molecular level. Docking offers critical insights into catalytic mechanisms that are challenging to obtain through experimental methods by simulating the binding orientation, affinity, and essential interactions between substrates and enzymes.

In this study, SeamDock—a web-based platform integrating multiple docking algorithms—was employed to perform both global and local docking simulations. The hierarchical docking method facilitates the identification of potential interaction sites on proteins, while the advanced 3D visualization allows for detailed examination of ligand–protein complexes (Murail et al., 2021). SeamDock offers a more straightforward, user-centric, and educationally oriented approach while maintaining accurate prediction compared to alternatives such as AutoDock, which may require multiple executions to circumvent local optima, and ArDock, which emphasizes interface ranking and is less suited for general docking studies (Reille et al., 2018; Atilgan & Hu, 2010).

Additionally, Webina, a web-based implementation of AutoDock Vina, was employed for its user-friendly interface and integrated 3D visualization capabilities. Operating entirely within a web browser, Webina automates docking calculations and simplifies parameter selection and result interpretation for users with limited computational expertise (Kochnev et al., 2020). In contrast, AutoDock necessitates command-line execution and several configuration steps (Allen et al., 2024; King et al., 2015; Rudnitskaya et al., 2010). Although Webina was initially designed for educational purposes, it provides reliable docking simulations suitable for research applications, including prediction of binding modes and affinities, structural interpretation, and active-site characterization (Kumar & Kumar, 2024; Di Muzio et al., 2017).

In summary, molecular docking enhances sequence analysis and structural modelling by offering direct insights into substrate–enzyme interactions. The application of computational approaches to tryptophan and TDC in *M. speciosa* illustrates the importance of connecting molecular characterization with functional understanding. Overall, this chapter reviewed *M. speciosa*, its alkaloid biosynthetic pathway, the role of TDC, and the integration of computational methods, including molecular docking.

CHAPTER THREE

METHODOLOGY

The methodology chapter is organized to achieve the following objectives: (1) isolating the *TDC* gene responsible for mitragynine biosynthesis in *M. speciosa*; (2) performing *in silico* characterization and structural analysis of the isolated *TDC* gene, including construction of its 3D protein structure; and (3) determining and validating the stability of binding interactions between the TDC protein and tryptophan through molecular docking analysis. The methodologies are summarized in the flowchart presented in Figure 3.1.

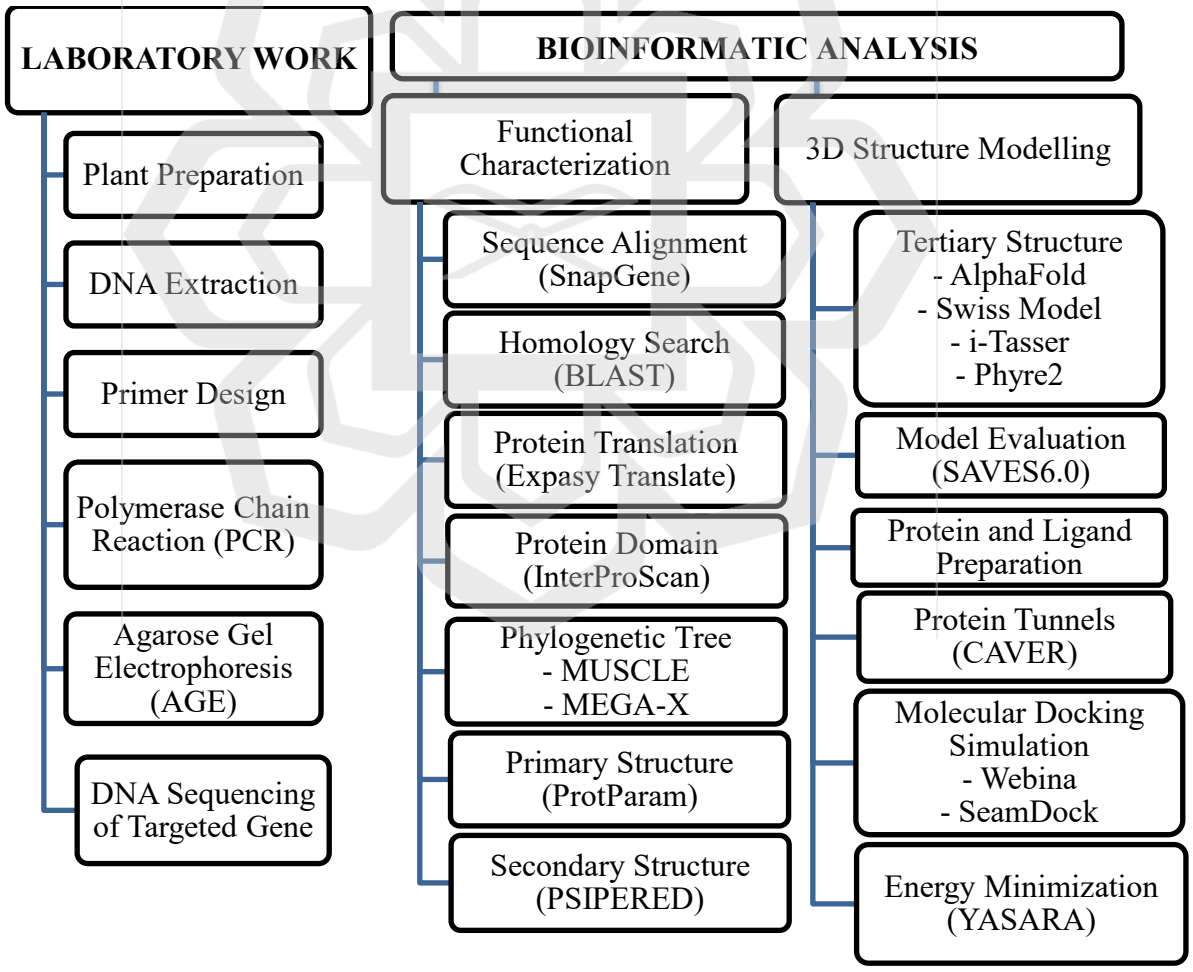


Figure 3.1 Flowchart methodology of this study

3.1 LABORATORY WORK

3.1.1 Preparation of Plant Material

The *M. speciosa* (kratom) plants used in this study were sourced from Kedah, Malaysia. The plants were cultivated under controlled conditions in the Plant Growth Room at the Kulliyyah of Science, International Islamic University Malaysia (Figure 3.2). The plants were maintained to ensure the production of healthy leaves, which were subsequently harvested for DNA extraction.

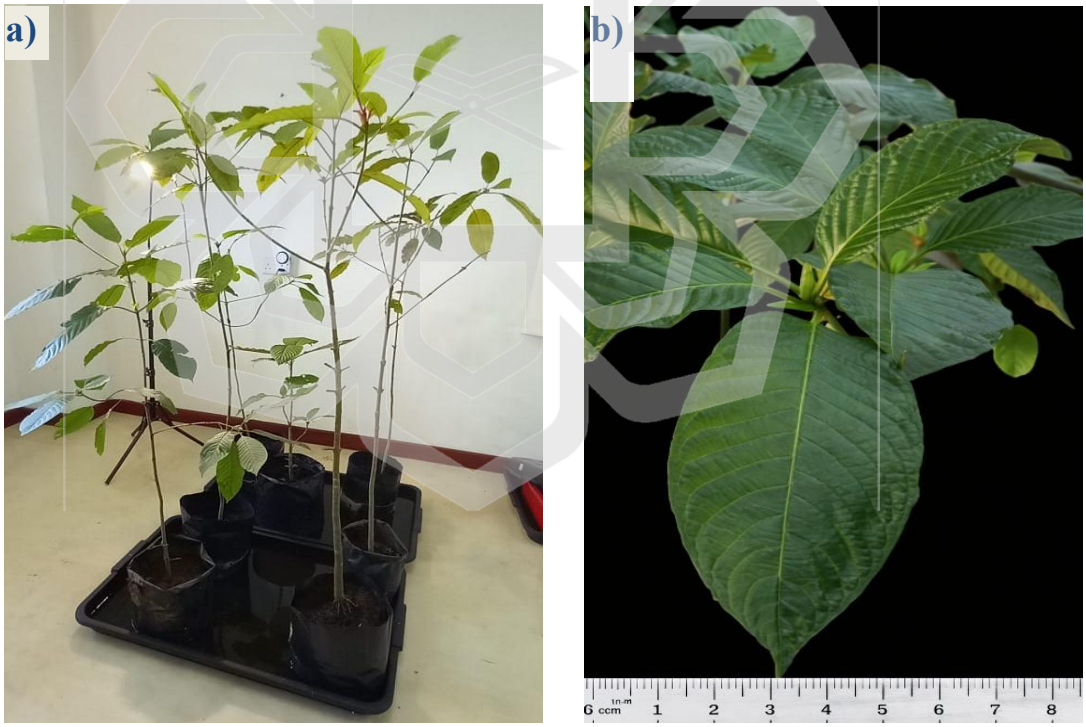


Figure 3.2 The *M. speciosa* plant (a) Growing in the faculty of IIUM (b) Leaves of the plant

3.1.2 DNA Extraction

The genomic DNA (gDNA) of fresh and mature *M. speciosa* leaves were extracted using the cetyltrimethylammonium bromide (CTAB) method, as first described by Doyle & Doyle (1990). This protocol has been widely used due to its efficiency and the ability to yield high DNA concentration.

The leaves were surface cleaned with 70% ethanol (v/v) to remove debris and residues. Then, the leaves were ground in liquid nitrogen using a porcelain pestle and mortar. A 650 µl extraction buffer (1 M Tris-HCl, pH 8.0; 0.5 M EDTA; CTAB; 5 M NaCl; and 0.2% β-mercaptoethanol) was added. The buffer was pre-heated at 65°C for 10 minutes before mixing with the ground tissue, followed by another incubation at the same temperature for 20 minutes. The tube was inverted and vortexed every 5 minutes to ensure the ground tissues were thoroughly mixed with the buffer, a crucial step to enhance the cell disruption and facilitate efficient release of genomic DNA. Next, 600 µl of chloroform: isoamyl-alcohol (CIA) (24:1, v/v) was added, and the mixture was homogenized by inverting and centrifuging at 12000 rpm for 10 minutes. The aqueous layer was transferred into a new tube, and 2/3 of the volume of chilled isopropanol was added to the clear solution. The two solutions were mixed and inverted multiple times until the genomic DNA precipitation formed. The tube was then incubated on ice for 5 minutes before being centrifuged at 12000 rpm for 15 minutes. The supernatant was discarded, and the pellet was mixed with 13 µl of nuclease-free water and 2 µl of RNase. The tube was incubated at 37°C for 5 minutes, followed by washing the pellet with a 70% (v/v) ethanol solution and centrifugation at 10000 rpm for 5 minutes. Following the final discard of the supernatant, the pellet was air-dried at room temperature. The genomic DNA was then mixed with 30 µl of Tris-EDTA (TE) buffer and heated to 65°C for 7 minutes. It was then stored at -20°C before sending for DNA sequencing. The purity of the extracted DNA was assessed using a UV-Vis spectrophotometer by measuring the absorbance ratio at 260/280 nm (A260/A280), where a ratio of ~1.8 is generally considered indicative of pure DNA, with lower ratios suggesting protein contamination (Wilfinger et al., 1997).

3.1.3 Gene-Specific Primers Design for TDC Sequence

Primers for the *TDC* gene were designed using Primer-BLAST based on the nucleotide sequence of the TDC mRNA from *M. speciosa* (complete coding sequence, CDS; accession no: JN643922.1), obtained from the NCBI database. This reference sequence originated from Thailand. The nucleotide sequence was converted from *.txt to *. fasta format for use as the template. The SnapGene software was used to evaluate primer pairs for suitable annealing temperatures (Tm) for PCR amplification.

Initially, three pairs of forward and reverse primers, TDC-01, TDC-02, and TDC-03, were used to amplify the *TDC* genes. Only the 5′ region was successfully amplified using TDC-02 forward and reverse primers (TDC-02_F and TDC-02_R). The middle part of TDC-03 (TDC-03_F and TDC-03_R) and the 3' region part of TDC-01 (TDC-01_F and TDC-01_R) did not amplify the DNA sequence. The amplification failure is due to the estimated length of the *TDC* gene, approximately 1620 base pairs.

To overcome this limitation, the reference *TDC* gene sequence was divided into six overlapping regions and additional primers were designed accordingly (Figure 3.3). A nested-PCR–inspired fragmented amplification strategy was employed to improve specificity and ensure complete coverage of the full-length gene. Six independent PCR reactions were performed, each using a specific primer pair and optimized annealing temperature (Ta) (Table 3.1), to generate individual overlapping DNA fragments. This approach was not conducted as a multiplex reaction; instead, each fragment was amplified in a separate PCR run and later assembled *in silico* to reconstruct the full-length *TDC* gene sequence.

The 5′ region was amplified using primer sets TDC-04 and TDC-06, which overlapped with the fragment generated by TDC-02 to ensure continuous sequence

coverage. The central region was successfully amplified using TDC-03 under optimized conditions, while the primer set TDC-05 was newly designed to improve amplification of the 3′ region previously targeted by TDC-01. Collectively, the six overlapping amplicons spanned the entire *TDC* gene, and the final DNA sequence alignment and assembly results are presented in Chapter 4.2.1.

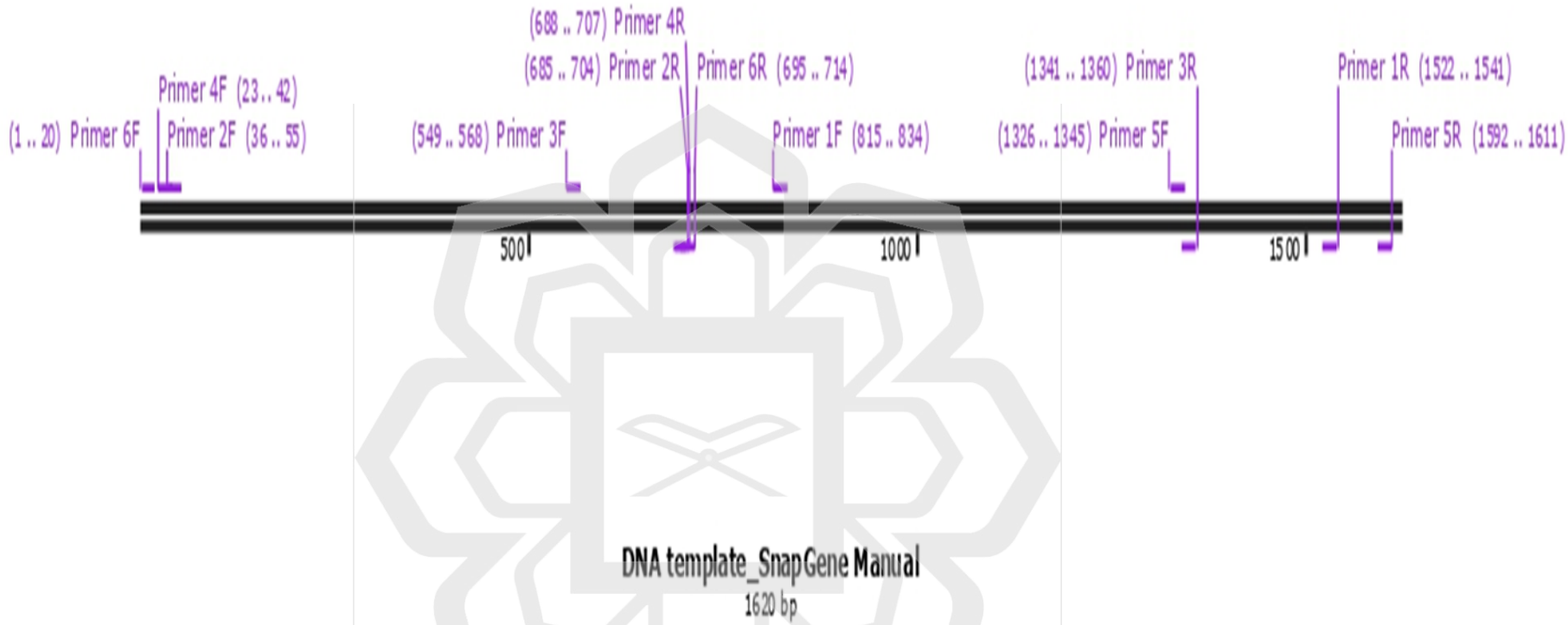


Figure 3.3 Primers mapping and overlapping fragment design used for independent PCR amplification of the 1620 bp *TDC* gene of *M. speciosa* prior to *in-silico* assembly.

3.1.4 Polymerase Chain Reaction (PCR) Technique

Polymerase chain reaction (PCR) is a standard laboratory method used to amplify certain DNA sequences from a sample. The method includes denaturing the DNA by heating and cooling the sample repeatedly, annealing the primers to the target sequence, and incorporating DNA through a heat-stable polymerase enzyme. By repeating this process, the original target sequence becomes increasingly efficient, and this allows it to be identified and analyzed in minute quantities of DNA. Each PCR reaction had a final volume of 25 µl, consisting of 12 µl of Q5 High-Fidelity 2X Master Mix (containing 2.0 mM Mg^{2+} and 200 µM of each dNTP) by New England Bioloabs (NEB), 10 µM of each primer, deionized water, and 100 ng/µl of genomic DNA. The PCR reactions were performed with an initial denaturation at 95 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 seconds, primer-specific annealing for 45 s, and extension at 72 °C for 1 min, with a final extension at 72 °C for 10 min. The annealing temperature (Ta) was optimized individually for each primer pair and ranged from 49–60 °C, as summarized in Table 3.1. Successful amplification was confirmed by the presence of clear bands under UV transillumination following agarose gel electrophoresis.

Table 3.1 Characteristics of gene-specific primers for TDC

Primer Name	**Sequences (5'-3')**	**Length (bp)**	**Annealing Temperature (°C)**	**GC Content (%)**	**Amplicon Size (bp)**
TDC-01_F	TGTGCTACTGTGGGGACAAC	20	59	55	727
TDC-01_R	CAGCGTCAGCGCATTGTTTA	20	58	50	

TDC-02_F	TCTCCAAGAGAA TGGGCAGC	20	59	55	669
TDC-02_R	TGCATGGGAAGA TACCAGCC	20	59	55	
TDC-03_F	CCAGTGAGGCCA TTCTTTGC	20	58	55	812
TDC-03_R	TTCGTTGGATCCT CCCGAAG	20	55	55	
TDC-04_F	CTGAACCAATTA CTCTCCAA	20	51	40	685
TDC-04_R	TATTGCATGGGA AGATACCA	20	52	40	
TDC-05_F	GTTTTAGGCTTA ACCCTTCG	20	54	45	286
TDC-05_R	GTACCCTTTTGA AGGATCAC	20	54	45	
TDC-06_F	CTTCACCCTCATT AAATTTG	20	49	35	714
TDC-06_R	ATTCTTATATTGC ATGGGAA	20	48	30	

3.1.5 Agarose Gel Electrophoresis (AGE) Technique

PCR products were separated on a 1.0% (w/v) agarose gel prepared in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0). ViSafe Green (Vivantis) nucleic acid staining solution was used to stain the gel. The DNA molecular markers used were 100 bp and 1 kb, determined by the size of the anticipated PCR product. All samples were pipetted and mixed with 6X loading dye prior to loading 10 µl per well. Electrophoresis was conducted for 60 minutes at 100 volts, and the gel was visualized under UV light (Gel Documentation System Vilber). Samples with distinct amplification bands were sent for sequencing to Apical Scientific Sdn. Bhd.

3.2 BIOINFORMATICS ANALYSIS

3.2.1 Functional Characterization of TDC in *M. speciosa*

3.2.1.1 Sequence Alignment

SnapGene software was used to align the *TDC* gene sequence obtained from Apical Scientific Sdn. Bhd. with the reference *TDC* gene template sequence to assess accuracy. This step was essential for identifying conserved regions and possible sequence variations. The alignment confirmed that the amplified gene fragment closely matched the template, thereby validating both the sequencing results and the primer design (Chapter 4.2.1). The *TDC* gene sequence obtained in this study (‘Malaysian TDC’) and the reference sequence (‘Thailand TDC’) were further analyzed using the BLAST tool for comparative analysis.

3.2.1.2 Homology Search using BLASTn

The Basic Local Alignment Search Tool (BLAST) was used to align the TDC nucleotide sequence and to identify homologous protein sequences in the non-redundant (nr) database via (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Homologous nucleotides were selected based on species variation and three key parameters: (i) Sequence similarity ranging from 80% to 100%, (ii) E-value ≤ 0.0, and (iii) total score. A higher percentage identity signifies greater similarity between target and query sequences, whereas a lower E-value denotes a more favourable match. All identified sequences were stored in FASTA format for subsequent analysis. The TDC query identified 16 nucleotide sequences from different species, which are presented in Chapter 4.2.5

3.2.1.3 Protein Translation

The TDC nucleotide sequence obtained from alignment was translated into its corresponding protein sequence using Expasy Translate (https://web.expasy.org/translate/). According to the universal genetic code, three nucleotide bases (codons) correspond to a single amino acid. This translation allowed the identification of the Open Reading Frame (ORF), a continuous sequence of codons with the potential to encode a functional protein. The ExPASy Translate tool was used to detect ORFs within the DNA sequence, and the longest ORF was selected, beginning with an ATG start codon and terminating with one of the stop codons (TAA, TGA, or TAG). This analysis also verified the presence of expected TDC protein domains, ensuring the accuracy of gene characterization (Gasteiger et al., 2003).

3.2.1.4 Protein Domains and Families using InterProScan

The InterProScan database (https://www.ebi.ac.uk/interpro/) was used to analyze the domains in the protein sequence. The sequence was submitted in FASTA format for the prediction and annotation of protein domains, families, and functional sites. InterProScan interrogates data from 13 database sources, including Protein families (Pfam), to determine the positions of protein domains within the sequence. Both Pfam and domains that were found through this study, together with Gene Ontology (GO) terms, provide insight into the biological processes and molecular roles of the protein. The results were stored for subsequent investigation of the structural and functional characteristics of the TDC protein (Jones et al., 2014).

3.2.1.5 Phylogenetic Analysis

Phylogenetic analysis was performed to investigate the evolutionary relationship between the target protein sequences, including the TDC protein of *M. speciosa*. The analysis comprised two main steps: (1) multiple sequence alignment (MSA) and (2) phylogenetic tree construction. A total of 16 homologous protein sequences were selected based on consensus quality and sequence similarity percentages. The MUSCLE program (https://www.ebi.ac.uk/Tools/msa/muscle/) was used to align the sequences and identify conserved regions, given its high accuracy (Edgar, 2004). The alignment was stored in *.meg format for subsequent analysis.

Then, phylogenetic tree construction was carried out using the Molecular Evolutionary Genetic Analysis (MEGA X) software version 10.2.0 (https://www.megasoftware.net/). Evolutionary distances were calculated using the

Maximum Likelihood algorithm based on the Jones-Taylor-Thornton (JTT) matrix-based model, a widely accepted empirical model for amino acid substitution (Jones et al., 1992). Bootstrap analysis with 1000 replicates was performed to assess the robustness of the phylogenetic tree. The analysis provided insights into the evolutionary divergence and potential common ancestry of the protein sequences.

3.2.2 3D Structure Model Development

3.2.2.1 Primary Structure and Physicochemical Properties

The FASTA format of the TDC protein sequence was submitted to the ExPASy ProtParam server (https://web.expasy.org/protparam/) to predict its physicochemical properties. The analysis calculated parameters such as molecular weight, theoretical isoelectric point (pI), amino acid composition, extinction coefficient, instability index, aliphatic index, and the grand average of hydropathicity (GRAVY) (Gasteiger et al., 2005).

3.2.2.2 Secondary Structure Prediction

In order to estimate the secondary structure of the TDC protein, the sequence was examined using the PSIPRED (Protein Structure Prediction Informatics of Secondary Structure) server. This tool identifies local folding patterns such as α-helices, β-sheets, and coil regions based on position-specific scoring matrices (Jones, 1999). The PSIPRED method consists of two stages: (1) a neural network estimates the probability of each amino acid residue forming a helix, strand, or coil, and (2) a hidden Markov model refines these predictions to generate the final secondary structure. The result includes a graphical

depiction of the anticipated structure, confidence ratings for each residue, and a full breakdown of secondary structural properties.

3.2.2.3 Tertiary Structure Prediction

The tertiary structure of the TDC protein from *M. speciosa* was predicted using four different computational tools: AlphaFold, SWISS-MODEL, Phyre2, and I-TASSER. Comparing outputs from multiple modelling platforms improved the reliability of the structural analysis. For homology-based modelling, the SWISS-MODEL server automatically searched the PDB for suitable experimental templates. The crystal structure of *Catharanthus roseus* tryptophan decarboxylase in complex with L-tryptophan (PDB ID: 6EEW) was identified as the most suitable template and retrieved from the Protein Data Bank (RCSB PDB) for structural comparison and subsequent superimposition analysis.

First, the 3D structure was predicted using AlphaFold, a deep learning-based protein structure prediction tool developed by DeepMind. The AlphaFold Colab notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb) was used to submit the amino acid sequence of the TDC protein. Following the instructions in the notebook, the prediction was made in Google Colaboratory, and the resulting model was selected based on the predicted Local Distance Difference Test (pLDDT) score, which reflects the confidence level of the prediction.

Next, the protein structure was generated with SWISS-MODEL, an automated homology modelling server that uses templates from the Protein Data Bank (PDB). The SWISS-MODEL workspace (https://swissmodel.expasy.org/interactive) received the TDC amino acid sequence and automatically found appropriate templates. To assess the dependability of the anticipated structures, the server offered models together with quality

scores like GMQE (Global Model Quality Estimation) and QMEAN. The TDC protein structure was also predicted with Phyre2, a widely used web-based protein modelling tool. The server (https://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) received the TDC sequence and applied hidden Markov model profiles to choose the best templates. The extent of sequence coverage and confidence percentage were used to select the final model.

Additionally, the structure prediction server I-TASSER, which is based on threading and iterative assembly simulations, was also used. The server (https://zhanggroup.org/I-TASSER/) received the TDC sequence and returned up to five models with their RMSD values, TM-scores, and C-scores. The model with the highest C-score was selected for further analysis. Finally, all predicted protein structures were downloaded in PDB format and visualised using PyMOL to determine the folding patterns, structural features, and potential active sites.

