ABSTRACT

YUE, ZHU. Biosynthesis of Flavanol and Oligomeric Flavonoids and Their Medicinal Values (Under the direction of Dr. De-Yu Xie).

Proanthocyanidin (PA) and flavan-3-ol (flavanol) are two groups of main plant flavonoids. PAs are oligomeric or polymeric flavan-3-ols. These two groups play multiple roles in the defense of plants against biotic and abiotic stresses. They also have multiple health benefits for humans. To date, although flavan-3-ols have gained appropriate studies in their pathway elucidation and genetic regulation, the biogenesis of their stereo configurations remains unclear. This stereo configuration is important, given that this knowledge is essential to understand the biosynthesis of PAs. Furthermore, how plants polymerize flavan-3-ols to PAs remains unknown. To enhance addressing these questions, my dissertation research focused on cloning and functional characterization of genes associated with PA and flavan-3-ol biosynthesis.

My dissertation includes one chapter reporting the current research progresses of PAs and flavan-3-ols and four chapters reporting research data to provide evidence showing roles of flavan-3-ols involved in the biosynthesis of PAs and a novel group of plant flavonoids as well as medicinal values. Chapter 1 reviewed the current understanding of flavan-3-ols and PAs biosynthesis as well as their medicinal values. Chapter II reported cloning and functional analysis of a 4-dihydroflavonol reductase gene from *Vitis bellula* (*VbDFR*), a key gene involved in the formation of stereo configurations of flavan-3-ols and the biosynthesis of PA. VbDFR catalyzed three dihydroflavonols to their corresponding leucoanthocyanidins *in vitro*. The overexpression o of *VbDFR* increased anthocyanin production in flowers and produced new tobacco anthocyanidins, precursors of flavan-3-ols. Chapter III reported a gene encoding a novel

flavanol polymerase (FP) in red 6R cells engineered from red PAP1 tobacco plants. FP catalyzed flavanols, such as epicatechin and catechin as substrates, to produce dimeric, trimeric, and higher degree oligomeric compounds. Chemical, physic, and MS/MS analyses characterized that these compounds were novel flavonoids. NMR experiments elucidated that these compounds consisted of flavanols such as epicatechin as a starter unit and flav-2-en-3-ol as extension units. Herein, we name these compounds papanridins, a novel group of plant flavonoids. Further metabolic profiling of plants and phylogenetic analysis revealed that papanridin is prevalent in the plant kingdom. Chapter IV reported medicinal activities of dimeric PAs and flavanols in inhibiting the main protease (M^{pro}) of SARS-Cov-2. Docking simulation was performed to screen flavan-3-ols and dimeric PAs. The resulting simulation showed positive results that all these simulated compounds could bind to the catalytic dyad of M^{pro}. *In vitro* assays demonstrated that five compounds inhibited the activity of M^{pro}. Furthermore, a screening of flavan-3-ols and PA-rich foods showed that extracts of green tea, muscadine grape, cacao, and dark chocolate products promisingly inhibited the activity of M^{pro}, suggesting potential pharmaceutical values in interfering COVID-19. Chapter V reported medicinal values of flavonols and dihydroflavonols in interfering COVID-19 caused by SARS-COV-2 and flu caused by human coronavirus 229E (hCov 229E). As completed in Chapter IV, docking simulation showed that flavonols and dihydroflavonols could bind to the catalytic dyad of M^{pro} of SARS-COV-2. *In vitro* assays showed that flavonols and dihydroflavonols promisingly inhibited the activity of M^{pro}. Infection assays of Huh-7 cells with hCov 229E and the resulting TCID50 data showed that flavanol gallates, B-type dimeric PA, flavonols, and dihydroflavonols inhibited the replication of hCov 229E in host cells. These findings indicate that these compounds are candidates with promising values in the therapy of COVID-19 and hCov 229E-infected flu.

Biosynthesis of Flavanol and Oligomeric Flavonoids and Their Medicinal Values

by
Yue Zhu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Plant Science

Raleigh, North Carolina
2021

APPROVED BY:

_____	_____
De-Yu Xie (Committee Chair)	Robert Franks
_____	_____
Jose Alonso	Colleen Doherty

BIOGRAPHY

Yue Zhu, Ph.D.

a) Professional Preparation

Jishou University, Hunan, China, Aug. 2006- May. 2010

Jishou University, Hunan, China, Aug. 2011- May. 2014

North Carolina State University, North Carolina, May. 2016-Present

b) Profession appointment

May. 2020- May. 2021 Ph.D. teaching assistant, Department of Plant and Microbial Biology, North Carolina State University.

May. 2016-present, Ph.D. graduate research assistant, Department of Plant and Microbial Biology, North Carolina State University.

May. 2014- Dec. 2015, research fellow member, State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China.

c) Products (publication)

1. Yue Zhu, De-Yu Xie. 2020, Docking Characterization and In vitro Inhibitory Activity of Flavan-3-ols and Dimeric Proanthocyanidins against the Main Protease Activity of SARS-Cov-2. Frontiers in Plant Science. 11:601316.

2. Yue Zhu, Qingzhong Peng, Kegang Li, De-Yu Xie. 2018, Molecular Cloning and Functional Characterization of a Dihydroflavonol 4-Reductase from Vitis bellula. Molecules. 23(4):861

3. Dongming Ma, Gui Li, Fatima Alejos-Gonzalez, Yue Zhu, Zhen Xue, Aimin Wang, Hui Zhang, Xing Li, Hechun Ye, Hong Wang, Benye Liu, and De-Yu Xie. 2017, Overexpression of a type I isopentenyl pyrophosphate isomerase of Artemisia annua in the cytosol leads to high arteannuin B production and artemisinin increase. The Plant Journal. (2017) 91,466–479.

4. Dongming Ma, Gui Li, Yue Zhu, De-Yu Xie. 2017. Overexpression and Suppression of Artemisia annua 4-Hydroxy-3-Methylbut-2-enyl Diphosphate Reductase 1 Gene (AaHDR1) Differentially Regulate Artemisinin and Terpenoid Biosynthesis. Frontiers in Plant Science. 8(e37149).

5. Le Wang, Shu-Ming Wu, Yue Zhu, Qiang Fan, Zhen-Nan Zhang, Guang Hu, Qing-Zhong, Peng, Jia-He Wu. 2017. Functional characterization of a novel jasmonate ZIM-domain interactor (NINJA) from upland cotton (Gossypium hirsutum). Plant Physiology and Biochemistry. 112 (2017) 152-160.

6. Guo-Liang Zhang, Yue Zhu, Wei-Dong Fu, Peng Wang, Rui-Hai Zhang, Yan-Lei Zhang, Zheng Song, Gui-Xian Xia, Jia-He Wu. 2016. ITRAQ Protein Profile Differential Analysis of Dormant and Germinated Grassbur Twin Seeds Reveals that Ribosomal Synthesis and Carbohydrate Metabolism Promote Germination Possibly Through the PI3K Pathway. Plant and Cell Physiology. 57(6):pcw074.

7. Le Wang, Yue Zhu, Peng Wang, Qiang Fan, Yao Wu, Qing-Zhong Peng, Gui-Xian Xia, Jia-He Wu. 2016, Functional Characterization of a Dihydroflavanol 4-Reductase from the Fiber of Upland Cotton (*Gossypium hirsutum*). Molecules. 21 (2):32.

8. Yue Zhu, Haiyun Wang, QingZhong Peng, Yuntao Tang, Guixian Xia, Jiahe Wu, De-Yu Xie. 2015, Functional Characterization of an Anthocyanidin Reductase Gene from the Fibers of Upland Cotton (*Gossypium hirsutum*). Planta, 241 (5).

9. Yue Zhu, Qing-Zhong Peng, Ke-Gang Li, De-Yu Xie. 2014, Molecular cloning and functional Characterization of the Anthocyanidin Reductase Gene from *Vitis Bellula*. Planta. 240 (2).

10. Jinhuan Pang, Yue Zhu, Qing Li, Jinzhi Liu, Yingchuan Tian, Yule Liu, Jiahe Wu. 2013, Development of Agrobacterium-Mediated Virus-Induced Gene Silencing and

Performance Evaluation of Four Marker Genes in Gossypium barbadense. PLoS ONE, 8(9):e73211

11. Yue Zhu, Qing-Zhong Peng, Ci Du, Ke-Gang Li, and De-Yu Xie. 2013, Characterization of Flavan-3-ols and Expression of MYB and Late Pathway Genes Involved in Proanthocyanidin Biosynthesis in Foliage of Vitis bellula. Metabolites 2013, 185-233.

12. Qing-Zhong Peng, Yue Zhu, Zhong Liu, Ci Du, Ke-Gang Li, and De-Yu Xie. 2012, An integrated approach to demonstrating the ANR pathway of proanthocyanidin biosynthesis in plants. Planta, 236(3):901-18.

d) Synergistic activities

Chinese Governmental Scholarship, 2016- 2020.

ACKNOWLEDGMENTS

Yue Zhu, PhD

08/09-2021

I joined the Department of Plant and Microbial Biology (PMB), North Carolina State University to pursue my PhD study with my great excitement and happiness in May 2016. When I received my offer letter from PMB on December, 2015, we started to image my wonderful and challenging study life. Today, when I recalled that day, that happy moment occurred just like yesterday. When I sit in front my desk thinking of my 5-year's study, I have a list of numerous deep thanks to all people who have given me tremendous helps in my thesis research, living, teaching, learning, improvement, and many other things. When I wrote my thesis, I was full of appreciations to every single help and encouragement that people have kindly offered me. All those kind supports not only have greatly helped my PhD study at NCSU, but also taught me how to develop my future career to support other people.

I am grateful to my PhD advisor, Dr. De-Yu Xie, for his tireless supervision. When Dr. Xie was an adjunct professor at Jishou University, China, he co-advised my master degree study majoring in Plant Ecology. He inspired me to pursue my PhD study at NCSU. During the past 5 years, his encouragement further motivated me to work on this challenging project to address important questions in the plant flavonoid biosynthesis. Under his mentoring, I have learned how to independently raise scientific questions, form hypothesis, design experiments to address questions, analyze and present my data, design future projects, and overcome problems. Whenever I emailed him questions during middle nights, he always replied me in 1-5 min. During the pandemic of COVID-19, like everybody, I experienced extreme difficulties and

stresses. Dr. Xie gave me considerate supports for me to perform experiments, analyze data, and draft manuscripts and my dissertation. When our campus was locked down, he requested time for him and me to access lab so that I could obtain two hours per day to finish experiments. Furthermore, he encouraged me to present my research data in local and international conferences, from which I gained opportunity to exchange my research with peers. All these productive experiences will greatly benefit my future career development.

I am deeply thankful to my committees, Drs. Robert Franks, Colleen Doherty, and Jose Alonso, for their kind mentoring. I always anticipated and loved my committee meetings, given that I could obtain numerous great suggestions about learning and research. When I had questions in development, cell biology, transcriptomes, and genetics, I could obtain excellent advice from them. I am grateful to my committees for their patience with my English challenges. Their advice has helped a lot in my English speaking, vocabulary, and writing.

I thank Dr. Seyit Yuzuak, our lab's former PhD student, for part of his dissertation research in cell culture and purification of Flavanol Polymerase from suspension cells. His solid protein experiments formed a platform for me to clone genes, characterize enzymes, and identify compounds. I thank Dr. Mingzhuo Li for helping confocal microscopy analysis. I thank lab's members, Ms. Yilun Dong, Mr. Salvador Matus, Dr. Rika Judd (a former PhD student), Ms. Emily Ye, and Ms. Bryanna Cruz, for their kind helps and suggestions in my dissertation research and presentation preparation.

I deeply express my gratitude to Drs. Jennifer Sun and Dr. Peter Thompson from The Department of Chemistry for helping Nuclear Magnet Resonance (NMR) experiments to elucidate new compound structures and for training me how to use NMR data to assign H, C and O. I deeply thank Dr. Hanna Gracz (an emeritus professor) for helping train me structural assignment and analyze NMR data. All these activities were completed during the pandemic. Their help greatly inspired me to work on my dissertation research during my difficult time.

I am greatly grateful to Professors Qingzhong Peng and Dr. Kegang Li at Jishou University, China, for their collaboration in gene cloning and functional analysis of DFR from *Vitis ballula*.

I am grateful to Professor Larry Blanton, our Graduate Program Director, for his kind instructions in every single step to follow all graduate school guides in time during my PhD study. I am thankful to Professor Robert Frank, our departmental head, for his support after my scholarship ended. I thank Mr. Dwayne Barnes for his kind instructions that help me follow all guides in time. I thank Mrs. Sue Vitello, our former graduate student secretary, for her helps when I joined PMB in beginning. I thank Mrs. Catherine Freeman, our department executive secretary, for her kind help in every single International Office service. I thank Mrs. Laurie Edwards for her help in any financial needs. I thank Dr. Jeff Gillikin for his advice in all supply purchases. I thank all other graduate students for their helps. I thank Phytotron and Greenhouse staff for offering me space to grow plants during my research.

I thank Professors Qingzhong Peng (my supervisor of master degree study) and Kegang Li (a vice-president) at Jishou University and Dr. Xie at NCSU, for helping my successful application

for my scholarship from the Chinese Scholarship Council (CSC). I am grateful to CSC for awarding my scholarship that supported my stipend for me to complete my PhD study at NCSU. I am also thankful to Dr. Xie for using grants to support my tuition to complete PhD study.
I express my deep appreciation to my graduate school representative, Dr. David Jordan, for witness my prelim and final exam.

I want to deeply express my gratitude to my parents and my young sister for their support during my PhD study. Since I joined NCSU, I have not visited them. I own them too much. I am thankful to them for their understanding. I thank all my family members for their support.

Finally, I am grateful to our excellent PMB for offering me this great opportunity to pursue my PhD study at NCSU. I am grateful to our great Wolfpack for my PhD study. Go Wolfpack!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Flavan-3-ol, Proanthocyanidin, and Their Medicinal Values

Abstract

Proanthocyanidin (PA) and flavan-3-ol (flavanol) are two groups of plant flavonoids. They play multiple roles in plant's defense against biotic and abiotic stresses. They also have multiple health benefits for humans. In the past decades, flavan-3-ols and PAs have gained a systematic investigation in chemistry, biochemistry, genetics, and evolution. Herein, I review the current understanding of flavan-3-ol and PA biosynthesis as well as their medicinal applications. In addition, several key questions are discussed to enhance the elucidation of the biosynthesis of PAs in the future.

1. Introduction

Flavan-3-ols and proanthocyanidins (PAs) are two groups of plant flavonoids. PAs, also named as condensed tannins, are oligomers or polymers of flavan-3-ols and prevalent in plants [1, 2]. These two groups are prevalent in the plant kingdom. They are biosynthesized in leaves, stems, roots, seeds, or/and fruits. In general, the two groups of compounds provide plant protection against UV-irradiation caused damage, insects, pathogens, and herbivores [1, 2]. Flavan-3-ols and PAs are powerful antioxidants. The uptake of flavan-3-ols and PAs gives multiple health benefits to humans. They can inhibit cancer, prevent from cardiovascular diseases, decrease the oxidation of cholesterol linoleate in LDL, inhibit virus infection, and inhibit SARS-Cov-2 replication [3-6]. Herein I reviewed and discussed the progresses in natural resource, analytic methods, biosynthetic pathway, and medicinal benefits of flavan-3-ols and PAs. Questions related to the unknown steps of the PA biosynthesis are discussed to provide information for future researches.

2. Flavan-3-ols and PAs in common food and diet products

PAs and flavan-3-ols exist in daily diets such as tea, grape, cacao, and apple. Due to their potent nutritional values, numerous studies were completed to quantify their contents in different food and diet products. Green tea is a major source of flavan-3-ols and oligomeric PAs [7-9]. The major flavan-3-ols in green tea are flavan-3-ol (flavanol) gallates such as epigallocaechin-3-O-gallate (EGCG), the most abundant catechins in tea, though research data might be variable from different tea cultivars. One study showed that the EGCG contents young leave of green tea were

approximately 20% (dry weight, DW) [10]. Other flavanols include epicatechin, epigallocatechin, catechin, gallocatechin, and epicatechin-3-O-gallate. In addition, B-type and A-type PA oligomers as well as PA gallates exist in tea leaves [5]. Grape berries are rich in flavan-3-ols and PAs. Although cultivars, climates, development of berries, and agricultural practices affect the formation of PAs and flavan-3-ols [11-13], the total contents of flavan-3-ols and PAs can reach approximately 1.2 g/kg (fresh weight, FW) and PAs 4g/kg (FW), respectively. In wine industry, PAs and flavan-3-ols are main compounds associated with the high quality of red wine products. They react with red anthocyanidin pigments via condensation and oxidation reactions to stabilize red color and provide the astringent and bitter tastes of high valuable red wines [14]. Cacao and apple are abundant in flavanols and PAs. Crude cacao powder contains approximately 3.5 to 4.5 mg/g (DW) flavan-3-ols and 30 to 50 mg/g PAs (DW) [15-17]. Green and unripe apples were reported to contain 0.82 mg/g FW flavan-3-ols and 15.1 mg/g FW PAs, although the contents of two decreased to 0.43 mg/g FW and 4.5 mg/g FW after fully ripe, respectively [18, 19]. In addition, flavan-3-ols and PAs commonly exist in strawberry, blue berry, cranberry, raspberry, and numerous other fruits.

3. Structures of flavan-3-ols and proanthocyanidins

Flavan-3-ols consist of two benzo rings (A and B-ring) and one heterocyclic ring (C-ring). In common flavan-3-ols, the A ring contains two hydroxyl groups at C5 and C7 positions and the C ring contains one hydroxyl group at C3 position. The hydroxyl groups numbers on the B-ring classify flavan-3-ols to three common categories. One hydroxyl group at C4' position define structures to be afzelechin or epiafzelechin. Two hydroxyl groups at C4' and C5' positions

classify structures to be catechin and epicatechin. Three hydroxyl groups at C3', C4', and C5' positions term structures to be gallocatechin or epigallocatechin [1, 2] (Fig. 1). Flavan-3-ols are featured with two chiral structures at C2 and C3 positions, which creates four diastereoisomers, 2R, 3S-*trans*-, 2S, 3R-*trans*-, 2R, 3R-*cis*-, and 2S, 3S-*cis* flavan-3-ols. Of these, 2,3-*trans*-2R, 3S (e.g. (+)-catechin) and 2,3-*cis*-2R, 3R (e.g. (-)-epicatechin) are more common than 2, 3- *trans*-2S, 3R (e.g. (-)-catechin) and 2,3-*cis*-2S, 3S (e.g. (+)-epicatechin) in plants [20, 21].

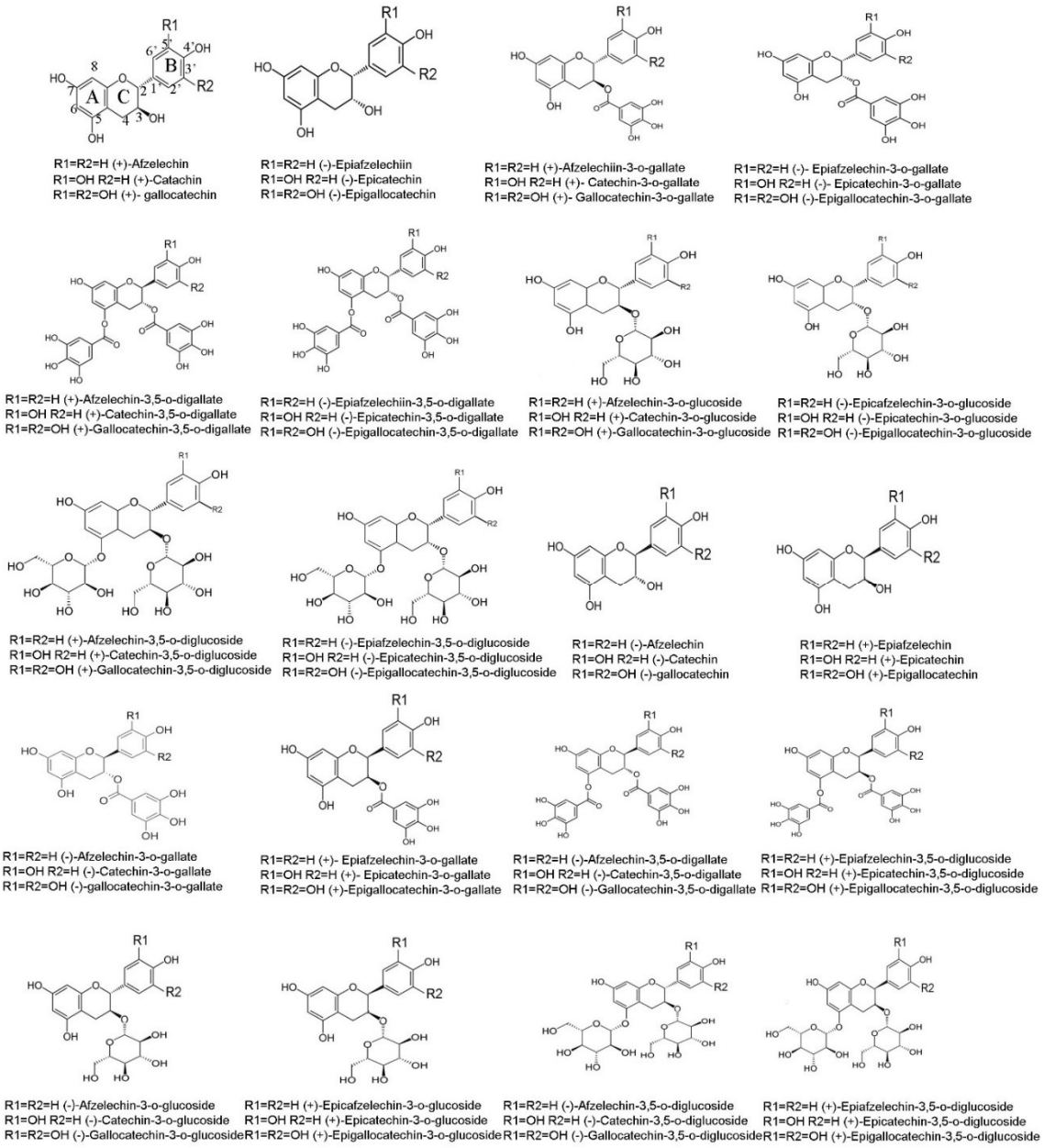


Figure 1 Structures of flavan-3-ol examples and their derivatives.

Flavan-3-ols are diversified by different modifications at the side groups. One common modification is acylation occurred at the –OH group at C3. Gallic acid and monosaccharides are

the main donors to form ethyl structures. In tea plants, the galloylation of EGC and EC form (-)-epigallocatechin 3-O-gallate (EGCG) and ECG [8, 22]. Other gallate derivatives reported in green tea include (+)-gallocatechin-3-O-gallate (GCG), and (+)-catechin-3-O-gallate (CAG) [8, 23]. The acylation of flavan-3-ols is also a common modification in other crops, such as (-)-epicatechin-3-O-gallate in bunch grape [24] and EGCG in muscadine grape [25]. In addition to –OH at C3 position, the –OH group at C5 is another acylation location. Two example compounds are epicatechin-3, 5-O-digallate and epigallocatechin-3,5-O-digallate reported in tea plants [26]. Glycosylation of flavan-3-ols is common in plants. Flavan-3-ol glycosides isolated from plants include epicatechin-3-O-glucoside [27-30], catechin 5-O-glucoside, and catechin-3-glycoside [31-33], and catechin 7-glucoside [27].

PAs are oligomers or polymers of flavan-3-ols. Based on the number of linkage between two flavan-3-ol units, PAs are classified into two different types: B-type and A-type. B-type is characterized by one interflavan bond linkage between the C4 position of an upper unit of extension unit and the C8 or C6 position of a lower unit or the start unit (Figure 2). A-type PAs are featured with an interflavan bond and an ether linkage between the C2 position of an upper unit and the hydroxyl group at the C7 position of a lower unit. PAs are further classified into different groups based on the number of hydroxyl groups on the B-ring, procyanidin, propelargonidin, and prodelphinidin, which are composed of catechin and/or epicatechin with two hydroxyl groups at C’3 and C’4, afzelechin and/or epiafzelechin with one hydroxyl group at C’4, and gallocatechin and/or epigallocatechin with three hydroxyl groups at C’3, C’4 and C’5, respectively. Based on PA structures identified, B-type procyanidins are dominant in the plant kingdom. In PAs, a number of dimeric ones are important nutrients in diets. Four common

nutritional dimeric procyanidins include procyanidin B1, B2, B3 and B4, which consist of catechin and epicatechin with a linkage between C4 to C8. Four other dimeric procyanidins are B5, B6, B7 and B8 consisting of catechin and epicatechin with a linkage between C4 to C6 [1, 2]. Two common dimeric A-types are procyanidin A1 and A2 (Fig. 2), which are abundant in cranberry [34, 35].

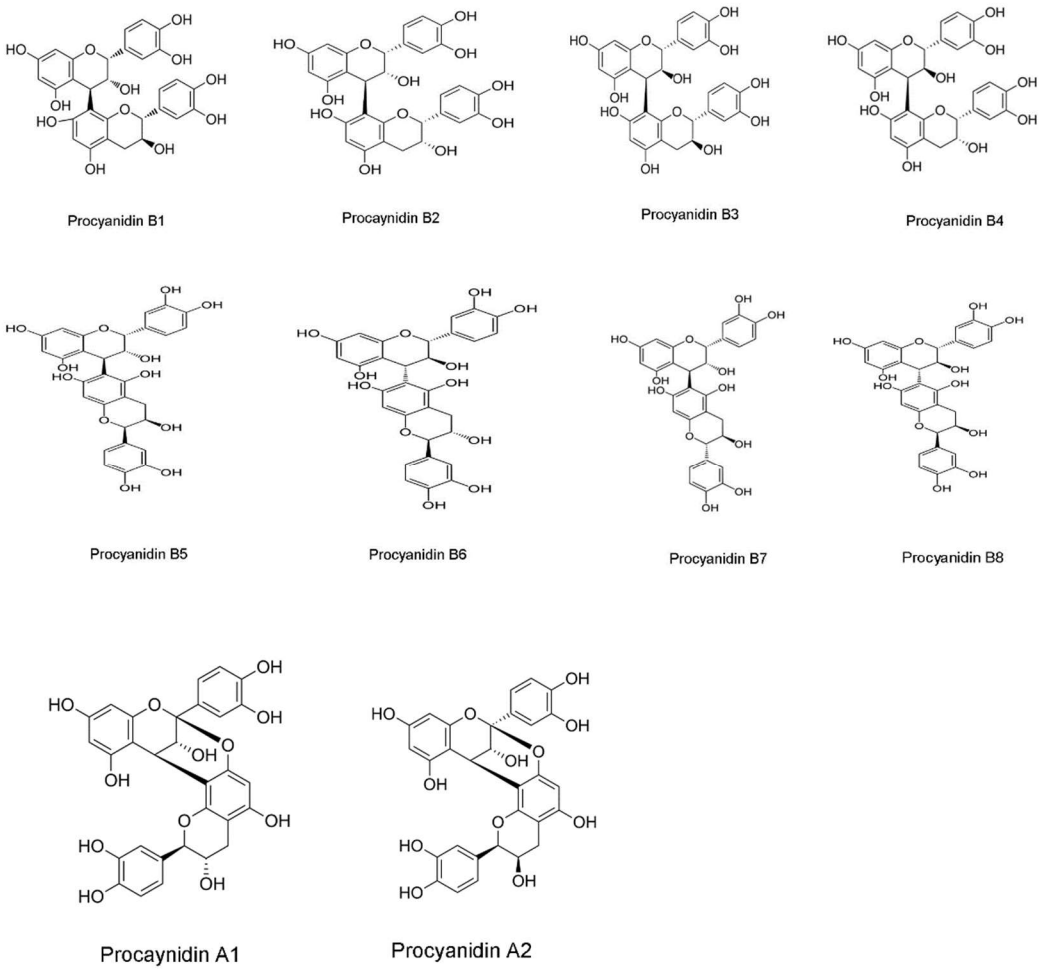


Figure 2 Structures of dimeric procyanidin B1-B8, and procyanidin A1 and A2.

4. Methodology for analysis of flavan-3-ols and PAs

Hot acidic alcohol cleavage is a classic approach to analyze PAs. This method uses acidic alcohol to boil PAs [36]. To date, this method has been modified to use 5% HCl in butanol to boil PAs [37, 38]. After boiling, PAs release anthocyanidins. This was the reason that PAs were considered as precursors of anthocyanidins until the discovery of ANR, which was reported anthocyanidins to be the precursors of flavan-3-ols and PAs [39]. In addition, given that the acidic boiling released three types of anthocyanidins, cyanidin, pelargonidin, and delphinidin, PAs were termed as procyanidin, propelargonidin, and prodelphinidin [38]. The chemical mechanism behind this method is still unclear. One hypothesis is that the extension units of PAs are released in the acidic alcohol as a carbocation form, which is further oxidized to anthocyanidin. An indirect evidence is that oligomeric PAs in red wine spontaneously release their extension units that form flavan-3-ol carbocation. Unfortunately, no evidence shows that the direct oxidation of flavan-3-ol carbocation produces anthocyanidins in this condition.

Acid based catalytic degradation of B-type PAs is an approach to characterize extension units and polymerization degrees. The mechanism is that PAs release their extension units as flavan-3-ols carbocation in acidic alcohol. When nucleophilic compounds, such as phloroglucinol or cysteine, are added to the reaction, they attack carbocation to form flavan-3-ol conjugates. This method is not applicable for A-type PAs [40, 41].

Vanillin-HCl reaction is a method to visualize the presence of PAs and flavan-3-ols. Vanillin reacts with these compounds to generate red pigments. This method is used to examine of PAs

and flavan-3-ols on thin layer chromatography (TLC) plates and in plant tissues [42]. However, the disadvantage of this method is that vanillin can react with other compounds, which lead to false positive results.

Dimethylaminocinnamaldehyde (DMACA) based bluish coloration is an effective approach for visualization and estimation of total PAs and flavan-3-ols [43-45]. DMACA reacts with PAs and flavan-3-ols to generate bluish coloration in the reactions. It is an effective approach to detect PAs and flavanols on TLC plates and in plant tissues [39, 46, 47]. The bluish compounds formed between DMACA and flavan-3-ols and PAs have a specific absorption at 640 nm for quantification of these compounds [48]. In addition, these bluish compounds are useful to analyze PAs and flavan-3-ols with HPLC.

Column chromatography (CC) is a common method to purify flavan-3-ols and PAs. Two types of CCs are commonly used, silica gel and LH-20 Sephadex gel. The separation of compounds on silica gel is based on their polarity. Two separation methods with silica gel regular CC and HPLC. LH-20 gel is used for size CC. Compounds with different molecular weights can be effectively separate with LH-20 CC. These methods are normally combined to purify flavan-3-ols and PAs extracted from plants.

HPLC, HPLC-MS, and LC-qTOF-MS/MS are effective approaches to identify known and annotate unknown PAs and flavan-3-ols. HPLC and HPLC-MS are effective for both identification and quantification of known flavan-3ols and PAs. LC-qTOF-MS/MS is efficient to annotate unknown compounds. Our laboratory recently developed protocols to annotate flavan-

3-ols and PAs [25]. The protocols include ionization modes, voltage, and collision induced dissociation (CID) for metabolic profiling. MS fragmentation patterns of flavan-3-ols and PAs provide useful information to analyze flavan-3-ols and PAs. Unique tandem mass fragmentations of PAs and flavan-3-ols were characterized by four types of structural fission patterns, heterocyclic ring fission (HRF), quinone methide (QM), Retro-Diels-Alder (RDA), and interflavan bond fission (IBF) [25]. These patterns can allow to annotate skeletons of flavan-3-ols and PAs. In addition, Public databases, such as SpactraBase (https://spectrabase.com/), Human Metabolome Database (HMDB) (https://hmdb.ca/), and MassBank of North America (MoNA) (https://mona.fiehnlab.ucdavis.edu/) provide useful MS references for predicting potential new flavan-3-ols and PAs in plant extracts.

5. Biosynthesis of flavan-3-ols and PAs

The biosynthesis of flavan-3-ols and PAs has been studied for decades. Herein, I summarize main biosynthetic steps that have been elucidated biochemically, genetically, and molecularly. Most steps are shared with the biosynthesis of anthocyanins and in general, the entire pathway is classified into three main sections, beginning, early, and late steps [49].

The beginning steps are shared with lignin and other phenylpropanoids starting from phenylalanine or tyrosine to 4-coumaroyl-CoA. Phenylalanine is converted to cinnamic acid by phenylalanine ammonia-lyase (PAL) [50, 51], followed by the conversion of cinnamic acid to 4-coumaric acid catalyzed by cinnamate 4-hydroxylase (C4H), a cytochrome P450 enzyme [52-54]. The steps from tyrosine is a side pathway, in which a tyrosine ammonia-lyase (TAL)

catalyzes L-tyrosine to 4-coumaric acid [54, 55]. A 4-coumarate: CoA ligase links CoA to 4-coumaric acid's side chain to form 4-coumaroyl-CoA, from which which the pathway switches to different branches to biosynthesize different polyphenols, such as lignan, lignin, and other phenylpropanoids [56, 57].

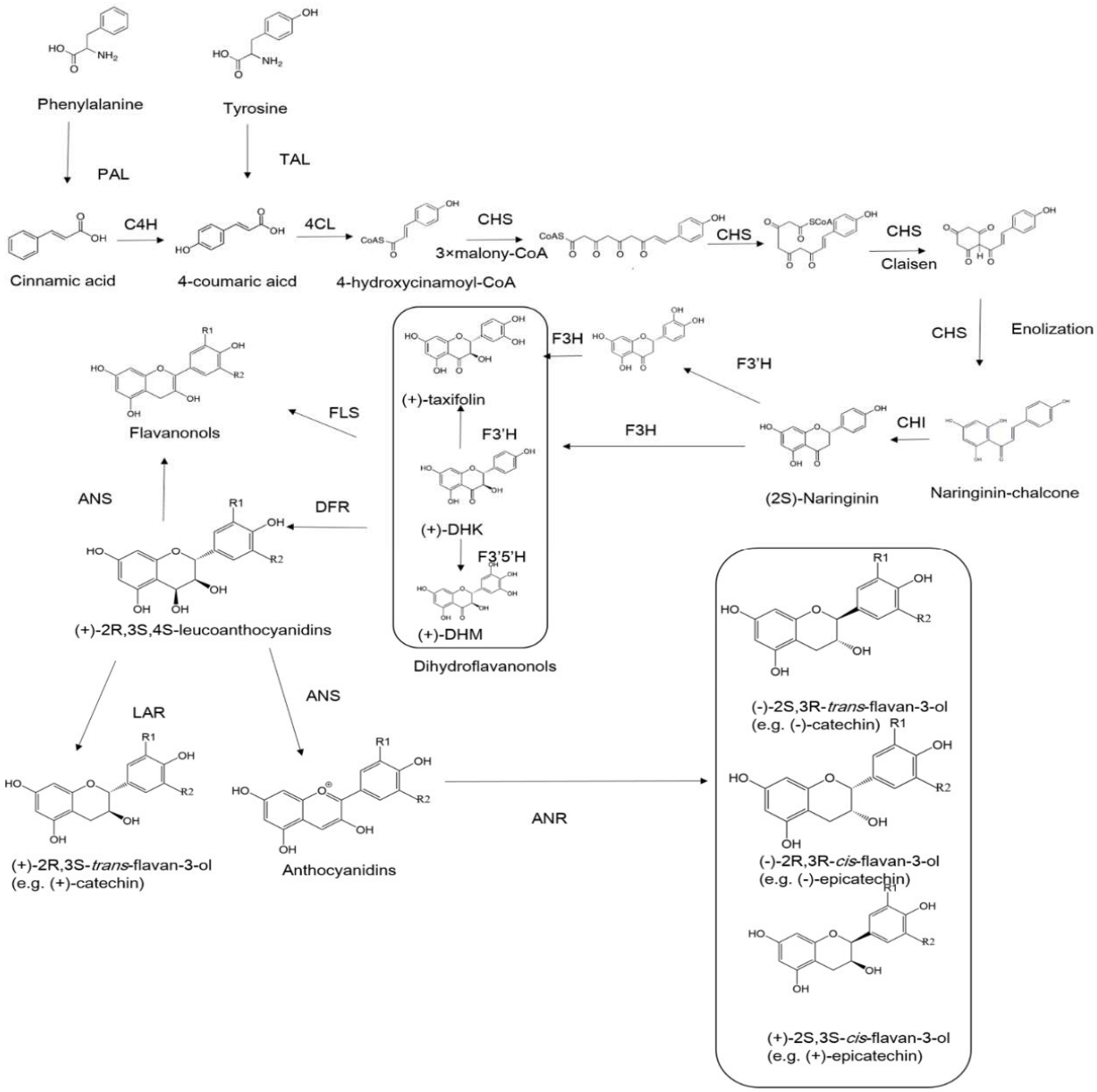
The early steps start with 4-coumaroyl-CoA. Three malonyl-CoA molecules are condensed to 4-courmaroyl-CoA to form chalcone catalyzed by a chalcone synthase, a member of type III polyketide synthase [58, 59]. Next, chalcone isomerase (CHI) converts chalcone to naringenin [60], a key flavonone, from which three branches are separated to flavones, isoflavonoids, and dihydroflavonols. Toward flavan-3-ols and PAs, flavonone-3-hydroxylase (F3H), a P450 member, hydroxylates flavonone at C-3 to a dihydroflavonol (DHK) molecule, dihydrokaempferol [61, 62], which is further hydroxylated on the B-ring to form dihydroquercetin (DHQ) and dihydromyricetin (DHM) by flavoid-3'-hydrosylase and flavonoid-3'5'-hydroxylase [63-67]. In brief, F3'H and F3'5'H hydroxylate B-ring at C3' to form DHQ (taxifolin) and C3' and C5' positions to form DHM, respectively. Then, followed are two branches, one to flavonols catalyzed by flavonol synthase (FLS) [68, 69] and the other to the late steps of anthocyanins, PAs, and flavan-3-ols described in the following.

The late steps include two crossing braches. One is from dihydroflavonols to three stereo types of flavan-3-ols and PAs via leucoanthocyanidins and anthocyanidins. The other is from dihydroflavonols to one stereo type of flavan-3-ol and PAs via leucoanthocyanidins. Enzymes that catalyze the first one include dihydroflavonol reductase (DFR), anthocyanidin synthase/leucoanthocyanidin dioxygenase (ANS/LDOX), and anthocyanidin reductase (ANR)

[39, 70, 71]. DFR reduces the carbonyl group of dihydroflavonols to hydroxyl group in the presence of NADPH or NADH to form leucoanthocyanidins [72]. ANS/LDOX dehydroxylates leucoanthocyanidins to anthocyanins [73]. ANR reduces anthocyanidins to three stereo types, 2R, 3R-*cis*, 2S,4R-*trans*, and 2S, 3S-*cis* flavan-3-ols [39, 48, 71, 74-76]. All steps in the first branch have been demonstrated by integrated studies of biochemistry, genetics, molecular biology, and metabolic engineering. Enzymes that catalyzes the second one include DFR and leucoanthocyanidin reductase (LAR) [70, 77]. LAR reduces leucoanthocyanidins to 2R, 3S-trans-flavan-3-ols. However, the genetic and transgenic evidence for LAR is still lacking.

Figure 3 The biosynthetic pathway from phenylalanine or tyrosine to flavan-3-ols and PAs.

PAL: phenylalanine ammonia-lyase. TAL: tyrosine ammonia-lyase. C4H: cinnamate 4-hydroxylase. 4CL: 4-coumaroyl: CoA-ligase. CHS: chalcone synthase. CHI: chalcone isomerase. F3'H: flavonoid 3'-hydroxylase. F3H: flavanone 3-hydroxylase. F3'5'H: flavonoid 3'5'-hydroxylase. FLS: flavonols synthase. DFR: dihydroflavonols 4-reductase. LAR: leucoanthocyanidin reductase. ANS: anthocyanidin synthase. ANR: anthocyanidin reductase. (+)-taxifolin: (+)-2R,3R-*trans*-taxifolin. (+)-DHK: (+)-2R,3R-*trans*-dihydrokaempferol. (+)-DHM: (+)-2R,3R-*trans*-dihydromyricetin.


Phenylalanine
Tyrosine
PAL
TAL
C4H
4CL
CHS
3×malony-CoA
CHS
CHS
Claisen
Cinnamic acid
4-coumaric aicd
4-hydroxycinamoyl-CoA
CHS
Enolization
F3H
F3'H
(+)-taxifolin
F3H
CHI
FLS
F3'H
Flavanonols
(2S)-Naringinin
Naringinin-chalcone
ANS
DFR
(+)-DHK
F3'5'H
(+)-DHM
(+)-2R,3S,4S-leucoanthocyanidins
Dihydroflavanonols
LAR
ANS
ANR
(-)-2S,3R-*trans*-flavan-3-ol
(e.g. (-)-catechin)
(+)-2R,3S-*trans*-flavan-3-ol
(e.g. (+)-catechin)
Anthocyanidins
(-)-2R,3R-*cis*-flavan-3-ol
(e.g. (-)-epicatechin)
(+)-2S,3S-*cis*-flavan-3-ol
(e.g. (+)-epicatechin)

Recently, the steps from leucoanthocyanidins to flavan-3-ols were interestingly revised due to unexpected functions of LAR and LDOX as well as a dual function of ANR. LAR was found to catalyze cysteine-epicatechin to epicatechin, which was proposed to associate with the non-extractable PA biosynthesis in *Medicago truncatula* [78]. LDOX was demonstrated to use 2R, 3S-*trans*-flavan-3-ols (e.g. (+)-catechin) as substrate to produce anthocyanidins in *M. truncatula* [79]. Flav-2-en-3-ol was proposed to be an intermediate during the catalysis of anthocyanidins to flavan-3-ols by ANR [80]. Recent findings demonstrated this hypothesis [81, 82]. Furthermore, a dual function of ANR is that it also reduces anthocyanidins to 2, 3-*cis*-leucoanthocyanidins, which are proposed to be precursors of PA extension units [81]. These new progresses provide insights to appropriately understand the biosynthesis of PAs.

6. Biogenesis of chiral configurations of flavan-3-ols

Flavan-3-ols have four different chiral configurations at C2 and C3 positions. These include 2R,3R-2,3-*cis*-(e.g. (-)-epicatechin), 2R,3S-2,3-*trans*-(e.g. (+)-catechin), 2S, 3R-2,3-*trans*-(e.g. (-)-catechin), and 2S, 3S-2,3-*cis*- (e.g. (+)-epicatechin), which are introduced by either ANR or LAR. Anthocyanidins have a double bond between C2 and C3, which has no chiral configurations. When ANR reduces anthocyanidins, one hydrogen is introduced at C2 from either alpha or beta phase to form 2S or 2R and the other is introduced at C3 from alpha or beta phase to form 3S or 3R. As a result, 2R, 3R-2, 3-*cis*-, 2S, 3R-2, 3-trans-, and 2S, 3S-2, 3-*cis*-flavan-3-ol isomers have been reported from different ANR assays [39, 80, 83]. Of these three, the 2R, 3R-2, 3-*cis*- configuration commonly exists in different plants. In comparison, 2S, 3S-2, 3-*cis*-flavan-3-ol is uncommon in plants. The biogenesis of 2R,3S-2,3-*trans*- flavan-3-ol formed

by the LAR catalysis starts with the conversion of naringenin-chalcone to naringenin, (2S)-4',5,7-trihydroxyflavan-4-one), which introduces a 2S (equivalent to 2R in flavan-3-ols) chirality at C2 position. In the next step, F3H introduces the second chirality at C3 position, leading to form a (+)-2R, 3R-*trans* configuration (e.g. (+)-taxifolin), the dominant configuration of dihydroflavonols. The catalysis of DFR prefers to (+)-2R, 3R-*trans* configuration and maintain this to leucoanthocyanidins. In addition, DFR adds the third chirality at C4 position. As a result, leucoanthocyanidins have two main chiral configurations characterized by chemical syntheses, (+)-2,3-*trans*-3,4-*cis* (2R,3S,4S-*cis*) and (+)-2,3-*trans*-3,4-*trans* (2R,3S,4R-*trans*). LAR has been demonstrated to prefer to reduce C4 of (+)-2,3-*trans*-3,4-*cis*-leucoanthocyanidins to 2,3-*trans*-flavan-3-ols, such as (+)-catechin. Besides, LAR has a strict stereo preference. For example, a recombinant LAR from grape was shown to accept (+)-2,3-*trans*-3,4-*cis*-leucocyanidin as the only substrate to produce (+)-catechin [84]. In comparison, ANS has a relative flexibility to catalyze substrate. For example, ANS homologs from *Arabidopsis* can use Both (+)-2,3-*trans*-3,4-*cis*-leucocyanidin and (+)-2,3-*trans*-3,4-*trans* (2R,3S,4R-*trans*) as substrates. On the one hand, ANS was shown to convert (+)-2,3-*trans*-3,4-*cis*-leucocyanidin to quercetin [85]. On the other hand, ANS uses (+)-2,3-*trans*-3,4-*trans*-leucocyanidin as substrate to produce cyanidin *in vitro* [86]. The two chiral configurations of leucocyanidin were observed in their precursors, dihydroflavonol and derivatives, which were extracted from plants. (-)-2S, 3S-*trans*-dihydroquercetin-3'-O-D-glucopyranoside was isolated from *Pinus densiflora* [87], (-)-2S, 3S-*trans*-taxifolin-3-O-D-glucopyranoside were identified from root-sprouts of *Agrimonia pilosa* [88]. (+)-2S, 3R-*cis*-taxifolin and (-)-2R,3S-*cis*-taxifolin 3-O-D-xylopyranosides were isolated from leaf extracts of *Thujopsis dolabrata* [89].

7. Role of flavan-3-ol carbocation in the formation of PAs

How plants polymerize flavan-3-ols to PAs is still a mystery. One main problem unsolved is that although it is clear that the start units of PAs are flavan-3-ols, the origin of extension units remain unknown. To address this problem, a few hypotheses were proposed in the past. Unstable flavan-3-ol carbocation was proposed to be the key intermediate to form the extension units of PAs [90]. To date, this carbocation is the key to form hypotheses to interpret the biosynthesis of PAs. One hypothesis is that flavan-3-ol carbocation is attacked by nucleophilic flavan-3-ols, the result of which is the formation of an interflavan bond between C4 of an upper unit and C8 of a bottom unit. The first evidence for this hypothesis resulted from a chemical synthesis that used (+)-leucocyanidin and (+)-catechin to produce dimeric catechin-catechin. In a 0.1N HCl methanol solvent, (+)-leucocyanidin was converted to (+)-catechin carbocation. When (+)-catechin was added to the mixture, a dimeric procyanidin consisting of catechin-catechin was produced [90]. A mechanism proposed for this chemical synthesis was that leucoanthocyanidins was converted to 4-quinone methide, followed by a spontaneous proton shift to form flavan-3-ol carbocation (Figure 4). Although no biochemical and genetic experiments proved this mechanism, (-)-epicatechin carbocation was recently trapped from plant tissues. This datum demonstrated the presence of flavan-3-ol carbocation in seeds of *Arabidopsis thaliana* and *CsANR* transgenic tobacco [82].

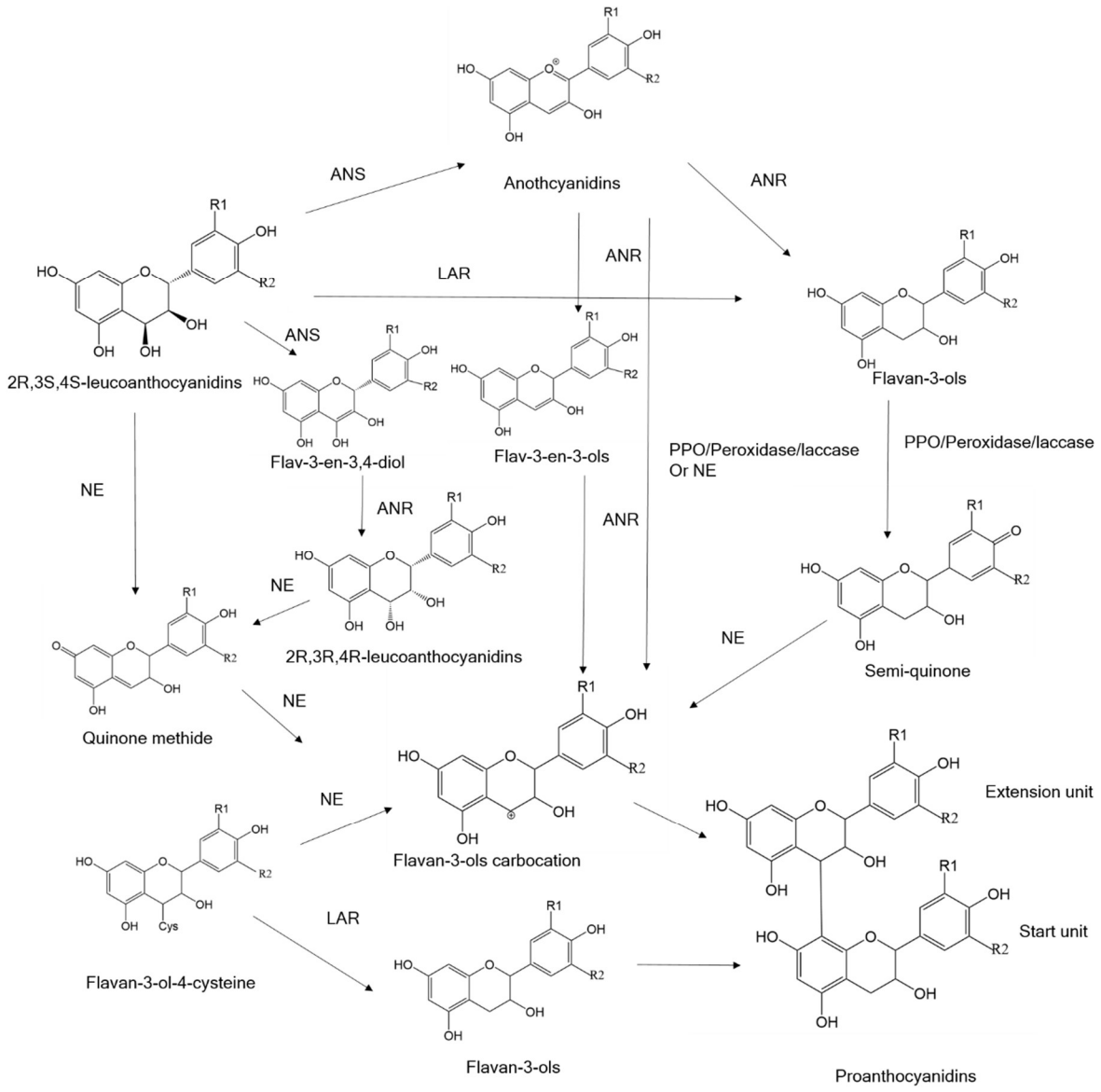


Figure 4 A scheme showing the formation of flavan-3-ol carbocation proposed from different substrates.

LAR: leucoanthocyanidin reductase. ANS: anthocyanidin synthase. ANR: anthocyanidin reductase. PPO: polyphenol oxidase. NE: non-enzymatic reaction.

Although the success of chemical synthesis of dimeric PAs proposes that leucoanthocyanidins be precursors of flavan-3-ol carbocations, this hypothesis cannot explain the origin of the 2R,3R-*cis* configuration that dominates in extension units of PAs. In addition, the biosynthesis of PAs in *Arabidopsis* does not support this hypothesis. On the one hand, *Arabidopsis* only produces (-)-

epicatechin with the 2R,3R-*cis* configuration and PAs in this model plant consist of (-)-epicatechin (e.g. procyanidin B2) only. On the other hand, *Arabidopsis* ANS was shown to only use 2R, 3S-*tran*s leucoanthocyanidin as substrate [91]. This observation suggests that 2R, 2R-2, 3-cis-flavan-3-ol carbocation must result from other flavonoids [92]. In 2005, Xie et al proposed a new model to interpret the biogenesis of flavan-3-ol carbocation in the plants. Flavan-3-ols and anthocyanidins could be two precursors of flavan-3-ol carbocation. Polyphenol oxidase, laccase, or peroxidase oxidizes flavan-3-ols and anthocyanidins to form semiquinone and quinone methides, which are spontaneously converted to flavan-3-ol carbocation via flav-3-en-3-ols. To date, on the one hand, the evidence of flavan-3-ol carbocation has been obtained from biochemical and chemical experiments and transgenic plants. Trapping experiments detected flavan-3-ol carbocation in ANR reactions, acidic degradation of PB, and plant tissues. On the other hand, enzymes with the proposed function have not been found from plants [1, 2]. This unknown mechanism indicates that, continuous efforts are necessary to understand the polymerization mechanism of PAs.

Flavan-3-ol carbocation was recently trapped in the ANR catalysis in vitro and ANR transgenic plants for the first time. Flavan-3-ol carbocation was proposed to exist in the reduction of anthocyanidins catalyzed by ANR [80]. A recent study reported the trapping of flavan-3-ol carbocation with nucleophilic cysteine [82]. When cysteine was added into the enzymatic incubation of cyanidin and ANR, 4-cystainyl-(-)-epicatechin was formed. In addition, when an extraction buffer supplemented with cysteine was used to extract samples of *ANR* transgenic tobacco, in addition to (-)-epicatechin and PAs, a new 4-cystainyl-(-)-epicatechin was extracted from samples [82]. These trapping experiments provided evidence that (-)-epicatechin

carbocation formed *in vitro* and *in vivo*. Another recent experiment showed a dual activity of ANR that it reduced flav-3-en-3,4-diol to 2,3-*cis*-3,4-*cis*-leucoanthocyanidin. Then 2,3-*cis*-3,4-*cis*-leucoanthocyanidin lose the –OH at C4 to form flavan-3-ol carbocation, which was proposed to form extension units of PAs [81]. Moreover, Flavan-3-ol-4-cysteine was observed in a *lar* mutant of *M. truncatula*. The mutation of LAR led to the increase of flavan-3-ols but the reduction of non-extractable PAs. Enzymatic assays demonstrated that LAR reduced flavan-3-ol-4-cysteine to flavan-3-ol likely via its flavan-3-ol carbocation [93].

8. Health benefits of flavan-3-ols and PAs

Flavan-3-ols and PAs are two main top groups of flavonoids that have gained a large number of pharmaceutical and nutritional studies to understand their benefits to human health. The main rationale is that two groups are strong antioxidants existing in daily diets . In addition to numerous compliments in evaluating their antioxidative activity, a large number of studies have been completed to understand their functions in anti-cardiovascular, anti-cancer, anti-diabetes, and antiviral diseases.

