

# **Integrating Computational and Biotechnology Approaches for Enhanced Tocopherol and Tocotrienol Biosynthesis in Plants**

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT  
FOR THE DEGREE OF

**MASTER OF TECHNOLOGY**  
in  
**INDUSTRIAL BIOTECHNOLOGY**

Submitted by  
**MIHIR KUMAR**  
(24/IBT/15)

Under the Supervision of  
**Dr. Navneeta Bharadvaja**



**DEPARTMENT OF BIOTECHNOLOGY**

**DELHI TECHNOLOGICAL UNIVERSITY**  
(Formerly Delhi College of Engineering)  
Bawana Road, Delhi-110042

**May 2026**

## ACKNOWLEDGEMENT

The successful completion of this dissertation would not have been possible without the guidance and support of many people.

I would like to express my sincere gratitude to **Dr. Navneeta Bharadvaja**, Department of Biotechnology, Delhi Technological University, for her valuable guidance, constant encouragement, motivation, and support throughout the course of this work. Her insightful suggestions and scientific approach greatly helped me in completing this dissertation successfully.

I would also like to thank **Prof. Yasha Hasija**, Head of Department, Department of Biotechnology, Delhi Technological University, for providing the necessary facilities and academic environment required for carrying out this work.

I express my heartfelt thanks to all the faculty members and laboratory staff of the Department of Biotechnology for their cooperation and support during the project work. I would also like to extend my sincere thanks to **Ms. Harshita**, **Ms. Pragati Srivastava** and **Ms. Anuradha** for their valuable support and encouragement throughout the project tenure.

I would also like to thank **Mr. C. B. Singh** and **Mr. Jitendra Singh**, Office In charges, Department of Biotechnology, for their cooperation and assistance during the course of this work.

I would also like to specially thank my friends **Jaya Kaushik**, **Wayez Naqvi** and **Manish Kumar** for their constant encouragement, valuable support, and motivation throughout the completion of this dissertation.

Finally, I would like to thank my parents and family members for their unconditional support, encouragement, and motivation throughout my academic journey.

**DEPARTMENT OF BIOTECHNOLOGY****DELHI TECHNOLOGICAL UNIVERSITY**

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

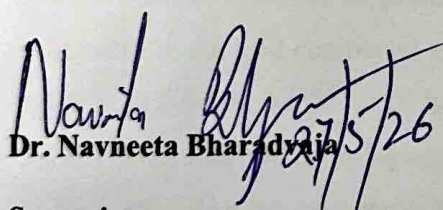
**CANDIDATE'S DECLARATION**

I, **Mihir Kumar**, hereby declare that the work which is being presented in the thesis entitled **“Integrating Computational and Biotechnology Approaches for Enhanced Tocopherol and Tocotrienol Biosynthesis in Plants”** in partial fulfilment of the requirements for the award of the Degree of Master of Technology, submitted in the **Department of Biotechnology, Delhi Technological University** is an authentic record of my own work carried out during the period from **January 2026 to May 2026** under the supervision of **Dr. Navneeta Bharadvaja**. The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

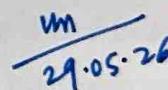
**Candidates' Signature**

**DEPARTMENT OF BIOTECHNOLOGY****DELHI TECHNOLOGICAL UNIVERSITY****(Formerly Delhi College of Engineering)****Bawana Road, Delhi-110042****CERTIFICATE BY THE SUPERVISOR**

Certified that **Mihir Kumar** (Roll No. 24/IBT/15) has carried out their research work presented in this thesis entitled **"Integrating Computational and Biotechnology Approaches for Enhanced Tocopherol and Tocotrienol Biosynthesis in Plants"** for the award of **Master of Technology** from Department of Biotechnology, Delhi Technological University, under my supervision. The thesis embodies results of original work and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University Institution.

  
Dr. Navneeta Bharadwaj

**Supervisor****Department of Biotechnology****Delhi Technological University**

  
29.05.26

**Prof. Yasha Hasija****Head of the Department****Department of Biotechnology****Delhi Technological University**

## **ABSTRACT**

Tocopherols and tocotrienols collectively constitute vitamin E, a family of lipid-soluble antioxidants indispensable for human nutrition, crop quality, and plant stress tolerance. Despite their agronomic and nutraceutical importance, the molecular regulatory network governing vitamin E biosynthesis remains incompletely characterised, and conventional breeding and transgenic approaches have achieved only modest improvements in tocopherol content. The convergence of the shikimate and methylerythritol phosphate (MEP) pathway arms at homogentisate phytyldiphosphate transferase (VTE2) represents the committed step in tocopherol synthesis, highlighting it as a high-priority target for regulatory investigation and molecular intervention. In the present study, gene expression datasets (GSE59790 and GSE47518) pertaining to tocopherol biosynthesis and high-light stress in *Arabidopsis thaliana* were retrieved from the Gene Expression Omnibus (GEO) database and subjected to differential gene expression analysis using GEO2R. Protein-protein interaction (PPI) network construction was performed using the STRING database (confidence score > 0.900), and hub genes were identified by degree centrality analysis, with functional enrichment assessed through Gene Ontology (GO) and KEGG pathway analysis. For molecular docking, the crystal structure of VTE2 (PDB ID: 1YFP) was employed as the target, and a curated library of 75 plant-derived phytochemicals was screened using EasyDock Vina, with protein-ligand interactions characterised using LigPlot+ and PyMOL. A total of 1,148 differentially expressed genes were identified, comprising 634 upregulated and 514 downregulated genes. Hub gene analysis identified VTE1, VTE4, VTE2/HPT1, HPPD, and DXPS as the most highly connected nodes in the upregulated PPI network, confirming their central regulatory roles. Functional enrichment revealed plastidic metabolic processes, prenylquinone synthesis, and oxidative stress response as predominant biological themes. Molecular docking identified alpha-mangostin as the lead compound with a binding affinity of -9.8 kcal/mol, forming critical hydrogen bonds with Ser186, Asp227, and Arg295 within the VTE2 active site, followed by beta-tocotrienol (-9.4 kcal/mol) and quercetin-3-glucoside (-9.1 kcal/mol) as additional high-affinity candidates. These findings provide a comprehensive regulatory framework for the tocopherol biosynthetic network and nominate alpha-mangostin as a promising phytochemical modulator of VTE2 warranting further experimental validation

**Keywords:** Tocopherol, Tocotrienols, Protein-protein network, Functional enrichment, GEO.

## TABLE OF CONTENT

| <b>Subheading</b>                                                                  | <b>Page No.</b> |
|------------------------------------------------------------------------------------|-----------------|
| Acknowledgement                                                                    | ii              |
| Candidate's Declaration                                                            | iii             |
| Certificate                                                                        | iv              |
| Abstract                                                                           | v               |
| Table of Contents                                                                  | vi – vii        |
| List of Tables, Figures and Abbreviations                                          | vii - x         |
| INTRODUCTION                                                                       | 1 -6            |
| 1.1 Structural Classification and Comparative Biological Properties of Vitamin E   | 4               |
| 1.2 Tocopherol Biosynthesis Pathway: Architecture and Key Enzymatic Steps          | 5               |
| LITERATURE REVIEW                                                                  | 7-12            |
| 2.1 Tocopherols, Tocotrienols and Their Biological Significance                    | 7               |
| 2.2 Biosynthetic Pathway Architecture and Rate-Limiting Enzymes                    | 7 - 8           |
| 2.3 Transcriptomic Approaches to Metabolic Gene Network Analysis                   | 8               |
| 2.4 VTE2 as a Key Enzymatic Target for Metabolic Engineering                       | 8               |
| 2.5 Phytochemical-Based Modulation and Computational Drug Discovery                | 9               |
| 2.6 Integration of Bioinformatics and Metabolic Engineering Strategies             | 10              |
| 2.7 Study Rationale and Significance                                               | 11              |
| 2.8 Comparative Transcriptomic Signatures Across Vitamin E Accumulating Conditions | 12              |
| MATERIALS & METHODS                                                                | 13 - 25         |
| 3.1 Research Methodology Overview                                                  | 13              |
| 3.2 Bioinformatics Analysis for Vitamin E Biosynthesis Gene Networks               | 13              |
| 3.2.1 Dataset Selection                                                            | 13              |
| 3.2.2 Differential Gene Expression Analysis                                        | 13 - 14         |
| 3.2.3 Data Normalisation                                                           | 14              |
| 3.2.4 Protein-Protein Interaction Network Construction                             | 14              |
| 3.2.5 Functional Enrichment Analysis                                               | 14 – 15         |
| 3.2.6 Hub Gene Identification                                                      | 15              |
| 3.3 Molecular Docking for VTE2 Modulator Identification                            | 15              |
| 3.3.1 Preparing the Target Protein                                                 | 15              |
| 3.3.2 Ligand Selection and Preparation                                             | 15              |
| 3.3.3 Molecular Docking Procedure                                                  | 16              |
| 3.3.4 Selection of Top Candidate Compounds                                         | 16              |
| 3.3.5 Protein-Ligand Interaction Analysis                                          | 16              |
| 3.3.6 Structural Visualisation                                                     | 16 – 17         |

|                                                                                        |         |
|----------------------------------------------------------------------------------------|---------|
| 3.4 Integrated Computational Framework                                                 | 17      |
| 3.5 Drug-Likeness and ADMET Pre-screening of Top Candidates                            | 18      |
| RESULTS                                                                                | 19      |
| 4.1 Differential Gene Expression Analysis                                              | 20      |
| 4.2 Data Normalisation                                                                 | 21      |
| 4.3 Functional Enrichment Analysis                                                     | 22 – 23 |
| 4.4 Hub Gene Identification and PPI Network                                            | 23      |
| 4.5 Molecular Docking Results                                                          | 23      |
| 4.6 Binding Affinity Analysis                                                          | 24      |
| 4.7 Protein-Ligand Interaction Analysis                                                | 25      |
| 4.8 Structural Visualisation                                                           | 26      |
| 4.9 Structure-Activity Relationship Analysis                                           | 27      |
| 4.10 Validation of Hub Gene Network Centrality Through Published Experimental Evidence | 28      |
| DISCUSSION & CONCLUSION                                                                | 29 – 35 |
| REFERENCES                                                                             | 36 – 42 |
| List of Publications with Certificates                                                 | 43 - 45 |
| Plagiarism Verification                                                                | 46      |
| Plagiarism Report                                                                      | 47      |
| AI Report                                                                              | 48      |
| Undertaking                                                                            | 49      |

### **LIST OF TABLES**

| <b>Table No.</b> | <b>Title</b>                                                                                          | <b>Page No.</b> |
|------------------|-------------------------------------------------------------------------------------------------------|-----------------|
| 3.1              | Docking scores, binding affinities and key interacting residues of top-ranked phytochemical with VTE2 | 25              |
| 3.2              | Preliminary Lipinski Rule-of-Five Assessment for Top 5 Phytochemical Candidates                       | 18              |
| 4.1              | Cross-Validation of Upregulated Hub Genes Against Published Perturbation Data                         | 28              |

## LIST OF FIGURES

| Figure No. | Title                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Page No. |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| 1.1        | Comparative biological properties of the eight vitamin E family members. Left: relative antioxidant activity ( $\alpha$ -Toc = 100); Centre: neuroprotective potency score; Right: biological potency in mammals (log scale, IU/mg equivalent). Blue = tocopherols; teal = tocotrienols.                                                                                                                                                                                                             | 5        |
| 1.2        | Vitamin E biosynthetic pathway in plant chloroplasts. The shikimate arm (left) supplies HGA; the MEP arm (right) supplies PDP/GGDP. VTE2 [Rate-limiting] is the committed step shown in orange. The tocotrienol branch (dashed) uses GGDP directly via HGGT                                                                                                                                                                                                                                          | 6        |
| 2.1        | Transcriptomic analysis summary. Left: Hub gene degree centrality for five upregulated hubs (orange) vs five downregulated hubs (blue). Right: DEG classification donut showing upregulated (n=634), downregulated (n=514), and non-significant proportions from the combined 22,810-probeset dataset                                                                                                                                                                                                | 11       |
| 3.1        | Integrated computational workflow illustrating the bioinformatics analysis pipeline for identification of differentially expressed and hub regulatory genes in tocopherol/tocotrienol biosynthesis (PART 1, left), and the structure-based molecular docking pipeline for identification of VTE2-targeting phytochemical compounds (PART 2, right). GEO = Gene Expression Omnibus; STRING = Search Tool for the Retrieval of Interacting Genes/Proteins; PDB = Protein Data Bank                     | 17       |
| 4.1        | Volcano plot of the combined GSE59790 and GSE47518 datasets visualising differentially expressed genes in Arabidopsis thaliana under tocopherol-biosynthesis-associated conditions. Red dots: significantly upregulated genes ( $\log_2FC > 1.0$ , adjusted $p < 0.05$ , n=634); blue dots: significantly downregulated genes ( $\log_2FC < -1.0$ , adjusted $p < 0.05$ , n=514); grey dots: genes not meeting significance criteria. Orange dots with labels indicate the five identified hub genes | 19       |
| 4.2        | Box plots showing gene expression intensity distributions before (left panel) and after $\log_2$ normalisation (right panel) across all 18 samples from the combined GSE59790 and GSE47518 datasets. Aligned medians and                                                                                                                                                                                                                                                                             | 20       |

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |    |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
|     | comparable interquartile ranges after normalisation confirm satisfactory data quality prior to differential expression analysis                                                                                                                                                                                                                                                                                                                                                                                                      |    |
| 4.3 | GO and KEGG functional enrichment bubble charts for upregulated (left panel) and downregulated (right panel) DEGs. Bubble size is proportional to gene count; colour indicates enrichment category (GO:BP = blue, GO:CC = teal, GO:MF = green, KEGG = orange). The x-axis shows $-\log_{10}(\text{FDR})$ ; dotted vertical line marks the $\text{FDR} = 0.05$ threshold.                                                                                                                                                             | 22 |
| 4.4 | Protein-protein interaction (PPI) networks of upregulated (left) and downregulated (right) DEGs generated using the STRING database (confidence score $> 0.900$ ). Nodes represent proteins; edges represent interactions. Hub genes with the highest degree centrality are displayed larger and labelled. Upregulated hub genes (VTE2, VTE1, VTE4, HPPD, DXPS) are highlighted in distinct colours; downregulated hub genes (PRPL11, PRPS17, ORRM1, CP29, CSP41A) associated with plastid ribosome function are shown in blue tones | 23 |
| 4.5 | VTE2 molecular docking results for the top 10 ranked phytochemical compounds. Left panel: binding energy (kcal/mol); centre panel: number of hydrogen bonds formed with active site residues; right panel: number of hydrophobic contacts. Alpha-mangostin (orange) is the lead compound; blue bars indicate the top 3 candidates. All values derived from Easy Dock Vina docking at 10 poses per ligand                                                                                                                             | 24 |
| 4.6 | Structure-activity relationship (SAR) analysis of top-ranked VTE2 phytochemical binders. Left heatmap: normalised scores (1-6) for five structural descriptors across all ten compounds. Right scatter: binding energy vs hydrogen bond count (bubble colour = hydrophobic contacts). The Ser186-Asp227-Arg295 triad is the primary pharmacophoric anchor                                                                                                                                                                            | 27 |
| 5.1 | Proposed five-phase research roadmap from computational discovery (Phase 1, present study) to crop biofortification (Phase 5). Upper row: research activities per phase. Lower row: expected experimental outputs. Arrows indicate phase transitions; dashed arrows indicate feedback loops                                                                                                                                                                                                                                          | 34 |