3.2.2.4 Model Evaluation

The SAVES 6.0 server (Structural Analysis and Verification Server) was used to validate the predicted TDC protein model structure (https://saves.mbi.ucla.edu/). This server is widely used to evaluate the quality of structural models based on energy profiles, residue interactions, geometric, and stereochemical properties. It integrates several programs, including ERRAT, PROCHECK, WHATCHECK, Verify3D, Ramachandran plot analysis, and PROVE (Eisenberg et al., 1997; Colovos & Yeates, 1993; Laskowski et al., 1993).

First, the ERRAT tool was used to analyze the protein structure based on non-bonded atom interactions, providing an overall quality score (Colovos & Yeates, 1993). Next, the PROCHECK tool was used to generate the Ramachandran plot, which assesses

the stereochemical quality of the backbone conformation by identifying residues in allowed and disallowed regions (Laskowski et al., 1993). Finally, the Verify3D program was applied to evaluate the compatibility of the atomic model (3D) with its amino acid sequence (1D profile) (Eisenberg et al., 1997).

3.2.3 Molecular Docking

3.2.3.1 Pocket Tunnel Prediction

The initial step in the molecular docking workflow involved identifying and analyzing potential binding pockets and tunnels within the target protein structure. These potential ligand-binding tunnels were predicted using the CAVER (Analysis of Channels, CAVities, and EneRgies) web tool. CAVER provides a user-friendly interface to identify structural features, predict potential active sites, and guide docking simulations. The predicted tunnels provided insight into the potential ligand entry routes and substrate-binding orientation within the protein (Marques et al., 2025).

In this study, the AlphaFold model was uploaded to CAVER (https://loschmidt.chemi.muni.cz/caverweb/) to detect suitable tunnels and analyze ligand transport pathways. The analysis involved several steps, including identifying, analyzing, and visualizing the molecular pathways. CAVER was used to identify pathways connecting buried active sites to the protein surface by calculating tunnel geometries, including length, bottleneck radius, throughput, and curvature. The procedure of the CAVER Web tool is illustrated in Figure 3.3.

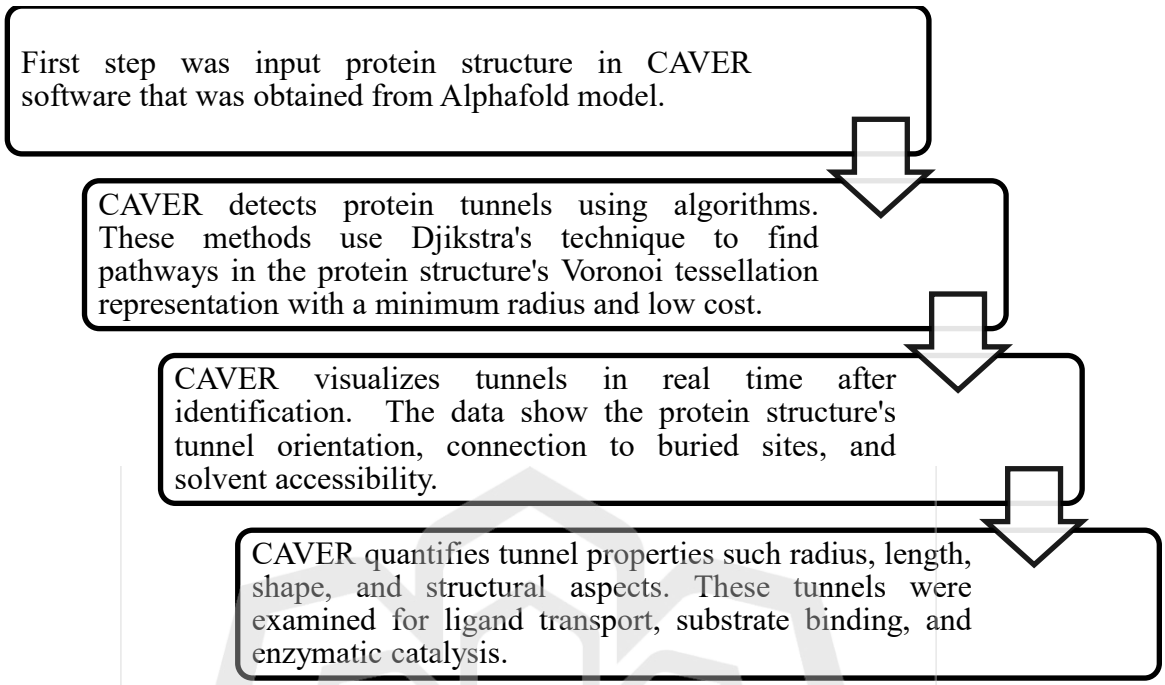


Figure 3.4 Workflow of pocket tunnel prediction using the CAVER web tool.

The CAVER Web analysis identified several potential binding tunnels suitable for ligand interactions. L-tryptophan was specifically selected as the ligand because the predicted binding site corresponded with the functional annotation from the InterProScan analysis. The two-dimensional structure of L-tryptophan was retrieved from the PubChem database and converted into the appropriate format for docking. The protein and ligand structures in PDB format were then used as input files, and the predicted tunnels were visualized in PyMOL software for structural interpretation. This analysis revealed possible access routes to the active site, contributing to a deeper understanding of the functional characteristics of the TDC protein.

3.2.3.2 Protein and Ligand Preparation

Prior to docking simulations, both the protein target and ligand molecules were carefully prepared to ensure the accuracy and reliability of the results. The tertiary structure of the TDC protein was obtained from the AlphaFold model, which was previously validated using the Ramachandran plot and ERRAT score. The protein structure was processed by removing water molecules, adding hydrogen atoms, and optimizing the geometry to ensure compatibility with docking algorithms. The ligand, L-tryptophan, was retrieved from the PubChem database (pubchem.ncbi.nlm.nih.gov) in Structure Data File (.sdf) format. Both the protein and ligand were converted into PDB format using PyMOL software, and the finalized structures were prepared for subsequent molecular docking simulations.

3.2.3.3 Molecular Docking Simulations

After protein and ligand preparation, molecular docking was performed to predict the binding affinity and binding pose of L-tryptophan within the TDC protein. To ensure reliability and consistency of the predictions, docking was conducted using two independent web-based platforms: Webina and SeamDock. Employing multiple docking tools allowed for cross-validation of the results and reduced potential bias associated with any single docking algorithm.

For the Webina docking protocol, the protein and ligand structures in PDB format were converted into Protein Data Bank, Partial Charge & Atom Type (.pdbqt) format by adding hydrogen atoms at physiological pH (7.4). A general grid box was constructed with coordinates of 0.8, 0.08, and 1.11 in the x, y, and z directions, respectively, and a box size of 19.39 × 20.23 × 18.03 Å (Angstrom). Once the parameters were defined, the docking simulations were executed, and the results were recorded for further analysis in Chapter 4.

For the SeamDock docking protocol, the protein and ligand structures were uploaded directly through the web interface. The receptor grid was centered at (0, 0, 0) with dimensions of 20 Å in each direction. The docking procedure was then performed, and the generated docking poses were collected for subsequent evaluation and visualization.

3.2.3.4 Energy Minimization in YASARA

After molecular docking, the protein–ligand complexes were optimized using the YASARA Energy Minimization Server to obtain a more accurate representation of the interactions. YASARA uses force-field computations to remove undesirable bond geometries, alleviate tension, eliminate steric interactions, and optimize atomic locations. By lowering the potential energy of the docked complexes, this procedure stabilizes the protein–ligand conformations and produces structurally reliable models suitable for subsequent interpretation and analysis (Krieger & Vriend, 2014).

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 EVALUATION OF POTENTIAL *TDC* GENE

The *TDC* gene was validated from the mature leaves of the *M. speciosa* plant through DNA extraction, PCR amplification, and sequencing. The CTAB method was used for genomic DNA extraction as described in methodology (Section 3.1.2). This approach is well-known for its ability to isolate high-quality DNA from plants that are rich in polysaccharides and secondary metabolites, which commonly interfere with downstream analyses (Gill et al., 2025; Haque et al., 2008). Furthermore, the CTAB method demonstrated high DNA concentrations and better integrity of nucleic acids than the extraction kits, making it useful for plants with complex secondary metabolites (Riascos-España et al., 2025; Liu et al., 2024; Doyle & Doyle, 1990).

DNA concentration and purity are both critical in molecular biology applications since contamination in nucleic acids can lead to inaccurate quantification of DNA and interfere with the downstream process. In this study, the nucleic acids concentration and purity were established using NanoDrop spectrophotometer, based on the absorbance ratios 260nm to 280nm (A260/A280) and 260 nm to 230 nm (A260/A230) ratios. As shown in Table 4.1, the nucleic acid concentrations obtained from six *M. speciosa* leaf samples ranged from 733.4 to 2751.3 ng/µL, respectively. The variation in yield is consistent with previous reports, as factors such as leaf physiology, developmental stage, and metabolite content can influence extraction efficiency (Hazrati et al., 2024; Lück et al., 2020). In this study, mature leaves were used since they contain more mitragynine compound and indole alkaloids than young leaves, and prior research had demonstrated a correlation between the expression of biosynthetic genes for *TDC* and the alkaloid synthesis (Veeramohan et al., 2023; Wungsintaweekul et al., 2012).

Additionally, Table 4.1 also provides the absorbance ratios (A260/A280 and A260/A230) of the extracted DNA, offering an indication of the purity of the samples. The A260/A280 ratios for the six DNA samples were 1.89, 1.87, 1.78, 1.85, 1.69, and 1.84, while their corresponding A260/A230 ratios were 1.62, 1.60, 2.19, 2.16, 1.71, and 2.25. These variations in the absorbance ratios may reflect inefficiencies in the CTAB extraction process, potentially indicating contamination by proteins, polysaccharides, or residual chemicals, as they will interfere with the next molecular analysis; hence, rigorous analysis of DNA purity is unavoidable (Chávez-Medina et al., 2019; Matlock, 2015). From the results, the A260/A280 ratios indicate that DNA_1, 2, 3, 4, and 6 can be considered as good quality for DNA, as their values were close to or above 1.8. In contrast, DNA_5, with a ratio below 1.7, may suggest possible contamination by proteins or phenolic compounds. For the A260/A230 ratios, DNA_3, 4, and 6 recorded values above 2.0, indicating minimal contamination by organic compounds or salts. In contrast, DNA_1, 2, and 5 showed lower values (<2.0), suggesting possible carryover of polysaccharides or other impurities (De Silva et al., 2025; Ning et al., 2009). Based on the explanation above, DNA_3 and 4 were selected for PCR analysis. These samples were thus used for downstream PCR amplification to ensure high-fidelity gene recovery.

Table 4.1 Results of the Nanodrop Spectrometer for *M. speciosa*'s leaves

Plant Sample (mature leaves)	Nucleic Acid Concentration (ng/µl)	A260/A280 ratios	A260/A230 ratios
DNA_1	878.5	1.89	1.62
DNA_2	767.4	1.87	1.60
DNA_3	2751.3	1.78	2.19
DNA_4	1317.6	1.85	2.16
DNA_5	1372.2	1.69	1.71
DNA_6	733.4	1.84	2.25

After selecting the best DNA extract, the DNA was subjected to the next analysis, PCR amplification, using the multiple working primers designed from the BLAST-primers database. Since the *TDC* gene sequence was too long to be amplified in a single reaction, it is crucial to divide it into six overlapping fragments, as presented in Table 3.1. Therefore, Figure 4.1 illustrates the PCR amplification results of the six primer fragments, where clear bands were observed within the expected size range. The estimated fragment sizes were determined by comparing the amplified bands with the DNA ladder, and these estimates corresponded closely with the expected sizes listed in Table 3.1. The primers specified in Table 3.1 were employed to amplify fragments measuring approximately 727 bp (TDC-01_F and TDC-01_R), 669 bp (TDC-02_F and TDC-02_R), 812 bp (TDC-03_F and TDC-03_R), 685 bp (TDC-04_F and TDC-04_R), 286 bp (TDC-05_F and TDC-05_R), and 714 bp (TDC-06_F and TDC-06_R). This agreement suggests a high likelihood that the amplified fragments represent the correct target regions of the *TDC* gene, thereby confirming that the PCR amplification was successful. The amplified fragments were subsequently sent for DNA sequencing to ascertain the precise arrangement of the nucleotides (adenine, thymine, cytosine, and guanine) within the PCR products. The resulting sequences were assembled and aligned with the reference *M. speciosa* TDC from Thailand to identify and extract the exon regions. The confirmed coding sequence was then translated *in silico* to generate the corresponding TDC amino acid sequence.

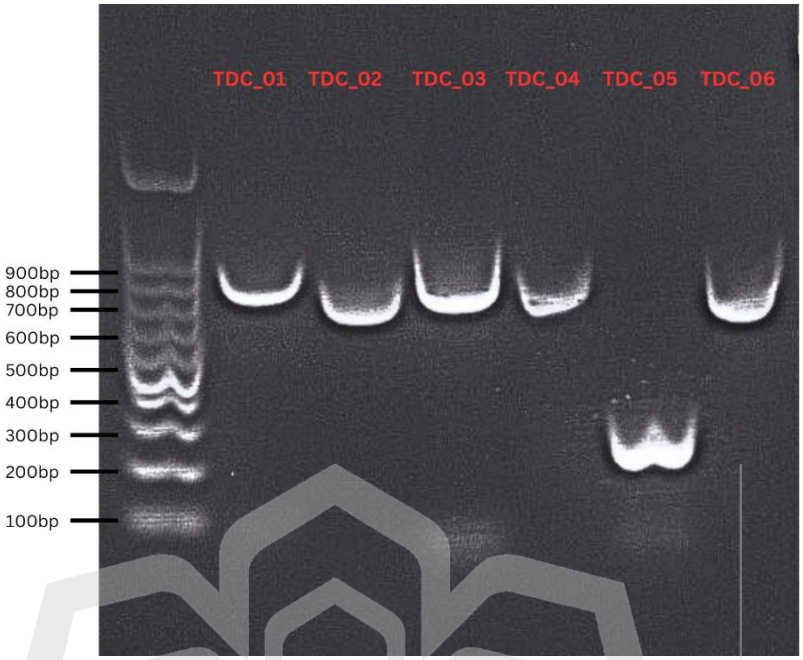


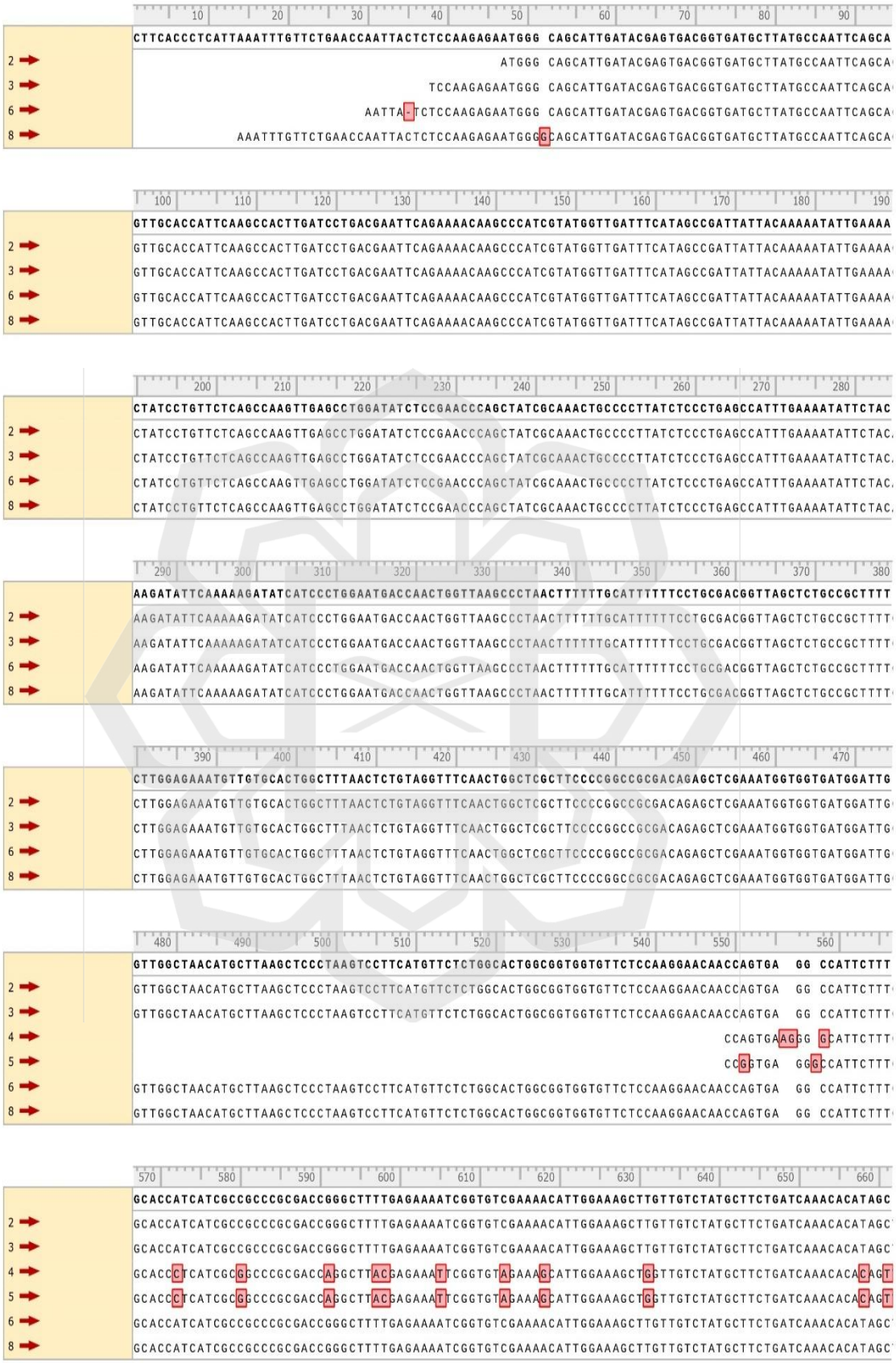
Figure 4.1 PCR extract isolated from the TDC of *M. speciosa* sequence identified using 100 bp DNA ladder and gene-specific primers.

4.2 IN SILICO ANALYSIS OF TDC GENE AND PROTEIN PROPERTIES

4.2.1 Sequence Alignment using SnapGene Software

After successful PCR amplification using six sets of primers that covered the full length of the *TDC* gene (1620 bp), the resulting fragments were assembled into a complete sequence using SnapGene software in Figure 4.2. Based on Figure 4.2, eight high-quality sequencing reads were obtained after trimming and assembly: fragment 1 from TDC_1 (forward read only), fragments 2 and 3 from TDC_2 (forward and reverse reads), fragments 4 and 5 from TDC_3 (forward and reverse reads), fragments 6 and 7 from TDC_4 (forward and reverse reads), and fragment 8 from TDC_5 (forward read only). The unequal number of reads

reflects differences in sequencing quality after trimming, where only high-quality reads were retained for assembly. Specifically, redundancy was observed, with TDC_2, TDC_4, and TDC_6 aligning at the 5′ end, TDC_3 positioned in the middle, and TDC_1 and TDC_5 overlapping at the 3′ end. Overlapping sequencing reads enhance consensus sequence reliability by providing redundancy that enables verification of base calls and reduces sequencing errors during contig assembly (Heather & Chain, 2016; Ewing & Green, 1998). The assembled sequence demonstrated high similarity to the TDC template from Thailand, but also revealed several nucleotide substitutions or additions, including transitions and transversions. This observation of codon changes may reflect genetic variations of *M. speciosa*, translation efficiency, and enzyme activity. Prior research indicates that the modification in codon usage can affect protein folding and function, ultimately altering metabolic pathways (Parvathy et al., 2022; Liu et al., 2021; Walsh et al., 2020). The successful reassembly of the TDC sequence in this study highlights the value of using multiple sets of overlapping primers to overcome PCR limitations. According to Hilgarth and Lanigan (2020), this method is particularly effective in amplifying extensive and challenging genomic regions, ensuring complete coverage and enhancing sequence accuracy. Thus, the complete sequence alignment is presented in Appendix B to support future applications and reference.


10 20 30 40 50 60 70 80 90
CTTCACCCTCATTAAATTTGTTCTGAACCAATTACTCTCCAAGAGAATGGG CAGCATTGATACGAGTGACGGTGATGCTTATGCCAATTCAGCA
2 ATGGG CAGCATTGATACGAGTGACGGTGATGCTTATGCCAATTCAGCA
3 TCCAAGAGAATGGG CAGCATTGATACGAGTGACGGTGATGCTTATGCCAATTCAGCA
6 AATTA-TCTCCAAGAGAATGGG CAGCATTGATACGAGTGACGGTGATGCTTATGCCAATTCAGCA
8 AAATTTGTTCTGAACCAATTACTCTCCAAGAGAATGGGGCAGCATTGATACGAGTGACGGTGATGCTTATGCCAATTCAGCA
100 110 120 130 140 150 160 170 180 190
GTTGCACCATTCAAGCCACTTGATCCTGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTATTACAAAAATATTGAAAA
2 GTTGCACCATTCAAGCCACTTGATCCTGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTATTACAAAAATATTGAAAA
3 GTTGCACCATTCAAGCCACTTGATCCTGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTATTACAAAAATATTGAAAA
6 GTTGCACCATTCAAGCCACTTGATCCTGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTATTACAAAAATATTGAAAA
8 GTTGCACCATTCAAGCCACTTGATCCTGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTATTACAAAAATATTGAAAA
200 210 220 230 240 250 260 270 280
CTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATCTCCGAACCCAGCTATCGCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAATATTCTAC
2 CTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATCTCCGAACCCAGCTATCGCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAATATTCTAC
3 CTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATCTCCGAACCCAGCTATCGCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAATATTCTAC
6 CTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATCTCCGAACCCAGCTATCGCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAATATTCTAC
8 CTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATCTCCGAACCCAGCTATCGCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAATATTCTAC
290 300 310 320 330 340 350 360 370 380
AAGATATTCAAAAAGATATCATCCCTGGAATGACCAACTGGTTAAGCCCTAACTTTTTTGCATTTTTTCCTGCGACGGTTAGCTCTGCCGCTTTT
2 AAGATATTCAAAAAGATATCATCCCTGGAATGACCAACTGGTTAAGCCCTAACTTTTTTGCATTTTTTCCTGCGACGGTTAGCTCTGCCGCTTTT
3 AAGATATTCAAAAAGATATCATCCCTGGAATGACCAACTGGTTAAGCCCTAACTTTTTTGCATTTTTTCCTGCGACGGTTAGCTCTGCCGCTTTT
6 AAGATATTCAAAAAGATATCATCCCTGGAATGACCAACTGGTTAAGCCCTAACTTTTTTGCATTTTTTCCTGCGACGGTTAGCTCTGCCGCTTTT
8 AAGATATTCAAAAAGATATCATCCCTGGAATGACCAACTGGTTAAGCCCTAACTTTTTTGCATTTTTTCCTGCGACGGTTAGCTCTGCCGCTTTT
390 400 410 420 430 440 450 460 470
CTTGGAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTTCAACTGGCTCGCTTCCCCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTG
2 CTTGGAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTTCAACTGGCTCGCTTCCCCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTG
3 CTTGGAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTTCAACTGGCTCGCTTCCCCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTG
6 CTTGGAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTTCAACTGGCTCGCTTCCCCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTG
8 CTTGGAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTTCAACTGGCTCGCTTCCCCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTG
480 490 500 510 520 530 540 550 560
GTTGGCTAACATGCTTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCAAGGAACAACCAGTGA GG CCATTCTTT
2 GTTGGCTAACATGCTTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCAAGGAACAACCAGTGA GG CCATTCTTT
3 GTTGGCTAACATGCTTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCAAGGAACAACCAGTGA GG CCATTCTTT
4 CCAGTGAAGGG GCATTCTTT
5 CCGGTGA GGGCCATTCTTT
6 GTTGGCTAACATGCTTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCAAGGAACAACCAGTGA GG CCATTCTTT
8 GTTGGCTAACATGCTTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCAAGGAACAACCAGTGA GG CCATTCTTT
570 580 590 600 610 620 630 640 650 660
GCACCATCATCGCCGCCCGCGACCGGGCTTTTGAGAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCTGATCAAACACATAGC
2 GCACCATCATCGCCGCCCGCGACCGGGCTTTTGAGAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCTGATCAAACACATAGC
3 GCACCATCATCGCCGCCCGCGACCGGGCTTTTGAGAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCTGATCAAACACATAGC
4 GCACCCTCATCGCGGCCCGCGACCAGGCTTACGAGAAATTCGGTGTAGAAAGCATTGGAAAGCTGGTTGTCTATGCTTCTGATCAAACACACAGT
5 GCACCCTCATCGCGGCCCGCGACCAGGCTTACGAGAAATTCGGTGTAGAAAGCATTGGAAAGCTGGTTGTCTATGCTTCTGATCAAACACACAGT
6 GCACCATCATCGCCGCCCGCGACCGGGCTTTTGAGAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCTGATCAAACACATAGC
8 GCACCATCATCGCCGCCCGCGACCGGGCTTTTGAGAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCTGATCAAACACATAGC


670 680 690 700 710 720 730 740 750
TTTTTCGTCAAGACTTGCAAATTGGCTGGTATCTT CCCATGCAATATAAGAATAATCCCTACTACTGCT GATGACAACTTTTCCATGTCACC
2 TTTTTCGTCAAGACTTGCAAATTGGCTGGTATCTTTCCCCATGCAA
3 TTTTTCGTCAAGACTTGCAAATTGGCTGGTATCTT CCCATGCAA
4 TTTTTCGTCAAGACTTGCAAATTGGCAGGTATTTT CCCGCAAAATATAAGAATAATCCCTACTACTGCTAGA-GAGAACTTTTCCATGTCCCC
5 TTTTTCGTCAAGACTTGCAAATTGGCAGGTATTTT CCCGCAAAATATAAGAATAATCCCTACTACTGCTAGA-GAGAACTTTTCCATGTCCCC
6 TTTTTCGTCAAGACTTGCAAATTGGCTGGTAT
8 TTTTTCGTCAAGACTTGCAAATTGGCTGGTATCTT CCCATGCAA
760 770 780 790 800 810 820 830 840
TGATGCTCTACGTAAACAAATTGAAGCTGACGTG GAAGATGGGCTGGTCCCACTTTTTATCTGTGCTACTGTGGGGACAACTTCTACCACTGCA
1 ACTGTGGGGACAACTTCCACCACTGCC
4 TCAGGCTGTACGTCGACAAATGCAAGCTGACGTGAGAA-ATGGGCTGGTCCCACTTTTCCTCTGTGCTACTGTGGGAACCACTTCCACCACTGCC
5 TCAGGCTGTACGTCGACAAATGCAAGCTGACGTGAGAA-ATGGGCTGGTCCCACTTTTCCTCTGTGCTACTGTGGGAACCACTTCCACCACTGCC
850 860 870 880 890 900 910 920 930 940
ATTGACCCAGTGAGTGAACTAGCCGATGTTGCAAATGATTTCAACGTGTGGATTCACGTTGATGCTGCCTACGCAGGTAGTGCATGCATATGTCC
1 ATTGACCCGGTGAGTCAACTAGCTGAAATTGCAAATAATTTCAACGTGTGGATTCACGTTGATGGTGCTTATGCAGGCAGTGCATGCATATGTCC
4 ATTGACCCGGTGAGTCAACTAGCTGAAATTGCAAATAATTTCAACGTGTGGATTCACGTTGATGGTGCTTATGCAGGCAGTGCATGCATATGTCC
5 ATTGACCCGGTGAGTCAACTAGCTGAAATTGCAAATAATTTCAACGTGTGGATTCACGTTGATGGTGCTTATGCAGGCAGTGCATGCATATGTCC
950 960 970 980 990 1000 1010 1020 1030
CGAGTTCAGGCAATACTTGGATGGGATTGAGCGAGTAGACTCTTTGAGTTTGAGCCCTCATAAGTGGCTACTTTGTTACTTGGACTGTTGTTGCT
1 CGAGTTCAGGCAATACTTGGAAGGGATCGAGCGAGTTGACTCGTTGAGCCTGAGCCCTCACAAGTGGCTACTTTGTTACTTGGACTGTTGTTGCT
4 CGAGTTCAGGCAATACTTGGAAGGGATCGAGCGAGTTGACTCGTTGAGCCTGAGCCCTCACAAGTGGCTACTTTGTTACTTGGACTGTTGTTGCT
5 CGAGTTCAGGCAATACTTGGAAGGGATCGAGCGAGTTGACTCGTTGAGCCTGAGCCCTCACAAGTGGCTACTTTGTTACTTGGACTGTTGTTGCT
040 1050 1060 1070 1080 1090 1100 1110 1120 1130
TGTGGGTGAAGAAAACAGACTTGCTGGTGAAGGCATTAGCCACCAATCCCGAGTACTTGAGAAACAAAAGAAGCGAGTTCGACTCAGTTGTGGAT
1 TGTGGGTGAAGAAAACAGACTTGCTGGTGAAGGCATTAGCCACCAATCCCGAGTACTTGAGAAACAAAAGAAGCGAGTTCGACTCAGTTGTGGAT
4 TGTGGGTGAAGAAAACAGACTTGCTGGTGAAGGCATTAGCCACCAATCCCGAGTACTTGAGAAACAAAAGAAGCGAGTTCGACTCAGTTGTGGAT
5 TGTGGGTGAAGAAAACAGACTTGCTGGTGAAGGCATTAGCCACCAATCCCGAGTACTTGAGAAACAAAAGAAGCGAGTTCGACTCAGTTGTGGAT
1140 1150 1160 1170 1180 1190 1200 1210 1220
TACAAAGACTGGCAAATTGGCACAGGTAAACGTTTCAGGGCATTACGATTATGGCTTGTTATGCGTTGTTATAGAGTTGCAAACCTACAAAGCCA
1 TACAAAGACTGGCAAATTGGCACAGGTAAACGTTTCAGGGCATTACGATTATGGCTTGTTATGCGTTGTTATGGAGTTGCAAACCTACAAAGCCA
4 TACAAAGACTGGCAAATTGGCACAGGTAAACGTTTCAGGGCATTACGATTATGGCTTGTTATGCGTTGTTATGGAGTTGCAAACCTACAAAGCCA
5 TACAAAGACTGGCAAATTGGCACAGGTAAACGTTTCAGGGCATTACGATTATGGCTTGTTATGCGTTGTTATGGAGTTGCAAACCTACAAAGCCA
230 1240 1250 1260 1270 1280 1290 1300 1310 1320
CATCCGATCCGACGTTCAAATGGCCAAGATGTTTGAAGGCTTTGTGAAATCTGACCCGAGATTTGAAATGATAGTACCACGTGCATTTTCACTCG
1 CATCCGATCCGACGTTCAAATGGCCAAGATGTTTGAAGGCTTTGTGAAATCTGACCCAAGATTTGAAATGATAGTACCACGTGCATTTTCACTCG
4 CATCCGATCCGACGTTCAAATGGCCAAGATGTTTGAAGGCTTTGTGAAATCTGACCCAAGATTTGAAATGATAGTACCACGTGCATTTTCACTCG
5 CATCCGATCCGACGTTCAAATGGCCAAGATGTTTGAAGGCTTTGTGAAATCTGACCCAAGATTTGAAATGATAGTACCACGTGCATTTTCACTCG
1330 1340 1350 1360 1370 1380 1390 1400 1410
TATGTTTTAGGCT TAACCCTTCGGGAGGA TCCAACGAAGCAGACTTGGAGCTTCTAAACAAGAAGCTACTCGACCGGGTCAACTCAACAGGCC
1 TATGTTTTAGGCT TAACCCTTCGGGAGGA TCCAACGAAGCAGACTTGGAGCTTCTAAACAAGAAGCTACTCGACCGGGTCAACTCAACAGGCC
4 TATGTTTTAGGCT TAACCCTTCGGGAGGAATCCAACGAA
5 TATGTTTTAGGCT TAACCCTTCGGGAGAA T
7 TGTTTTAGGCCCTAACCCTTCGGGAGGA TCCAACGAAGCAGACTTGGAGCTTCTAAACAAGAAGCTACTCGACCGGGTCAACTCAACAGGCC

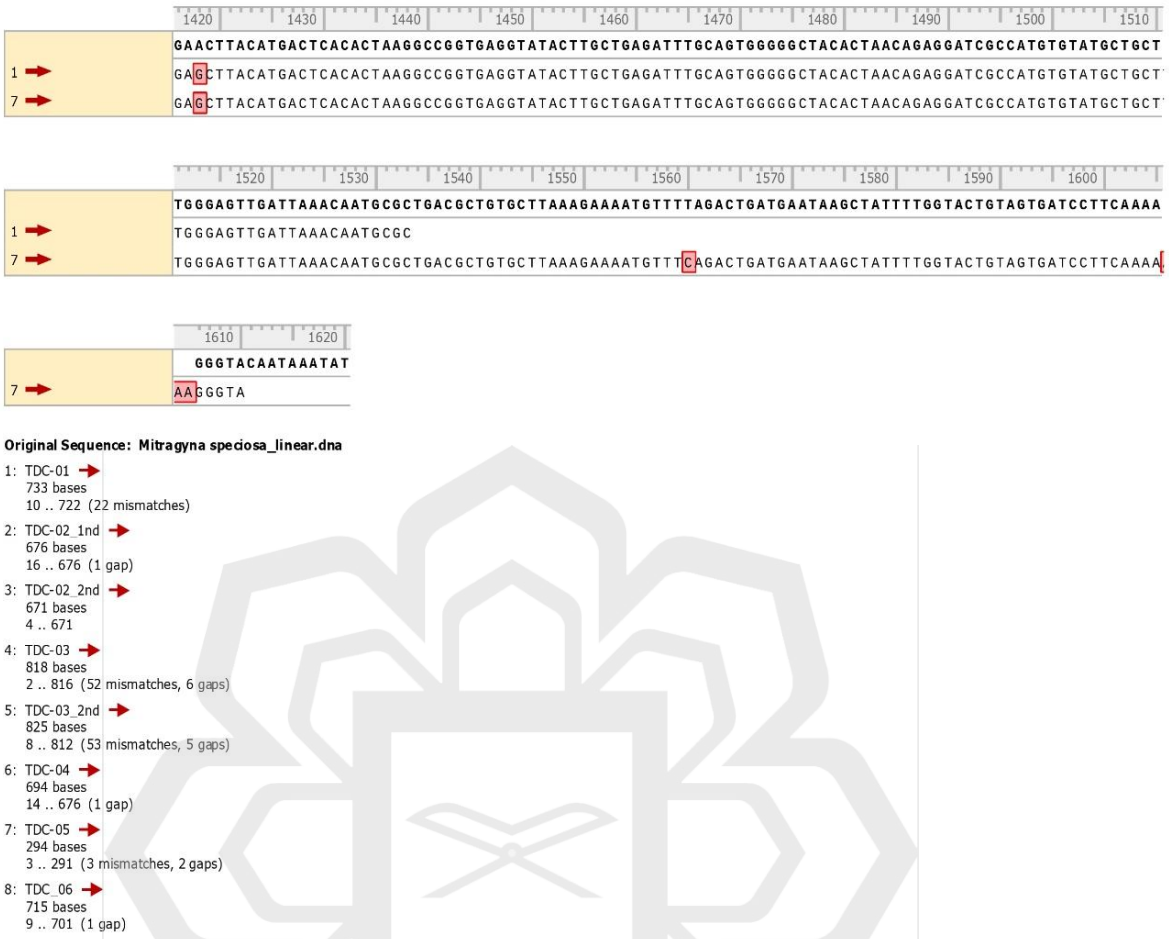


Figure 4.2 Alignment of six PCR fragments with the TDC template sequence. Red boxes indicate nucleotide variations, including substitutions, insertions, and transitions/transversions.

4.2.2 Protein Translation using Expasy Translate

After assembling the full-length *TDC* gene sequence, the next step was to predict its corresponding protein sequence. The analysis revealed the longest ORF to be 1599 bp in length, encoding 506 amino acid residues based on 5'3' Frame 1, which represents the putative TDC protein (Figure 4.3). Such approaches are commonly employed in molecular biology research to extract a reference protein sequence from DNA transcripts, as they provide reliable predictions of protein-coding regions. The identification of the longest ORF also enables the primer pairs design for gene amplification, thereby supporting future studies into protein function and phylogenetic relationships. The assembled sequence showed strong alignment with the published *M. speciosa* TDC coding sequence, indicating that the fragments correspond to exonic regions and allowing direct ORF prediction.

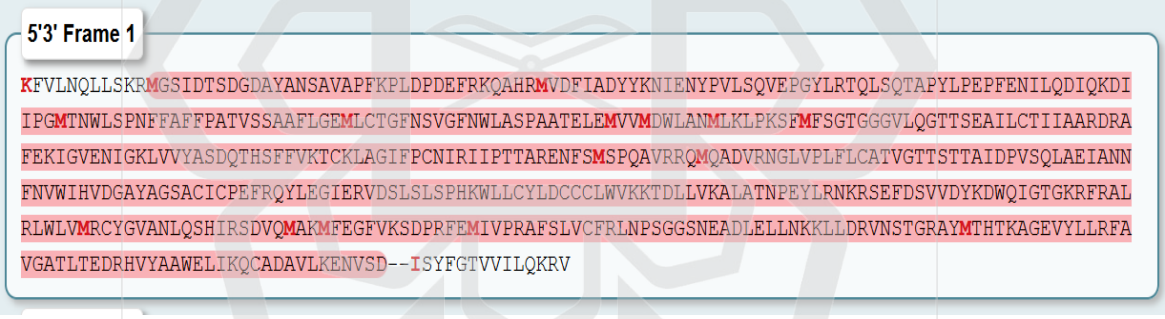
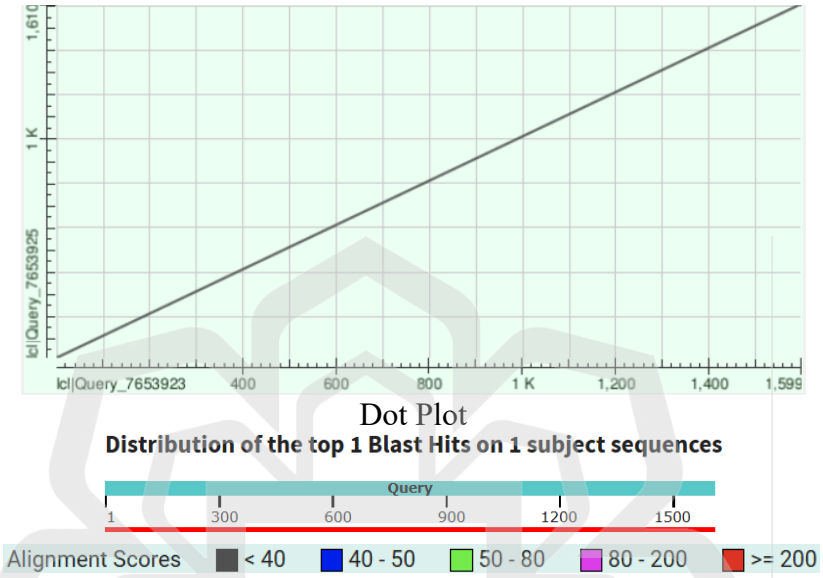


Figure 4.3 The longest ORF for the TDC protein sequence in the 5′–3′ direction (Frame 1).

4.2.3 Validation of Sequence using Needleman-Wunsch Alignment

The assembly of the TDC sequences against the Thailand TDC template sequences from both DNA and proteins were subjected to a homology search using the Needleman-Wunsch (NW) alignment, which performs a global alignment across the entire sequence length (Needleman & Wunsch, 1970). The results revealed a high level of similarity between the assembled sequences and the reference TDC sequence, supporting the successful reconstruction of the full-length gene. These findings are consistent with previous studies reporting that global alignment is suitable for end-to-end comparison of complete gene sequences (Muhamad et al., 2015).

When the assembled TDC DNA sequence was aligned with the Thai reference template (Figure 4.4), the resulting dot plot displayed a continuous diagonal line with an alignment score exceeding 200 across approximately 1610 bp. This pattern indicates that the sequences are highly conserved, with no evidence of extensive rearrangements, inversions, or large insertions/deletions. The presence of such a consistent diagonal line confirms strong homology across the full sequence length, providing confidence that the assembled *TDC* gene was accurately reconstructed and corresponds well to the reference gene. Complementing the dot plot, the alignment statistics summarized in Table 4.2 showed a NW score of 1414, with 96% sequence identity (1557 out of 1622 bp) and only 1% gaps (25 out of 1622 bp). Manual inspection further verified that the assembled sequence length (1599 bp) closely matched the Thai reference (1620 bp), with minor differences occurring mainly at the terminal regions. Importantly, the presence of only 1% gaps did not disrupt the overall alignment or gene structure, indicating smooth continuity between the sequences. Collectively, these results confirm that the assembled DNA sequence is highly similar to the Thai reference *TDC* gene, reflecting close evolutionary relatedness while maintaining high sequence integrity (Wungsintaweekul et al., 2012).



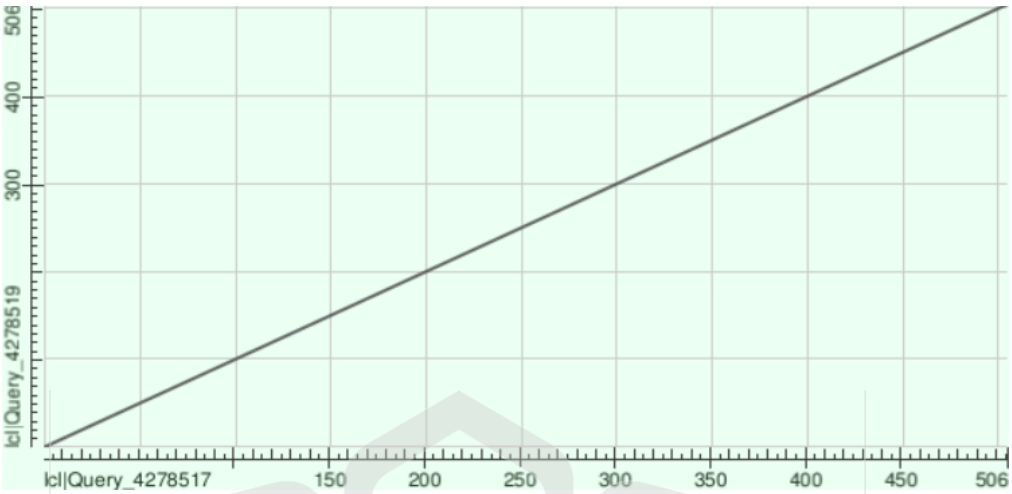
Query - TDC DNA sequence isolated from *M. speciosa* (Malaysia)
Subject - Reference TDC DNA sequence from *M. speciosa* (Thailand, NCBI database)

Figure 4.4 Pairwise global alignment (Needleman–Wunsch) between the assembled TDC DNA sequence from *M. speciosa* (Malaysia) and the reference TDC DNA sequence from *M. speciosa* (Thailand).

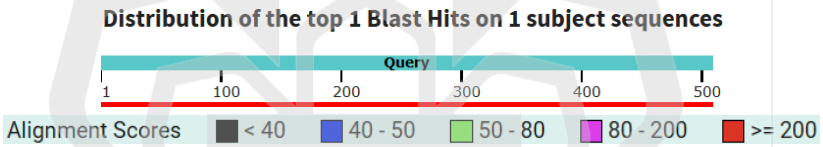
Table 4.2 Pairwise alignment statistics of the assembled TDC DNA sequence from *M. speciosa* (Malaysia) and the reference TDC DNA sequence from *M. speciosa* (Thailand) using the Needleman–Wunsch algorithm.

NW Score	Identities	Gaps	Strand
1412	1557/1622 (96%)	25/1622 (1%)	Plus/Plus

Next, the translated protein sequence was also aligned with the Thai reference protein using the same global alignment method and showed high sequence conservation (Figure 4.5). The graph revealed a continuous straight line with a score above 200 throughout the complete 506 amino acid residues. This indicated strong conservation with minimal insertions or deletions. Additionally, the alignment statistics were summarized in Table 4.3, which confirmed such observation with a NW score of 2582, 96% sequence similarity (487/506 residues), 98% positives (499/506 residues) and no gaps revealed. The absence of gaps and the high sequence identity between two proteins with almost identical basic structures are strongly indicative of conservation of enzymatic function. Sangar et al (2007) found that sequence identity over 50% is generally associated with functional conservation. This data strongly suggests that the Malaysia TDC protein is identical to the Thailand reference sequence, thereby verifying its validity as a good representation of the enzyme.



Dot Plot



Query - Translated TDC protein sequence from *M. speciosa* (Malaysia)
Subject - Reference TDC protein sequence from *M. speciosa* (Thailand)

Figure 4.5 Pairwise global alignment (Needleman–Wunsch) between the translated TDC protein sequence from *M. speciosa* (Malaysia) and the reference TDC protein sequence from *M. speciosa* (Thailand).

Table 4.3 Pairwise alignment statistics of the translated TDC protein sequence from *M. speciosa* (Malaysia) and the reference TDC protein sequence from *M. speciosa* (Thailand) using the Needleman–Wunsch algorithm.

NW Score	Identities	Positives	Gaps
2582	487/506 (96%)	499/506 (98%)	0/506 (0%)

The findings from both DNA and protein alignments clearly show that the TDC sequence from Malaysian *M. speciosa* is very similar to the Thai reference, affirming its validity as the orthologous gene. It revealed high similarity at both the DNA and protein levels, 96% identities with 98% positives, which implies good conservation, whereas the absence of gaps within the protein sequences indicates the structural elements and catalytic domains are conserved. Prior research found that enzymes with high sequence conservation tend to keep their structural conformations and enzymatic activities, since insertions or deletions within functional domains are rare (Panchenko et al., 2004; Nair & Rost, 2002). In conclusion, all these results showed that the Malaysian TDC sequence can be used as a foundation for later uses, like structure models, evolutionary studies and substrate-binding analysis.

4.2.4 Protein Domain Analysis using InterProScan

Protein domains are fundamental structural and functional units of proteins that can fold and evolve independently. They are important for understanding how proteins function, are classified, and how they are related through evolution. Identifying domains is also crucial for figuring out biological roles, protein evolution, and catalytic function (Triant & Pearson, 2015). InterPro is one of the most comprehensive resources for protein domain annotation, covering a wide range of domains accurately through the integration of 13 member databases, including Pfam, SMART, PROSITE Patterns and more (Paysan-Lafosse et al., 2023).

In this study, InterProScan reported the presence of two conserved domains, IPR002129 and PF00282, between amino acid positions 57 and 430 (Figure 4.5). The IPR002129 entry is for the pyridoxal phosphate (PLP)-dependent decarboxylase domain, which is a type of pyridoxal-dependent carboxylase in group II. Aromatic L-amino acid

decarboxylases are part of this enzyme family as they help decarboxylate aromatic amino acids, from tryptophan to tryptamine. The presence of the PLP-dependent decarboxylase domain in TDC confirms its classification as a PLP-dependent enzyme. PLP serves as a coenzyme in many enzymatic processes such as decarboxylation, transamination, deamination and others (Liang et al., 2019). This TDC domain catalyzes the decarboxylation of tryptophan to tryptamine, a key precursor in indole alkaloid biosynthesis. Thus, this conserved domain reveals the significance of TDC in the production of tryptamine for mitragynine synthesis in *M. speciosa.*

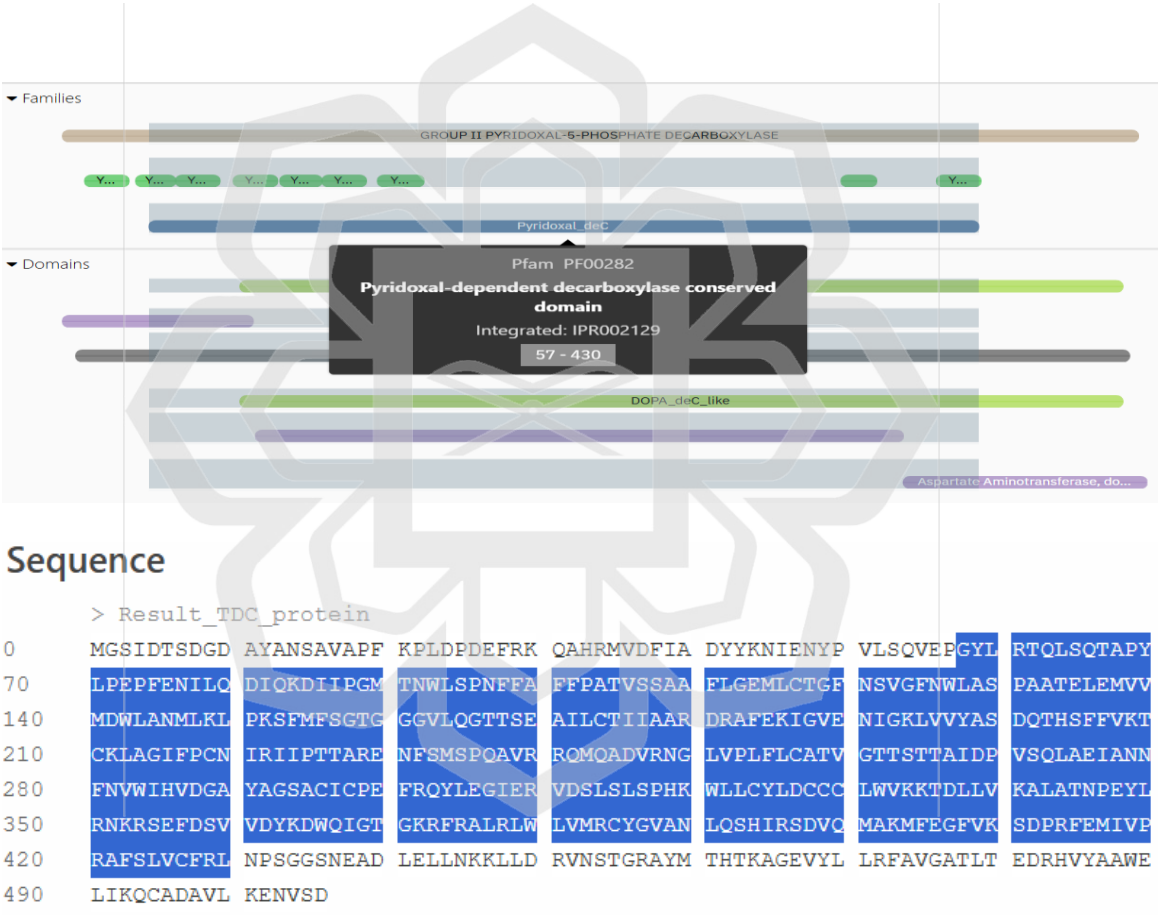


Figure 4.6 TDC protein domain (aa 57–430) involved in the decarboxylation of L-tryptophan to tryptamine.

InterProScan was further used to analyze the TDC protein sequence to identify functional domains and key active sites. The analysis revealed that the binding site, as shown in Figure 4.7, is located between amino acid positions 313 and 334, forming a region critical for substrate recognition and catalysis. This segment is predicted to accommodate L-tryptophan, the substrate for TDC, enabling the enzyme to catalyze its decarboxylation to form tryptamine. The precise positioning of this binding site underscores its importance for substrate specificity and catalytic efficiency and aligns with conserved motifs characteristic of PLP-dependent decarboxylases. This domain analysis provides the structural basis for further exploration of catalytic mechanisms through docking studies.

Figure 4.7 Predicted binding site of TDC protein showing amino acid positions 313–334.

4.2.5 Phylogenetic Tree

The Multiple Sequence Alignment (MSA) was performed using MUSCLE to align the TDC protein sequences obtained from *M. speciosa* with those of its related species. MSA is a fundamental bioinformatics tool used for the identification of conserved regions, functional motifs, and evolutionary relationships among protein sequences (Reddy & Fields, 2022; Patel et al., 2012; Thompson et al., 2005). The aligned sequences identified regions of high similarity and divergence that were used for subsequent phylogenetic analysis. Sixteen TDC protein sequences were selected from BLAST X as listed in Table

4.4 based on their plant species, percentage identities, total scores, E-values and accession numbers.