The potential application for preventing against cardiovascular diseases is of significance to human health [94]. A few problems, such as excessive free radicals, oxidative stress, high blood cholesterol levels, and high LDL levels, are associated with different cardiovascular diseases [94, 95]. Any improvements of these problems can reduce the risk of cardiovascular diseases, such as high blood pressure and atherosclerosis. The past studies demonstrated that flavan-3-ols and PAs effectively removed free radicals such as superoxide and nitric oxide in plasma [96], lowered

down the levels of total cholesterol, inhibited LDL oxidation [97-99], and inhibited angiotensin I converting enzyme (ACE) to improve high blood pressure [100].

Flavan-3-ols and PAs have been shown to possess promising anti-cancer activities. *In vitro* studies showed that PAs and flavan-3-ols from grape and cranberry inhibited the proliferation of H460 human large cell lung carcinoma and K562 chronic myelogenous leukemia cell [101-103]. Mechanistic studies reported that two groups of compounds induced apoptosis of cancerous cells [103]. Molecular studies showed PAs and flavan-3-ols increased the expression of Bax and Caspase-3, two proteins required for DNA fragmentation and apoptosis of Hela cells [104]. PAs and flavan-3-ols are potential pharmaceuticals to help improve type 2 diabetes. Improvement of blood sugar concentrations is an important treatment of diabetes. PAs and flavan-3-ols extracted from persimmon were shown to not only reduce high blood sugar concentration but also improve hyperlipidemia of mice with type 2 diabetes [105]. PA and flavan-3-ol extracts from grape seeds were demonstrated to significantly reduce glycation end products, blood urea nitrogen, and creatinine in diabetic rats [106, 107]. Another study showed crude PAs and flavan-3-ols extracts from longan flowers decreased the systolic blood pressure and enhanced the expression of several insulin-signaling related proteins in insulin-resistant rats [108].

PAs and flavan-3-ols have promising anti-viral activities. These compounds or their extracts have been tested for activities against influenza, hepatitis, herpes virus, HIV, and SARS. SP-303, an oligomeric PA extract from *Croton lechleri*, was shown to inhibit replication of both DNA and RNA viruses in cell cultures, such as respiratory syncytial virus (RSV), influenza A (FLU-

A), parainfluenza virus (PIV), herpesvirus (HSV) type 1 and 2, and hepatitis A and B [109]. Further tests with animal models demonstrated that SP-303 inhibited RSV in infected cotton rats and African green monkeys, PIV-3 in infected mice and pigs, and FLU-A in mice [109]. Mechanistic studies revealed that PAs bound to viral envelope protein and inhibited viral attachment to the host cell membrane. Another mechanism was shown that A-type PAs inhibited viral RNA synthesis in host cells. For example, one experiment demonstrated that procyanidin A2 eliminated canine distemper virus replication in VERO cells by inhibiting its RNA-dependent RNA polymerase's activity [110]. A great number of experiments have demonstrated that the authentic EGCG is a fundamental pharmaceuticals to prevent or treat virus-infected diseases. Experimental data revealed that EGCG inhibited Zika virus, herpes virus, FLU-A, RSV, and HIV [111-114]. Furthermore, EGCG is effective to inhibit SARS [115, 116]. We recently reported that EGCG, three other flavan-3-ol derivatives, and PB2 effectively inhibited the main proteinase of SARS-Cov-2 [6]. Furthermore, our new data showed that these compounds inhibited coronavirus 229E replication in human cells.

9. Questions

The past studies have elucidated all biosynthetic steps from phenylalanine or tyrosine to flavan-3-ols and demonstrated that flavan-3-ols are the starter units of PAs. Recent experimental evidence further showed that the hypothetical flavan-3-ol carbocation existed in the ANR catalysis and plant cells. All these data greatly enhance understanding the biosynthesis of PAs. In particular, the demonstration of flavan-3-ol carbocation provides fundamental information to further elucidate the polymerization of PAs. Meanwhile, all these findings indicate that a few

important questions need to be addressed to fully elucidate the polymerization mechanism of PAs in the future.

If leucoanthocyanidins are the precursors of PA extension units, how are their stereo configurations enzymatically formed by DFR? If DFR can introduce these stereo configurations in vitro, do chiral configurations of leucoanthocyanidins exist in plant? A condensing synthase (CS) was proposed to be members of PPOs, lacases, or peroxidases. Which one is CS and how can it catalyze the polymerization of PAs?

Flava-3-ols, leucoanthocyanidins, and anthocyanidins are proposed to be precursors of PA extension units. Which groups are the main one, why and how?

Additional questions include how plants biosynthesize non-extractable cell wall-bound PAs and extractable PAs, how PAs bind to cell wall, and others.

REFERENCES

1. Dixon, R.A., D.Y. Xie, and S.B. Sharma, *Proanthocyanidins--a final frontier in flavonoid research?* New Phytol, 2005. **165**(1): p. 9-28.
2. Xie, D.Y. and R.A. Dixon, *Proanthocyanidin biosynthesis--still more questions than answers?* Phytochemistry, 2005. **66**(18): p. 2127-44.
3. Bagchi, D., et al., *Protective effects of grape seed proanthocyanidins and selected antioxidants against TPA-induced hepatic and brain lipid peroxidation and DNA fragmentation, and peritoneal macrophage activation in mice.* Gen Pharmacol, 1998. **30**(5): p. 771-6.
4. Renaud, S. and M. de Lorgeril, *Wine, alcohol, platelets, and the French paradox for coronary heart disease.* Lancet, 1992. **339**(8808): p. 1523-6.
5. Yamakoshi, J., et al., *Proanthocyanidin-rich extract from grape seeds attenuates the development of aortic atherosclerosis in cholesterol-fed rabbits.* Atherosclerosis, 1999. **142**(1): p. 139-49.
6. Zhu, Y. and D.-Y. Xie, *Docking characterization and in vitro inhibitory activity of flavan-3-ols and dimeric proanthocyanidins against the main protease activity of SARS-Cov-2.* Frontiers in Plant Science, 2020. **11**(1884).
7. Molan, A.L., S. De, and L. Meagher, *Antioxidant activity and polyphenol content of green tea flavan-3-ols and oligomeric proanthocyanidins.* Int J Food Sci Nutr, 2009. **60**(6): p. 497-506.
8. Jiang, X.L., et al., *Tissue-specific, development-dependent phenolic compounds accumulation profile and gene expression pattern in tea lant Camellia sinensis.* Plos One, 2013. **8**(4): p. 14.

9. Musial, C., A. Kuban-Jankowska, and M. Gorska-Ponikowska, *Beneficial properties of green tea catechins.* International Journal of Molecular Sciences, 2020. **21**(5): p. 11.

10. Hu, J., D. Zhou, and Y. Chen, *Preparation and antioxidant activity of green tea extract enriched in epigallocatechin (EGC) and epigallocatechin gallate (EGCG).* J Agric Food Chem, 2009. **57**(4): p. 1349-53.

11. Poudel, P.R., K. Koyama, and N. Goto-Yamamoto, *Evaluating the influence of temperature on proanthocyanidin biosynthesis in developing grape berries (Vitis vinifera L.).* Mol Biol Rep, 2020. **47**(5): p. 3501-3510.

12. Cortell, J.M., et al., *Influence of vine vigor on grape (Vitis vinifera L. Cv. Pinot Noir) and wine proanthocyanidins.* J Agric Food Chem, 2005. **53**(14): p. 5798-808.

13. Rinaldi, A., et al., *Metabolic and RNA profiling elucidates proanthocyanidins accumulation in Aglianico grape.* Food Chem, 2017. **233**: p. 52-59.

14. Dallas, C., J. Ricardo-da-Silva, and O. Laureano, *Degradation of oligomeric procyanidins and anthocyanins in a Tinta Roriz red wine during maturation.* Vitis, 1995. **34**(1): p. 51-56.

15. Ioannone, F., et al., *Flavanols, proanthocyanidins and antioxidant activity changes during cocoa (Theobroma cacao L.) roasting as affected by temperature and time of processing.* Food chemistry, 2015. **174**: p. 256-262.

16. Natsume, M., et al., *Analyses of polyphenols in cacao liquor, cocoa, and chocolate by normal-phase and reversed-phase HPLC.* Bioscience, biotechnology, and biochemistry, 2000. **64**(12): p. 2581-2587.

17. Seigler, D. and R. Lieberei, *3 Proanthocyanidins in Cacao.* Natural Antioxidants and Biocides from Wild Medicinal Plants, 2013: p. 35.

18. Hellstrom, J.K. and P.H. Mattila, *HPLC determination of extractable and unextractable proanthocyanidins in plant materials.* Journal of agricultural and food chemistry, 2008. **56**(17): p. 7617-7624.

19. Lister, C.E., et al., *Developmental changes in the concentration and composition of flavonoids in skin of a red and a green apple cultivar.* Journal of the Science of Food and Agriculture, 1994. **64**(2): p. 155-161.

20. Inanami, O., et al., *Oral administration of (-) catechin protects against ischemia-reperfusion-induced neuronal death in the gerbil.* Free radical research, 1998. **29**(4): p. 359-365.

21. Geiss, F., et al., *Proanthocyanidins with (+)-epicatechin units from Byrsonima crassifolia bark.* Phytochemistry, 1995. **39**(3): p. 635-643.

22. Liu, Y.J., et al., *Purification and Characterization of a Novel Galloyltransferase Involved in Catechin Galloylation in the Tea Plant (Camellia sinensis).* Journal of Biological Chemistry, 2012. **287**(53): p. 44406-44417.

23. Davis, A.L., et al., *1H and 13C NMR assignments of some green tea polyphenols.* Magnetic resonance in Chemistry, 1996. **34**(11): p. 887-890.

24. Da Silva, J.M.R., et al., *Procyanidin dimers and trimers from grape seeds.* Phytochemistry, 1991. **30**(4): p. 1259-1264.

25. Yuzuak, S., J. Ballington, and D.-Y. Xie, *HPLC-qTOF-MS/MS-based profiling of flavan-3-ols and dimeric proanthocyanidins in berries of two muscadine grape hybrids FLH 13-11 and FLH 17-66.* Metabolites, 2018. **8**(4): p. 57.

26. Kim, J.H., et al., *Inhibitory activity of (−)-epicatechin-3, 5-O-digallate on α-glucosidase and in silico analysis.* International journal of biological macromolecules, 2018. **107**: p. 1162-1167.

27. Vihakas, M., et al., *Flavonoid metabolites in the hemolymph of european pine sawfly (Neodiprion sertifer) larvae.* Journal of chemical ecology, 2012. **38**(5): p. 538-546.

28. Tang, J.S., et al., *Antioxidative Phenylpropanoid-Substituted Epicatechin Glycosides from Parabarium huaitingii.* Planta Medica, 2009. **75**(15): p. 1586-1590.

29. Grabber, J.H., D. Ress, and J. Ralph, *Identifying New Lignin Bioengineering Targets: Impact of Epicatechin, Quercetin Glycoside, and Gallate Derivatives on the Lignification and Fermentation of Maize Cell Walls.* Journal of Agricultural and Food Chemistry, 2012. **60**(20): p. 5152-5160.

30. Cui, C.B., et al., *Constitutents of fern, Davallia mariessi Moore. 4. Isolation and structures of a novel norcarotane sesquiterpene glycoside, a chromone glucuronide, and two epicatechin glycosides.* Chemical & Pharmaceutical Bulletin, 1992. **40**(8): p. 2035-2040.

31. Nonaka, G.-I., et al., *Flavanol glucosides from rhubarb and Rhaphiolepis umbellata.* Phytochemistry, 1983. **22**(7): p. 1659-1661.

32. Hosny, M., et al., *(-)-Catechin glycosides from Ulmus davidiana.* Archives of Pharmacal Research, 2014. **37**(6): p. 698-705.

33. Li, Y., et al., *A novel phenylpropanoid-substituted catechin glycoside and a new dihydrochalcone from Sarcandra glabra.* Chinese Chemical Letters, 2006. **17**(2): p. 207-210.

34. Sprygin, V.G. and N.F. Kushnerova, *Cranberry: a new source of oligomeric proanthocyanidins.* Pharmaceutical Chemistry Journal, 2004. **38**(2): p. 100.

35. Howell, A.B., et al., *A-type cranberry proanthocyanidins and uropathogenic bacterial anti-adhesion activity.* Phytochemistry, 2005. **66**(18): p. 2281-2291.

36. Swain, T. and W. Hillis, *The phenolic constituents of Prunus domestica. I.—The quantitative analysis of phenolic constituents.* Journal of the Science of Food and Agriculture, 1959. **10**(1): p. 63-68.

37. Dalzell, S.A. and G.L. Kerven, *A rapid method for the measurement of Leucaena spp proanthocyanidins by the proanthocyanidin (butanol/HCl) assay.* Journal of the Science of Food and Agriculture, 1998. **78**(3): p. 405-416.

38. Porter, L.J., L.N. Hrstich, and B.G. Chan, *The conversion of procyanidins and prodelphinidins to cyanidin and delphinidin.* Phytochemistry, 1985. **25**(1): p. 223-230.

39. Xie, D.-Y., et al., *Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis.* Science, 2003. **299**(5605): p. 396-399.

40. Matthews, S., et al., *Method for estimation of proanthocyanidins based on their acid depolymerization in the presence of nucleophiles.* Journal of Agricultural and Food Chemistry, 1997. **45**(4): p. 1195-1201.

41. Kennedy, J.A. and G.P. Jones, *Analysis of proanthocyanidin cleavage products following acid-catalysis in the presence of excess phloroglucinol.* Journal of Agricultural and Food Chemistry, 2001. **49**(4): p. 1740-1746.

42. Butler, L.G., M.L. Price, and J.E. Brotherton, *Vanillin assay for proanthocyanidins (condensed tannins): modification of the solvent for estimation of the degree of polymerization.* Journal of Agricultural and Food Chemistry, 1982. **30**(6): p. 1087-1089.

43. Thies, M. and R. Fischer, *New color reactin for microchemical detection and quantitative determination of catechins.* Mikrochim Acta, 1971. **59**(1): p. 9-13.

44. Treutter, D., *Chemical reaction detection of catechins and proanthocyanidins with 4-dimethylaminocinnamaldehyde.* Journal of Chromatography, 1989. **467**: p. 185-193.

45. Nagel, C.W. and Y. Glories, *Use of a modified dimethylaminocinnamaldehyde reagent for analysis of flavanols.* Am. J. Enol.Vitic, 1991. **42**: p. 364-366.

46. Xie, D.-Y., et al., *Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor.* Plant J, 2006. **45**(6): p. 895-907.

47. Zhu, Y., et al., *Molecular cloning and functional characterization of the anthocyanidin reductase gene from Vitis bellula.* Planta, 2014. **240**(2): p. 381-398.

48. Zhu, Y., et al., *Molecular cloning and functional characterization of the anthocyanidin reductase gene from Vitis bellula.* Planta, 2014. **240**(2): p. 381-398.

49. Shi, M.-Z. and D.-Y. Xie, *Biosynthesis and metabolic engineering of anthocyanins in Arabidopsis thaliana.* Recent Pat Biotechnol, 2014. **8**(1): p. 47-60.

50. Camm, E.L. and G.N. Towers, *Phenylalanine ammonia lyase.* Phytochemistry, 1973. **12**(5): p. 961-973.

51. Tanaka, Y., et al., *Structure and characterization of a cDNA clone for phenylalanine ammonia-lyase from cut-injured roots of sweet potato.* Plant physiology, 1989. **90**(4): p. 1403-1407.

52. Bell-Lelong, D.A., et al., *Cinnamate-4-hydroxylase expression in Arabidopsis (regulation in response to development and the environment).* Plant Physiology, 1997. **113**(3): p. 729-738.

53. Mizutani, M., D. Ohta, and R. Sato, *Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from Arabidopsis and its expression manner in planta.* Plant Physiology, 1997. **113**(3): p. 755-763.

54. Louie, G.V., et al., *Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases.* Chemistry & biology, 2006. **13**(12): p. 1327-1338.

55. Schwede, T.F., J. Rétey, and G.E. Schulz, *Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile.* Biochemistry, 1999. **38**(17): p. 5355-5361.

56. Hu, W.-J., et al., *Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (Populus tremuloides).* Proceedings of the National Academy of Sciences, 1998. **95**(9): p. 5407-5412.

57. Lavhale, S.G., R.M. Kalunke, and A.P. Giri, *Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants.* Planta, 2018. **248**(5): p. 1063-1078.

58. Austin, M.B. and J.P. Noel, *The chalcone synthase superfamily of type III polyketide synthases.* Natural product reports, 2003. **20**(1): p. 79-110.

59. Napoli, C., C. Lemieux, and R. Jorgensen, *Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans.* The plant cell, 1990. **2**(4): p. 279-289.

60. Jez, J.M., et al., *Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase.* Nature structural biology, 2000. **7**(9): p. 786-791.

61. BRITSCH, L. and H. GRISEBACH, *Purification and characterization of (2S)-flavanone 3-hydroxylase from Petunia hybrida.* European Journal of Biochemistry, 1986. **156**(3): p. 569-577.

62. Pelletier, M.K. and B.W. Shirley, *Analysis of flavanone 3-hydroxylase in Arabidopsis seedlings (Coordinate regulation with chalcone synthase and chalcone isomerase).* Plant physiology, 1996. **111**(1): p. 339-345.

63. Xu, B.-B., et al., *Cloning and molecular characterization of a functional flavonoid 3'-hydroxylase gene from Brassica napus.* Journal of plant physiology, 2007. **164**(3): p. 350-363.

64. Schoenbohm, C., et al., *Identification of the Arabidopsis thaliana flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme.* 2000.

65. Brugliera, F., et al., *Isolation and characterization of a flavonoid 3'-hydroxylase cDNA clone corresponding to the Ht1 locus of Petunia hybrida.* The Plant Journal, 1999. **19**(4): p. 441-451.

66. de Vetten, N., et al., *A cytochrome b5 is required for full activity of flavonoid 3', 5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors.* Proceedings of the National Academy of Sciences, 1999. **96**(2): p. 778-783.

67. Seitz, C., et al., *Cloning, functional identification and sequence analysis of flavonoid 3'-hydroxylase and flavonoid 3', 5'-hydroxylase cDNAs reveals independent evolution of flavonoid 3', 5'-hydroxylase in the Asteraceae family.* Plant molecular biology, 2006. **61**(3): p. 365-381.

68. Downey, M.O., J.S. Harvey, and S.P. Robinson, *Synthesis of flavonols and expression of flavonol synthase genes in the developing grape berries of Shiraz and Chardonnay (Vitis vinifera L.).* Australian Journal of Grape and Wine Research, 2003. **9**(2): p. 110-121.

69. Owens, D.K., et al., *Functional analysis of a predicted flavonol synthase gene family in Arabidopsis.* Plant physiology, 2008. **147**(3): p. 1046-1061.

70. Xie, D.-Y. and R.A. Dixon, *Proanthocyanidin biosynthesis - still more questions than answers?* Phytochemistry, 2005. **66**(18): p. 2127-2144.

71. Peng, Q.Z., et al., *An integrated approach to demonstrating the ANR pathway of proanthocyanidin biosynthesis in plants.* Planta, 2012. **236**(3): p. 901-18.

72. Xie, D.-Y., et al., *Molecular and biochemical analysis of two cDNA clones encoding dihydroflavonol-4-reductase from Medicago truncatula.* Plant Physiol, 2004. **134**(3): p. 979-994.

73. Saito, K., et al., *Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red forma of Perilla frutescens.* The Plant Journal, 1999. **17**: p. 181-189.

74. Zhu, Y., et al., *Functional characterization of an anthocyanidin reductase gene from the fibers of upland cotton (Gossypium hirsutum).* Planta, 2015. **241**(5): p. 1075-1089.

75. Xie, D.-Y., S.B. Sharma, and R.A. Dixon, *Anthocyanidin reductases from Medicago truncatula and Arabidopsis thaliana.* Archives of Biochemistry and Biophysics, 2004. **422**(1): p. 91-102.

76. Thirugnanasambantham, K., S. Muralidaran, and A.K.A. Mandal, *Molecular cloning, computational and expression analysis of anthocyanidin reductase in tea (Camellia sinensis).* Applied biochemistry and biotechnology, 2014. **174**(1): p. 130-145.

77. Tanner, G.J., et al., *Proanthocyanidin biosynthesis in plants. Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA.* J Biol Chem, 2003. **278**: p. 31647-56.

78. Liu, C.G., et al., *A role for leucoanthocyanidin reductase in the extension of proanthocyanidins.* Nat. Plants, 2016. **2**(12).

79. Jun, J.H., et al., *Proanthocyanidin subunit composition determined by functionally diverged dioxygenases.* Nature Plants, 2018. **4**(12): p. 1034-1043.

80. Xie, D.-Y., S.B. Sharma, and R.A. Dixon, *Anthocyanidin reductases from Medicago truncatula and Arabidopsis thaliana.* Arch Biochem Biophys, 2004. **422**(1): p. 91-102.

81. Jun, J.H., et al., *Dual activity of anthocyanidin reductase supports the dominant plant proanthocyanidin extension unit pathway.* Science Advances, 2021. **7**(20): p. eabg4682.

82. Wang, P., et al., *Functional demonstration of plant flavonoid carbocations proposed to be involved in the biosynthesis of proanthocyanidins.* The Plant Journal, 2020. **101**(1): p. 18-36.

83. Gargouri, M., et al., *Structure and epimerase activity of anthocyanidin reductase from Vitis vinifera.* Acta Crystallogr Sect D-Biol Crystallogr, 2009. **65**: p. 989-1000.

84. Zhang, J.R., et al., *Production of 3,4-cis- and 3,4-trans-Leucocyanidin and Their Distinct MS/MS Fragmentation Patterns.* J Agric Food Chem, 2018. **66**(1): p. 351-358.

85. Turnbull, J.J., et al., *Are anthocyanidins the immediate products of anthocyanidin synthase?* Chemical Communications, 2000(24): p. 2473-2474.

86. Turnbull, J.J., et al., *The C-4 stereochemistry of leucocyanidin substrates for anthocyanidin synthase affects product selectivity.* Bioorg Med Chem Lett, 2003. **13**(21): p. 3853-7.

87. Sato, M., et al., *Flavanonol glucoside and proanthocyanidins: oviposition stimulants for the cerambycid beetle, Monochamus alternatus.* Journal of Pesticide Science, 1999. **24**(2): p. 123-129.

88. Pei, Y., et al., *Studies on the structure of a new flavanonol glucoside of the root-sprouts of Agrimonia pilosa Ledeb.* Yao xue xue bao= Acta pharmaceutica Sinica, 1990. **25**(4): p. 267-270.

89. Nonaka, G.-I., et al., *Tannins and related compounds. LII. Studies on the constituents of the leaves of Thujopsis dolabrata Sieb. et Zucc.* Chemical and pharmaceutical bulletin, 1987. **35**(3): p. 1105-1108.

90. Creasy, L. and T. Swain, *Structure of condensed tannins.* Nature, 1965. **208**(5006): p. 151-153.

91. Wilmouth, R.C., et al., *Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana.* Structure (Camb), 2002. **10**(1): p. 93-10.

92. Routaboul, J.-M., et al., *Flavonoid diversity and biosynthesis in seed of Arabidopsis thaliana.* Planta, 2006. **224**(1): p. 96-107.

93. Liu, C., et al., *A role for leucoanthocyanidin reductase in the extension of proanthocyanidins.* Nature Plants, 2016. **2**(12): p. 1-7.

94. Aron, P.M. and J.A. Kennedy, *Flavan-3-ols: Nature, occurrence and biological activity.* Molecular nutrition & food research, 2008. **52**(1): p. 79-104.

95. Folts, J.D., *Potential health benefits from the flavonoids in grape products on vascular disease.* Flavonoids in cell function, 2002: p. 95-111.

96. Havsteen, B.H., *The biochemistry and medical significance of the flavonoids.* Pharmacology & therapeutics, 2002. **96**(2-3): p. 67-202.

97. Bagchi, D., et al., *Molecular mechanisms of cardioprotection by a novel grape seed proanthocyanidin extract.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2003. **523**: p. 87-97.

98. Vitseva, O., et al., *Grape seed and skin extracts inhibit platelet function and release of reactive oxygen intermediates.* Journal of cardiovascular pharmacology, 2005. **46**(4): p. 445-451.

99. Grassi, D., et al., *Cocoa and cardiovascular health. The sweet heart protection.* AGRO FOOD INDUSTRY HI TECH, 2006. **17**(1): p. XIII.

100. Fernández, K. and J. Labra, *Simulated digestion of proanthocyanidins in grape skin and seed extracts and the effects of digestion on the angiotensin I-converting enzyme (ACE) inhibitory activity.* Food chemistry, 2013. **139**(1-4): p. 196-202.

101. Ye, X., et al., *The cytotoxic effects of a novel IH636 grape seed proanthocyanidin extract on cultured human cancer cells*, in *Stress adaptation, prophylaxis and treatment*. 1999, Springer. p. 99-108.

102. Neto, C.C., et al., *MALDI-TOF MS characterization of proanthocyanidins from cranberry fruit (Vaccinium macrocarpon) that inhibit tumor cell growth and matrix metalloproteinase expression in vitro.* Journal of the Science of Food and Agriculture, 2006. **86**(1): p. 18-25.

103. Suganya, M., et al., *Antitumor effect of proanthocyanidin induced apoptosis in human colorectal cancer (HT-29) cells and its molecular docking studies.* BMC chemistry, 2019. **13**(1): p. 1-14.

104. Roy, A.M., et al., *Grape seed proanthocyanidins induce apoptosis through p53, Bax, and caspase 3 pathways.* Neoplasia, 2005. **7**(1): p. 24-36.

105. Lee, Y.A., et al., *Inhibitory activities of proanthocyanidins from persimmon against oxidative stress and digestive enzymes related to diabetes.* Journal of Nutritional Science and Vitaminology, 2007. **53**(3): p. 287-292.

106. Li, B.Y., et al., *Back-regulation of six oxidative stress proteins with grape seed proanthocyanidin extracts in rat diabetic nephropathy.* Journal of cellular biochemistry, 2008. **104**(2): p. 668-679.

107. Lee, Y.A., et al., *Ameliorative effects of proanthocyanidin on oxidative stress and inflammation in streptozotocin-induced diabetic rats.* Journal of agricultural and food chemistry, 2007. **55**(23): p. 9395-9400.

108. Hu, Y., et al., *Tartary buckwheat flavonoids ameliorate high fructose-induced insulin resistance and oxidative stress associated with the insulin signaling and Nrf2/HO-1 pathways in mice.* Food & function, 2017. **8**(8): p. 2803-2816.

109. Ubillas, R., et al., *SP-303, an antiviral oligomeric proanthocyanidin from the latex of Croton lechleri (Sangre de Drago).* Phytomedicine, 1994. **1**(2): p. 77-106.

110. Gallina, L., et al., *Inhibition of viral RNA synthesis in canine distemper virus infection by proanthocyanidin A2.* Antiviral research, 2011. **92**(3): p. 447-452.

111. Carneiro, B.M., et al., *The green tea molecule EGCG inhibits Zika virus entry.* Virology, 2016. **496**: p. 215-218.

112. Zhao, M., et al., *A proprietary topical preparation containing EGCG-stearate and glycerin with inhibitory effects on herpes simplex virus: Case study.* Inflammation & Allergy-Drug Targets (Formerly Current Drug Targets-Inflammation & Allergy)(Discontinued), 2012. **11**(5): p. 364-368.

113. Costa, S.S., et al., *Flavonoids in the therapy and prophylaxis of flu: a patent review.* Expert opinion on therapeutic patents, 2012. **22**(10): p. 1111-1121.

114. Rrapo, E., et al., *Green Tea-EGCG reduces GFAP associated neuronal loss in HIV-1 Tat transgenic mice.* American journal of translational research, 2009. **1**(1): p. 72.

115. Chen, C.N., et al., *Inhibition of SARS-CoV 3C-like protease activity by theaflavin-3,3 '-digallate (TF3).* Evidence-Based Complementary and Alternative Medicine, 2005. **2**(2): p. 209-215.

116. Singh, S.P. and B.K. Konwar, *Molecular docking, DFT and ADME-toxicity studies on analogues of epigallocatechin gallate as SARS coronavirus 3CL protease inhibitors.* Journal of Bioinformatics and Intelligent Control, 2013. **2**(1): p. 1-10.

CHAPTER 2

Molecular Cloning and Functional Characterization of a Dihydroflavonol 4-Reductase from *Vitis bellula*

Abstract: *Vitis bellula* is a new grape crop in southern China. Berries of this species are rich in antioxidative anthocyanins and proanthocyanidins. This study reports cloning and functional characterization of a cDNA encoding a *V. bellula* dihydroflavonol reductase (VbDFR) involved in the biosynthesis of anthocyanins and proanthocyanidins. A cDNA including 1014 bp was cloned from young leaves and its open reading frame (ORF) was deduced encoding 337 amino acids, highly similar to *V. vinifera* DFR (VvDFR). Green florescence protein fusion and confocal microscopy analysis determined the cytosolic localization of VbDFR in plant cells. A soluble recombinant VbDFR was induced and purified from *E. coli* for enzyme assay. In the presence of NADPH, the recombinant enzyme catalyzed dihydrokaempferol (DHK) and dihydroquercetin (DHQ) to their corresponding leucoanthocyanidins. The *VbDFR* cDNA was introduced into tobacco plants via *Agrobacterium*-mediated transformation. The overexpression of *VbDFR* increased anthocyanin production in flowers. Anthocyanin hydrolysis and chromatographic analysis revealed that transgenic flowers produced pelargonidin and delphinidin, which were not detected in control flowers. These data demonstrated that the overexpression of VbDFR produced new tobacco anthocyanidins. In summary, all data demonstrate that *VbDFR* is a useful gene to provide three types of substrates for metabolic engineering of anthocyanins and proanthocyanidins in grape crops and other crops.

This work was published in Molecules, 2018.

Yue Zhu, Qingzhong Peng, Kegang Li, De-Yu Xie. 2018, Molecular Cloning and Functional Characterization of a Dihydroflavonol 4-Reductase from *Vitis bellula*. Molecules. 23(4):861

1. Introduction

Dihydroflavonol 4-reductase (DFR) is a key late enzyme in the plant flavonoid pathway toward both anthocyanins and proanthocyanidins (Figure 1). It catalyzes the key step from dihydroflavonols, such as dihydrokaempferol (DHK), dihydroquercetin (DHQ), and dihydromyricetin (DHM), to leucoanthocyanidins, such as leucopelargonidin, leucocyanidin, and leucodelphinidin [1]. DFR genes have been cloned from multiple plants and its mutation has been demonstrated to cause the loss of anthocyanins and proanthocyanidins in plants [2–5]. To date, *DFR* is an economically important plant gene, given that anthocyanins and proanthocyanidins are two groups of antioxidants relating to high nutritional values of crop, food, and beverage (such as wine and green tea) products [6–9]. Particularly, anthocyanins are one of the richest plant natural pigments with significant economic values in the horticulture industry [6,10–13]. Since the first success of *Petunia* flower color engineering was achieved using the first *DFR* cDNA cloned from maize [13], multiple homologs have been cloned to engineer anthocyanins [1,14,15] and proanthocyanidins [16,17]. DFR also plays a key role in transcription factor-based anthocyanin engineering. A common metabolic phenotype is that the activation or enhancement of *DFR* expression is necessary in MYB transcription factor-based anthocyanin engineering. For example, the regulatory function of the Production of Anthocyanin Pigmentation 1 (PAP1, a MYB75) depends upon the expression of *DFR* [18–24].

Vitis bellula (namely Bellula here) is a relative of *V. vinifera* (in Vitaceae), which is the main grape crop for both wine and non-alcoholic beverage products [25–27]. Although the cropping of *V. bellula* is limited to southern China [28], this species has become an emerging new grape crop

to develop potential wine and beverage products due to its high nutritional values. Our recent studies have revealed that its berry is rich in flavan-3-ols and proanthocyanidins [28,29] (Figure 1). We have demonstrated that the formation of proanthocyanidins and flavan-3-ols in *V. bellula* is via two pathways, the anthocyanidin reductase (ANR) and leucoanthocyanidin reductase (LAR) pathways [28,30]. Metabolic profiling revealed that Bellula's berries produce two types of configurations of flavan-3-ols, 2R, 3S-2, 3-*trans*-flavan-3-ols such as (+)-catechin and (+)-gallocatechin and 2R, 3R-2, 3-*cis*-flavan-3-ols such as (−)-epicatechin (and (−)-epigallocatechin. Both butanol:HCl cleavage and chromatograph analyses have revealed that the building units of proanthocyanidins include three types of flavan-3-ols, which are characterized by one –OH, two –OH, and three –OH groups in the B ring, such as (−)-epifazelechin, (−)-epicatechin, and (−)-epigallocatechin (Figure 1). These three types of structures are correspondingly derived from leucopelargonidin, leucocyanidin, and leucodelphinidin via either the ANR or LAR pathway (Figure 1).

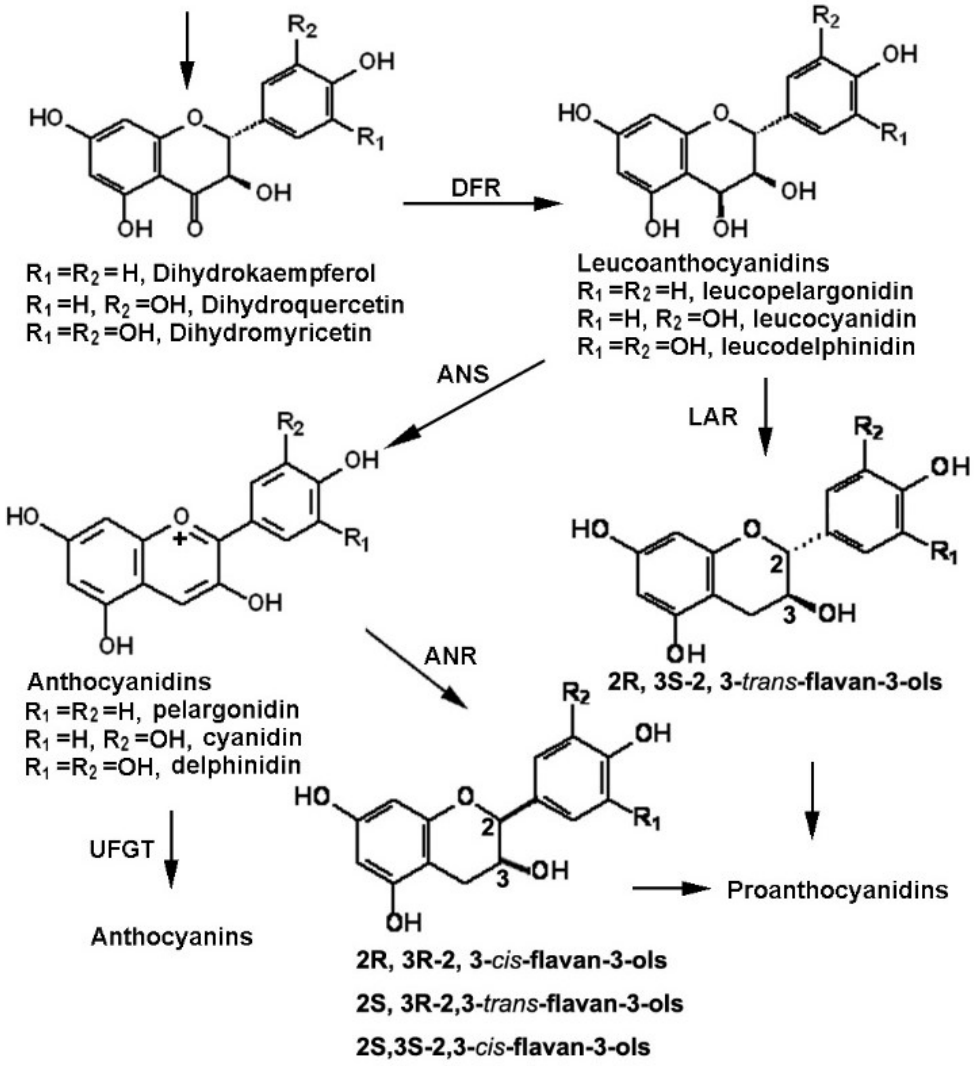


Figure 1. Late pathway steps of flavonoids from dihydroflavonol to anthocyanins and proanthocyanidins. DFR: dihydroflavonol reductase, ANS: anthocyanidin synthase, ANR: anthocyanidin reductase, LAR: leucoanthocyanidin reductase, and UFGT: UDP-glucose flavonoid 3-*O*-glucosyltransferase. UDP: uridine diphosphate.

To date, cDNA(s) encoding DFR has not be cloned from *V. bellula*. In this article, we report the cloning and characterization of *V. bellula DFR* (*VbDFR*). A *DFR* cDNA homolog, namely *VbDFR*, was cloned from young leaves. Sequence alignment using deduced amino acids showed that only three amino acids were different between VbDFR and *V. vinifera* DFR (VvDFR), demonstrating their high identity. In vitro enzymatic assay, transgenic analysis, and metabolic

profiling of transgenic plants showed that VbDFR catalyzes the step from dihydroflavonols to leucoanthocyanidins. Particularly, the overexpression of *VbDFR* in tobacco flowers led to the formation of pelargonidin, cyanidin, and delphinidin in plants. These results indicate that *VvDFR* is a useful gene to metabolically engineer three types of anthocyanidins and their corresponding anthocyanins.

2. Results

2.1. Cloning of DFR Gene from Leaf of V. bellula

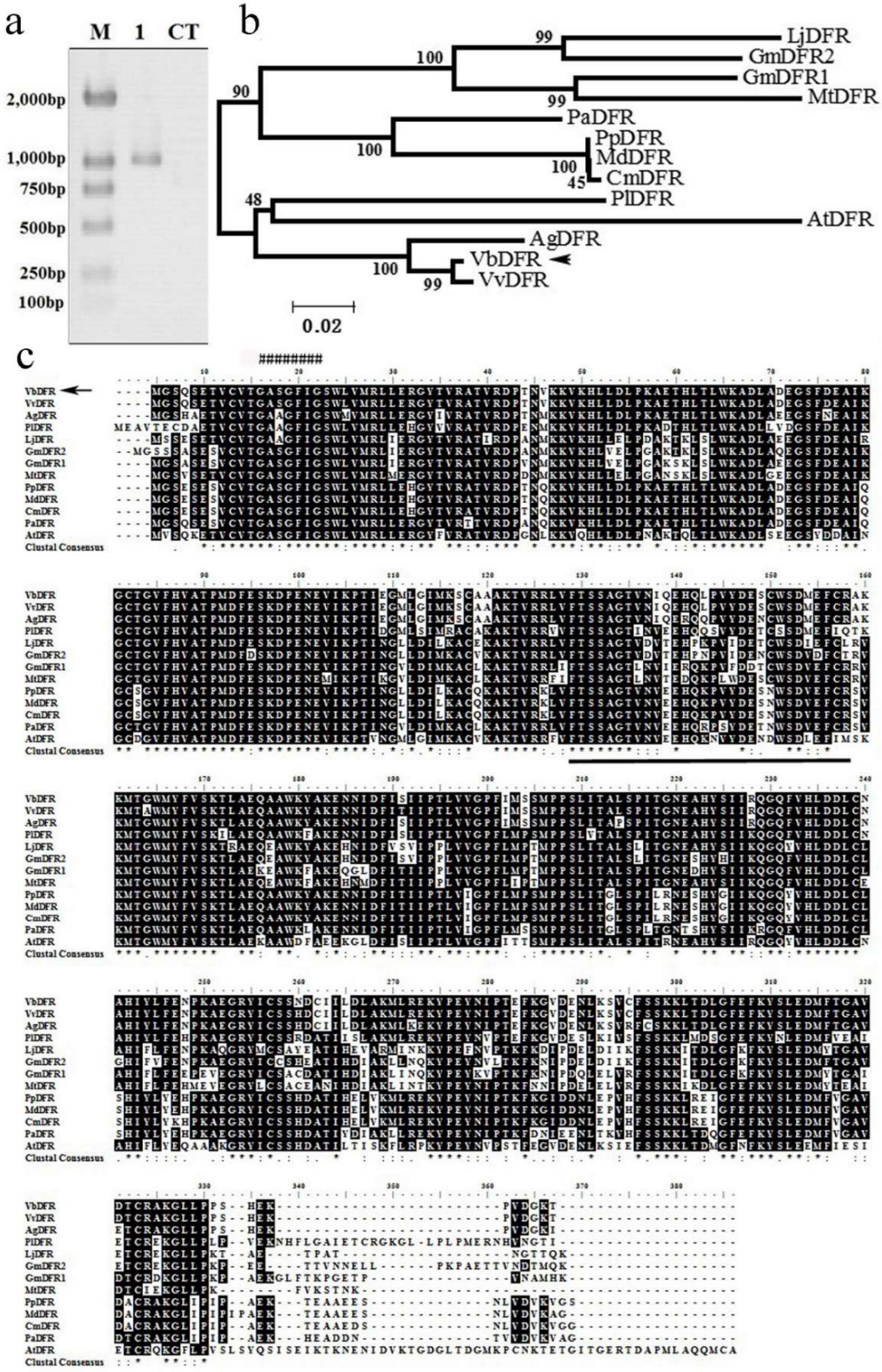
Based on *VvDFR* sequence (GenBank: CAA53578.1), we designed a pair of primers, which contained the start and stop codon nucleotides. RT-PCR was carried out to amplify a *DFR* cDNA from a leaf cDNA library of *V. bellula*. We named it *VbDFR*. This amplification produced an approximately 1 kb cDNA fragment (Figure 2a). Sequencing confirmed that this fragment was composed of 1014 nucleotides including the full length of open reading frame (ORF) from the start codon to the stop codon, which was deduced to translate 337 amino acids. An unrooted phylogenetic tree was built using amino acid sequences of 13 DFR homologs. The resulting tree clustered them into two main clades (Figure 2b). VbDFR, VvDFR, AtDFR, AgDFR, and PlDFR were clustered in one clade, which VbDFR and VvDFR are directly clustered together, indicating their close similarity. Further amino acid sequence alignment using 13 DFR homologs revealed three different amino acids between VbDFR and *V. vinifera* DFR (VvDFR). In addition, this alignment showed that the deduced VbDFR amino acid sequence included the conserved glycine-rich Rossmann NADPH/NADH-binding domain and a substrate specificity domain (Figure 2c).

Figure 2. Cloning, amino acid sequence alignment, and an unrooted phylogenetic tree of DFR obtained from deduced amino acid sequences of 13 DFR homologs.

a, a *DFR* cDNA fragment was amplified from *V. bellula* leaf tissue by RT-PCR. M: DNA marker, 1: *DFR* cDNA fragment, CT: negative control.

b, an unrooted phylogenetic tree was built from amino acid sequences of 13 DFR homologs.

c, amino acid sequence alignment were develop from 13 DFR homologs; "*": the same amino acid in all sequences; ":": conserved amino acid residues; ".": half conserved amino acid residues; "######": potential NADPH/NADH binding domain; amino acids underlined form a potential substrate specificity domain of DFR. AtDFR: *Arabidopsis thaliana* DFR; CmDFR: *Crataegus monogyna* DFR; GmDFR1: *Glycine max* DFR1; GmDFR2: *Glycine max* DFR2; LjDFR: *Lotus japonicas* DFR; MdDFR: *Malus domestica* DFR; MtDFR: *Medicago truncatula* DFR; NgDFR: *Nekemias grossedentata* DFR; PaDFR: *Prunus avium* DFR; PpDFR: *Pyrus pyrifolia* DFR; PlDFR: *Paeonia lactiflora* DFR; VvDFR: *Vitis vinifera* DFR.


a
M
1
CT
2,000bp
1,000bp
750bp
500bp
250bp
100bp
b
LjDFR
GmDFR2
GmDFR1
MtDFR
PaDFR
PpDFR
MdDFR
CmDFR
PlDFR
AtDFR
AgDFR
VbDFR
VvDFR
0.02
c
########
VbDFR
VvDFR
AgDFR
PlDFR
LjDFR
GmDFR2
GmDFR1
MtDFR
PpDFR
MdDFR
CmDFR
PaDFR
AtDFR
Clustal Consensus

2.2. Recombinant VbDFR Expression and Enzymatic Assay

The ORF was cloned into a pET 28a (+) vector to obtain a recombinant pET 28a (+)-VbDFR vector. This construct was transformed into BL21 (DE3) plysS strain to induce recombinant VbDFR protein. The pET 28a (+) empty vector was used as control. After induced with 1.0 mM IPTG, extraction of *E. coli* revealed that recombinant VbDFR protein existed on both inclusion body and buffer (Figure 3a). However, the recombinant protein was not produced from the empty vector control (Figure 3a). Next, the recombinant protein was purified by using a Ni-NTA column. SDS-PAGE analysis showed one major recombinant VbDFR band from this purification (Figure 3b). The purified recombinant VbDFR was stored at −20 °C for further enzymatic assays. (−)-Taxifolin and (−)-dihydrokaempferol (DHK) are two DFR substrates (Figure 1). We used these two substrates to examine the enzymatic activity of the recombinant VbDFR. The enzymatic reaction was carried out in 500 µL volume that included 15 µg substrates, 50 µg protein, 50 mM citrate buffer at pH 6, and 20 µg NADPH. After reactions were extracted with ethyl acetate, products were dissolved in methanol for HPLC analysis. HPLC profiles recorded at 280 nm showed that one new peak was observed from reactions using taxifolin (Figure 3c-1) and DHK (Figure 3d-1), respectively. However, these new peaks were neither observed from control reactions using denatured protein nor existed in standard samples (Figure 3c-2,d-2). These results demonstrated that the recombinant VbDFR used these two metabolites as substrates. Although leucoanthocyanidin standards were not commercially available, based on our previous elucidation method for DFR's products [1], these two new peaks from taxifolin and DHK were annotated to be leucocyanidin and leucopelargonidin, respectively.

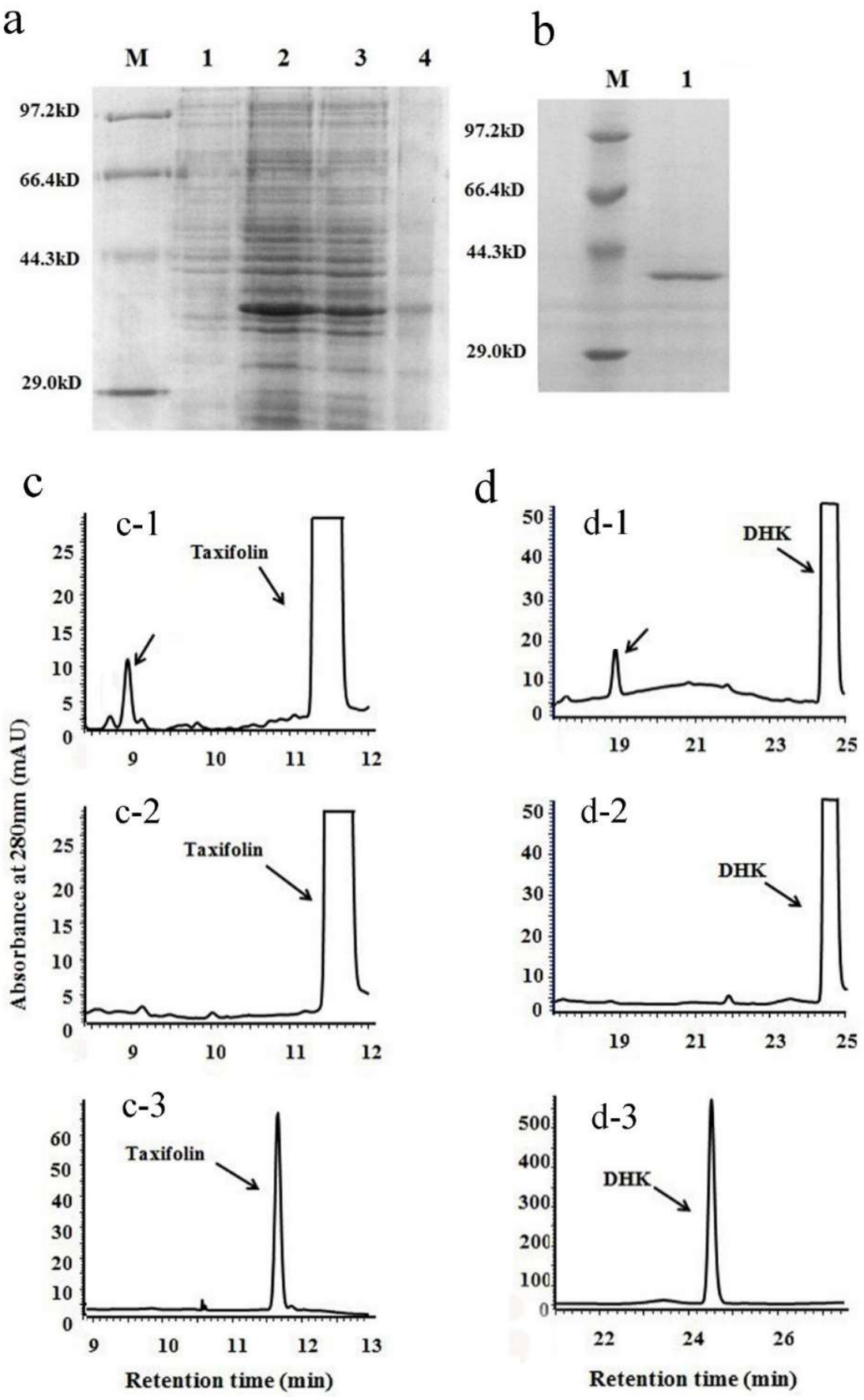
Figure 3. Recombinant protein expression and enzymatic analysis of VbDFR.

a, a SDS-PAGE image shows recombinant VbDFR protein induced in *E. coli* BL21 (DE3) plysS strain. Lane 1: 20 μg crude protein extracts from BL21 (DE3) plysS/pET28a (+) vector control; lane 2: 20 μg crude protein extracts from BL21 (DE3) plysS/pET28a (+)-VbDFR. Lane 3: insoluble crude protein extracts from BL21 (DE3) plysS/pET28a (+)-VbDFR. Lane 4: soluble crude protein extracts from BL21 (DE3) plysS/pET28a (+)-VbDFR. M: protein molecular weight marker.

b, an image shows recombinant VbDFR purified.

c, HPLC profiles show one product formed from the incubation of taxifolin and recombinant VbDFR (c-1) but not denatured VbDFR (c-2); c-3, taxifolin standard. (d), HPLC profiles show one product formed from the incubation of dihydrokaempferol (DHK) and recombinant VbDFR (d-1) but not denatured recombinant VbDFR (d-2); d-3 DHK standard.

2.3. Subcellular Localization of VbDFR


a
M
1
2
3
4
97.2kD
66.4kD
44.3kD
29.0kD
b
M
1
97.2kD
66.4kD
44.3kD
29.0kD
c
c-1
Taxifolin
c-2
Taxifolin
c-3
Taxifolin
Absorbance at 280nm (mAU)
Retention time (min)
d
d-1
DHK
d-2
DHK
d-3
DHK
Retention time (min)

After the stop codon was removed, *VbDFR* was fused to the N-terminus of *GFP* to obtain a new plasmid, namely pBI121-VbDFR-GFP (Figure 4a). In addition, one pBI121-GFP plasmid was used as control. Two plasmids were introduced to onion epidermal cells via a gene gun transformation. Examination of epidermal cells under confocal microscope revealed that green fluorescence signal was mainly located in the entire cytosol of epidermal cells transformed with the pBI121-VbDFR-GFP plasmid (Figure 4a). The same result was observed in epidermal cells transformed with the pBI121-GFP plasmid (Figure 4b). These results showed the cytosolic localization of VbDFR.

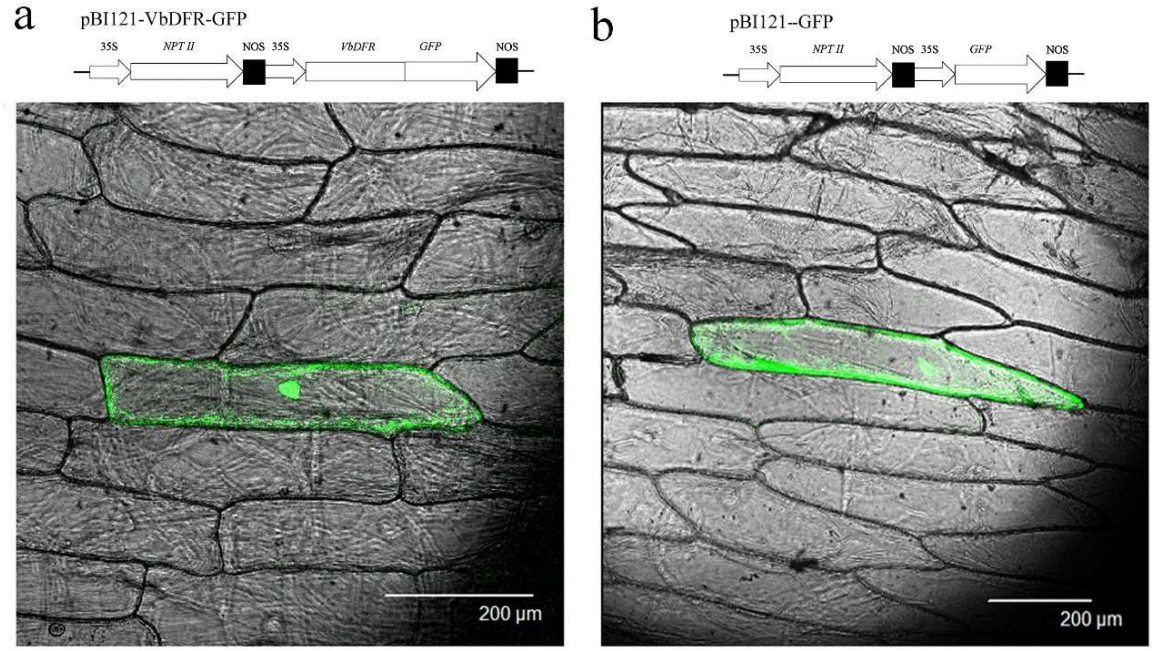


Figure 4. Transient expression using GFP fusion and confocal microscope images.

a, a cassette showing *VbDFR* and *GFP* fusion and a confocal microscopy image showing subcellular localization of VbDFR-GFP;

b, a cassette showing GFP as positive control and a confocal microscopy image showing subcellular localization of GFP.