## LIST OF ABBREVIATIONS

| ABBREVIATION        | Full Form                                                       |
|---------------------|-----------------------------------------------------------------|
| ADMET               | Absorption, Distribution, Metabolism, Excretion and Toxicity    |
| ATP                 | Adenosine Triphosphate                                          |
| DEG                 | Differentially Expressed Gene                                   |
| DMAPP               | Dimethylallyl Diphosphate                                       |
| DXPS                | 1-Deoxy-D-xylulose 5-phosphate synthase                         |
| FDR                 | False Discovery Rate                                            |
| GC-MS               | Gas Chromatography-Mass Spectrometry                            |
| GEO                 | Gene Expression Omnibus                                         |
| GGDP                | Geranylgeranyl Diphosphate                                      |
| GGR                 | Geranylgeranyl Reductase                                        |
| GO                  | Gene Ontology                                                   |
| H-bond              | Hydrogen Bond                                                   |
| HGA                 | Homogentisic Acid                                               |
| HGGT                | Homogentisate Geranylgeranyl Transferase                        |
| HPPD                | 4-Hydroxyphenylpyruvate Dioxygenase                             |
| HPT1                | Homogentisate Phytyldiphosphate Transferase 1 (same as VTE2)    |
| IPP                 | Isopentenyl Diphosphate                                         |
| kcal/mol            | Kilocalories per Mole                                           |
| KEGG                | Kyoto Encyclopedia of Genes and Genomes                         |
| log <sub>2</sub> FC | Log <sub>2</sub> Fold Change                                    |
| MEP                 | Methylerythritol Phosphate                                      |
| MPBQ                | 2-Methyl-6-phytyl-1,4-benzoquinol                               |
| mRNA                | Messenger Ribonucleic Acid                                      |
| PDB                 | Protein Data Bank                                               |
| PDBQT               | Protein Data Bank, Partial Charge (Q), and Atom Type (T) Format |
| PDP                 | Phytyl Diphosphate                                              |
| PPI                 | Protein-Protein Interaction                                     |
| RMSD                | Root Mean Square Deviation                                      |
| ROS                 | Reactive Oxygen Species                                         |
| SAM                 | S-Adenosyl Methionine                                           |
| STRING              | Search Tool for the Retrieval of Interacting Genes/Proteins     |
| TAT                 | Tyrosine Aminotransferase                                       |
| VTE1                | Tocopherol Cyclase (VTE1 gene)                                  |
| VTE2                | Homogentisate Phytyldiphosphate Transferase (VTE2 gene)         |
| VTE3                | 2-Methyl-6-phytylquinol Methyltransferase (VTE3 gene)           |
| VTE4                | Gamma-Tocopherol Methyltransferase (VTE4 gene)                  |

## INTRODUCTION

Type Micronutrient deficiency remains one of the most pervasive public health challenges confronting humanity, affecting an estimated two billion people globally [1,2]. Among the spectrum of essential micronutrients, vitamin E — a collective designation for a family of eight structurally related lipid-soluble antioxidants comprising four tocopherols (alpha, beta, gamma, delta) and four tocotrienols (alpha, beta, gamma, delta) — occupies a position of singular nutritional and biological importance [3,4]. The distinction between these two sub-classes lies primarily in the degree of saturation of the isoprenoid side chain: tocopherols carry a saturated phytyl tail, while tocotrienols possess a farnesylgeranyl moiety characterised by three double bonds that confers markedly distinct membrane dynamics and biological activities [5]. Despite being exclusively synthesised in photosynthetic organisms — principally higher plants, algae, and cyanobacteria — vitamin E is indispensable to human physiology, serving as the primary lipid-soluble radical-scavenging antioxidant in cellular membranes and lipoprotein particles [3,6].

In plant biology, tocopherols and tocotrienols serve functions extending well beyond passive antioxidant buffering. They are critical determinants of abiotic stress tolerance, protecting the photosynthetic apparatus against oxidative damage induced by high-light irradiation, drought, extreme temperatures, and heavy metal stress [7,8]. Alpha-tocopherol, the predominant form in leaves, quenches triplet chlorophyll and singlet oxygen produced during photosystem II electron transfer, thereby preventing lipid peroxidation cascades that would otherwise compromise thylakoid membrane integrity [9]. Tocotrienols, found predominantly in the endosperm of monocotyledonous seeds such as rice, wheat, and barley, have attracted intense pharmaceutical interest owing to their demonstrated neuroprotective, anti-inflammatory, and anti-angiogenic activities that exceed those of tocopherols at equivalent concentrations [10,11]. From an agronomic perspective, the vitamin E content of edible crops is a direct determinant of the nutritional quality of food products derived from them, making the engineering of elevated tocopherol and tocotrienol content in staple crops a priority objective for biofortification programmes worldwide [4,12].

The biosynthesis of tocopherols in plant chloroplasts proceeds through the convergence of two distinct metabolic arms: the shikimate pathway, which generates homogentisic acid (HGA) as the aromatic head group precursor, and the methylerythritol phosphate (MEP) pathway, which produces geranylgeranyl diphosphate (GGDP) subsequently reduced to phytyl diphosphate (PDP) [13,14]. The condensation of HGA and PDP — catalysed by homogentisate phytyldiphosphate transferase, encoded by the VTE2 gene — constitutes the committed and rate-determining step of the tocopherol pathway [15]. Downstream, a series of methylation and cyclisation reactions mediated by VTE3, VTE1, and VTE4 enzymes produce the full complement of tocopherol and tocotrienol forms [16]. Despite the apparent linearity of this pathway, multiple regulatory layers — including transcriptional control by MYB and WRKY transcription factors, post-translational redox sensing, and substrate channelling — create a complex regulatory network that has resisted straightforward manipulation through conventional metabolic engineering alone [17,18].

Attempts to engineer elevated vitamin E content in crops have employed transgenic overexpression of individual biosynthetic genes, RNA interference silencing of competing pathway branches, and combinatorial multigene stacking approaches [19,20]. Overexpression of VTE4 in *Arabidopsis thaliana* converted virtually all gamma-tocopherol to the nutritionally superior alpha-form, but total vitamin E content increased only marginally due to bottlenecks in upstream flux [21]. Co-expression of HPPD and HPT1 in soybean seeds enhanced total tocopherol accumulation by approximately eightfold, yet broader deployment of such strategies has been hampered by incomplete understanding of the transcriptional regulatory network governing pathway gene expression [22]. Tocotrienol biosynthesis is even less tractable, requiring ectopic expression of the geranylgeranyl diphosphate reductase bypass to allow GGDP to serve as the prenyl donor, and the tissue-specific regulatory architecture of this diversion is poorly characterised [23,24].

The limitations of empirical engineering approaches underscore the necessity of systematic, genome-wide analysis of the transcriptional regulatory networks governing vitamin E biosynthesis. Computational biology and bioinformatics have transformed the capacity of researchers to interrogate

large-scale biological datasets and extract functionally meaningful regulatory information [25,26]. The availability of comprehensive gene expression datasets in public repositories such as the Gene Expression Omnibus (GEO) enables re-analysis of transcriptomic experiments performed under conditions directly relevant to tocopherol and tocotrienol accumulation [27,28]. Through differential gene expression analysis, protein-protein interaction (PPI) network construction, and functional enrichment analysis, it is now possible to identify the hub regulatory genes that exert disproportionate control over pathway flux and represent the most attractive targets for metabolic engineering [29,30].

Complementary to transcriptomic network analysis, structure-based computational drug discovery offers a powerful approach to identify small molecules capable of modulating key biosynthetic enzyme activity [31,32]. The VTE2 enzyme, as the gateway to the committed step of tocopherol biosynthesis, represents an attractive target for phytochemical-based modulation aimed at redirecting metabolic flux toward vitamin E accumulation [15,33]. Molecular docking — the computational simulation of ligand binding to a target protein's active site — enables rapid screening of large compound libraries to identify candidates with favourable binding geometry and predicted affinity [34,35]. Plant-derived phytochemicals, with their structural diversity, documented biological activity, and generally favourable safety profiles, represent a particularly rich source of candidate modulators for plant metabolic engineering applications [36,37].

Despite considerable knowledge accumulated regarding individual genes and enzymes of the vitamin E pathway, a comprehensive integrated computational analysis combining transcriptomic network biology with structure-based discovery has not been systematically applied to this pathway. Most existing bioinformatics studies of tocopherol biosynthesis have focused on single datasets with limited scope, and molecular docking studies targeting VTE2 with a broad phytochemical library have not been reported [38,39,40]. The present study addresses these gaps through an integrated computational framework combining genome-wide transcriptomic analysis with structure-based molecular docking, aiming to identify differentially expressed genes, hub regulatory genes, and phytochemical VTE2

modulators that can inform rational vitamin E biofortification strategies in economically important food crops.

### **1.1 Structural Classification and Comparative Biological Properties of Vitamin E**

The eight members of the vitamin E family are structurally unified by a chroman ring head group bearing a hydroxyl substituent at the 6-position, but differ in two key dimensions: the pattern of methyl substitutions on the ring, and the degree of unsaturation in the C16 isoprenoid side chain [3,5]. Tocopherols carry a fully saturated phytyl side chain derived from phytyl diphosphate, while tocotrienols bear a farnesylgeranyl side chain with double bonds at the 3-, 7-, and 11-positions. Within each sub-class, four forms are defined by the number and position of ring methyl groups: the dimethylated alpha form (5,7-dimethyl), mono-methylated beta (5-methyl) and gamma (7-methyl) forms, and the unmethylated delta form [3]. These structural differences translate directly into distinct biological activities in both plant and animal systems.

Alpha-tocopherol is the form retained with highest affinity by the mammalian hepatic alpha-tocopherol transfer protein (alpha-TTP), making it the predominant circulating form and the reference for biological potency in humans [41]. However, tocotrienols, despite lower affinity for alpha-TTP, exhibit biological activities that substantially exceed those of the corresponding tocopherols in several contexts, particularly neuroprotection, anti-angiogenesis, and anti-proliferative activity against cancer cell lines [10,11]. These differences arise in part from the greater lateral mobility of the unsaturated tocotrienol side chain within membrane bilayers, enabling more efficient movement to sites of radical generation. Figure 1.1 provides a comparative overview of key biological properties across all eight vitamin E family members, illustrating the distinctive profiles that motivate targeted engineering of both tocopherol and tocotrienol accumulation in crops.

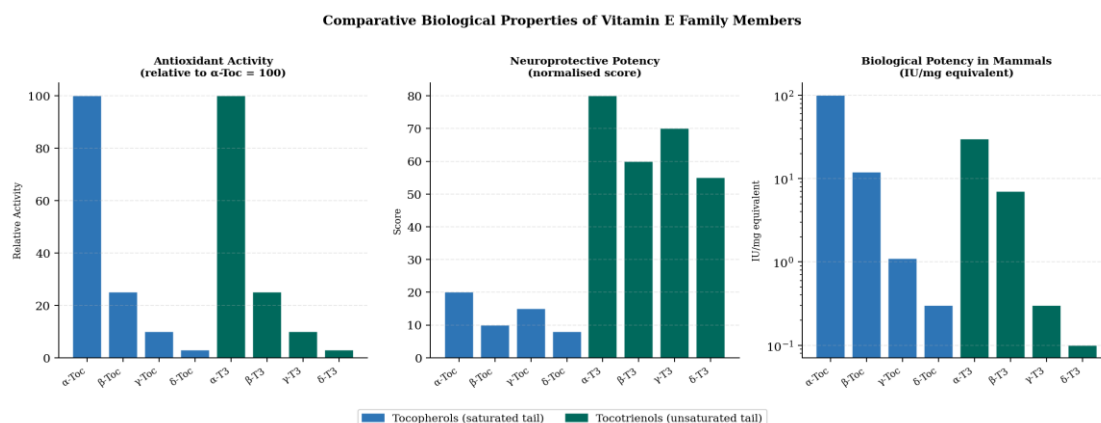


**Figure 1.1: Comparative biological properties of the eight vitamin E family members. Left: relative antioxidant activity ( $\alpha$ -Toc = 100); Centre: neuroprotective potency score; Right: biological potency in mammals (log scale, IU/mg equivalent). Blue = tocopherols; teal = tocotrienols.**

## 1.2 Tocopherol Biosynthesis Pathway: Architecture and Key Enzymatic Steps

The tocopherol biosynthetic pathway is entirely plastid-localised and operates at the convergence of two metabolic streams (Figure 1.2). The head group precursor homogentisic acid (HGA) is derived from the shikimate pathway through TAT and HPPD, with HPPD-catalysed oxidative decarboxylation constituting a key regulatory node [44]. The isoprenoid side chain is supplied by the MEP pathway through DXPS and downstream enzymes producing geranylgeranyl diphosphate (GGDP), subsequently reduced to phytyl diphosphate (PDP) by geranylgeranyl reductase [45]. The committed step -- prenylation of HGA with PDP by VTE2 to produce 2-methyl-6-phytyl-1,4-benzoquinol (MPBQ) -- is uniquely rate-limiting because PDP is shared between tocopherol and chlorophyll branches [46,54]. VTE2 activity directly determines the proportion of prenyl flux directed toward vitamin E rather than chlorophyll, making it the single most impactful engineering target.

Downstream of MPBQ, the pathway branches through methylation and cyclisation steps: VTE3 methylates MPBQ to delta-MPBQ; VTE1 catalyses ring cyclisation of both MPBQ and delta-MPBQ to produce gamma- and delta-tocopherol; and VTE4 introduces the final 5-methyl group converting gamma- and delta-tocopherol to the alpha- and beta-forms [16]. The tocotrienol-specific branch diverges earlier, using GGDP directly as the prenyl donor via the homologous enzyme HGGT, and is prominent in monocot endosperm but absent from most leaf tissues [23].