Table 4.4 summarizes the top 16 homologous protein sequences identified from the BLASTx search. The top three homologous sequences showed high similarity to the *M. speciosa* TDC protein from Malaysia, with the sequence identities of 96.25% (*M. speciosa* from Thailand), 88.69% (*Uncaria rhynchophylla*), and 85.32% (*Coffea arabica*). Sequences from other species belonging to the plant families such as Rubiaceae, Solanaceae and other families ranged from 70.86% to 85.15% identity. These findings indicate that the TDC protein, which has more than 50% identity, was well-preserved in related Rubiaceae species and has moderate similarity to proteins of related plant families (Sangar et al., 2007). The homology data were subsequently used for phylogenetic tree construction to elucidate the evolutionary relationships and conserved regions among the sequences.

Table 4.4. 16 top hits of the homology search of TDC using BLAST X

ACCESSIONS (Sequence ID)	**DESCRIPTION**	**IDENTITY (%)**	**TOTAL SCORE**	**E-VALUE**
AEQ01059.1	tryptophan decarboxylase [*Mitragyna speciosa*]	96.25	1024	0.0
WRW51055.1	tryptophan decarboxylase [*Uncaria rhynchophylla*]	88.69	905	0.0
XP_027061221.1	tryptophan decarboxylase TDC2-like [*Coffea arabica*]	85.32	897	0.0
XP_027172386.1	tryptophan decarboxylase TDC2-like [*Coffea eugenioides*]	85.12	895	0.0

XP_009760717.2	tryptophan decarboxylase TDC1-like [*Nicotiana sylvestris*]	76.54	840	0.0
XP_009622485.1	tryptophan decarboxylase TDC1-like [*Nicotiana tomentosiformis*]	76.94	837	0.0
XP_059284124.1	tryptophan decarboxylase TDC1-like [*Lycium ferocissimum*]	75.75	829	0.0
XP_060189824.1	tryptophan decarboxylase TDC1-like [*Lycium barbarum*]	75.35	822	0.0
XP_015081169.1	tryptophan decarboxylase TDC1-like [*Solanum pennellii*]	73.21	809	0.0
XP_016580730.1	tryptophan decarboxylase TDC1 [*Capsicum annuum*]	73.27	806	0.0
XP_004243301.1	tryptophan decarboxylase TDC1 [*Solanum lycopersicum*]	72.82	801	0.0
XP_057956277.1	tryptophan decarboxylase TDC2-like [*Malania oleifera*]	74.30	798	0.0
XP_049353196.1	tryptophan decarboxylase TDC2-like [*Solanum verrucosum*]	72.08	794	0.0
XP_057488279.1	tryptophan decarboxylase TDC2-like [*Actinidia eriantha*]	74.75	774	0.0
QBK17431.1	tryptophan decarboxylase [*Actinidia deliciosa*]	74.16	770	0.0
KAH9792819.1	Tryptophan decarboxylase 2 [*Citrus sinensis*]	70.86	770	0.0

The TDC protein sequences from *M. speciosa* and the other 16 related species were aligned using MSA by MUSCLE to evaluate the conservation of amino acid residues. The sequences of TDC protein as listed in Table 4.4 were used to generate the MSA, as illustrated in Figure 4.8, which displays the alignment by species names and BLASTX-derived sequences. The figure highlights the presence of the IPR002129 domain in yellow colour, positioned between 57 and 430 aa, which occurs predominantly in highly conserved regions. This indicates that the domain is functionally important and likely under strong evolutionary pressure to maintain its structure and role in TDC activity. MSA therefore provides an overall vision of sequence conservation and facilitates the identification of functional motifs and domains that contribute to protein stability and catalytic efficiency (Singh, 2015; Thompson et al., 2005).

In addition, three conservation symbols were observed: (1) Asterisks (*) symbol in red boxes represent identical residues in all sequences. (2) Colons (:) symbol in green boxes indicate strongly conserved substitutions, whereas (3) Periods (.) symbol in blue boxes denote weakly-conserved substitutions, which may reflect evolutionary divergence among species while preserving partial functionality (Thompson, 2005; Chenna et al., 2003). The presence of both strict and flexible regions implies that although the TDC protein core structure is conserved, some peripheral areas are more adaptable, presumably indicating functional diversity across plant families.

Phylogenetic tree analysis was conducted to provide a visual representation of evolutionary relationships among protein sequences from various species (Kaur & Kaur, 2024; Edwards, 2019; Sharma et al., 2019). This study aligned TDC protein sequences with MUSCLE and constructed a phylogenetic tree in MEGA X using the Maximum Likelihood method and the JTT matrix-based model, establishing a solid framework for assessing evolutionary relationships. The stability of the reconstructed tree was evaluated using bootstrap analysis with 1,000 replicates. Bootstrap (BS) values were employed to estimate the reliability of each node such that higher BS values indicate more significant statistical support for the inferred evolutionary relationships (Wiegert et al., 2024; Douady et al.,

2003). Generally, values greater than 95% are considered to reflect strong support since the probability of accepting a wrongly inferred tree under these conditions is less than 5% (Zharkikh & Li, 1992). Conversely, low bootstrap values are equivalent to weak support and cannot rightly establish evolutionary relationships.

Figure 4.8 Multiple sequence alignment of TDC sequences performed using MUSCLE. Asterisks (*) indicate identical residues across all sequences (red box), colons (:) represent strongly conserved substitutions (green box), and periods (.) denote weakly conserved substitutions (blue box). The *M. speciosa* sequence from Malaysia is highlighted with a purple box, while conserved domains are highlighted in yellow.

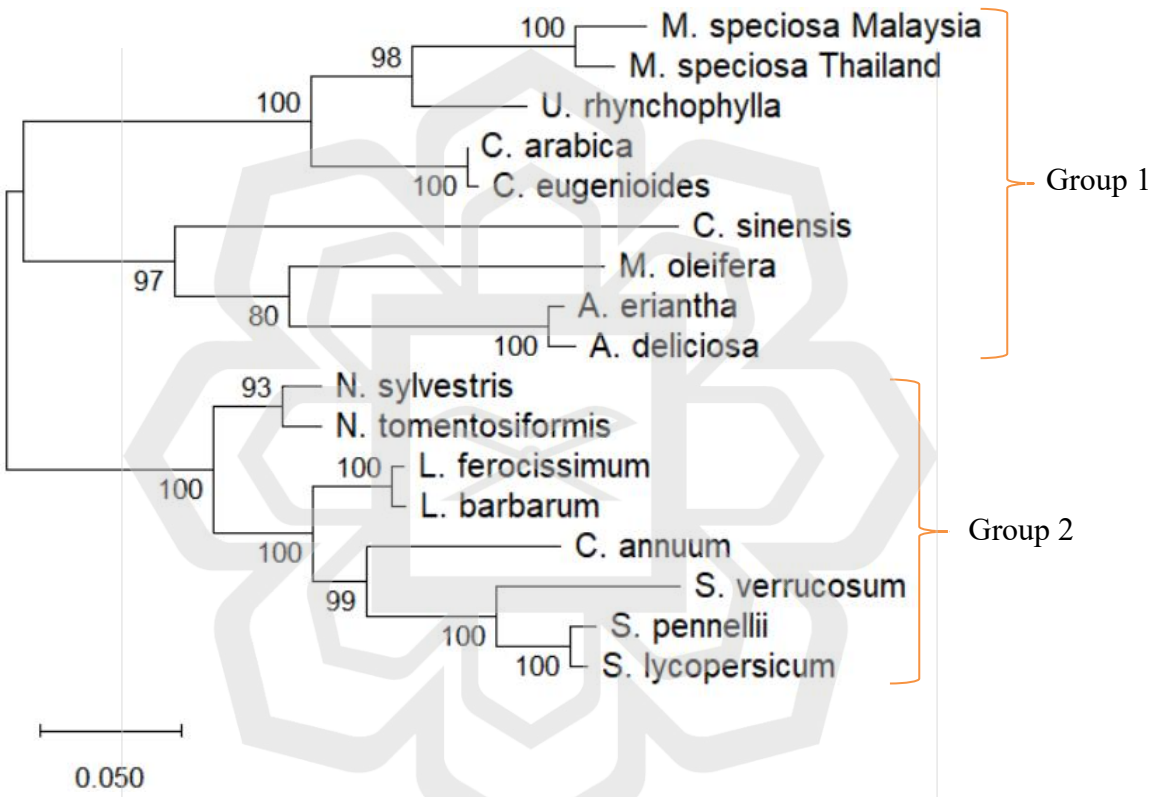


Figure 4.9 Phylogenetic tree of TDC *M. speciosa* using MEGA X.

The analysis involved 17 TDC protein sequences, which were clustered into two main groups based on plant family affiliations and sequence similarity (Figure 4.9), including *M. speciosa* from Malaysia. Group 1, located in the upper part of the tree, consists primarily of members of the Rubiaceae family, including *Mitragyna speciosa* (Malaysia and Thailand), *Uncaria rhynchophylla*, *Coffea arabica*, and *Coffea eugenioides*. These species clustered together with bootstrap support values (98–100), indicating a close evolutionary relationship. Interestingly, this group also included several non-Rubiaceae species such as *Citrus sinensis* (Rutaceae), *Malania oleifera* (Olacaceae), and *Actinidia eriantha* and *Actinidia deliciosa* (Actinidiaceae), which branched within the same cluster. Although these species belong to different families, they are grouped with Rubiaceae TDCs with moderately high bootstrap support values (80–97), suggesting some degree of sequence similarity despite their distinct taxonomic origins.

In contrast, Group 2 is positioned in the lower part of the tree and comprises exclusively members of the Solanaceae family. This includes *Nicotiana sylvestris*, *Nicotiana tomentosiformis*, *Lycium ferocissimum*, *Lycium barbarum*, *Capsicum annuum*, *Solanum verrucosum*, *Solanum pennellii*, and *Solanum lycopersicum*. The clustering in this group is strongly supported, with bootstrap values ranging from 93 to 100, reflecting high confidence in the evolutionary relationships within Solanaceae.

The scale bar (0.050) represents substitutions per site, with shorter branches indicating higher sequence similarity. Notably, *M. speciosa* from Malaysia clustered tightly with other Rubiaceae members, including *M. speciosa* (Thailand), *Coffea arabica*, *Coffea eugenioides*, and *Uncaria rhynchophylla*. This cluster exhibited strong bootstrap support (98–100), supporting the hypothesis that TDC plays a conserved role in indole alkaloid biosynthesis within the Rubiaceae family. These results align with previous studies showing close genetic and functional relationships among Rubiaceae members (Guo et al., 2021; Bremei, 2009).

4.2.6 Primary Structure and Physicochemical Properties of TDC

The ExPASy ProtParam tool was used to analyze the physicochemical properties of the TDC protein. The analysis of results, as presented in Figure 4.9, include parameters such as molecular weight, extinction coefficient, grand average of hydropathicity (GRAVY), theoretical isoelectric point (pI), aliphatic index, instability index, and the numbers of positively and negatively charged residues. These characteristics provide crucial information on the amino acid structure of TDC protein, emphasizing its molecular weight and isoelectric point (pI) in particular.

Table 4.5 Physicochemical characteristics of TDC by ExPasy ProtParam.

Parameters	TDC protein
Number of amino acids	506
Molecular weight (Da)	56.58 kDa
Aliphatic index	87.33
Theoretical Isoelectric points (pI)	6.05
Instability index	43.47 (unstable)
Grand average of hydropathicity (GRAVY)	-0.017
Total number of positively charged residues (Arg + Lys)	50
Total number of negatively charged residues (Asp + Glu)	54

Table 4.5 shows that the TDC protein in *M. speciosa* comprises 506 amino acids, with a molecular weight of 56.58 kDa. This weight is characteristic of plant aromatic amino acid decarboxylases and aligns with the average size for this enzyme family (Bisello et al., 2023). The theoretical isoelectric point (pI) of 6.05 indicates that the protein is slightly acidic, which influences its solubility, stability, and interactions with cellular molecules, thereby affecting enzymatic activity and metabolic processes (Mohanta et al., 2021; Kozłowski, 2017). The aliphatic index of 87.33 indicates a significant hydrophobic residue

content, which may contribute to thermal stability by preserving the protein's 3D structure under diverse conditions (Dutta et al., 2018). The instability index of 43.47, marginally exceeding 40, suggests a potential for *in vitro* instability, as proteins with elevated instability indices typically possess amino acid compositions that promote flexibility and disorder (Kaur et al., 2020). The elevated aliphatic index suggests thermal stability; however, it does not guarantee overall stability in physiological conditions (Gupta et al., 2012). Next, the GRAVY value of –0.017 indicates that the protein exhibits moderate hydrophilicity, which enhances its solubility in aqueous environments and is essential for enzymatic activity (Sariyatun et al., 2015). The charge residue composition includes 50 positive and 54 negative amino acids, indicating a slight predominance of acidic residues. This composition may improve solubility and facilitate interactions with other cellular molecules, while also contributing to structural flexibility (Das & Pappu, 2013).

Overall, the TDC protein exhibits physicochemical characteristics comparable to other plant decarboxylases, with a moderate molecular weight, slight acidity, good hydrophilicity, and high aliphatic content indicative of thermal stability, although its instability index suggests potential vulnerability under in vitro conditions. The molecular weight is moderate, typical of plant decarboxylases. These characteristics are crucial for understanding protein behaviour in functional studies or biotechnological applications.

4.2.7 Secondary Structure Prediction using the PSIPRED

Secondary structure prediction was conducted to determine the folding structures of the TDC protein, including the distribution of alpha helices, beta strands, and coil regions, which contribute to its stability and catalytic function. The role of secondary structure elements in the functional regions of proteins is crucial for their ability to interact with ligands or other proteins (Zhao et al., 2020). The secondary structural components were predicted using the PSIPRED server, which applies a neural network-based approach to the

amino acid sequence of TDC. Figure 4.10 illustrates that alpha helices are represented in pink, beta strands in yellow, and coil regions in grey. The protein exhibits a higher proportion of alpha helices, interspersed with beta strands and coil regions. The higher proportion of alpha helices in the TDC protein indicates a stable and compact structure, attributed to significant internal hydrogen bonding between backbone amide and carbonyl groups, which is crucial for preserving protein integrity under physiological conditions (Doig et al., 2001).

Alpha helices present distinct surfaces for interaction due to their amphipathic characteristics, which feature both hydrophobic and hydrophilic faces, facilitating binding with ligands, substrates, or other proteins (Buchwald et al., 2012). Meanwhile, although alpha helices contribute to stability, they exhibit a degree of flexibility that allows proteins to undergo local conformational changes essential for enzymatic activity and dynamic interactions (Meek et al., 2023). The properties outlined elucidate the predominance of alpha helices over beta strands and coil regions, thereby reinforcing the structural resilience and functional versatility of the TDC protein.The predominance of α-helices supports the compact folding observed in tertiary models.

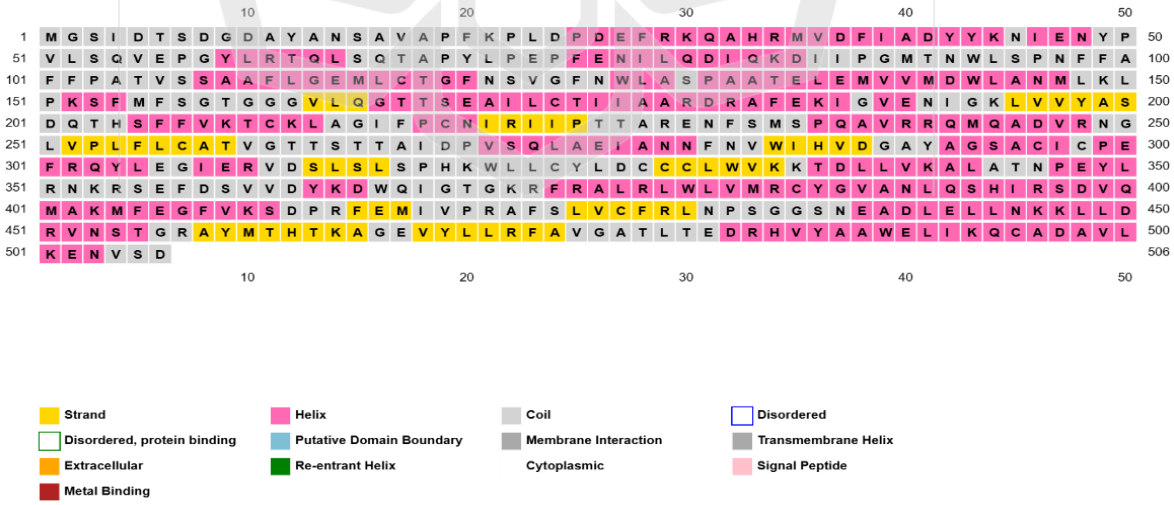


Figure 4.10 Sequence Plot of PSIPRED

4.2.8 Tertiary Structure Prediction and Evaluation

The tertiary structure of the TDC protein was predicted using multiple computational tools, including template-based predictions (TDC), SWISS-MODEL, AlphaFold, Phyre2, and I-TASSER (Figure 4.11). The overall stereochemical quality of the TDC protein models generated by the various template-based modelling servers was evaluated using ERRAT and PROCHECK through the SAVES 6.0 server. PROCHECK was employed to assess the stereochemical quality of the protein by evaluating the backbone geometry, including phi (ϕ) and psi (ψ) angles, and identifying residues in favoured, allowed, and disallowed regions of the Ramachandran plot. ERRAT, on the other hand, was used to evaluate the overall quality of the protein structure by analyzing non-bonded atom-atom interactions. Overall, assessment of the TDC models based on backbone geometry, residue interactions, and energy profiles confirmed their structural reliability for subsequent computational analyses.

Subsequently, the 3D models of the TDC protein were evaluated using ERRAT scores, which assess the reliability of a protein's tertiary structure based on non-bonded atomic interactions. As shown in Table 4.6, the SWISS-MODEL produced the highest ERRAT score of 96.17%, closely followed by AlphaFold at 95.80%, indicating that these models possess well-optimized atomic geometry and are likely to represent the true protein conformation accurately, consistent with Rashid et al. (2024), where higher ERRAT scores correspond to greater structural reliability. In comparison, Phyre2 yielded a moderate score of 86.36%, whereas I-TASSER exhibited an extremely low score of 8.05%, indicating poor structural reliability and substantial inconsistencies in the predicted conformation. Reliable tertiary structures are critical for understanding the functional mechanisms of TDC, such as substrate binding and catalysis, since inaccuracies in the 3D model could misrepresent active sites or folding patterns. Based on these results, both SWISS-MODEL and AlphaFold were identified as the most reliable high-quality models. The next step involved assessing the stereochemical quality of these high-quality models using the Ramachandran

plot, which will guide the selection of the most structurally robust model for subsequent binding site prediction and molecular docking studies.

Table 4.6 ERRAT scores of TDC protein structures generated by four modelling tools.

Protein model	ERRAT score (%)
AlphaFold	95.80
SWISS-MODEL	96.17
Phyre2	86.36
I-TASSER	8.05

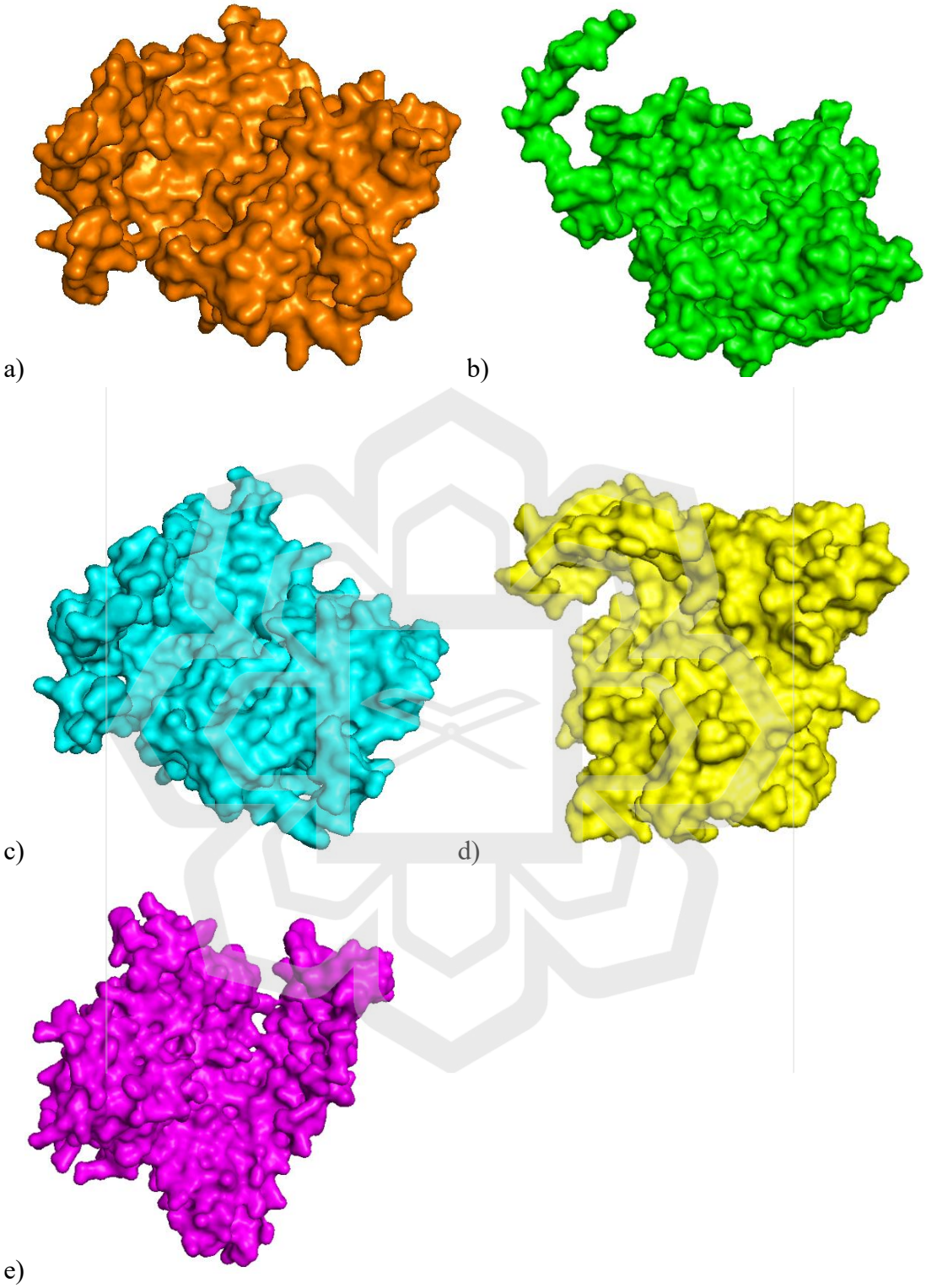


Figure 4.11 Tertiary structure of the TDC protein modelled by a) template-based prediction of TDC (orange), b) AlphaFold (green), c) SWISS-MODEL (blue), d) Phyre2 (yellow), and e) I-TASSER (purple). The active-site region (residues 313–334) is highlighted in hot pink.

The tertiary structure of the TDC protein was further evaluated using Ramachandran plot analysis through the PROCHECK server to assess the stereochemical quality of the models generated by four different modelling tools (Figure 4.12). The Ramachandran plot divides the φ–ψ dihedral angles into four regions: most favoured, additional allowed, generously allowed, and disallowed regions, as shown in Table 4.7. In an ideal model, a high-quality protein model should have the majority of residues in the most favoured regions, with minimal residues in disallowed regions. Among the models, AlphaFold exhibited the highest proportion of residues (93.3%) in the most favoured regions, with no residues in disallowed regions, indicating excellent stereochemical quality and structural reliability. These findings are consistent with the ERRAT score (95.80%), which further supports AlphaFold as the most accurate and representative tertiary model. For visual comparison, the structural superimposition of all predicted TDC models, including the template-based prediction, is shown in Appendix F. The figure demonstrates that, while all models share a similar overall fold, the AlphaFold model exhibits optimal stereochemical quality and a conserved active-site geometry, supporting its selection for molecular docking, as indicated by ERRAT and Ramachandran analyses.

Table 4.7 Ramachandran plot statistics for TDC protein models from four modelling tools.

Protein Model	**Most favoured regions (%)**	**Additional allowed regions (%)**	**Generously allowed regions (%)**	**Disallowed regions (%)**
Alphafold	93.3	6.2	0.4	0.0
SWISS-MODEL	92.7	6.9	0.1	0.3
Phyre2	91.6	7.2	0.9	0.2
iTasser	78.3	12.2	7.1	2.4

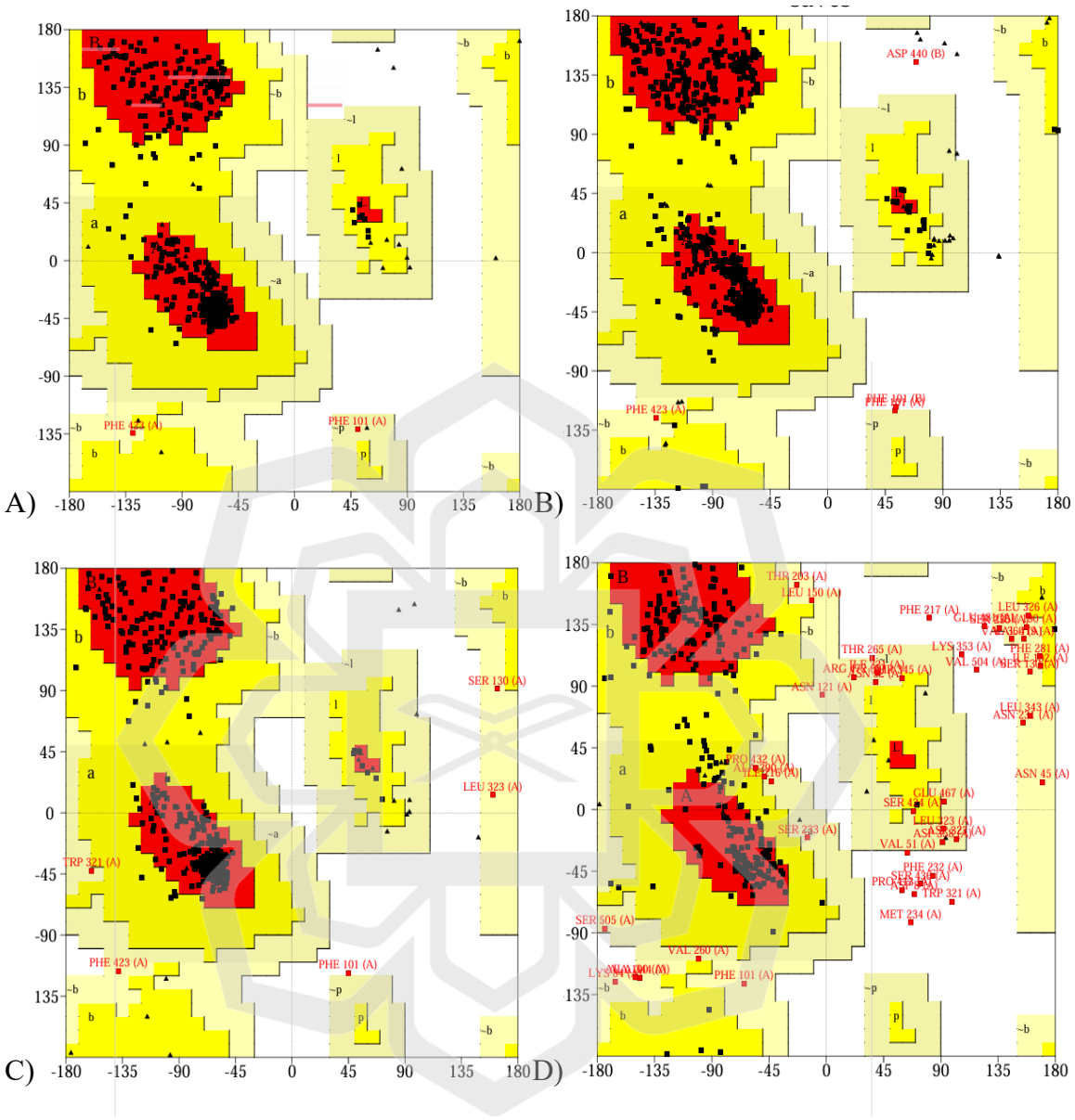


Figure 4.12 Ramachandran plot of TDC protein models from a) AlphaFold, b) SWISS-MODEL, c) Phyre2, and d) I-TASSER. This plot analysis shows the presence of amino acid residues in favourable region (red), additional allowed region (yellow), generously allowed region (light brown), and disallowed region (white).