2.4. Overexpression of VbDFR Tobacco Enhances Production of Anthocyanins

VbDFR was overexpressed in tobacco plants to examine its function *in planta*. The ORF of *VbDFR* was cloned into the pBI121 binary vector under the control of a CaMV35S promoter (Figure 5a) and transformed to tobacco plants as reported previously [1]. More than 10 kanamycin-resistant plantlets were regenerated from tissue culture, planted in pot soil, and maintained in the greenhouse to grow up for flowers. In addition, pBI121 vector transgenic and wild-type plants were grown as control to compare plant growth and flower coloration. PCR analysis showed that the transgene was integrated into the transgenic plant genome and RT-PCR analysis demonstrated the ectopic expression of the *VbDFR* transgene (Figure 5b). Obvious flower color changes were observed in *VbDFR* transgenic plants. The red pigmentation of *VbDFR* transgenic petals was deeper than that of control petals (Figure 5a). Further quantification using methanol: HCl extraction was carried out on a UV spectrophotometer. Absorbance values recorded at 530 nm showed that the contents of anthocyanins are significantly higher in *VbDFR* transgenic flowers than in control ones (Figure 5c). These demonstrated that the overexpression of *VbDFR* increased anthocyanin biosynthesis in transgenic flowers.

To further characterize effects of the *VbDFR* transgene on anthocyanidin profiles in transgenic flowers, anthocyanin extracts were hydrolyzed using butanol: HCl boiling. The hydrolyzed products were subject to TLC profiling and HPLC analysis. The resulting TLC profiles showed that in addition to a strong spot with the same Rf value as cyanidin, red colors were observed in spots with the same Rf values as pelargonidin and delphinidin (Figure 5d). To further characterize anthocyanidins corresponding to these spots, HPLC analysis were performed to

compare them with three authentic standards, cyanidin, pelargonidin, and delphinidin, which have the maximum absorbance value at 524 nm, 515 nm, and 530 nm, respectively. Based on the retention time and UV spectrum properties, cyanidin, pelargonidin, and delphinidin were detected in *VbDFR* transgenic flower extracts, while only cyanidin was detected from wild-type control flowers (Figure 5e). Moreover, the peak area of cyanidin from the *VbDFR* transgenic flowers were significantly bigger than that from wild type flowers. We noticed that on the TLC plate, the light pink spot at the same retention time as pelargonidin was obvious, however HPLC only showed a relatively small shouldered peak (Figure 5e-2). This likely resulted from the overlapping of an unidentified anthocyanidin and pelargonidin on the TLC plate, two of which were separated by HPLC. These results further demonstrated that the overexpression of *VbDFR* in tobacco not only increased cyanidin contents, but also produced two new tobacco anthocyanidins.

Figure 5 Increase of anthocyanins in transgenic tobacco flowers overexpressing *VbDFR*.

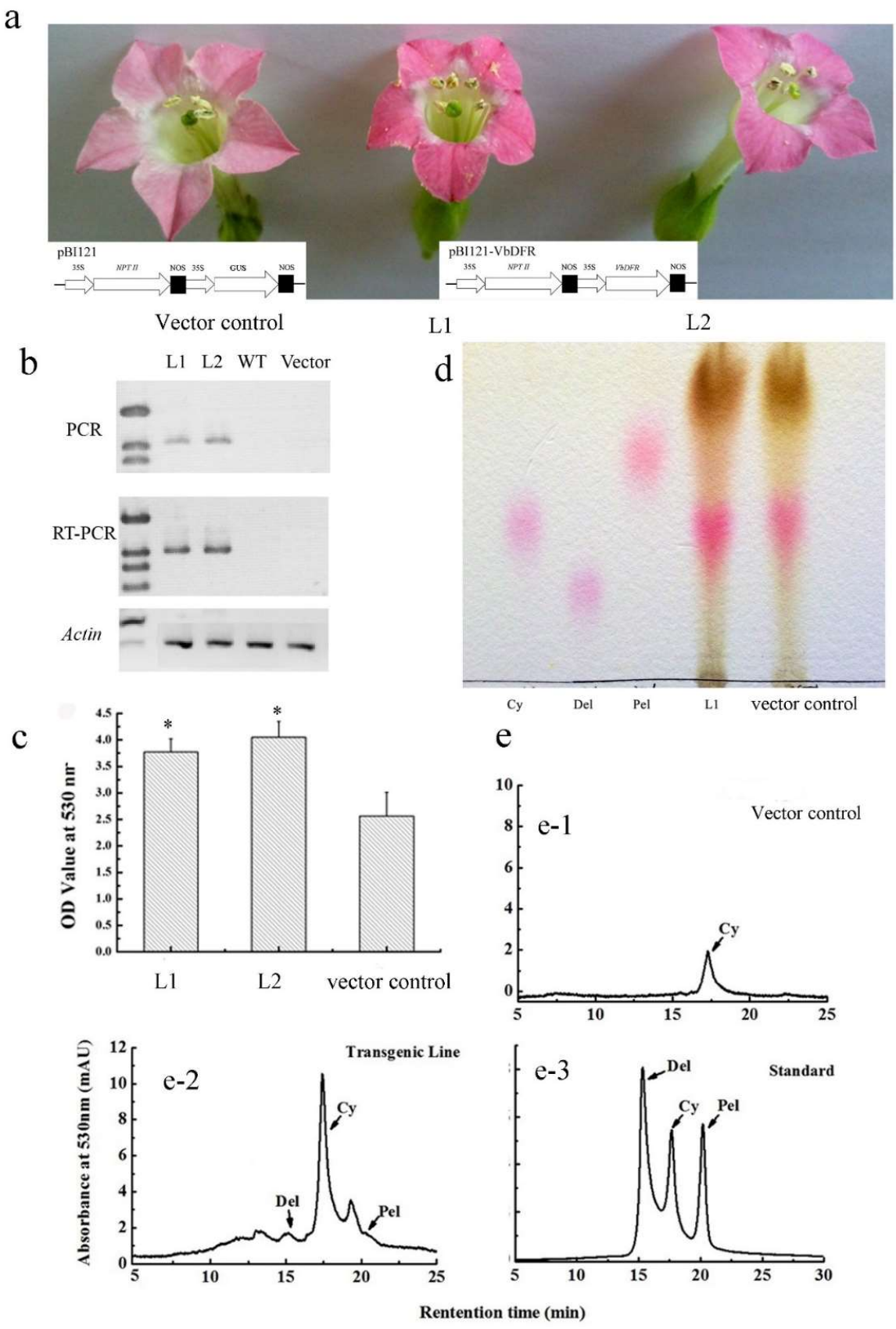
a, cassettes of pBI121 and pBI121-VbDFR vectors show *VbDFR* controlled by 35S promoter for overexpression, and red pigmentation was enhanced in flowers of two *VbDFR* transgenic tobacco lines (L1 and L2) compared with vector control flower.

b, images show fragments amplified by genomic DNA-based PCR and RT-PCR.

c, absorbance values for anthocyanin extracts were significantly higher from flowers of *VbDFR* transgenic tobacco lines (L1 and L2) than from vector control flowers.

d, TLC profiles show anthocyanidins from the butanol:HCl hydrolysis of anthocyanins extracts from transgenic tobacco line (L1) and vector control plants.

e, HPLC profiles show anthocyanidins from butanol:HCl hydrolysis of anthocyanins extracted from transgenic tobacco line (e-1) and vector control (e-2) plants. e-3: authorized standards, Del: delphinidin, Cy: cyanidin, Pel: pelargonidin.


a
pBI121
35S NPT II NOS 35S GUS NOS
pBI121-VbDFR
35S NPT II NOS 35S VbDFR NOS
Vector control
L1
L2
b
L1 L2 WT Vector
PCR
RT-PCR
Actin
d
Cy
Del
Pel
L1
vector control
c
OD Value at 530 nm
L1
L2
vector control
e
e-1
Vector control
Cy
e-2
Transgenic Line
Cy
Del
Pel
e-3
Standard
Del
Cy
Pel
Absorbance at 530nm (mAU)
Rentention time (min)

3. Discussion

The study of more *DFR* homologs from different *Vitis* species is necessary in order to appropriately understand this key gene's function in the biosynthesis of plant flavonoids for grape agriculture. Since the *VvDFR* cDNA sequence was cloned from *V. vinifera* two decades ago [31], to our knowledge, no *DFR* homologs have been characterized from other *Vitis* species. To date, *VvDFR* is the only *Vitis* DFR homolog that has been appropriately characterized to understand the step from dihydroflavonols to leucoanthocyanidins in the biosynthetic pathways of anthocyanins and proanthocyanidins in grape berries [32–36]. Different transcriptional studies have showed that the expression of *VvDFR* is regulated by different factors. High temperature during night time was found to inhibit the expression of *VvDFR* [34]. Light was found to induce the activation of the DFR promoter [32]. The synthetic auxin, 2,4-dichlorophynoxyacetic acid, was found to inhibit the expression of *VvDFR* in anthocyanin producing cells, which abscisic acid was shown to upregulate the expression of *VvDFR* [37]. Other factors such as calcium and sucrose were also demonstrated to induce the expression of *VvDFR* [32]. A crystal structural study has characterized the NADPH-binding Rossmann domain at the N-terminus and substrate-binding specificity in the variable C-terminus [36]. The amino acid region from 131–15 has been characterized to be the substrate binding site. In particular, the Asn or Asp residue at position 133 has been characterized to associate with substrate recognition although the variant N133D may not be solely specific to recognize the three hydroxylation patterns in the B-ring of dihydroflavonols (Figure 1) [36]. In summary, on the one hand, these data provide useful information for understanding the regulation of biosynthesis of anthocyanins and proanthocyanidins in grape berries in the field. On the other hand, the function of *VvDFR in*

planta, such as in transgenic plants, still remains for further investigation. The present study characterizes the function of *VbDFR,* a new homolog, to understand the biosynthesis of plant flavonoids in the *Vitis* species. It is interesting that VbDFR and VvDFR only have three different amino acids, revealing the high conservation of sequences. Accordingly, as we expected, an in vitro assay showed that the recombinant VbDFR used DHK and DHQ as substrates to produce their corresponding leucopelargonidin and leucocyanidin (Figure 3d,e). Although we couldn't find DHM for a substrate assay, *VbDFR* transgenic but not control tobacco flowers produced delphinidin, providing evidence that VbDFR uses DHM as substrate. Our confocal analysis showed the subcellular localization of VbDFR in the cytosol (Figure 4), revealing the in vivo catalytic localization in cells. To date, the function of VvDFR has not been characterized in transgenic plants. In our study, we overexpressed *VbDFR* in tobacco plants to understand this *Vitis* DFR's functions *in planta*. The ectopic expression of *VbDFR* increased anthocyanin production (Figure 5a,d). More importantly, transgenic flowers produced pelargonidin and delphinidin (Figure 5c,e), which were not produced in control flowers, demonstrating that the VbDFR produced new tobacco anthocyanidins. Given that anthocyanidins characterized from *V. vinifera* mainly include cyanidin and delphinidin and their derivatives such as peonidin, petunidin, and malvidin [35], while pelargonidin and pelargonins are uncommonly identified from this crop, our transgenics data provide new information to understand the structural diversity of anthocyanidins in Bellula grape plants and suggest that studying more grape species can enhance the discovery of new anthocyanins for wine and non-alcoholic beverage industries.

4. Materials and Methods

4.1. Cloning of VbDFR from Young Leaves of V. bellula

Fifty milligrams of young leaf sample of *V. bellula* were collected and ground into fine powder in liquid nitrogen. Total RNA was extracted using a Plant RNA extract Kit (Sangon, Shanghai, China) and then digested with 1.0 μg DNase to remove genomic DNA contamination. The resulting DNA-free RNA sample was used as a template to synthesize the first strand cDNA with MMLV Reverse transcriptase (Takara, Japan) and oligo $(dT)_{12}$ primer. All steps followed those manufacturers' protocols.

Based on *V. vinifera DFR* sequences, one pair of primers containing the start and stop codons was designed to amplify its homolog from *V. bellula*, namely *VbDFR*. The forward primer was 5′-ATGGGTTCACAAAGTGAAAC-3′, and the reverse primer was 5′-CTAGGTCTTGCCATCTACAG-3′. Two μL of the 1st strand cDNA was used as template for polymerase chain reaction (PCR) to amplify *VbDFR* cDNA. Ex-Taq polymerase (Takara, Japan) was used for PCR following this manufacturer's protocol. The thermal cycle was composed of 94 °C 5 min, 30 cycles of 94 °C 45 s, 57 °C 45 s, and 72 °C 45 s. The final extension step was 10 min at 72 °C. The amplified cDNA was separated on an agarose gel by electrophoresis and visualized using EB dye. The band was excised from gel, from which the cDNA fragment was purified using a Qiagen DNA purification kit by following the manufacturer's protocol. Next, the isolated cDNA was cloned into a T-easy vector (Promega, Madison, WI, USA) to obtain a T-VbDFR plasmid by following the manufacturer's protocol. The new plasmid was introduced into competent cells of *E. coli* DH5α strain. One single colony was selected for suspension culture.

The resulting *E. coli* culture was used to isolate the plasmid with QIAprep Spin Miniprep Kit (Qiagin, Hilden, Germany). The resulting plasmid was used for sequencing.

4.2. Expression of Recombinant VbDFR in E. coli and Purification

A pair of primers was designed to clone *VbDFR* to pET28a (+) (Novagen, Madison, WI, USA), a protein expression vector. The forward primer was 5′-CGGAATTCATGGGTTCACAAAGTG AAACCG-3′, in which the underlined region is an introduced *EcoR* I restriction site. The reverse primer was 5′-CCGCTCGAGTTAGTGATGGTGATGATGGTGGGTCTTGCCATCCTACAGG-3′, in which the underlined CTCGAG region is an introduced *Xhol* restriction site and the other is an introduced His-tag encoding site consisting of 18 nucleotides. Approximately 1.0 ng of the T-VbDFR plasmid was used as template for PCR, which was carried out using Ex-taq DNA polymerase (Takara, Japan) as described above. The resulting PCR product was purified as described above and then digested with *Xhol* and *EcoR* I (Takara, Japan) by following the manufacturer's protocol. The digested products were purified via gel purification as described above and then ligated into the pET28a (+) vector, which was also digested by *Xhol* and *EcoR* I. Ligated products were introduced into competent cells of BL21 (DE3) plysS *E. coli* strain, which were streaked on agar-solidified LB medium plate containing 200 mg/L ampicillin. Single positive colonies were selected via PCR-based screening and then used for suspension culture to isolate the plasmid. The resulting plasmid was used for sequencing to obtain a sequence without mutations. The empty vector was also introduced to *E. coli* as a control.

A positive colony was identified and then inoculated to 10 mL autoclaved liquid LB medium supplemented with 200 mg/L ampicillin in a 50 mL tube, which was placed on a rotary shaker at a speed of 250 rpm at 37 °C. After an overnight culture, one mL suspension was inoculated to 100 mL fresh liquid LB containing 200 mg/L ampicillin in a 500 mL Erlenmeyer flask, which was placed on the same shaker at speed of 120 rpm at 37 °C. When the optical density (OD) of suspension culture measured at 600 nm reached 0.7, the incubation temperature was reduced to 30 °C and isopropyl-1-thio-β-d-galactopyranoside (IPTG) was added to cell suspension to a final concentration of 10 mM. Next, cell suspension was continuously cultured 4 h to reach approximately OD_{600} 1.1. Cells were harvested by centrifugation at 6000 rpm 5 min at 4 °C. The supernatant was disposed of to a waste container and the remaining pellet was immediately used for enzyme extraction and purification described below.

Cell pellets were thoroughly suspended in 10.0 mL extraction buffer consisting of pH 8.0 20 mM Tris-HCl in a 50.0 mL tube. Lysozyme was added into the mixture to a final concentration of 100 μM to treat cells 1 min at room temperature. The mixture was sonicated 30 s on ice. The resulting sticky mixture was configured at 10,000× *g* 10 min. The upper supernatant containing proteins was used for protein purification. The purification of recombinant VbDFR was carried out using a Ni-NTA agarose column. One mL agarose resin (Qiagen, Hilden, Germany) was loaded into a 1× 10 cm syringe (diameter × height). After the resin had fully sunk to the bottom, it was washed with 10 volumes of extraction buffer (20 mM pH 8.0 Tris-HCl, 10 mM imidazole). Ten mL supernatant cell lysate was loaded onto the top of the resin and flowed through by gravity. The column was then washed using extraction buffer until no proteins were detected from elution. Two mL elution buffer (pH 8.0 20 mM Tris-HCl, 250 mM imidazole) was

added to the column to elute recombinant VbDFR. The protein concentration was estimated by Bradford protein assay and the quality of protein was examined on a 10% SDS-PAGE.

4.3. Enzymatic Assay

Enzymatic assay was carried out to examine the catalytic activity of the recombinant VbDFR in a 500 μL reaction volume in a 1.5 mL tube. Two commercially available metabolites, taxifolin (dihydroquercetin, DHQ) and (−)-dihydrokaempferol (DHK), were used as substrates. NADPH was used as co-enzyme. Each reaction was composed of 50 mM pH 6 citrate buffer, 20 μg NADPH, 15 μg substrates, and 50 μg purified recombinant VbDFR. The reaction time and temperature were 30 min and 45 °C. The reaction was stopped by addition of 1.0 mL ethyl acetate (EA) and vigorously vortexed. After centrifugation at 10,000 rpm 2 min, the upper EA phase was transferred into a new tube and evaporated in a speedy vacuum at room temperature. The remaining residue was dissolved in 50 μL methanol for HPLC analysis.

HPLC analysis of VbDFR products was performed on a Shimadzu LC-20AT instrument equipped with an SPD-M20A photodiode array detector (Shimadzu, Japan). Metabolites were separated with a diamonsil C18 reversed phase column and detected at 280 nm. The elution solvent system consisted of 1% phosphate (solvent A) and methanol (solvent B). For taxifolin assay, a linear gradient program was developed to elute metabolites, which was composed of three gradient ratios of B:A, 0–8 min: solvent B from 15% to 60%, 9–15 min: solvent B from 60–15%, and 16–25 min: 15% solvent B at a flow rate 1.5 mL/min. For DHK assay, a different linear gradient program was developed for metabolite separation, which was composed of three

gradient ratios of B:A, 0–20 min: solvent B from 15–60%, 21–28 min: solvent B from 60–15%, and 29–35 min: solvent B 15% at a flow rate of 1 mL/min.

4.4. GFP Fusion and Transient Expression

The construction of VbDFR-GFP fusion in pBI121 binary vector [1] was carried out with NEBuilder® HiFi DNA Assembly Cloning Kit (NEB, Herts, UK) following the manufacturer's protocol. In detail, two primer pairs were designed for PCR to amplify *VbDFR* and *GFP* cDNA sequences. For *VbDFR* cDNA amplification, the forward primer was 5′-CTATGACCATGATTACGCCAATGGGTTCACAAAGTGAAACCG-3′, in which the underlined region was overlapped with 20 nucleotides in the 5-terminus of Hind III restrict site in pBI121. The reverse primer was 5′-GCTCCTCGCCCTTCGACATGGTCTTGCCATCTACAGG-3′, in which the stop codon of *VbDFR* was excluded and the underlined region was overlapped with 19 nucleotides of the 5-terminus of the GFP codon sequence. For GFP amplification, the forward primer was 5′-CCTGTAGATGGCAAGACCATGTCGAAGGGCGAGGAGC-3′, in which the underlined region is overlapped with 19 nucleotides in the 3-terminus of *VbDFR* cDNA without stop codon. The reverse primer was 5′-CGATCGGGGAAATTCGAGCTCTACTTGTACAGCTCGTC-3′, in which the underlined region was overlapped with 20 nu0cleotides of the 3-terminus of sac I restrict site in pBI121. The resulting two DNA fragments were then ligated to Hind III and sac I sites in digested linear pBI121 using DNA Assembly Cloning Kit. The new binary vector was named as pBI121-VbDFR-GFP, in which VbDFR-GFP was controlled by a 35 promoter (Figure 4a). In addition, the *GFP* gene was introduced into pBI121 vector to create pBI121-GFP construct (Figure 4b) as a positive control.

These two recombinant vectors were introduced into competent cells of *E. coli* DH5α strain, respectively. *E. coli* cells were spread on agar-solidified LB medium plates supplemented with 50 mg/L kanamycin. One positive clone for each was inoculated to 10 mL liquid LB medium supplemented with 50 mg/L kanamycin in 50 mL tube, which was placed on a rotary shaker at speed of 250 rpm at 37 °C overnight. *E. coli* cultures were harvested by centrifugation of tubes at 4000 rpm 10 min. The resulting *E. coli* pellets were used to extract plasmids using a QIAprep Spin Miniprep Kit. Plasmids were coated with gold particles for transient expression as reported previously [28]. Onion epidermis was pre-cultured on agar-solidified Murashige & Skoog (MS) medium plates overnight at 25 °C, then were bombarded with plasmid-coated gold particles. All epidermises bombarded were incubated on MS medium plate overnight at 25 °C, followed by examination under a confocal laser scanning microscope (Leica TCS SP2, Leica, Germany). The light wavelength was set at 488 nm to observe GFP signal. Cell images and GFP signal were photographed to visualize subcellular localization of protein.

4.5. Overexpression of VbDFR in Tobacco

As reported for two *M. trucatula DFRs* previously [1], the ORF of *VbDFR* was cloned to the pBI121 binary vector by replacing *GUS*. This cloning generated a new recombinant binary vector, pBI121-VbDFR, in which *VbDFR* is controlled a 35S promoter. Both pBI121-VbDFR and pBI121 vectors were introduced into *Agrobacterium tumefaciens* strain EHA105 for genetic transformation. MS medium, plant hormones used, and all steps of transformation and selection of transgenic plants followed our protocols published previously [1]. Leaves of sterile seedling were used as explants for genetic transformation. *VbDFR* and vector control transgenic plants were selected using 50 mg/L kanamycin. Transgenic plantlets and control plants were grown in

pot soil and placed side by side in green house to develop flowers. Total RNA was isolated from young leaves of transgenic and control plants as described above. RT-PCR were carried out for genotyping with a pair of gene specific primers. The forward primer was 5′-ATGGGTTCACAAAGTGAAAC-3′, and the reverse primer was 5′-CTAGGTCTTGCCATCTACAG-3′. The thermal cycle was composed of 94 °C 5 min, 30 cycles of 94 °C 45 s, 57 °C 45 s, and 72 °C 45 s. The final extension step was 10 min at 72 °C. In addition, *ACTIN* gene was used as the control. Seeds from T0 plants were collected and germinated on MS medium containing 50 mg/L kanamycin. Resistant T1 seedlings were grown in greenhouse to develop flowers for anthocyanin analysis.

4.6. Anthocyanin Extraction, Hydrolysis, TLC Assay, and HPLC Analysis

One hundred milligrams of fresh petal tissues were homogenized into fine powder in liquid nitrogen and transferred into a 1.5 mL tube. One mL extraction buffer (0.5% HCl in methanol: water, 1:1) was added into the tube. The powder was thoroughly suspended by vortexing. The tube was placed in the dark 30 min at room temperature, followed by centrifugation at 10,000 rpm 10 min. The supernatant was pipetted into a new tube that contained 0.2 mL chloroform. The tube was vortexed 30 s, followed by centrifugation at 10,000 rpm 2 min. The chloroform phase containing non-polar compounds was pipetted to a waste bottle. This step was repeated once. The resulting upper methanol and water phase was pipetted into a new tube for estimation of anthocyanins at 530 nm on a UV spectrophotometer (UV-2550, Shimadzu, Kyoto, Japan). Then, 50 µL anthocyanins extract was mixed with 950 µL butanol:HCl (95:5, *v*/*v*) in a 1.5 mL tube. The mixture was boiled 1 h, cooled to room temperature, and then evaporated in a speed

vacuum. The remaining anthocyanidin residue was suspended in 50 μL of methanol with 0.1% HCl for TLC and HPLC analysis described below.

TLC assay for anthocyanidins was described previously [18]. In brief, 10 μL of methanol extract of anthocyanidins was loaded onto a cellulose F-200 μM plate. Three authentic standards, pelargonidin chloride, cyanidin chloride, and delphinidin chloride, were also loaded onto TLC plates as positive controls.

The same HPLC instrument and column described above were used to analyze anthocyanidins. The elution solvent system was composed of 0.1% acetic acid (solvent A) and acetonitrile (solvent B). The gradient program composed of different ratios of solvent A to solvent B to elute anthocyanidins. The program consisted of 90:10 to 83:17 (0–5 min), 83:17 to 77:23 (5–10 min), 77:23 to 71:29 (10–15 min), 71:29 to 68:32 (15–20 min), 68:32 to 65:35 (20–25 min), 60:35 (25–39 min), 65:35 to 50:50 (39–45 min), 50:50 to 70:30 (45–50 min), and 70:30 to 90:10 (50–55 min), and then followed by 10 min of column washing. The injection volume and flow rate was 5 μL and 1 mL/min, respectively. Chromatograph was recorded at 530 nm. Three authentic standards, pelargonidin chloride, cyanidin chloride, and delphinidin chloride, were injected as positive controls.

REFERENCES

1. Xie, D.-Y.; Jackson, L.A.; Cooper, J.D.; Ferreira, D.; Paiva, N.L. Molecular and biochemical analysis of two cDNA clones encoding dihydroflavonol-4-reductase from Medicago truncatula. Plant Physiol. 2004, 134, 979–994.

2. Shirley, B.W.; Hanley, S.; Goodman, H.M. Effects of ionizing radiation on a plant genome: Analysis of two arabidopsis transparent testa mutations. Plant Cell 1992, 4, 333–347.

3. Peer, W.A.; Brown, D.E.; Tague, B.W.; Muday, G.K.; Taiz, L.; Murphy, A.S. Flavonoid accumulation patterns of transparent testa mutants of Arabidopsis. Plant Physiol. 2001, 126, 536–548.

4. Watanabe, K.; Kobayashi, A.; Endo, M.; Sage-Ono, K.; Toki, S.; Ono, M. CRISPR/Cas9-mediated mutagenesis of the dihydroflavonol-4-reductase-B (DFR-B) locus in the Japanese morning glory Ipomoea (Pharbitis) nil. Sci. Rep. 2017, 7, 10028.

5. Kim, S.; Yoo, K.S.; Pike, L.M. Development of a PCR-based marker utilizing a deletion mutation in the dihydroflavonol 4-reductase (DFR) gene responsible for the lack of anthocyanin production in yellow onions (Allium cepa). Theor. Appl. Genet. 2005, 110, 588–595.

6. Shi, M.-Z.; Xie, D.-Y. Biosynthesis and metabolic engineering of anthocyanins in Arabidopsis thaliana. Recent Pat. Biotechnol. 2014, 8, 47–60.

7. Xie, D.-Y.; Dixon, R.A. Proanthocyanidin biosynthesis—Still more questions than answers? Phytochemistry 2005, 66, 2127–2144.

8. Sanchez-Moreno, C.; Cao, G.; Ou, B.; Prior, R.L. Anthocyanin and proanthocyanidin content in selected white and red Wines. Oxygen radical absorbance capacity comparison with nontraditional wines obtained from Highbush blueberry. J. Agric. Food Chem. 2003, 51, 4889–4896.

9. Demeule, M.; Michaud-Levesque, J.; Annabi, B.; Gingras, D.; Boivin, D.; Jodoin, J.; Lamy, S.; Bertrand, Y.; Beliveau, R. Green tea catechins as novel antitumor and antiangiogenic compounds. Curr. Med. Chem. Anti-Cancer Agents 2002, 2, 441–63.

10. Harborne, J.B.; H. Baxter, Anthocyanins. The Handbook of Natural Flavonoids; Harborne, J.B., Baxter, H., Eds.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1999; Volume I, pp. 1–115.

11. Kayesh, E.; Shangguan, L.F.; Korir, N.K.; Sun, X.; Bilkish, N.; Zhang, Y.P.; Han, J.; Song, C.N.; Cheng, Z.M.; Fang, J.G. Fruit skin color and the role of anthocyanin. Acta Physiol. Plant 2013, 35, 2879–2890.

12. Zhang, Y.; Butelli, E.; Martin, C. Engineering anthocyanin biosynthesis in plants. Curr. Opin. Plant Biol. 2014, 19, 81–90.

13. Meyer, P.; Heidmann, I.; Forkmann, G.; Saedler, H. A new petunia flower colour generated by transformation of a mutant with a maize gene. Nature 1987, 330, 677–678.

14. Rosati, C.; Simoneau, P.; Treutter, D.; Poupard, P.; Cadot, Y.; Cadic, A.; Duron, M. Engineering of flower color in forsythia by expression of two independently-transformed dihydroflavonol 4-reductase and anthocyanidin synthase genes of flavonoid pathway. Mol. Breed. 2003, 12, 197–208.

15. Davies, K.M.; Schwinn, K.E.; Deroles, S.C.; Manson, D.G.; Lewis, D.H.; Bloor, S.J.; Bradley, J.M. Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. Euphytica 2003, 131, 259–268.

16. Bavage, A.D.; Davies, I.G.; Robbins, M.P.; Morris, P. Expression of an Antirrhinum dihydroflavonol reductase gene results in changes in condensed tannin structure and accumulation in root cultures of Lotus corniculatus (bird's foot trefoil). Plant Mol. Biol. 1997, 35, 443–458.

17. Robbins, M.P.; Bavage, A.D.; Strudwicke, C.; Morris, P. Genetic manipulation of condensed tannins in higher plants. II. Analysis of birdsfoot trefoil plants harboring antisense dihydroflavonol reductase constructs. Plant Physiol. 1998, 116, 1133–1144.

18. Xie, D.-Y.; Sharma, S.B.; Wright, E.; Wang, Z.-Y.; Dixon, R.A. Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. Plant J. 2006, 45, 895–907.

19. Shi, M.Z.; Xie, D.Y. Engineering of red cells of Arabidopsis thaliana and comparative genome-wide gene expression analysis of red cells versus wild-type cells. Planta 2011, 233, 787–805.

20. Zhou, L.-L.; Zeng, H.-N.; Shi, M.-Z.; Xie, D.-Y. Development of tobacco callus cultures over expressing Arabidopsis PAP1/MYB75 transcription factor and characterization of anthocyanin biosynthesis Planta 2008, 229, 37–51.

21. Zhou, L.L.; Shi, M.Z.; Xie, D.Y. Regulation of anthocyanin biosynthesis by nitrogen in TTG1-GL3/TT8-PAP1-programmed red cells of Arabidopsis thaliana. Planta 2012, 236, 825–837.

22. Liu, Z.; Shi, M.Z.; Xie, D.Y. Regulation of anthocyanin biosynthesis in Arabidopsis thaliana red pap1-D cells metabolically programmed by auxins. Planta 2014, 239, 765–781.

23. Butelli, E.; Titta, L.; Giorgio, M.; Mock, H.-P.; Matros, A.; Peterek, S.; Schijlen, E.G.W. M.; Hall, R.D.; Bovy, A.G.; Luo, J.; et al. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat. Biotechnol. 2008, 26, 1301–1308.

24. Zuluaga, D.L.; Gonzali, S.; Loreti, E.; Pucciariello, C.; Degl'Innocenti, E.D.; Guidi, L.; Alpi, A.; Perata, P. Arabidopsis thaliana MYB75/PAP1 transcription factor induces anthocyanin production in transgenic tomato plants. Funct. Plant Biol. 2008, 35, 606–618.

25. Deed, R.C.; Fedrizzi, B.; Gardner, R.C. Influence of fermentation temperature, yeast strain, and grape juice on the aroma chemistry and sensory profile of Sauvignon Blanc wines. J. Agric. Food Chem. 2017, 65, 8902–8912.

26. Sambucci, O.; Alston, J.M. Estimating the value of California wine grapes. J. Wine Econ. 2017, 12, 149–160.

27. Read, P.E. Grape and wine production in diverse regions. In Proceedings of the Xxix International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes, Brisbane, Australia, 25 November 2016; Volume 1115, pp. 1–2.

28. Zhu, Y.; Peng, Q.Z.; Li, K.G.; Xie, D.Y. Molecular cloning and functional characterization of the anthocyanidin reductase gene from Vitis bellula. Planta 2014, 240, 381–398.

29. Zhu, Y.; Peng, Q.-Z.; Du, C.; Li, K.-G.; Xie, D.-Y. Characterization of flavan-3-ols and expression of MYB and late pathway genes involved in proanthocyanidin biosynthesis in foliage of Vitis bellula. Metabolites 2013, 3, 185–203.

30. Peng, Q.Z.; Zhu, Y.; Liu, Z.; Du, C.; Li, K.G.; Xie, D.Y. An integrated approach to demonstrating the ANR pathway of proanthocyanidin biosynthesis in plants. Planta 2012, 236, 901–918.

31. Sparvoli, F.; Martin, C.; Scienza, A.; Gavazzi, G.; Tonelli, C. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (Vitis vinifera L.). Plant Mol. Biol. 1994, 24, 743–755.

32. Gollop, R.; Even, S.; Colova-Tsolova, V.; Perl, A. Expression of the grape dihydroflavonol reductase gene and analysis of its promoter region. J. Exp. Bot. 2002, 53, 1397–1409.

33. El-Kereamy, A.; Chervin, C.; Roustan, J.P.; Cheynier, V.; Souquet, J.M.; Moutounet, M.; Raynal, J.; Ford, C.; Latche, A.; Pech, J.C.; et al. Exogenous ethylene stimulates the long-term expression of genes related to anthocyanin biosynthesis in grape berries. Physiol. Plant. 2003, 119, 175–182.

34. Mori, K.; Sugaya, S.; Gemma, H. Decreased anthocyanin biosynthesis in grape berries grown under elevated night temperature condition. Sci. Hortic. 2005, 105, 319–330.

35. Boss, P.K.; Davies, C.; Robinson, S.P. Analysis of the expression of anthocyanin pathway genes in developing Vitis vinifera L cv Shiraz grape berries and the implications for pathway regulation. Plant Physiol. 1996, 111, 1059–1066.

36. Petit, P.; Granier, T.; d'Estaintot, B.L.; Manigand, C.; Bathany, K.; Schmitter, J.-M.; Lauvergeat, V.; Hamdi, S.; Gallois, B. Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis. J. Mol. Biol. 2007, 368, 1345–1357.

37. Ban, T.; Ishimaru, M.; Kobayashi, S.; Shiozaki, S.; Goto-Yamamoto, N.; Horiuchi, S. Abscisic acid and 2,4-dichlorophenoxyacetic acid affect the expression of anthocyanin biosynthetic pathway genes in 'Kyoho' grape berries. J. Hortic. Sci. Biotechnol. 2003, 78, 586–589.

CHAPTER 3

Flavanol Polymerase Catalyzes the Biosynthesis of Novel Oligomeric Plant Flavonoids

Abstract

Flavonoids are a large group of plant secondary metabolites with approximately 13,000 compounds described from the plant kingdom. Based on the structures and biosynthesis, flavonoids are classified into 13 groups, in which proanthocyanidins are oligomeric or polymeric flavanols. In this study, we reported a novel group of oligomeric flavonoids and their biosynthesis. A flavanol polymerase (FP) was discovered in red 6R cells of PAP1 tobacco plants and a cDNA encoding this enzyme was cloned for functional analysis. Sequence and biochemical analyses showed that FP was a type III peroxidase. It used flavanols, such as epicatechin and catechin, as substrate to produce dimeric, trimers and higher degree oligomers. Chemical, physic, and MS/MS analyses characterized that these compounds were novel flavonoids. NMR experiments elucidated that these compounds consisted of flavonols such as epicatechin as a starter unit and flav-2-en-3-ol as extension units. Herein, we name these compounds papanridin, a novel group of plant flavonoids. Further metabolic profiling of plants and phylogenetic analysis revealed that papanridin is prevalent in the plant kingdom.

1. Introduction

Plant secondary metabolites (natural products) (PSM or PNP) play important roles in the interactions of plant and environment. A great number of PSMs play essential roles for plants to adapt arid land environment from aquatic conditions [1], given that many of them such as flavonoids protect plants from UV-irradiation caused damages [2-5]. A large array of PSMs particularly protect plants against numerous herbivores and pathogens [1, 6-8]. During the adaptation to all stressful environments, plants have evolved to produce a great number of different molecular skeletons and diverse PSMs. By 2003, more than 100,000 natural product structures were estimated to be isolated from the plant kingdom [9-15]. Since the plant biology entered the post genomics era, the plant kingdom was predicted to be able to produce more than 200,000 PSMs [15]. Shortly later, an statistical estimation showed that approximately 200,000 PSMs were found in the plant kingdom [8, 16-18]. Recently, the plant kingdom was predicted to produce approximately 1,000,000 [4, 19, 20]. The quick evolution of these PSM numbers indicates that there is a long way to fully understand the molecular diversity of PSMs in plants.

Flavonoids are a large group of plant-specialized PSMs [14, 21, 22]. To date, approximately 13,000 flavonoids have been reported from the plant kingdom [23]. Based on their structures and biogenesis from phenylalanine, flavonoids are classified to 13 main groups, chalcone, flavanones, flavones, aurones, dihydroflavonols, flavonols, flavan-3,4-diols, anthocyanins, flavanols, proanthocyanidins (PAs), isoflavonoids, bioflavonoids, and neoflavonoids (Fig. 1). Of these groups, PA is the only oligomeric or polymeric flavonoids that are biosynthesized from flavanols [22, 24]. Flavonoids play important roles in plant adaptation, pollination, and

reproduction during the evolution [2-5]. Anthocyanins are pink, red, purple, violet, or blue pigments that play key roles to attract pollinators and seed dispensers for reproduction of plants [25]. Flavanols and PAs protect plants against diseases caused by pathogens and damages caused by UV-radiation [26]. Moreover, flavonoids are fundamental nutrients that are consumed by humans daily [27]. In general, they are strong antioxidants and have antiviral, anti-cardiovascular, anti-cancer, and anti-aging diseases [28-33]. Recently, we reported that flavanol-gallates and dimeric PAs had a promising activity to inhibit SARS-Cov-2 and showed a potential to treat COVID-19 [34]. In summary, hunting for novel flavonoids is still one of main approaches to discover and innovate new medicines for human health.

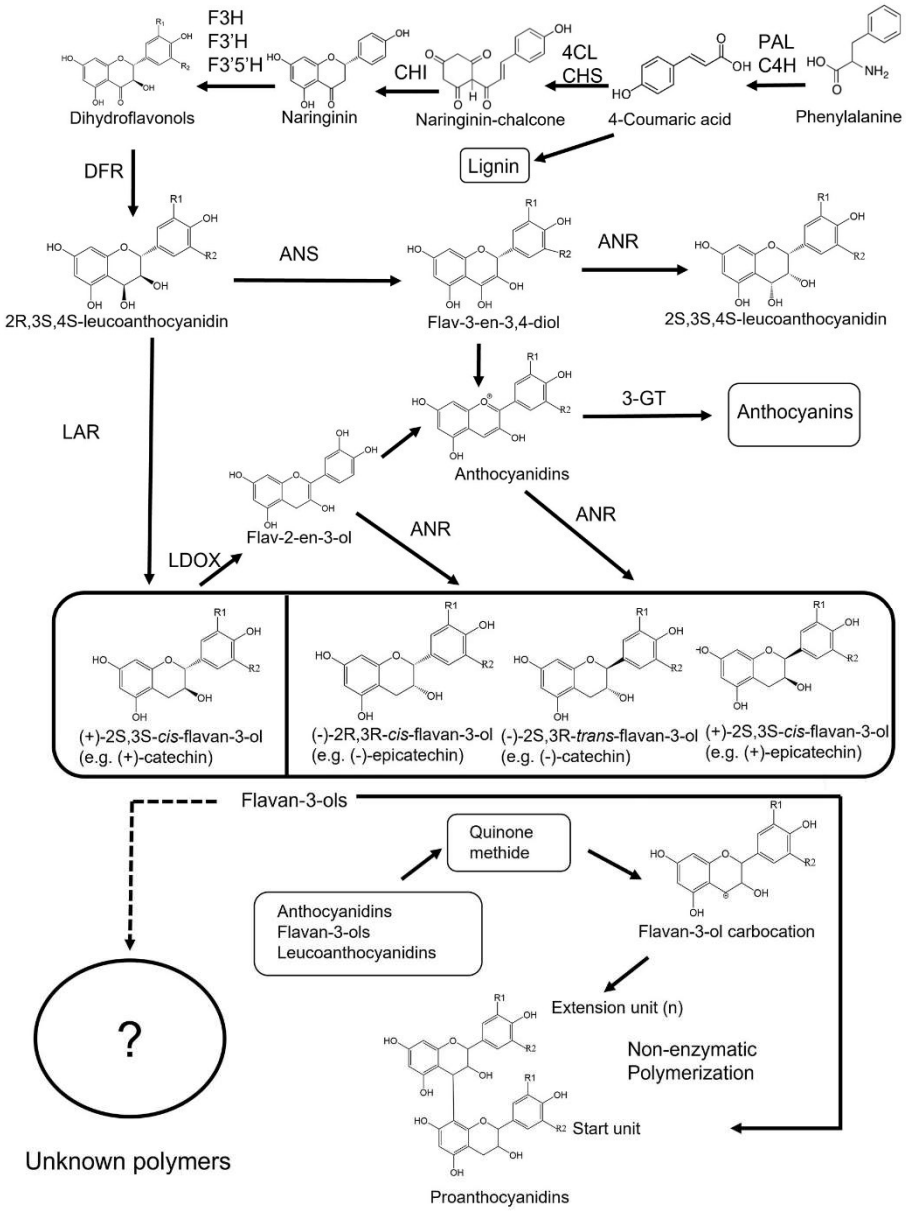


Figure 1 Biosynthesis of flavanols and proanthocyanidins starting with phenylalanine. Proanthocyanidins are the only group of oligomeric or polymeric flavonoids from flavanols linked by an interflavan bond. The question mark means whether or not flavanols can form other oligomers or polymers. Abbreviation, PAL: phenylalanine ammonia-lyase, C4H: coumarate 4-hydroyxlase, 4CL: 4 coumarate-CoA ligase, CHS: chalcone synthase, CHI: chalcone isomerase, F3H: flavonoid 3-hydroxylase, F3'H: flavonone 3'hydroxylase, F3'5'H: flavonone 3'5' hydroxylase, DFR: dihydroflavonol reductase, LAR: leucoanthocyanidin reductase, ANS: anthocyanidin synthase, LDOX: leucoanthocyanidin dioxygenase, and ANR: anthocyanidin reductase.

The biosynthesis of flavonoids has gained intensive studies over the past three decades. Although how plants polymerize PAs remains open for studies, pathways to each group have been biochemically, genetically, and molecularly elucidated (Fig. 1). The main pathway includes beginning (from phenylalanine to coumaryl-CoA), early (coumaryl-CoA to dihydroflavonols), and late steps (dihydroflavonols to flavonols, anthocyanins, flavanols, and PAs) [25]. Main pathway genes include phenylalanine ammonia-lyase (*PAL*), coumarate 4-hydroyxlase (*C4H*), 4 coumarate-CoA ligase (*4CL*) chalcone synthase (*CHS*), chalcone isomerase (CHI), flavonone 3-hydroxylase (F3H), isoflavonoid synthase (IFS), flavone synthase (FNS), flavonoid 3'hydroxylase (F3'H), flavonoid 3'5' hydroxylase (F3'5'H), dihydroflavonol reductase (DFR), flavonol synthase (FLS), leucoanthocyanidin reductase (LAR), anthocyanidin synthase (ANS), leucoanthocyanidin dioxygenase (LDOX), anthocyanidin reductase (ANR) (Fig. 1). [26, 35-38]. All known flavonoids are biosynthesized from the main pathway.

PAs, hydrolysable tannins, and lignin are only three groups of oligomeric or polymeric PSMs. PAs are oligomers or polymers of flavanols (flavan-3-ols) polymerized via an interflavan bond formed between the C4 of an upper unit and the C8 of a bottom unit [24] (Fig. 1). PAs and flavanols protect plants against damages caused by pathogens and UV-radiation [26]. Furthermore, oligomeric PAs and flavonols are potent antioxidants with multiple health benefits, such as anti-viral, anti-cancer, anti-atherosclerosis, anti-heart, anti-aging diseases, and anti-COVID activities [34, 39-45]. To date, how plants polymerize flavanols to PAs remain open for investigation. Our previous hypotheses were that polyphenol oxidases, lacases, or peroxidases might be responsible for the catalysis of this step [24]. We proposed that these enzymes might oxidize flavanols to quionone methides to form flavanol carbocation, which was attacked by

flavanols to form PAs (Fig. 1) [24]. On the one hand, unfortunately, no evidence has been obtained to demonstrate these hypotheses. On the other hand, recent evidence has revealed that the polymerization of PAs might be more complicated than the previous hypotheses (Fig. 1) [46]. A new function of LAR was found to catalyze beta-cysteinyl epicatechin to epicatechin, which was a non-oxidoreduction catalysis associating with the formation of non-extractable PAs [47]. LDOX was unexpectedly shown to catalyze catechin to epicatechin via flav-2-en-3-ol [48]. Although ANR was demonstrated to catalyze anthocyanidins to flavonols via proposed flav-2-en-3-ol, flav-3-en-3-ol, and flavan-3-one [49, 50], a recent follow up study revealed that ANR could also convert flav-3-en-3, 4-diol to 2S, 3S, 4S-leucoanthocyanidins [51]. We recently for the first time demonstrated the presence of flavanol carbocation involved in the formation of PAs in plant cells [52]. These new discoveries and the multiple roles of flavanols indicate that additional efforts are necessary to fully understand the polymerization mechanism of PAs.

Although the current understanding of plant flavonoids is relatively comprehensive, there are still many questions that remain for answers, given that the majority of plants have not been phytochemically studied. Are there new novel skeletons of flavonoids in the plant kingdom? Are flavanols, anthocyanins, flavonols, and flavones are end products in each branch? In addition to PAs, can flavanols be precursors of other unknown flavonoids? How many new flavonoids are unknown and what are their roles? All these questions remain for massive investigations. Based on plant taxonomy, although how many plants that the plant kingdom has remains for further studies, the plant list database recorded 350,699 species [53-55], the majority of which remains for phytochemical investigation. Although one main necessary approach is to study each plant one by one, metabolic engineering of plants has shown a power to discover new compounds that

plants may not produce in natural conditions. For example, *Arabidopsis thaliana* mainly produced two anthocyanins [56]. However, our studies showed that the dominant *pap1-D* mutants overexpressing *PAP1* gene (encoding MYB75, a R2R3-MYB transcription factor) could produce 27 additional anthocyanins [57]. In addition, we have showed that the overexpression of PAP1 created new anthocyanins in tobacco plants [58-61]. Therefore, an integration of cutting-edge technologies and phytochemical approaches is important to discover new flavonoids.

In this study, we report to take advantage of our PAP1 and PAP1-BAN (or ANR) tobacco plants to discover a novel type of flavonoid skeleton and a group of novel oligomeric flavonoids in the plant kingdom. PAP1 tobacco was created by the overexpression of *PAP1* [24]. This novel genotype highly biosynthesizes anthocyanins in all tissues. *BAN* encodes ANR that is a key reductase catalyzing the biosynthesis of three stereo types of flavanols (Fig. 1) [24, 49, 62]. The cross of PAP1 and ANR tobacco plants has created PAP1-ANR (BAN) genotypes that can produce flavanols and PAs in leaves and flowers [63]. However, although PAP1-ANR tobacco produces high contents of anthocyanin and appropriate levels of flavanols, the production of PAs is relatively low. In addition, our preliminary data of metabolic profiling showed that PAP1-ANR plants produced a number of new compounds compared to wild-type ones. Based on these phytochemical phenotypes, we hypothesized that PAP1 and PAP1-ANR tobacco plants might express unknown pathways leading to the biosynthesis of new metabolites. In addition, since PAs are only produced in red cells [49], we further hypothesized that red cells expressed enzymes toward the biosynthesis of both PAs and unknown metabolites. To test these hypotheses, we engineered red 6R cell and P3 control cell lines from PAP1 and wild-type tobacco plants, respectively [59]. Red 6R cells produce high levels of anthocyanins, while P3

cells do not produce anthocyanins. These cell lines were used to identify enzymes and genes for functional analysis. A protein was demonstrated to be a flavanol polymerase (FP). A cDNA encoding this FP was cloned from red cells for further functional characterization. Transgenic, phytochemical, metabolic profiling, NMR, and chemical data demonstrated that FP catalyzed flavanols to produce novel dimeric and oligomeric flavonoids. These novel flavonoids are prevalent in the plant kingdom. Since FP was discovered in red PAP1 cells and novel flavonoids were formed in PAP1-ANR tobacco plants, we named this novel group of flavonoids as papanridin.

2. Results

2.1 Isolation of flavanol polymerase from red 6R suspension cells

6R red cells (Supplementary Fig. S1a) were engineered from a novel red PAP1 tobacco variety (*Nicotiana tabacum* Xanthi) that was programed to highly produce anthocyanins in all tissues by the overexpression of *PAP1* [59, 63]. In addition, P3, a control cell line, was engineered from wild-type plant as control. Based on our success of proanthcoyanin (PA) engineering in PAP1 tobacco plants [63], we hypothesized that red cells expressed genes encoding enzymes to polymerize flavanol to dimeric or oligomeric PAs. red 6R and white P3 cells were used to develop cell suspension culture (Fig. 1a and Supplementary Fig. 1b). Both suspension cells and liquid medium were collected to isolate crude proteins (Fig. 1b and supplementary Fig. S2a), which were tested for a polymerase activity with epicatechin and catechin. As hypothesized, the crude protein extracts from both cell lines and liquid media converted epicatechin and catechin to yellowish compounds (Supplementary Fig. 2c and d). In addition, proteins were boiled for 1-10

min as controls. An unexpected result was that after boiling for 1-3 min, SDS-PAGE analysis showed the presence of multiple proteins from crude protein extracts of both 6R and P3 cells (Fig. 1b and Supplementary Fig. 2 a). Further enzyme tests with both catechin and epicatechin revealed that enzymes boiled for 1-3 min still converted substrates to yellowish compound(s) (Supplementary Fig. 2c and d) although intensity of yellowish color was reduced compared with that of unboiled enzymes. Further TLC analysis stained with DMACA showed that unboiled and 1-3 min boiled proteins converted epicatechin and catechin to dimeric PA-like compounds (Fig. 2 c and supplementary Fig. S2e). To understand whether these compounds resulted from a real catalysis of proteins, crude, boiled, and proteinase K(PK)-digested proteins were separated on native gels. Then, the gels were placed in the pH7 Tris-HCl buffer supplemented with epicatechin. In 30 min, those lanes with crude, 3 and 6-min boiled, and PK-digested 6R proteins showed yellowish to deep yellowish coloration, those lanes with crude, 3-min boiled, and PK-digested P3 proteins showed light yellowish coloration, but two lanes with 6-min digested P3 proteins and BSA did not show any yellowish coloration (Supplementary Fig. 2f). These results not only demonstrated that crude proteins and 3-6 min boiled 6R proteins contained active enzymes, but also revealed that red 6R protein extracts had different active enzymes from P3 proteins (Supplementary Fig. S2).

Figure 2 A flavonol polymerase isolated from red tobacco cell engineered by PAP1.

a, a suspension culture of 6R red cells was developed to produce proteins.

b, a SDS-PAGE gel image shows band profiles of crude protein extracts from red 6R cell. UB: unboiled protein and 1m, 2m, and 3m: boiled for 1, 2, and 3 min.

c, a TLC image shows dimeric PA-like compounds produced from incubations consisted of crude proteins and catechin and epicatechin. CA: catechin, EC: epicatechin, B1 and B2: procyanidin B1 and B2, EC+E (I+II): reactions of epicatechin and two unboiled protein extracts (I and II), CA+E: reactions of catechin and two unboiled protein extracts (III and IV), EC+BE: reactions of epicatechin and two 3-min boiled protein extracts (V and VI), CA+BE: reactions of catechin and two 3-min boiled protein extracts (VII and VIII), E: two enzyme extract controls (IX and X) without adding substrates.

d, a SDS-PAGE image shows protein bands separated from a DEAE Sephadex A-20-120 column with elution buffer supplemented with 0.1 M NaCl. N/A: buffer without adding NaCl.