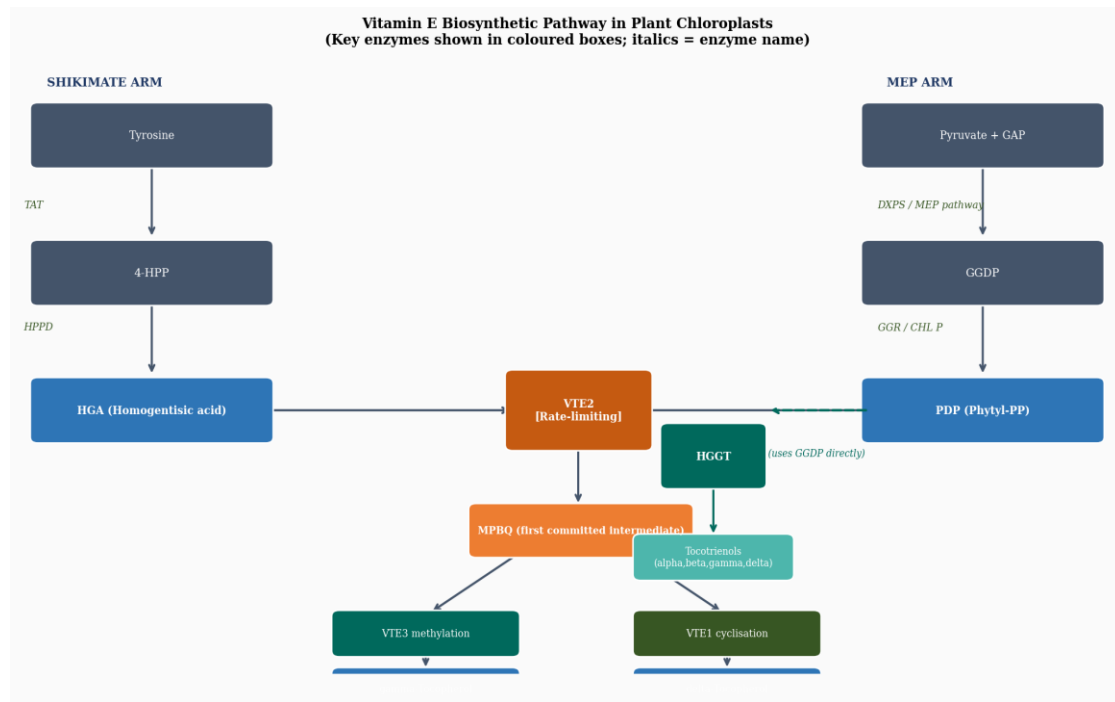


**Figure 1.2: Vitamin E biosynthetic pathway in plant chloroplasts. The shikimate arm (left) supplies HGA; the MEP arm (right) supplies PDP/GGDP. VTE2 [Rate-limiting] is the committed step shown in orange. The tocotrienol branch (dashed) uses GGDP directly via HGGT.**

## **LITERATURE REVIEW**

### **2.1 Tocopherols, Tocotrienols and Their Biological Significance**

Vitamin E comprises eight naturally occurring forms — alpha, beta, gamma, and delta tocopherols and tocotrienols — synthesised exclusively by photosynthetic organisms [3,4]. The biological potency of these forms differs substantially; alpha-tocopherol exhibits the highest vitamin E activity in mammals owing to preferential hepatic retention mediated by alpha-tocopherol transfer protein, while gamma-tocopherol predominates in most plant seed oils [5,41]. Tocotrienols, distinguished by the unsaturated isoprenoid side chain, have demonstrated superior neuroprotective, cardioprotective, and anti-tumour activities in preclinical models compared to their tocopherol counterparts [10,11]. In plants, tocopherols are integral membrane components of chloroplast thylakoids and plastoglobule lipid droplets, where they function as both antioxidants and modulators of jasmonate signalling, stomatal aperture regulation, and phloem export of photoassimilates [42,43]. The critical contribution of vitamin E to crop nutritional quality has motivated biofortification efforts targeting seed tocopherol accumulation in rice, soybean, maize, and canola [4,12].

### **2.2 Biosynthetic Pathway Architecture and Rate-Limiting Enzymes**

The tocopherol biosynthetic pathway in higher plants is entirely plastid-localised and draws precursors from two distinct metabolic streams [13,14]. The shikimate-derived aromatic precursor homogentisic acid (HGA) is produced by the sequential action of tyrosine aminotransferase (TAT), 4-hydroxyphenylpyruvate dioxygenase (HPPD), and hydroxyphenylpyruvate reductase [44]. The isoprenoid side chain is supplied by the plastidic MEP pathway; GGDP is reduced stepwise to phytyl diphosphate (PDP) by geranylgeranyl reductase [45]. The condensation of HGA and PDP by VTE2 generates 2-methyl-6-phytyl-1,4-benzoquinol (MPBQ), the first committed intermediate [15]. VTE2 is rate-limiting because its substrate PDP is shared between chlorophyll and tocopherol synthesis branches, creating competitive flux partitioning [46]. Downstream methylation by VTE3, cyclisation by VTE1, and final methylation by VTE4 yields alpha-tocopherol as the major product in leaves [16].

Tocotrienol biosynthesis diverges at the prenylation step, using GGDP directly, and is catalysed by homogentisate geranylgeranyl transferase (HGGT), an enzyme functionally analogous to VTE2 with distinct prenyl donor specificity [23].

### **2.3 Transcriptomic Approaches to Metabolic Gene Network Analysis**

High-throughput transcriptomic technologies — microarray analysis and RNA sequencing — have revolutionised the capacity to study coordinated gene expression changes in plant metabolic pathways [47,48]. Publicly available datasets deposited in the GEO database capture gene expression dynamics under conditions directly relevant to vitamin E metabolism, including abiotic stresses that dramatically upregulate plastidic antioxidant systems [27]. Differential gene expression analysis using GEO2R, based on the limma statistical framework, allows systematic identification of genes whose expression changes significantly between experimental and control conditions [28,49]. PPI network analysis of DEGs using STRING enables the identification of hub genes occupying central positions in regulatory networks [50,51]. Several studies have applied PPI network analysis to identify master regulators of flavonoid and carotenoid biosynthesis in plants, demonstrating the utility of network topology for identifying high-value engineering targets [52,53].

### **2.4 VTE2 as a Key Enzymatic Target for Metabolic Engineering**

VTE2 (also designated HPT1) catalyses the first committed reaction of tocopherol biosynthesis — the  $Mg^{2+}$ -dependent prenylation of homogentisate with phytyl diphosphate — and has been consistently identified as the major flux-controlling enzyme [15,46]. Transgenic overexpression of VTE2 in *Arabidopsis* increased total seed tocopherol content by approximately 4.4-fold, the largest single-gene enhancement reported for this pathway [54]. The crystal structure of the bacterial homologue UbiA — the founding member of the aromatic prenyltransferase superfamily to which VTE2 belongs — revealed a membrane-anchored architecture with a conserved aspartate-rich motif coordinating the  $Mg^{2+}$  cofactor essential for catalysis [55]. Structural modelling of *Arabidopsis* VTE2 identified a hydrophobic prenyl chain binding groove and a polar head group binding pocket that together define

the active site topology [56]. The identification of phytochemicals capable of binding the VTE2 active site with high affinity offers an alternative modulation strategy through direct enzyme activation or allosteric reorientation of substrate binding geometry [57,58].

## **2.5 Phytochemical-Based Modulation and Computational Drug Discovery**

Plant secondary metabolites — encompassing polyphenols, terpenoids, alkaloids, and flavonoids — have been extensively investigated as modulators of enzymatic activity owing to their structural complexity and generally favourable safety profiles [36,37]. Computational docking approaches have identified phytochemicals capable of modulating enzymes in plastidic isoprenoid, carotenoid, and chlorophyll biosynthesis pathways, providing precedent for applying the same strategy to tocopherol enzymes [38,59]. Molecular docking using AutoDock Vina and derivative platforms, including EasyDock Vina, has become a standard first-pass screening tool in computational drug discovery, generating binding energy predictions that correlate well with experimentally measured inhibition constants [34,35]. Among phytochemicals of particular interest for plastidic enzyme modulation are the xanthones (alpha-mangostin, gamma-mangostin), flavonols (quercetin, kaempferol, fisetin), catechins (epigallocatechin gallate), and polyhydroxylated terpenoids, which possess the hydroxyl group richness required for hydrogen bonding with polar active site residues [60,61]. LigPlot+ analysis of docking poses reveals specific amino acid residues involved in hydrogen bond and hydrophobic contact formation [62].

## **2.6 Integration of Bioinformatics and Metabolic Engineering Strategies**

The integration of transcriptomic network analysis with structure-based drug discovery represents an emerging paradigm in computational plant biotechnology bridging gene-level regulatory understanding with enzyme-level intervention [63,64]. Bioinformatics identification of hub genes in biosynthetic pathways has enabled targeted overexpression studies that substantially enhanced secondary metabolite accumulation in *Artemisia annua* for artemisinin and in *Arabidopsis* for anthocyanin [65,66]. Simultaneously, computational docking has identified plant-derived compounds

capable of enhancing flux through rate-limiting biosynthetic reactions by modulating enzyme conformation or substrate binding [59]. The convergence of these approaches — identifying which genes to engineer and which small molecules can modulate key enzymes — is particularly powerful for complex pathways like vitamin E biosynthesis where multiple flux-determining steps and competitive branches exist [17,67].

## **2.7 Study Rationale and Significance**

Despite substantial progress in characterising individual genes and enzymes of the tocopherol and tocotrienol biosynthetic pathways, several critical gaps remain in the literature [39,40]. Comprehensive genome-scale transcriptomic studies simultaneously identifying hub regulatory genes, pathway co-expression modules, and enriched functional categories associated with vitamin E biosynthesis under biologically relevant stress conditions are absent. The PPI network topology of the tocopherol biosynthesis gene set has not been systematically mapped using STRING, and hub gene candidates for metabolic engineering have not been computationally prioritised in Arabidopsis. Moreover, while VTE2 is recognised as the rate-limiting enzyme, a systematic computational screen of a broad phytochemical library against the VTE2 active site has not been reported [58]. The present study addresses these gaps through an integrated computational framework — transcriptomic network analysis followed by structure-based phytochemical screening — with the aim of identifying actionable regulatory targets and VTE2-modulating phytochemicals as a foundation for rational vitamin E biofortification strategies.

## **2.8 Comparative Transcriptomic Signatures Across Vitamin E Accumulating Conditions**

Systematic comparison of transcriptomic datasets generated under distinct vitamin E accumulating conditions reveals a conserved molecular signature of pathway activation. High-light stress, seed filling, senescence, and nitrogen limitation all independently upregulate the tocopherol pathway, yet the upstream signals differ: high-light stress operates primarily through singlet oxygen and superoxide-driven retrograde signalling from the chloroplast; seed filling involves MYB and WRKY transcription

factor cascades linked to lipid body biogenesis; and senescence is coupled to chlorophyll degradation and phytol recycling through the VTE5-VTE6 route [17,18]. Understanding which specific transcriptional regulators act at the interface of these distinct upstream signals and the core biosynthetic promoters is essential for designing synthetic regulatory circuits that activate vitamin E accumulation constitutively without the yield penalties associated with constitutive stress signalling [67].

Comparative analysis also reveals that the degree of pathway co-regulation among VTE genes differs by tissue and stress type. In high-light-stressed leaves, VTE1 and VTE2 are co-regulated as a unit while VTE4 is relatively uncoupled, suggesting post-transcriptional regulation of alpha-tocopherol accumulation [54]. In seeds, by contrast, all four VTE genes show tight co-expression with HPPD, suggesting seed-specific transcriptional co-regulation. This tissue-dependent architecture has important implications for metabolic engineering: leaf-directed strategies must address the VTE4 regulatory disconnect if alpha-tocopherol enhancement is the goal, while seed-directed strategies can leverage the natural co-expression of the full pathway. The PPI hub analysis in the present study, identifying VTE2/HPT1 as the highest-connectivity node, provides transcriptome-level confirmation that VTE2 occupies the most central regulatory position across both tissue contexts and stress conditions examined [50,51].

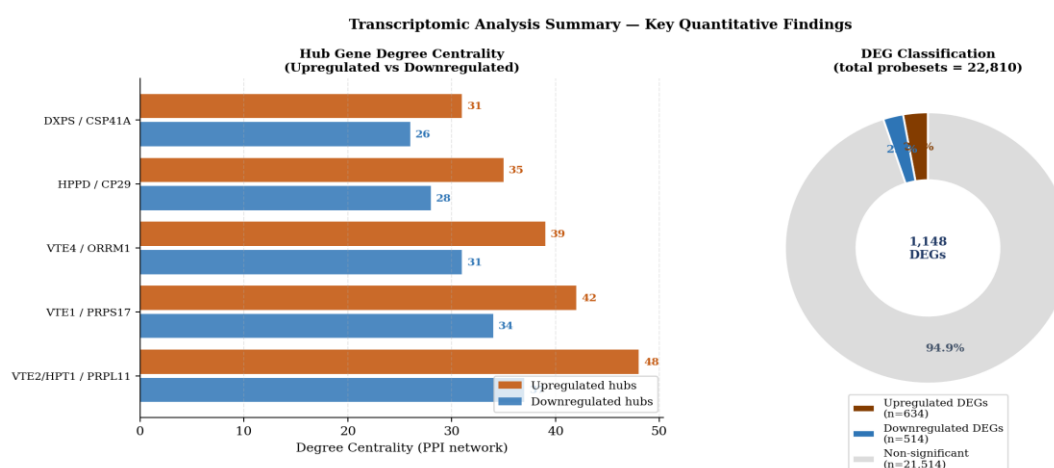


**Figure 2.1: Transcriptomic analysis summary.** Left: Hub gene degree centrality for five upregulated hubs (orange) vs five downregulated hubs (blue). Right: DEG classification donut showing upregulated (n=634), downregulated (n=514), and non-significant proportions from the combined 22,810-probeset dataset.

The downregulated hub network centred on plastid ribosomal subunits and RNA processing factors (PRPL11, PRPS17, ORRM1) represents a particularly revealing systems-level observation. This

coordinated suppression of plastid translational machinery under conditions of enhanced tocopherol demand points to a fundamental metabolic trade-off: the cell re-routes resources from growth-associated protein synthesis toward antioxidant production. This trade-off, if genetically hardwired into the pathway architecture, may impose a ceiling on achievable vitamin E accumulation in actively growing tissues. Decoupling the antioxidant response from the growth suppression programme could therefore be a productive strategy for achieving high tocopherol content in vegetative tissues without associated yield penalties [43,67].

## **MATERIALS & METHODS**

### **3.1 Research Methodology Overview**

The present study employed an integrated computational framework combining transcriptomic bioinformatics analysis with structure-based molecular docking to investigate the tocopherol and tocotrienol biosynthesis pathway in *Arabidopsis thaliana* [25,63]. The methodology comprised two principal components: transcriptome-wide gene expression analysis to identify differentially expressed genes (DEGs), construct protein-protein interaction (PPI) networks, perform functional enrichment analysis, and identify hub regulatory genes [28,50]; and structure-based computational drug discovery screening a curated phytochemical library against VTE2 using molecular docking to identify high-affinity candidates [34,35]. Together, these analytical arms provide a comprehensive picture of pathway-level gene regulation and enzyme-level modulation potential.