4.3 MOLECULAR DOCKING

4.3.1 Active Site Pocket Tunnel Prediction of TDC Protein

Tunnel prediction was performed to identify the potential entry pathway and the binding pocket where the substrate, L-tryptophan, interacts with the TDC protein. This analysis provides insights into the accessibility of the active site, the arrangement of catalytic residues, and the potential routes for ligand transport within the enzyme structure (Kozlikova et al., 2014). For this study, the tertiary structure of the TDC protein, predicted by AlphaFold, was analyzed using CAVER Web v2.0, a computational tool for identifying tunnels and cavities in protein structures.

The CAVER analysis produced several potential pockets, which were ranked based on relevance score, cavity volume, and druggability index (Table 4.8). Among the predicted pockets, pocket ID 1 exhibited the highest relevance score (100%) and a druggability index of 0.826; however, this pocket was restricted to structural residues that did not align with known catalytic motifs, and visualization further confirmed that pocket ID 1 was located away from the catalytic cavity (Figure 4.13a). In contrast, pocket ID 2 showed a lower relevance score (89%) yet uniquely included key residues associated with catalytic activity, specifically H204, D288, S317, and H319 from Caver web. Although the druggability index is relatively low (0.053), the presence of these catalytic residues indicates that pocket ID 2 represented the functional active-site tunnel. Other predicted pockets (ID 3 and ID 13) were less relevant and did not contain catalytic residues. Consequently, pocket ID 2 was used for further analysis.

The predicted pockets were then visualized on the TDC protein using the AlphaFold model and compared with the ligand that performed in PyMOL software. Notably, the CAVER pocket ID 2 coincided with the ligand-binding position predicted by PyMOL, where L-tryptophan was located within the protein cavity. Both independent methods

indicated the same spatial location, confirming that the binding tunnel identified by CAVER overlapped precisely with the ligand-binding site visualized in PyMOL. This overlap strengthens the reliability of the finding, providing strong evidence that pocket ID 2 accurately represents the true catalytic cavity of the TDC protein (Figure 4.13b and c).

Table 4.8 Top pockets ID tunnels of CAVER

Pocket ID	Catalytic residues inside the pocket	Relevance score (%)	Volume [Å^3]	Druggability
1	F102, P103, A104	100	881	0.826
2	F102, T168, S169, H204, T263, D288, S317, H319	89	583	0.053
3	F102, P103, A104, H204, T263, H319	61	1102	0.379
13	S169	21	306	0.023

An analysis of the residues in pocket ID 2 identified multiple binding sites exhibiting approximately 70% identity to the cofactor pyridoxal 5'-phosphate (PLP) or the substrate L-tryptophan as predicted by CAVER Web. Among all the catalytic residues, H319 was prioritized for further discussion due to several factors: (i) it is located within the predicted binding site region (313–334 aa) as described in Section 4.2.4, (ii) it exhibited the highest identity score (70%) compared to other residues (38–70%), (iii) it is specifically annotated as an L-tryptophan binding site, and (iv) it is situated within the local SPHKW motif, which could offers a favourable chemical environment for substrate recognition as mentioned in section 4.. The features highlight that H319 serves as a crucial residue for anchoring L-tryptophan during catalysis, aligning with its anticipated function in the decarboxylation mechanism (Table 4.9)

Table 4.9. Annotation of predicted binding residues in pocket ID 2 from the CAVER web

Position	Residue	Acc. Code	Identity	Type	Description	Neighbourhood
319	H	P17770	70 %	binding site	L-tryptophan	SPHKW SPHKW

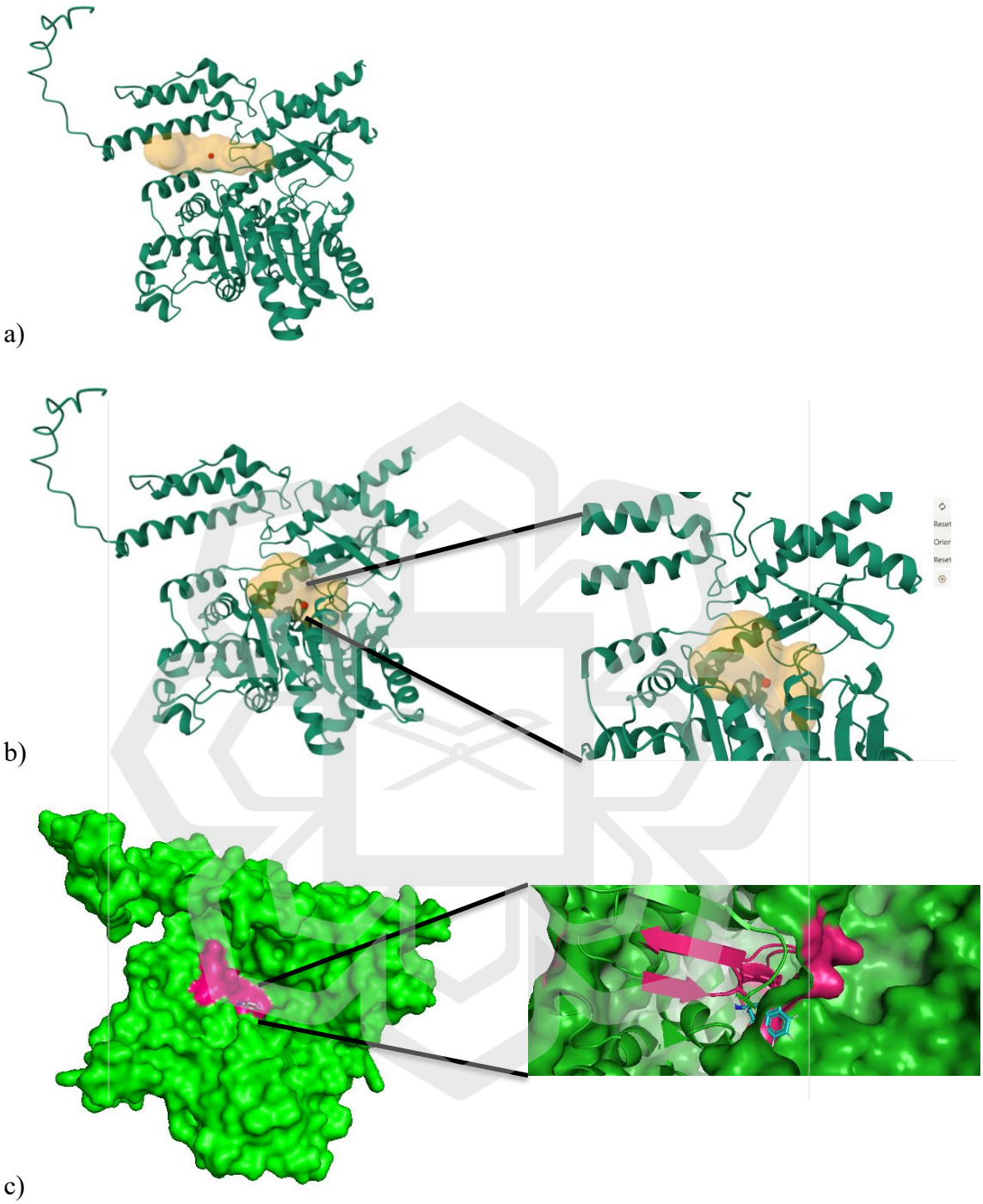


Figure 4.13 Visualization of the predicted active-site tunnel of the TDC protein using the AlphaFold model: (a) Pocket ID 1 predicted by CAVER Web, (b) Pocket ID 2 predicted by CAVER Web. The red dot indicates the potential ligand binding site in both pockets. (c) Protein-ligand docking in PyMOL software.

Following active-site pocket prediction and docking, ligand accessibility within the TDC protein was further investigated. As mentioned earlier, pocket ID 2 was selected for detailed interpretation due to its association with catalytic residues. The tunnel geometry parameters predicted by the CAVER algorithm are summarized in Table 4.10 (Stourac et al., 2019). The active-site tunnel displayed a minimum bottleneck radius of 2.25 Å and a tunnel length of 4.7 Å, suggesting a relatively direct pathway from the protein surface to the catalytic pocket. Although the bottleneck radius appears smaller than the approximate dimensions of L-tryptophan (~5–6 Å across the indole ring), surface representation and docking analysis (Figure 4.13c) confirm that the substrate can be accommodated within the tunnel, likely adopting a flexible conformation to traverse the bottleneck. The tunnel curvature of 1.17 indicates an almost linear path, thereby minimizing steric hindrance during substrate transport. Furthermore, the throughput value of 0.89 reflects a highly accessible and efficient transport channel. In addition, the maximum bottleneck radius (BR) was 3.02 Å, confirming that the tunnel was sufficiently wide to accommodate the entry of L-tryptophan into the catalytic cavity. Collectively, these parameters demonstrate that the predicted tunnel possesses structural features favourable for ligand accessibility and enzymatic catalysis.

Table 4.10 Tunnel geometry parameters by the CAVER algorithm

Parameters	Values
Bottleneck radius	2.25 Å
Maximum BR	3.02 Å
Tunnel length	4.7 Å
Curvature	1.17
Throughput	0.89

In summary, the integration of pocket prediction, structural visualization, residue annotation, and tunnel geometry provides substantial evidence that pocket ID 2 serves as the functional active-site tunnel of the TDC protein. The alignment of CAVER and PyMOL predictions, along with the identification of H319 as the crucial L-tryptophan binding

residue, offers supporting evidence for the localization of the catalytic site. The results validate L-tryptophan as the natural ligand and confirm the mechanistic role of the *TDC* gene in the biosynthesis pathway.

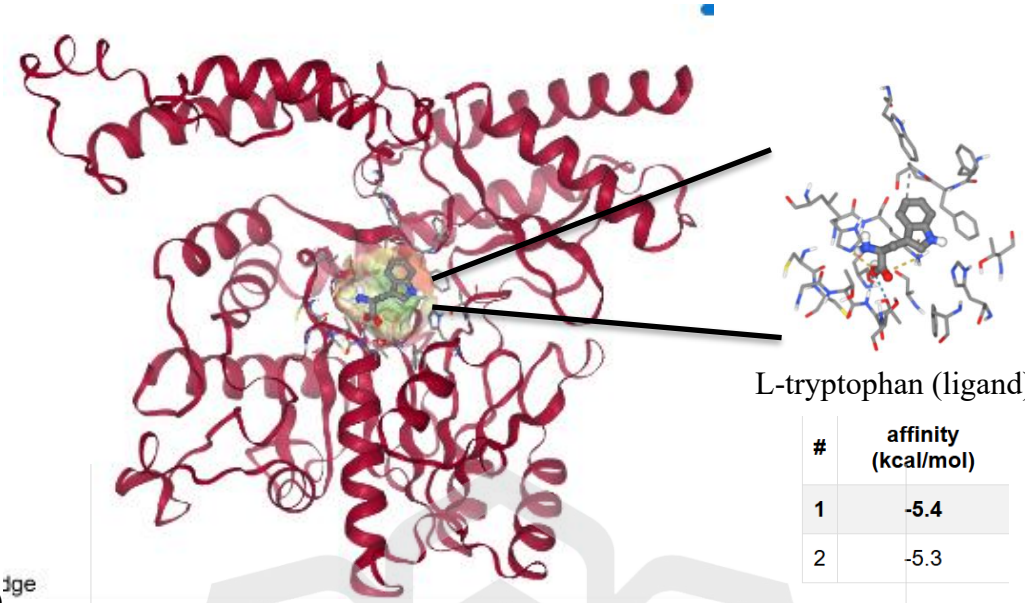
4.3.2 Molecular Docking

Molecular docking is a computational technique used to predict the binding pose and affinity of ligands with target proteins, providing insights into molecular recognition and interaction strength (Kumar & Kumar, 2019). It is extensively utilized in drug discovery and virtual screening, where various docking programs employ algorithms and scoring functions to assess binding geometries and affinities (Muhammed & Aki-Yalcin, 2024; Blanes-Mira et al., 2022). In this study, molecular docking was performed using two web-based platforms, SeamDock and Webina, to evaluate the binding affinity and interactions between L-tryptophan and the TDC protein model generated by AlphaFold.

The first platform, SeamDock, produced binding affinity values of –5.4 and –5.3 kcal/mol, with docking visualization confirming the ligand's position within the protein binding pocket (Figure 4.14a). In contrast, Webina generated stronger interactions, with the optimal docking pose (Mode 1) exhibiting a binding affinity of –9.901 kcal/mol (Figure 4.14b). The remaining docked poses ranged from –8.883 to –7.773 kcal/mol, which are provided in Appendix G. Binding affinity values indicate the strength and stability of a protein–ligand interaction, where more negative scores correspond to stronger binding (Halyal et al., 2025).

Docking visualization revealed that L-tryptophan was oriented within the predicted binding site in close proximity to the catalytic residues (Figure 4.14b). This observation is further supported by the tunnel analysis performed using CAVER (Section 4.3.1), which

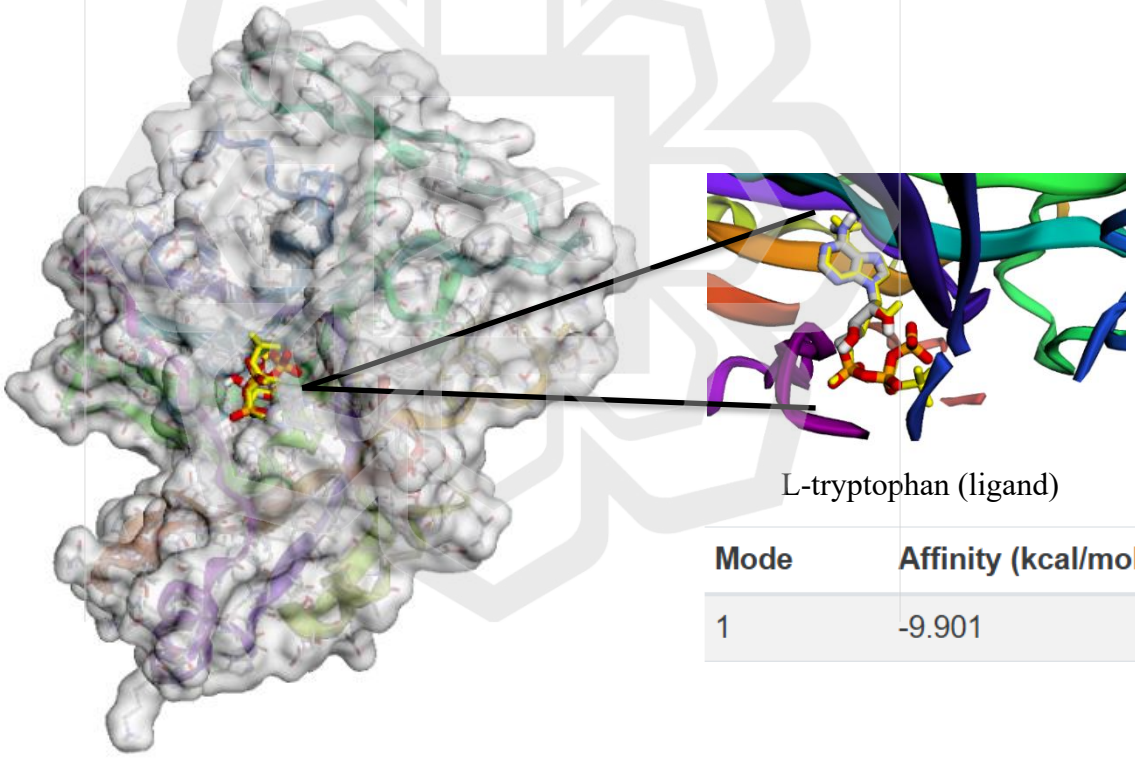
identified a feasible access pathway for L-tryptophan to reach the active site. The consistency between docking and tunnel predictions provides strong evidence that TDC specifically accommodates L-tryptophan as its natural substrate, thereby reinforcing the functional reliability of the modelled TDC structure.



L-tryptophan (ligand)

#	affinity (kcal/mol)
1	-5.4
2	-5.3

A)



L-tryptophan (ligand)

Mode	Affinity (kcal/mol)
1	-9.901

B)

Figure 4.14 Protein–ligand interaction of the TDC protein with L-tryptophan visualized using (a) SeamDock (binding affinity -5.4 and -5.3 kcal/mol) and (b) Webina (best binding affinity -9.901 kcal/mol, mode 1).

4.3.3 Energy Minimization using YASARA

Energy minimization was performed on the docked complex of the TDC protein structure from the AlphaFold model with the tryptophan ligand to refine interactions and eliminate unfavourable steric interactions that remained after docking. As shown in Figure 4.15, the initial unminimized structure exhibited a high energy value ($+1.106 \times 10^{10}$ kcal/mol), indicating substantial structural strain and instability within the complex. After minimization in YASARA, the energy decreased significantly to -6.688×10^{4} kcal/mol. This significant reduction indicates that the refinement process successfully optimized the geometry, resulting in a more stable and energetically favourable conformation. Furthermore, this reduction demonstrates structural relaxation and stabilization of the enzyme–substrate complex, supporting the reliability of the docking result. Steric clashes arise when atoms are positioned in close proximity, resulting in significant repulsive forces. These interactions frequently lead to high-energy, unrealistic conformations in docked complexes. Energy minimization addresses these clashes by modifying atomic positions to appropriate distances, which reduces strain and stabilizes the complex. Briceño-Vargas et al. (2022) and Meitei & Heßelmann (2017) reported analogous findings, emphasizing the influence of steric repulsion on molecular stability.

Additionally, Ramachandran et al. (2011) illustrated that minimization tools can effectively mitigate steric clashes in protein structures. Unlike most studies that perform energy minimization prior to docking, this study applied post-docking minimization to stabilize the final binding poses, reduce steric clashes, and enhance the reliability of predicted interactions. A similar post-docking energy minimization approach has been reported to enhance docking accuracy and reliability by refining docked complexes to achieve geometries closer to experimental structures (Lindström et al., 2011). The substantial reduction in energy indicates that the docked protein–ligand complex attained a stable and refined conformation, thereby enhancing the reliability of the docking results.

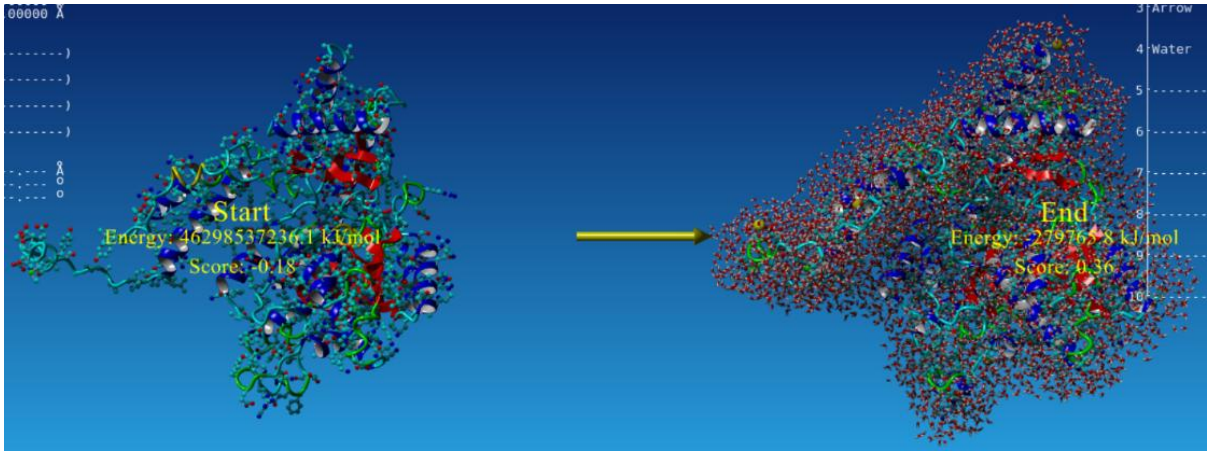


Figure 4.15 Energy minimization using YASARA

In short, the computational analyses successfully elucidated the *in silico* structural and functional characteristics of the TDC enzyme in *M. speciosa*. Sequence alignment and domain prediction confirmed that the gene encodes a pyridoxal phosphate (PLP)-dependent decarboxylase, consistent with its role in alkaloid biosynthesis. Phylogenetic analysis demonstrated close evolutionary relationships with other plant TDCs, indicating strong conservation across species. Physicochemical characterization revealed that the TDC protein is moderately stable, hydrophilic, and predominantly α-helical in structure. The AlphaFold-predicted tertiary model exhibited the highest structural quality based on ERRAT and Ramachandran analyses and was therefore selected for molecular docking. Docking and subsequent energy minimization confirmed stable and favourable interactions between TDC and L-tryptophan, highlighting key residues involved in substrate binding and catalysis. Collectively, these computational findings provide a comprehensive structural framework for understanding the enzymatic mechanism of TDC and lay the foundation for future molecular engineering aimed at enhancing alkaloid biosynthesis in *M. speciosa*.

CHAPTER FIVE

CONCLUSION

5.1 CONCLUSION

This research focused on the functional role and genetic organization of TDC in M. speciosa, which is a key enzyme for mitragynine biosynthesis, one of the major bioactive alkaloids. By combining molecular and bioinformatics approaches, the study aimed to achieve the three main objectives: (i) to isolate the *TDC* gene responsible for mitragynine biosynthesis in *M. speciosa,* (ii) to perform *in silico* molecular and structural analysis of the isolated *TDC* gene, encompassing 3D protein structure construction, and (iii) to determine and validate the stability of binding interactions between the TDC protein and tryptophan through molecular docking analysis. The first objective was successfully achieved through isolation of the *TDC* gene using the CTAB method and PCR amplification, which yielded >700 ng/µl DNA and produced clear PCR bands. Additionally, sequence analysis revealed a 506–amino acid protein with identifiable functional properties. This analysis identified the Pyridoxal-Dependent Decarboxylase domain and classified TDC as a polar protein, supporting its role in alkaloid biosynthesis.

For the second objective, this research successfully modelled the predicted 3D structure using bioinformatics tools and revealed that AlphaFold was the most reliable model with the highest structural validation scores through ERRAT and Ramachandran plot analyses. The resulting model provides a high-quality structural framework for further analysis. The third objective revealed a measurement for the ligand tunnel of 4.7 Å in length with a bottleneck radius of 2.25 Å, indicating favourable conditions for substrate transport. Webina showed a strong binding interaction with −9.901 kcal/mol compared to −5.4 kcal/mol for SeamDock. Lastly, YASARA energy minimization improved the complex

stability by reducing the total potential energy from 1.106×10^{10} kcal/mol to -6.688×10^{4} kcal/mol.

Overall, the study successfully integrated molecular and structural characterization of TDC in *M. speciosa*. The results are consistent with the enzyme's role in catalyzing the conversion of tryptophan into tryptamine and show that certain structural characteristics and residues are necessary for its action. These findings enhance current understanding of mitragynine biosynthesis and provide a foundation for future protein engineering and pathway modification efforts.

5.2 FUTURE WORK

These findings have provided relevant data and understanding of the structure and function of the TDC enzyme in *M. speciosa*, thus providing a foundation for future research. First and foremost, the wet lab is needed to obtain the exact gene sequence amplified from *M. speciosa* gDNA for data-driven computational analyses, ensuring the accuracy and relevance of the data. Future experimental approaches such as site-directed mutagenesis, enzyme assays, or heterologous expression should be employed to confirm the predicted protein–ligand interactions and docking results. These validations would make the *in silico* results more reliable and help to understand the mechanism of the TDC pathway more deeply. Nevertheless, despite its inherent limitations, *in-silico* structural methods such as 3D structure prediction and molecular docking provide speedy and critical preliminary insights to lay the foundation of enzyme function prior to the more time and cost-consuming validations via experimental methods.

In addition, the protein-ligand interactions should be further investigated to clarify substrate specificity, binding dynamics, and the role of conserved residues in catalysis,

which may contribute to the possibilities of modifying TDC activity to enhance tryptamine production. Furthermore, future studies should also investigate the effects of *TDC* gene expression under various physiological and stress conditions, on enzyme-ligand interactions that may provide important insights into the regulation of TDC in plants and its role in the overall alkaloid biosynthesis.

In conclusion, the findings of this study provide avenues for the application of TDC in synthetic biology and metabolic engineering. Altering or reconstructing the mitragynine biosynthesis pathway in microbial or plant-based systems could enable the safe and controlled production of mitragynine and related alkaloids. Such approaches could overcome issues related to plant toxicity and legal limitations, creating opportunities for new drug development.