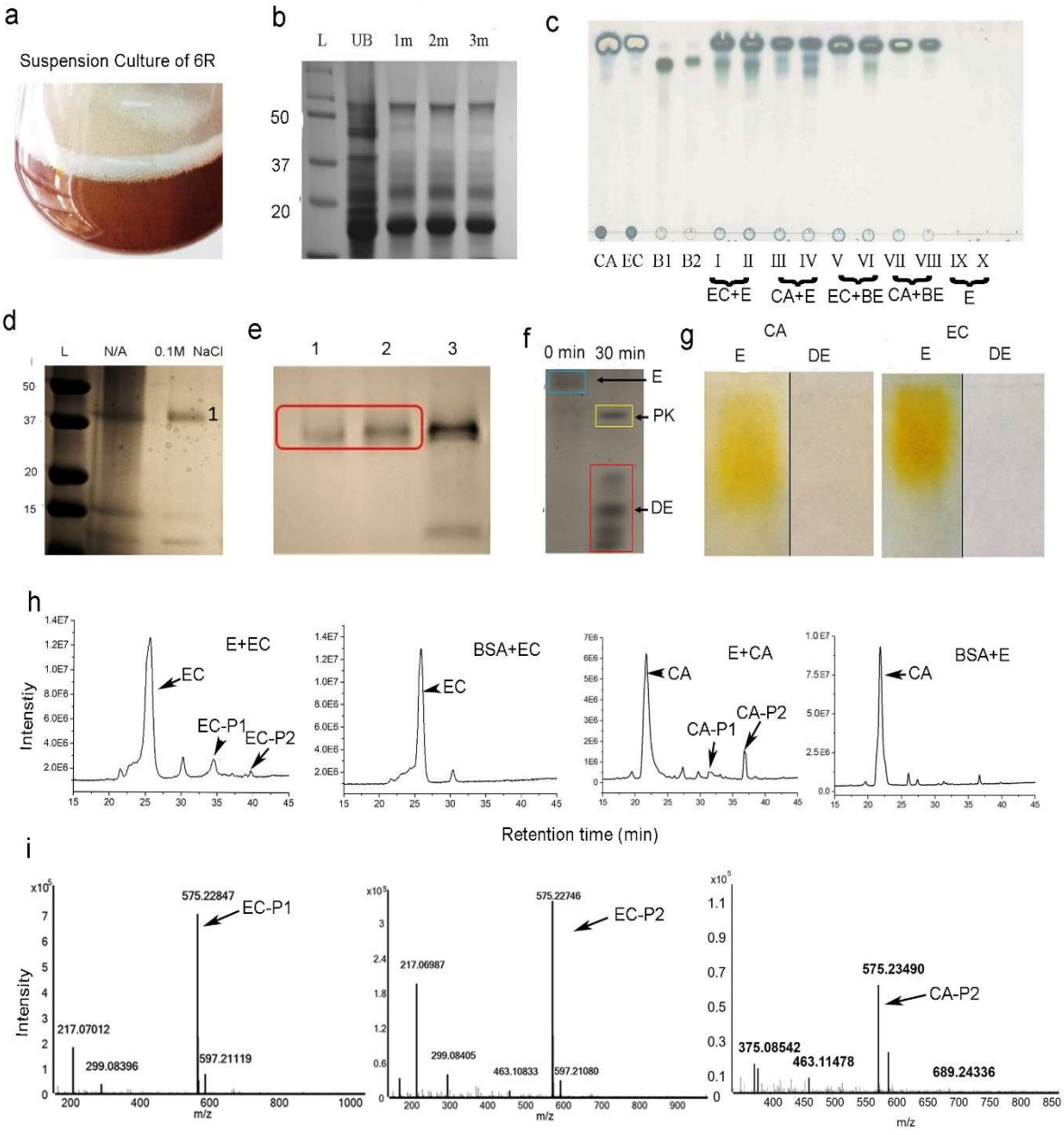
e, a SDS-PAGE image shows protein bands of three fractions (1-3) isolated from a Sephadex G-75 column.

f, a SDS-PAGE image shows the purified protein (0 min) and its peptides resulted from the digestion of proteinase K (PK) for 30 min (30 min).

g, native PAGE images show that the protein purified from the Sephadex G-75 column (f) catalyzes catechin and epicatechin into yellowish compounds on the gel but PK-digested products (f) cannot. CA: catechin, EC: epicatechin, E: enzyme, and DE: digested enzyme.

h, total ion chromatograms that the purified enzyme converts epicatechin (E+EC) to two products (EC-P1 and EC-P2), the purified enzyme converts catechin (CA+E) to two new products (CA-P1 and CA-P2), but BSA control does not catalyze epicatechin and catechin (BSA+EC and BSA+CA). CA: catechin, EC: epicatechin, EC-P1 and EC-P2: epicatechin products 1 and 2, CA-P1 and CA-P2: catechin products 1 and 2.

i, the mass spectra characterized that EP-P1, EP-P2, and CA-P2 had the same the mass to charge ratio 575 $[m/z]^-$.


a
Suspension Culture of 6R
b
L UB 1m 2m 3m
50
37
20
c
CA EC B1 B2 I II III IV V VI VII VIII IX X
EC+E
CA+E
EC+BE
CA+BE
E
d
L N/A 0.1M NaCl
50
37
20
15
1
e
1 2 3
f
0 min 30 min
E
PK
DE
g
CA
EC
E DE E DE
h
E+EC
EC
EC-P1
EC-P2
BSA+EC
EC
E+CA
CA
CA-P1
CA-P2
BSA+E
CA
Intenstiy
Retention time (min)
i
575.22847
EC-P1
217.07012
299.08396
597.21119
575.22746
EC-P2
217.06987
299.08405
463.10833
597.21080
575.23490
CA-P2
375.08542
463.11478
689.24336
Intensity
m/z

Next, we boiling, ion-exchange column, and Sephadex column to purify one protein (Fig. 2d-e), which was separated with native gels. In situ reaction on gel showed that the purified protein on the gels converted catechin and epicatechin to yellowish compounds, while proteinase K-digested protein did not (Fig. 2f-g and Supplementary Fig. S3). LC-MS analysis detected two new compounds from the enzymatic reactions, which were not produced from negative control incubations (Fig. 2h). These compounds from epicatechin and catechin were named as EC-P1, EC-P2, CA-P1 and CA-P2, which had different retention time. Further MS analysis showed that compounds had the same 575.23 $[m/z]^-$ value. This value indicates that their molecular weight is 576 Dalton, the same as that of procyanidin A1 and A2 but two protons less than dimeric PB1-8, 578 Dalton, suggesting that they are not dimeric B-type PAs. Given that these new yellowing compounds, like flavan-3-ols and PAs, react with DMACA to form bluish coloration, we named the purified protein Flavanol Polymerase (FP).

2.2 Gene cloning and functional analysis of FP

We used two steps to clone a cDNA encoding FP. First, the purified FP protein was sequenced to obtain seven peptides, which were used for blast search at GenBank (Supplementary Fig. S4). The blasting results showed that all peptides hit one single tobacco protein, annotated to be a peroxidase (Supplementary Fig. S4). Second, we designed a pair of primers to clone approximately 1 kb cDNA from red 6R cells. Meanwhile RT-PCR analysis revealed that it was not expressed in leaves, stems, roots of tobacco plants and white 6W cells (Fig. 3a). A sequence alignment with deduced amino acids and modeling showed that it belonged to a member of the class III of the peroxidase family I (Supplementary Fig. S5a), which featured with four conserved disulphide bridges and two calcium ions [64, 65]. In addition, the N-end has a

secretory signal peptide. A C-end fusion of GFP and confocal analysis of transgenic roots showed the vacuolar localization of FP (Supplementary Fig. S6 a-b). A recombinant FP was purified from its expression induced in *E. coli* (Fig. 3b and Supplementary Fig. S7). The incubation of the purified recombinant FP and epicatechin in an oxygenated buffer produced yellowish compounds (Fig. 3c), supporting the results from the native FP (Fig. 2). Further in situ reaction on native gel demonstrated the catalysis of the recombinant FP (Fig. 3d). LC-MS analysis for the yellowish compounds revealed EC-P1and EC-P2 from epicatechin and CA-P1 and CA-P2 from catechin (Fig. 3e), which supported the results from the native FP analysis (Fig. 2). Binary constructs (Supplementary Fig. S6a) were generated to obtain *FP* transgenic PAP1-BAN tobacco plants (PAP1-BAN-FP) for the overexpression of *FP* (Fig. 3f-h). LC-MS profiling was completed to analyze these new compounds and EC in transgenic plants (Fig. 3g and h). This analysis identified EC-P1 and EC-P2 from the leaf extracts of PAP1-BAN-FP plants, while only EC-P1 in those of control transgenic PAP1-BAN plants (Fig. 3g). In comparison, the peak area size of EC-P1 was much bigger in PAP1-BAN-FP extracts than in control ones. Further estimation showed the significant higher contents of these two compounds in PAP1-BAN-FP extracts than in control ones (Fig. 3i). In contrast, the contents of EC were significantly lower in in PAP1-BAN-FP extracts than in control ones (Fig. 3h and j). These data demonstrate that the transgenic FP convert EC into EC-P1 and EC-P2 compounds in planta. LC-MS analysis also detected PB2 and quantitative estimation showed a higher content in PAP1-BAN-FP plants than in control ones (Supplementary Fig. S8a-b). A further butanol-HCl boiling showed the increase of total PAs in PAP1-BAN-FP plants than in control ones (Supplementary Fig. S8 c).

The *in vitro* catalysis of FP was further characterized with different reaction conditions. First, it catalysis is dependent upon the presence of oxygen. Three pH conditions, pH 6, 7 and 8, were

tested to understand the necessity of oxygen. When the reaction buffer was neither oxygenated by vortexing strongly or adding H_2O_2, almost its catalysis was not observed or only a detectable level (Supplementary Fig. S9a-d). In addition, deoxygenation of buffer via dissolving nitrogen led to the complete loss of the catalysis of FP. By contrast, in addition to oxygenation via vortexing buffer described above, an addition of H_2O_2 strongly initiated and enhanced the catalysis of FP in incubations at pH 6 and 7, while the reaction at pH8 was different (Supplementary Fig. S9 a-d). In the conditions of pH 6 and 7, EC was not converted to yellowish compounds in the absence of H_2O_2, while EC was quickly converted to brown-yellowish compounds in the presence of H_2O_2 (Supplementary Fig. S9a-d); two negative controls with BSA did not convert epicatechin to yellowish compounds; TLC analysis could detect PA-like compounds in the presence of H_2O_2, but hardly detected those compounds in the absence of H_2O_2 (Supplementary Fig. S9b and d); LC-MS analysis showed that FP converted to epicatechin to EC-P1, EC-P2, and another 861 $[m/z]^-$ compound (Supplementary Fig. S9e) and catechin to CA-P1 and CA-P2 (Supplementary Fig. S10), while BSA controls did not covert epicatechin and catechin to corresponding compounds. In the pH8 condition (Supplementary Fig. S9a and b), in the absence of H_2O_2, epicatechin was converted to yellowish compounds, while the presence of H_2O_2, this catalysis was strongly enhanced; in two negative control reactions with BSA, epicatechin was also converted to slight yellowish compounds; however, TLC analysis could not or hardly detect PA-like compounds (Supplementary Fig. S9a-b); LC-MS analysis only detected EC-P1 from EC in the absence of H_2O_2 (Supplementary Fig. S9f) but could not detect these compounds from EC in the presence of H_2O_2 and negative control conditions; LC-MS analysis could not detect CA-P1 and CA-P2 from catechin in all conditions (Supplementary Fig. S10). Based on these data, this yellowish coloration at pH8 likely mainly resulted from the degradation

of EC and CA, the products of which were extremely water soluble but could hardly be extracted by ethyl acetate for TLC and LC-MS analysis. Second, the velocity of the FP catalysis was linear during the incubation times from 30-120 min; NADPH, beta-mercaptoethanol, and vitamin C, were inhibitors of the catalysis of FP; the optimum pH value based on the reduction of substrate was pH8; and the optimum of temperature was 50ºC (Supplementary Fig. S11 a-d). Third, PB used catechol as a substrate but did not use phloroglucinol as a substrate (Supplementary Fig. S12).

Figure 3 Cloning of gene and functional analysis of recombinant flavanol polymerase (FP).

a, a cDNA of FP was amplified from suspension red 6R cells by RT-PCR, namely *FP*. M, DNA ladder, L: leaves, R: roots, S: stems, and 6W: a white cell line engineered together with red 6R cells from the same explant.

b, a SDS-PAGE gel image shows the recombinant GST-FP-His protein purified.

c, GST-FP-His (E) protein catalyzed epicatechin (EC) to yellowish compounds, but BSA control did not. The incubation was consisted of FP or BSA and epicatechin in 50 mM pH 7 Tris-HCl buffer that was oxygenated by strong vortex. 0min, 30min, 1h, 2h, and o/n: reaction 0 and 3o min, 1 and 2 hrs, and overnight at room temperature.

d, an image shows that the purified GST-FP-His protein on the native-PAGE gel converted epicatechin to yellowish compound(s), but not the BSA control did not after the gels were incubated with epicatechin for 30 min. Gels were placed in 50 mM pH 7 Tris-HCl supplemented with 1.0 μg/μl contained in petri dishes, which were shaken on a rotary shaker.

e, TIC profiles of 30 min-reaction products revealed that GST-FP-His (E) converted epicatechin into two products (EC-P1 and EC-P2) and catechin into one main product (CA-P2), but BSA control did not have these activities.

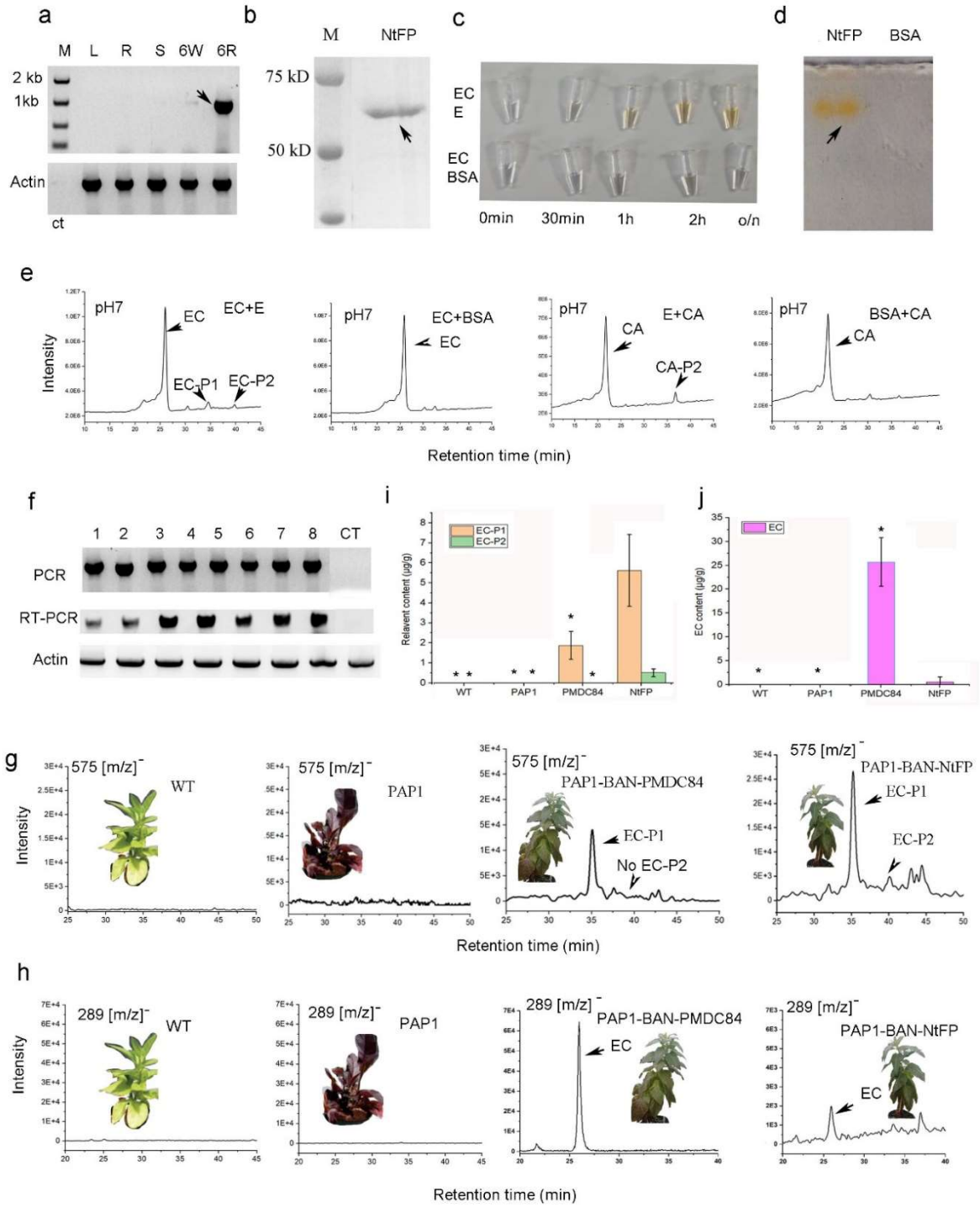
f, PCR and RT-PCR demonstrate the integration and expression of *FP* transgene in transgenic PAP1-ANR plants but in PMDC84 vector control transgenic plants (CT).

g, elective ion chromatogram (EIC) analysis detected EC-P1 and EC-P2 from leaf extracts of FP transgenic PAP1-BAN plants and EC-P2 from PMDC84 transgenic PAP1-BAN plants, but did not detect these compounds from wild type and PAP1 tobacco plants. Plant picture inserts show their phenotypes. There were not phenotypic differences in PAP1-BAN-MPDC84 and PAP1-BAB-NtFP transgenic plants.

h, EIC analysis detected the reduction of epicatechin (EC) in leaves of FP transgenic PAP1-BAN plants compared to PMDC84 control transgenic PAP1-BAN plants. EC was not detected from wild type and PAP1 plants. Plant picture inserts show their phenotypes.

i, the levels of both EC-P1 and EC-P2 were significantly higher in leaves of FP transgenic plants than in those of PDMC84 transgenic ones. However, these compounds were not detected from leaves of wild type and PAP1 plants.

j, the content of epicatechin (EC) was significantly lower in leaves of transgenic PAP1-BAN plants than in those of PDMC84 transgenic ones. However, EC was not detected from leaves of wild type and PAP1 plants.


a
M L R S 6W 6R
2 kb
1kb
Actin
ct
b
M NtFP
75 kD
50 kD
c
EC
E
EC
BSA
0min 30min 1h 2h o/n
d
NtFP BSA
e
pH7 EC EC+E EC-P1 EC-P2
pH7 EC+BSA EC
pH7 E+CA CA CA-P2
pH7 BSA+CA CA
Intensity
Retention time (min)
f
1 2 3 4 5 6 7 8 CT
PCR
RT-PCR
Actin
i
EC-P1
EC-P2
Relavent content (μg/g)
WT PAP1 PMDC84 NtFP
j
EC
EC content (μg/g)
WT PAP1 PMDC84 NtFP
g
575 [m/z]⁻
WT
PAP1
PAP1-BAN-PMDC84
EC-P1
No EC-P2
PAP1-BAN-NtFP
EC-P1
EC-P2
Intensity
Retention time (min)
h
289 [m/z]⁻
WT
PAP1
PAP1-BAN-PMDC84
EC
PAP1-BAN-NtFP
EC
Intensity
Retention time (min)

2.3 EC-P1, EC-P2, CA-P1, and CA-P2 are novel dimeric flavonoids

We performed LC-qTOF-MS/MS-based fragmentation, phloroglucinol degradation, and butanol-HCl analyses to characterize EC-P1, EC-P2, CA-P1, and CA-P2. First, to annotate their potential structures, we used PB2 and PA2 as references for MS/MS analysis. Three characteristic fragmentation types of PB2 include heterocyclic ring fission (HRF), C4 and C8 interflavan bond fission (IBF), and Retro-Diels-Alder (RDA) fission [66](Fig. 4a-b). Skeleton based-main fragments of PB2 generated by collision induced dissociation (CID) were 125 and 451 from HRF, 289 from IBF, and 407 from RDA (Fig. 4a-b). The main fragments derived from the CID of EC-P1 included 125 and 449 from HRF, 287 and 289 from IBF, and 407 from RDA (Fig. 4c). Another main fragment 271 was likely derived from the removal of one oxygen from the fragment of 287. The main fragments from PA2 included 125 and 449 from HRF, 285 (or 286) and 289 from IBF, and 423 from RDA. These main finger printings indicate that on the one hand, EC-P1 has some similar structural features as PA2 and PB2, on the other hand, it is different from these standards. Second, in the condition of methanol-HCl and phloroglucinol, PB2 and other B-type PAs degradation to produce epicatechin-phloroglucinol (Fig. 4e) or catechin-phloroglucinol [52]. This degradation was performed for PB2, PA2, and EC-P1. As expected, the degradation of PB2 produced epicatechin-phloroglucinol, with a 413 $[m/z]^-$ (Fig. 4f). However, this reaction did not degrade EC-P1 to produce epicatechin-phloroglucinol (Fig. 4g). In addition, this reaction did not degrade CA-P1 and PA2 (Supplementary Fig. 13). These results indicate that the acidic methanol and phloroglucinol reaction cannot degrade EC-P1. Third, all A- and B-type PAs are cleaved to anthocyanidins by butanol-HCl boiling. We performed this experiment. The results showed that butanol-HCl boiling produced red cyanidin pigment from PB2 but could not produce any red pigment from EC-P1 (Fig. 4i). This result

indicates that butanol-HCl could not cleave EC-P1 to anthocyanidins. Moreover, the same results were obtained from EC-P2, CA-P1, and CA-P2. In conclusion, EC-P1, EC-P2, CA-P1, and CA-P2 are novel dimeric flavonoids.

Figure 4 EC-P1, a new dimeric compound different from PAs

MS, phloroglucinol analysis, and butanol-HCl boiling were completed to compare EC-P1 and PB2.

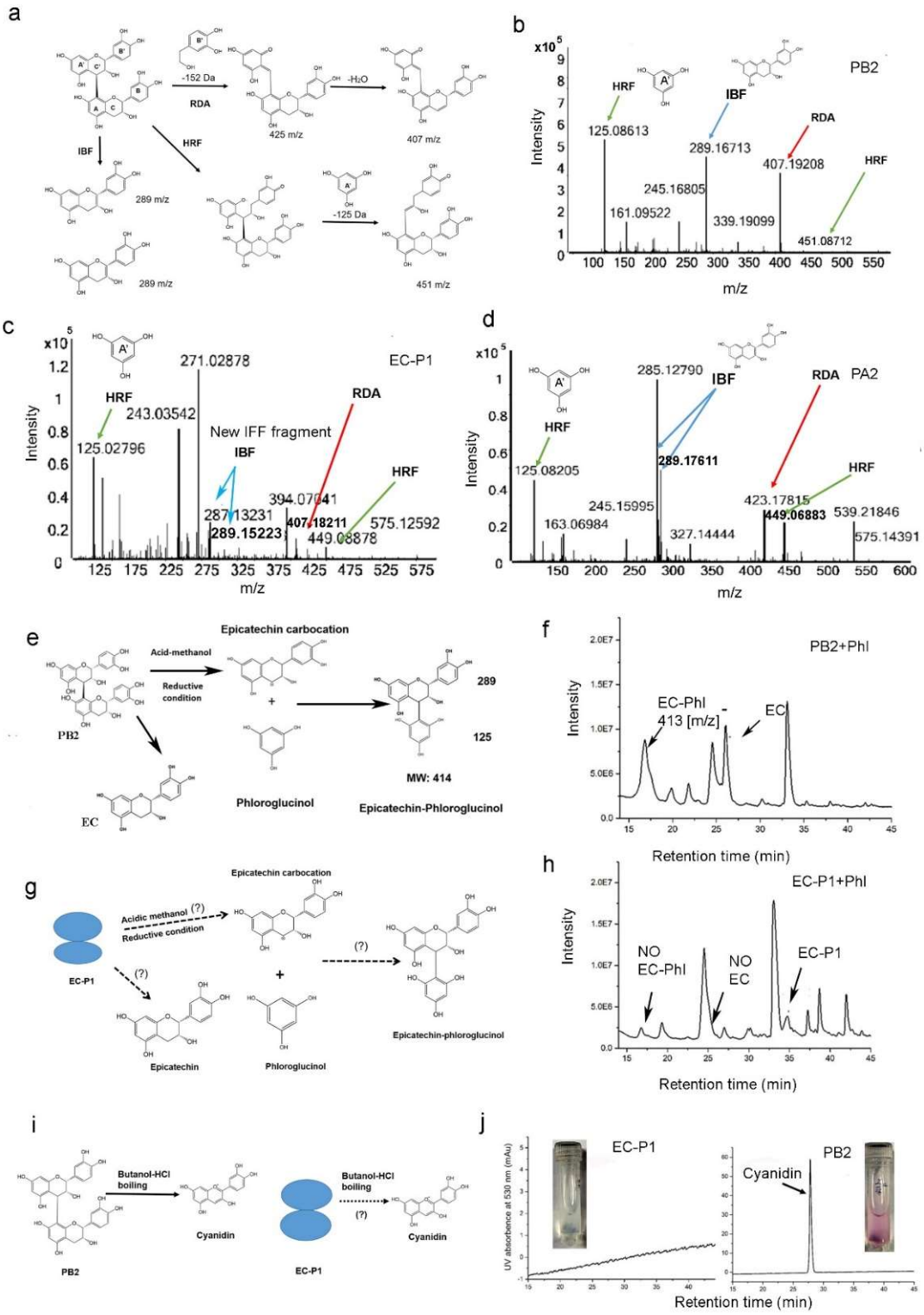
a, a scheme characterizes three fragmentation mechanisms of PB2 by collision induced dissociation (CID), Retro-Diels-Alder (RDA), heterocyclic ring fission (HRF), and interflavan bond fission (IBF).

b-d, MS/MS profiles show that PB2 (b), EC-P1 (c), and PA2 (d) have the same fragmentation mechanisms, RDA, HRF, and QM. However, except for the fragment of 125 resulted from HRF, main fragments from QM and RDA are different. These fragment patterns indicate that on the one hand, EC-P1 is a different compound from PB2 and PA; on the other hand, they have the same A and A' rings.

e-f, phloroglucinol-HCl based degradation of PB2 produces epicatechin (EC) and epicatechin-phloroglucinol (EC-Phl). (e), a scheme shows the theory that PB2 is degraded to produce epicatechin (EC) and epicatechin-phloroglucinol (EC-Phl). (f), the TIC profile detects EC-Phl with 413 $[m/z]^-$ and epicatechin from PB2.

g-h, phloroglucinol-HCl based degradation demonstrates that EC-P1 is not a dimeric PA compound. (g), a scheme proposes whether EC-P1 can degraded to produce epicatechin (EC) and epicatechin-phloroglucinol (EC-Phl). (h) a TIC profile shows that there are no EC-Phl and EC formed from the phloroglucinol-HCl based degradation of EC-P1.

i-j, the butanol-HCl boiling based cleavage shows that EC-P1 is not a PA compound. (i), this scheme shows the theory that the butanol-HCl boiling of PB2 produces cyanidin and a hypothesis whether or not that of EC-P1 can produce anthocyanidins. (j), HPLC profiles and coloration of the boiling show that EC-P1 cannot produce anthocyanidins (left) but PB2 can produce cyanidin (right) after the butanol-HCl boiling.


a
-152 Da
RDA
-H2O
425 m/z
407 m/z
IBF
HRF
289 m/z
289 m/z
-125 Da
451 m/z
b
PB2
HRF
IBF
RDA
HRF
125.08613
289.16713
407.19208
245.16805
161.09522
339.19099
451.08712
Intensity
m/z
c
EC-P1
271.02878
HRF
243.03542
RDA
New IFF fragment
IBF
HRF
125.02796
394.07041
287.13231
407.18211
289.15223
575.12592
449.08878
Intensity
m/z
d
PA2
285.12790
IBF
RDA
HRF
125.08205
289.17611
HRF
245.15995
423.17815
449.06883
539.21846
163.06984
327.14444
575.14391
Intensity
m/z
e
Epicatechin carbocation
Acid-methanol
Reductive condition
PB2
289
125
MW: 414
EC
Phloroglucinol
Epicatechin-Phloroglucinol
f
PB2+Phl
EC-Phl
413 [m/z]
EC
Intensity
Retention time (min)
g
Epicatechin carbocation
Acidic methanol (?)
Reductive condition
EC-P1
(?)
(?)
Epicatechin
Phloroglucinol
Epicatechin-phloroglucinol
h
EC-P1+Phl
NO
EC-Phl
NO
EC
EC-P1
Intensity
Retention time (min)
i
Butanol-HCl
boiling
Cyanidin
PB2
Butanol-HCl
boiling
(?)
Cyanidin
EC-P1
j
EC-P1
PB2
Cyanidin
UV absorbance at 530 nm (mAu)
Retention time (min)

2.4 Structural elucidation by NMR and nomenclature of papanridin

We purified EC-P1 via preparation of TLC and HPLC. HPLC-MS analysis showed that EC-P1 was isolated with one single peak (Supplementary Fig. S14). The purified EC-P1 is yellowish and has a special absorbance at 388 nm (Supplementary Fig. S14). Like PB2, this compound could not be crystalized.

NMR analysis was performed to elucidate structures. Experiments performed included ROESY, HMBC, and thermal shifting. Given that EC-P1 shared features with PB2, we used this standard as a reference. All parameters were obtained for 30 carbons, all protons, 10 hydroxy groups, and interflavan bond in PB2 (Supplementary Figure S15 and Figure S16).

All data demonstrated that EC-P1 had 30 carbons, 13 protons, 10 hydroxy groups, and one interflavan bond between C4 and C8. Based these assignments, we demonstrated that EC-P1 was a novel dimeric compound consisting of epicatechin (the bottom unit) and flav-2-en-3-ol (the upper unit) linked by an interflavan bond between C4 of the upper unit and C8 of the bottom unit (Fig. 5). Given that this compound is formed from the catalysis of FP from PAP1 cells and biosynthesized in PAP1-ANR plants, we named this new type of compounds as papanridin. Subsequently, we named EC-P1 and EC-P2 from epicatechin as papanridin ECII-A and -B, and CA-P1 and CA-P2 as papanridin CAII-A and –B, in which “II” means dimers.

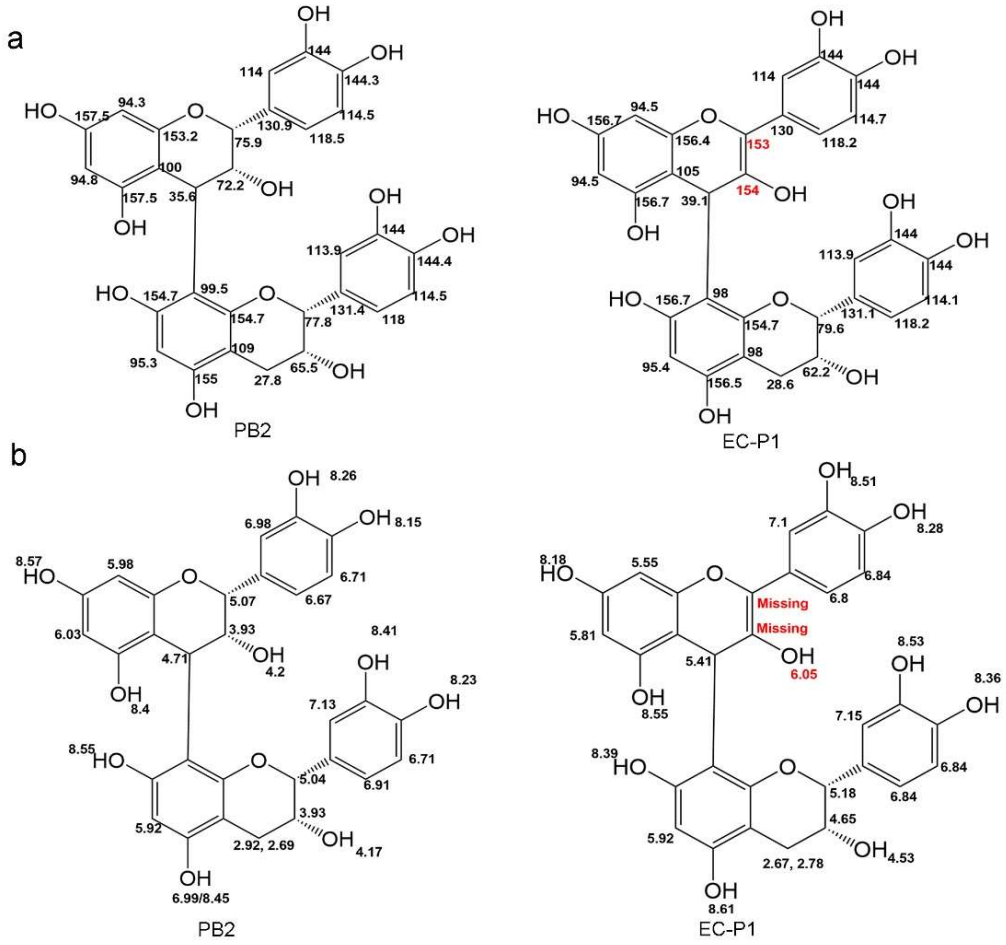


Figure 5 Novel dimeric flaven-flavan structures determined by nuclear magnetic resonance (NMR) analysis.

a, 13C NMR assignments of 30 carbons on PB2 and EC-P1. The black numbers in two compounds showed the similar PPM values of carbons on PB2 and EC-P1, which demonstrates these carbons on these two compounds are the same. The red numbers in EC-P1 showed the two carbons are different from PB2, which indicates an additional double bond in this position.

b, 1H NMR assignments of 15 protons on PB2, 13 protons on EC-P1, and 10 hydroxy groups on PB2 and EC-P1. The black numbers in two compounds showed the similar PPM values of protons and hydroxy groups on PB2 and EC-P1, which demonstrates these protons and hydroxy groups on these two compounds are the same. The red numbers in EC-P1 showed one hydroxy group is different from PB2, which indicates this position is different from PB2. In addition to hydroxy group, 2 protons on EC-P1 are not detected, which indicates an additional double bond is formed on this position.

2.5 Evidence of polymerization mechanisms and higher oligomeric papanridins

We hypothesized that FP polymerized papanridin via carbocation formed in the reactions. FP catalyzed epicatechin to flav-2-en-3-ol carbocation, and then epicatechin nucleophilically attacked the carbocation to form papanridin ECII A and B (Fig. 6 a). We also hypothesized that the flav-2-en-3-ol carbocation could be nucleophilically trapped by phloroglucinol to form flav-2-en-3-ol-phloroglucinol (MW, 412 Dalton) (Fig. 6a). To test these hypotheses, phloroglucinol was added to the incubation including epicatechin, FP, and H_2O_2. TIC showed that a bigger peak with a 411 $[m/z]^-$ overlaid with EC-P1 was formed from the reaction, but it was not observed from controls (Fig. 6b). Further EIC analysis showed this 411 $[m/z]^-$ compound from the incubations, but did not show this compound in the control incubations with phloroglucinol (Fig. 6c). In addition, papanridin ECII-A and -B and two trimers were produced in reactions with or without phloroglucinol (Fig. 6c). To characterize this 411 $[m/z]^-$ compound, MS/MS analysis revealed two characteristic fragments, 125 from phloroglucinol and 285 likely from flav-2-en-3-ol (Fig. 6d). This feature was substantiated by MS/MS analysis of epicatechin-phloroglucinol as a positive control, which produced 125 from phloroglucinol and 287 from epicatechin (Fig. 6d). To further demonstrate the presence of flav-2-en-3-ol carbocation, we used quercetin and taxifolin as controls, because these two flavonoids with a carbonyl group at C4 and cannot form carbocation at C4. We incubated FP, quercetin or taxifolin, and phloroglucinol. TIC analysis did not show the formation of quercetin-phloroglucinol and taxifolin-phloroglucinol from the reactions (Supplementary Fig. S17). These data demonstrated the presence of flav-2-en-3-ol carbocation that was formed in the enzymatic reaction of FP. Moreover, we hypothesized that trimers with a MW 698 could be formed from the phloroglucinol, FP and epicatechin incubations. EIC analysis revealed five 697 $[m/z]^-$ peaks (Supplementary Fig. S18a). MS/MS

analysis showed that the CID of 697 generated 571, 409, and 287 fragments, which corresponded to two flav-2-en-3-ol carbocation molecules (574 - 4), flav-2-en-3-ol→phloroglucinol (411-2), and flav-2-en-3-ol, respectively (Supplementary Fig. S18b). Therefore, these compounds are trimeric flav-2-en-3-ol→flav-2-en-3-ol→phloroglucinol (Supplementary Fig. s18c). Higher oligomers were observed in the enzymatic reactions of FP H_2O_2, and epicatechin (Supplementary Figs. S9b and S14a and Fig. 6c). EIC analysis showed that in addition to papanridin ECII-A and -B, two trimers, two tetramers, and one pentamer were featured with an 861 $[m/z]^-$, an 1147 $[m/z]^-$, and a 1433 $[m/z]^-$, respectively, however, these compounds were not observed from all control incubations (Supplementary Fig. S19a). MS/MS analysis further showed that these new compounds were trimeric flav-2-en-3-ol→ flav-2-en-3-ol epicatechin (MW, 862 Dalton), tetrameric flav-2-en-3-ol→flav-2-en-3-ol→flav-2-en-3-ol→epicatechin (MW, 1148 Dalton), and pentameric flav-2-en-3-ol→flav-2-en-3-ol→ flav-2-en-3-ol→flav-2-en-3-ol→epicatechin (MW, 1434 Dalton) (Supplementary Fig. S19b).

Figure 6 Evidence for flav-2-en-3-ol carbocation in enzymatic reactions
Enzymatic reactions consisted of enzyme or BSA (control), epicatechin, hydrogen peroxide, and phloroglucinol added or not added in pH6 buffer.

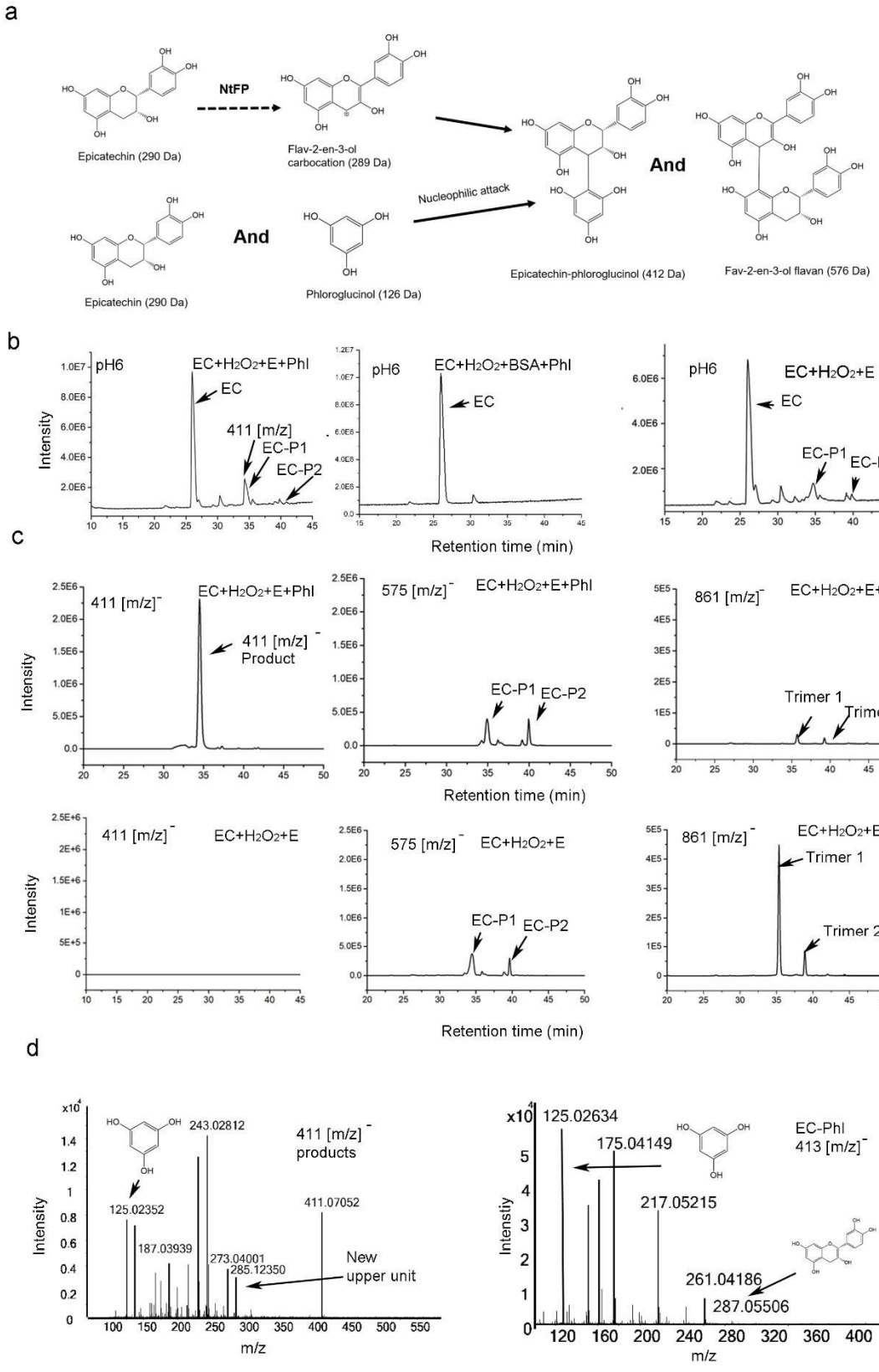
a, a scheme presents a hypothesis that flav-2-en-3-ol carbocation produced in the enzymatic reactions; nucleophilic compounds in the reactions can attack the carbocation at C4 to form dimeric or oligomeric compounds; and two nucleophilic compounds can competitively attack the carbocation.

b, TIC profiles showed that the incubation of enzyme, epicatechin, hydrogen peroxide, and phloroglucinol produced EC-P1, EC-P2, and a new compound with 411 $[m/z]^-$ (left) overlaid with EC-P1, but the incubation of BSA, epicatechin, hydrogen peroxide, and phloroglucinol did not produce these compounds (middle), and the incubation of enzyme, epicatechin, and hydrogen peroxide produced EC-P1 and EC-P2 but no new compounds with a 411 $[m/z]^-$ (right).

d, EIC profiles demonstrated the incubation of enzyme, epicatechin, hydrogen peroxide, and phloroglucinol produced the new 411$[m/z]^-$ compound (left), EC-P1 and EC-P2 with a 575 $[m/z]^-$ (middle), and two trimers with an 861 $[m/z]^-$ value (right).

e, EIC profiles demonstrated that the incubation of enzyme, epicatechin, and hydrogen peroxide did not produce the new 411$[m/z]^-$ compound (left), produced EC-P1 and EC-P2 (middle) and two trimers (right).

f, the new 411 $[m/z]^-$ compound is not epicatechin-phloroglucinol. MS/MS profiles showed the different main fragments between the 411 $[m/z]^-$ compound and epicatechin-phloroglucinol. Both have the 125.02 fragment, which is corresponding to phloroglucinol. The 411 $[m/z]^-$ compound has a main fragment of 285, while epicatechin-phloroglucinol has a main fragment of 287. Therefore, 411 was fragmented to 125 and 285 and epicatechin-phloroglucinol was fragmented to 125 and 287, indicating the presence of a new carbocation rather than epicatechin carbocation formed in the enzymatic reactions.


a
Epicatechin (290 Da)
NtFP
Flav-2-en-3-ol carbocation (289 Da)
Epicatechin (290 Da)
And
Phloroglucinol (126 Da)
Nucleophilic attack
Epicatechin-phloroglucinol (412 Da)
And
Fav-2-en-3-ol flavan (576 Da)
b
pH6
EC+H2O2+E+Phl
EC
411 [m/z]
EC-P1
EC-P2
pH6
EC+H2O2+BSA+Phl
EC
pH6
EC+H2O2+E
EC
EC-P1
EC-P2
Intensity
Retention time (min)
c
411 [m/z]-
EC+H2O2+E+Phl
411 [m/z]- Product
575 [m/z]-
EC+H2O2+E+Phl
EC-P1
EC-P2
861 [m/z]-
EC+H2O2+E+Phl
Trimer 1
Trimer 2
411 [m/z]-
EC+H2O2+E
575 [m/z]-
EC+H2O2+E
EC-P1
EC-P2
861 [m/z]-
EC+H2O2+E
Trimer 1
Trimer 2
Intensity
Retention time (min)
d
411 [m/z]- products
243.02812
125.02352
187.03939
273.04001
285.12350
411.07052
New upper unit
Intensity
m/z
125.02634
175.04149
217.05215
261.04186
287.05506
EC-Phl 413 [m/z]-
Intensity
m/z

2.5 Reduction of dimeric papanridins in fp mutant of Arabidopsis thaliana

To genetically support the biosynthesis of papanridins in plants, we blasted FP homologs in the Arabidopsis genome and identified one homolog, namely AtFP here. We further identified one *atfp* mutant. On the one hand, LC-MS/MS based profiling detected papanridin ECII-A and –B in seed extracts of wild-type *A. thaliana* but not in those of *atfp* mutants (Fig. 7a and Supplementary Fig. S20). On the other hand, the level of EC was higher in the mutant than in wild-type ones (Supplementary Fig. S21a-b). These results demonstrate that the polymerization pathway of papanridins in plants.

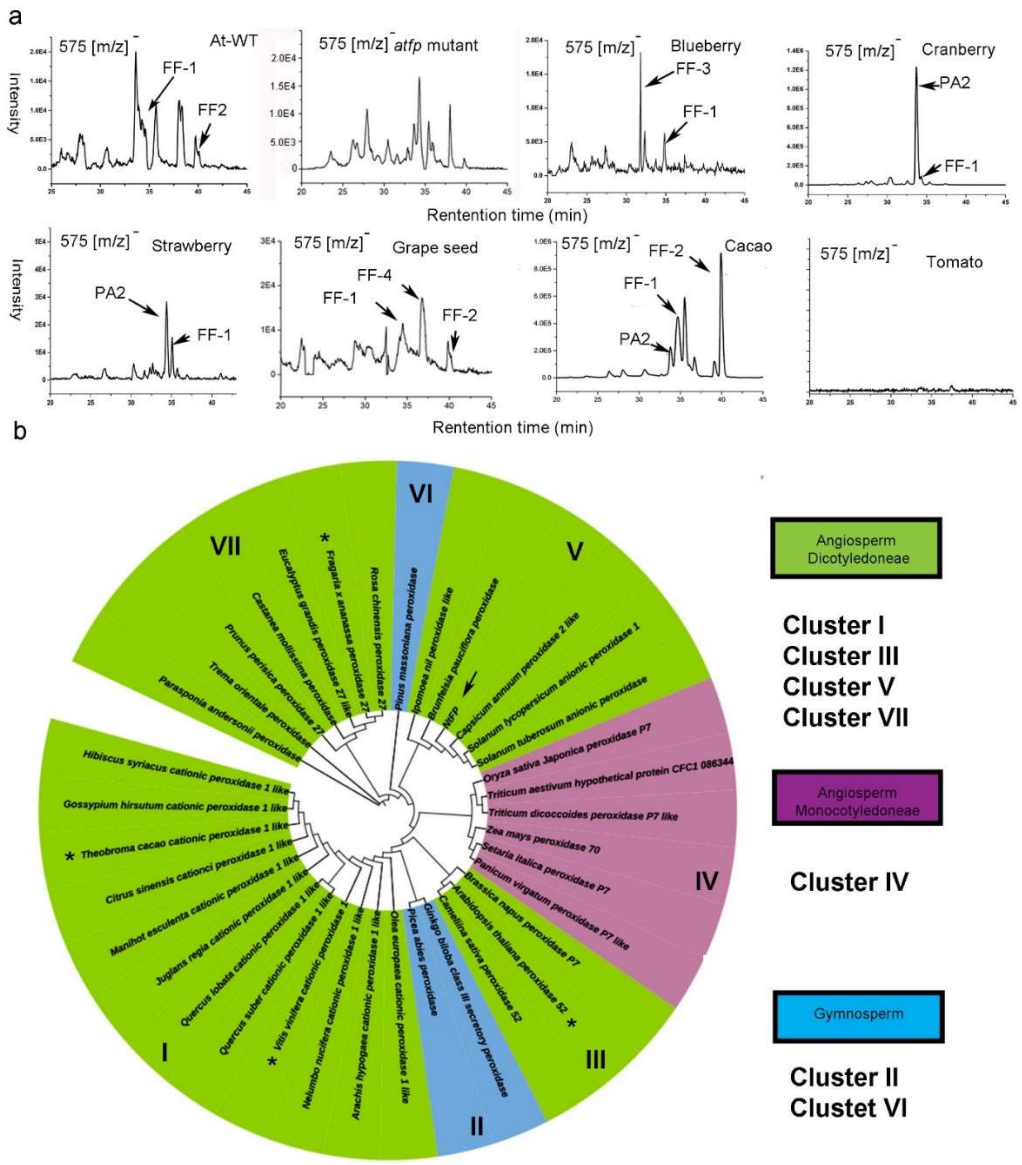


Figure 7 Prevalence of dimeric flav-2-en-flavan (FF) compounds in PA-producing tissues of plants. Mature seeds of *Arabidopsis thaliana* and the mutant *atfp*, berries of blueberry, cranberry, strawberry, and tomato, grape seeds, and cacao seeds were analyzed with LC-MS/MS.

a, EIC profiling detected FF1 and FF2 in seeds of Arabidopsis, FF1 and FF3 in blue berry, FF1 in cranberry and strawberry, FF1, FF2, and FF3 in grape seeds, FF1 and FF2 in cacao seeds, but could not detect any of these compounds in seeds of *atfp* mutant and berry of tomato. In addition, PA2 was detected in cranberry, strawberry, and cacao seeds.

b, an unrooted phylogenetic tree developed from amino acid sequences of FP and 22 homologs across gymnosperm and angiosperm. The numbers on branches are approximate likelihood ratio test values that are used to provide statistical support for the branches of the tree.

2.6 Papanridins in flavan-3-ol-rich plants and phylogenetic analysis

LC-MS/MS profiling was performed to analyze papanridinds in flavan-3-ol and PA-rich plants to understand if this polymerization pathway is prevalent in the plant kingdom. Plant tissues analyzed included berries of blueberry, cranberry, and strawberry, muscadine grape seeds, and cacao seeds. In addition, tomato berry lacking flavan-3-ols was used as a negative control. The resulting data showed that blueberry produced papanridin ECII-A and CAII-A, cranberry and strawberry had papanridin ECII-A, muscadine seeds produced papanridin ECII-A and –B, and CAII-D, and cacao seeds produced papanridin ECII-A and –B. However, tomato did not produce these compounds (Fig. 7a). Meanwhile, A-type procyanidin A2 was identified in cranberry, strawberry, and cacao seeds (Fig. 7a). EC was produced in blueberry, cranberry, strawberry, muscadine seeds, and cacao seeds, CA was produced in blueberry, cranberry, and muscadine seeds, while two were produced in tomato berry (Supplementary Fig. S21). These data indicate the polymerization pathway in these flavan-3-ol and PA-rich plants.

We further blasted FP homologs in plant genomes curated in the GenBank at NCBI. This search obtained 37 homologs across gymnosperms and angiosperms. Included were homologs from A. thaliana, cacao, bunch grape (*V. vinifera*), and strawberry, which produces dimeric papanridins. Due to the unavailability of genomic sequences for blueberry and cranberry, homologs were not obtained for these two plants. The 37 sequences were used to establish an unrooted phylogenetic tree (Fig. 7b), which was featured by seven clusters, I-VII. It was interesting that three gymnosperms were grouped into two clusters. Six monocotyledonous plants were grouped into one cluster. Dicotyledonous plants were grouped into four clusters. Cacao and grape FPs were grouped to cluster I. Arabidopsis, tobacco, and strawberry FPs were grouped into cluster III, V,

and VII. Although other homologs remain for studies, this phylogenetic analysis indicates that FP is prevalent in the plant kingdom.

3. Discussion

Our findings unearth papanridin, a novel group of oligomeric or polymeric flavonoids in the plant kingdom. Plant flavonoids are the largest group of plant polyphenols [23, 67-69]. By 2000, more than 5000 flavonoids were reported to be identified from the plant kingdom [28, 30, 31, 70]. In the past decade, as more plants were investigated, more than 8000 new flavonoids were discovered from the plant kingdom to increase the total number to be approximately 13000 compounds [23]. Given that all known flavonoids were identified from a limited number of plants in the past [3-5, 71-73], most of approximately 420,000 plant species [13, 74] are open for investigations to identify new flavonoids. Based on the structures and biosynthesis, flavonoids are classified to 13 main groups, chalcone, flavanones, flavones, aurones, dihydroflavonols, flavonols, flavan-3,4-diols, anthocyanins, flavanols, proanthocyanidins, isoflavonoids, bioflavonoids, and neoflavonoids [22, 23, 75-78]. Although thousands of new flavonoids were found from plants in the past 20 years [23, 69], most of them structurally fell in the 13 groups. In these groups, PAs are the only oligomeric or polymeric flavonoids that are synthesized from flavanols via an interflavan bond between starter and extension units [24]. Our findings show that papanridin is a novel group of oligomeric or polymeric flavonoids with these chemical features. 1) Papanridin is polymerized from flavanols as substrate but structurally characterized by flavanols as the starter unit and flav-2-en-3-ols extension unit (Figs. 5 and 8). 2) Like PAs, the bottom and upper units are linked by a interflavan bond formed at C8 and C4. 3) Unlike PAs,

dimeric papanridin ECII-A and –B, papanridin CAII-A and –B, trimers, and tetramers are yellowish and have a characteristic spectrum at 388 nm (Supplementary Fig. S14a and d). 4) Unlike PAs, neither can phloroglucinol in the acidic methanol condition cleave papanridin, nor can butanol-HCl boiling cleave them to produce anthocyanidins (Fig. 4 g-j). 5) papanridin is weak soluble in water, unstable, and can be oxidized to quinones. Furthermore, papanridin is prevalent in flavan-3-ol rich plants (Fig. 7). What functions papanridins have in the plants is unclear. Based on metabolic profiling of *FP* transgenic PAP1-BAN plants and *atfp* mutant plants, one hypothesis is that papanridin might be associated with PAs (Fig. 8a). As FP was overexpressed in PAP1-BAN plants, the contents of PB2 and total PAs was significantly increased (Supplementary Fig. S8), while the content of EC was significantly decreased (Fig. 3h and j). Since this is the first report, it can be anticipated that a great number of studies are necessary to understand the functions of papanridin in plants.

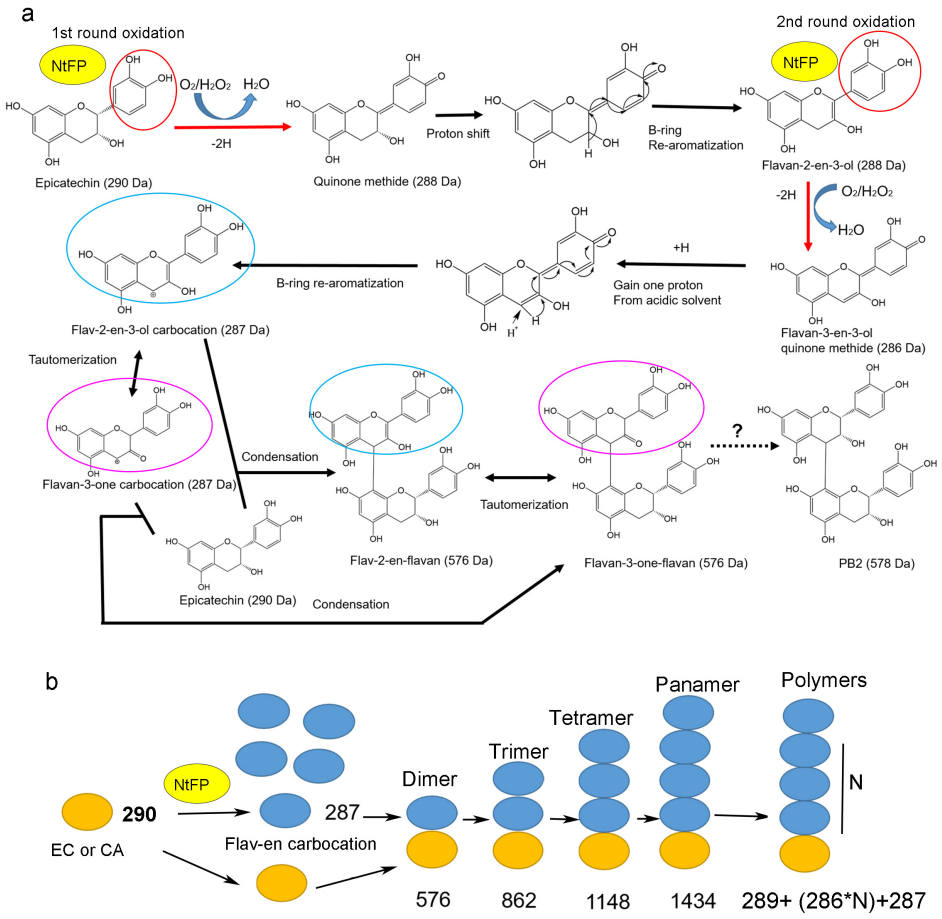


Figure 8 Hypothetic mechanisms proposed for the formation of dimeric and oligomeric flaven-flavans catalyzed by FP.

a, a pathway is proposed to start with the FP-catalyzed oxidation of epicatechin to form flav-2-en-3-ol carbocation and then dimeric FFs. First, FP oxidizes epicatechin to quinone methide, which undergoes spontaneous proton shit and B-ring re-aromatization to form flav-2-en-3-ol. Second, FP oxidizes flav-2-en-3-to its quinone methide, which undergoes addition of one H+ from a weak acidic condition and then B-ring re-aromatization to form flav-2-en-ol carbocation. Then, available epicatechin molecules attack the C4 of the carbocation to form dimeric flav-2-en-3-ol→epicatechin, which further tautomerizes to flavan-3-one→epicatechin and flav-3-en-3-ol→epicatechin. Although what functions dimeric FFs have is unknown, it can be speculated that FF might be reduced to produce PAs.

b, a carton is created to show the polymerization mechanism by which in the oxygen condition, plant cells use FP to biosynthesize trimers and higher degree oligomers. Once one dimeric flaven-flavan forms, it can nucleophilically attack flav-2-en-3-ol carbocation to form one trimer. Further sequential addition of flav-2-en-3-ol carbocation produces higher degree oligomers.