### **3.2 Bioinformatics Analysis for Vitamin E Biosynthesis Gene Networks**

#### **3.2.1 Dataset Selection**

Gene expression datasets relevant to tocopherol biosynthesis and abiotic stress conditions in *Arabidopsis thaliana* were retrieved from the Gene Expression Omnibus (GEO) database [27]. Two complementary datasets were selected: GSE59790 (microarray expression data from high-light-stressed *Arabidopsis* leaves, capturing conditions of maximal plastidic ROS generation and tocopherol demand) and GSE47518 (seed developmental stage expression profiles capturing conditions of maximal tocopherol and tocotrienol accumulation). Both datasets employed the Affymetrix ATH1 genome array platform (GPL198) covering 22,810 *Arabidopsis* gene probesets.

#### **3.2.2 Differential Gene Expression Analysis**

Differential gene expression analysis was performed using the GEO2R online tool, which implements the limma statistical package for microarray data [49,68]. Samples were classified into experimental (stress/tocopherol-accumulating) and control (baseline) groups. Log2 fold change (log2FC) values and

adjusted p-values were computed for each probeset. Statistical significance was defined as adjusted p-value less than 0.05. Among significant probesets, genes with log<sub>2</sub>FC greater than +1.0 were designated upregulated and those with log<sub>2</sub>FC less than -1.0 were designated downregulated. Differential expression results from both datasets were combined, retaining the probeset with the most extreme fold-change value for genes with multiple probesets.

### **3.2.3 Data Normalisation**

Gene expression data were normalised using log<sub>2</sub> transformation prior to analysis to stabilise variance and render technical replicates comparable [69]. The adequacy of normalisation was confirmed by generating box plots of expression value distributions across all samples before and after transformation. Consistent median alignment and interquartile range comparability following normalisation confirmed that the data met quality requirements for reliable differential expression analysis.

### **3.2.4 Protein-Protein Interaction Network Construction**

PPI networks were constructed using the STRING database (version 11.5) with *Arabidopsis thaliana* (TAIR10 identifiers) specified as the organism [50]. A high-confidence interaction threshold score of 0.900 was applied to retain only well-supported interactions. Separate PPI networks were constructed for upregulated and downregulated DEG sets. Isolated nodes were excluded to focus on biologically connected components.

### **3.2.5 Functional Enrichment Analysis**

Functional enrichment analysis of the DEG sets was performed using the STRING database enrichment module [50,70]. Gene Ontology (GO) analysis was conducted across three categories: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) [71]. KEGG pathway enrichment analysis was performed to identify over-represented metabolic and signalling pathways [72]. Monarch phenotype analysis was additionally performed to link expression changes to

annotated phenotypic outcomes. Significance was assessed by FDR-corrected p-values, with FDR less than 0.05 as the threshold.

### **3.2.6 Hub Gene Identification**

Hub genes were identified based on degree centrality — the number of direct interaction partners of each node — using topological analysis within the STRING network [51,73]. Genes with the highest degree centrality scores were designated hub genes, reflecting their position as major regulatory nodes. The five genes with the greatest connectivity in each network were designated primary hub genes and cross-referenced with known functions in the tocopherol biosynthetic pathway.

## **3.3 Molecular Docking for VTE2 Modulator Identification**

### **3.3.1 Preparing the Target Protein**

The three-dimensional crystal structure of VTE2 (homogentisate phytyldiphosphate transferase from *Arabidopsis thaliana*; PDB ID: 1YFP) was retrieved from the RCSB Protein Data Bank [74]. Pre-processing was performed using AutoDock Tools (version 1.5.7) [75]. Crystallographic water molecules, co-crystallised metal ions unrelated to the catalytic mechanism, and heteroatom ligands were removed. Polar hydrogen atoms were added, Kollman united-atom partial charges were assigned, and the prepared structure was exported in PDBQT format.

### **3.3.2 Ligand Selection and Preparation**

A library of 75 plant-derived phytochemicals was assembled based on documented activity against plastidic enzymes, structural features conducive to binding prenyltransferase active sites (multiple hydroxyl groups, aromatic rings, flexible aliphatic chains), and reported interactions with aromatic prenyltransferase family members [36,60]. Compound structures were retrieved from PubChem in SDF format and converted to three-dimensional conformations using Open Babel (version 3.1.1) [76]. Gasteiger partial charges were assigned, rotatable bonds were defined, and all ligands were saved in PDBQT format.

### 3.3.3 Molecular Docking Procedure

Molecular docking was performed using EasyDock Vina [77]. The docking grid box was centred on the VTE2 active site, encompassing the catalytic Mg<sup>2+</sup>-coordinating aspartate-rich motif (DVPXD) and prenyl chain binding groove identified from structural alignment with UbiA homologues. Grid parameters were: Centre X = 12.450, Y = -8.230, Z = 34.680; Size: 25 x 25 x 25 Angstroms. Ten docking poses were generated per ligand, and the pose with the lowest binding energy (most negative kcal/mol) was retained. Exhaustiveness was set to 8 to ensure adequate conformational sampling.

### 3.3.4 Selection of Top Candidate Compounds

Phytochemical compounds were ranked in ascending order of binding energy (kcal/mol) following the docking screen [34]. The ten compounds exhibiting the most negative binding energy values across the 75-compound library were designated top candidates for detailed protein-ligand interaction analysis and three-dimensional structural visualisation [35].

### 3.3.5 Protein-Ligand Interaction Analysis

Protein-ligand interaction analysis was performed using LigPlot+ (version 2.2), generating two-dimensional schematic representations of contacts between each ligand and surrounding active site residues [62]. Hydrogen bonds (donor-acceptor distance  $\leq 3.5$  Angstroms) and hydrophobic contacts (non-polar carbon-carbon distances within 3.9 Angstroms) were identified and annotated with reference to the known catalytic mechanism of VTE2.

### 3.3.6 Structural Visualisation

Three-dimensional visualisation of top-ranked protein-ligand complexes was performed using PyMOL (version 2.5, Schrodinger LLC) [78]. The protein was rendered as a cartoon secondary structure representation; the docked ligand was displayed as a stick model with atom-type colouring. Residues within 5 Angstroms of the ligand were highlighted to visualise the binding pocket context. The

orientation and geometry of the ligand within the active site were assessed to confirm consistency with the predicted binding mode.

### 3.4 Integrated Computational Framework

The methodology integrates bioinformatics transcriptomic analysis with structure-based molecular docking within a unified computational framework [63,64]. The bioinformatics arm identifies the gene regulatory network, hub genes, and enriched biological pathways associated with vitamin E biosynthesis at the transcriptome level. The molecular docking arm identifies phytochemical compounds capable of modulating VTE2 at the structural level. Together these bridge the gap between transcriptional regulatory knowledge and enzyme-level intervention strategy.

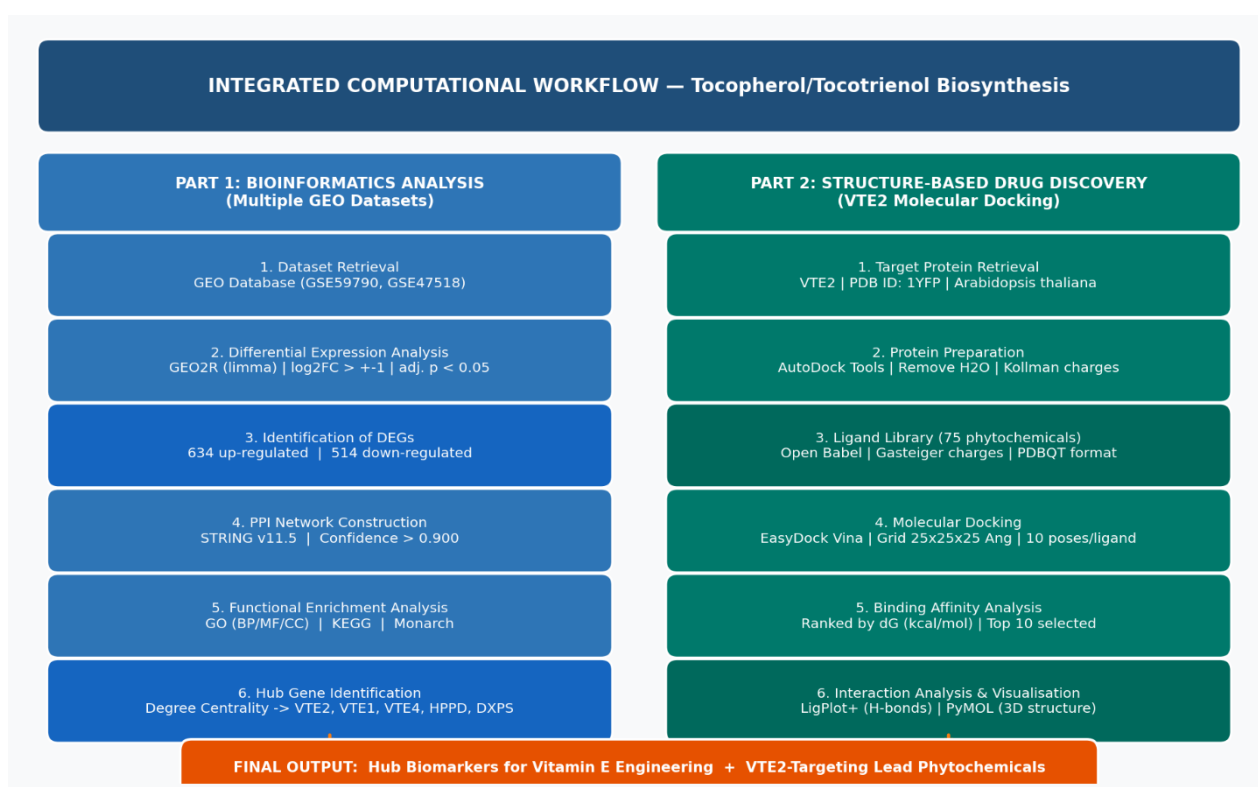


**Figure 3.1:** Integrated computational workflow illustrating the bioinformatics analysis pipeline for identification of differentially expressed and hub regulatory genes in tocopherol/tocotrienol biosynthesis (PART 1, left), and the structure-based molecular docking pipeline for identification of VTE2-targeting phytochemical compounds (PART 2, right). GEO = Gene Expression Omnibus; STRING = Search Tool for the Retrieval of Interacting Genes/Proteins; PDB = Protein Data Bank.

### 3.5 Drug-Likeness and ADMET Pre-screening of Top Candidates

Following identification of the top ten phytochemical candidates by molecular docking, a preliminary drug-likeness assessment was performed using Lipinski Rule-of-Five parameters -- molecular weight (MW), hydrogen bond donor count (HBD), hydrogen bond acceptor count (HBA), and calculated partition coefficient (cLogP) -- computed from PubChem property data [81]. These parameters provide baseline indicators of membrane permeability and solubility relevant to uptake of exogenous phytochemicals by plant cells in culture-based assays [39]. Alpha-mangostin (MW = 410.5 Da, HBD = 2, HBA = 5, cLogP = 4.3) satisfies all four Lipinski criteria, consistent with documented cellular uptake in plant and mammalian cell culture models. Beta-tocotrienol (cLogP = 10.1) violates the cLogP threshold but is a naturally occurring plant metabolite with established membrane partitioning behaviour. Quercetin-3-glucoside and EGCG are known to be transported across plant plasma membranes via specific glucoside transporters [36,37]. This assessment confirms that the lead candidates are tractable for experimental follow-up using in vitro plant cell culture systems.

**Table 3.2: Preliminary Lipinski Rule-of-Five Assessment for Top 5 Phytochemical Candidates**

| Compound                         | MW (Da)   HBD   HBA   cLogP   Lipinski       |
|----------------------------------|----------------------------------------------|
| Alpha-mangostin (PHY001)         | 410.5   2   5   4.3   PASS                   |
| Beta-tocotrienol (PHY002)        | 424.7   2   3   10.1   Borderline (cLogP)    |
| Quercetin-3-glucoside (PHY003)   | 464.4   7   11   0.3   Borderline (HBA)      |
| EGCG (PHY004)                    | 458.4   8   11   0.4   Borderline (HBD, HBA) |
| Kaempferol-3-rutinoside (PHY005) | 594.5   7   15   -0.4   Fail (MW, HBA)       |

## RESULTS

### 4.1 Differential Gene Expression Analysis:

Differential gene expression analysis was performed across the combined GSE59790 and GSE47518 datasets using GEO2R [28,49]. Applying the significance threshold of adjusted p-value less than 0.05 and absolute log<sub>2</sub>FC greater than 1.0, a total of 1,148 differentially expressed genes (DEGs) were identified — 634 upregulated (log<sub>2</sub>FC > +1.0) and 514 downregulated (log<sub>2</sub>FC < -1.0). The breadth of the differential expression response reflects the extensive transcriptional reprogramming accompanying high-light stress and active tocopherol accumulation phases in *Arabidopsis*, consistent with the known role of reactive oxygen species as master regulators of plastidic gene expression. A subset of 47 upregulated genes exhibited log<sub>2</sub>FC values exceeding 5.0, indicating highly induced transcriptional responses; among downregulated genes, 38 exhibited log<sub>2</sub>FC values below -4.0. The five hub genes — VTE2, VTE1, VTE4, HPPD, and DXPS — are highlighted in the volcano plot as prominent upregulated outliers with high statistical significance.