REFERENCES

Abu-Zaitoon, Y. (2014). Phylogenetic analysis of putative genes involved in the tryptophan-dependent pathway of auxin biosynthesis in rice. *Applied Biochemistry and Biotechnology, 172*(5), 2480-2495. https://doi.org/10.1007/s12010-013-0710-4

Adkins, J. E., Boyer, E. W., & McCurdy, C. R. (2011). *Mitragyna speciosa*, a psychoactive tree from Southeast Asia with opioid activity. *Current Topics in Medicinal Chemistry*, *11*(9), 1165–1175. https://doi.org/10.2174/156802611795371305

Ahmad, I., Prabowo, W. C., Arifuddin, M., Fadraersada, J., Indriyanti, N., Herman, H., Purwoko, R. Y., Nainu, F., Rahmadi, A., Paramita, S., Kuncoro, H., Mita, N., Narsa, A. C., Prasetya, F., Ibrahim, A., Rijai, L., Alam, G., Mun'im, A., & Dej-adisai, S. (2022). *Mitragyna* species as pharmacological agents: from abuse to promising pharmaceutical products. *Life, 12*(2), 193. https://doi.org/10.3390/life12020193

Alghalith, A., Nguyen, H., & Tennant, R. (2024). The complexities of kratom: insights on an increasingly frequent clinical encounter. *Cureus*. https://doi.org/10.7759/cureus.59650

Allen, P., Nguyen, N., Humphrey, N. D., Mao, J., Chavez-Bonilla, D., & Sorin, E. J. (2024). A hands-on collaboration-ready single- or interdisciplinary computational exercise in molecular recognition and drug design. *Education Sciences, 14*(2), 139. https://doi.org/10.3390/educsci14020139

Ameline, A., Gheddar, L., Arbouche, N., Blanchot, A., Raul, J.-S., & Kintz, P. (2024). Testing for Kratom alkaloids in fingernail clippings – not only mitragynine. *Journal of Pharmaceutical and Biomedical Analysis, 243*, 116078. https://doi.org/10.1016/j.jpba.2024.116078

Ameline, A., Raul, J.-S., & Kintz, P. (2023). Identification of mitragynine by UPLC-MS/MS in the hair and nails of two kratom users. *Toxicologie Analytique et Clinique, 35*(4), 362–367. https://doi.org/10.1016/j.toxac.2023.07.005

Arafa, A., Ragab, A., Ibrahim, A., Abdel-Mageed, W., & Emam, M. (2019). Cloning and overexpression of strictosidine β-d-glucosidase gene short sequence from *Catharanthus Roseus* in *Escherichia Coli*. *Advanced Pharmaceutical Bulletin, 9*(4), 655-661. https://doi.org/10.15171/apb.2019.076

Araújo-Júnior, J. X. d., Antheaume, C., Trindade, R. C. P., Schmitt, M., Bourguignon, J., & Santana, A. E. G. (2007). Isolation and characterisation of the monoterpenoid indole alkaloids of *Aspidosperma pyrifolium*. *Phytochemistry Reviews, 6*(1), 183-188. https://doi.org/10.1007/s11101-006-9044-y

Armano, G., Mancosu, G., Milanesi, L., Orro, A., Saba, M., & Vargiu, E. (2005). A hybrid genetic-neural system for predicting protein secondary structure. *BMC Bioinformatics*, *6*(S4). https://doi.org/10.1186/1471-2105-6-S4-S3

Atilgan, E., & Hu, J. (2010). Efficient protein-ligand docking using sustainable evolutionary algorithms. *In 2010, the 10th International Conference on Hybrid Intelligent Systems* (pp. 113–118). IEEE. https://doi.org/10.1109/his.2010.5600082

Bahieldin, A., Atef, A., Edris, S., Gadalla, N. O., Al-matary, M., Al-Kordy, M. A., Ramadan, A. M., Bafeel, S., Alharbi, M. G., Al-Quwaie, D. A. H., Sabir, J. S. M., Al-Zahrani, H. S., Nasr, M. E., & El-Domyati, F. M. (2018). Stepwise response of MeJA-induced genes and pathways in leaves of *Catharanthus roseus*. *Comptes Rendus. Biologies, 341*(9–10), 411–420. https://doi.org/10.1016/j.crvi.2018.10.001

Barleben, L., Panjikar, S., Ruppert, M., Koepke, J., & Stöckigt, J. (2007). Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *The Plant Cell, 19*(9), 2886-2897. https://doi.org/10.1105/tpc.106.045682

Begum, T., Nasrin, M., Ahsan, M., & Dash, P. R. (2025). A review on *Mitragyna speciosa* (Rubiaceae) as a prominent medicinal plant based on ethnobotany, phytochemistry and pharmacological activities. *Natural Product Research, 39*(6), 1636–1652. https://doi.org/10.1080/14786419.2024.2371564

Bhowal, B., Bhattacharjee, A., Goswami, K., Sanan-Mishra, N., Singla-Pareek, S. L., Kaur, C., & Sopory, S. (2021). Serotonin and melatonin biosynthesis in plants: Genome-wide

identification of the genes and their expression reveal a conserved role in stress and development. *International Journal of Molecular Sciences, 22*(20), 11034. https://doi.org/10.3390/ijms222011034

Bisello, G., Ribeiro, R. P., Perduca, M., Belviso, B. D., Polverino De' Laureto, P., Giorgetti, A., Caliandro, R., & Bertoldi, M. (2023). Human aromatic amino acid decarboxylase is an asymmetric and flexible enzyme: Implication in aromatic amino acid decarboxylase deficiency. *Protein Science: A Publication of the Protein Society*, *32*(8), e4732. https://doi.org/10.1002/pro.4732

Blanes-Mira, C., Fernández-Aguado, P., de Andrés-López, J., Fernández-Carvajal, A., Ferrer-Montiel, A., & Fernández-Ballester, G. (2022). Comprehensive survey of consensus docking for high-throughput virtual screening. *Molecules, 28*(1), 175. https://doi.org/10.3390/molecules28010175

Bowe, A., & Kerr, P. L. (2024). Endogenous opioid activity as the mechanism of action for *Mitragyna speciosa* (kratom): The current state of the evidence. *In Advances in Neurobiology* (pp. 287–313). Springer International Publishing. https://doi.org/10.1007/978-3-031-45493-6_15

Briceño-Vargas, F. M., Mirón-López, G., & Quijano-Quiñones, R. F. (2022). NCI study of the influence of steric hindrance on the conformational isomers of propane, butane, and ethane derivatives. *ChemistrySelect*, *7*(22). https://doi.org/10.1002/slct.202201024

Bremei, B. (2009). A review of molecular phylogenetic studies of Rubiaceae1. *Annals of the Missouri Botanical Garden, 96*(1), 4–26. https://doi.org/10.3417/2006197

Brose, J., Lau, K. H., Dang, T. T. T., Hamilton, J. P., Martins, L. do V., Hamberger, B., Hamberger, B., Jiang, J., O'Connor, S. E., & Buell, C. R. (2021). The *Mitragyna speciosa* (kratom) genome: a resource for data-mining potent pharmaceuticals that impact human health. *G3 Genes|Genomes|Genetics, 11*(4). https://doi.org/10.1093/g3journal/jkab058

Brown, P. N., Lund, J. A., & Murch, S. J. (2017). A botanical, phytochemical and ethnomedicinal review of the genus *Mitragyna* korth: Implications for products sold

as kratom. *Journal of Ethnopharmacology*, *202*, 302–325. https://doi.org/10.1016/J.JEP.2017.03.020

Buchwald, T., Niciejewski, K., Kozielski, M., Szybowicz, M., Siatkowski, M., & Krauss, H. (2012). Identifying compositional and structural changes in spongy and subchondral bone from the hip joints of patients with osteoarthritis using Raman spectroscopy. *Journal of Biomedical Optics, 17*(1), 017007. https://doi.org/10.1117/1.jbo.17.1.017007

Burge, S., Kelly, E., Lonsdale, D., Mutowo-Muellenet, P., McAnulla, C., Mitchell, A., Sangrador-Vegas, A., Yong, S. Y., Mulder, N., & Hunter, S. (2012). Manual GO annotation of predictive protein signatures: the InterPro approach to GO curation. *Database: The Journal of Biological Databases and Curation*, *2012*, bar068. https://doi.org/10.1093/database/bar068

Burlat, V., Oudin, A., Courtois, M., Rideau, M., & St-Pierre, B. (2004). Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus Roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *The Plant Journal, 38*(1), 131-141. https://doi.org/10.1111/j.1365-313x.2004.02030.x

Byeon, Y., Park, S., Lee, H., Kim, Y., & Back, K. (2014). Elevated production of melatonin in transgenic rice seeds expressing rice tryptophan decarboxylase. *Journal of Pineal Research, 56*(3), 275-282. https://doi.org/10.1111/jpi.12120

Can, T. (2013). Introduction to Bioinformatics. In *Methods in Molecular Biology* (pp. 51–71). Humana Press. https://doi.org/10.1007/978-1-62703-748-8_4

Cammisa, M., Correra, A., Andreotti, G., & Cubellis, M. V. (2013). Identification and analysis of conserved pockets on protein surfaces. *BMC Bioinformatics, 14*(S7). https://doi.org/10.1186/1471-2105-14-s7-s9

Cázares-Flores, P., Levac, D., & De Luca, V. (2016). *Rauvolfia serpentina* N-methyltransferases involved in ajmaline and Nβ-methylajmaline biosynthesis belong to a gene family derived from γ-tocopherol C-methyltransferase. *The Plant Journal, 87*(4), 335–342. https://doi.org/10.1111/tpj.13186

Charoonratana, T., Wungsintaweekul, J., Keawpradub, N., & Verpoorte, R. (2013a). Molecular cloning and expression of tryptophan decarboxylase from *Mitragyna speciosa*. *Acta Physiologiae Plantarum, 35*(8), 2611–2621. https://doi.org/10.1007/s11738-013-1296-8

Charoonratana, T., Wungsintaweekul, J., Pathompak, P., Georgiev, M. I., Choi, Y. H., & Verpoorte, R. (2013b). Limitation of mitragynine biosynthesis in *Mitragyna speciosa* (Roxb.) Korth. through tryptamine availability. *Zeitschrift Für Naturforschung C, 68*(9–10), 394–405. https://doi.org/10.1515/znc-2013-9-1007

Chávez-Medina, J. A., Ibarra-Gámez, J. C., Flores-Zamora, G. L., García-Negroe, C. B., Álvarez-Ruiz, P., Álvarez, S. P.-, Castro-Espinoza, L., Gutiérrez-Coronado, M. A., Garcia-Gutierrez, C., & Martínez-Carrillo, J. L. (2019). Evaluación de cinco métodos de extracción de ADN e identificación de biotipos de *Spodoptera frugiperda* (J.E. Smith)1 mediante PCR-RFLP. *Southwestern Entomologist, 44*(4), 935. https://doi.org/10.3958/059.044.0405

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research*, *31*(13), 3497–3500. https://doi.org/10.1093/nar/gkg500

Cinosi, E., Martinotti, G., Simonato, P., Singh, D., Demetrovics, Z., Roman-Urrestarazu, A., Bersani, F. S., Vicknasingam, B., Piazzon, G., Li, J. H., Yu, W. J., Kapitány-Fövény, M., Farkas, J., Di Giannantonio, M., & Corazza, O. (2015). Following "the roots" of kratom (*Mitragyna speciosa*): The evolution of an enhancer from a traditional use to increase work and productivity in Southeast Asia to a recreational psychoactive drug in Western countries. *BioMed Research International*, *2015*, 968786. https://doi.org/10.1155/2015/968786

Chen, M., Zhang, M., Liang, Z., & He, Q. (2022). Characterization and comparative analysis of chloroplast genomes in five *Uncaria* species endemic to China. *International Journal of Molecular Sciences*, *23*(19), 11617. https://doi.org/10.3390/ijms231911617

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science, 2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Das, R. K., & Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(33), 13392–13397. https://doi.org/10.1073/pnas.1304749110

De Luca, V., Fernandez, J. A., Campbell, D., & Kurz, W. G. (1988). Developmental regulation of enzymes of indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiology*, *86*(2), 447–450. https://doi.org/10.1104/pp.86.2.447

De Silva, S., Lee, S.-S., Dugan, M. B., & Anderson, J. L. (2025). Recent advancements and emerging techniques in nucleic acid isolation, amplification, and detection from diverse complex matrices of human interest. *Trends in Analytical Chemistry, 185*, 118172. https://doi.org/10.1016/j.trac.2025.118172

Di Muzio, E., Toti, D., & Polticelli, F. (2017). DockingApp: a user-friendly interface for facilitated docking simulations with AutoDock Vina. *Journal of Computer-Aided Molecular Design*, *31*(2), 213–218. https://doi.org/10.1007/s10822-016-0006-1

Doig, A. J., Andrew, C. D., Cochran, D. A., Hughes, E., Penel, S., Sun, J. K., Stapley, B. J., Clarke, D. T., & Jones, G. R. (2001). Structure, stability and folding of the alpha-helix. *Biochemical Society symposium*, (68), 95–110. https://doi.org/10.1042/bss0680095

Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F., & Douzery, E. J. (2003). Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biology and Evolution*, *20*(2), 248–254. https://doi.org/10.1093/molbev/msg042

Doyle, J. J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus, 12*(1), 13–15.

Dubouzet, J. G., Matsuda, F., Ishihara, A., Miyagawa, H., & Wakasa, K. (2013). Production of indole alkaloids by metabolic engineering of the tryptophan pathway in rice. *Plant Biotechnology Journal, 11*(9), 1103-1111. https://doi.org/10.1111/pbi.12105

Dudley, Q. M., Jo, S., Guerrero, D. A. S., Chhetry, M., Smedley, M. A., Harwood, W. A., Sherden, N. H., O'Connor, S. E., Caputi, L., & Patron, N. J. (2021). Reconstitution of monoterpene indole alkaloid biosynthesis in genome engineered *Nicotiana benthamiana*. *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/2021.08.12.456143

Dutta, B., Banerjee, A., Chakraborty, P., & Bandopadhyay, R. (2018). *In silico* studies on bacterial xylanase enzyme: Structural and functional insight. *Journal, Genetic Engineering & Biotechnology*, *16*(2), 749–756. https://doi.org/10.1016/j.jgeb.2018.05.003

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edwards, R. J. (2019). Phylogenetic tree rooting. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 727–735). Elsevier.

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology, 277*, 396–404. https://doi.org/10.1016/S0076-6879(97)77022-8

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research*, *8*(3), 175–185. https://doi.org/10.1101/gr.8.3.175

Farah Idayu, N., Taufik Hidayat, M., Moklas, M. A. M., Sharida, F., Nurul Raudzah, A. R., Shamima, A. R., & Apryani, E. (2011). Antidepressant-like effect of mitragynine

isolated from *Mitragyna speciosa* Korth in mice model of depression. *Phytomedicine, 18*(5), 402–407. https://doi.org/10.1016/j.phymed.2010.08.011

Facchini P. J. (2001). Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annual Review of Plant Physiology and Plant Molecular Biology*, *52*, 29–66. https://doi.org/10.1146/annurev.arplant.52.1.29

Facchini, P. J., & De Luca, V. (2008). Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. *The Plant Journal: for Cell and Molecular Biology*, *54*(4), 763–784. https://doi.org/10.1111/j.1365-313x.2008.03438.x

Facchini, P. J., Huber-Allanach, K. L., & Tari, L. W. (2000). Plant aromatic L-amino acid decarboxylases: evolution, biochemistry, regulation, and metabolic engineering applications. *Phytochemistry*, *54*(2), 121–138. https://doi.org/10.1016/s0031-9422(00)00050-9

Fluyau, D., & Revadigar, N. (2017). Biochemical benefits, diagnosis, and clinical risks evaluation of kratom. *Frontiers in Psychiatry*, *8*, 62. https://doi.org/10.3389/fpsyt.2017.00062

Fouad, A., Hegazy, A., Azab, E., Khojah, E., & Kapiel, T. (2021). Boosting of antioxidants and alkaloids in *Catharanthus roseus* suspension cultures using silver nanoparticles with expression of *CrMPK3* and *STR* genes. *Plants (Basel, Switzerland)*, *10*(10), 2202. https://doi.org/10.3390/plants10102202

Garba, S. A., Shaari, K., Abdul Manap, M. R., Lee, S. Y., Abdulazeez, I., & Mohd Faudzi, S. M. (2024). Quantitative analysis of selected alkaloids of *Mitragyna speciosa* using 1H quantitative nuclear magnetic resonance spectroscopy. *Magnetic Resonance in Chemistry*, *62*(11), 803–813. https://doi.org/10.1002/mrc.5477

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research, 31*(13), 3784–3788. https://doi.org/10.1093/nar/gkg563

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press.

Gerasimenko, I., Sheludko, Y., Ma, X., & Stöckigt, J. (2002). Heterologous expression of a Rauvolfia cDNA encoding strictosidine glucosidase, a biosynthetic key to over 2000 monoterpenoid indole alkaloids. *European Journal of Biochemistry*, *269*(8), 2204–2213. https://doi.org/10.1046/j.1432-1033.2002.02878.x

Ghazalli, M. N., Md Sah, M. S., Mat, M., Awang, K., Jaafar, M. A., Mirad, R., Zaini, A. Z., Nordin, A. R. M., Rusli, N. M., Mohamad, S. S., & Dalee, A. S. M. (2021). Leaf anatomy and micromorphology characteristics of ketum [*Mitragyna speciosa* (Korth.) Havil.] (Rubiaceae). *Tropical Life Sciences Research*, *32*(1), 107–117. https://doi.org/10.21315/tlsr2021.32.1.7

Gill, K., Negi, S., Kumar, P., & Irfan, M. (2025). Improved genomic DNA extraction from citrus species using a modified CTAB method. *Molecular Biology Reports*, *52*(1), 638. https://doi.org/10.1007/s11033-025-10763-1

Graham, K., Cantu, C., & Houston, R. (2024). Sequence variation of commercially available kratom products at universal DNA barcode regions. *Journal of Forensic Sciences, 69*(4), 1421–1428. https://doi.org/10.1111/1556-4029.15547

Grundmann, O., Hendrickson, R. G., & Greenberg, M. I. (2023). Kratom: History, pharmacology, current user trends, adverse health effects and potential benefits. *Disease-a-month, 69*(6), 101442. https://doi.org/10.1016/j.disamonth.2022.101442

Guirimand, G., Courdavault, V., Lanoue, A., Mahroug, S., Guihur, A., Blanc, N., Giglioli-Guivarc'h, N., St-Pierre, B., & Burlat, V. (2010). Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"?. *BMC Plant Biology*, *10*, 182. https://doi.org/10.1186/1471-2229-10-182

Guirimand, G., Guihur, A., Ginis, O., Poutrain, P., Héricourt, F., Oudin, A., Lanoue, A., St-Pierre, B., Burlat, V., & Courdavault, V. (2011). The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-

tonoplast translocations of intermediate metabolites. *The FEBS Journal, 278*(5), 749–763. https://doi.org/10.1111/j.1742-4658.2010.07994.x

Guo, X. M., Wang, Z. F., Zhang, Y., & Wang, R. J. (2021). Chromosomal-level assembly of the *Leptodermis oblonga* (Rubiaceae) genome and its phylogenetic implications. *Genomics*, *113*(5), 3072–3082. https://doi.org/10.1016/j.ygeno.2021.07.012

Gupta, S. K., Rai, A. K., Kanwar, S. S., & Sharma, T. R. (2012). Comparative analysis of zinc finger proteins involved in plant disease resistance. *PloS one*, *7*(8), e42578. https://doi.org/10.1371/journal.pone.0042578

Hilgarth, R. S., & Lanigan, T. M. (2020). Optimization of overlap extension PCR for efficient transgene construction. *MethodsX*, *7*, 100759. https://doi.org/10.1016/j.mex.2019.12.001

Halyal, P., Sujatha, C., Desai, P., & Ronad, P. (2025). Protein-ligand binding affinity prediction. In *Lecture Notes in Networks and Systems* (pp. 461–476). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-2703-5_31

Handayani, W., Imawan, C., Umar, A., Yasman, Yunilawati, R., & Djuhana, D. (2021). Structural, optical, and potential broad-spectrum antibacterial properties of CuO–Ag nanoparticles biosynthesized using the extract of *Diospyros discolor* Willd. *Advances in Natural Sciences: Nanoscience and Nanotechnology*, *12*(4), 045007. https://doi.org/10.1088/2043-6262/ac458a

Halim, S. A., Low, J. H., Chee, Y. C., & Alias, M. R. (2021). Seizures among young adults consuming kratom beverages in Malaysia: A case series. *Epilepsy & Behavior*, *121*(Pt A), 108057. https://doi.org/10.1016/j.yebeh.2021.108057

Halpenny G. M. (2017). *Mitragyna speciosa*: Balancing potential medical benefits and abuse. *ACS Medicinal Chemistry Letters*, *8*(9), 897–899. https://doi.org/10.1021/acsmedchemlett.7b00298

Haque, I., Bandopadhy, R., & Mukhopadhy, K. (2008). An optimised protocol for fast genomic DNA isolation from high secondary metabolites and gum containing

plants. *Asian Journal of Plant Sciences*, *7*(3), 304–308. https://doi.org/10.3923/ajps.2008.304.308

Harizal, S. N., Mansor, S. M., Hasnan, J., Tharakan, J. K. J., & Abdullah, J. (2010). Acute toxicity study of the standardized methanolic extract of *Mitragyna speciosa* Korth in rodent. *Journal of Ethnopharmacology*, *131*(2), 404–409. https://doi.org/10.1016/j.jep.2010.07.013

Hassan, Z., Muzaimi, M., Navaratnam, V., Yusoff, N. H. M., Suhaimi, F. W., Vadivelu, R., Vicknasingam, B. K., Amato, D., von Hörsten, S., Ismail, N. I. W., Jayabalan, N., Hazim, A. I., Mansor, S. M., & Müller, C. P. (2013). From Kratom to mitragynine and its derivatives: Physiological and behavioural effects related to use, abuse, and addiction. In *Neuroscience and Biobehavioral Reviews* (Vol. 37, Issue 2, pp. 138–151). https://doi.org/10.1016/j.neubiorev.2012.11.012

Hassan, Z., Singh, D., Suhaimi, F. W., Chear, N. J.-Y., Harun, N., See, C. P., Kaur, G., Mat, N. H., Bakar, S. N. S., Yusof, N. S. M., Kasinather, V. B., Chawarski, M. C., Murugaiyah, V., & Ramanathan, S. (2023). Evaluation of toxicity profile of kratom (*Mitragyna speciosa* Korth) decoction in rats. *Regulatory Toxicology and Pharmacology, 143,* 105466. https://doi.org/10.1016/j.yrtph.2023.105466

Hazrati, S., Mousavi, Z., & Nicola, S. (2024). Harvest time optimization for medicinal and aromatic plant secondary metabolites. *Plant Physiology and Biochemistry*, *212*(108735), 108735. https://doi.org/10.1016/j.plaphy.2024.108735

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hedhili, S., Courdavault, V., Giglioli-Guivarc'h, N., & Gantet, P. (2007). Regulation of the terpene moiety biosynthesis of *Catharanthus roseus* terpene indole alkaloids. *Phytochemistry Reviews, 6*(2-3), 341-351. https://doi.org/10.1007/s11101-006-9021-5

Hoffmann, F., Vancea, I., Kamat, S. G., & Strodel, B. (2014). Protein structure prediction: Assembly of secondary structure elements by basin-hopping. *ChemPhysChem, 15*(15), 3378–3390. https://doi.org/10.1002/cphc.201402247

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, *292*(2), 195–202. https://doi.org/10.1006/jmbi.1999.3091

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences: CABIOS*, *8*(3), 275–282. https://doi.org/10.1093/bioinformatics/8.3.275

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics, 30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kang, K., Kang, S., Lee, K., Park, M., & Back, K. (2008). Enzymatic features of serotonin biosynthetic enzymes and serotonin biosynthesis in plants. *Plant Signaling & Behavior, 3*(6), 389-390. https://doi.org/10.4161/psb.3.6.5401

Kang, K., Kim, Y., Park, S., & Back, K. (2009). Senescence-induced serotonin biosynthesis and its role in delaying senescence in rice leaves. *Plant Physiology, 150*(3), 1380–1393. https://doi.org/10.1104/pp.109.138552

Katoh, K. & Standley, D. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution, 30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kaur, A., & Kaur, M. (2024). Construction of phylogenetic tree using UPGMA method. In *Proceedings of the 2024 7th International Conference on Contemporary Computing and Informatics (IC3I)* (pp. 242–246). IEEE. https://doi.org/10.1109/IC3I61595.2024.10828834

Kaur, A., Pati, P., Pati, A., & Nagpal, A. (2020). Physico-chemical characterization and topological analysis of pathogenesis-related proteins from *Arabidopsis thaliana* and *Oryza sativa* using *in-silico* approaches. *Plos One, 15*(9), e0239836. https://doi.org/10.1371/journal.pone.0239836

Khalid, K., Ku Md Saad, S., Soelar, S. A., Mohamed Yusof, Z., & Warijo, O. (2021). Exploring adolescents' practice and perspective on the use and misuse of kratom in northwest Malaysia. *Journal of Ethnicity in Substance Abuse*, *22*(1), 121–132. https://doi.org/10.1080/15332640.2021.1906816

King, M. D., Phillips, P., Turner, M. W., Katz, M., Lew, S., Bradburn, S., Andersen, T., & McDougal, O. M. (2015). Computational exploration of a protein receptor binding space with student proposed peptide ligands. *Biochemistry and Molecular Biology Education, 44*(1), 63–67. https://doi.org/10.1002/bmb.20925

Kitajima, M., & Takayama, H. (2014). The botany of *Mitragyna speciosa* (Korth.) Havil. and related species. In *Kratom and Other Mitragynines* (pp. 74–93). CRC Press.