Our findings demonstrate that FP catalyzes the polymerization of papanridins from flavanols. Although flavanols are the substrates, the extension units of papanridins are flav-2-en-3-ols instead of flavanols. The trapping experiments with nucleophilic phloroglucinol demonstrated the presence of flav-2-en-3-ol carbocation in the catalytic reactions of FP and EC (Fig. 6). Although the mechanism by which this carbocation is formed from EC in the reactions remains for investigations, we hypothesize two oxidation steps by which FP oxidizes EC to flav-2-en-3-ol carbocation (Fig. 8a). The first oxidation step is that in the presence of oxygen, FP oxidizes EC to quinone methide. The proton at C3 shifts to the B-ring via C2 leading to B-ring re-aromatization that forms flav-2-en-3-ol. The second oxidation step is that FP oxidizes flav-2-en-3-ol to form flav-3-en-3-ol quinone methide, to which one proton from the acidic condition was added. This proton addition leads to the second time of B-ring re-aromatization to form flav-2-en-3-ol carbocation with a positive charge at C4. The EC then nucleophilically attacks flav-2-en-3-ol carbocation leading to a condensation via the formation an interflavan bond. Accordingly, a new flav-2-en-3-ol→EC with a MW 576 Dalton is formed. Given that two dimeric compounds are obtained with the same 576 Dalton, we further hypothesize that like B-type proanthocyanidins, there are two types of stereo configurations, flav-2-en-3-ol-(4α→8)-EC and flav-2-en-3-ol-(4β→8)-EC. To provide evidence that flav-2-ev-3-ol was produced in the incubations, we used EIC to analyze other potential metabolites from FP catalysis. The resulting data showed that four from EC and three from CA had a 287 $[m/z]^-$ (Supplementary Fig. S22a), which these compounds were not detected from negative controls. Given that only one substrate, FP, and H_2O_2 were used in the reactions, these compounds were formed from substrates catalyzed by FP. Besides, give that we used negative mode for ionization, the 287 $[m/z]^-$ value must result from non-charged compounds with a MW 288 Dalton, which is the MW of flav-2-en-

3-ol. Why were more than one 287 $[m/z]^-$ peaks from the catalysis? Based our previous report that flavan-3-one and flav-3-en-3-ol are two tautomers from the spontaneous tautomerization of flav-2-en-3-ol in the catalytic formation of EC [50, 52], we hypothesize that additional peaks result from the tautomerization. A comparison among the 287 $[m/z]^-$ peaks (Supplementary Fig. S22a), the peak 2 (P2) were obtained from both EC and CA. This indicates that the same compound was produced from EC and CA isomers. Based on this fact, P2 does not have a stereo configuration between C2 and C3. Therefore, the covalent bond between C2 and C3 is double. In another words, P2 is flav-2-en-3-ol (Supplementary Fig. S22b). In addition, we propose that three other peaks from EC are 2R-flav-3-en-3-ol, 2S-flav-3-en-3-ol, and 2R-flavan-3-one and two other peaks from CA are 2S-flavan-3-one and 2S-flav-3-en-3-ol (Supplementary Fig. S22b). We hypothesize that flav-2-en-3-ol can spontaneously tautomerizes into its tautomers, flavan-3-one and flav-3-en-3-ol. The rationale is that these tautomers exists in plants. Ours and other recent reports demonstrated that flav-2-en-3-ol was involved in the formation of PAs [48, 51, 52]. A flav-2-en-3-O-glycoside was reported to associate with the formation of cyanins in black soybean [79, 80]. Flav-3-en-3-ol glycosides were also isolated from a medicinal plant [81]. Based on our metabolic profiling of dimeric papanridins in Arabidopsis seeds and five flavanol-rich plants and FP homolog analysis (Fig. 7), we propose a model to show the polymerization of panpanridins (Fig. 8b). In this model, additional flav-2-en-3-ol carbocations are added to the upper unit of dimeric papanridins to form oligomers and then polymers. The evidence was trimers, tetramers, and pentamers that were observed in the enzymatic reactions of FP (Supplementary Fig. S19a).

4. Materials and methods

4.1 Plant materials and chemical reagents

Red/purple PAP1 transgenic 6R, wild-type P3 cell lines, *Nicotiana tabacum* var. Xanthi, lightly red PAP1-ANR (PAP1-BAN) tobacco, *Arabidopsis thaliana*, blue berries, cranberry, muscadine grape berries, raspberry, cacao seeds, and tomato fruit were used in this study. Red/purple PAP1 transgenic 6R and wild-type P3 cell lines were established from red PAP1 tobacco and wild-type tobacco plants, respectively [59]. Red cells constitutively overexpress Arabidopsis *PAP1* gene and produces high production of anthocyanins, while P3 cells do not. PAP1-ANR tobacco is a homozygous progeny of *PAP1* and *ANR* transgenic plant, which produces anthocyanins, flavanols, and proanthocyanidins in leaves and flowers [63]. Plants were grown in the greenhouse supplemented with natural light. Seeds and fruits were purchased from grocery stores.

Authentic (+)-catechin, (±)-catechin, (-)-epicatechin, (-)-epigallocatechin, procyanidin A2, procyanidin B2, phloroglucinol, and catechol were purchased from Sigma (St. Luis, USA). These compounds were dissolved in methanol to make 10 mg/ml stocking solutions, which were stored in a -20°C freezer. Organic solvents used are specified in each different experiment.

4.2 Callus culture and cell suspension culture

The medium, subculture of 6R and P3 cell lines, and tissue culture conditions were as reported previously [59]. Since 2008, two cell lines have been subcultured for studying anthocyanin biosynthesis and plant flavonoids. Briefly, the medium was composed of MS medium (pH 5.7)

containing 0.25 mg L^{-1} 2,4-dichlorophenoxyacetic acid, 0.10 mg L^{-1} kinetin, 30 g L^{-1} sucrose and 0.8% (w/v) phytoagar. Calli were aseptically sub-cultured with a 15-day interval in a growth chamber facilitated with 22 °C and a 16 h/8 h light/dark photoperiod (Supplementary Fig. S1) Cell suspension cultures were established from 6R and P3 cells. The liquid medium used was made of the subculture medium by the removal of agar and a 50 ml aliquot was contained in a 250 ml Erlenmeyer flask (Supplementary Fig. S1). Each flask was inoculated with 5 gram calli and then placed on a rotary shaker with a speed of 100 rpm in a growth chamber with the same growth condition as described above. After 15 days, suspension cells were filtered with autoclaved 10 and 50 μm sieves. Cells harvested between the two types of sieves were subcultured in 50 ml liquid medium. After 3 times of subculture, appropriate cell suspension cultures were established. Suspension cultures were developed for red 6R and wild-type P3 cells as comparison. To obtain sufficient amount of cells, 10 g (fresh weight, FW) suspension cells after filtering were cultured in 1-liter liquid medium contained in a 6-liter E-flask and cultured for 15 days (Supplementary Fig. S1). Cells were filtered (Supplementary Fig. S1), collected to liquid nitrogen, and then stored in an 80ºC freezer until the use for the following experiments. The remaining liquid medium was transferred to 500 ml tubes for 10 min of centrifugation at 8,000 ×g and 4ºC. The supernatant (Supplementary Fig. S1) was concentrated to then stored in 4ºC for the following experiments.

4.3 Protein Extraction

Frozen 6R and P3 cells were ground into fine powder in liquid nitrogen. The protein extraction buffer prepared was composed of 0.1 M Tris-Cl (pH 7.0) and 10 mM EDTA. Thirty grams of powder were suspended in 150 mL of the protein extraction buffer in a 500 ml tube and vortexed

for one min to mix completely. The mixture was sonicated for 30 min and then shaken on a rotary shaker at 100 rpm for 20 min at 4 °C. The suspension mixture was centrifuged for 20 min at 10,000 rpm at 4 °C. The supernatant was transferred to a clean beaker and then filtered through a Whatman filter paper. The clear protein extraction was pipetted into a Millipore 15 mL 10K filter, which was centrifuged. Accordingly, 150 protein extract was concentrated to obtain 15 ml for SDS-PAGE (10%) and concentration measurement using Coomassie Brilliant Blue, and then stored at a -20°C freezer for catalytic analysis and protein purification described below.

4.4 In vitro enzymatic assays and thin layer chromatography (TLC) analysis

A screening assay was developed to test the catalysis of crude protein extracts from cells and liquid medium. Enzymatic reactions were carried out in 15 ml of clean glass tubes containing a 5 ml of reaction reagents. The enzymatic reaction consisted of 15-115 µg of crude proteins (0.75 ml), 0.187 µM CuSO4 (6 µl of 156 µM or 0.025 mg/ml), 250-750 µg catechin or epicatechin (25-75 µl), and 4.169-4.214 ml of 1 M pH 7 oxygenated phosphate buffer. The oxygenation of the reaction buffer was performed by strong vortex for 5 min prior to the use. The reaction was incubated at room temperature for 2-10 min and stopped by addition of 5 ml ethyl acetate (EA) and strong vortex for 1 min. Tubes were centrifuged for 5 min at 3000 rpm under 4ºC. The upper EA phase (4.5 ml) was transferred into a clean glass tube. The EA extraction step was repeated twice and pooled into the same tube. EA was then removed with a nitrogen gas flow (purchased from Machine & Welding Supply Company) at room temperature. The residue was dissolved in 450 µl of MS grade methanol for TLC.

4.5 In vitro enzymatic assays and thin layer chromatography (TLC) analysis for screening native proteins

Methods were developed to screen native proteins with catalytic activity of crude native proteins. Substrates screened included (-)-epicatechin, (-)-epigallocatechin, phloroglucinol, catechol, or (+)-catechin. Crude protein extracts tested include 6R and P3 proteins. Buffers were oxygenated by strongly vortexing for 5 min at room temperature immediately prior to enzymatic incubations. Enzymatic reactions in 1.5-ml Eppendorf polyethylene tube were carried out in 500 µl volume, which was composed of 475 µl of 0.02 M Tris-Cl (pH 7.0 or pH 8.0) buffer, or 0.05 M phosphate-citric buffer (pH 6.0), 15 µl protein extracts (10 µg) and 5 µl substrate (0.5 mg). Four types of control reactions were performed, including substrate absence, enzyme absence, BSA as the substitute of enzyme, and 1-10 min boiled crude protein extracts. All reactions were incubated for 30 min at room temperature and stopped by adding one ml of ethyl acetate (EA). The tubes were then vortexed and centrifuged at 5,000 rpm for 5 min. The EA phase was transferred to new 1.5 ml tube and evaporated in a speed vacuum. The residues were diluted in 200 µl of methanol for TLC and HPLC analyses.

Thin layer chromatography was used to examine enzymatic products. Ten µl of methanol extraction was loaded onto an aluminum-backed silica Kieselgel 60 F254 TLC sheet (0.2 mm layer thickness, EM Sciences, USA). Authentic standards of catechin, epicatechin, procyanidin B1, and procyanidin B2 (0.1- 1 µg) were loaded onto the same TLC sheet as positive control. Samples were separated by a solvent consisting of formic acid: water: ethyl acetate (1:1:18, v/v). Compounds were visualized by spraying 0.1 % DMACA dissolved in ethanol: HCl (50%: 3M) on TLC plates.

4.6 Native enzyme purification

Based on catalytic screening results, the crude protein extracts from 6R cell line were subjected to three steps of purifications to isolate the active enzyme. First, based on the screening result that catalytic proteins were thermally stable, the crude protein extracts were boiled for one min and then centrifuged at 12,500 rpm for 10 min to remove all thermally instable proteins in the pellets. The resulting supernatant included active enzymes. Second, the boiled protein mixtures were subjected to a purification via an anion exchanger chromatography. Column for ion-exchange chromatography was prepared by loading 15 ml DEAE Sephadex A-20-120 in 25 ml syringes, washed with 15 ml ddH2O 5 times, and then equilibrated with 0.02 M pH 7.0 Tris-Cl buffer. The supernatant was loaded onto the column, which was washed with 0.02 M pH 7.0 Tris-Cl. After three volumes of buffer as that of column, proteins were eluted with three types three types of elution buffers, 0.02 M pH 7.0 Tris-Cl buffer added with 0 M, 0.1 M, and 0.2 M. For each buffer, three volumes as the column size were used to elute the protein. Each flow-through fraction was collected for catalytic test as described above. The most active fractions were combined for further purification. Third, this step was gel chromatography (size-exclusion) purification. Columns were prepared by loading 15 ml Sephadex G-75 into a 25-ml syringe and then washed with 45 ml 0.02 M Tris-Cl pH 7.0 buffer. The active fractions were loaded onto the G-75 column and then eluted with 0.02 M Tris-Cl pH 7.0 buffer. The flow-through were collected every 30 seconds to obtain different fractions. All fractions were analyzed by a 10% SDS-PAGE. All single band fractions with the same size were combined and concentrated to 5 ml via a Millipore 15 mL 10K filter as described above. Each single purified enzyme was quantified by Bio-Rad Bradford protein assay kit and analyzed via a 10% SDS-PAGE to examine the relative purity. The bands from SDS-PAGE were cut for the protein sequencing at

Duke Center for Genomics and Computational Biology (https://genome.duke.edu/cores-and-services/proteomics-and-metabolomics). Sequencing steps followed the center's protocol. Finally, all purified proteins were stored at -80 °C for next experiments. Native-polyacrylamide gel electrophoresis for *in Situ* assay of enzymatic reaction

Native-polyacrylamide gel electrophoresis (Native-PAGE) was completed to test enzymatic reactions on gel. Native-PAG (10.0 %) was freshly prepared prior to electrophoresis. Separation gel was prepared using 2.33 mL of 30% acrylamide/bisacrylamide solvent, 1.75 mL of 4X Tris (pH 8.8), 2.85 mL of dH_2O, 70 uL of 10% (w/v) ammonium persulphate (APS), and 5 uL of tetramethylethylenediamine (TEMED). Stacking gel was prepared using 500 uL of 30% acrylamide/bisacrylamide solvent, 375 uL of 4X Tris (pH 6.8), 2.1 mL of ddH_2O, 30 uL of 10% APS and 4.0 uL of TEMED. The size of comb created 40 µL volume wells for sample loading. Native-PAGE was prepared using a caster (10 x 8 cm). Native-PAG was then installed in a Bio-Rad Mini Protean® Tetra Cell Electrophoresis apparatus (Bio-Rad, USA). Ten µg of protein sample was mixed with 10 µL of protein loading buffer, which was composed of 50% glycerol, 0.5 M Tris-Cl (pH:7.0), 0.15 uM bromophenol blue (Bio-Rad), and dH_2O. In addition, the bovine serum albumin (BSA) and proteinase K digested protein was added as the negative control.

The resulting mixture was gently mixed and then loaded to wells of Native-PAGE. The polyacrylamide gel and 1-liter of 1 X Tris-Glycine-running buffer including 3.03 g of Tris and 14 g of glycine were placed in an electrophoresis chamber. In addition, 10 µg of Bio-Rad Precision Plus Protein™ ladder was loaded a molecular weight standard. The electrodes of the electrophoresis apparatus were wired to a power source. The electrophoresis apparatus was placed

in a container and surrounded with ice to avoid any protein degradation due to heating. The voltage of electrophoresis was set at 70 Volt for 30 min and then increased to 120 Volt for 90 min. After the completion of electrophoresis, the gel was rinsed using dH_2O and 0.02 M Tris-Cl buffer (pH:7.0) and then placed into in a glass tank for incubation with substrate. The gel was suspended in 0.02 M Tris-Cl buffer pH 7.0 containing 1 mg/ml (+)-catechin or (-)-epicatechin. The incubation was achieved by shaking the gel smoothly. Color changes on the gel were recorded at 0, 30 and 60 minutes after incubation started. The appearance of a brownish-yellowish color indicated protein bands with activity. Given that the protein catalyzed the formation of oligomeric compounds, herein, this enzyme was named to be flavanol polymerase (FP)

4.7 Cloning of FP cDNA from 6R cells

First, the oligomeric peptides obtained from protein sequencing were used to blast against tobacco sequences curated in the GenBank curated by NCBI. All oligomeric peptides hit one tobacco peroxidase gene (Supplementary Fig. S4). Based on the sequence, a pair of primers was designed to clone this gene from our red 6R cells. The forward primer was 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTTTTCGTTTGAGTCA-3' which contains attB1 recombinant sequence and ATG start codon. The reverse primer was 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGTGGTGGTGGTGGTGGTGCATAG AAGCCACAGAGCT-3' which contains attB2 recombinant sequence, 6-HIS tag encodes sequence, and TCA stop codon.

Total RNA were extracted from 6R calli using a TRI Reagent™ Solution (ThermoFisher, Catalog number: AM9738) and treated with DNase to remove genomic DNA as we reported

recently [61]. One μg of DNA-free RNA was reversely transcribed into the 1st strand cDNA in a 20 μl reaction volume using High capacity cDNA Reverse Transcription Kit (Applied Biosystems, USA) by following the manufacturer's protocol. Two μl of the 1st strand cDNA product was used as the template in PCR to amplify a cDNA. The thermal cycle consisted of 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 60 °C for 1 min, and 72°C for 30 s. The final extension step was 10 min at 72 °C. The amplified cDNA was then cloned into the gateway pDONR221 vector by using the Gateway BP Clonase II Enzyme Mix (Invitrogen, USA) following the manufacturer's protocol. The ligated products were introduced into competent *E. coli* ccdB Survival 2 $T1^R$ strain via a common heating shock transformation. Positive colonies were selected and inoculated to LB medium to purify the pDONR221-FP construct.

4.8 Expression of recombinant cDNA in E. coli and purification

The recombinant pDONR221-FP plasmid was used as the template to clone sequence to pDEST15 protein expression vector by using Gateway LR Clonase II Enzyme Mix (Invitrogen, USA) following the manufacturer's protocol. The ligation product was introduced into competent BL21 (DE3) plysS *E. coli* strain, which was inoculated on LB medium supplemented with 200 mg/L ampicillin and 34 mg/L chloramphenicol. A positive colony was used to isolate the positive recombinant pDEST15-FP-His vector, in which a *GST* tag and a *His* tag were cloned to the 5'-end and the 3'-end of the FP cDNA, which could express a fused GST-FP-His protein.

A single positive colony was inoculated to 10 ml of LB broth supplemented with 200 mg/L ampicillin and 34 mg/L chloramphenicol in a 50 ml tube. The tube was placed in an incubator at 200 rpm 37 °C for 16 hrs at. Then 1 ml of suspension culture was inoculated into 100 ml of fresh

LB broth medium containing 200 mg/L ampicillin and 34 mg/L chloramphenicol in a 250 ml Erlenmeyer (E) flask. The E flask was placed on an incubator and shaken at 120 rpm at 37 °C until the absorbance value measured at 600 nm was 0.7. The culture temperature was reduced to 30 °C and isopropyl-1-thiol-beta-D-galactopyranoside (IPTG) was then added to the cultures to a final concentration of 1 mM. After 4 h of induction, the suspension cultures were harvested by centrifuging at 6,000 rpm for 5 min. the remaining pellets were stored at -80°C to purify the recombinant protein late on.

The *E. coli* pellets were completely suspended in 10 ml of denaturing extraction buffer (8M urea, 20 mM Tris-Cl pH 8.0, 10 mM imidazole) then lysis by 5 min of supersonic. The lysate mixture was centrifuged for 20 min at 12,000 rpm. The supernatant containing crude proteins was transferred to a new tube for protein extraction.

The purification of recombinant protein was carried out on a Ni-NTA column. One ml of Ni-NTA Sefinose resin (G-Biosciences, USA) was loaded onto a plastic column. After the resin was stacked tightly, the column was washed three times with ten volumes of denaturing extraction buffer. Then, 10 ml of crude protein extracts were loaded onto the top of the resin column and eluted by gravity at a flow rate of 0.5 ml/min. The column was washed three times using four times the volume of the denature extraction buffer to wash off non-specific column-binding proteins. Then the column was moved to a cold room and washed by three times ten-volume of refolding buffer (20 mM Tris-Cl pH 8.0, 10 mM imidazole) to allow the recombinant protein full refolded on the column. The refolded recombinant protein was then eluted with 3 ml of elution

buffer (20 mM Tris-Cl pH8.0, 250 mM imidazole) to obtain GST-NtFP-His recombinant protein. The purified protein was evaluated by electrophoresis on a 12% SDS-PAGE.

4.9 Enzymatic and kinetic assays using recombinant enzyme

The enzymatic assay of recombinant protein followed the same protocol as described above. In brief, the enzymatic reaction was composed of 475 µl of 0.02 M Tris-Cl (pH 7.0 or pH8.0) buffer, or 0.05 M phosphate-citric buffer (pH 6.0), 0.5 mg of substrates, and 10 µg of recombinant protein in a total volume of 500 µl in 1.5 ml tubes. For hydrogen peroxide tests, 0.5 µl of 30% hydrogen peroxide was added. Also, control reactions without substrates, without protein, and with BSA were performed. All reactions were incubated for 30 min at 37 °C and stopped by adding one ml of ethyl acetate (EA). The tubes were then vortexed and centrifuged at 5,000 rpm for 5 min. The EA phase was transferred to new 1.5 ml tubes and drought in a speed vacuum. The residues were dissolved in 200 µl of methanol for the LC-qTOF-MS/MS analysis.

To identify the optimum reaction time, each enzymatic reaction was composed of 475 µl of 0.02 M Tris-Cl (pH 7.0) buffer, 0.5 mg (-)-epicatechin, and 10 µg of recombinant protein. The reactions were stopped at 30 min, 45 min, 60 min, 90 min, 120 min, and 180 min. The EA extracts of each reaction were detected by LC-qTOF-MS/MS analysis as described above. The velocity of reactions was calculated with the (-)-epicatechin consumption.

Several different buffer systems were tested to optimize a pH value. Every single reaction was composed of 475 µl phosphate sodium buffer (0.05 M, pH 6.2 to 8.0), or Tris-Cl buffer (pH 7.0

to 9.5), 0.5 mg (-)-epicatechin, and 10 µg of recombinant protein. All reactions were performed at 37 °C for 30 min and products were analyzed by LC-qTOF-MS/MS.

To optimize temperature value, every enzymatic reaction has consisted of 475 µl of 0.02 M Tris-Cl (pH 7.0) buffer, 0.5 mg (-)-epicatechin, and 10 µg of recombinant protein. The reactions were performed at 35, 40, 45, 50, 60, 70, and 80 °C for 30 min and analyzed by LC-qTOF-MS/MS.

To calculate kinetic parameters, reactions were composed of 475 µl of 0.02 M Tris-Cl (pH 7.0) buffer, 10 µg of recombinant protein, and 0.312, 0.5, 0.625, 0.825, 1, 1.2, 1.4, or 1.7 µM of (-)-epicatechin. The reactions were performed at 37 °C for 30 min and analyzed by LC-qTOF-MS/MS.

4.10 Purification of enzymatic products

The enzymatic product of EP-P1 was purified from large volumes of enzymatic reactions. Enzymatic reactions were composed 5 mg crude protein extract of 6R cells, 100 mg (-)-epicatechin, 500 ml 0.05 M phosphate-citric buffer (pH 6.0), and 0.5 ml 30% hydrogen peroxide. The reaction was performed at room temperature for 30 min and stopped by added 1 L of EA. The EA phase was transferred to a new E-flask and drought completely on a rotatory evaporator. The residue was then dissolved in 2 ml EA and loaded on a TLC silica gel 60 plates (2 mm thickness, 20 cm × 20 cm, EMD Millipore, USA). The TLC plates were developed in water: formic acid: ethyl acetate (1:1:18), v/v). After full development, bright yellowish bands were cut down from the plates. The resulting silica gel pieces containing the yellowish EP-P1 were grounded to a fine powder and mixed with 10 ml ddH_2O. Then, 25 ml EA was added to the

mixture and vortexed for 1 min. The mixture was centrifuged at 3,000 rpm for 5 min and the EA phase was transferred to a new 50 ml tube.

The EA extract was dried with nitrogen gas stream. The remaining residues were dissolved in 1 ml methanol for further purification with HPLC on Shimadzu LCMS-2010 EV instrument. In brief, compounds were separated on an Advantage 300 C18 column (250 × 10 mm, 5µm, Thomason Liquid Chromatography, USA). The mobile phase solvents used for elution included 1% acetic acid in water (solvent A, HPLC grade acetic acid, HPLC grade water) and 100% acetonitrile (solvent B, HPLC grade). A gradient solvent system, which was developed to separate compounds, was composed of ratios of solvent A to B: 80:20 (0–5 min), 80:20 to 70:30 (5–10 min), 70:30 to 55:45 (10–20 min), 55:45 to 50:50 (20–30 min), 50:50 to 45:55 (30–35 min), 45:55 to 80:20 (35-36 min). After the last gradient step, the column was equilibrated and washed for 10 min with solvent ratio A: B 80:20. The flow rate was 0.6 ml/min and the injection volume was 50 µl. Chromatograms were detected and recorded at 280 nm and 380 nm. All elutes were collected to different tubes labelled with numbers. The same elutes collected from different separations were combined and extracted by 50 ml of EA. The EA phase was transferred to new tubes and dried with nitrogen gas stream. The remained residues were dissolved in methanol and then examined by LC-qTOF-MS/MS.

4.11 Butanol:HCl cleavage of enzymatic products

Butanol: HCl (50:50) boiling is a classic method to cleave and calculate PA. As described previously [63], we used this method to cleave enzymatic products. In addition, PB2 was used as a positive control. In brief, 5 µl PB2 (50 µg) or a purified product (50 µg EP-P1 or EP-P2) was

mixed with 950 µl butanol:HCl in a 1.5 ml tube. The mixture was boiled for 1 hr and then cooled down to room temperature. The boiled mixture was used to measure absorbance at 550 nm. After measurement, the samples were evaporated to remove butanol and water in a speed vacuum. The remaining residue was suspended in 200 µl methanol with 0.1% HCl for LC-qTOF-MS/MS assay.

4.12 Acid-catalytic degradation and phloroglucinol competition assay

Acid-catalytic and phloroglucinol-based degradation is a classic method to determine extension units of PAs. The reaction was composed of 450 µl of 0.1 N HCl in methanol, 50 mg/ml phloroglucinol, 10 mg/ml ascorbic acid, and 50 µl of 1 mg/ml purified products (EP-P1 or CA-P1) in a 1.5 ml tube. Then the reaction tubes were incubated at 50 °C for 20 min. After cooled to to room temperature, the reaction products were examined with LC-qTOF-MS/MS. In addition, 50 µl of 1 mg/ml procyanidin B2 and procyanidin A2 were used as a controls.

4.13 Competition assays between phloroglucinol and epicatechin

The phloroglucinol and epicatechin competition experiment was carried out in a 500 µl volume in a 1.5 ml tube. The reaction was composed of 475 µl 0.05 phosphate-citric buffer (pH6.0), 1 mg/ml (-)-epicatechin, 1 mg/ml phloroglucinol, 10 µg native protein or BSA, and 0.5 µl 30% hydrogen peroxide. In addition, taxifolin and quercetin were used to substitute epicatechin as controls. The tube was incubated at room temperature for 30 min and the reaction was stopped by added 1 ml of EA. The tubes were then vortexed and centrifuged at 5,000 rpm for 5 min. The EA phase was transferred to a new 1.5 ml tube and the dried in a speed vacuum. The residues were dissolved in 200 µl of methanol for HPLC-MS analysis.

4.14 Overexpression of NtFP in PAP1-ANR tobacco and PA assay

The stop codon of *NtFP* ORF was removed by PCR. The resulting fragment was cloned to Pdonr223 vector using Gateway BP Clonase II Enzyme Mix (Invitrogen, USA) by following the manufacturer's protocol. This cloning obtained a Pdonr223-NtFP vector. The purified Pdonr223-NtFP plasmid was used to clone NtFP into a binary vector PMDC84 with Gateway LR Clonase II Enzyme Mix (Invitrogen, USA) by following the manufacturer's protocol. This cloning generated a recombinant binary vectors, PMBC84-NtFP, in which NtFP was fused to the 5'-end of *GFP* and *NtPAC-GFP* was driven by 35S promoter. Both PMDC84-NtFP and PMDC84 vectors were introduced into *Agrobacterium tumefaciens* strain GV3101 for genetic transformation. The PAP1-ANR(BAN) tobacco [63] was used for genetic transformation. MS medium, plant hormones used, and all steps of transformation and selection of transgenic plants followed our protocols published previously [63]. The PAP1-BAN-NtFP and vector control transgenic plants were screened on media supplemented with 50 mg/L hygromycin and further confirmed with PCR and RT-PCR. Transgenic and control plants were grown in pot soil and placed side by side in the greenhouse to develop flowers. Seeds were collected from T0 plants and germinated on MS medium containing 50 mg/L hygromycin. Resistant T1 seedlings were grown in the greenhouse to develop flowers for flavan-3-ols and PA analysis as described above.

4.15 Subcellular localization of NtFP

The T1 transgenic seeds and vector control seeds were grown on MS medium containing 50 mg/L hygromycin for five days to develop young seedlings. Half of transgenic seedlings were pretreated with 30% sucrose for 10 min to create plasmolysis of seedlings. In addition, seedlings pretreated with BCECF fluorescent dye were used as positive control. The protocol followed a

classic protocol of plasmolysis reported in the literature. Five-day old wild type tobacco seedlings were placed in perfusion solution (0.2 mM $CaCl_2$, 50 mM glucose, 3μM BCECF ~AM, and 10 mM MES buffer, pH 6.0) for 30 min at room temperature, and then washed three times with 0.1 mM $CaCl_2$. The seedlings were then held on perfusion solution without BCECF~AM for 1 hour. All seedlings were exanimated under a confocal laser scanning microscope (Leica TCS SP2, Leica, Germany). The light wavelength was set at 488 nm to observe GFP signal.

4.16 Extraction of proanthocyanidin and flavan-3-ols

Frozen samples were ground into fine powder in liquid nitrogen. Two hundred milligrams of powder were weighed into a 1.5 ml tube and then suspended in 1 ml 70% acetone. The tube was vortexed 30 seconds, followed by centrifugation for 5 min at 5,000 rpm. The supernatant was pipetted into a new 1.5 ml tube. The supernatant was then dried in a speed vacuum until the small water phase left in the bottom. The remained water phase was added 500 μl water, followed by addition of 200 μl chloroform. The mixture was then vortexed and centrifuged at 10,000 rpm to gain the upper water and bottom chloroform phases. The chloroform was pipetted to a dispensable bottle to remove chlorophyll, lipids, and fatty acids. This step was repeated once. The water phase in the tube was added 1 ml EA and vortexed for 1 min. The tube was centrifuged at 10,000 rpm for 5 min. The EA phase was then transferred into a new 1.5 ml tube and dried completely in a speed vacuum. The residues were dissolved in 200 μl of methanol for HPLC-qTOF-MS/MS assay.

4.17 TLC assay

TLC was used to separate compounds of EA extracts. 10 µl samples were loaded on aluminum-backed silica Kieselgel 60 F_{254} TLC sheets (0.2 mm layer thickness; EM Sciences, Busking Ridge, NJ, USA) which were developed in water: formic acid: ethyl acetate (1:1:18), v/v), and then sprayed with 0.1% (w/v) dimethylaminocinnamaldehyde (DMACA), which was freshly prepared in 6 M HCl:ethanol (1:1,v/v), to visualize the monomeric flavan-3-ols and PA compounds. 10 µl of 1 mg/ml (+)-catechin, (+)-epicatechin, and procyanidin B2 (PB2) were load to the plates as references.

4.18 LC-qTOF-MS/MS assay

HPLC-qTOF-MS/MS analysis was performed to profile the enzymatic products based on our previous reports [66]. HPLC-qTOF-MS/MS analysis was carried out by using Agilent Technologies (Santa Clara, CA, USA) 6520 time-of-flight-MS/MS. The mobile phase solvents optimized include 1% acetic acid in water (solvent A: 1% HPLC grade acetic acid in LC-MS grade water) and 100% acetonitrile (solvent B) (LC-MS grade). A gradient elution system is optimized with the two mobile solvents to separate generate phenolics in an Elipes XDB-C18 analytical column (250 × 4.6 mm, 5 µM, 25 ◦C, Agilent). The gradient solvent system is composed of ratios of solvent A to B: 95:5 (0–5 min), 95:5 to 90:10 (5–10 min), 90:10 to 85:15 (10–15 min), 85:15 to 45:55 (15–45 min), 45:55 to 25:75 (45–50 min), 25:75 to 95:5 (50–60 min). After the last gradient step, the column is equilibrated and washed for 10 min with solvents A: B (95:5). The flow rate is 0.4 mL/min. The injection volume of the samples is set up at 5.0 µl. The drying gas flow and the nebulizer pressure are set at 12 L/min and 50 psi, respectively. Metabolites are ionized with negative mode. The mass spectra are scanned from 100 to 3000

m/z. The acquisition rate is three spectra per second. Other MS conditions optimized include fragmentor: 150 V, skimmer: 65 V, OCT 1 RF Vpp: 750 V, and collision energy: 30. The authentic (-)-epicatechin, (+)-catechin, procyanidin B2, procyanidin A2 were used as the references.

4.19 Phylogeny analysis

The 34 homologues of NtFP amino acid sequences were obtained from Genebank by BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). A multiple Alignment were created by online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). The multiple alignment was used to generate a phylogenic tree by using NGphylogeny.fr tool (https://ngphylogeny.fr/).

REFERENCES

1. Iasona, G., *The role of plant secondary metabolites in mammalian herbivory: ecological perspectives.* Proceedings of the Nutrition Society, 2007. **64**: p. 123-131
2. Mouradov, A. and G. Spangenberg, *Flavonoids: a metabolic network mediating plants adaptation to their real estate.* Front Plant Sci, 2014. **5**.
3. Harborne, J.B., *Flavonoids and evolution of angiosperms.* Biochemical Systematics and Ecology, 1977. **5**(1): p. 7-22.
4. Yonekura-Sakakibara, K., Y. Higashi, and R. Nakabayashi, *The origin and evolution of plant flavonoid metabolism.* Frontiers in Plant Science, 2019. **10**(943).
5. Davies, K.M., et al., *The evolution of flavonoid biosynthesis: A byophyte perspective.* Frontiers in Plant Science, 2020. **11**(7).
6. Wink, M., *Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores.* Theoretical and Applied Genetics, 1988. **75**(2): p. 225-233.
7. Dixon, R.A., *Natural products and disease resistance.* Nature, 2001. **411**: p. 843-847.
8. Hartmann, T., *Plant-derived secondary metabolites as defensive chemicals in herbivorous insects: a case study in chemical ecology.* Planta, 2004. **219**(1): p. 1-4.
9. Srivastav, V.K., C. Egbuna, and M. Tiwari, *Chapter 1 - Plant secondary metabolites as lead compounds for the production of potent drugs*, in *Phytochemicals as Lead Compounds for New Drug Discovery*, C. Egbuna, et al., Editors. 2020, Elsevier. p. 3-14.
10. Wink, M., *Potential of DNA Intercalating Alkaloids and Other Plant Secondary Metabolites against SARS-CoV-2 Causing COVID-19.* Diversity, 2020. **12**(5): p. 175.

11. Verma, N. and S. Shukla, *Impact of various factors responsible for fluctuation in plant secondary metabolites.* Journal of Applied Research on Medicinal and Aromatic Plants, 2015. **2**(4): p. 105-113.

12. Russell, W. and G. Duthie, *Plant secondary metabolites and gut health: the case for phenolic acids.* Proceedings of the Nutrition Society, 2011. **70**(3): p. 389-396.

13. Wink, M., *Chapter 11 Importance of plant secondary metabolites for protection against insects and microbial infections*, in *Advances in Phytomedicine*, M. Rai and M.C. Carpinella, Editors. 2006, Elsevier. p. 251-268.

14. Croteau, R., T. Kutchan, and N. Lewis, *Natural Products (Secondary Metabolites)*, in *Biochemistry & Molecular Biology of Plants* B.B. Buchanan, W. Gruissem, and R.L. Jones, Editors. 2000, John Wiley & Sons, Inc: Somerset, NJ 08875-1272. p. 1250-1342.

15. Dixon, R.A. and D. Strack, *Phytochemistry meets genome analysis, and beyond.* Phytochemistry, 2003. **62**(6): p. 815-6.

16. J. Schmidt, T., et al., *The Potential of Secondary Metabolites from Plants as Drugs or Leads Against Protozoan Neglected Diseases - Part II.* Current Medicinal Chemistry, 2012. **19**(14): p. 2176-2228.

17. Wink, M., *Plant secondary metabolism: Diversity, function and its evolution.* Nat Prod Commun, 2008. **3**(8): p. 1205-1216.

18. Mohiuddin, A., *Chemistry of secondary metabolites.* Ann Clin Toxicol. 2019; 2 (1), 2019. **1014**.

19. Gutsch, A., et al., *Long-Term Cd Exposure Alters the Metabolite Profile in Stem Tissue of Medicago sativa.* Cells, 2020. **9**(12).

20. Delfin, J.C., M. Watanabe, and T. Tohge, *Understanding the function and regulation of plant secondary metabolism through metabolomics approaches.* Theoretical and Experimental Plant Physiology, 2019. **31**(1): p. 127-138.

21. Manthey, J.A., B.l. Buslig, and S. Symposium on Flavonoids in the Living, *Flavonoids in the living system*. 1998, New York: New York : Plenum Press, [1998].

22. Harborne, J.B. and H. Baxter, *The handbook of natural flavonoids*. 1999: John Wiley and Sons.

23. Buckingham, J. and V.R.N. Munasinghe, *Dictionary of flavonoids : with CD-ROM*. 2015, Boca Raton, Florida: Boca Raton, Florida : CRC Press, 2015.

24. Xie, D.-Y. and R.A. Dixon, *Proanthocyanidin biosynthesis - still more questions than answers?* Phytochemistry, 2005. **66**(18): p. 2127-2144.

25. Shi, M.-Z. and D.-Y. Xie, *Biosynthesis and metabolic engineering of anthocyanins in Arabidopsis thaliana.* Recent Pat Biotechnol, 2014. **8**(1): p. 47-60.

26. Dixon, R.A., D.Y. Xie, and S.B. Sharma, *Proanthocyanidins- a final frontier in flavonoid research?* New Phytologist, 2005. **165**: p. 9-28.

27. Peterson, J. and J. Dwyer, *Flavonoids: Dietary occurrence and biochemical activity.* Nutrition Research, 1998. **18**(12): p. 1995-2018.

28. Ross, J.A. and C.M. Kasum, *Dietary flavonoids: Bioavailability, metabolic effects, and safety.* Annual Review of Nutrition, 2002. **22**: p. 19-34.

29. Schewe, T., H. Kuhn, and H. Sies, *Flavonoids of Cocoa Inhibit Recombinant Human 5-Lipoxygenase.* J. Nutr., 2002. **132**(7): p. 1825-1829.

30. Beecher, G.R., *Overview of dietary flavonoids: Nomenclature, occurrence and intake.* Journal of Nutrition, 2003. **133**(10): p. 3248S-3254S.

31. Pretorius, J.C., *Flavonoids: A Review of Its Commercial Application Potential as Anti-Infective Agents.* Current Medicinal Chemistry - Anti-Infective Agents, 2003. **2**(4): p. 335-353.

32. Ishige, K., D. Schubert, and Y. Sagara, *Flavonoids protect neuronal cells from oxidative stress by three distinct mechanisms.* Free Radical Biology and Medicine, 2001. **30**(4): p. 433-446.

33. Spencer, J.P.E., et al., *The Beneficial Effects of Dietary Flavonoids: Sources, Bioavailability and Biological Functions*. Health Benefits of Organic Food: Effects of the Environment, ed. D.I. Givens, et al. 2006, Wallingford: Cabi Publishing-C a B Int. 207-239.

34. Zhu, Y. and D.-Y. Xie, *Docking characterization and in vitro inhibitory activity of flavan-3-ols and dimeric proanthocyanidins against the main protease activity of SARS-Cov-2.* Frontiers in Plant Science, 2020. **11**(1884).

35. Deavours, B.E. and R.A. Dixon, *Metabolic engineering of isoflavonoid biosynthesis in alfalfa.* Plant Physiol., 2005. **138**(4): p. 2245-2259.

36. Wang, X.Q., *Structure, function, and engineering of enzymes in isoflavonoid biosynthesis.* Functional & Integrative Genomics, 2011. **11**(1): p. 13-22.

37. Lee, Y.J., et al., *Characterization of flavone synthase I from rice.* Bmb Reports, 2008. **41**(1): p. 68-71.

38. Yun, C.S., et al., *Expression of parsley flavone synthase I establishes the flavone Biosynthetic pathway in Arabidopsis thaliana.* Bioscience Biotechnology and Biochemistry, 2008. **72**(4): p. 968-973.

39. Park, J., et al., *Therapeutic potential of EGCG, a green tea polyphenol, for treatment of coronavirus diseases.* Life, 2021. **11**(3): p. 197.

40. Li, J.R., et al., *Antiviral Effect of Epigallocatechin Gallate via Impairing Porcine Circovirus Type 2 Attachment to Host Cell Receptor.* Viruses-Basel, 2020. **12**(2): p. 18.

41. Rogovskii, V.S., et al., *The possibility of preventive and therapeutic use of green tea catechins in prostate cancer.* Anti-Cancer Agents in Medicinal Chemistry, 2019. **19**(10): p. 1223-1231.

42. Mekky, R.Y., et al., *Epigallocatechin gallate (EGCG) and miR-548m reduce HCV entry through repression of CD81 receptor in HCV cell models.* Archives of Virology, 2019. **164**(6): p. 1587-1595.

43. Suganuma, M., A. Saha, and H. Fujiki, *New cancer treatment strategy using combination of green tea catechins and anticancer drugs.* Cancer Science, 2011. **102**(2): p. 317-323.

44. Kelishadi, R., *Cacao to cocoa to chocolate: Healthy food?* Arya Atherosclerosis, 2005. **1**(1): p. 28-34.

45. Rauf, A., et al., *Proanthocyanidins: A comprehensive review.* Biomedicine & Pharmacotherapy, 2019. **116**: p. 108999.

46. Dixon, R.A. and S. Sarnala, *Proanthocyanidin biosynthesis-a matter of protection.* Plant Physiology, 2020. **184**(2): p. 579-591.

47. Liu, C.G., et al., *A role for leucoanthocyanidin reductase in the extension of proanthocyanidins.* Nat. Plants, 2016. **2**(12).

48. Jun, J.H., et al., *Proanthocyanidin subunit composition determined by functionally diverged dioxygenases.* Nature Plants, 2018. **4**(12): p. 1034-1043.

49. Xie, D.-Y., et al., *Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis.* Science, 2003. **299**(5605): p. 396-399.

50. Xie, D.-Y., S.B. Sharma, and R.A. Dixon, *Anthocyanidin reductases from Medicago truncatula and Arabidopsis thaliana.* Arch Biochem Biophys, 2004. **422**(1): p. 91-102.

51. Jun, J.H., et al., *Dual activity of anthocyanidin reductase supports the dominant plant proanthocyanidin extension unit pathway.* Science Advances, 2021. **7**(20): p. eabg4682.

52. Wang, P., et al., *Functional demonstration of plant flavonoid carbocations proposed to be involved in the biosynthesis of proanthocyanidins.* The Plant Journal, 2020. **101**(1): p. 18-36.

53. Org, C.G., *The Plant List with literature*. 2016.

54. Page, R., *The plant list with literature*. 2016.

55. Bibi Sadeer, N., et al., *Chemical profiling, antioxidant, enzyme inhibitory and molecular modelling studies on the leaves and stem bark extracts of three African medicinal plants.* Journal of Pharmaceutical and Biomedical Analysis, 2019. **174**: p. 19-33.

56. Bloor, S.J. and S. Abrahams, *The structure of the major anthocyanin in Arabidopsis thaliana.* Phytochemistry, 2002. **59**(3): p. 343-346.

57. Shi, M.-Z. and D.-Y. Xie, *Features of anthocyanin biosynthesis in pap1-D and wild-type Arabidopsis thaliana plants grown in different light intensity and culture media conditions.* Planta, 2010. **231**: p. 1385-1400.

58. He, X., et al., *Metabolic engineering of anthocyanins in dark tobacco varieties* Physiol Plant, 2017. **159**: p. 2-12.

59. Zhou, L.-L., et al., *Development of tobacco callus cultures over expressing Arabidopsis PAP1/MYB75 transcription factor and characterization of anthocyanin biosynthesis* Planta, 2008. **229**: p. 37-51.

60. Li, G., et al., *Creation of elite growth and development features in PAP1-programmed red Nicotiana tabacum Xanthi via overexpression of synthetic geranyl pyrophosphate synthase genes.* Molecular Breeding, 2019. **39**(4): p. 63.

61. Li, M., et al., *A de novo regulation design shows an effectiveness in altering plant secondary metabolism.* Journal of Advanced Research, 2021. **in press**.

62. Peng, Q.Z., et al., *An integrated approach to demonstrating the ANR pathway of proanthocyanidin biosynthesis in plants.* Planta, 2012. **236**(3): p. 901-18.

63. Xie, D.-Y., et al., *Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor.* Plant J, 2006. **45**(6): p. 895-907.

64. Hiraga, S., et al., *A large family of class III plant peroxidases.* Plant Cell Physiol, 2001. **42**(5): p. 462-468.

65. Almagro, L., et al., *Class III peroxidases in plant defence reactions.* J Exp Bot, 2009. **60**(2): p. 377-390.

66. Yuzuak, S., J. Ballington, and D.-Y. Xie, *HPLC-qTOF-MS/MS-based profiling of flavan-3-ols and dimeric proanthocyanidins in berries of two muscadine grape hybrids FLH 13-11 and FLH 17-66.* Metabolites, 2018. **8**(4): p. 57.

67. Quideau, S., et al., *Plant Polyphenols: Chemical Properties, Biological Activities, and Synthesis.* Angewandte Chemie-International Edition, 2011. **50**(3): p. 586-621.

68. Cheynier, V., *Phenolic compounds: from plants to foods.* Phytochemistry Reviews, 2012. **11**(2-3): p. 153-177.

69. Andersen, Ø.M. and K.R. Markham, *Flavonoids : chemistry, biochemistry, and applications*. 2006, Boca Raton, FL: Boca Raton, FL : CRC Taylor & Francis, 2006.

70. Madhuri, G. and A.R. Reddy, *Plant Biotechnology of Flavonoids.* Plant Biotechnology, 1999. **16**(3): p. 179-199.

71. Stafford, H.A., *Flavonoid evolution: an enzymic approach.* Plant Physiol, 1991. **96**: p. 680-685.

72. Winkel-Shirley, B., *Flavonoid Biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology.* Plant Physiol, 2001. **126**: p. 485-493.

73. Azimova, S.S., V.I. Vinogradova, and I. ebrary, *Natural compounds Flavonoids : plant sources, structure and properties*. 2013, New York, NY: New York, NY : Springer, c2013.

74. Vuorela, P., et al., *Natural products in the process of finding new drug candidates.* Curr Med Chem, 2004. **11**(11): p. 1375-1389.

75. Mabry, T.J., K.R. Markham, and M.B. Thomas, *The systematic identification of the flavonoids*. 1970, New York: Springer-Verlag. 354.

76. Harborne, J.B., The Flavonoids : advances in research since 1980. 1988, London New York: London ; New York : Chapman and Hall, 1988.

77. Harborne, J.B., The Flavonoids : advances in research since 1986. 1994, London New York: London ; New York : Chapman & Hall, 1994.

78. Harborne, J.B. and T.J. Mabry, The Flavonoids : advances in research. 1982, London New York: London ; New York : Chapman and Hall, 1982.

79. Fukami, H., Y. Yano, and T. Iwashita, *Isolation of a reduced form of cyanidin 3-O-beta-D-glucoside from immature black soybean (Glycine max (L.) Merr.) and its reducing properties.* Journal of Oleo Science, 2013. **62**(8): p. 623-629.

80. Yoshida, K., et al., *5,7,3',4'-Tetrahydroxyflav-2-en-3-ol 3-O-glucoside, a new biosynthetic precursor of cyanidin 3-O-glucoside in the seed coat of black soybean, Glycine max.* Scientific Reports, 2020. **10**(1): p. 10.

81. Chaipech, S., et al., *New flav-3-en-3-ol glycosides, kaempferiaosides C and D, and acetophenone glycosides, kaempferiaosides E and F, from the rhizomes of Kaempferia parviflora.* Journal of Natural Medicines, 2012. **66**(3): p. 486-492.

Supplemental data

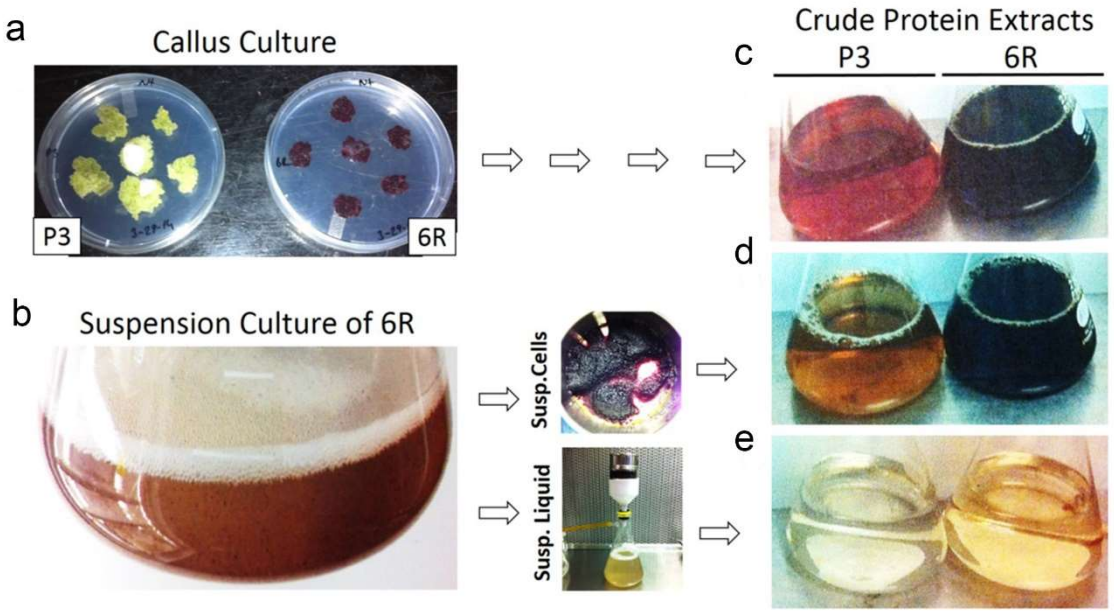


Figure S1 Callus and suspension culture of red and wild-type cells for protein extraction. a, callus culture of wild-type tobacco (P3) and PAP1 transgenic (6R) tobacco cells grown on the subculture media; b, suspension culture of 6R tobacco cells growing in liquid medium, c-e, crude proteins extracted from P3 and 6R cultures, calli (c), suspension cells harvested (left) (d), and liquid medium (left) (e).