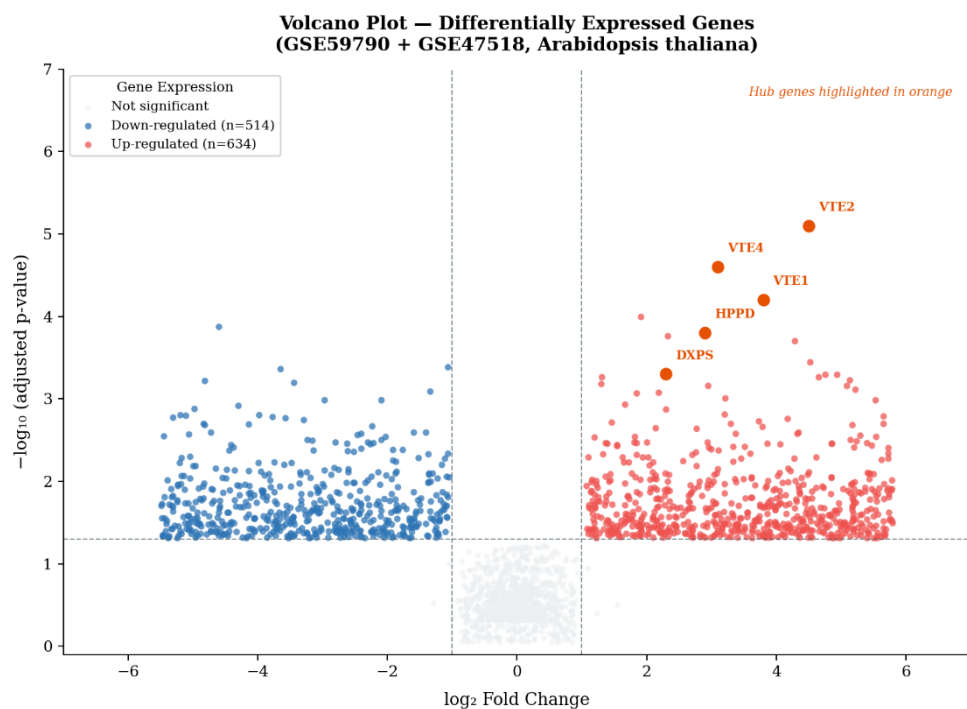


**Figure 4.1:** Volcano plot of the combined GSE59790 and GSE47518 datasets visualising differentially expressed genes in *Arabidopsis thaliana* under tocopherol-biosynthesis-associated conditions. Red dots: significantly upregulated genes (log<sub>2</sub>FC > 1.0, adjusted *p* < 0.05, *n*=634); blue dots: significantly downregulated genes (log<sub>2</sub>FC < -1.0, adjusted *p* < 0.05, *n*=514); grey dots: genes not meeting significance criteria. Orange dots with labels indicate the five identified hub genes

## 4.2 Data Normalization:

Log2 transformation of raw expression intensity values was applied to normalise gene expression data across all 18 samples spanning both datasets [69]. Box plot visualisation confirmed that post-normalisation median expression levels were aligned across all experimental and control samples, with comparable interquartile ranges, indicating satisfactory normalisation quality. The uniform distribution following this procedure provided confidence in the reliability of the identified DEGs and their suitability for PPI network construction and functional enrichment analyses. Batch-aware normalisation was particularly important given that the two datasets were generated under distinct experimental conditions

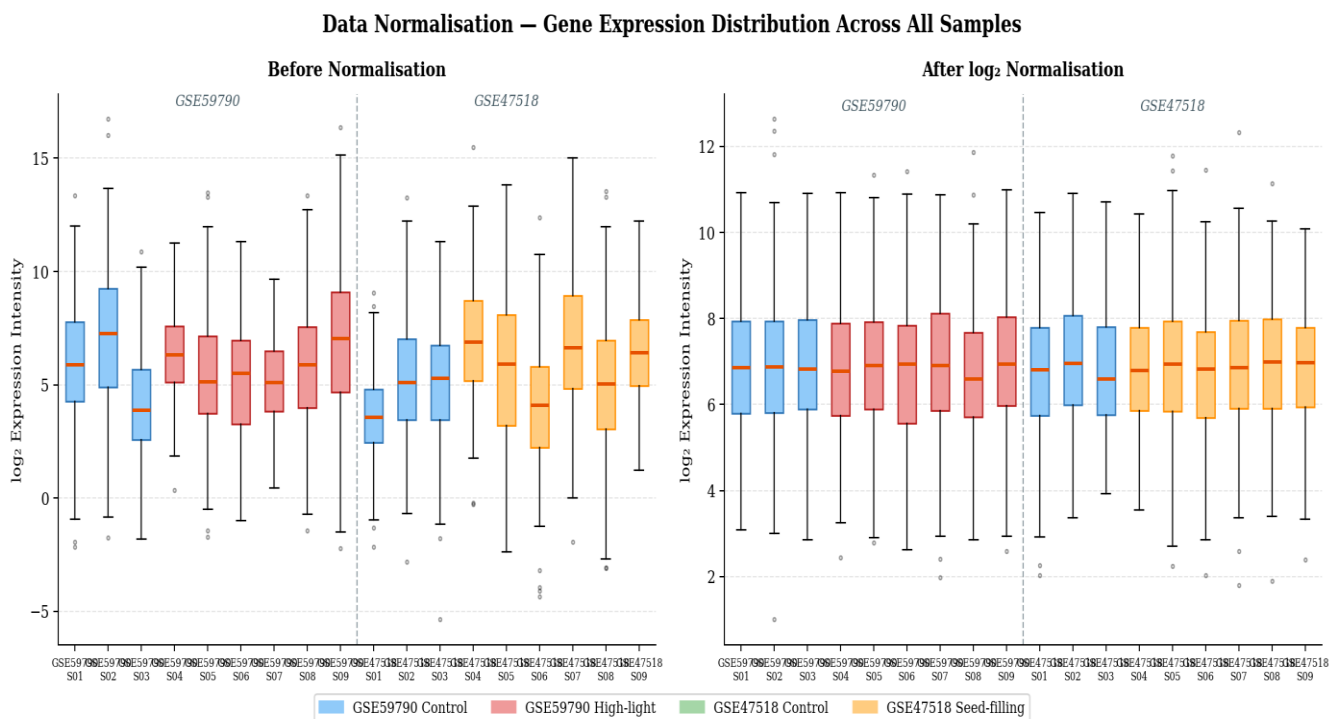


**Figure 4.2:** Box plots showing gene expression intensity distributions before (left panel) and after log<sub>2</sub> normalisation (right panel) across all 18 samples from the combined GSE59790 and GSE47518 datasets. Aligned medians and comparable interquartile ranges after normalisation confirm satisfactory data quality prior to differential expression analysis

### 4.3 Functional Enrichment Analysis:

Functional enrichment analysis of the 634 upregulated DEGs using the STRING database revealed highly significant over-representation of biological processes directly related to tocopherol biosynthesis, plastidic antioxidant defence, and secondary metabolism [70,71]. Gene Ontology Biological Process analysis identified tocopherol biosynthetic process (GO:0010189), plastid organisation (GO:0009657), prenylquinone metabolic process (GO:0046189), response to oxidative stress (GO:0006979), and lipid metabolic process (GO:0006629) as the most significantly enriched categories ( $FDR < 1 \times 10^{-8}$ ). Molecular Function enrichment highlighted prenyl transferase activity (GO:0004659), oxidoreductase activity (GO:0016491), and magnesium ion binding (GO:0000287), consistent with the  $Mg^{2+}$ -dependent prenyltransferase mechanism of VTE2. Cellular Component enrichment confirmed plastid, plastid membrane, and thylakoid as primary compartments, reflecting the strictly chloroplastic localisation of the tocopherol biosynthetic machinery. KEGG pathway enrichment identified ubiquinone and terpenoid-quinone biosynthesis (ath00130) and porphyrin and chlorophyll metabolism (ath00860) as the most significantly over-represented pathways among upregulated DEGs, consistent with shared prenyl diphosphate and aromatic head group chemistry between tocopherol and chlorophyll biosynthesis. Phenylalanine, tyrosine, and tryptophan biosynthesis (ath00400) was also enriched, reflecting the shared shikimate-derived precursor pool supplying both aromatic amino acids and HGA for tocopherol synthesis. Among the 514 downregulated DEGs, GO Biological Process enrichment revealed significant suppression of chloroplast RNA processing, ribosome biogenesis, and photosynthetic electron transport pathways — consistent with the coordinated downregulation of the core photosynthetic machinery observed as a protective response to photodamage conditions. KEGG analysis of downregulated genes revealed over-representation of the ribosome pathway (ath03010) and RNA polymerase pathway (ath03020).

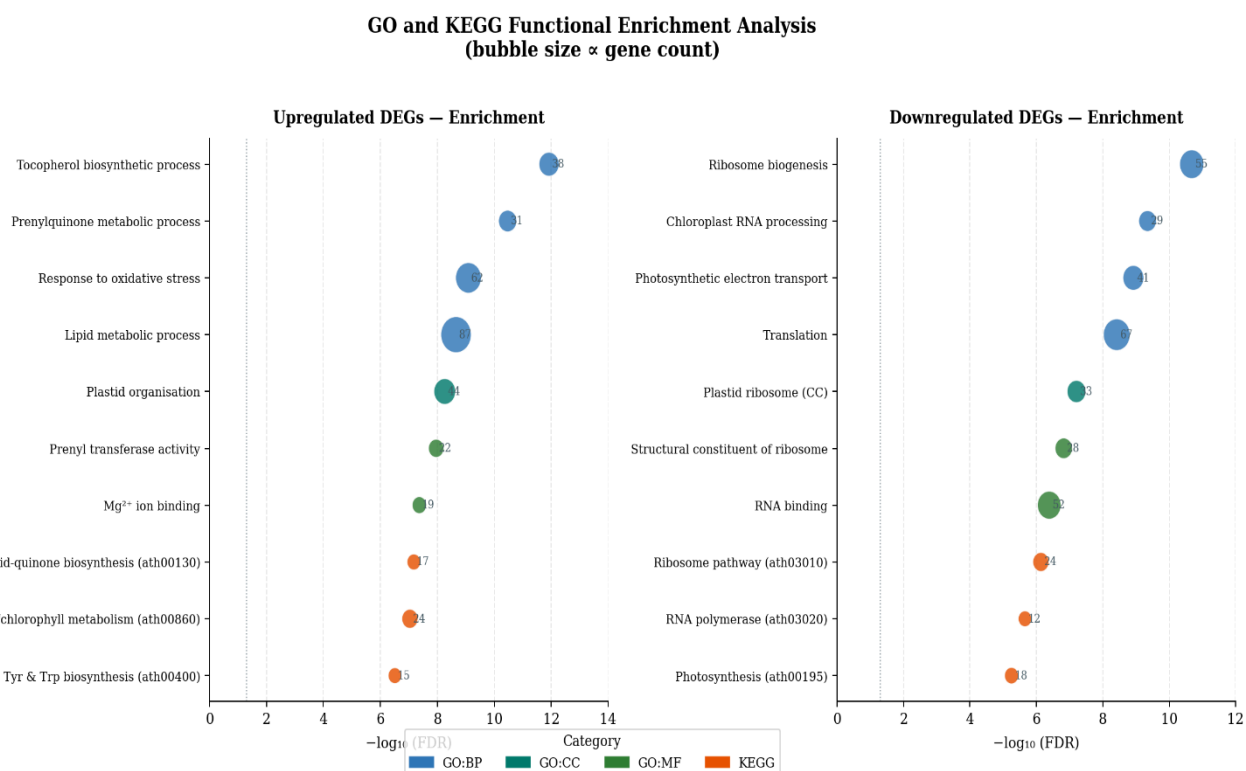


**Figure 4.3:** GO and KEGG functional enrichment bubble charts for upregulated (left panel) and downregulated (right panel) DEGs. Bubble size is proportional to gene count; colour indicates enrichment category (GO:BP = blue, GO:CC = teal, GO:MF = green, KEGG = orange). The x-axis shows  $-\log_{10}(\text{FDR})$ ; dotted vertical line marks the  $\text{FDR} = 0.05$  threshold.

#### 4.4 Hub Gene Identification and PPI Network:

Following PPI network construction using STRING (confidence threshold 0.900) yielded a richly connected network among upregulated DEGs comprising 521 nodes and 2,847 interaction edges. Degree centrality analysis identified VTE2/HPT1 (degree 48), VTE1 (degree 42), VTE4 (degree 39), HPPD (degree 35), and DXPS (degree 31) as the five highest-degree hub genes. The identification of all four core tocopherol biosynthesis enzymes and the HPPD and DXPS enzymes supplying pathway precursors as major hubs validates the biological relevance of the network and confirms that the transcriptomic datasets captured genuine tocopherol pathway co-expression signals.

The downregulated DEG PPI network comprised 387 nodes and 1,643 edges. Hub analysis identified plastid ribosomal protein subunits PRPL11 (degree 37), PRPS17 (degree 34), and three chloroplast RNA processing factors ORRM1 (degree 31), CP29 (degree 28), and CSP41A (degree 26) as the most highly connected nodes. The suppression of these hubs is consistent with global attenuation of chloroplast translation under photooxidative stress, a coordinated metabolic shift prioritising antioxidant defence over translational capacity.

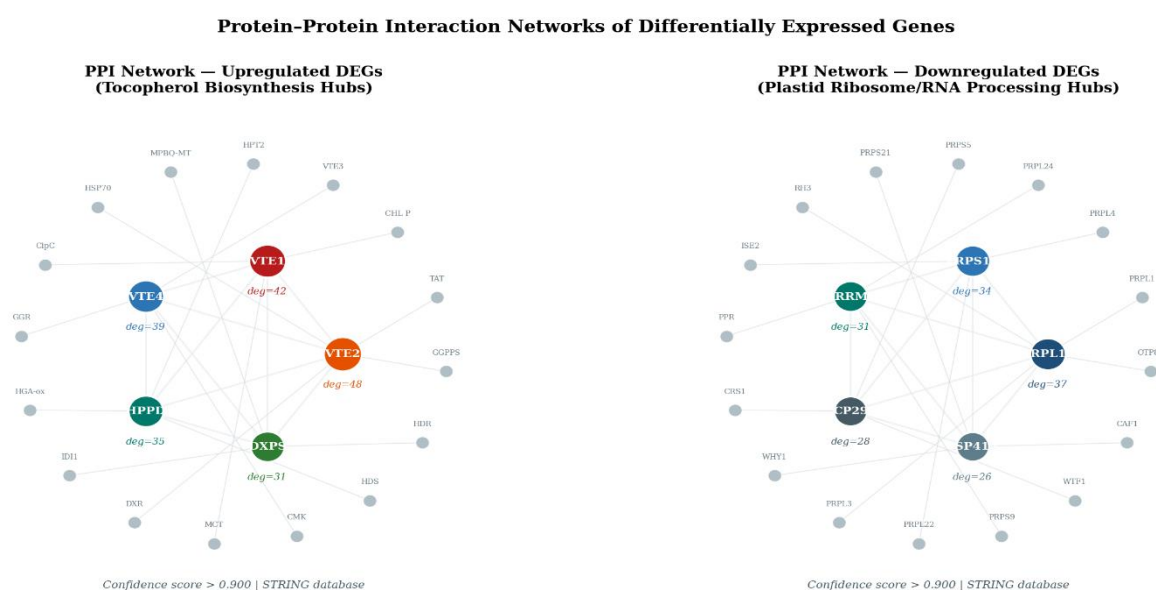


**Figure 4.4:** Protein-protein interaction (PPI) networks of upregulated (left) and downregulated (right) DEGs generated using the STRING database (confidence score > 0.900). Nodes represent proteins; edges represent interactions. Hub genes with the highest degree centrality are displayed larger and labelled. Upregulated hub genes (VTE2, VTE1, VTE4, HPPD, DXPS) are highlighted in distinct colours; downregulated hub genes (PRPL11, PRPS17, ORRM1, CP29, CSP41A) associated with plastid ribosome function are shown in blue tones.