Kochnev, Y., Hellemann, E., Cassidy, K. C., & Durrant, J. D. (2020). Webina: an open-source library and web app that runs AutoDock Vina entirely in the web browser. *Bioinformatics, 36*(16), 4513–4515. https://doi.org/10.1093/bioinformatics/btaa579

Kozlikova, B., Sebestova, E., Sustr, V., Brezovsky, J., Strnad, O., Daniel, L., Bednar, D., Pavelka, A., Manak, M., Bezdeka, M., Benes, P., Kotry, M., Gora, A., Damborsky, J., & Sochor, J. (2014). CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics (Oxford, England)*, *30*(18), 2684–2685. https://doi.org/10.1093/bioinformatics/btu364

Kozłowski, Ł. (2017). Proteome-pi: proteome isoelectric point database. *Nucleic Acids Research, 45*(D1), D1112-D1116. https://doi.org/10.1093/nar/gkw978

Krieger, E., & Vriend, G. (2014). YASARA View—Molecular graphics for all devices—from smartphones to workstations. *Bioinformatics, 30*(20), 2981–2982. https://doi.org/10.1093/bioinformatics/btu426

Kruegel, A. C., Gassaway, M. M., Kapoor, A., Váradi, A., Majumdar, S., Filizola, M., Javitch, J. A., & Sames, D. (2016). Synthetic and receptor signaling explorations of the *Mitragyna* alkaloids: Mitragynine as an atypical molecular framework for opioid receptor modulators. *Journal of the American Chemical Society, 138*(21), 6754–6764. https://doi.org/10.1021/jacs.6b00360

Kumar, P., & Kumar, A. (2024). 4 State-of-the-art modeling techniques in performing docking algorithms and scoring. In *Computational Drug Discovery* (pp. 65–80). De Gruyter. https://doi.org/10.1515/9783111207117-004

Kumar, S., & Kumar, S. (2019). Molecular Docking: A structure-based approach for drug repurposing. In *In Silico Drug Design* (pp. 161–189). Elsevier. https://doi.org/10.1016/b978-0-12-816125-8.00006-7

Kurgan, L., Cios, K., & Chen, K. (2008). SCPRED: accurate prediction of protein structural class for sequences of twilight-zone similarity with predicting sequences. *BMC Bioinformatics*, *9*(1), 226. https://doi.org/10.1186/1471-2105-9-226

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography, 26*(2), 283–291. https://doi.org/10.1107/S0021889892009944

Lehmann, T., Hoffmann, M., Hentrich, M., & Pollmann, S. (2010). Indole-3-acetamide-dependent auxin biosynthesis: a widely distributed way of indole-3-acetic acid production?. *European Journal of Cell Biology*, *89*(12), 895–905. https://doi.org/10.1016/j.ejcb.2010.06.021

Leksungnoen, N., Andriyas, T., Ngernsaengsaruay, C., Uthairatsamee, S., Racharak, P., Sonjaroon, W., Kjelgren, R., Pearson, B. J., McCurdy, C. R., & Sharma, A. (2022). Variations in mitragynine content in the naturally growing Kratom (*Mitragyna speciosa*) population of Thailand. *Frontiers in Plant Science, 13*. https://doi.org/10.3389/fpls.2022.1028547

Leksungnoen, N., Andriyas, T., Ku-Or, Y., Chongdi, S., Tansawat, R., Aramrak, A., Ngernsaengsaruay, C., Uthairatsamee, S., Sonjaroon, W., Thongchot, P., Ardsiri, S., & Pongchaidacha, P. (2025). The effect of light intensity and polyethylene-glycol-induced water stress on the growth, mitragynine accumulation, and total alkaloid content of kratom (*Mitragyna speciosa*). *Horticulturae, 11*(3), 272. https://doi.org/10.3390/horticulturae11030272

Lesiak, A. D., Cody, R. B., Dane, A. J., & Musah, R. A. (2014). Rapid detection by direct analysis in real time-mass spectrometry (DART-MS) of psychoactive plant drugs of abuse: The case of *Mitragyna speciosa* aka "Kratom." *Forensic Science International, 242*, 210–218. https://doi.org/10.1016/j.forsciint.2014.07.005

Li, L., Zheng, M., Long, H., Deng, G., Ishihara, A., Liu, F., Liang, J., Pan, Z., & Yu, M. (2015). Molecular cloning and characterization of two genes encoding tryptophan decarboxylase from *Aegilops variabilis* with resistance to the cereal cyst nematode (*Heterodera avenae*) and root-knot nematode (*Meloidogyne naasi*). *Plant Molecular Biology Reporter, 34*(1), 273–282. https://doi.org/10.1007/s11105-015-0909-3

Li, M., Tadfie, H., Darnell, C. G., & Holland, C. K. (2023). Biochemical investigation of the tryptophan biosynthetic enzyme anthranilate phosphoribosyl transferase in plants. *Journal of Biological Chemistry, 299*(10), 105197. https://doi.org/10.1016/j.jbc.2023.105197

Liang, J., Han, Q., Tan, Y., Ding, H., & Li, J. (2019). Current advances on structure-function relationships of pyridoxal 5'-phosphate-dependent enzymes. *Frontiers in Molecular Biosciences*, *6*, 4. https://doi.org/10.3389/fmolb.2019.00004

Liu, J., Cai, J., Wang, R., & Yang, S. (2016). Transcriptional regulation and transport of terpenoid indole alkaloid in *Catharanthus roseus*: exploration of new research directions. *International Journal of Molecular Sciences, 18*(1), 53. https://doi.org/10.3390/ijms18010053

Liu, Y., Li, D. W., Zhang, Z. H., & Zu, Y. G. (2014). Low Light Affacting Alkaloids Accumulation and Related Biosynthetic Pathway Genes Expression in Leaves of <i>Catharanthus roseus</i> Seedlings. *Applied Mechanics and Materials, 522–524,* 311–315. https://doi.org/10.4028/www.scientific.net/amm.522-524.311

Liu, Y., Zhong, X., Wang, Y., Wang, K., Li, C., Liang, X., Lin, S., Zhou, R., & Xie, C. (2024). Variations in gut microbiome profiles based on different DNA extraction methods: A comparative study. *Advanced Gut & Microbiome Research, 2024*, 1–12. https://doi.org/10.1155/2024/6653560

Liu, Yi, Yang, Q., & Zhao, F. (2021). Synonymous but not silent: The Codon usage code for gene expression and protein folding. *Annual Review of Biochemistry*, *90*(1), 375–401. https://doi.org/10.1146/annurev-biochem-071320-112701

Lindström, A., Edvinsson, L., Johansson, A., Andersson, C. D., Andersson, I. E., Raubacher, F., & Linusson, A. (2011). Postprocessing of docked protein-ligand complexes using implicit solvation models. *Journal of Chemical Information and Modeling*, *51*(2), 267–282. https://doi.org/10.1021/ci100354x

Lu, J., Wei, H., Wu, J., Jamil, M. F. A., Tan, M. L., Adenan, M. I., Wong, P., & Shim, W. (2014). Evaluation of the cardiotoxicity of mitragynine and its analogues using human induced pluripotent stem cell-derived cardiomyocytes. *PLoS ONE, 9*(12), e115648. https://doi.org/10.1371/journal.pone.0115648

Lück, M., Bergs, F., & Jupke, A. (2020). Solvent accessibility limitation by plant matrix compounds in extraction of rutin from *Solanum lycopersicum*. *Separation Science Plus*, *3*(3), 63–71. https://doi.org/10.1002/sscp.201900074

Manda, V., Avula, B., Ali, Z., Khan, I., Walker, L., & Khan, S. (2014). Evaluation of *in vitro* absorption, distribution, metabolism, and excretion (ADME) properties of mitragynine, 7-hydroxymitragynine, and mitraphylline. *Planta Medica, 80*(07), 568–576. https://doi.org/10.1055/s-0034-1368444

Marques, S. M., Borko, S., Vavra, O., Dvorsky, J., Kohout, P., Kabourek, P., Hejtmanek, L., Damborsky, J., & Bednar, D. (2025). Caver Web 2.0: Analysis of tunnels and ligand transport in dynamic ensembles of proteins. *Nucleic Acids Research* 53 (W1), W132–W142.

Matlock, B. (2015). *Assessment of nucleic acid purity.* Technical Bulletin NanoDrop Spectrophotometers, 1–2.

Mayanja, R., Kintu, C., Diabate, O., Soremekun, O., Oluwagbemi, O. O., Wele, M., Kalyesubula, R., Jjingo, D., Chikowore, T., & Fatumo, S. (2022). Molecular dynamic simulation reveals structure differences in APOL1 variants and implication in pathogenesis of chronic kidney disease. *Genes, 13*(8), 1460. https://doi.org/10.3390/genes13081460

Meek, K., Buehl, C., Goff, N., Mikhova, M., Hardwick, S., Blundell, T., Modesti, M., Schmidt, J., & Chaplin, A. (2023). *Unravelling the complexities of DNA-PK activation by structure-based mutagenesis.* Springer Science and Business Media LLC. https://doi.org/10.21203/rs.3.rs-3627471/v1

Meireles, V., Rosado, T., Barroso, M., Soares, S., Gonçalves, J., Luís, Â., Caramelo, D., Simão, A. Y., Fernández, N., Duarte, A. P., & Gallardo, E. (2019). *Mitragyna speciosa*: clinical, toxicological aspects and analysis in biological and non-biological samples. *Medicines, 6*(1), 35. https://doi.org/10.3390/medicines6010035

Meitei, O. R., & Heßelmann, A. (2017). Intramolecular interactions in sterically crowded hydrocarbon molecules. *Journal of Computational Chemistry*, *38*(29), 2500–2508. https://doi.org/10.1002/jcc.24908

Miranda-Ham, M. D. L., Islas-Flores, I., & Vázquez-Flota, A. F. (2007). Accumulation of monoterpenoid indole alkaloids in periwinkle seedlings (*Catharanthus roseus*) as a model for the study of plant–environment interactions. *Biochemistry and Molecular Biology Education, 35*(3), 206-210. https://doi.org/10.1002/bmb.60

Mohammad Farris, I. L. A., & Darshan, S. (2021). Assessment of cardiovascular functioning among regular kratom (*Mitragyna speciosa* Korth) users: A case series. *Frontiers in Pharmacology (Vol. 12)*. Frontiers Media SA. https://doi.org/10.3389/fphar.2021.723567

Mohamad Zuldin, N. N., Said, I. M., Mohd Noor, N., Zainal, Z., Jin Kiat, C., & Ismail, I. (2013). Induction and analysis of the alkaloid mitragynine content of a *Mitragyna speciosa* suspension culture system upon elicitation and precursor feeding. *The Scientific World Journal*, *2013*(1), 209434. https://doi.org/10.1155/2013/209434

Mohanta, T., Mishra, A., Mohanta, Y., & Al-Harrasi, A. (2021). Virtual 2D mapping of the viral proteome reveals host-specific modality distribution of molecular weight and isoelectric point. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-00797-3

Moreno-Arribas, V., & Lonvaud-Funel, A. (2001). Purification and characterization of tyrosine decarboxylase of *Lactobacillus brevis* IOEB 9809 isolated from wine. *FEMS*

Microbiology Letters, 195(1), 103–107. https://doi.org/10.1111/j.1574-6968.2001.tb10505.x

Muhamad, F. N., Ahmad, R. B., Mohd. Asi, S., & Murad, M. N. (2015). Reducing the search space and time complexity of Needleman-Wunsch algorithm (global alignment) and Smith-Waterman algorithm (local alignment) for DNA sequence alignment. *Jurnal Teknologi*, *77*(20). https://doi.org/10.11113/jt.v77.6564

Muhammed, M. T., & Aki-Yalcin, E. (2024). Molecular Docking: Principles, advances, and its applications in drug discovery. *Letters in Drug Design & Discovery, 21*(3), 480–495. https://doi.org/10.2174/1570180819666220922103109

Murail, S., de Vries, S. J., Rey, J., Moroy, G., & Tufféry, P. (2021). SeamDock: An interactive and collaborative online docking resource to assist small compound molecular docking. *Frontiers in Molecular Biosciences, 8*. https://doi.org/10.3389/fmolb.2021.716466

Nair, R., & Rost, B. (2002). Sequence conserved for subcellular localization. *Protein Science, 11*(12), 2836–2847. https://doi.org/10.1110/ps.0207402

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*(3), 443–453. https://doi.org/10.1016/0022-2836(70)90057-4

Nguyen, L., Schmidt, H., Haeseler, A., & Minh, B. (2014). Iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution, 32*(1), 268-274. https://doi.org/10.1093/molbev/msu300

Ning, J., Liebich, J., Kästner, M., Zhou, J., Schäffer, A., & Burauel, P. (2009). Different influences of DNA purity indices and quantity on PCR-based DGGE and functional gene microarray in soil microbial community study. *Applied Microbiology and Biotechnology*, *82*(5), 983–993. https://doi.org/10.1007/s00253-009-1912-0

Norliana-Izzati, M., Lee, C., Fauziah, A., Zunoliza, A., Ling, S., Mohd-Hafidz-Hadi, A., Ng, K., Ng, C., Tnah, L., Ng, C., Goh, H.-H., & Lee, S. (2024). HPLC-PDA method for the quantification of mitragynine in fresh kratom (*Mitragyna speciosa*) leaf.

Journal Of Tropical Forest Science, 36(4), 434–443. https://doi.org/10.26525/jtfs2024.36.4.434

Nützmann, H., Huang, A., & Osbourn, A. (2016). Plant metabolic clusters – from genetics to genomics. *New Phytologist, 211*(3), 771-789. https://doi.org/10.1111/nph.13981

Pan, Q., Wang, Q., Yuan, F., Xing, S., Zhao, J., Choi, Y. H., Verpoorte, R., Tian, Y., Wang, G., & Tang, K. (2012). Overexpression of ORCA3 and G10H in *Catharanthus roseus* plants regulated alkaloid biosynthesis and metabolism revealed by NMR-metabolomics. *PLoS ONE, 7*(8), e43038. https://doi.org/10.1371/journal.pone.0043038

Panchenko, A. R., Kondrashov, F., & Bryant, S. (2004). Prediction of functional sites by analysis of sequence and structure conservation. *Protein Science: A Publication of the Protein Society*, *13*(4), 884–892. https://doi.org/10.1110/ps.03465504

Papadi, G., Bakhiya, N., & Ildico Hirsch-Ernst, K. (2022). Assessment of the possible health risks associated with the consumption of botanical preparations of *Mitragyna speciosa* (kratom). *EFSA Journal, 20*(EU-FORA Series 4). https://doi.org/10.2903/j.efsa.2022.e200415

Parammal, A., Kumar, M., Singh, S., Xavier, J. S., & Subramanian, P. (2023). The total synthesis of aspidofractinine and related alkaloids. *European Journal of Organic Chemistry, 27*(3). https://doi.org/10.1002/ejoc.202300960

Parvathy, S. T., Udayasuriyan, V., & Bhadana, V. (2022). Codon usage bias. *Molecular Biology Reports*, *49*(1), 539–565. https://doi.org/10.1007/s11033-021-06749-4

Patel, S., Panchal, H., & Anjaria, K. (2012). Phylogenetic analysis of some leguminous trees using CLUSTALW2 bioinformatics tool. *2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops*.

Patil, S., & Gupta, A. D. (2022). Bioinformatics and its application in computing biological data. In *Information Retrieval in Bioinformatics* (pp. 133–154). Springer Nature Singapore.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Lázaro Pinto, B., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H., Bateman, A., & InterPro Consortium. (2023). InterPro in 2022. *Nucleic Acids Research, 51*(D1), D418–D427. https://doi.org/10.1093/nar/gkac993

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, *Chapter 3*(1), 3.1.1-3.1.8. https://doi.org/10.1002/0471250953.bi0301s42

Petřek, M., Košinová, P., Koča, J., & Otyepka, M. (2007). Mole: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure, 15*(11), 1357-1363. https://doi.org/10.1016/j.str.2007.10.007

Phoong, S. W., Phoong, S. Y., Mohd Maasom, H., & Ahmad, M. A. (2025). Associations between kratom risk perceptions and intention to legalize in Malaysia. *Journal of Substance Use*, *30*(4), 662–667. https://doi.org/10.1080/14659891.2024.2378759

Pootakham, W., Yoocha, T., Jomchai, N., Kongkachana, W., Naktang, C., Sonthirod, C., Chowpongpang, S., Aumpuchin, P., & Tangphatsornruang, S. (2022). A chromosome-scale genome assembly of *Mitragyna speciosa* (kratom) and the assessment of its genetic diversity in Thailand. *Biology, 11*(10), 1492. https://doi.org/10.3390/biology11101492

Prozialeck, W. C. (2016). Update on the pharmacology and legal status of Kratom. *Journal of the American Osteopathic Association*, *116*(12), 802–809. https://doi.org/10.7556/jaoa.2016.156

Qi, Y., Cheng, X., Han, W., Jo, S., Schulten, K., & Im, W. (2014). CHARMM-GUI PACE CG builder for solution, micelle, and bilayer coarse-grained simulations. *Journal of Chemical Information and Modeling, 54*(4), 1003–1009.

Qu, Y., Easson, M. E. A. M., Simionescu, R., Hajicek, J., Thamm, A. M. K., Salim, V., & De Luca, V. (2018). Solution of the multistep pathway for assembly of corynanthean, strychnos, iboga, and aspidosperma monoterpenoid indole alkaloids from 19 E-

geissoschizine. *Proceedings of the National Academy of Sciences, 115*(12), 3180–3185. https://doi.org/10.1073/pnas.1719979115

Rahmdel, S., Purkayastha, M., Nega, M., Liberini, E., Li, N., Luqman, A., Brüggemann, H., & Götz, F. (2024). Diversity of neurotransmitter-producing human skin commensals. *International Journal of Molecular Sciences, 25*(22), 12345. https://doi.org/10.3390/ijms252212345

Ramachandran, S., Kota, P., Ding, F., & Dokholyan, N. V. (2011). Automated minimization of steric clashes in protein structures. *Proteins*, *79*(1), 261–270. https://doi.org/10.1002/prot.22879

Rashid, N. N. A., Nasir, M. H. M., Hamzah, N., Ismail, C. M. K. H., Hishamuddin, S. A. S. N., Suffian, I. F., & Hamid, A. A. A. (2024). Three-dimensional structure of Human Epididymis Protein 4 (HE4): A protein modelling of an ovarian cancer biomarker through *in silico* approach. *Journal of Tropical Life Science, 14*(02), 331–348. https://doi.org/10.11594/jtls.14.02.13

Reddy, B., & Fields, R. (2022). Multiple sequence alignment algorithms in bioinformatics. In *Lecture Notes in Networks and Systems* (pp. 89–98). Springer Singapore.

Reille, S., Garnier, M., Robert, X., Gouet, P., Martin, J., & Launay, G. (2018). Identification and visualization of protein binding regions with the ArDock server. *Nucleic Acids Research, 46*(W1), W417–W422. https://doi.org/10.1093/nar/gky472

Riascos-España, A. F., Cuastumal, B. T. A., Zambrano, M. C. I., Arteaga, J. Z. C., & Velasquez-Vasconez, P. A. (2025). Optimized protocol for DNA extraction in three Theobroma species. *Bio-Protocol, 15*(9), e5297. https://doi.org/10.21769/BioProtoc.5297

Rudnitskaya, A., Török, B., & Török, M. (2010). Molecular docking of enzyme inhibitors. *Biochemistry and Molecular Biology Education, 38*(4), 261–265. https://doi.org/10.1002/bmb.20392

Sangar, V., Blankenberg, D. J., Altman, N., & Lesk, A. M. (2007). Quantitative sequence-function relationships in proteins based on Gene Ontology. *BMC Bioinformatics*, *8*(1), 294. https://doi.org/10.1186/1471-2105-8-294

Saingam, D., Singh, D., Geater, A. F., Assanangkornchai, S., Jitpiboon, W., & Latkin, C. (2023). The health impact of long-term kratom (*Mitragyna speciosa*) use in Southern Thailand. *Substance Use & Misuse, 58*(10), 1212–1225. https://doi.org/10.1080/10826084.2023.2215301

Salim, H. M., Puspitarini, M. D., Setiwati, Y., & Shimabukuro, M. (2021). Antibacterial mechanism of kratom (*Mitragyna speciosa*) methanol extract on *Streptococcus pneumoniae* and *Escherichia coli* bacteria. *Biomolecular and Health Science Journal, 4*(2), 98. https://doi.org/10.20473/bhsj.v4i2.28933

Saran, A., Chandra, A., Jain, S., Pareek, K., Paliwal, S., & Sharma, S. (2023). Exploring multifaceted applications of bioinformatics. In *Bioinformatics and Computational Biology* (pp. 313–326). Chapman and Hall/CRC.

Sariyatun, R., Reviono, R., Prasetyo, A., & Kageyama, S. (2015). Molecular analysis of a complete hiv-1 pol gene isolated in Central Java, Indonesia. *Kne Life Sciences, 2*(1), 34. https://doi.org/10.18502/kls.v2i1.113

Schotte, C., Jiang, Y., Grzech, D., Dang, T.-T. T., Laforest, L. C., León, F., Mottinelli, M., Nadakuduti, S. S., McCurdy, C. R., & O'Connor, S. E. (2023). Directed biosynthesis of mitragynine stereoisomers. *Journal of the American Chemical Society, 145*(9), 4957–4963. https://doi.org/10.1021/jacs.2c13644

Sharma, A., Kamble, S. H., León, F., Chear, N. J. -Y., King, T. I., Berthold, E. C., Ramanathan, S., McCurdy, C. R., & Avery, B. A. (2019). Simultaneous quantification of ten key Kratom alkaloids in *Mitragyna speciosa* leaf extracts and commercial products by ultra-performance liquid chromatography−tandem mass spectrometry. *Drug Testing and Analysis, 11*(8), 1162–1171. https://doi.org/10.1002/dta.2604

Sharma, M. A., Anand Vihar, Thakur, D. R. S. & Jaloree, D. S. (2019). Computational analysis of distance-based phylogenetic tree for Azotobacter species. *International*

Journal of Innovative Technology and Exploring Engineering, 8(10), 810–814. https://doi.org/10.35940/ijitee.j9006.0881019

Sheng, X. & Himo, F. (2020). Computational study of pictet–spenglerase strictosidine synthase: reaction mechanism and origins of enantioselectivity of natural and non-natural substrates. *Acs Catalysis, 10*(22), 13630-13640. https://doi.org/10.1021/acscatal.0c03758

Singh, D., Grundmann, O., Murugaiyah, V., Rahim, A. B. M., Chawarski, M., & Balasingam, V. (2020). Improved sexual functioning of long-term daily users of *Mitragyna speciosa* (Korth.). *Journal of Herbal Medicine, 19*, 100293. https://doi.org/10.1016/j.hermed.2019.100293

Singh, G. B. (2015). Multiple Sequence Alignment. In *Fundamentals of Bioinformatics and Computational Biology* (pp. 143–158). Springer International Publishing.

Singh, M. (2005). Predicting protein secondary and supersecondary structure. In *Chapman & Hall/CRC Computer & Information Science Series* (pp. 29-1-29–29). Chapman and Hall/CRC.

Smith, K., Sharma, A., Grundmann, O., & McCurdy, C. (2023a). Kratom alkaloids: a blueprint?. *Acs Chemical Neuroscience, 14*(2), 195-197. https://doi.org/10.1021/acschemneuro.2c00704

Smith, K. E., Grundmann, O., Swogger, M. T., & Garcia-Romeu, A. (2023b). Kratom (*Mitragyna speciosa*). In *The Oxford Handbook of Opioids and Opioid Use Disorder* (pp. 386–418). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197618431.013.15.

Stourac, J., Vavra, O., Kokkonen, P., Filipovic, J., Pinto, G., Brezovsky, J., Damborsky, J., & Bednar, D. (2019). Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Research*, *47*(W1), W414–W422. https://doi.org/10.1093/nar/gkz378

Tan, K., Nguyen, T., Patel, S., Varadarajan, R., & Madhusudhan, M. (2013). Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding

cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Research, 41*(W1), W314-W321. https://doi.org/10.1093/nar/gkt503

Thompson, Julie D. (2005). The Clustal series of programs for multiple sequence alignment. In *The Proteomics Protocols Handbook* (pp. 493–502). Humana Press.

Thompson, J. D., Koehl, P., Ripp, R., & Poch, O. (2005). BAliBASE 3.0: latest developments of the multiple sequence alignment benchmark. *Proteins*, *61*(1), 127–136. https://doi.org/10.1002/prot.20527

Triant, D. A., & Pearson, W. R. (2015). Most partial domains in proteins are alignment and annotation artifacts. *Genome Biology*, *16*(1), 99. https://doi.org/10.1186/s13059-015-0656-7

van der Fits, L., Hilliou, F., & Memelink, J. (2001). T-DNA activation tagging as a tool to isolate regulators of a metabolic pathway from a genetically non-tractable plant species. *Transgenic Research, 10*(6), 513–521. https://doi.org/10.1023/a:1013087011562

Veeramohan, R., Zamani, A. I., Azizan, K. A., Goh, H.-H., Aizat, W. M., Razak, M. F. A., Yusof, N. S. M., Mansor, S. M., Baharum, S. N., & Ng, C. L. (2023). Comparative metabolomics analysis reveals alkaloid repertoires in young and mature *Mitragyna speciosa* (Korth.) Havil. leaves. *PloS One*, *18*(3), e0283147. https://doi.org/10.1371/journal.pone.0283147

Veltri, C., & Grundmann, O. (2019). Current perspectives on the impact of Kratom use. *Substance Abuse and Rehabilitation*, *Volume 10*, 23–31. https://doi.org/10.2147/sar.s164261

Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A., & Clark, P. L. (2020). Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(7), 3528–3534. https://doi.org/10.1073/pnas.1907126117

Wiegert, J., Höhler, D., Haag, J., & Stamatakis, A. (2024). Predicting phylogenetic bootstrap values via machine learning. *Molecular Biology and Evolution, 41*(10). https://doi.org/10.1093/molbev/msae215

Wilfinger, W. W., Mackey, K., Krug, D., & Krug, R. (1997). RNA integrity number (RIN)—Standardization of RNA quality control. *BioTechniques, 22*(3), 550-558.