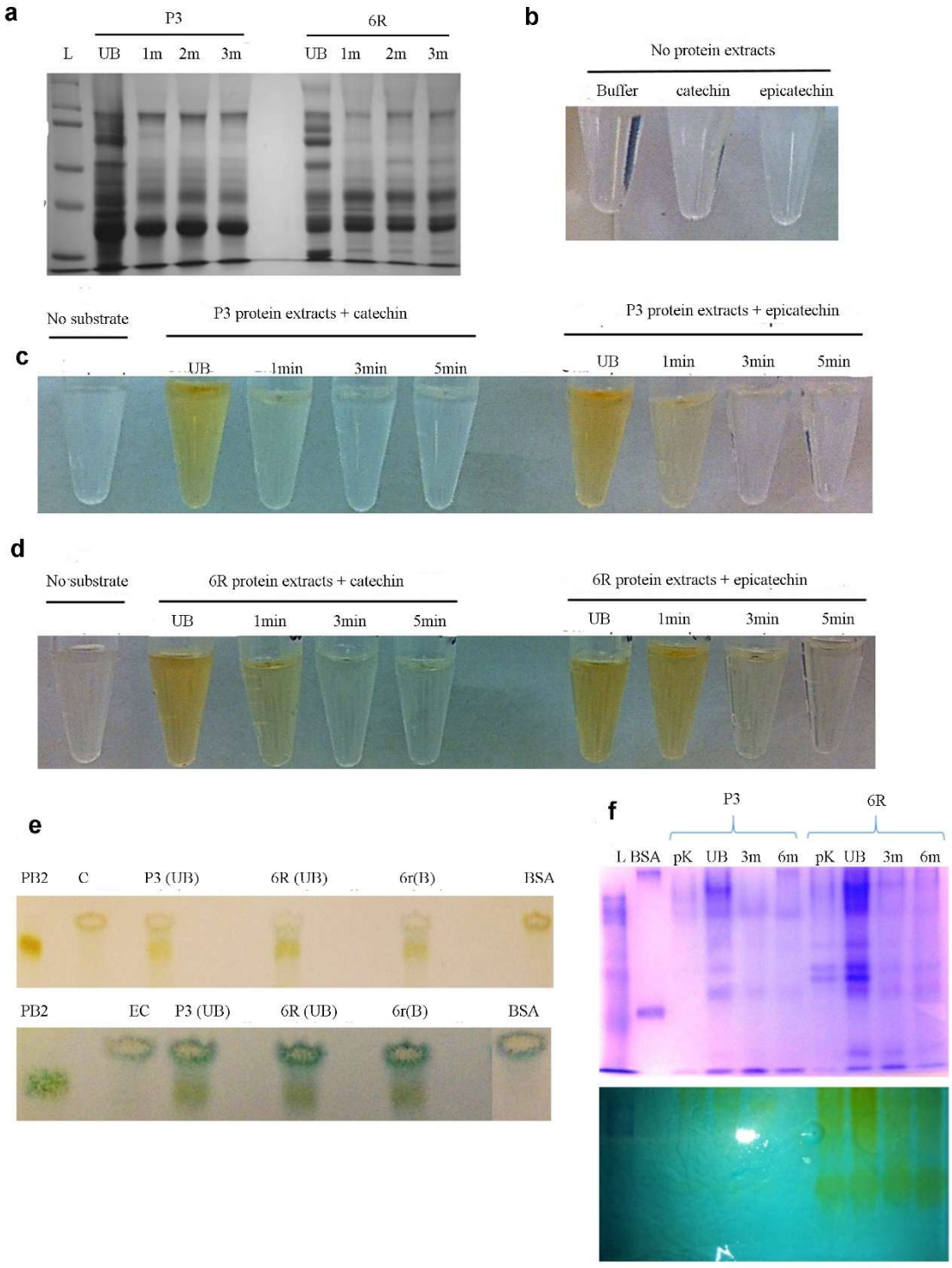
Figure S2 Crude proteins extracted from 6R and P3 cell suspension cultures and incubation with substrates.

a, A SDS-PAGE image shows different profiles of crude proteins extracted from 6R (shown in Figure 2 too) and P3 cell suspension cultures and treated with boiling. L, protein ladder (kDa), UnB: Un-boiled crude protein extracts, 1-3m, proteins after crude protein extracts were boiled for 1, 2, and 3 minutes at 100°C.

b-d, Incubation of crude protein extractions with catechin or epicatechin as substrate produce new yellowish or brownish-yellowish compounds. Crude proteins used here include un-boiled and boiled (1, 3, and 5 min, labeled as 1m, 3m, and 5m). b, buffer and two substrates in the buffer are colorless. c; un-boiled and 1 min-boiled P3 crude proteins converted catechin and epicatechin to yellowish or light brownish compounds. d, un-boiled and 1 and 3 min-boiled 6R crude proteins converted catechin and epicatechin to yellowish or light brownish compounds.

e, TLC and DMACA staining showed PA-like oligomers from formed from in vitro incubations of crude protein extractions of 6R and P3 suspension cells with (+)-catechin and (-)-epicatechin. Both crude 6R protein extracts and 1 min-boiled proteins were tested for catalytic activity. Bovine Serum Albumin (BSA) was used as protein control. In addition, the negative (-) labelling denoted that only proteins were used in the incubations without substrates added as controls. Standards include procyanidin B2 (PB2), (+)-catechin (C) and (-)-epicatechin (EC). Unb; un-boiled.

f, Native-PAGE images show the protein band profiles from un-boiled, 3 and 6 min-boiled, and protease K digested crude proteins of P3 and 6R. Gels were stained with Coomassie blue and incubated with 100 μM epicatechin. L, protein ladder (kDa), pK, proteinase K digested. The incubation with epicatechin not only demonstrated that crude proteins converted epicatechin to yellowish compounds, but also revealed the catalytic differences between P3 and 6R proteins.


a
P3
6R
L UB 1m 2m 3m UB 1m 2m 3m
b
No protein extracts
Buffer catechin epicatechin
c
No substrate
P3 protein extracts + catechin
P3 protein extracts + epicatechin
UB 1min 3min 5min UB 1min 3min 5min
d
No substrate
6R protein extracts + catechin
6R protein extracts + epicatechin
UB 1min 3min 5min UB 1min 3min 5min
e
PB2 C P3 (UB) 6R (UB) 6r(B) BSA
PB2 EC P3 (UB) 6R (UB) 6r(B) BSA
f
P3
6R
L BSA pK UB 3m 6m pK UB 3m 6m

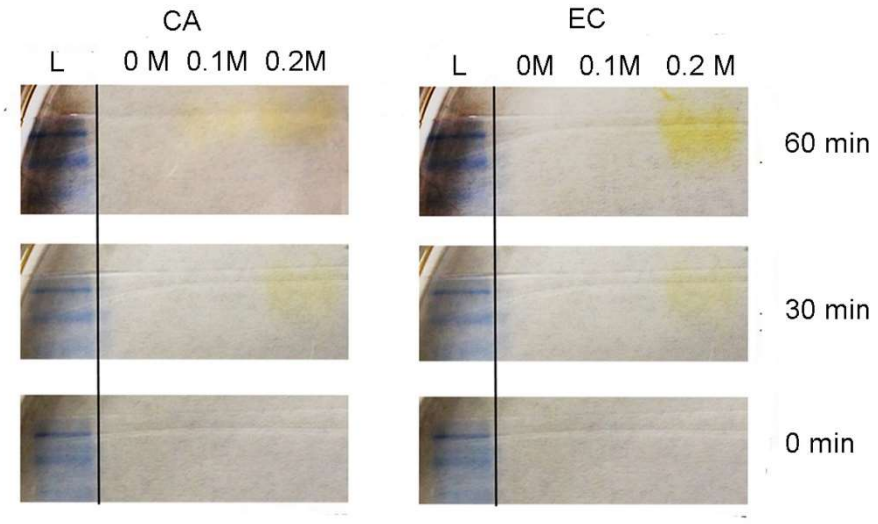
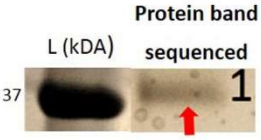


Figure S3 Protein fractions isolated from ion-exchange columns (IECs) and examination of their activity via an incubation of native-PAGE gels and substrates.
Proteins were eluted through IECs and loaded onto a Native-PAGE with buffer supplemented with 0, 0.1 M, and 0.2 M NaCl. L; protein ladder (kDa). Native PAGE gel images show the dynamics of yellowing products formed on the gels after incubated with catechin and epicatechin for 0, 30 and 60 min. These data demonstrate the presence of an active enzyme.



A0A1S4C1T9_TOBAC (100%), 38,666.8 Da
Peroxidase OS=Nicotiana tabacum GN=LOC107814252 PE=3 SV=1
17 exclusive unique peptides, 50 exclusive unique spectra, 297 total spectra, 111/360 amino acids (31% coverage)

MAFRLSHLSL ALSLVALALA GVAIYRNTYE AMSKGFQTLS PEFDLLESAA
SILTLNNNAD QQNSDSKLTQ PLSPSACIFS AVRGVVNSAI DRERRMGASL
IRLHFHDCFV DGCDGGVLLD DIPGSFQGEK TSPPNNNSAR GFEVIEQAKQ
RVKDTCPNTP VSCADILAIA ARDSVVKLGG QGYNVALGRR DARTANFTGA
LTQLPAPFDN LTVQLRKFND KNFNAREMVA LAGAHTVGFA RCATVCNSNN
VNPAARLQCN CSVTQNDTNL QQLDITPAVF DRVYFQDLNR NQGILFSDQV
LTGNTTTAAI VTTYSNNGAV FFGDFAAAMI KMGNLPPSQG VQLEIRDVCS
RVNPSSVASM

Figure S4 Results from peptide sequencing of an active fraction and the blast. The amino acid sequences of six oligomers highlighted with a yellow background were obtained from proteomics. The blast search with the six oligomers hit one protein in the GenBank, which were annotated a peroxidase.

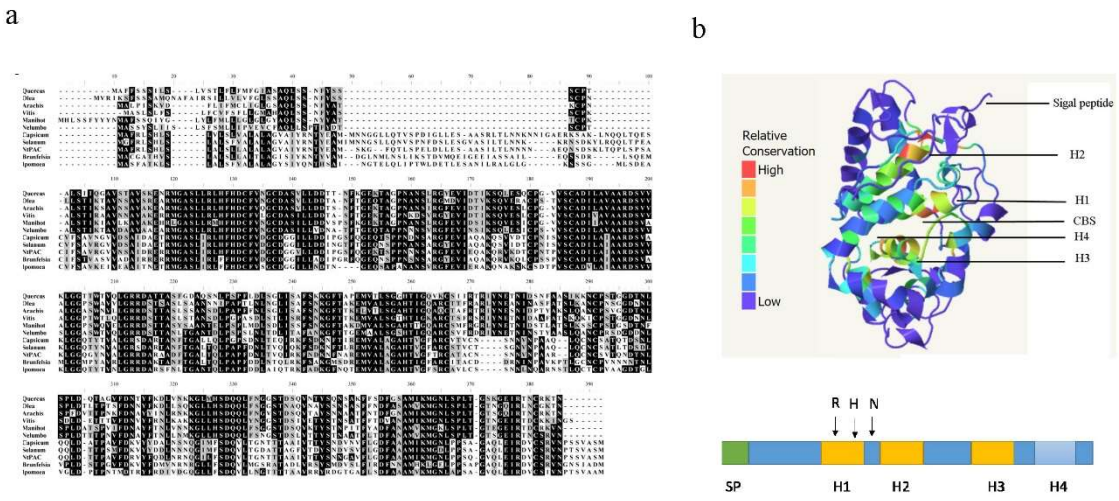
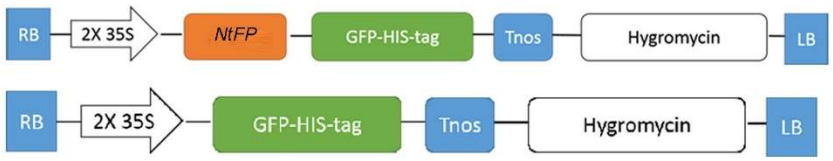


Figure S5 Amino acid sequence alignment and 3D modeling of FP.
a, an amino acid sequence alignment was developed using FP and 10 homologs.
b, a 3D modeling was developed based on a peanut peroxidase as the template. SP: signal peptide, H1, H2, H3, and H4: helix. CBS: calcium binding sites including R, H, and N residues.

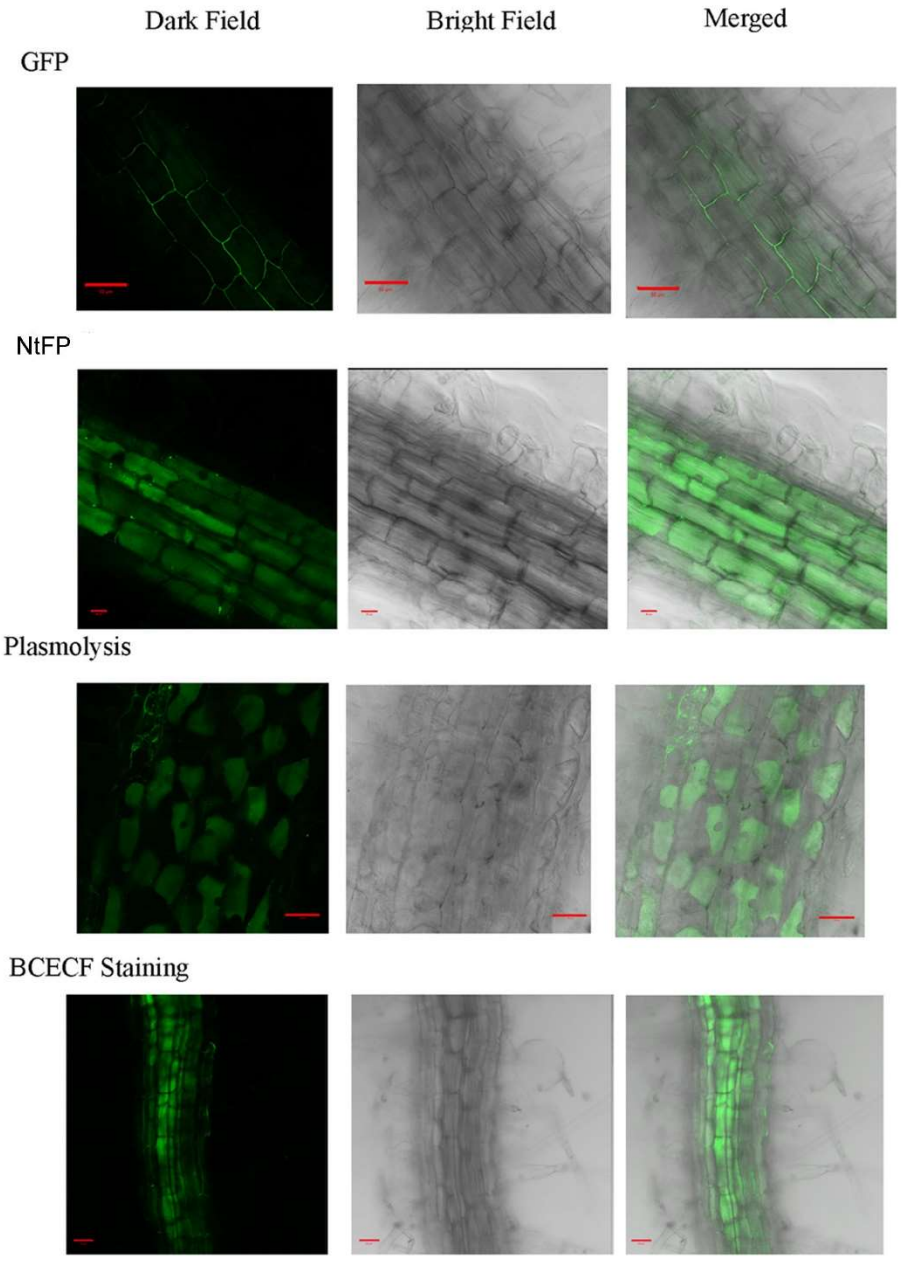
Figure S6 Constructs and subcellular localization of NtFP in stably transgenic plants created using PAP1-BAN (ANR) tobacco.

a, Two T-DNA cassettes were developed for genetic transformation, the T-DNA cassette of the PMDC84 binary vector and the T-DNA cassette of a recombinant binary vector developed from PMDC84 by fusing *NtFP* to the 5'-end of *GFP* gene for subcellular localization. Two T-DNA cassettes were transformed into PAP1-ANR plants. PMDC84 was used as a vector control.
b, confocal microscopy images showed the vacuolar localization of the GFP fluorescence signal of NtFP-GFP in root cells of transgenic tobacco seedling. GFP: GFP control; NtFP: fused NtFP-GFP in vacuoles; plasmolysis: fused NtFP-GFP in vacuoles; BCECF staining: positive control of vacuolar localization.

a


RB
2X 35S
NtFP
GFP-HIS-tag
Tnos
Hygromycin
LB
RB
2X 35S
GFP-HIS-tag
Tnos
Hygromycin
LB

b


Dark Field
Bright Field
Merged
GFP
NtFP
Plasmolysis
BCECF Staining

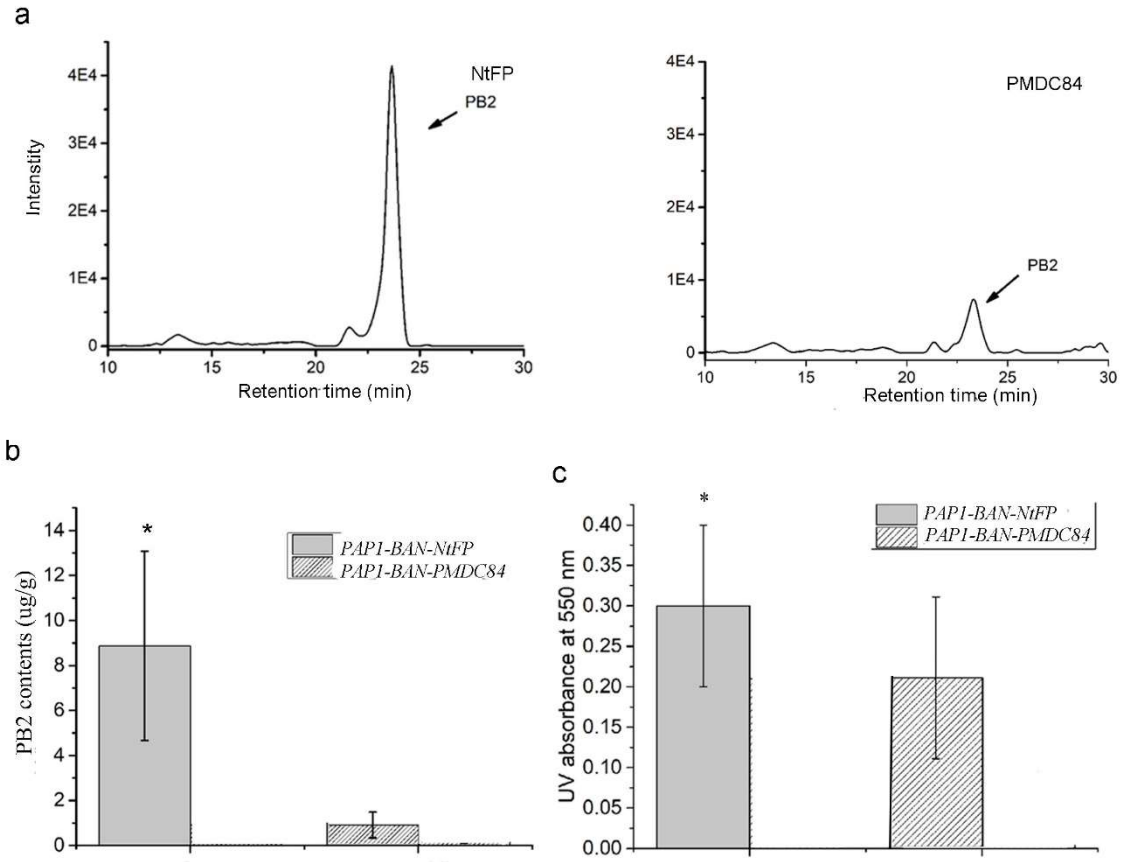


Figure S7 Analysis of proanthocyanidins in *NtFP* transgenic PAP1-BAN plants.
a, EIC profiles show PB2 detected in leaves of both *PAP1-BAN-NtFP* and PMDC84 empty vector control transgenic tobacco plants.
b, the PB2 content was significantly higher in leaves of *PAP1-BAN-NtFP* than in those of PMDC84 empty vector control transgenic tobacco.
c, the UV absorbance values of boiled extracts in butanol-HCl were measured at 550 nm. The absorbance values were higher in leaf extracts of *PAP1-BAN-NtFP* plants than in those of *PAP1-BAN-PMDC84* plants.

Figure S8 Effects of pH values and hydrogen peroxide on the formation of yellowish compounds from (-)-epicatechin

a, effects of pH6 and pH8 as well as hydrogen peroxide on yellow color formation from the *in vitro* incubation of native FP with epicatechin. Buffer was not vortexed for oxygenation prior to adding enzyme and epicatechin. At pH6, only the incubation with enzyme and hydrogen peroxide converted epicatechin to yellowish compound(s). At pH8, all incubations except for BSA protein control converted epicatechin to light yellowish to slightly brownish compounds. These results indicate that in a weak acidic condition, hydrogen peroxide is required for the catalysis to produce yellowish compounds, while in a weak basic condition, hydrogen peroxide enhances producing yellowish and brownish compounds.

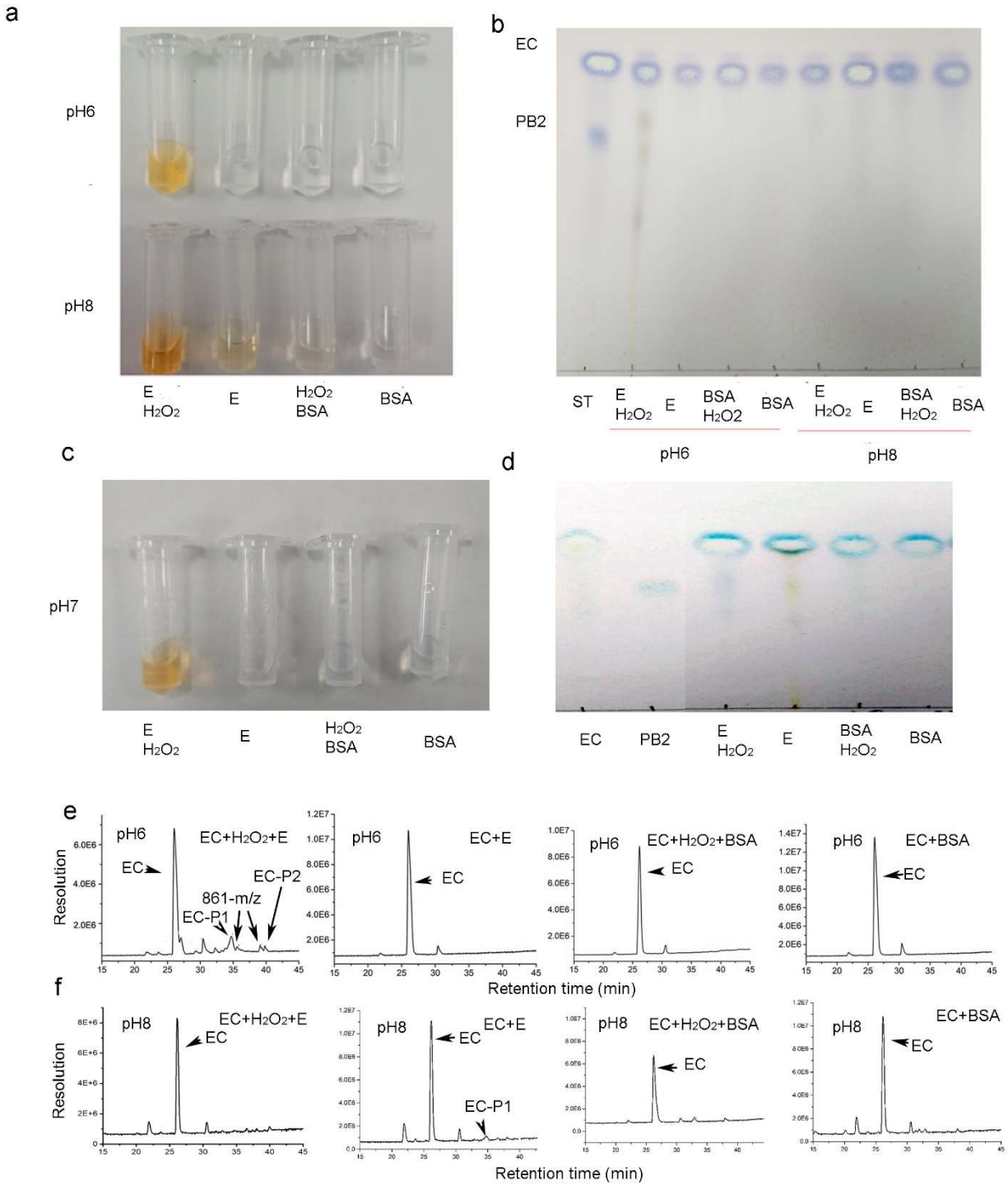
b, TLC analysis detected PA-like products formed from the incubation of the recombinant FP and epicatechin at pH6 and hydrogen peroxide but hardly detected these compounds from other conditions. The incubations were extracted with ethyl acetate, which was evaporated. The residues were dissolved in methanol for TLC. TLC plates were sprayed with DMACA reagents.

c-d, hydrogen peroxide promoted the formation of yellowing compound in the incubations at pH7 (c) and TLC analysis detected PA-like compounds formed from the incubation of recombinant FP and epicatechin but did not detect compounds from two control incubations (d).

e, a, TIC profiles show EC-P1, EC-P2 and other peaks formed from the incubation consisting of native FP, epicatechin, and hydrogen peroxide (EC+H2O2+E) but not produced from three incubation controls, native enzyme and epicatechin (EC+E), BSA, epicatechin and hydrogen peroxide (EC+H2O2+BSA), and 4) BSA and epicatechin (EC+BSA) at pH 6

f, TIC profiles only show EP-P1 formed from the incubation consisting of native NtFP and epicatechin (EC+E) but not formed in three others types of incubations, NtFP, epicatechin, and hydrogen peroxide (EC+H2O2+E), (3) BSA, epicatechin, and hydrogen peroxide

(EC+H2O2+BSA), and BSA and epicatechin (EC+BSA).



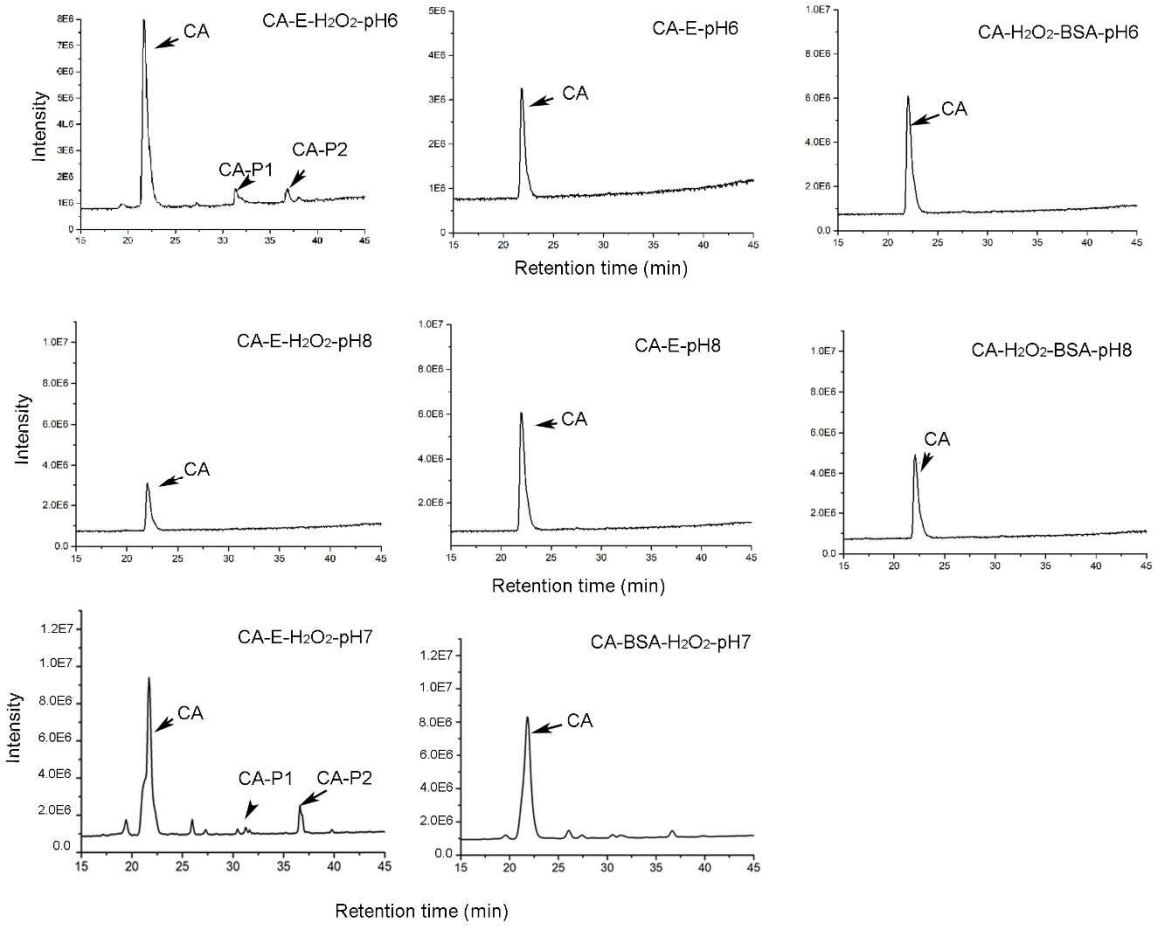


Figure S9 Effects of pH values and hydrogen peroxide on the formation of yellowish products from the incubation of FP and catechin.

a, TIC profiles show CA-P1 and CA-P2 products formed from the incubation consisting of native NtFP and catechin and hydrogen peroxide (left: CA+E+H2O2, pH6) but not formed from two control incubations, native enzyme and catechin (middle: CA+E-pH6) and BSA, catechin, and hydrogen peroxide (right: CA+H2O2+BSA, pH6) at pH6.

b, TIC profiles show CA-P1 and CA-P2 products formed from the incubation consisting of native NtFP and catechin and hydrogen peroxide (left: CA+E+H2O2, pH7) but not formed from the control incubation, enzyme catechin, and hydrogen peroxide (right: CA+BSA+H2O2, pH7) at pH7.

c, TIC profiles do not show products formed from three incubations at pH 8, native NtFP, catechin, and hydrogen peroxide (left: CA+E+H2O2, pH8), native NtFP and catechin (middle: CA+E, pH8), and BSA, catechin, and hydrogen peroxide (right CA+H2O2+BSA, pH8).

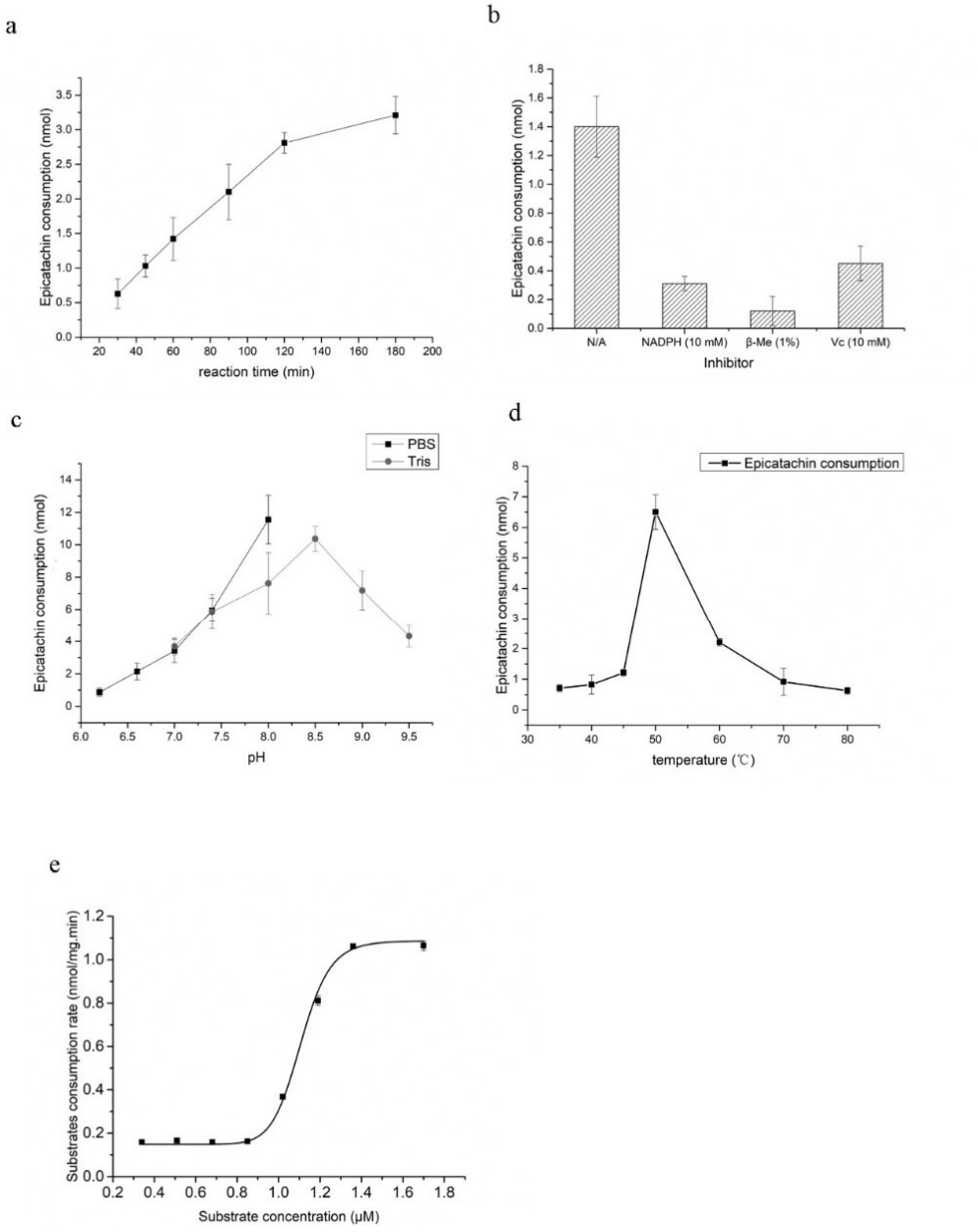


Figure S10 Effects of reaction time, reducing reagents, pH values, and temperatures on enzyme activity and sigmoid kinetics.

Recombinant enzyme was used to test these conditions. Due to at least two products formed in the reactions, the reduction of epicatechin concentrations was used to calculate values of reaction time, reducing agents, pH and temperature optima, and Km value.

a-d, plots show effects of reaction time (a), reducing reagents (b), pH values (c), and temperature (d) on the activity of recombinant enzyme.

e, a plot of velocity versus epicatechin contraction shows a sigmoid kinetics of recombinant NtFP activity.

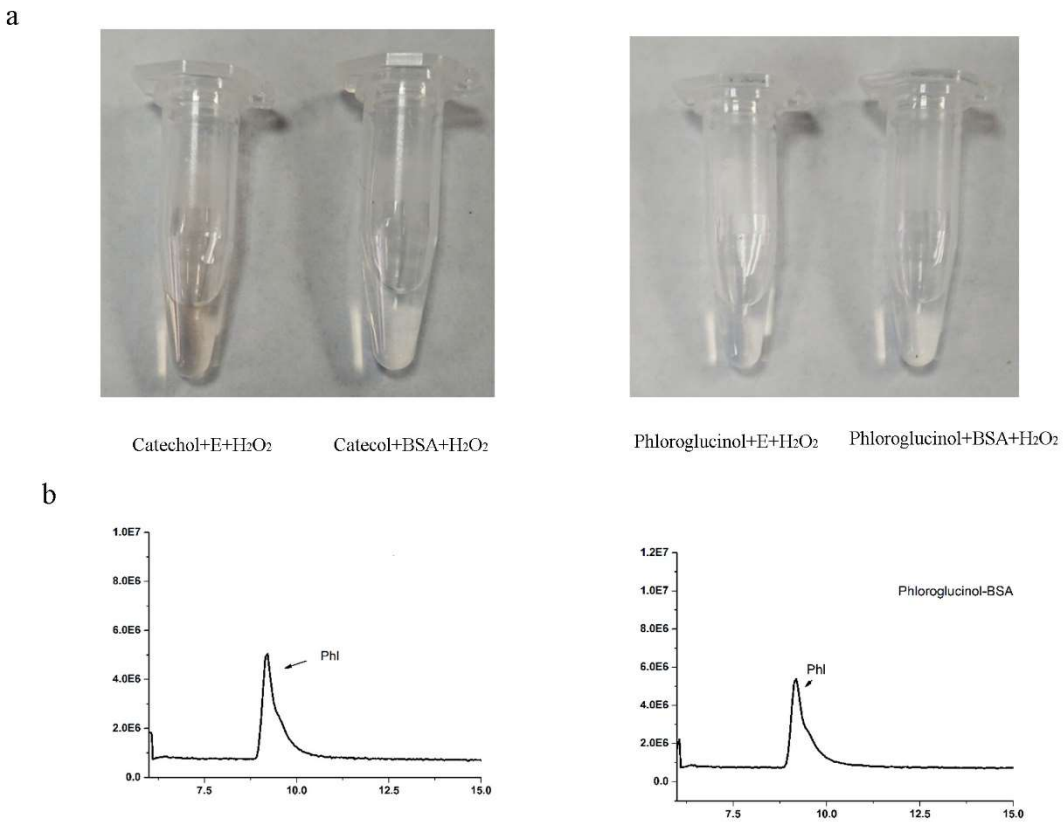


Figure S11 Enzyme catalysis of catechol but no enzyme catalysis of phloroglucinol by NtFP
Incubations were composed of catechol or phloroglucinol, enzyme or BSA control, and hydrogen peroxide in pH 6 buffer. Incubations were extracted with ethyl acetate
a, images showed that the recombinant enzyme converted catechol to yellowish compounds but BSA control did not; the recombinant enzyme and BSA control did not produce color from the incubation of phloroglucinol.
b, TIC profiles showed the level of phloroglucinol was not changed between the enzymatic and BSA control incubations, indicating that NtFP cannot use this compound as a substrate. HPLC profiles for catechol reaction could not detect substrate and products because catechol was completely catalyzed into water soluble quinone, which could not be extracted with ethyl acetate.

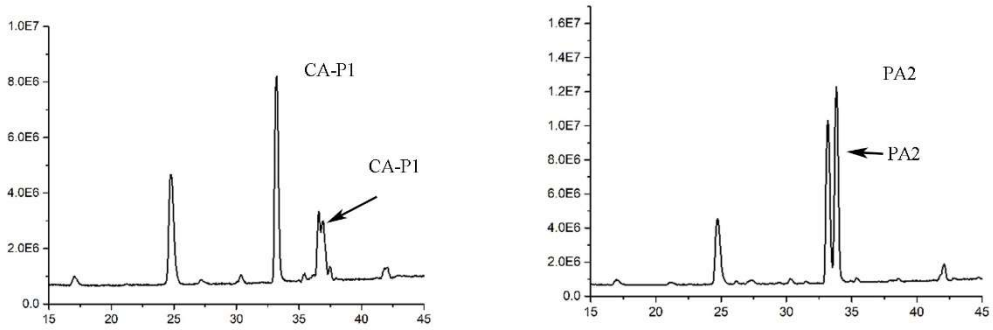


Figure S12 No catechin-phloroglucinol and epicatechin-phloroglucinol produced from hydrochloric acid (HCl)-catalyzed cleavage of CA-P1 and procyanidin A2 (PA2) in presence of phloroglucinol (support Figure 4)

TIC profiling did not detect catechin-phloroglucinol and epicatechin-phloroglucinol formed from the methanol HCl-catalyzed cleavage CA-P1 (left) and PA2 (right). These data indicate that HCl-methanol does not cleave CA-P1 to catechin and catechin carbocation and PA2 to epicatechin and epicatechin carbocation, therefore, phloroglucinol cannot nucleophilically attack epicatechin carbocation to form epicatechin-phloroglucinol and catechin carbocation to form catechin-phloroglucinol.

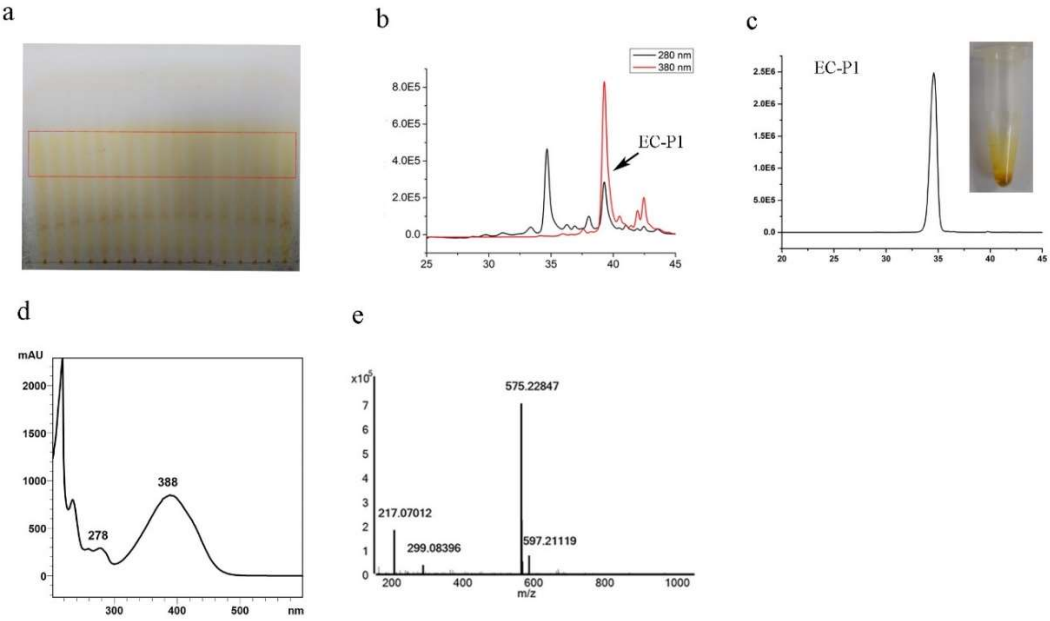


Figure S13 Purification of EC-P1 produced from enzymatic reactions and its features of UV spectrum and mass spectrum

a-c, TLC preparation (a) and HPLC separation (b) were used to isolate EP-P1. TIC showed EC-P1 purified (c).

d, the UV-spectrum profile shows a featured absorption of EP-P1at 388 nm.

e, the mass spectrum shows an accurate mass-to-charge ratio of EP-P1, 575.22847 $[m/z]^-$.

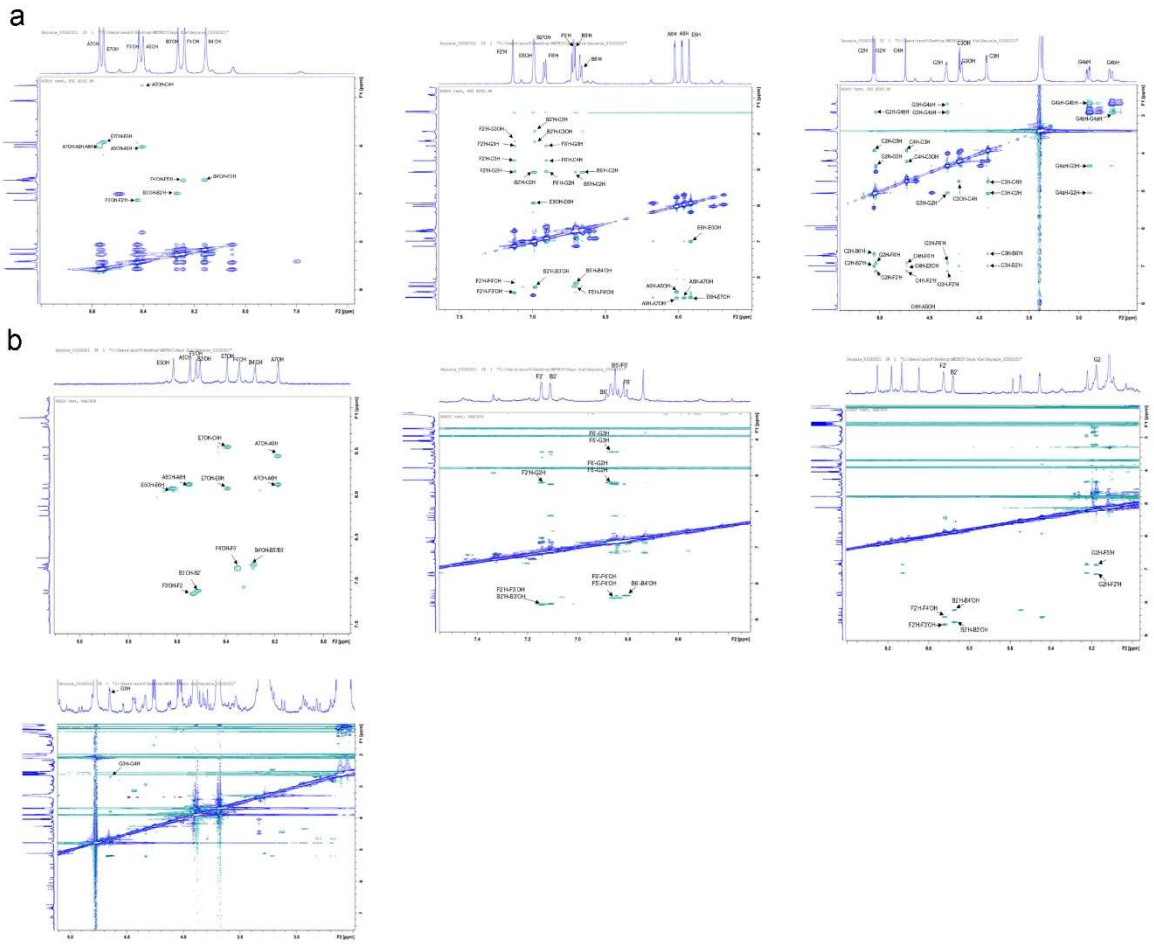


Figure S14 NMR ROESY assignments of PB2 and EC-P1
a, ROESY assignments of PB2.
b, ROESY assignments of EC-P1.

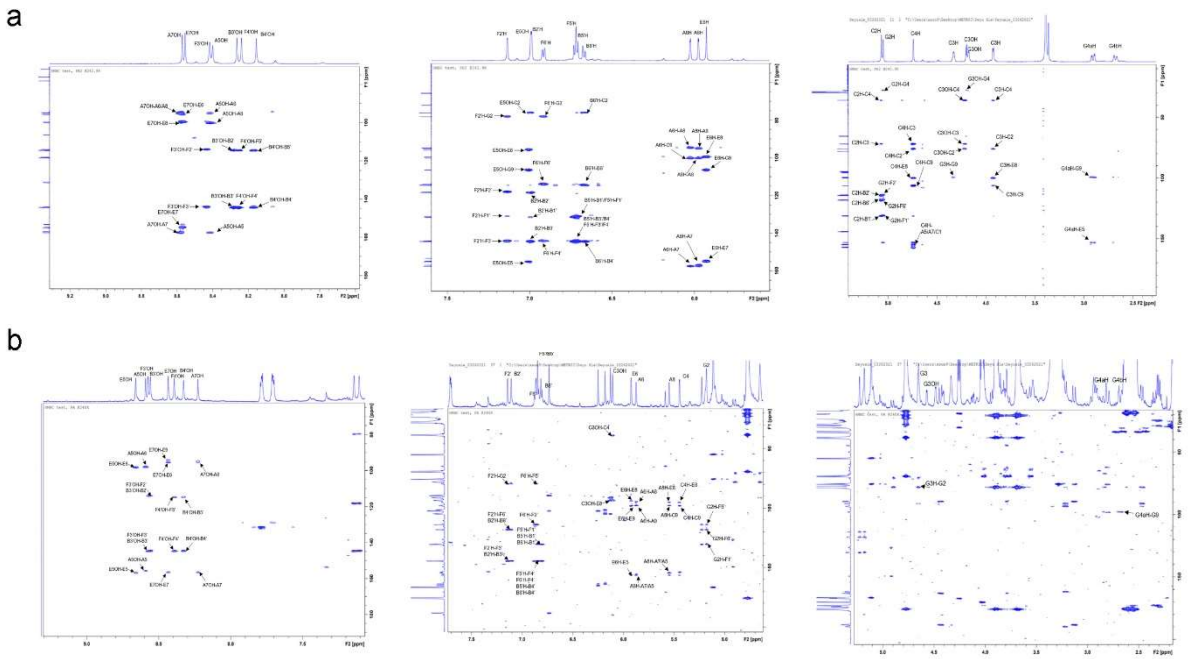


Figure S15 NMR HMBC assignments of PB2 and EC-P1
a, HMBC assignments of PB2.
b, HMBC assignments of EC-P1.

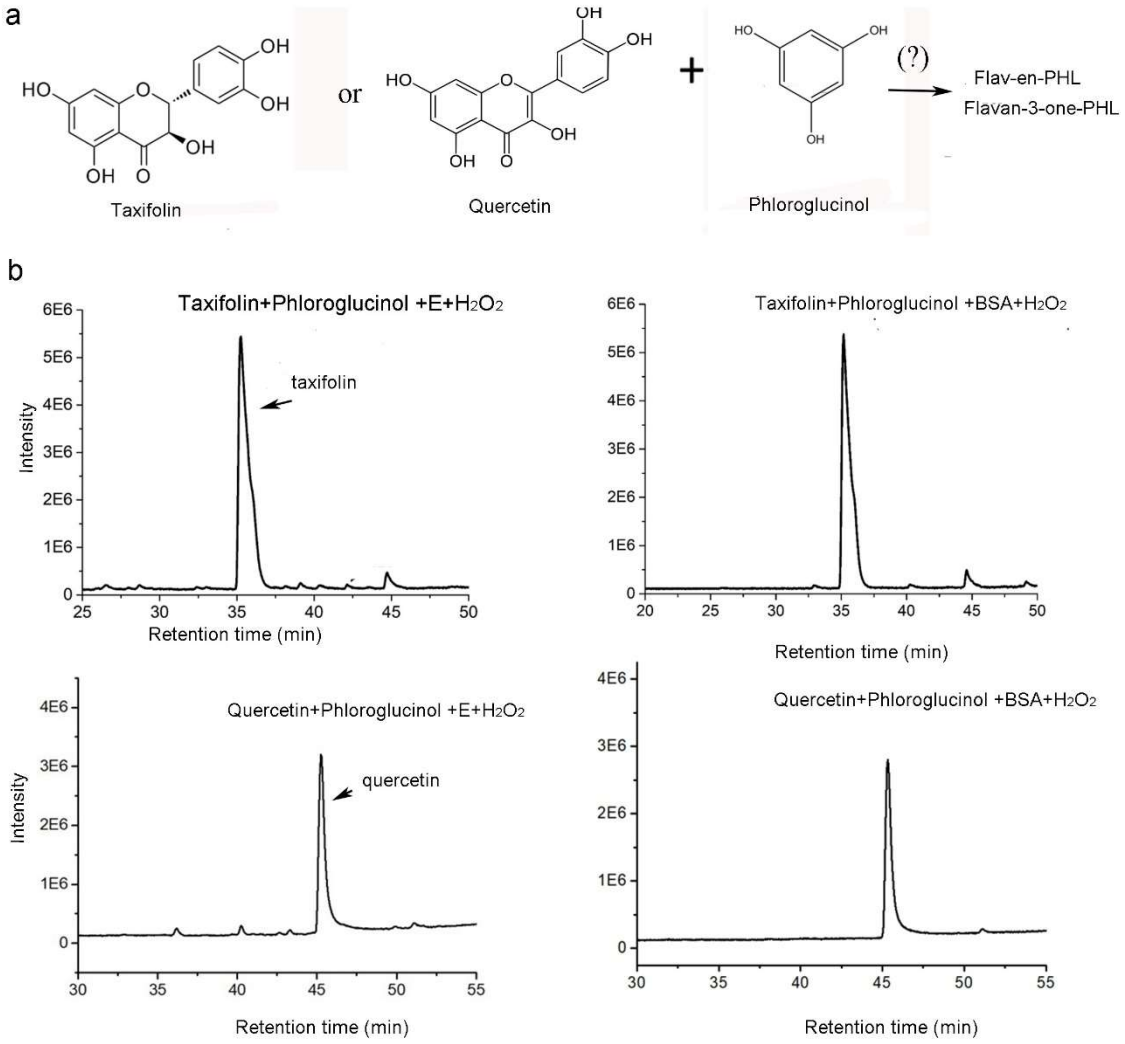


Figure S16 No formation of the 412 Dalton compound formed from incubations of quercetin or taxifolin, phloroglucinol, and NtFP

a, a scheme presents a hypothesis whether the incubation with enzyme, quercetin or taxifolin, phloroglucinol, and hydrogen peroxide produces conjugates in pH 6 buffer.

b, TIC profile showed that no molecular weight 412 Dalton compound (411 $[m/z]^-$) was detected from the incubations of enzyme or BSA (control), quercetin, phloroglucinol, and hydrogen peroxide.

c, TIC profiles showed that no molecular weight 412 Dalton compound (411 $[m/z]^-$) was detected from the incubations of enzyme or BSA (control), taxifolin, phloroglucinol, and hydrogen peroxide.

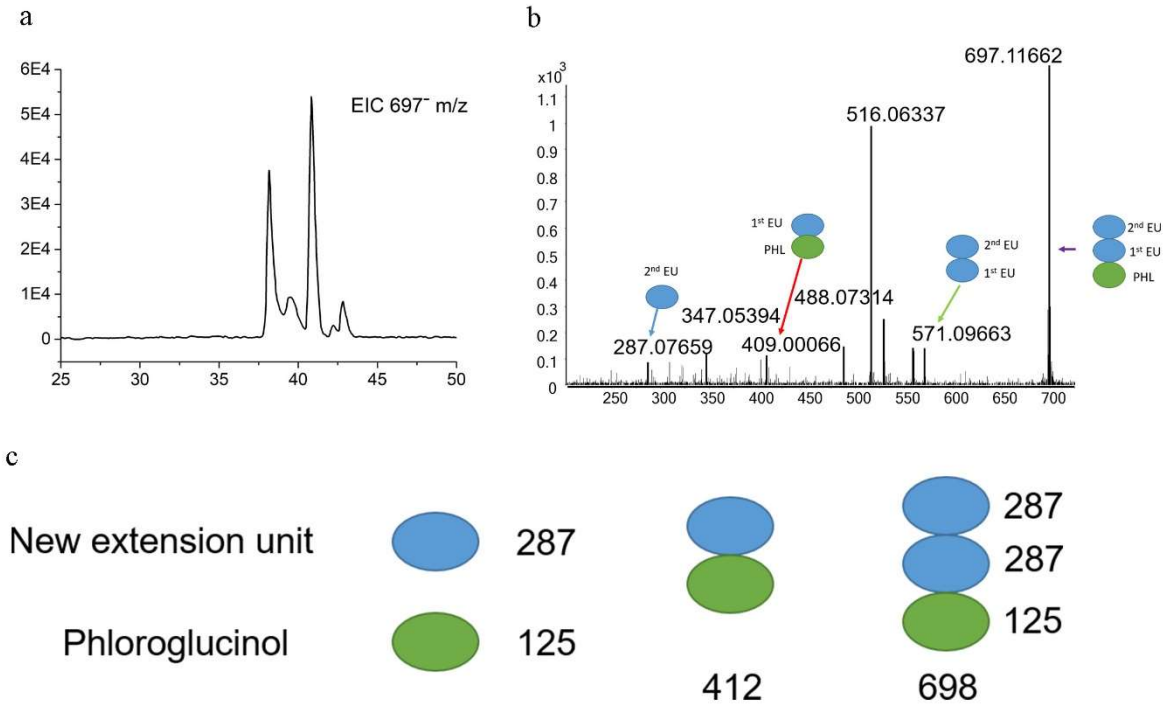


Figure S17 Formation of new trimeric compounds from phloroglucinol added in the incubation of NtFP and epicatechin.

Extracted-ion chromatogram (EIC) was performed to characterize MS of compounds formed from the incubation of enzyme, epicatechin, hydrogen peroxide, and phloroglucinol.

a, an EIC profile showed five peaks characterized by a 697 $[m/z]^-$ produced from the incubations.

b, MS/MS fragmentation showed main fragments from all 697 $[m/z]^-$ compounds, including 571, 516, 409, and 287.

c, a carton scheme presents a mechanism by which the incubation of enzyme, epicatechin, hydrogen peroxide, and phloroglucinol produces trimeric compounds. Phloroglucinol forms the starter unit.