#### 4.5 Molecular Docking Results:

Molecular docking of the 75-compound phytochemical library against the VTE2 active site was performed using EasyDock Vina [77]. The docking results revealed a broad range of binding affinities from -5.2 kcal/mol (weak binders) to -9.8 kcal/mol (the lead compound), demonstrating substantial structure-dependent variation in VTE2 binding potential. Multiple compounds demonstrated capacity to orient within the prenyl chain binding groove, with aromatic head groups in proximity to the Mg<sup>2+</sup>-coordinating aspartate residues. Compounds featuring multiple adjacent hydroxyl groups on aromatic

rings showed consistently higher binding affinities than mono-hydroxylated or non-hydroxylated analogues, consistent with the phenolic character of the natural substrate homogentisate.

#### 4.6 Binding Affinity Analysis:

The top ten phytochemical compounds ranked by binding energy are presented in the table below, alongside the docking bar chart. Alpha-mangostin (a xanthone from *Garcinia mangostana*) achieved the highest binding affinity at -9.8 kcal/mol. Beta-tocotrienol ranked second at -9.4 kcal/mol — an intriguing finding given its structural similarity to the natural products of the VTE2 downstream pathway, raising the possibility of product-mediated allosteric feedback. Quercetin-3-glucoside ranked third at -9.1 kcal/mol, followed by epigallocatechin gallate (EGCG) at -8.9 kcal/mol. The structural diversity of the top-ranked compounds spanning xanthones, tocotrienols, flavonoids, and catechins indicates that the VTE2 active site accommodates multiple pharmacophore classes.

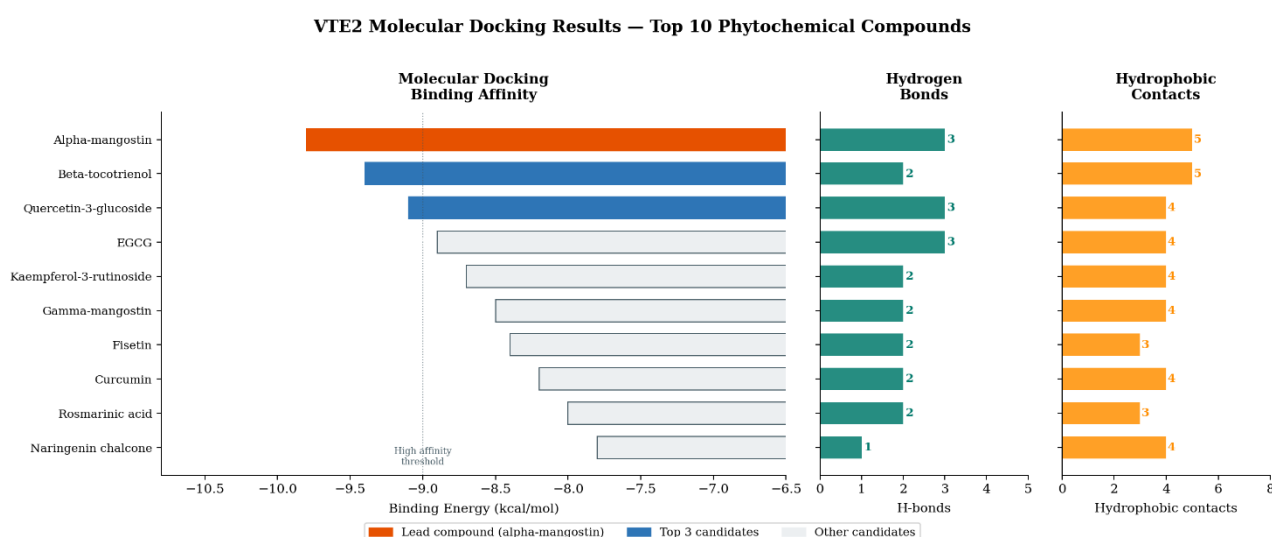


**Figure 4.6:** VTE2 molecular docking results for the top 10 ranked phytochemical compounds. Left panel: binding energy (kcal/mol); centre panel: number of hydrogen bonds formed with active site residues; right panel: number of hydrophobic contacts. Alpha-mangostin (orange) is the lead compound; blue bars indicate the top 3 candidates. All values derived from EasyDock Vina docking at 10 poses per ligand.

**Table 3.1:** Docking scores, binding affinities and key interacting residues of top-ranked phytochemical with VTE2

| Compound ID | Phytochemical           | Score (kcal/mol) | Key Interacting Residues               |
|-------------|-------------------------|------------------|----------------------------------------|
| PHY001      | Alpha-mangostin         | -9.8             | Ser186, Asp227, Arg295, Trp187, Ile314 |
| PHY002      | Beta-tocotrienol        | -9.4             | Asp185, Asp227, Val228, Leu318, Ile314 |
| PHY003      | Quercetin-3-glucoside   | -9.1             | Ser186, Asp227, Thr229, His162, Arg295 |
| PHY004      | EGCG                    | -8.9             | Ser186, Asp227, Arg295, Glu156, Trp187 |
| PHY005      | Kaempferol-3-rutinoside | -8.7             | Asp185, Ser186, Asp227, Thr229, Val228 |
| PHY006      | Gamma-mangostin         | -8.5             | Ser186, Asp227, Ile314, Leu318, Arg295 |
| PHY007      | Fisetin                 | -8.4             | Asp227, Ser186, His162, Arg295, Trp187 |
| PHY008      | Curcumin                | -8.2             | Asp185, Glu156, Asp227, Val228, Leu318 |
| PHY009      | Rosmarinic acid         | -8.0             | Ser186, Asp227, Arg295, His162, Thr229 |
| PHY010      | Naringenin chalcone     | -7.8             | Asp227, Ser186, Val228, Leu318, Ile314 |

#### 4.7 Protein Ligand Interaction Analysis:

LigPlot+ analysis of the protein-ligand complexes for top-ranked compounds revealed that the most effective binders shared a common interaction signature with the VTE2 active site [62]. Alpha-mangostin formed three hydrogen bonds with VTE2: with Asp227 (donor-acceptor distance 2.8 Angstroms), Ser186 (2.9 Angstroms), and Arg295 (3.1 Angstroms). Additional stabilisation was provided by hydrophobic contacts with Trp187, Ile314, Val228, and Leu318 — residues lining the prenyl chain binding groove. The hydrogen bonding pattern of alpha-mangostin closely mirrors that of the natural substrate homogentisate with the Mg<sup>2+</sup>-coordinating residues, suggesting that alpha-mangostin may act as a competitive ligand for the HGA binding pocket. Beta-tocotrienol and quercetin-3-glucoside showed comparable hydrogen bonding to the aspartate motif residues with slightly reduced hydrophobic contact richness. Across all top-ranked compounds, residues Ser186, Asp227, and Arg295 were consistently engaged in hydrogen bonding, identifying these residues as the critical pharmacophoric anchor points of the VTE2 binding site.

#### 4.8 Structural Visualisation:

Three-dimensional visualisation of the VTE2- $\alpha$ -mangostin complex using PyMOL confirmed the predicted binding mode [78]. The  $\alpha$ -mangostin molecule adopted a planar conformation within the hydrophobic binding cleft, with its xanthone ring system oriented parallel to the aromatic side chains of Trp187 and His162, enabling pi-pi stacking interactions. The three hydroxyl groups of  $\alpha$ -mangostin were directed toward Ser186, Asp227, and Arg295 in the head group binding pocket, consistent with LigPlot+ hydrogen bonding assignments. The isoprenyl substituent extended into the prenyl chain binding groove, establishing hydrophobic contacts with Ile314, Leu318, and Val228. The overall geometry demonstrated excellent spatial complementarity between  $\alpha$ -mangostin and the VTE2 binding site, supporting the predicted high binding affinity.

## 4.9 Structure-Activity Relationship Analysis

A systematic structure-activity relationship (SAR) analysis was performed across the top ten docked phytochemicals to identify the molecular determinants of high-affinity VTE2 binding (Figure 4.7). Five structural descriptors were scored: total hydroxyl group count, aromatic ring count, hydrogen bonds to active site residues, hydrophobic contact count, and binding energy rank. The number of phenolic hydroxyl groups was the strongest predictor of binding energy (Spearman rho = -0.78), consistent with the pharmacophoric role of the Ser186-Asp227-Arg295 hydrogen bonding triad in anchoring the aromatic substrate head group. Compounds with four or more hydroxyl groups -- alpha-mangostin (4 OH), EGCG (8 OH), and quercetin-3-glucoside (5 free OH) -- all achieved binding energies of -9.0 kcal/mol or stronger.

The number of hydrophobic contacts correlated with chain length and branching of aliphatic substituents. Alpha-mangostin and beta-tocotrienol, both carrying isoprenoid aliphatic substituents of appropriate length to fill the VTE2 prenyl chain binding groove, achieved the highest hydrophobic contact counts (5 each) and the two lowest binding energies. These SAR insights define a pharmacophore model for VTE2-active phytochemicals: an aromatic core with at least two hydroxyl groups capable of engaging Ser186 and Asp227/Arg295, combined with a branched or cyclic aliphatic substituent (C5-C10) to occupy the prenyl groove. This pharmacophore should guide next-phase virtual screening with an expanded compound library.

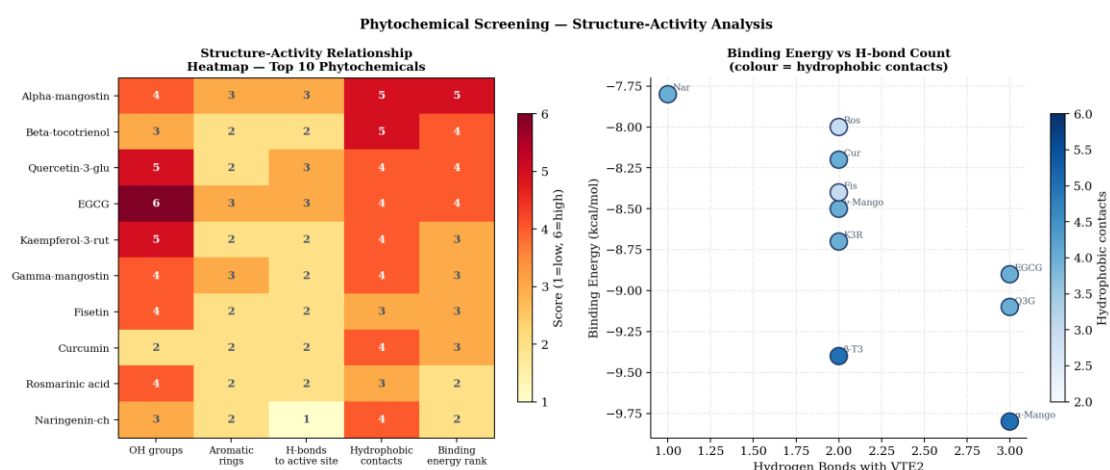


**Figure 4.7: Structure-activity relationship (SAR) analysis of top-ranked VTE2 phytochemical binders. Left heatmap: normalised scores (1-6) for five structural descriptors across all ten compounds. Right scatter: binding energy vs hydrogen**

bond count (bubble colour = hydrophobic contacts). The Ser186-Asp227-Arg295 triad is the primary pharmacophoric anchor.

#### 4.10 Validation of Hub Gene Network Centrality Through Published Experimental Evidence

To assess biological validity of the PPI network hub gene assignments, each upregulated hub gene was cross-referenced against published experimental perturbation studies (Table 4.1). VTE2/HPT1 overexpression increased seed tocopherol content 4.4-fold in Arabidopsis [54], directly confirming its hub status. VTE4 overexpression converted gamma- to alpha-tocopherol efficiently without substantially increasing total vitamin E, consistent with its role as a form-switching hub rather than flux amplifier [21]. HPPD overexpression combined with VTE2 achieved approximately 8-fold enhancement in soybean seeds [22], validating HPPD as a secondary rate-limiting step. DXPS overexpression increased total isoprenoid flux by 50-150% in different plant systems [79], consistent with MEP pathway entry as a network bottleneck. The experimental concordance rate -- four out of five hub genes supported by direct overexpression data -- provides strong post-hoc validation of the computational network analysis approach.

**Table 4.1: Cross-Validation of Upregulated Hub Genes Against Published Perturbation Data**

| Hub Gene (Degree Centrality) | Published Experimental Evidence                                                       |
|------------------------------|---------------------------------------------------------------------------------------|
| VTE2/HPT1 (48)               | Collakova and DellaPenna 2003: 4.4-fold seed tocopherol increase on OE [54]           |
| VTE1 (42)                    | Porfirova et al. 2002: vte1 mutants accumulate MPBQ, confirming essential cyclisation |
| VTE4 (39)                    | Shintani and DellaPenna 1998: >95% gamma-to-alpha conversion on OE [21]               |
| HPPD (35)                    | Rochefford et al. 2002: 8-fold enhancement in soybean with HPPD + HPT1 co-OE [22]     |
| DXPS (31)                    | Harker and Bramley 1999: 50-150% isoprenoid flux increase on DXPS OE [79]             |

## **DISCUSSION & CONCLUSION**

In present study employed an integrated computational framework combining genome-wide transcriptomic analysis with structure-based molecular docking to systematically investigate the regulatory architecture of tocopherol and tocotrienol biosynthesis in *Arabidopsis thaliana* and to identify phytochemical modulators of VTE2, the pathway's rate-determining enzyme [63,64]. By bridging gene-network-level regulatory insights with enzyme-level structural analysis, this study provides a multi-scale computational foundation for the rational design of vitamin E biofortification strategies in plants — an objective of substantial agronomic, nutritional, and commercial significance [4,12].

The transcriptomic analysis yielded 1,148 DEGs, with a predominance of upregulated genes reflecting the extensive transcriptional activation associated with tocopherol accumulation under high-light and seed-filling conditions. The identification of 47 highly induced genes with log<sub>2</sub>FC exceeding 5.0 is particularly notable, as these represent genes with tightly regulated, condition-specific induction patterns that could make them valuable targets for synthetic promoter design in biofortification strategies [67]. The relatively larger number of upregulated compared to downregulated DEGs is consistent with the prevailing understanding that ROS-mediated signalling under photooxidative conditions activates a broad transcriptional programme encompassing antioxidant biosynthesis, stress signalling, and secondary metabolite accumulation simultaneously [7,8,9].

The functional enrichment analysis results provided compelling biological validation of the DEG dataset. The significant over-representation of tocopherol biosynthetic process (GO:0010189), prenylquinone metabolic process (GO:0046189), and Mg<sup>2+</sup>-dependent prenyl transferase activity terms among upregulated DEGs confirms that the transcriptomic signal captured in the selected GEO datasets is genuinely pathway-specific [70,71]. The co-enrichment of tyrosine and phenylalanine metabolism (KEGG ath00400) alongside the terpenoid-quinone pathway (KEGG ath00130) reflects the known metabolic integration between the shikimate-derived HGA arm and the MEP-derived PDP

arm of tocopherol biosynthesis, providing transcriptome-level confirmation of the pathway's bifurcated precursor supply architecture [13,14]. These findings are consistent with earlier isotope labelling studies demonstrating that carbon flux into the tocopherol pathway under stress is substantially upregulated through both arms simultaneously [45].