Wungsintaweekul, J., Choo-Malee, J., Charoonratana, T., & Keawpradub, N. (2012). Methyl jasmonate and yeast extract stimulate mitragynine production in *Mitragyna speciosa* (Roxb.) Korth. shoot culture. *Biotechnology Letters*, *34*(10), 1945–1950. https://doi.org/10.1007/s10529-012-0968-6

Yaffe, E., Fishelovitch, D., Wolfson, H., Halperin, D., & Nussinov, R. (2008). Molaxis: efficient and accurate identification of channels in macromolecules. *Proteins Structure Function and Bioinformatics, 73*(1), 72-86. https://doi.org/10.1002/prot.22052

Yang, L., Zou, H., Zhu, H., Ruppert, M., Gong, J., & Stöckigt, J. (2010). Improved expression of his6-tagged strictosidine synthase cDNA for chemo-enzymatic alkaloid diversification. *Chemistry & Biodiversity, 7*(4), 860-870. https://doi.org/10.1002/cbdv.201000052

You, D., Feng, Y., Wang, C., Sun, C., Wang, Y., Zhao, D., & Kai, G. (2021). Cloning, characterization, and enzymatic identification of a new tryptophan decarboxylase from *Ophiorrhiza pumila*. *Biotechnology and Applied Biochemistry*, *68*(2), 381–389. https://doi.org/10.1002/bab.1935

Yusoff, N. H. M., Suhaimi, F. W., Vadivelu, R. K., Hassan, Z., Rümler, A., Rotter, A., Amato, D., Dringenberg, H. C., Mansor, S. M., Navaratnam, V., & Müller, C. P. (2014). Abuse potential and adverse cognitive effects of mitragynine (kratom). *Addiction Biology, 21*(1), 98–110. https://doi.org/10.1111/adb.12185

Zailan, N. F. Z., Sarchio, S. N. E., & Hassan, M. (2022). Evaluation of phytochemical composition, antioxidant and anti-diabetic activities of *Mitragyna speciosa* methanolic extract (MSME). *Malaysian Journal of Medicine and Health Sciences, 18*(s21), 93–100. https://doi.org/10.47836/mjmhs.18.s21.15

Zhang, M., Sharma, A., León, F., Avery, B., Kjelgren, R., McCurdy, C. R., & Pearson, B. J. (2020). Effects of nutrient fertility on growth and alkaloidal content in *Mitragyna speciosa* (Kratom). *Frontiers in Plant Science, 11*. https://doi.org/10.3389/fpls.2020.597696

Zhao, Y., Zhang, H., & Liu, Y. (2020). Protein secondary structure prediction based on generative confrontation and convolutional neural network. *IEEE Access, 8*, 199171-199178. https://doi.org/10.1109/access.2020.3035208

Zharkikh, A., & Li, W. H. (1992). Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Molecular Biology and Evolution*, *9*(6), 1119–1147. https://doi.org/10.1093/oxfordjournals.molbev.a040782

Zhu, W., Yang, B., Komatsu, S., Lu, X., Li, X., & Tian, J. (2015). Binary stress induces an increase in indole alkaloid biosynthesis in *Catharanthus roseus*. *Frontiers in Plant Science, 6*. https://doi.org/10.3389/fpls.2015.00582

Zhu, X., Zeng, X., Sun, C., & Chen, S. (2014). Biosynthetic pathway of terpenoid indole alkaloids in *Catharanthus roseus*. *Frontiers of Medicine, 8*(3), 285-293. https://doi.org/10.1007/s11684-014-0350-2

APPENDIX

A. DNA Sequence

>Result_TDC_*M. speciosa*_from Malaysia

AAATTTGTTCTGAACCAATTACTCTCCAAGAGAATGGGCAGCATTGATACGA
GTGACGGTGATGCTTATGCCAATTCAGCAGTTGCACCATTCAAGCCACTTGAT
CCTGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTA
TTACAAAAATATTGAAAACTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATC
TCCGAACCCAGCTATCGCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAAT
ATTCTACAAGATATTCAAAAAGATATCATCCCTGGAATGACCAACTGGTTAA
GCCCTAACTTTTTTGCATTTTTTCCTGCGACGGTTAGCTCTGCCGCTTTTCTTG
GAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTTCAACTGGCTCGCTTCC
CCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTGGTTGGCTAACATGC
TTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCAA
GGAACAACCAGTGAGGCCATTCTTTGCACCATCATCGCCGCCCGCGACCGGG
CTTTTGAGAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCT
GATCAAACACATAGCTTTTTCGTCAAGACTTGCAAATTGGCTGGTATCTTCCC
ATGCAATATAAGAATAATCCCTACTACTGCTAGAGAGAACTTTTCCATGTCCC
CTCAGGCTGTACGTCGACAAATGCAAGCTGACGTGAGAAATGGGCTGGTCCC
ACTTTTCCTCTGTGCTACTGTGGGGACAACTTCCACCACTGCCATTGACCCGG
TGAGTCAACTAGCTGAAATTGCAAATAATTTCAACGTGTGGATTCACGTTGAT
GGTGCTTATGCAGGCAGTGCATGCATATGTCCCGAGTTCAGGCAATACTTGGA
AGGGATCGAGCGAGTTGACTCGTTGAGCCTGAGCCCTCACAAGTGGCTACTTT
GTTACTTGGACTGTTGTTGCTTGTGGGTGAAGAAAACAGACTTGCTGGTGAAG
GCATTAGCCACCAATCCCGAGTACTTGAGAAACAAAAGAAGCGAGTTCGACT
CAGTTGTGGATTACAAAGACTGGCAAATTGGCACAGGTAAACGTTTCAGGGC
ATTACGATTATGGCTTGTTATGCGTTGTTATGGAGTTGCAAACCTACAAAGCC
ACATCCGATCCGACGTTCAAATGGCCAAGATGTTTGAAGGCTTTGTGAAATCT
GACCCAAGATTTGAAATGATAGTACCACGTGCATTTTCACTCGTATGTTTTAG
GCTTAACCCTTCGGGAGGATCCAACGAAGCAGACTTGGAGCTTCTAAACAAG
AAGCTACTCGACCGGGTCAACTCAACAGGCCGAGCTTACATGACTCACACTA
AGGCCGGTGAGGTATACTTGCTGAGATTTGCAGTGGGGGCTACACTAACAGA
GGATCGCCATGTGTATGCTGCTTGGGAGTTGATTAAACAATGCGCTGACGCTG
TGCTTAAAGAAAATGTTTCAGACTGATGAATAAGCTATTTTGGTACTGTAGTG
ATCCTTCAAAAAAGGGTA

B. Protein Sequence

>Result_TDC_*M. speciosa*_Amino acid_from Malaysia

MGSIDTSDGDAYANSAVAPFKPLDPDEFRKQAHRMVDFIADYYKNIENYPVLSQ
VEPGYLRTQLSQTAPYLPEPFENILQDIQKDIIPGMTNWLSPNFFAFFPATVSSAAF
LGEMLCTGFNSVGFNWLASPAATELEMVVMDWLANMLKLPKSFMFSGTGGGV
LQGTTSEAILCTIIAARDRAFEKIGVENIGKLVVYASDQTHSFFVKTCKLAGIFPCN
IRIIPTTARENFSMSPQAVRRQMQADVRNGLVPLFLCATVGTTSTTAIDPVSQLAE
IANNFNVWIHVDGAYAGSACICPEFRQYLEGIERVDSLSLSPHKWLLCYLDCCCL
WVKKTDLLVKALATNPEYLRNKRSEFDSVVDYKDWQIGTGKRFRALRLWLVM
RCYGVANLQSHIRSDVQMAKMFEGFVKSDPRFEMIVPRAFSLVCFRLNPSGGSNE
ADLELLNKKLLDRVNSTGRAYMTHTKAGEVYLLRFAVGATLTEDRHVYAAWEL
IKQCADAVLKENVSD

C. Pairwise Alignment Validation for TDC *M. speciosa* from Malaysia

Pairwise Alignment of TDC *M. speciosa* Full DNA Sequence

JN643922.1_TDC_TEMPLATE

Sequence ID: **Query_7653925** Length: **1620** Number of Matches: **1**

Range 1: 1 to 1620 Graphics ▼ Next Match ▲ P

NW Score	Identities	Gaps	Strand
2929	1557/1622(96%)	25/1622(1%)	Plus/Plus

```
Query  1    ------------AAATTTGTTCTGAACCAATTACTCTCCAAGAGAATGGGCAGCATTGA  47
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    CTTCACCCTCATTAAATTTGTTCTGAACCAATTACTCTCCAAGAGAATGGGCAGCATTGA  60

Query  48   TACGAGTGACGGTGATGCTTATGCCAATTCAGCAGTTGCACCATTCAAGCCACTTGATCC  107
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   TACGAGTGACGGTGATGCTTATGCCAATTCAGCAGTTGCACCATTCAAGCCACTTGATCC  120

Query  108  TGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTATTACAAAAA  167
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  TGACGAATTCAGAAAACAAGCCCATCGTATGGTTGATTTCATAGCCGATTATTACAAAAA  180

Query  168  TATTGAAAACTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATCTCCGAACCCAGCTATC  227
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  TATTGAAAACTATCCTGTTCTCAGCCAAGTTGAGCCTGGATATCTCCGAACCCAGCTATC  240

Query  228  GCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAATATTCTACAAGATATTCAAAAAGA  287
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAAACTGCCCCTTATCTCCCTGAGCCATTTGAAAATATTCTACAAGATATTCAAAAAGA  300

Query  288  TATCATCCCTGGAATGACCAACTGGTTAAGCCCTAACTTTTTTGCATTTTTTCCTGCGAC  347
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  TATCATCCCTGGAATGACCAACTGGTTAAGCCCTAACTTTTTTGCATTTTTTCCTGCGAC  360

Query  348  GGTTAGCTCTGCCGCTTTTCTTGGAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTT  407
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GGTTAGCTCTGCCGCTTTTCTTGGAGAAATGTTGTGCACTGGCTTTAACTCTGTAGGTTT  420

Query  408  CAACTGGCTCGCTTCCCCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTGGTTGGC  467
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  CAACTGGCTCGCTTCCCCGGCCGCGACAGAGCTCGAAATGGTGGTGATGGATTGGTTGGC  480

Query  468  TAACATGCTTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCA  527
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TAACATGCTTAAGCTCCCTAAGTCCTTCATGTTCTCTGGCACTGGCGGTGGTGTTCTCCA  540

Query  528  AGGAACAACCAGTGAGGCCATTCTTTGCACCATCATCGCCGCCCGCGACCGGGCTTTTGA  587
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  AGGAACAACCAGTGAGGCCATTCTTTGCACCATCATCGCCGCCCGCGACCGGGCTTTTGA  600

Query  588  GAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCTGATCAAACACATAG  647
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601  GAAAATCGGTGTCGAAAACATTGGAAAGCTTGTTGTCTATGCTTCTGATCAAACACATAG  660

Query  648  CTTTTTCGTCAAGACTTGCAAATTGGCTGGTATCTTCCCATGCAATATAAGAATAATCCC  707
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661  CTTTTTCGTCAAGACTTGCAAATTGGCTGGTATCTTCCCATGCAATATAAGAATAATCCC  720

Query  708  TACTACTGCTAGAGAGAACTTTTCCATGTCCCCTCAGGCTGTACGTCGACAAATGCAAGC  767
            ||||||||||  || |||||||||||||| ||| | ||| |||||  ||||||  |||||
Sbjct  721  TACTACTGCTGATGACAACTTTTCCATGTCACCTGATGCTCTACGTAAACAAATTGAAGC  780

Query  768  TGACGTGAGAAATGGGCTGGTCCCACTTTTCCTCTGTGCTACTGTGGGGACAACTTCCAC  827
            |||||||  | ||||||||||||||||||  |||||||||||||||||||||||||| ||
Sbjct  781  TGACGTGGAAGATGGGCTGGTCCCACTTTTTATCTGTGCTACTGTGGGGACAACTTCTAC  840

Query  828  CACTGCCATTGACCCGGTGAGTCAACTAGCTGAAATTGCAAATAATTTCAACGTGTGGAT  887
            |||||| |||||||| ||||||| |||||||| ||  |||||||| ||||||||||||||
Sbjct  841  CACTGCAATTGACCCAGTGAGTGAACTAGCCGATGTTGCAAATGATTTCAACGTGTGGAT  900
```

Query - TDC DNA sequence isolated from *M. speciosa* (Malaysia)

Subject- Reference TDC DNA sequence from *M. speciosa* (Thailand, NCBI database)

Pairwise Alignment of TDC *M. speciosa* Full Protein Sequence

Template_TDC_M. speciosa

Sequence ID: Query_4278519 **Length:** 506 **Number of Matches:** 1

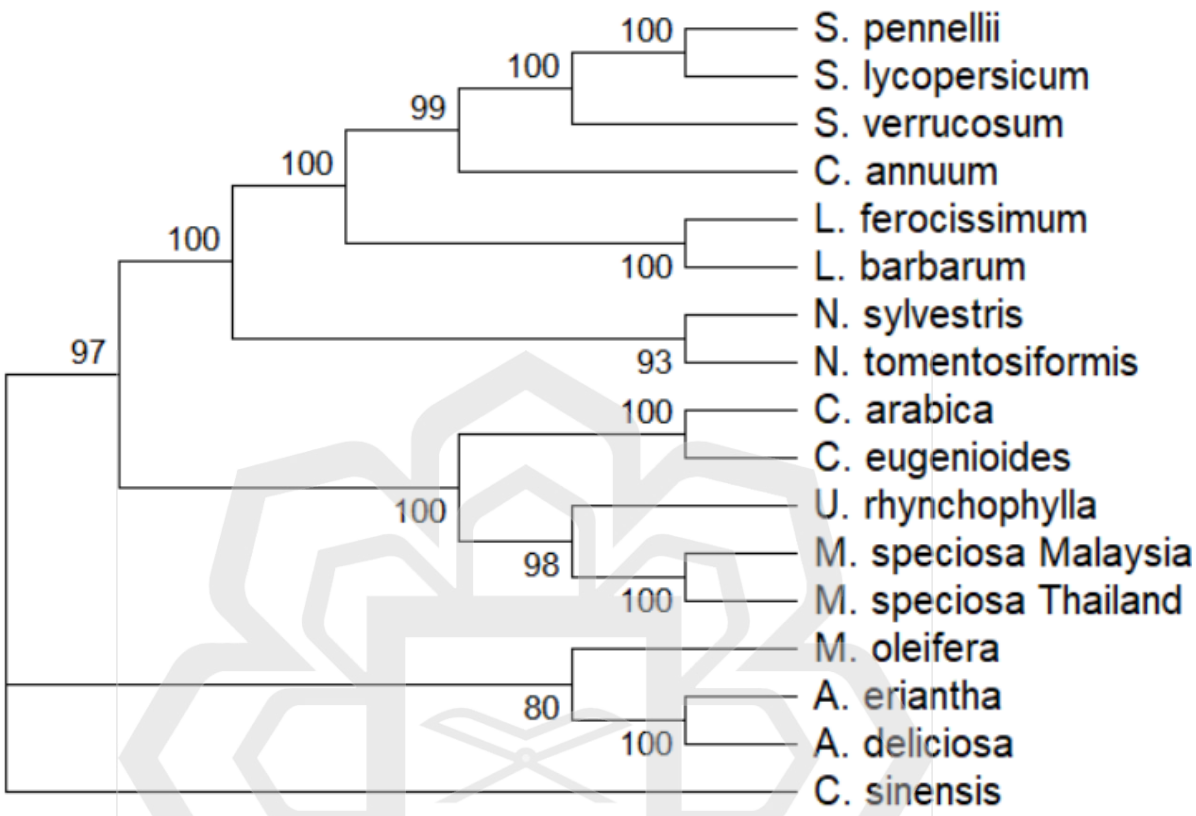
Range 1: 1 to 506 Graphics ▼ Next Match

NW Score	Identities	Positives	Gaps
2582	487/506(96%)	499/506(98%)	0/506(0%)

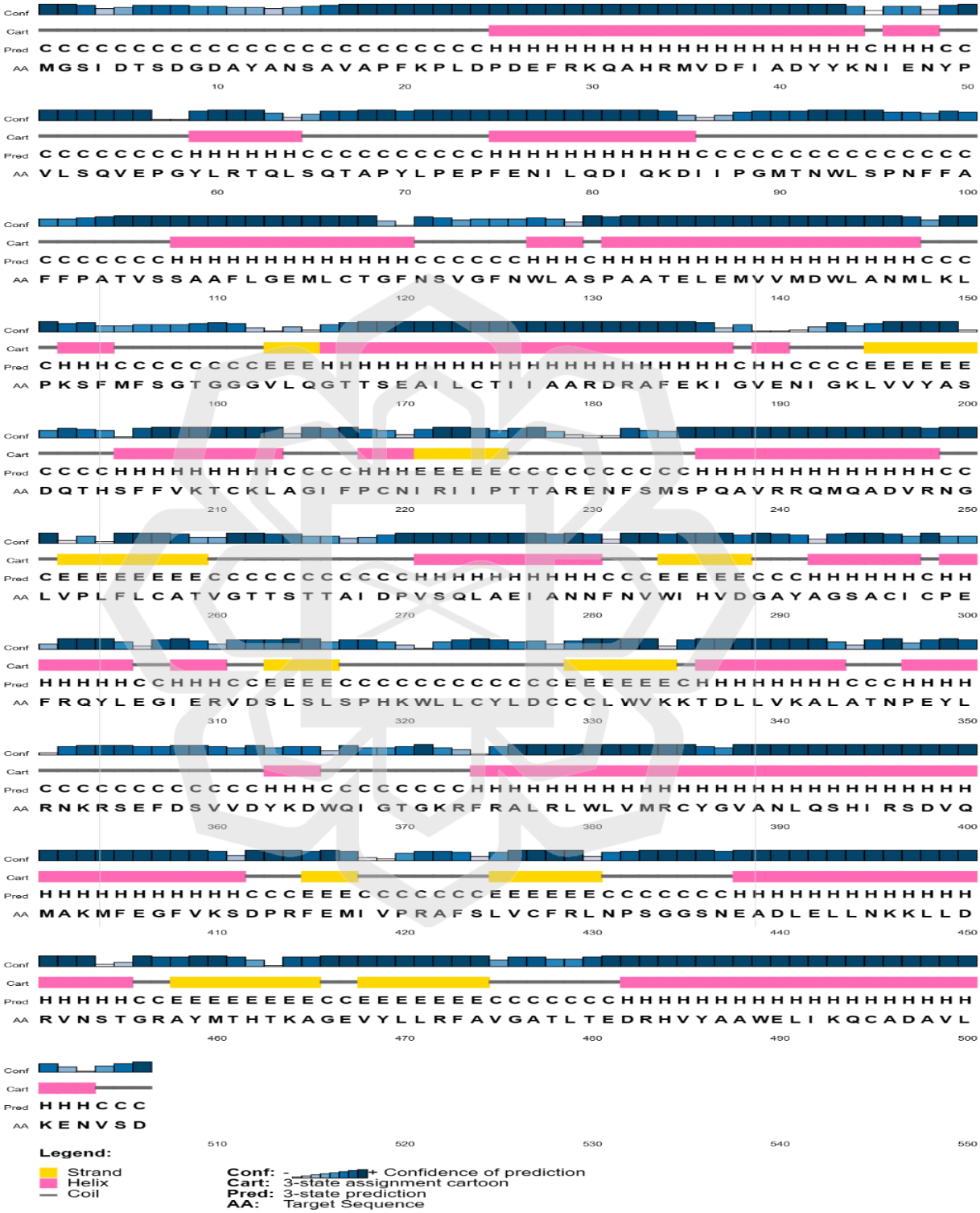
```
Query  1    MGSIDTSDGDAYANSAVAPFKPLDPDEFRKQAHRMVDFIADYYKNIENYPVLSQVEPGYL  60
            MGSIDTSDGDAYANSAVAPFKPLDPDEFRKQAHRMVDFIADYYKNIENYPVLSQVEPGYL
Sbjct  1    MGSIDTSDGDAYANSAVAPFKPLDPDEFRKQAHRMVDFIADYYKNIENYPVLSQVEPGYL  60

Query  61   RTQLSQTAPYLPEPFENILQDIQKDIIPGMTNWLSPNFFAFFPATVSSAAFLGEMLCTGF  120
            RTQLSQTAPYLPEPFENILQDIQKDIIPGMTNWLSPNFFAFFPATVSSAAFLGEMLCTGF
Sbjct  61   RTQLSQTAPYLPEPFENILQDIQKDIIPGMTNWLSPNFFAFFPATVSSAAFLGEMLCTGF  120

Query  121  NSVGFNWLASPAATELEMVVMDWLANMLKLPKSFMFSGTGGGVLQGTTSEAILCTIIAAR  180
            NSVGFNWLASPAATELEMVVMDWLANMLKLPKSFMFSGTGGGVLQGTTSEAILCTIIAAR
Sbjct  121  NSVGFNWLASPAATELEMVVMDWLANMLKLPKSFMFSGTGGGVLQGTTSEAILCTIIAAR  180

Query  181  DRAFEKIGVENIGKLVVYASDQTHSFFVKTCKLAGIFPCNIRIIPTTARENFSMSPQAVR  240
            DRAFEKIGVENIGKLVVYASDQTHSFFVKTCKLAGIFPCNIRIIPTTA +NFSMSP A+R
Sbjct  181  DRAFEKIGVENIGKLVVYASDQTHSFFVKTCKLAGIFPCNIRIIPTTADDNFSMSPDALR  240

Query  241  RQMQADVRNGLVPLFLCATVGTTSTTAIDPVSQLAEIANNFNVWIHVDGAYAGSACICPE  300
            +Q++ADV +GLVPLF+CATVGTTSTTAIDPVS+LA++AN+FNVWIHVD AYAGSACICPE
Sbjct  241  KQIEADVEDGLVPLFICATVGTTSTTAIDPVSELADVANDFNVWIHVDAAYAGSACICPE  300

Query  301  FRQYLEGIERVDSLSLSPHKWLLCYLDCCCLWVKKTDLLVKALATNPEYLRNKRSEFDSV  360
            FRQYL+GIERVDSLSLSPHKWLLCYLDCCCLWVKKTDLLVKALATNPEYLRNKRSEFDSV
Sbjct  301  FRQYLDGIERVDSLSLSPHKWLLCYLDCCCLWVKKTDLLVKALATNPEYLRNKRSEFDSV  360

Query  361  VDYKDWQIGTGKRFRALRLWLVMRCYGVANLQSHIRSDVQMAKMFEGFVKSDPRFEMIVP  420
            VDYKDWQIGTGKRFRALRLWLVMRCY VANLQSHIRSDVQMAKMFEGFVKSDPRFEMIVP
Sbjct  361  VDYKDWQIGTGKRFRALRLWLVMRCYRVANLQSHIRSDVQMAKMFEGFVKSDPRFEMIVP  420

Query  421  RAFSLVCFRLNPSGGSNEADLELLNKKLLDRVNSTGRAYMTHTKAGEVYLLRFAVGATLT  480
            RAFSLVCFRLNPSGGSNEADLELLNKKLLDRVNSTGR YMTHTKAGEVYLLRFAVGATLT
Sbjct  421  RAFSLVCFRLNPSGGSNEADLELLNKKLLDRVNSTGRTYMTHTKAGEVYLLRFAVGATLT  480

Query  481  EDRHVYAAWELIKQCADAVLKENVSD  506
            EDRHVYAAWELIKQCADAVLKENV D
Sbjct  481  EDRHVYAAWELIKQCADAVLKENVLD  506
```

Query - Translated TDC protein sequence from *M. speciosa* (Malaysia)
Subject - Reference TDC protein sequence from *M. speciosa* (Thailand)

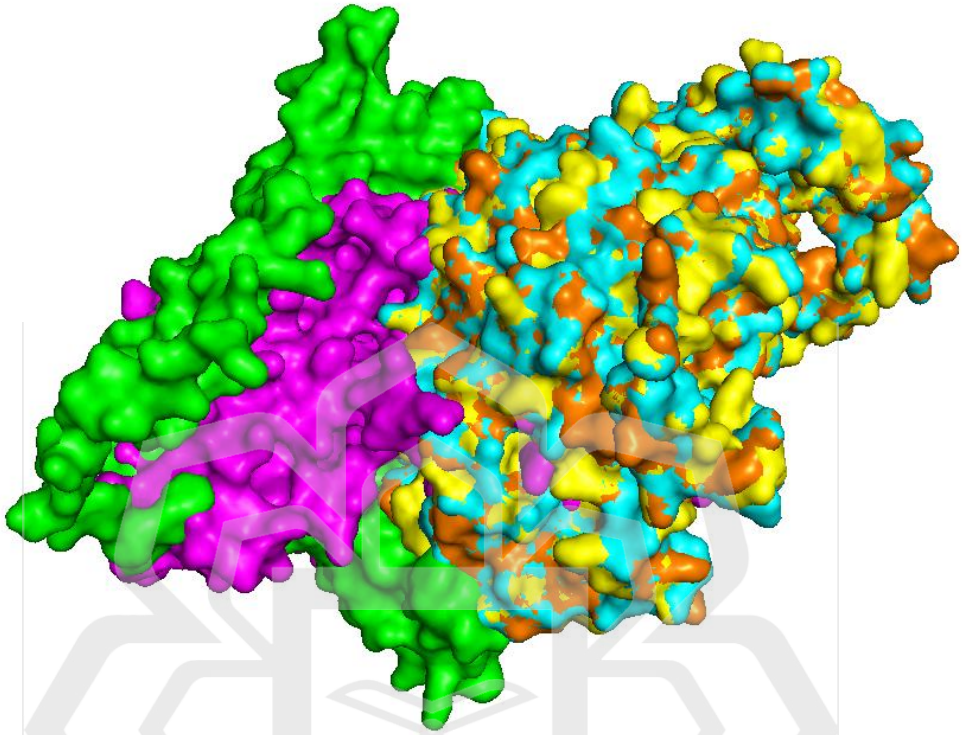
D. Bootstrap Consensus Tree



E. Secondary Structure Prediction



F. Superimposed 5 Predicted Modelling Structure



The colour surface represents template-based prediction of TDC (orange), AlphaFold (green), SWISS-MODEL (blue), c) Phyre2 (yellow), and d) I-TASSER (purple).

G. Predicted Binding Affinities and Poses from Webina

Mode	Affinity (kcal/mol)	Distance from Rmsd L.B.	Distance from Rmsd U.B.
1	-9.901	0	0
2	-8.883	1.308	3.309
3	-8.732	1.973	3.086
4	-8.498	1.066	2.096
5	-8.341	1.438	2.011
6	-8.105	2.597	3.766
7	-7.986	5.79	8.068
8	-7.822	2.061	3.466
9	-7.773	1.505	2.558