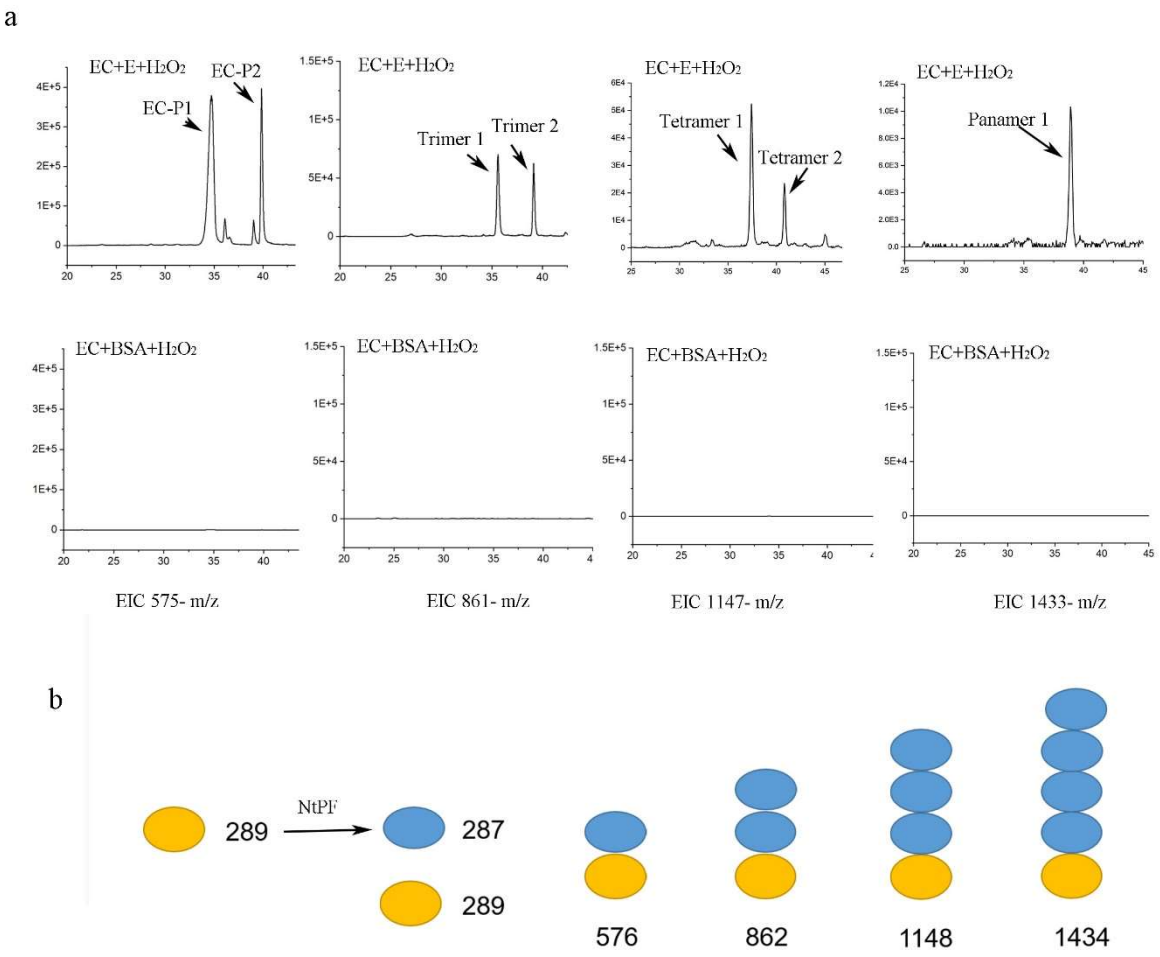


Figure S18 A series of new oligomeric compounds detected from the incubations of enzyme, and epicatechin.

Extracted ion chromatograms (EIC) were recorded to detect compounds from the incubation of enzyme, epicatechin, and hydrogen peroxide.

a, EIC showed 575, 861, 1147, and 1433 $[m/z]^-$ corresponding to dimers (EC-P1 and EC-P2) timers, tetramers, and pentemers formed from enzymatic reactions but not from control incubations.

b, a carton diagram shows a model by which the incubations form oligomers from the enzymatic reactions of epicatechin.

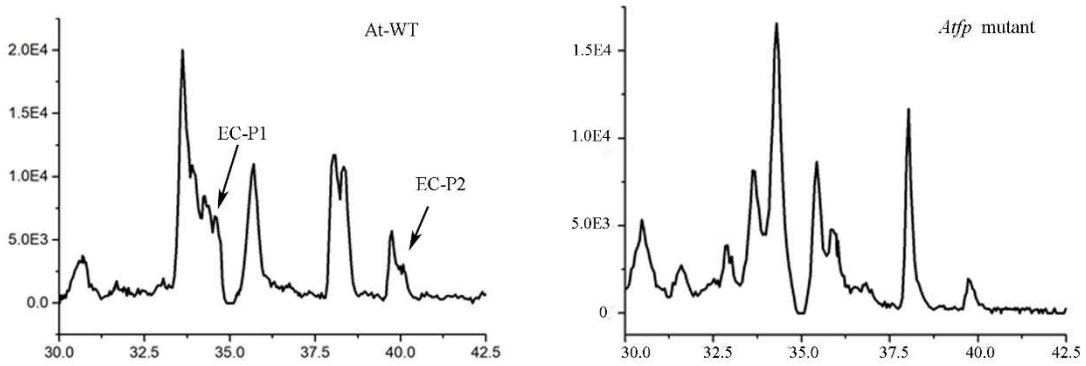


Figure S19 A high resolution snapshot showing EC-P1 and EC-P2 in seeds of wild-type plants but not in seeds of *atfp* mutants.

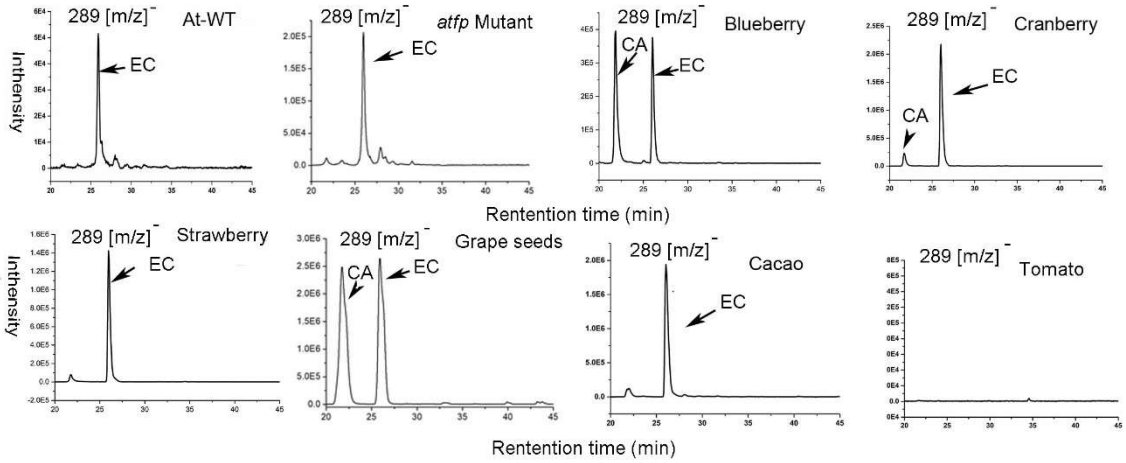


Figure S20 Analysis of epicatechin and catechin in tissues of eight plants.
EIC profiling detected epicatechin (EC) in seeds of *A. thaliana* and *atfp* mutants and berry of blueberry, cranberry, strawberry, grape seeds, and cacao seeds. Catechin (CA) was detected from blueberry, cranberry, and grape seeds. EC and CA were not detected from tomato.

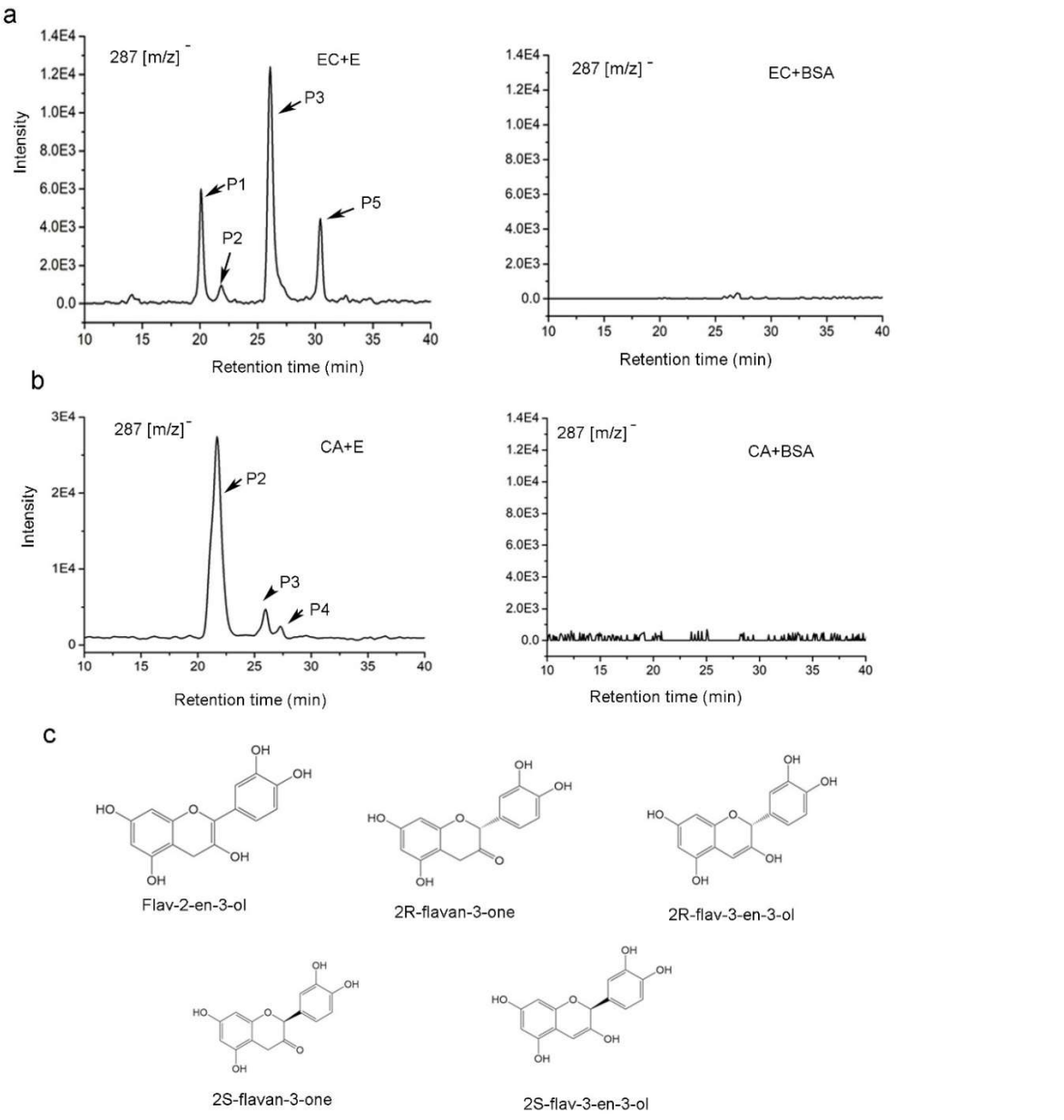


Figure S21 Detection of five 287 [m/z]- compounds from the incubations of NtFP and epicatechin and catechin

a, EIC profiles show that four 287 [m/z]- compounds are formed in the incubations of NtFP, epicatechin (EC), and hydrogen peroxide, but are not formed from the control incubation consisting of NtFP, BSA and EC.

b, EIC profiles show that three 287 [m/z]- compounds are formed in the incubations of NtFP, catechin (CA), and hydrogen peroxide, but are not formed from the control incubation consisting of NtFP, BSA and CA.

c, five structures are predicted to be flav-2-en-3-ol, 2R-flavan-3-one, 2S-flavan-3-one, 2R-flav-3-en-3-ol, and 2S-flav-3-en-3-ol.

CHAPTER 4

Virtual docking characterization and *in vitro* inhibitory activity of flavan-3-ols and dimeric proanthocyanidins against the main protease activity of SARS-Cov-2

Abstract

We report to use the main protease (M^{pro}) of SARS-Cov-2 to screen plant flavan-3-ols and proanthocyanidins. Twelve compounds, (-)-afzelechin (AF), (-)-epiafzelechin (EAF), (+)-catechin (CA), (-)-epicatechin (EC), (+)-gallocatechin (GC), (-)-epigallocatechin (EGC), (+)-catechin-3-O-gallate (CAG), (-)-epicatechin-3-O-gallate (ECG), (-)-gallocatechin-3-O-gallate (GCG), (-)-epigallocatechin-3-O-gallate (EGCG), procyanidin A2 (PA2), and procyanidin B2 (PB2), were selected for virtual docking analysis. The resulting data predicted that all 12 metabolites could bind to M^{pro}. PA2 and PB2 were predicted to have the best affinity scores, -9.2, followed by ECG, GCG, EGCG and CAG, -8.3 to -8.7, and then six flavan-3-ol aglycones, -7.0 to -7.7. Virtual docking analysis predicted that these compounds bound to three or four subsites (S1, S1', S2, and S4) in the binding pocket of M^{pro} via different spatial ways and various formation of one to four hydrogen bonds. *In vitro* analysis with 10 available compounds showed that CAG, ECG, GCG, EGCG, and PB2 inhibited the M^{pro} activity with an IC_{50} value, 2.98±0.21, 5.21±0.5, 6.38±0.5, 7.51±0.21, and 75.3±1.29 μM, respectively, while CA, EC, EGC, GC, and PA2 did not have inhibitory activities. To further substantiate the inhibitory activities, extracts prepared from green tea (GT), two muscadine grapes (MG), cacao, and dark chocolate (DC), which are rich in CAG, ECG, GAG, EGCG or/and PB2, were used for inhibitive assay. The resulting data showed that GT, two MG, cacao, and DC extracts inhibited the M^{pro} activity with an IC_{50} value, 2.84 ±0.25, 29.54 ±0.41, 29.93±0.83, 153.3±47.3, and 256.39 ±66.3 μg/ml, respectively. These findings indicate that on the one hand, structural features of flavan-3-ols are

closely associated with the affinity scores; on the other hand, the galloylation and oligomeric types of flavan-3-ols are critical in creating the inhibitory activity against the M^{pro} activity.

This work was published in Frontiers in Plant Science, 2020.

Yue Zhu, De-Yu Xie. 2020, Docking Characterization and In vitro Inhibitory Activity of Flavan-3-ols and Dimeric Proanthocyanidins against the Main Protease Activity of SARS-Cov-2. Frontiers in Plant Science. 11:601316.

1. Introduction

SARS-CoV-2 is a novel severe acute respiratory syndrome-related coronavirus strain and highly contagious to humans (Fig. 1). Its infection on humans was reported in the early December of 2019 in Wuhan, China [1, 2]. On January 30, 2020, The World Health Organization (WHO) declared the outbreak of the coronavirus disease 2019 (COVID-19) emergency[3]. On March 11, 2020, WHO announced the COVID-19 pandemic [4]. As of Aug. 19, 2020, it has infected approximately 22.3 million people and caused approximately 784,107 deaths worldwide (data source from Johns Hopkins Coronavirus Resource Center). These numbers are continuously increasing every day. Extremely sadly, the humans do not know when this pandemic will end. Although hydroxychloroquine, chloroquine, and remdesivir were recommended for treating COVID-19 [5-7], their therapeutic effects still remain for studies. For example, a new study showed that chloroquine could not treat COVID-19 [8]. Although 300 trials with different anti-virus medicines are being performed [9], such as lopinavir and ritonavir [10], the fact is that no medicines show therapeutic effectiveness. Although in June 2020, dexamethasone, a steroid, was publically reported to decrease the death risk of COVID-19 patients [11, 12], more trials need to be performed to demonstrate its effectiveness. Furthermore, different proteins or genes of SARS-Cov-2, such as the main protease (Fig. 1), have been targeted to screen medicines, however, no small molecules have been conclusively shown to be able to treat COVID-19 patients [13, 14]. Currently, the humans are placing a hope on vaccines. However, we do not have one yet and the effectiveness and potential risks of vaccine remain for continuous investigations. Making matters worse, more studies have shown the originality and the transmission of this contagious virus are more complicated than the humans know. Studies have shown that in addition to aerosol

transmission, this virus can be transmitted through gastrointestinal infection [15, 16] and can stably stay for three hours in air and 72 hours on plastic and steal surfaces [17]. In addition to causing lung diseases, this virus has been found to cause other health complications, such as abdominal pain [16] and neurologic abnormality [18]. To the whole world surprise, one newest study revealed that SARS-Cov-2 existed in wastewater that had been stored in Barcelona, Spain since March 2019 [19], the time of which was nine months earlier than the first report from Wuhan. This finding implies that SARS-Cov-2 might have been transmitted in humans before the outbreak. In summary, no medicines can treat COVID-19 and no vaccines can prevent this contagious disease. Therefore, effective treatments and preventions are urgently needed.

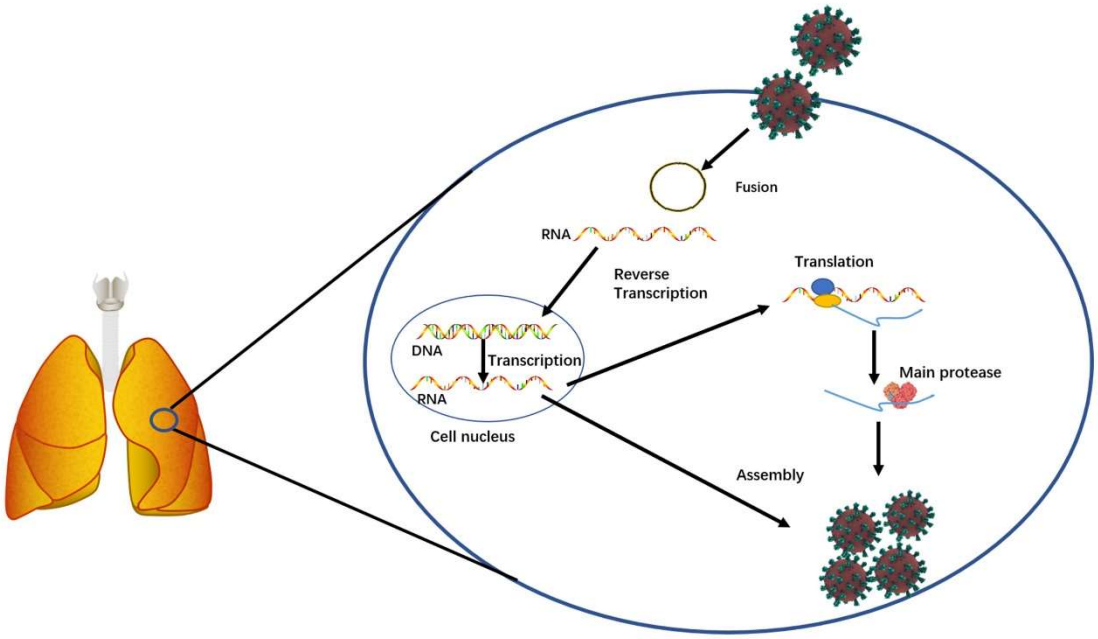


Figure 1 A simplified diagram showing the essential function of the main protease involved in SARS-Cov-2 replication in infected host cells.

Flavan-3-ols and proanthocyanidins (PAs) are two groups of plant flavonoids (Fig. 2) [20]. They commonly exist in fruits, food products, and beverages, such as grape [21-24], strawberry [25-27], persimmon [28], cranberry [29, 30], blueberry [31], cacao nuts [32, 33], chocolate [34, 35], green tea [36-38], and wines [39]. Common flavan-3-ol aglycones in these plant products include (-)-epicatechin (EC), (+)-catechin (CA), (-)-epigallocatechin (EGC), (+)-gallocatechin (GC), (-)-epiafzelechin (EAF), and (+)-afzelechin (AF) (Fig. 1) [20]. Common flavan-3-ol gallates include (-)-epicatechin-3-O-gallate (ECG), (+)-catechin-3-O-gallate (CAG), (-)-gallatechin-3-O-gallate (GCG), (-)-epigallocatechin-3-O-galloate (EGCG), which are highly abundant in green tea [38, 40]. PAs are oligomeric or polymeric flavan-3-ols. In PAs, the lowest and upper units are called as the starter and extension units, which are linked by interflavan bonds formed between the C_8 or C_6 of a lower unit and the C_4 of an upper unit (Fig. 1). In addition, a second linkage is an ether C_2-O-C_7 bond between the starter unit and the extension unit. Based on the linkage numbers, PAs are classified into two types of structures, the dominant B-type characterized with an interflavan bond only and the uncommon A-type featured with an interflavan bond and an ether linkage [20]. Common dimeric B-types in fruits and beverages include procyanidin B1, B2, B3 and B4. Two examples of dimeric A-type PAs are procyanidin A1 and A2 [20]. Flavan-3-ols and PAs are potent antioxidants with multiple benefits to human health [20]. Multiple compounds from these two groups, such as CA, EPC, EGC, EGCG, procyanidin B2, and procyanidin A2 (Fig. 2), have been shown to have antiviral function [41, 42], antibacterial activity [43, 44], anticancer [45, 46], anti-cardiovascular diseases [47-49], and anti-aging diseases [50-52]. All these features suggest that flavan-3-ols and PAs are appropriate targets for screening potential anti-SARS-Cov-2 medicines.

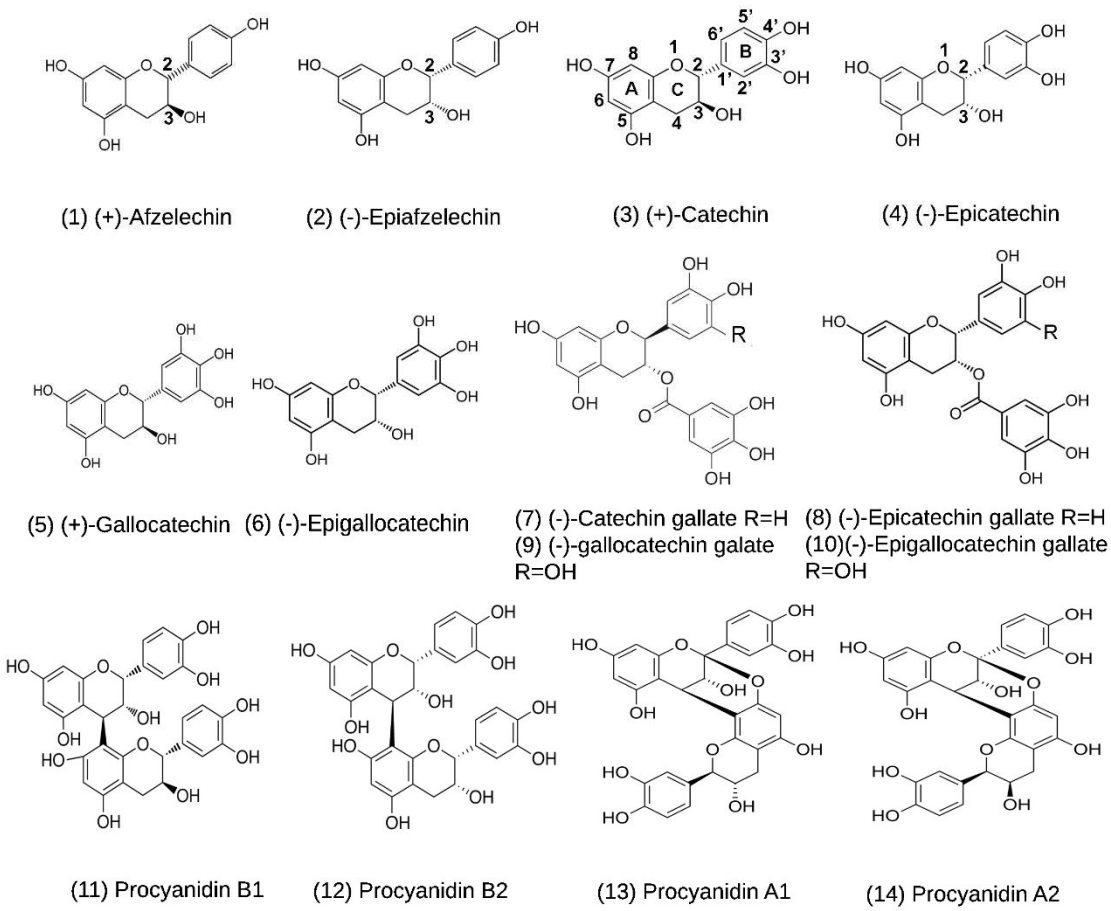


Figure 2 Structures of (+)-afzelechin, (-)-epiafzelechin, (+)-catechin, (-)-epicatechin, (+)-gallocatechin, (-)-epigallocatechin, (-)-catechin-3-O-gallate, (-)-epicatechin-3-O-gallate, (-)-gallocatechin-3-O-gallate, and (-)-epigallocatechin-3-O-gallate, procyanidin B1 and B2, and procyanidin A1 and A2.

In this study, we performed virtual studies to screen flavan-3-ol and PA to identify potential active compounds and characterized virtual binding similarity and difference among different structures. Next, based on positive virtual screening results, we used the M^{pro} of SARS-Cov-2 as a key target to perform *in vitro* inhibitory experiments. Five compounds were identified to have anti-M^{pro} activity. Crude extracts from early spring green tea, cacao, and two muscadine grapes, which are rich in these compounds, also showed inhibitory effects on the M^{pro} activity.

2. Results

2.1 Virtual compound binding to M^{pro}

Theoretical compound screening for inhibitors of the SARS-Cov-2 M^{pro} is an effective approach to identify potential candidates that can be used for trials to test their inhibitory activity against SARS-Cov-2 [13, 14, 53-55]. To understand whether flavan-3-ol aglycones, their gallates, and dimeric PAs (Fig. 1) could inhibit the M^{pro} activity, we performed docking analysis via two publically available software, Dock Prep tool of UCSF-Chimera and AutoDock vina. The resulting docking data showed that all flavan-3-ol aglycones, flavan-3-ol gallates, and dimeric procyanidins tested (PA2 and PB2) (Fig. 1) could bind to the same location of M^{pro}, which is bound by N3 inhibitor (Fig. 3 a-f). In addition, two antiviral control compounds, ebselen and cinanserin, bound to the same binding pocket (Fig. 3 g and h). In contrast, the antiviral lopinavir bound to a different location (Fig. 3 i).

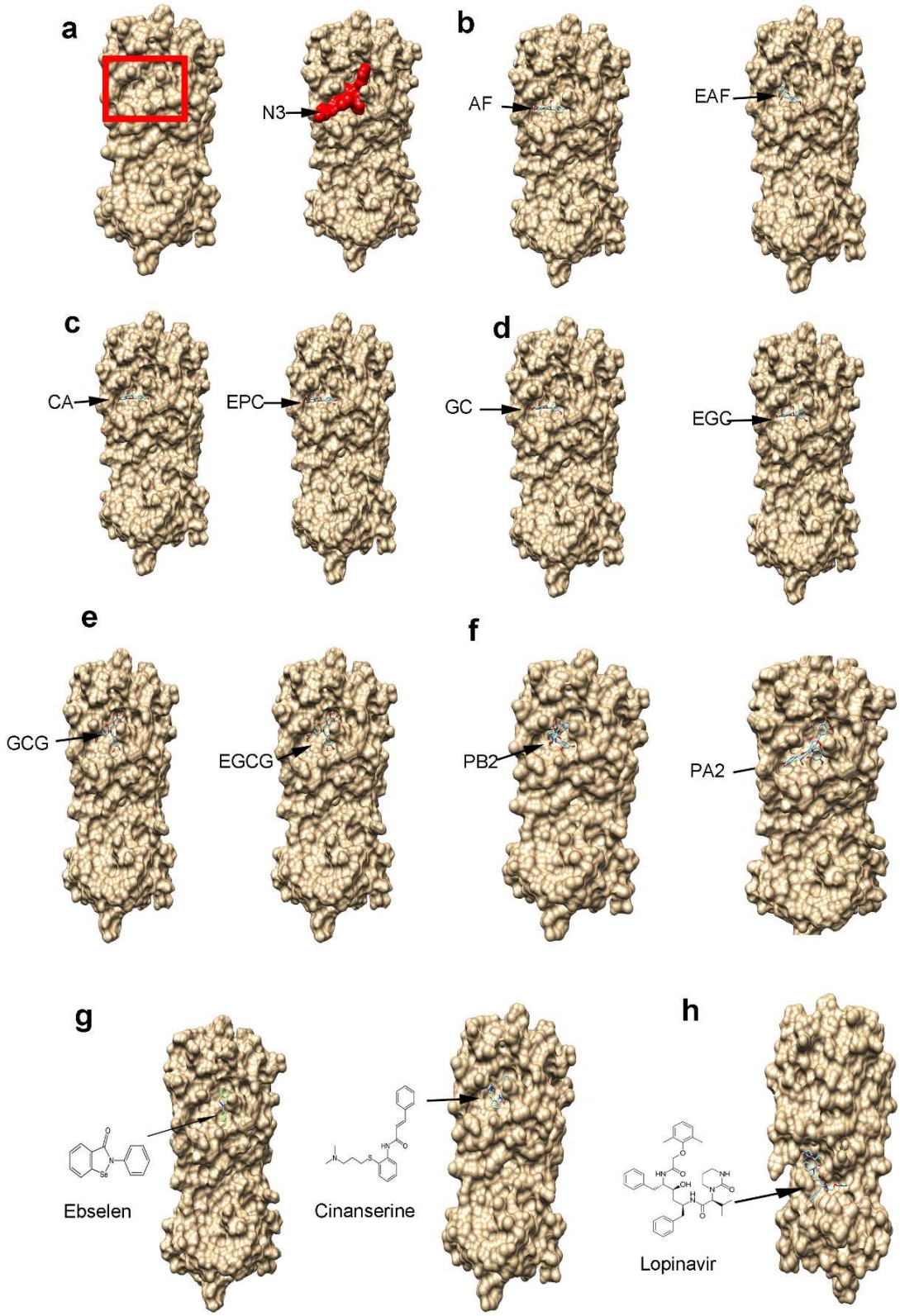
Figure 3 The binding of flavan-3-ols, procyanidin B2, and three potential anti-SARS-Cov-2 drugs to the main protease (M^{pro}) shown by protein-ligand docking.

a, the 3D surface view of the SARS-Cov-2 M^{pro} without and with the inhibitor peptide N3. The red rectangle frame shows the M^{pro} substrate-binding pocket and the red amino acid residues show the N3 binding to the M^{pro} substrate-binding pocket.

b-f, 10 compounds bind to the M^{pro} substrate-binding pocket, (+)-afzelechin (AF) and (-)-epiafzelechin (EAF) (b), (+)-catechin (CA) and (-)-epicatechin (EPC) (c), (+)-gallocatechin (GC) and (-)-epigallocatechin (EGC) (d), (-)-gallocatechin gallate (GCG) and (-)-epigallocatechin gallate (EGCG) (e), and procyanidin B2 and A2 (f).

g, ebselen (EBS) and cinanserin (INN) bind to the M^{pro} substrate-binding pocket.

h, lopinavir (LOP) binds to M^{pro} at a different position.


a
N3
b
AF
EAF
c
CA
EPC
d
GC
EGC
e
GCG
EGCG
f
PB2
PA2
g
Ebselen
Cinanserine
h
Lopinavir

The resulting affinity scores were -7.0 to -7.7 for six flavan-3-ol aglycones, -8.3 to -8.7 for four flavan-3-ol gallates, and -9.2 for two dimeric proanthocyanidins (Table 1). All these score values were lower than those of ebselen and cinanserin, -6.6 and -5.4. The affinity scores of the dimeric procyanidins and flavan-3-ol-gallates were lower than -8.0 for lopinavir. These data suggested that flavan-3-ols, flavan-3-ol gallates, and dimeric PAs were appropriate candidates with inhibitory effects on the M^{pro} activity.

Table 1 The affinity scores of AF, EAF, CA, EPC, GC, EGC, GCG, EGCG, PA2 and PB2 and three putative anti-COVID-19 molecules binding to the main protease.

Molecules	Affinity score (kcal/mol)
Procyanidin B2 (PB2)	-9.2
Procyanidin A2 (PA2)	-9.2
(-)-epigallocatechin-3-O-gallate (EGCG)	-8.7
(-)-gallocatechin-3-O-gallate (GCG)	-8.7
(-)-epicatechin-3-O-gallate (ECG)	-8.7
(+)-catechin-3-O-gallate (CAG)	-8.3
(-)-epigallocatechin (EGC)	-7.7
(+)-gallocatechin (GC)	-7.6
(-)-epicatechin (EPC)	-7.5
(+)-catechin (CA)	-7.5
(-)-epiafzelechin (EAF)	-7.5
(-)-afzelechin (AF)	-7.0
Lopinavir (LOP)	-8.0
Ebselen (EBS)	-6.6
Cinanserin (CIN)	-5.4

2.2 Spatial binding characteristics at different subsites of M^{pro}

The substrate binding pocket of M^{pro} has four subsites, S1', S1, S2, and S4 [13]. Substrate-binding analysis was completed for ebselen (positive control) and 14 compounds and (Fig. 1). The results predicted different binding features among ebeslen, six flavan-3-ol aglycones, four flavan-3-ol-gallates, and two dimeric PAs. The two benzene rings of ebselen were predicted to

face S1 and S1’ (Fig. 5 a), while the 14 compounds were predicted to face three or four subsites (Fig. 5 b-e). The A-ring and B-ring of EGC were predicted to face S2/S4 and S1, respectively (Fig. 5 b). The same patterns were predicted for AF, EAF, CA, EPC, and GC. The A-ring, B-ring, and the gallate ester group of EGCG were predicted to face to S2, S1’ and S1, respectively (Fig. 5 c). The same patterns were obtained for CAG, ECG, and GCG. In PA1 and PA2 binding, the A-ring and B-ring of the starter unit and the B-ring of the upper unit were predicted to face to S1, S4, and S1’, respectively (Fig. 5 d). In PB1 and B2 docking modeling, the B-ring of the starter unit and the A-ring and B-ring of upper unit were predicted to face S1’, S1 and S2, respectively (Fig. 5 e);

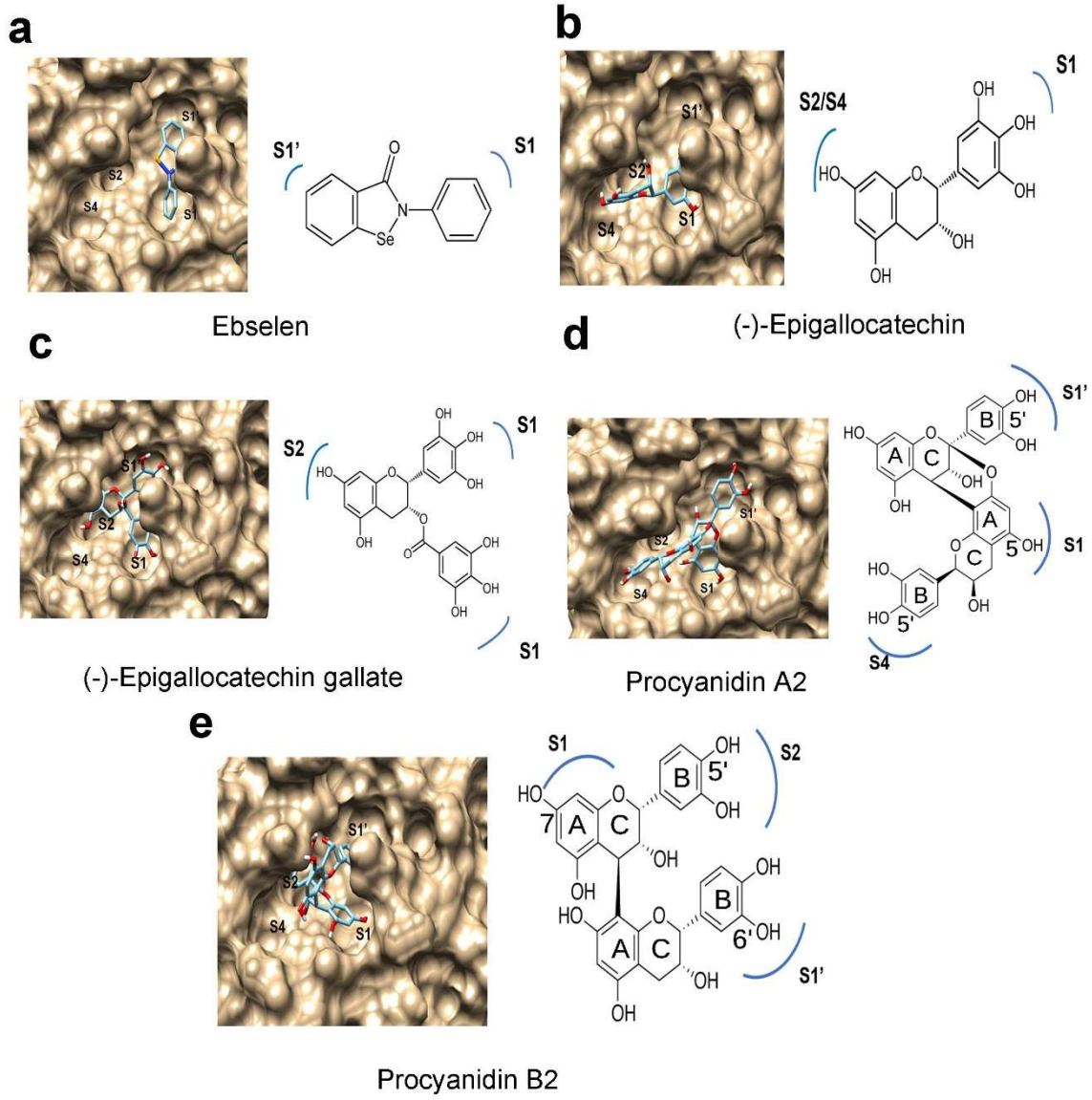


Figure 4 Features of ebselen, (-)-epigallocatechin, (-)-epigallocatechin 3-O-gallate (EGCG), procyanidin A2 (PA2), and procyanidin B2 (PB2) binding to four subsites in M^{pro} predicted by protein-ligand docking modeling.

a, ebselen binds to S1 and S1' subsites.

b, (-)-epigallocatechin binds to the M^{pro} S1, S2/S4 subsites.

c, (-)-epigallocatechin 3-O-gallate binds to the M^{pro} S1, S1', and S2 subsites,

d, procyanidin A2 binds to the M^{pro} S1', S1, and S4 subsites.

e, procyanidin B2 binds to the M^{pro} S1, S1', and S2 subsites.

2.3 Potential hydrogen bonds between compounds and the binding pocket of M^{pro}

This docking analysis predicted that the binding of these compounds to the substrate pocket was via the formation of hydrogen bonds. Based on the best affinity score of each compound (Table 1), the modeling analysis could predict the most potential hydrogen bonds formed between each compound and the binding pocket of M^{pro}. The resulting data showed different features of the hydrogen bond number and potential linkage positions between ligand compounds and amino acids of M^{pro} (Fig 5 a-j). Three hydrogen bonds were predicted to be formed between AF, EAF, CA, or EC and M^{pro} (Fig. 5 a-d). The linkages of those three hydrogen bonds were formed between AF and M^{pro} via O_1-Glu_{166}, C_3-O-Glu_{166}, and C_7-O-Leu_{141} (Fig. 5 a), between EAF and M^{pro} via O_1-Glu_{166}, C_7-O-Leu_{141}, and $C_{4'}$-O-Gln_{189} (Fig. 5 b), between CA and M^{pro} via O_1-Glu_{166}, C_3-O-Glu_{166}, and $C_{4'}$-O-Leu_{141} (Fig. 5 c), and between EC and M^{pro} via O_1-Glu_{166}, C_7-O-Thr_{190}, and $C_{4'}$-O-Gln_{189} (Fig. 5 d). GC, EGC, and GCG were predicted to form one hydrogen bond with M^{pro} via O_1-Glu_{166} (Fig. 5 e-f). EGCG and M^{pro} were predicted to form four hydrogen bonds via O_1-Glu_{166}, C_7-O-Thr_{190}, $C_{5'}$-O-Phe_{140}, and gallate-3-O-Gly_{143} (Fig. 5 g). Procyanidin A1 and M^{pro} were predicted to form one hydrogen bond via $C_{3'}$-O-Gly_{143}. Procyanidin B2 and M^{pro} were predicted to form three hydrogen bounds via C_5-O-Gly_{143} and G_5-O-Cys_{145} on the B ring of the upper unit and C_3-O-Glu_{166} on the C-ring of the starter unit. Furthermore, all hydrogen-binding distances were different from 1.833Å between PA2 and M^{pro} (Fig. 5 i) to 2.541Å between EGC and M^{pro} (Fig. 5 f).

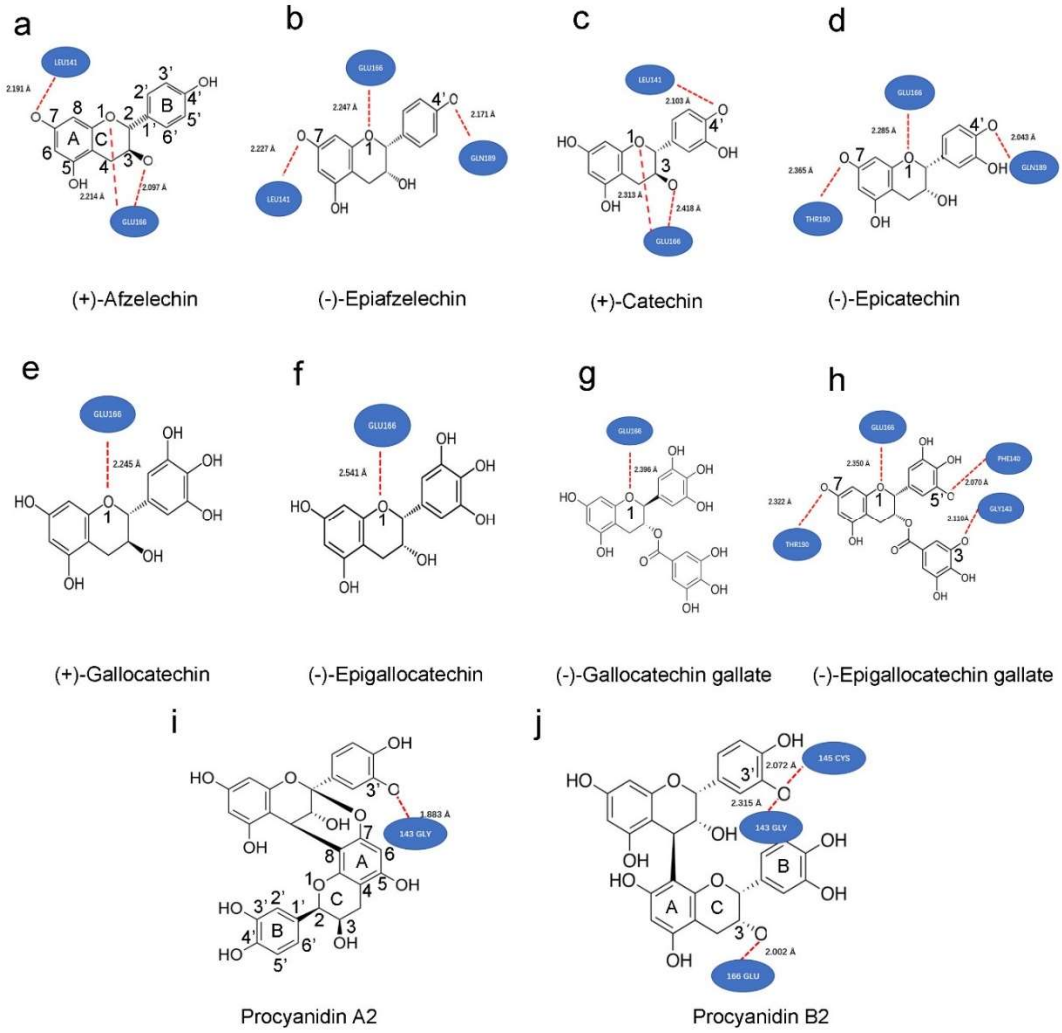


Figure 5 Hydrogen bonds formed between M^{pro} and 10 compounds predicted by ligand-protein modeling. These virtual bond positions and numbers formed between 10 various metabolites and the binding pocket of M^{pro} were predicted based on the best affinity scores of each compound (Table 1).

a-d, (+)-afzelechin (AF) (a), (-)-epiafzelechin (b), (+)-catechin (CA) (c), and (-)-epicatechin (d) were predicted form three hydrogen bonds with M^{pro}. The hydrogen patterns were the same between AF and CA, while those were the same between EAF and EC.

e-g, (+)-gallocatechin (e), (-)-epigallocatechin (f), and (-)-gallocatechin gallate (g) were predicted to form one hydrogen with M^{pro}.

h, (-)-epigallocatechin gallate was predicted to form three hydrogen bonds with M^{pro}.

i-j, procyanidin A2 (i) and B2 (j) were predicted to form one and three hydrogen bonds with M^{pro}.

2.4 In vitro inhibitory effects of ten compounds on the M^{pro} activity

To understand the effects of these compounds on the M^{pro} activity, we used CA, EPC, GC, EGC, CAG, GCG, EGCG, PA2, and PB2 to perform *in vitro* inhibition assay. The resulting data showed that CAG, ECG, GCG, EGCG, and PB2 inhibited the M^{pro} activity (Fig. 6). The values of the half maximal inhibitory concentration (IC_{50}) for CAG, ECG, GCG, EGCG, and PB2 were approximately 2.98±0.21, 5.21±0.5, 6.38±0.5, 7.51±0.21, and 75.3±1.29 μM, respectively (Fig. 6 a-e). The resulting data showed when the concentrations of these compounds were increased, such as 100 μM, the inhibition of PB2 was more effective than that of CAG, GCG, and EGCG (Fig. 6 f). In contrast, the in vitro analysis did not detect that CA, EPC, GC, EGC, and PA2 in a range of concentrations from 0-500 μM tested could inhibit the M^{pro} activity (Fig. 6 f and Fig. 7 a-e).

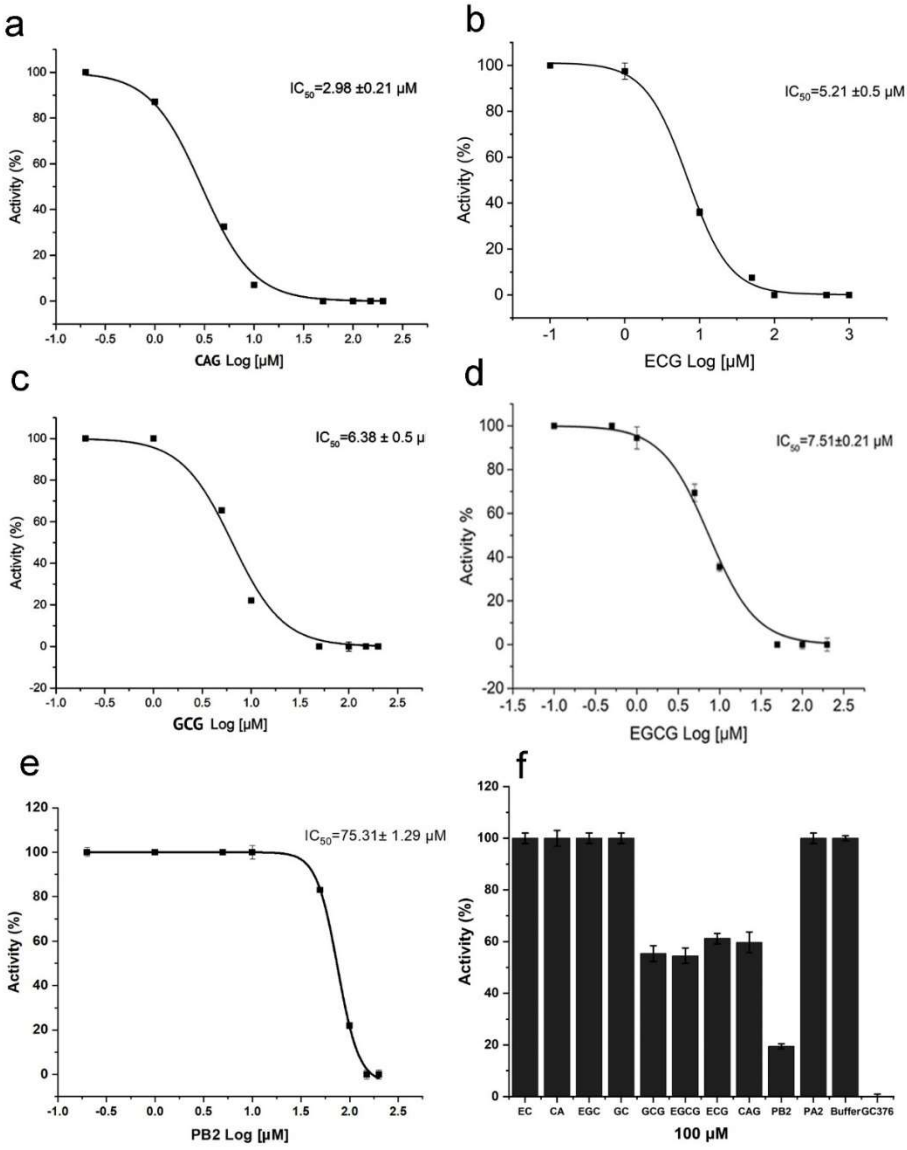


Figure 6 Inhibitory effects of four four-3-ol-gallates, four flavan-3-ol aglycones, procyanidin A2 (PA2), and procyanidin B2 (PB2) on the M^{pro} activity of SARS-COv-2.

a-e, dynamic curves show that (+)-catechin-3-O-gallate (CAG) (a), (-)-epicatechin-3-O-gallate (ECG) (b), (-)-gallocatechin-3-O-gallate (GCG) (c), (-)-epigallocatechin-3-O-gallate (EGCG) (d), and PB2 (e) inhibit the activity of M^{pro} with an IC_{50} 2.98±0.21 μM, IC_{50} 5.21±0.5 μM, IC_{50} 6.38±0.5 μM, IC_{50} 7.51±0.21 μM, and IC_{50} 75.31±1.29 μM, respectively.

f, inhibitory effects of 100 μM (-)-epicatechin (EC), (+)-catechin (CA), EGC, CAG, ECG, GCG, EGCG, PA2, and PB2 on the activity of M^{pro}. Compared with the negative control, the activity of M^{pro} is inhibited by 40%, 40%, 44.4%, 50%, 81.5%, and 100% by 100 μM CAG, ECG, EGCG, GCG, PB2, and positive control GC376.

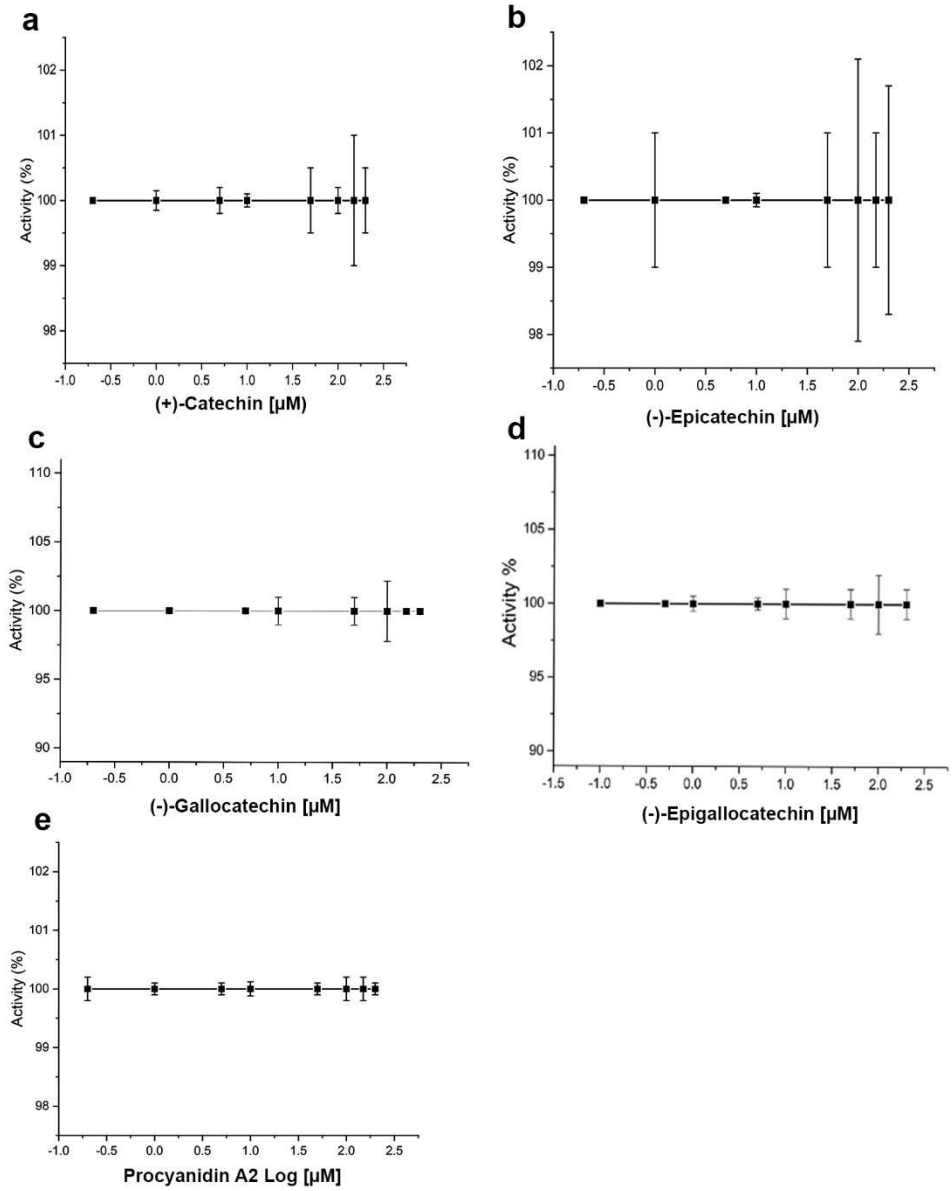


Figure 7 No inhibitory effects of four flavan-3-ols and procyanidin A2 on the activity of M^{pro}. Eight concentrations (0-200 μM) tested for (+)-catechin (a), (-)-epicatechin (b), (-)-gallocatechin (c), (-)-epigallocatechin (d), and procyanidin A2 (e) did not show inhibitory effects on the activity of M^{pro}.