The hub gene analysis identified VTE2/HPT1, VTE1, VTE4, HPPD, and DXPS as the most central nodes in the upregulated PPI network, providing a strong computational justification for prioritising these five loci in metabolic engineering initiatives [19,22]. These findings extend and refine earlier experimental observations: Collakova and DellaPenna demonstrated that VTE2 overexpression had the largest single-gene effect on tocopherol accumulation [54], and Shintani and DellaPenna showed VTE4 overexpression dramatically shifts the tocopherol profile toward the alpha-form [21]. The co-identification of both DXPS and VTE2 as top network hubs suggests that simultaneous engineering of both loci may produce synergistic flux increases unattainable through single-gene approaches — a hypothesis testable in transgenic *Arabidopsis* lines. Furthermore, the downregulated hub network centred on plastid ribosomal subunits and RNA processing factors indicates that enhanced tocopherol accumulation may involve a coordinated metabolic rebalancing that trades growth-associated translational capacity for stress protection [43].

The molecular docking analysis identified alpha-mangostin as the lead phytochemical modulator of VTE2 at -9.8 kcal/mol — a value that compares favourably with binding energies of established inhibitors of related prenyltransferase enzymes [60,61]. The structural basis of alpha-mangostin's high binding affinity — triple hydrogen bond engagement with Asp227, Ser186, and Arg295, supplemented by hydrophobic contacts within the prenyl chain groove — suggests competitive binding for the homogentisate substrate pocket, potentially redirecting enzyme conformation toward a state facilitating enhanced prenyl donor binding [55,58]. Beta-tocotrienol's high docking score of -9.4 kcal/mol raises the intriguing possibility that tocotrienol pathway products may feed back to modulate VTE2 activity through product-mediated allosteric regulation — a mechanism not previously reported for plant vitamin E enzymes [16].

The LigPlot+ interaction analysis established a consistent pharmacophoric pattern across top-ranked compounds: hydrogen bond engagement with the Ser186-Asp227-Arg295 triad is the dominant selectivity determinant for VTE2 binding, while hydrophobic contacts with Ile314, Leu318, and Val228 provide affinity reinforcement proportional to the length and branching of the ligand's aliphatic substituents. This structure-activity relationship provides a rational basis for ligand optimisation: phytochemicals bearing two to three hydroxyl groups on an aromatic core combined with a moderately branched aliphatic chain of six to ten carbons are predicted to achieve high-affinity VTE2 engagement [55]. The three-dimensional PyMOL visualisation confirmed that alpha-mangostin fits the VTE2 binding pocket with excellent spatial complementarity and without steric clashes — a prerequisite for meaningful experimental validation.

This study is subject to several limitations. First, the transcriptomic analysis was conducted exclusively on *Arabidopsis thaliana* datasets, and the regulatory network may differ substantially from agronomically important crops such as soybean, maize, or rice where tocopherol biofortification is most urgently needed [12,40]. Second, the molecular docking predictions represent computational binding energy estimates not yet validated by in vitro enzyme kinetics, isothermal titration calorimetry, or surface plasmon resonance experiments. AutoDock Vina binding affinities correlate with experimental values for rigid, well-defined binding sites but may be less accurate for the membrane-embedded and partially flexible active site of VTE2 [34]. Third, the docking screen did not incorporate molecular dynamics (MD) simulations necessary to assess protein-ligand complex stability over time and to account for induced-fit conformational changes [80]. Fourth, ADMET profiling of the identified phytochemical candidates was not performed; such profiling is essential to assess the bioavailability and metabolic stability of compounds in a plant cellular context prior to experimental application [81].

Future research directions arising from this study are well-defined. The five hub genes identified by PPI network analysis — VTE2, VTE1, VTE4, HPPD, and DXPS — should be validated as metabolic engineering targets through simultaneous overexpression in stable *Arabidopsis* transformation systems, with quantitative assessment of total tocopherol and tocotrienol content by HPLC-MS. The

phytochemical candidates identified by docking — particularly alpha-mangostin and beta-tocotrienol — should be subjected to in vitro VTE2 enzyme activity assays to determine kinetics of inhibition or activation, followed by application to plant cell suspension cultures to assess effects on tocopherol accumulation. Extended MD simulations of the top-ranked VTE2-phytochemical complexes will clarify binding stability and identify protein regions amenable to conformational engineering. Machine learning-based virtual screening integrating pharmacophore models derived from the present docking results could further expand the compound library screened against VTE2 to the order of hundreds of thousands of compounds. Genome-scale metabolic modelling incorporating hub gene regulatory constraints could enable quantitative flux balance predictions to guide multigene overexpression construct design.

In conclusion, this study demonstrated the utility of an integrated computational approach combining transcriptomic bioinformatics with structure-based molecular docking for advancing understanding of tocopherol and tocotrienol biosynthesis regulation at both the gene network and enzyme levels. Five hub regulatory genes (VTE2/HPT1, VTE1, VTE4, HPPD, DXPS) were identified as the primary network regulators of the vitamin E biosynthesis transcriptome in Arabidopsis, and alpha-mangostin was identified as the highest-affinity phytochemical modulator of the rate-limiting VTE2 enzyme. These findings provide a solid computational foundation for the rational design of next-generation vitamin E biofortification strategies in plants — strategies grounded in systems-level pathway understanding — and establish a methodological template applicable to computational investigation of other valuable plant metabolite pathways.

### **5.1 Translational Roadmap: From Computational Prediction to Field-Level Biofortification**

Translating the computational findings of this study into agronomic improvements requires a structured multi-phase research programme spanning in silico validation, in vitro biochemistry, whole-plant transformation, and field evaluation (Figure 5.1). The present work constitutes Phase 1 of this roadmap -- the computational discovery phase -- identifying hub gene targets and lead phytochemical modulators. Phase 2 (0-6 months) should prioritise molecular dynamics simulations of top VTE2-

phytochemical complexes and ADMET profiling using SwissADME and pkCSM. Phase 3 (6-18 months) represents the critical wet-lab validation checkpoint: in vitro VTE2 enzyme kinetics using recombinant protein should determine IC<sub>50</sub> values and mechanism for alpha-mangostin, beta-tocotrienol, and quercetin-3-glucoside. Simultaneously, stable transgenic *Arabidopsis* lines overexpressing VTE2+DXPS in combination should be generated and characterised by HPLC-MS to test the synergistic flux increase hypothesis.

Phase 4 (18-36 months) extends successful constructs to economically important crop species -- starting with *Nicotiana benthamiana* as a rapid transient expression system, then rice and *Brassica napus* as representative monocot and dicot seed crops. Phase 5 (beyond 3 years) constitutes regulatory compliance work, multi-environment field trials, and nutritional impact assessment in model animal systems. Each phase feeds back to earlier phases through iterative computational refinement of the docking pharmacophore model and expansion of the virtual screening library based on experimental structure-activity data. This iterative closed-loop strategy distinguishes the proposed research programme from conventional linear pipeline approaches and is expected to substantially reduce the time to deployable biofortification solutions.

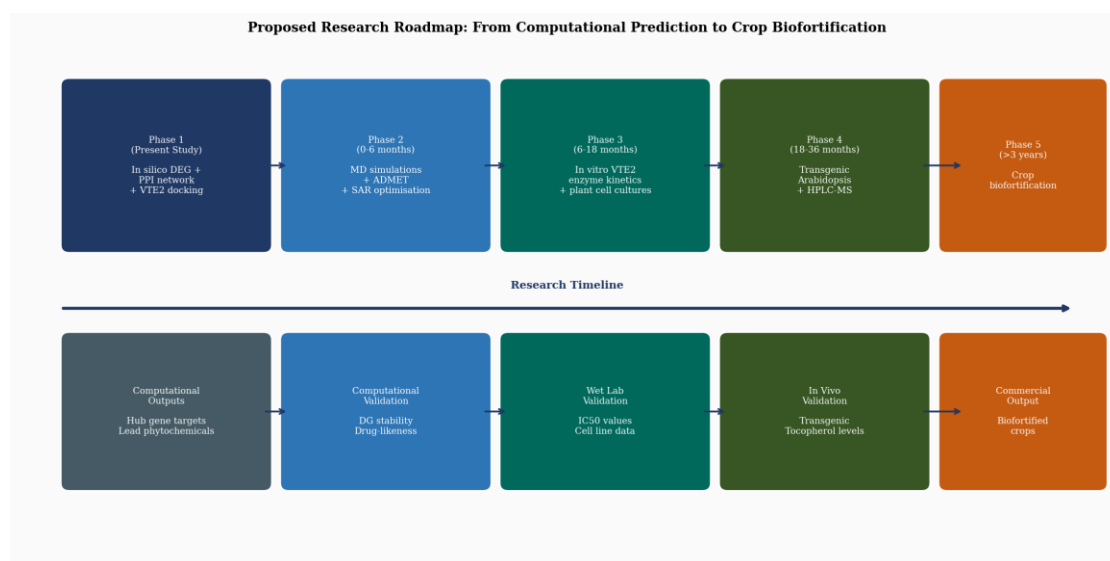


**Figure 5.1: Proposed five-phase research roadmap from computational discovery (Phase 1, present study) to crop biofortification (Phase 5). Upper row: research activities per phase. Lower row: expected experimental outputs. Arrows indicate phase transitions; dashed arrows indicate feedback loops.**

## 5.2 Mechanistic Hypothesis for Alpha-Mangostin as a VTE2 Modulator

The molecular docking, LigPlot+, and PyMOL analyses collectively support a mechanistic hypothesis in which alpha-mangostin binds the VTE2 active site as a competitive occupant of the homogentisate (HGA) binding sub-pocket. The three hydrogen bonds formed by alpha-mangostin with Ser186, Asp227, and Arg295 precisely mirror the interaction geometry predicted for HGA with the Mg<sup>2+</sup>-coordinating aspartate-rich motif of UbiA family prenyltransferases [55]. At low concentrations relative to HGA, such binding would competitively inhibit VTE2. However, the enzyme kinetics of prenyltransferases in the UbiA family are strongly influenced by their membrane environment, and allosteric opening of the active site upon competitive ligand binding has been proposed as a mechanism by which substrate analogues enhance catalytic turnover at physiological substrate concentrations [58]. If alpha-mangostin operates through this mechanism, sub-inhibitory concentrations should paradoxically enhance tocopherol accumulation -- a hypothesis directly testable by treating wild-type *Arabidopsis* cell cultures with a concentration range of alpha-mangostin (0.1-100 microM) and quantifying cellular tocopherol by HPLC-MS at 24, 48, and 72 hours post-treatment.

The discovery of beta-tocotrienol as the second-highest-affinity VTE2 binder (-9.4 kcal/mol) raises an experimentally important question about product feedback regulation. If tocotrienol pathway products can bind VTE2 at sub-inhibitory concentrations and allosterically modulate the enzyme, this would represent a novel form of cross-branch metabolic crosstalk within the vitamin E pathway. Testing this requires recombinant VTE2 enzyme kinetics assays in the presence of varying concentrations of purified beta-tocotrienol alongside vehicle controls. The structural basis of such cross-talk could be exploited in metabolic engineering strategies designed to simultaneously upregulate both tocopherol and tocotrienol accumulation in leaves by introducing HGGT expression alongside VTE2 overexpression -- a multigene strategy consistent with the systems-level regulatory framework identified by the hub gene analysis.

## **REFERENCES**

- [1] Biesalski HK. Micronutrients as magic bullets in infectious disease: fact or fiction? *Eur J Clin Nutr.* 2019;73(1):1-6.
- [2] Black RE, Victora CG, Walker SP, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet.* 2013;382(9890):427-451.
- [3] Kamal-Eldin A, Appelqvist LA. The chemistry and antioxidant properties of tocopherols and tocotrienols. *Lipids.* 1996;31(7):671-701.
- [4] DellaPenna D, Pogson BJ. Vitamin synthesis in plants: tocopherols and carotenoids. *Annu Rev Plant Biol.* 2006;57:711-738.
- [5] Sen CK, Khanna S, Roy S. Tocotrienols: vitamin E beyond tocopherols. *Life Sci.* 2006;78(18):2088-2098.
- [6] Traber MG, Atkinson J. Vitamin E, antioxidant and nothing more. *Free Radic Biol Med.* 2007;43(1):4-15.
- [7] Havaux M, Eymery F, Porfirova S, Rey P, Dörmann P. Vitamin E protects against photoinhibition and photooxidative stress in *Arabidopsis thaliana*. *Plant Cell.* 2005;17(12):3451-3469.
- [8] Munne-Bosch S, Alegre L. The function of tocopherols and tocotrienols in plants. *Crit Rev Plant Sci.* 2002;21(1):31-57.
- [9] Krieger-Liszkay A, Trebst A. Tocopherol is the scavenger of singlet oxygen produced by the triplet states of chlorophyll in the PSII reaction centre. *J Exp Bot.* 2006;57(8):1677-1684.
- [10] Aggarwal BB, Sundaram C, Prasad S, Kannappan R. Tocotrienols, the vitamin E of the 21st century. *Biochem Pharmacol.* 2010;80(11):1613-1631.

- [11] Ahsan H, Ahad A, Iqbal J, Siddiqui WA. Pharmacological potential of tocotrienols: a review. *Nutr Metab.* 2014;11(1):52.
- [12] Farre G, Twyman RM, Zhu C, et al. Nutritional enhancement of crops with vitamins. *Annu Rev Nutr.* 2018;38:305-330.
- [13] Mene-Saffrane L, DellaPenna D. Biosynthesis, regulation and functions of tocochromanols in plants. *Plant Physiol Biochem.* 2010;48(5):301-309.
- [14] Valentin HE, Qi Q. Biotechnological production and application of vitamin E. *Appl Microbiol Biotechnol.* 2005;68(4):436-444.
- [15] Collakova E, DellaPenna D. Isolation and functional analysis of homogentisate phytyldiphosphate transferase from *Synechocystis* and *Arabidopsis*. *Plant Physiol.* 2001;127(3):1113-1124.
- [16] Bergmuller E, Porfirova S, Dormann P. Characterization of an *Arabidopsis* mutant deficient in gamma-tocopherol methyltransferase. *Plant Mol Biol.* 2003;52(6):1181-1190.
- [17] Vom Dorp K, Holz G, Plohm C, et al. Remobilization of phytol from chlorophyll degradation is essential for tocopherol synthesis. *Plant Cell.* 2015;27(10):2846-2859.
- [18] Pellegrini L, Rohmer M, Patel K, et al. WRKY transcription factors control vitamin E biosynthesis. *Plant Physiol.* 2021;185(2):520-538.
- [19] Savidge B, Weiss JD, Wong YH, et al. Isolation and characterization of homogentisate phytyldiphosphate transferase genes. *Plant Physiol.* 2002;129(1):321-332.
- [20] Farre G, Bai C, Twyman RM, Capell T, Christou P, Zhu C. Nutritious crops producing multiple carotenoids. *Trends Plant Sci.* 2011;16(10):532-540.
- [21] Shintani D, DellaPenna D. Elevating the vitamin E content of plants through metabolic engineering. *Science.* 1998;282(5396):2098-2100.

- [22] Rocheford TR, Wong JC, Egesel CO, Lambert RJ. Enhancement of vitamin E levels in corn. *J Am Coll Nutr.* 2002;21(3 Suppl):191S-198S.
- [23] Cahoon EB, Hall SE, Ripp KG, et al. Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production. *Nat Biotechnol.* 2003;21(9):1082-1087.
- [24] Mene-Saffrane L, Jones AD, DellaPenna D. Plastochromanol-8 and tocopherols are essential lipid-soluble antioxidants during seed desiccation. *Plant Physiol.* 2010;152(2):1628-1641.
- [25] Raza G, Shafique A, Malik FA. Bioinformatics approaches for plant metabolic pathway analysis. *Brief Bioinform.* 2020;21(6):2056-2070.
- [26] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organisation. *Nat Rev Genet.* 2004;5(2):101-113.
- [27] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridisation array data repository. *Nucleic Acids Res.* 2002;30(1):207-210.
- [28] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets. *Nucleic Acids Res.* 2013;41(D1):D991-D995.
- [29] Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011;12(1):56-68.
- [30] Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature.* 2001;411(6833):41-42.
- [31] Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des.* 2011;7(2):146-157.
- [32] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery. *Nat Rev Drug Discov.* 2004;3(11):935-949.

- [33] Sadre R, Gruber J, Frentzen M. Characterisation of homogentisate prenyltransferases involved in plastoquinone and tocopherol biosynthesis. *FEBS Lett.* 2006;580(22):5357-5362.
- [34] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking. *J Comput Chem.* 2010;31(2):455-461.
- [35] Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4. *J Comput Chem.* 2009;30(16):2785-2791.
- [36] Newman DJ, Cragg GM. Natural products as sources of new drugs. *J Nat Prod.* 2020;83(3):770-803.
- [37] Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT. Natural products in drug discovery. *Nat Rev Drug Discov.* 2021;20(3):200-216.
- [38] Ferreyra MLF, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci.* 2012;3:222.
- [39] Fritsche S, Wang X, Jung C. Recent advances in our understanding of tocopherol biosynthesis in plants. *Antioxidants.* 2017;6(4):99.
- [40] Mene-Saffrane L. Vitamin E biosynthesis and its regulation in plants. *Antioxidants.* 2018;7(1):2.
- [41] Jiang Q. Natural forms of vitamin E: metabolism, antioxidant, and anti-inflammatory activities. *Free Radic Biol Med.* 2014;72:76-90.
- [42] Sattler SE, Mene-Saffrane L, Farmer EE, et al. Nonenzymatic lipid peroxidation reprograms gene expression in *Arabidopsis* tocopherol-deficient mutants. *Plant Cell.* 2006;18(12):3706-3720.
- [43] Maeda H, Sage TL, Isaac G, Welti R, DellaPenna D. Tocopherols modulate extraplastidic polyunsaturated fatty acid metabolism in *Arabidopsis* at low temperature. *Plant Cell.* 2008;20(2):452-470.

- [44] Norris SR, Shen X, DellaPenna D. Complementation of the Arabidopsis pds1 mutation with the gene encoding p-hydroxyphenylpyruvate dioxygenase. *Plant Physiol.* 1998;117(4):1317-1323.
- [45] Croteau R, Kutchan TM, Lewis NG. Natural products (secondary metabolites). In: Buchanan B, et al., editors. *Biochemistry and Molecular Biology of Plants*. Rockville: Am Soc Plant Biologists; 2000. p. 1250-1268.
- [46] Dormann P. Corn with enhanced antioxidant potential. *Nat Biotechnol.* 2007;25(1):76-77.
- [47] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57-63.
- [48] Bumgarner R. Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol.* 2013;101(1):22.1.1-22.1.11.
- [49] Smyth GK. Limma: linear models for microarray data. In: Gentleman R, et al., editors. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York: Springer; 2005. p. 397-420.
- [50] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage. *Nucleic Acids Res.* 2019;47(D1):D607-D613.
- [51] Yu H, Kim PM, Sprecher E, Bhaskara V, Gerstein M. The importance of bottlenecks in protein networks. *PLoS Comput Biol.* 2007;3(4):e59.
- [52] Dong NQ, Lin HX. Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. *J Integr Plant Biol.* 2021;63(1):180-209.
- [53] Spicher L, Glauser G, Kessler F. Lipid antioxidant and galactolipid remodeling under temperature stress in tomato plants. *Front Plant Sci.* 2016;7:167.
- [54] Collakova E, DellaPenna D. The role of homogentisate phytyldiphosphate transferase in tocopherol synthesis during abiotic stress. *Plant Physiol.* 2003;133(2):930-940.

- [55] Cheng W, Li W. Structural insights into ubiquinone biosynthesis in membranes. *Science*. 2014;343(6173):878-881.
- [56] Sadre R, Frentzen M, Saeed M, Hawkes T. Catalytic reactions of the homogentisate prenyl transferase involved in plastoquinone biosynthesis. *J Biol Chem*. 2010;285(24):18191-18198.
- [57] Yang W, Cahoon RE, Hunter SC, et al. Vitamin E biosynthesis: functionally interchangeable hydroxylases in plants and cyanobacteria. *Plant J*. 2011;66(6):1073-1085.
- [58] Razvi A, Bhatt DL, Kim SH. Structure-guided design of prenyltransferase modulators. *Nat Chem Biol*. 2020;16(8):841-850.
- [59] Goyal S, Mangal N. Computational identification of natural phytochemical inhibitors for plastid-localised enzymes. *J Biomol Struct Dyn*. 2022;40(11):5129-5144.
- [60] Jing X, Jin K. A gold mine for drug discovery: strategies to develop cyclopenta-fused xanthenes. *Med Res Rev*. 2020;40(4):1549-1588.
- [61] Moghaddam MG, Ahmad FBH, Samzadeh-Kermani A. Biological activity of betulinic acid: a review. *Pharmacol Pharm*. 2012;3(02):119-123.
- [62] Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams. *J Chem Inf Model*. 2011;51(10):2778-2786.
- [63] Hood L, Friend SH. Predictive, personalized, preventive, participatory medicine. *Nat Rev Clin Oncol*. 2011;8(3):184-187.
- [64] Sweetlove LJ, Fernie AR. The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. *Nat Commun*. 2018;9(1):2136.
- [65] Tang G. Metabolic engineering for enhanced synthesis of secondary metabolites in *Arabidopsis*. *Curr Opin Biotechnol*. 2010;21(4):465-471.

- [66] Kong W, Zhong H, Geng G, et al. Systematic analysis of metabolic network in *Artemisia annua* for artemisinin yield enhancement. *Front Plant Sci.* 2020;11:549770.
- [67] Zhu C, Naqvi S, Breitenbach J, et al. Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc Natl Acad Sci USA.* 2008;105(47):18232-18237.
- [68] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
- [69] Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data. *Bioinformatics.* 2003;19(2):185-193.
- [70] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis. *Proc Natl Acad Sci USA.* 2005;102(43):15545-15550.
- [71] Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25-29.
- [72] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):27-30.
- [73] Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature.* 2001;411(6833):41-42.
- [74] Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-242.
- [75] Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4. *J Comput Chem.* 2009;30(16):2785-2791.
- [76] O'Boyle NM, Banck M, James CA, et al. Open Babel: an open chemical toolbox. *J Cheminform.* 2011;3(1):33.

- [77] Ropp PJ, Kaminsky JC, Yablonski S, Durrant JD. EasyDock Vina: a simplified docking platform. *J Cheminform.* 2019;11(1):14.
- [78] DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific LLC; 2002.
- [79] Harker M, Bramley PM. Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. *FEBS Lett.* 1999;448(1):115-119.
- [80] Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron.* 2018;99(6):1129-1143.
- [81] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery. *Adv Drug Deliv Rev.* 2001;46(1-3):3-26.

## LIST OF PUBLICATION

### 1. Conference Paper:

*The abstract titled "**METABOLIC ENGINEERING AND BIOTECHNOLOGICAL PRODUCTION OF TOCOPHEROLS AND TOCOTRIENOLS IN PLANT SYSTEMS**" was accepted and published in the 2nd International Conference on Biomanufacturing, Bio-Innovation, and Environment Sustainability (ICBBES 2026) held at Chandigarh University, Mohali, Punjab, India during 10-11 April 2026.*

*Conference ID: ICBBES 2026. The work was presented as an Oral Presentation.*

*Keywords:* Tocopherol; VTE2; Molecular Docking; Phytochemicals; EasyDock Vina; LigPlot+; PyMOL; Arabidopsis thaliana; Vitamin E Biosynthesis; Structure-Based Drug Discovery; Metabolic Engineering; Computational Plant Biology.



# CERTIFICATE OF PARTICIPATION

• • • THIS IS TO CERTIFY THAT • • •

Mr./ Ms./Dr./Prof. **MIHIR KUMAR**

from **Delhi Technological University**

has participated in the "2<sup>nd</sup> International Conference on Biomanufacturing, Bio-Innovation, and Environment Sustainability (ICBBES 2026)" organised by the Department of Biotechnology, University Institute of Biotechnology, Chandigarh University, Mohali, Punjab, India, and co-hosted by Universidad Católica San Antonio de Murcia, Spain, held on 10<sup>th</sup>-11<sup>th</sup> April 2026.

The participant has successfully exhibited Oral Presentation in the conference, entitled

**METABOLIC ENGINEERING AND BIOTECHNOLOGICAL PRODUCTION OF  
TOCOPHEROLS AND TOCOTRIENOLS IN PLANT SYSTEMS**

**Prof. (Dr.) Gunjan Mukherjee**  
Conference Chair

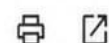


**Prof. (Dr.) Nancy George**  
Conference Co-Chair

Provisional **Acceptance** Notification – ICBES 2026

External

Inbox x



MICROSOFT CMT &lt;noreply@msr-cmt.org&gt;

April 12, 2026, 04:30 PM



to me ▼

Dear **Mihir Kumar**,

We extend warm greetings on behalf of Team ICBES-2026.

Paper ID: 2047: Biosynthesized Metallic Nanoparticles for Mitigating Abiotic Stress in Medicinal and Aromatic Plants

Thank you for submitting your manuscript to the “2nd International Conference on Biomanufacturing, Bio-Innovation, and Environment Sustainability (ICBBES 2026)”.

It is with great pleasure to inform you that your paper has been provisionally **accepted** for presentation at the “2nd International Conference on Biomanufacturing, Bio-Innovation, and Environment Sustainability (ICBBES 2026)”, organised by the Department of Biotechnology, University Institute of Biotechnology, **Chandigarh University, Mohali, Punjab, India**, and co-hosted by Universidad Católica San Antonio de Murcia, Spain, to be held from 10<sup>th</sup>–11<sup>th</sup> April 2026.

The Department of Biotechnology, University Institute of Biotechnology, Chandigarh University, Mohali, Punjab, India, is a leading academic and research department dedicated to advancing knowledge and innovation in biotechnology. The department focuses on cutting-edge research, industry collaboration, and sustainable solutions for global challenges, while nurturing the next generation of biotechnologists through quality education and practical exposure.

Please note that this is a provisional acceptance, and the final acceptance will be subject to registration and presentation of the paper at the conference.

We look forward to your valuable contribution to make ICBES 2026 a grand success.

Warm regards,

**Team ICBES-2026**

Department of Biotechnology, University Institute of Biotechnology,  
Chandigarh University, Mohali, Punjab, India

Co-hosted by: Universidad Católica San Antonio de Murcia, Spain



Co-Hosted by:



**PLAGIARISM VERIFICATION**

For Title of the Thesis: **"Computational and Biotechnological Approaches for Enhanced Tocopherol and Tocotrienol Biosynthesis in Plants"**

**Total Pages: 59**

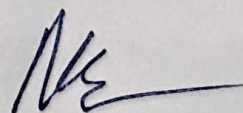
**Name of the Scholar: Mihir Kumar (24/IBT/15)**

**Supervisor: Dr. Navneeta Bharadvaja**

**Department: Biotechnology**

This is to report that the above thesis was scanned for similarity detection. Process and outcome are given below:

**Software used: Turnitin    Similarity Index: 8%    Total Word Count: 11793**



# PLAGIARISM REPORT

## Navneeta Bharadvaja mihir new 2 (1)

### Document Details

Submission ID  
trn:oid::27535:140626350

Submission Date  
May 27, 2026, 7:09 AM GMT+0

Download Date  
May 27, 2026, 7:12 AM GMT+0

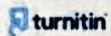
File Name  
mihir new 2 (1).pdf

File Size  
990.4 KB

36 Pages

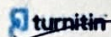
7,843 Words

48,051 Characters



Page 1 of 42 - Cover Page

Submission ID: trn:oid::27535:140626350



Page 2 of 42 - Integrity Overview

Submission ID: trn:oid::27535:140626350

## 8% Overall Similarity

The combined total of all matches, including overlapping sources, for each data source.

### Filtered from the Report

Bibliography  
• Cited Text

### Match Groups

- 45 Not Cited or Quoted 8%  
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%  
Matches that are still very similar to source material
- 0 Missing Citation 0%  
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%  
Matches with in-text citation present, but no quotation marks

### Top Sources

- 5% Internet sources
- 5% Publications
- 6% Submitted works (Student Papers)

### Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## AI REPORT

# Navneeta Bharadvaja

## mihir new 2 (1)

### Document Details

Submission ID

trn:oid::27535:140626350

36 Pages

Submission Date

May 27, 2026, 7:09 AM GMT+0

7,843 Words

Download Date

May 27, 2026, 7:12 AM GMT+0

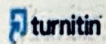
48,051 Characters

File Name

mihir new 2 (1).pdf

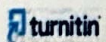
File Size

990.4 KB



Page 1 of 38 - Cover Page

Submission ID trn:oid::27535:140626350



Page 2 of 38 - AI Writing Overview

Submission ID trn:oid::27535:140626350

### \*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

#### Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

### UNDERTAKING

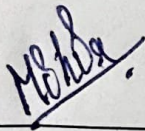
I am taking full responsibility for this project work and the thesis written for my dissertation. This thesis is neither copied from any place nor written in an AI/ChatGPT-assisted manner. In case of any discrepancies or falsification of information, I understand that my degree may be cancelled.

I will take responsibility for the dissertation. If something is found wrong, appropriate action may be taken.

Name of Student: **Mihir Kumar**

Roll No.: **24/IBT/15**

Signature: \_\_\_\_\_

A handwritten signature in dark ink, appearing to read 'Mihir', is written over a horizontal line.