2.5 In vitro inhibitory effects of plant extracts on the M^{pro} activity

Green tea [37, 38], muscadine grape berry (including seeds and skin) [24], and dark chocolate and cacao [56, 57] are rich in flavan-3-ol gallates and dimeric PAs. To understand whether these plant products could inhibit the M^{pro} activity of SARS-Cov-2, we extracted numerous samples

from these four products and completed *in vitro* assays. The resulting data showed that the M^{pro} activity was inhibited by all extracts of these products (Fig. 8 a-f)). The green tea extracts showed the highest inhibitory activity with an IC_{50} 2.84±0.25 µg/ml. At 10 µg/ml, the green tea extract tea completely inhibited the M^{pro} activity (Fig. 6 a). The high content of EGCG, relatively high contents of ECG, and appropriate contents of dimeric procyanidin B1, B2, B3, and B4 in green tea [37, 38, 58] were associated with the inhibitory activity. The extracts of two muscadine grapes FLH 13-11 and FLH17-66 inhibited the M^{pro} activity with an IC_{50} 29.54±0.41 µg/ml and an IC_{50} 29.93±0.83 µg/ml. At 100 µg/ml, the extracts of two muscadine grapes completely inhibited the M^{pro} activity (Fig. 6 b, c, and f). EGCG, ECG, and procyanidin B1-B3 and B5-B8 in the extracts [24] were associated with the inhibitory activity. Cacao and dark chocolate extracts were shown to inhibit the M^{pro} activity when the extract concentrations tested were higher than 10 µg/ml and their IC_{50} values were 153.3 ±47.3 µg/ml and 256.39 ±66.3 µg/ml, respectively (Fig. 8 d-e). ECG, EGCG, and PB2 in the cacao and dark chocolate extracts [56, 57] were associated with the inhibitory activity. We further compared the inhibitory activity of all extracts at 100 µg/ml tested, the results showed that the extracts of green tea and two muscadine grapes completely inhibited the M^{pro} activity and the extracts of cacao and dark chocolate reduced the M^{pro} activity by 40-50% (Fig. 6 f).

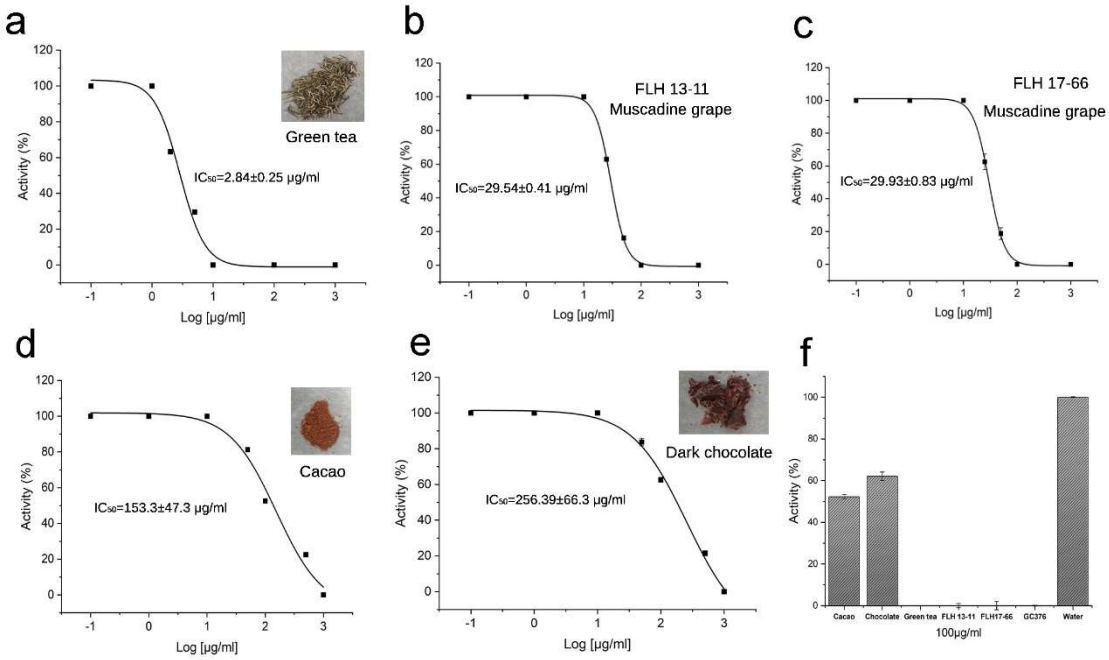


Figure 8 Inhibitory effects of five extracts on the M^{pro} activity.

a-e, dynamic curves show inhibitory activities of different concentrations of extracts from green tea with IC_{50} 2.84 ±0.25 μg/ml (a), FLH13-11 muscadine berry IC50 29.54 ±0.41 μg/ml (b), FLH17-66 berry with IC_{50} 29.93 ±0.83 μg/ml (c), cacao with IC_{50} 153.3 ±47.3 μg/ml (d), and dark chocolate IC_{50} 256.39 ±66.3 μg/ml (e).

f, inhibition of the M^{pro} activity by 100 μg/ml extracts of green tea, cacao, chocolate, FLH13-11 muscadine berry, and FLH17-66 muscadine berry. GC376 (100 μg/ml) and 1% DMSO in water were used as positive and negative controls. All values are averaged from five replicates.

3. Discussion

Both virtual docking and *in vitro* assays showed that the stereo configurations, galloylation, and oligomeric types of flavan-3-ols affected the ligand-protein binding features and inhibitory activity. Flavan-3-ols and PAs are two groups of plant flavonoids [20]. PAs are oligomeric or polymeric flavan-3-ols. Flavan-3-ol aglycones have four different stereo configurations at C_2 and C_3 on the C-ring (Fig. 2). For example, (+)-Catechin at C_2 and C_3 has a *2R, 3S*-2, 3-trans configuration (Fig. 2 c). Its isomer (-)-epicatechin has a *2R, 3R*-2, 3-cis configuration. (-)-

Catechin and (+)-epicatechin have a *2S, 3R*-2, 3-trans and a *2S, 3S*-2, 3-cis configuration, respectively [20]. In addition, the number of hydroxyl groups on the B-ring, such as one on AF, two on CA, and three on GA (Fig. 2), diversifies structures. We used both virtual docking and *in vitro* assay to screen reprehensive flavan-3-ol aglycones, flavan-3-ol gallates, and dimeric PAs. Six flavan-3-ol aglycones, including AF, EAF, CA, EC, GC, and EGC (Fig. 2), were used for virtual docking analysis. All were predicted to be able to bind to the binding pocket of M^{pro} (Fig. 3). It was interesting that the affinity scores decreased as the number of hydroxyl groups on the B-ring of flavan-3-ols increased (Table 1). This datum indicates that more hydroxygroups on the B-ring increase the binding capacity to M^{pro}. Furthermore, the docking analysis showed that the galloylation at C_3-OH of CA, EC, GC, and EGC and dimerization of EC decreased the affinity scores (Table 1), indicating that the galloylation and oligomerization could increase the binding capacity to M^{pro}. Based on the virtual docking and affinity scores, we hypothesized that 12 compounds (Fig.1) could have an inhibitory activity against the M^{pro} activity. However, our *in vitro* assays only detected the positively inhibitory effects of galloylated flavan-3-ols, including CAG, ECG, GCG, and EGCG, on the M^{pro} activity (Fig. 6). However, no inhibitory activity was detected in the assays of CA, EC, GC, and EGC (Fig. 7), although their affinity scores were from -7.5 to -7.7. These data indicate that positive virtual docking data are useful to screen candidates, however, *in vitro* test experiments are necessary to show an activity. Furthermore, these data show that the galloylation of these flavan-3-ols create the inhibitory activity against M^{oro}. In addition, it was interesting that although the affinity scores of PA2 and PB2 were the same, -9.2, PB2 but not PA2 showed an inhibitory activity against M^{oro} (Fig. 6 e and Fig. 7). This datum indicates that the dimeric PA types are closely associated with the inhibitory activities.

Although it was difficult to test more flavan-3-ol gallates and dimeric PAs due to the unavailability of compounds, we could take advantage of natural sources to understand the inhibitory effects of more flavan-3-ol gallate and dimeric PA diversity on the M^{pro} activity. We previously reported that green tea (GT) produced not only EGCG and PB2 but also ECG, GCG, CAG, procyanidin B1 (PB1), PB3, and PB4 [37, 38, 58]. We prepared extracts from GT and tested their activity. The resulting inhibitory activity was effective and the concentration at 10 μg/ml could completely inhibit the M^{pro} activity (Fig. 8 a). These data suggest that in addition to ECG, GCG, CAG, EGCG, and PB2 tested, PB1, PB3, and PB4 may provide additive inhibitory activity against the M^{pro} activity. We also reported that the berries (including seeds and peels) of two muscadine grapes, FLH 13-11 FL and FLH 17-66 FL, produced four flavan-3-ol aglycones, 18 galloylated or glycosylated conjugates, and eight dimeric procyanidins [24]. The extracts of these two muscadine grapes showed an effective inhibitory activity against the M^{pro} activity (Fig. 8 b and c), indicating that those untested dimeric PAs and flavan-3-ol conjugates might enhance the inhibitory activity. We recently analyzed cacao nut and DW extracts and identified PB1 and PB2 in the extracts. These two extracts also showed appropriately inhibitory activity against the M^{pro} activity. Based on these flavan-3-ol, flavan-3-ol gallate, and dimeric PA-containing natural resources, we hypothesize that those flavan-3-ol gallates, dimeric PAs, and oligomeric PAs untested have inhibitory effects on the M^{pro} activity. It is valuable that continuous screening of more flavan-3-ol gallates and other derivatives as well as PAs can enhance identifying effective compounds for anti-SARS-Cov-2 efforts.

4. Materials and methods

4.1 Candidates of flavan-3-ols and PAs for virtual analysis

Main flavan-3-ol aglycones include (+)-afzelechin, (-)-epiafzelechin, (+)-catechin (CA), (-)-epicatechin (EC), (+)-gallocatechin (GC), and (-)-epigallocatechin (EGC). Galloylated flavan-3-ol conjugates include (-)-epicatechin-3-O-gallate (ECG), (-)-catechin-3-O-gallate, (-)-gallocatechin-3-O-gallate (GCG), and (-)-epigallocatechin-3-gallate (EGCG). Dimeric proanthocyanidins include procyanidin A1, A2, B1, and B2. Fourteen structures are listed in Figure 1. In addition, three known anti-viral medicines, ebselen, cinanserin, and lopinavir, were used as positive metabolite control.

4.2 Plant materials

Three types of plant products rich in flavan-3-ols and dimeric PAs were used for extraction. "*Xin-Yang-Mao-Jian*" is one of famous green tea (*Camellia sinensis*) products in China. This product is composed of newly leaves (0.8-1.2 cm in length) of early spring sprouts (harvested around April 5 every year). This green tea product was produced in 2019. Cacao (*Theobroma cacao*) seed powder used were obtained from Ecuador in 2019. Ripen muscadine grape berries of FLH 13-11 and FLH 17-66 were collected in 2011 and 2012, ground into powder in liquid nitrogen, freeze dried, and stored in -80ºC freezer [24].

4.3 M^{pro} docking analysis

To understand whether flavan-3-ol aglycones, gallates, and dimeric PA molecules (Fig. 2) could have anti-SARS-Cov-2 activity, we used M^{pro} to perform docking analysis via two publically

available software, Dock Prep tool of UCSF-Chimera (https://www.cgl.ucsf.edu/chimera/docs) and AutoDock vina (http://vina.scripps.edu/). In addition, three reported potential anti-COVID-19 candidates, ebselen [59], cinanserin [60, 61], and lopinavir [10], were used as positive controls for virtual docking. We used four steps to complete docking. First, we obtained a SARS-Cov-2 M^{pro} (PDB ID: 6LU7) structure curated in the Protein Data Bank (https://www.rcsb.org/), which consisted of M^{pro} and an inhibitor N3 (Fig. 2 a-b) [59]. Second, we used the Dock Prep tool of UCSF-Chimera to perform receptor (M^{pro}) preparation. During the receptor preparation, the inhibitor N3 was removed. In addition, hydrogens were added and receptor charge was optimized, which allowed determining the histidine protonation state. The resulting M^{pro} structure file was saved in the mol2 format. Third, the 3D structures of flavan-3-ols and other interesting compounds used in this study were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/) and used as ligands. All ligand structures were then uploaded to the Minimize structure tool of UCSF-Chimera and minimized, during which charges and hydrogens were added to each ligand. The resulting ligand files were also saved in the mol2 format. Fourth, both M^{pro} and ligand files were uploaded to AutoDock vina for docking. During modelling, the M^{pro} protein was framed in one box for receptor-ligand docking. The entire box size was x=50, y=55, and z=50 and the origin center of the box was at x=-27, y=13, and z=58.

4.4 Extraction of flavan-3-ols and PAs from green tea, cacao, and muscadine berry powders

One gram of dry powder was suspended in 10 ml acetone: deionized water (70:30) contained in a 50 ml falcon tube. The tube was strongly vortexed for 5 min and then sonicated in a water bath for 10 min at the room temperature, followed by centrifugation at 4000 rpm for 20 min. The resulting supernatant was pipetted to a new 50 ml tube. The remained pellet was extracted once

with the same steps. Two extractions were pooled together to obtain 20 ml in volume, which was reduced to 5.0 ml by removing acetone with a nitrogen gas flow at the room temperature. The tube containing the remained water phase was added 1.0 ml chloroform and then strongly vortexed for 2.0 min, followed by centrifugation at 4000 rpm for 5 min. The resulting lower chloroform phase containing non-polar compounds was removed. This step was repeated once. The tube that contained the remained water phase including flavan-3-ols and PAs was added 5.0 ml ethyl acetate, strongly vortexed for 3.0 min, and then centrifuged at 4000 rpm for 5.0 min. The resulting upper ethyl acetate phase containing flavan-3-ols and PAs was transferred to a new 50 ml tube. This ethyl acetate extraction step was repeated two times. The three times of ethyl acetate extractions were pooled together and then dried completely with nitrogen flow at the room temperature. The remained residue was dissolved in DMSO to obtain 50 mg/ml extracts and stored in a -20ºC freezer until use.

4.5 In vitro inhibition assay of M^{pro} activity

CA, EC, GC, EGC, CAG, GAG, ECG, EGCG, PA2, and PB2 were purchased from Sigma-Aldrich (St. Louis, MO) and dissolved in DMSO to prepare 1.0 mM stock solution. A 3CL Protease (M^{pro}) (SARS-CoV-2) Assay Kit (BPS bioscience, https://bpsbioscience.com/) was used to test the inhibitory activity of the ten compounds. The steps of *in vitro* assay followed the manufacturer's protocol. In brief, each reaction was completed in a 25 µl volume in 96-well plates. Each reaction solution contains 150 ng recombinant M^{pro} (the final concentration in the reaction, 6 ng/ µl), 1.0 mM DDT, 50 µM fluorogenic substrate, and CA, EPC, GC, EGC, CAG, GAG, ECG, EGCG, PA2, PB2 (final concentrations, 0, 0.1, 0.5, 1, 5, 10, 50, 100, 150, and 200 µM) or one plant extract (final concentrations, 0, 1, 10, 100, 1000 µg/ml) in pH 8.0 50 mM Tris-

HCl and 5.0 μM EDTA buffer. GC376 (50 μM) was used as positive control and Tris-HCl-EDTA buffer was used as negative control. The reaction mixtures were incubated for 4.0 hours at the room temperature. The fluorescence intensity of each reaction was measured and recorded on a microtiter plate-reading fluorimeter (BioTek's Synergy H4 Plate Reader for detect fluorescent and luminescent signals). The excitation wavelength was 360 nm and the detection emission wavelength was 460 nm. Each concentration of compounds and extracts was tested five times. A mean value was calculated using five replicates. One way analysis of variance (ANOVA) was performed to evaluate the statistical significance. The P-value less than 0.05 means significant differences.

REFERENCES

1. Wu, Y.-C., C.-S. Chen, and Y.-J. Chan, *The outbreak of COVID-19: An overview.* Journal of the Chinese Medical Association, 2020. **83**(3): p. 217-220.

2. Wang, C., et al., *A novel coronavirus outbreak of global health concern.* Lancet (London, England), 2020. **395**(10223): p. 470-473.

3. WHO, *Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)*. 2020, World Health Organization: Geneva, Switzerland.

4. WHO, *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. 2020, World Health Organization.

5. Wang, M., et al., *Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro.* Cell research, 2020. **30**(3): p. 269-271.

6. Colson, P., J.-M. Rolain, and D. Raoult, *Chloroquine for the 2019 novel coronavirus SARS-CoV-2.* International journal of antimicrobial agents, 2020. **55**(3): p. 105923-105923.

7. Sehailia, M. and S. Chemat, *In-silico studies of antimalarial-agent artemisinin and derivatives portray more potent binding to Lys353 and Lys31-binding hotspots of SARS-CoV-2 spike protein than hydroxychloroquine: Potential repurposing of artenimol for COVID-19.* Preprints (chemrxiv.org), 2020.

8. Hoffmann, M., et al., *Chloroquine does not inhibit infection of human lung cells with SARS-CoV-2.* Nature, 2020.

9. Sanders, J.M., et al., *Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review.* JAMA, 2020. **323**(18): p. 1824-1836.

10. Cao, B., et al., *A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19.* The New England journal of medicine, 2020. **382**(19): p. 1787-1799.

11. Selvaraj, V., et al., *Short-Term Dexamethasone in Sars-CoV-2 Patients.* R H O D E I S L A N D M E D I C A L J O U R N A L, 2020. **103**: p. 39-43.

12. Khan, S.U. and T.-T. Htar, *Deciphering the binding mechanism of Dexamethasone against SARS-CoV-2 Main Protease: Computational molecular modelling approach.* chemrxiv.org, 2020.

13. Dai, W., et al., *Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease.* Science, 2020. **368**(6497): p. 1331-1335.

14. Zhang, L., et al., *Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors.* Science, 2020. **368**(6489): p. 409-412.

15. Xiao, F., et al., *Evidence for Gastrointestinal Infection of SARS-CoV-2.* Gastroenterology, 2020. **158**(6): p. 1831-1833.e3.

16. Lamers, M.M., et al., *SARS-CoV-2 productively infects human gut enterocytes.* Science, 2020. **369**(6499): p. 50-54.

17. van Doremalen, N., et al., *Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1.* New England Journal of Medicine, 2020. **382**(16): p. 1564-1567.

18. Helms, J., et al., *Neurologic Features in Severe SARS-CoV-2 Infection.* New England Journal of Medicine, 2020. **382**(23): p. 2268-2270.

19. Chavarria-Miró, G., et al., *Sentinel surveillance of SARS-CoV-2 in wastewater anticipates the occurrence of COVID-19 cases.* medRxiv, 2020: p. 2020.06.13.20129627.

20. Xie, D.-Y. and R.A. Dixon, *Proanthocyanidin biosynthesis - still more questions than answers?* Phytochemistry, 2005. **66**(18): p. 2127-2144.

21. Zhu, Y., et al., *Molecular cloning and functional characterization of the anthocyanidin reductase gene from Vitis bellula.* Planta, 2014. **240**(2): p. 381-398.

22. Rousserie, P., A. Rabot, and L. Geny-Denis, *From Flavanols Biosynthesis to Wine Tannins: What Place for Grape Seeds?* Journal of Agricultural and Food Chemistry, 2019. **67**(5): p. 1325-1343.

23. Iacopini, P., et al., *Catechin, epicatechin, quercetin, rutin and resveratrol in red grape: Content, in vitro antioxidant activity and interactions.* J. Food Compos. Anal., 2008. **21**(8): p. 589-598.

24. Yuzuak, S., J. Ballington, and D.Y. Xie, *HPLC-qTOE-MS/MS-Based Profiling of Flavan-3-ols and Dimeric Proanthocyanidins in Berries of Two Muscadine Grape Hybrids FLH 13-11 and FLH 17-66.* Metabolites, 2018. **8**(4).

25. Fossen, T., S. Rayyan, and O.M. Andersen, *Dimeric anthocyanins from strawberry (Fragaria ananassa) consisting of pelargonidin 3-glucoside covalently linked to four flavan-3-ols.* Phytochemistry, 2004. **65**(10): p. 1421-1428.

26. Fischer, T.C., et al., *Premature and ectopic anthocyanin formation by silencing of anthocyanidin reductase in strawberry (Fragaria x ananassa).* New Phytologist, 2014. **201**(2): p. 440-451.

27. Lopez-Serrano, M. and A.R. Barcelo, *Kinetic properties of (+)-catechin oxidation by a basic peroxidase ioenzyme from strawberries.* Journal of Food Science, 1997. **62**(4): p. 676-723.

28. Akagi, T., et al., *Expression balances of structural genes in shikimate and flavonoid biosynthesis cause a difference in proanthocyanidin accumulation in persimmon (Diospyros kaki Thunb.) fruit.* Planta, 2009. **230**(5): p. 899-915.

29. Foo, L.Y., et al., *The structure of cranberry proanthocyanidins which inhibit adherence of uropathogenic P-fimbriated Escherichia coli in vitro.* Phytochemistry, 2000. **54**: p. 173-181.

30. Foo, L.Y., et al., *A-type proanthocyanidin trimers from cranberry that inhibit adherence of uropathogenic P-fimbriated Escherichia coli.* Journal of Natural Products, 2000. **63**: p. 1225-1228.

31. Gu, L., et al., *Fractionation of polymeric procyanidins from lowbush blueberry and quantification of procyanidins in selected foods with an optimized normal-phase HPLC-MS fluorescent detection method.* J. Agric. Food Chem., 2002. **50**: p. 4852-4860.

32. Murphy, K.J., et al., *Dietary flavanols and procyanidin oligomers from cocoa (Theobroma cacao) inhibit platelet function.* Am J Clin Nutr, 2003. **77**(6): p. 1466-1473.

33. Miller, K.B., et al., *Survey of Commercially Available Chocolate- and Cocoa-Containing Products in the United States. 2. Comparison of Flavan-3-ol Content with Nonfat Cocoa Solids, Total Polyphenols, and Percent Cacao.* Journal of Agricultural and Food Chemistry, 2009. **57**(19): p. 9169-9180.

34. Schewe, T., H. Kuhn, and H. Sies, *Flavonoids of Cocoa Inhibit Recombinant Human 5-Lipoxygenase.* J. Nutr., 2002. **132**(7): p. 1825-1829.

35. Serafini, M., et al., *Plasma antioxidants from chocolate.* Nature, 2003. **424**: p. 1013.

36. Liu, Y.J., et al., *Purification and Characterization of a Novel Galloyltransferase Involved in Catechin Galloylation in the Tea Plant (Camellia sinensis).* Journal of Biological Chemistry, 2012. **287**(53): p. 44406-44417.

37. Zhao, L., et al., *Metabolic Characterization of the Anthocyanidin Reductase Pathway Involved in the Biosynthesis of Flavan-3-ols in Elite Shuchazao Tea (Camellia sinensis) Cultivar in the Field.* Molecules, 2017. **22**(12): p. 21.

38. Wang, P., et al., *Functional demonstration of plant flavonoid carbocations proposed to be involved in the biosynthesis of proanthocyanidins.* The Plant Journal, 2020. **101**(1): p. 18-36.

39. Monagas, M., et al., *Monomeric, oligomeric,and polymeric flavan-3-ol composition of wines and grapes from Vitis vinifera L. cv. Graciano, Tempranillo, and Cabernet Sauvignon.* Journal of Agricultural and Food Chemistry, 2003. **51**: p. 6475-6481.

40. Dai, X., et al., *Discovery and characterization of tannase genes in plants: roles in hydrolysis of tannins.* New Phytologist, 2020. **in press**.

41. De Bruyne, T., et al., *Biological evaluation of proanthocyanidin dimers and related polyphenols.* J. Nat. Prod., 1999. **62**(7): p. 954-958.

42. Iwasawa, A., et al., *Antiviral activity of proanthocyanidin against feline calicivirus used as a surrogate for noroviruses, and coxsackievirus used as a representative enteric virus.* Biocontrol Science, 2009. **14**(3): p. 107-111.

43. Molan, A.L., et al., *The effect of condensed tannins from Lotus pedunculatus and Lotus corniculatus on the growth of proteolytic rumen bacteria in vitro and their possible mode of action.* Can J Microbiol, 2001. **47**(7): p. 626-633.

44. Howell, A.B., et al., *A-type cranberry proanthocyanidins and uropathogenic bacterial anti-adhesion activity.* Phytochemistry, 2005. **66**(18): p. 2281-2291.

45. Ohata, M., et al., *Effects of tea constituents on cell cycle progression of human leukemia U937 cells.* Biomed Res. , 2005. **26**(1): p. 1-7.

46. Suganuma, M., A. Saha, and H. Fujiki, *New cancer treatment strategy using combination of green tea catechins and anticancer drugs.* Cancer Science, 2011. **102**(2): p. 317-323.

47. Panneerselvam, M., et al., *Dark chocolate receptors: epicatechin-induced cardiac protection is dependent on delta-opioid receptor stimulation.* American Journal of Physiology-Heart and Circulatory Physiology, 2010. **299**(5): p. H1604-H1609.

48. Loke, W.M., et al., *Pure dietary flavonoids quercetin and (-)-epicatechin augment nitric oxide products and reduce endothelin-1 acutely in healthy men.* Am J Clin Nutr, 2008. **88**(4): p. 1018-1025.

49. MacRae, K., et al., *Epicatechin's cardiovascular protective effects are mediated via opioid receptors and nitric oxide.* European Journal of Nutrition, 2019. **58**(2): p. 515-527.

50. Li, M.-H., et al., *Protective effects of oligomers of grape seed polyphenols against beta-amyloid-induced oxidative cell death.* Ann NY Acad Sci, 2004. **1030**(1): p. 317-329.

51. Levites, Y., et al., *Neuroprotection and neurorescue against Aβtoxicity and PKC-dependent release of non-amyloidogenic soluble precursor protein by green tea polyphenol (-)- epigallocatechin-3-gallate.* FASEB J., 2003. **17**: p. 952-954.

52. Weinreb, O., et al., *Neurological mechanisms of green tea polyphenols in Alzheimer's and Parkinson's diseases.* J Nutriti Biochem., 2004. **15**(9): p. 506-516.

53. Jin, Z., et al., *Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur.* Nature Structural & Molecular Biology, 2020. **27**(6): p. 529-532.

54. Xue, X., et al., *Structures of two coronavirus main proteases: Implications for substrate binding and antiviral drug design.* Journal of Virology, 2008. **82**(5): p. 2515-2527.

55. Pillaiyar, T., et al., *An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy.* Journal of Medicinal Chemistry, 2016. **59**(14): p. 6595-6628.

56. Schewe, T., et al., *Polyphenols of cocoa: inhibition of mammalian 15-lipoxygenase.* Biol Chem. , 2001. **382**(12): p. 1687-96.

57. Takahashi, T., et al., *Procyanidin oligomers selectively and intensively promote proliferation of mouse hair epithelial cells in vitro and activate hair follicle growth in vivo.* J Invest Dermatol, 1999. **112**(3): p. 310-6.

58. Dai, X., et al., *Discovery and characterization of tannase genes in plants: roles in hydrolysis of tannins.* New Phytologist, 2020. **226**(4): p. 1104-1116.

59. Jin, Z., et al., *Structure-based drug design, virtual screening and high-throughput screening rapidly identify antiviral leads targeting COVID-19.* bioRxiv, 2020: p. 2020.02.26.964882.

60. Jin, Z., et al., *Structure of M^{pro} from COVID-19 virus and discovery of its inhibitors.* bioRxiv, 2020: p. 2020.02.26.964882.

61. Chen, L., et al., *Cinanserin Is an Inhibitor of the 3C-Like Proteinase of Severe Acute Respiratory Syndrome Coronavirus and Strongly Reduces Virus Replication In Vitro.* Journal of Virology, 2005. **79**(11): p. 7095-7103.

CHAPTER 5

Flavonols and Dihydroflavonols Inhibit the Main Protease Activity of SARS-CoV-2 and the Replication of Human Coronavirus 229E

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) cuases COVID-19. To date, vaccines are available to prevent the infection of this virus, however, medicines are necessary to help control COVID-19. Human coronavirus 229E (HCoV-229E) causes the common cold. The main protease (M^{pro}) is an essential enzyme required for the multiplication of these two viruses in the host cells. Herein, we report docking simulation with two M^{pro} enzymes and flavonols and dihydroflavonols, *in vitro* inhibition of the SARS-CoV-2 M^{pro}, and *in vitro* inhibition of the HCoV 229E replication. The docking simulation results predicted that (+)-dihydrokaempferol, (+)-dihydroquercetin, (+)-dihydromyricetin, kaempferol, quercetin, myricentin, isoquercetin, and rutin could inhibitively bind to at least two subsites (S1, S1', S2, and S4) in the binding pocket of SARS-CoV-2 M^{pro} with affinity scores from -8.8 to -7.4. Likewise, these compounds were predicted to inhibitively bind to the HCoV-229E M^{pro} with affinity scores from -7.8 to -7.1. *In vitro* inhibition assays showed that seven available compounds effectively inhibited the SARS-CoV-2 M^{pro} activity and their IC50 values ranged from 0.125 to 12.9 μM. Five compounds inhibited the replication of HCoV-229E in Huh-7 cells. These findings indicate that these antioxidative flavonols and dihydroflavonols are promising candidates for curbing these two viruses.

This work is under review.

Yue Zhu, Frank Scholle, Samantha C. Kisthardt, De-Yu Xie, 2021 Flavonols and dihydroflavonols inhibit the main protease activity of SARS-CoV-2 and the replication of human coronavirus 229E, bioRxiv 2021.07.01.450756; doi: https://doi.org/10.1101/2021.07.01.450756

1. Introduction

SARS-CoV-2 is the abbreviation of the novel severe acute respiratory syndrome coronavirus 2. This virus was firstly reported to cause a severe pneumonia in December of 2019 in Wuhan, China [1-3]. On February 11, 2020, the World Health Organization (WHO) designated this pneumonia as coronavirus disease 2019 (COVID-19). COVID-19 the rapidly spread different countries. On March 11, 2020, WHO announced the COVID-19 pandemic [4,5]. This pandemic has rapidly spread across all over the world. By June 21, 2021, based on the COVID-19 Dashboard by Center for Systems Science and Engineering at Johns Hopkins Coronavirus Resource Center, 117,553,726 infected cases and 3,867,641 deaths have been reported from more than 200 countries or regions. No strategy to stop the spread of this virus was available until January 2021, when several vaccines started to be approved for vaccination in several countries [6-11]. On one hand, since the start of vaccination, the number of infections has started to decrease. On the other hand, due to the insufficient vaccine quantities and vaccine hesitancy even where available, in the first week of February, 2021, the daily infection cases and deaths were still more than 400,000 and 11,000, respectively. By Feb. 9, the case numbers still increased by more than 300,000 daily. Meanwhile, the use of vaccines has also indicated that developing effective medicines is necessary to stop COVID-19. A recent study showed that mutations in the spike protein of SARS-CoV-2 might cause the escape of new variants from antibody [12]. The variant B.1.351 found in South Africa was reported to be able to escape vaccines developed by AstraZeneca, Johnson & Johnson (J&J), and Novavax [13]. Merck & Co has stopped their race for vaccines due to the lack of effectiveness of their products, instead, they continue to focus on antiviral drug development [14]. Unfortunately, to date, effective medicines

are still under screening. Although chloroquine and hydroxychloroquine were reported to be potentially effective in helping to improve COVID-19 [15], the use of these two anti-malarial medicines has been arguable in USA because of potential risk concerns [16]. Other potential candidate medicines are the combination of α-interferon and anti-HIV drugs lopinavir/ritonavir [17], and remdesivir [15,18]. Given that the efficacy of all these medicines being repurposed has not been conclusive, further studies are necessary to apply them for treating COVID-19.

SARS-CoV-2 is a single stranded RNA virus. Its genomic RNA contains around 30,000 nucleotides and forms a positive sense strand with a 5′ methylated cap and a 3′ polyadenylated tail that encodes at least six open reading frames (ORF) [19,20]. This feature allows it to be able to use the ribosomes of the host cells to translate proteins. The longest ORF (ORF1a/b) translates two polyproteins, which are cleaved by one main protease (M^{pro}, a 3C-like protease, $3CL^{pro}$) and another papain-like protease (PL2pro) into 16 nonstructural proteins (NSPs), which include RNA-dependent RNA polymerase (RdRp, nsp12), RNA helicase (nsp13), and exoribonuclease (nsp14). The NSPs subsequently produce structural and accessory proteins. The structural proteins include an envelope protein, membrane protein, spike (S) protein, and nucleocapsid protein (Fig.1) [21,22]. The S protein is a type of glycoprotein and plays an essential role in the attachment and the infection of the host cells [23]. It binds to the human angiotensin converting enzyme 2 to help the virus enter the human cells [24,25]. Since May 2020, the mutations of amino acids of the S protein has created a large number of variants, of which the emergence of alpha, beta, gamma, and delta variants has shown more pathogenic and transmissible, thus caused potential challenges to use vaccines to completely control the pandemic [26-30].

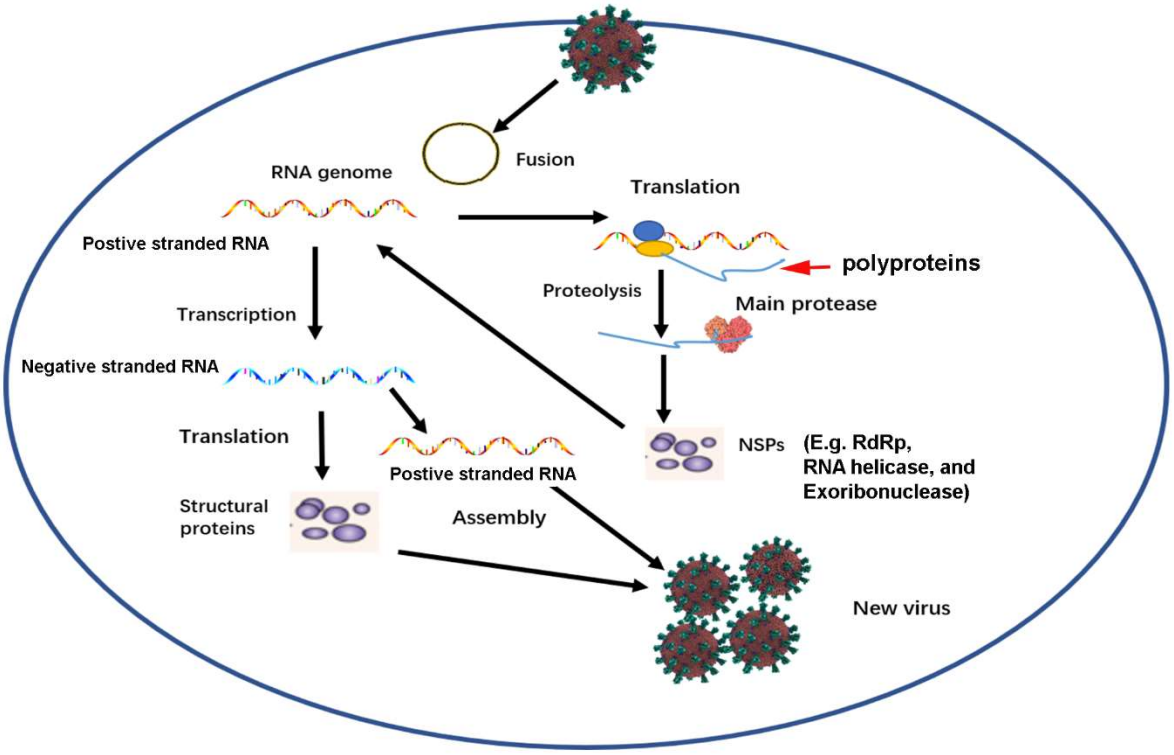


Figure 1 A diagram showing the function of the SARS-CoV-2 main protease in the virus replication in the host cells. Once the virus enters into the host cells. Its positive sense and single stranded RNA uses the ribosomes to translate open reading frames 1a and 1b to polyproteins (PP), in which the main protease and papain-like protease cleaves PPs to non-structural proteins (NSPs). Three NSPs, RNA dependent RNA polymerase (RdRp), RNA helicase, and exoribonuclease, are involved in the transcription of the positive RNA to negative sense and single stranded RNA, which is further transcribed to positive sense and single stranded RNA. Finally, structural proteins and a positive single stranded RNA assembly together to form a virus progeny.

Human coronavirus 229E (HCoV-229E) is a pathogenic virus in the genus *Alphacoronavirus* [31]. It is one of the causative viral agents of the common cold [32,33]. Its genome consists of a positive sense and single-stranded RNA with 27,317 nucleotides (nt). Its genome size commonly varies in different clinical isolates. For example, HCoV-229E strains 0349 and J0304 were two clinical isolates causing the common cold [34]. The entire genome of these two clinical isolates were reported to be about 27, 240 nt, which included 38.07 % GC content is in 0349 and 38.13 % GC content in J0304. In general, the genome of HCoV-229E is characterized with a gene

order of 5'-replicase ORF1a/b, spike (S), envelope (E), membrane (M), nucleocapsid (N)-3' [35], Like SARS-CoV-2, the spike protein is the determinant of infections to host cells [36]. The ORF1a/b of HCoV-229E encodes 16 non-structural proteins (NSPs). The *NSP5* encodes the M^{pro} that is required for the replication in the host cells [34]. Given that HCoV-229E is allowed to be studied in BSL2 laboratories, this pathogenic virus is an appropriate model to screen therapeutics for the treatment of both common cold and COVID-19.

Given that the SARS-CoV-2 M^{pro} not only plays a vital role in the cleavage of polyproteins, and there is no human homolog, it is an ideal target for anti-SARS-CoV-2 drug screens and development [37,38]. It belongs to the family of cysteine proteases and has a Cys-His catalytic dyad, which is an appropriate site to design and screen antiviral drugs [39]. Its high-resolution crystal structure was elucidated in April 2020 [40]. Based on the crystal structure, screening the existing antiviral medicines or designed chemicals revealed that cinanserin, ebselen, GC376, 11a, and 11b showed inhibitory effects on the M^{pro} activity [39-43]. A common feature is that these molecules deliver their carbonyl group (aldehyde group or ketone group) to the thiol of the 145 cysteine residue to form a covalent linkage, thus inhibit the M^{pro} activity. The potential application of these molecules is still under studies to evaluate their effectiveness and side effects. In addition, we recently found that flavan-3-ol gallates, such as (-)-epigallotechin-3-gallate, (-)-catechin-3-gallate, and (-)-epicatechin-3-gallate, and dimeric procyanidins promisingly inhibited the M^{pro} activity [44]. Docking simulation indicated that their inhibitory activity likely resulted from the formation of hydrogen bonds between these compounds and several amino acids in the binding domain of M^{pro}.

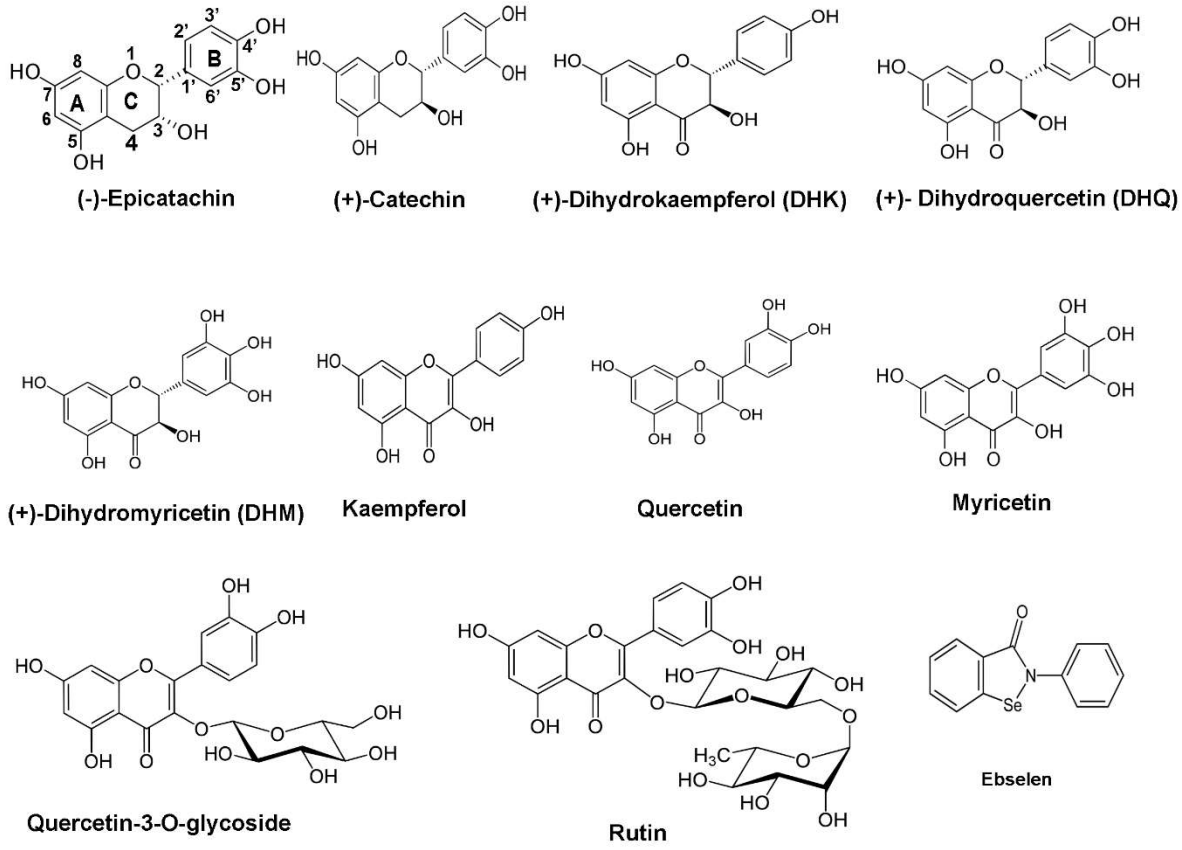


Figure 2 Structures of ebselen and 10 flavonoids. Two flavan-3-ols: (-)-epicatechin and (+)-catechin; three dihydroflavonol aglycones: (+)-dihydroquercetin, (+)-dihydrokaempferol, and (+)-dihydroquercetin; three flavonols aglycones, kaempferol, quercetin, and myricetin; two glycosylated flavonols: quercetin-3-O-glycoside (isoquercitrin), and rutin.

Flavonols and dihydroflavonols (Fig. 2) are two main groups of plant flavonoids [45,46]. Quercetin, kaempferol, and myricetin are three flavonol molecules widely existing in plants. Likewise, dihydroquercetin, dihydrokaempferol, and dihydromyricetin are three dihydroflavonol molecules in plants [47,48]. In general, flavonols and dihydroflavonols are strong antioxidants with multiple benefits to human health [49-58]. Furthermore, studies have reported that quercetin and its derivatives have antiviral activity [59-61]. Based on these previous findings, we hypothesized that flavonols and dihydroflavonols might inhibit the M^{pro} activity of SARS-CoV-2

and HCoV-229E. In this study, to test this hypothesis, we performed docking simulation for three dihydroflavonols, three flavonols, and two glycosylated quercetins. Then, we tested these compound's inhibition against the recombinant M^{pro} activity of SARS-CoV-2 *in vitro*. More importantly, five available compounds were evaluated to determine their inhibitive activity against the replication of HCoV-229E in Huh-7 cells. The resulting data showed eight compounds effectively inhibited the M^{pro} activity of SARS-CoV-2 and five tested compounds inhibited the replication of HCoV-229E in Huh-7 cells.

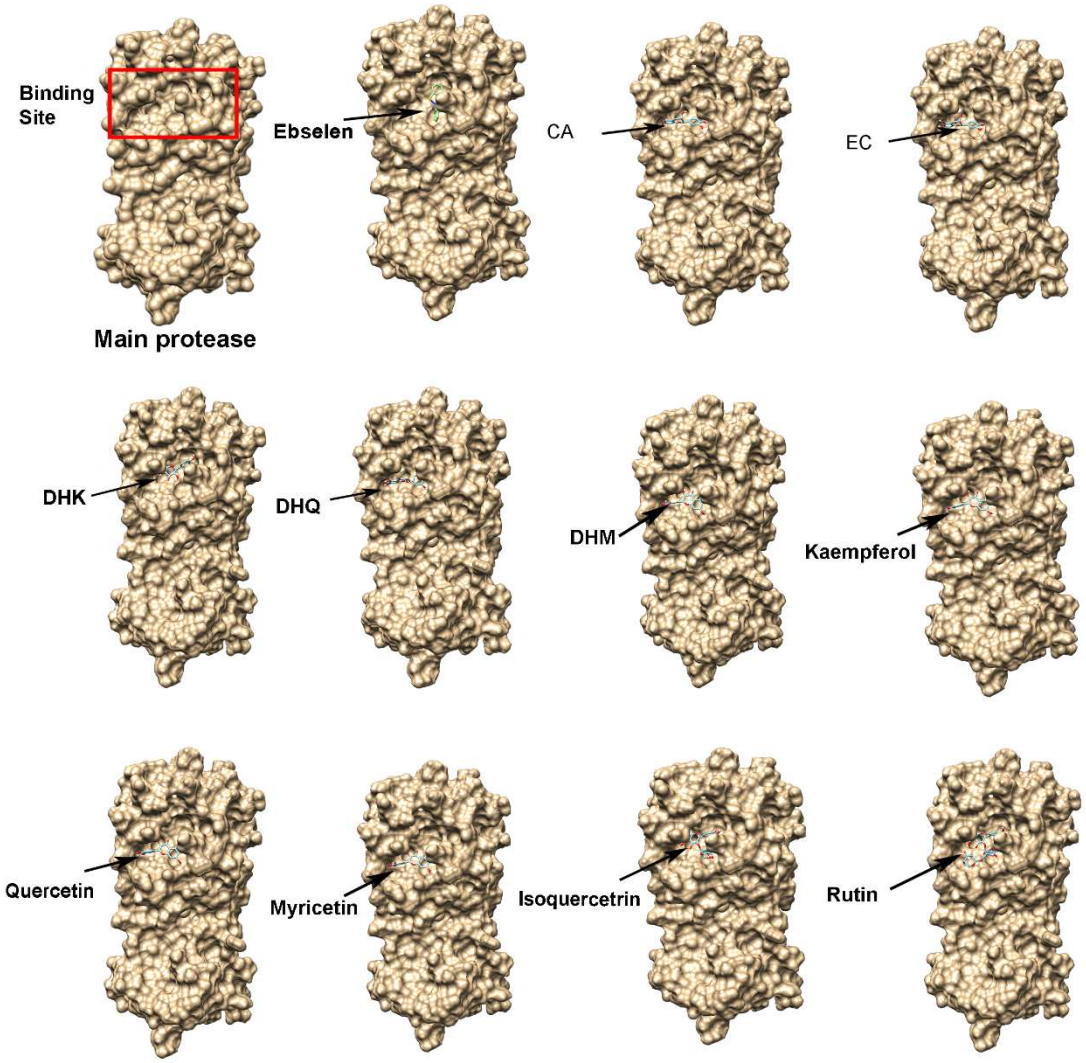


Figure 3 Ligand-receptor docking modeling showing the binding of eleven compounds to the substrate pocket of the SARS-CoV-2 main protease (M^{pro}). The first image shows the 3D surface view of the SARS-CoV-2 M^{pro}, on which the red rectangular frame indicates the substrate-binding pocket. Eleven flavonoids and ebselen bind to this pocket. Two flavan-3-ols: (+)-catechin (CA) and (-)-epicatechin (EC); three dihydroflavonol aglycones: (+)-dihydroquercetin (DHQ), (+)-dihydrokaempferol (DHK), and (+)-dihydroquercetin (DHM); three flavonols aglycones, kaempferol, quercetin, and myricetin; two glycosylated flavonols: quercetin-3-O-glycoside (isoquercitrin), and rutin.

2. Results

2.1 Ligand-receptor docking of flavonols and dihydroflavonols to the M^{pro} of SARS-CoV-2

Docking simulation was completed with the UCSF-Chimera and AutoDock Vina software to evaluate the binding abilities of flavonols and dihydroflavonols to the SARS-CoV-2 M^{pro}. The M^{pro} structure is featured with a substrate-binding pocket (Fig.3). When the 3D structure of the protein was downloaded from the public database, the peptide inhibitor N3 was shown to bind to this pocket. During protein preparation, N3 was removed for docking. The simulation results showed that (+)-DHQ, (+)-DHK, (+)-DHM, quercetin, kaempferol, myricetin, quercetin-3-O-glycoside, rutin, (-)-epicatechin, (+)-catechin, and ebselen bound to the binding pocket (Fig. 3). The resulting affinity scores for (+)-DHQ, (+)-DHK, (+)-DHM, quercetin, kaempferol, myricetin, quercetin-3-O-glycoside, and rutin ranged from -8.8 to -7.4, lower and better than the score of ebselen (-6.6) (Table 1). The scores among the aglycones of (-)-epicatechin, (+)-catechin, three dihydroflavonols, and three flavonol aglycones were close, either -7.4 or -7.5. These data suggested that dihydroflavonols, flavonols, and glycosylated flavonols could potentially inhibit the M^{pro} activity.

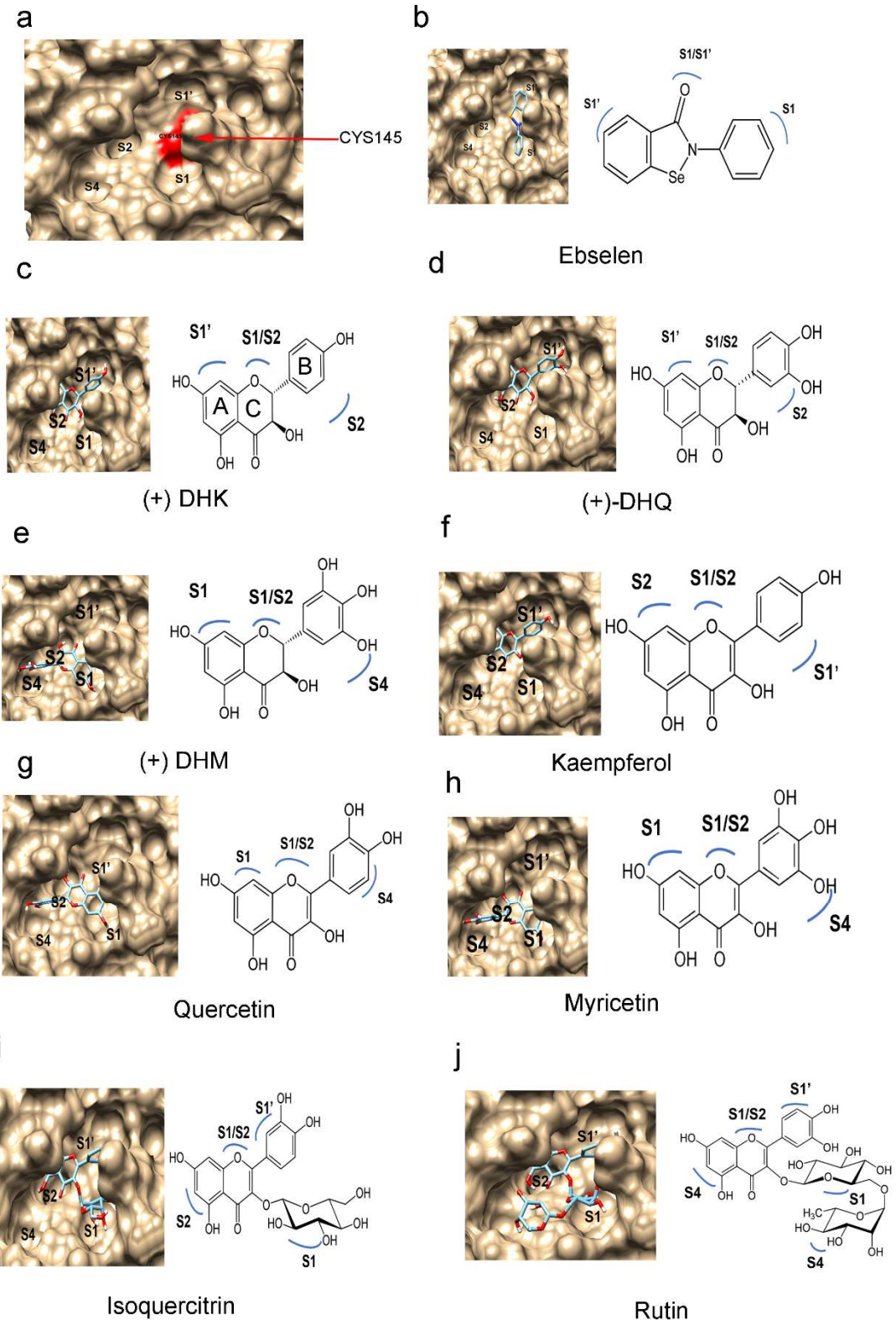
Table 1 Affinity scores of 11 compounds binding to the main proteases of SARS-CoV-2 and HuCoV-229E

Compounds	Affinity score (SARS-CoV-2)	Affinity score (229E)	Molecular weight (Da)
Rutin	-8.8	-7.8	610.5
Isoquercitrin	-8.7	-7.5	464.1
Kaempferol	-7.7	-7.6	286.2
DHK	-7.6	-7.6	288.2
DHM	-7.5	-7.6	320.2
Myricetin	-7.4	-7.1	318.2
Quercetin	-7.4	-7.7	302.2
(+)-Taxifolin	-7.4	-7.8	304.2
(+)-catechin	-7.5	-7.6	290.2
(-)-epicatechin	-7.5	-7.6	290.2
Ebselen	-6.6	-6.0	274.2

Figure 4 Orientation features of compounds binding to subsites.

a, a surface image shows the four subsites in the binding pocket.

b-j, images show the binding positions of nine compounds. Two flavan-3-ols: (+)-catechin (CA) and (-)-epicatechin (EC); three dihydroflavonol aglycones: (+)-dihydroquercetin (DHQ), (+)-dihydrokaempferol (DHK), and (+)-dihydroquercetin (DHM); three flavonols aglycones, kaempferol, quercetin, and myricetin; two glycosylated flavonols: quercetin-3-O-glycoside (isoquercitrin), and rutin.


a
S1'
S2
S4
S1
CYS145
b
S1/S1'
S1'
S1
O
N
Se
Ebselen
c
S1'
S1/S2
OH
HO
A
B
C
OH
OH
O
S2
(+) DHK
d
S1'
S1/S2
OH
HO
OH
OH
OH
O
S2
(+)-DHQ
e
S1'
S2
S4
S1
S1
S1/S2
OH
HO
OH
OH
OH
O
S4
(+) DHM
f
S1'
S2
S4
S1
S2
S1/S2
OH
HO
OH
OH
O
S1'
Kaempferol
g
S1'
S2
S4
S1
S1
S1/S2
OH
HO
OH
OH
OH
O
S4
Quercetin
h
S1'
S2
S4
S1
S1
S1/S2
OH
HO
OH
OH
OH
O
S4
Myricetin
i
S1'
S2
S4
S1
S1'
S1/S2
OH
HO
OH
S2
OH
O
HO
OH
OH
S1
Isoquercitrin
j
S1'
S2
S1
S1'
S1/S2
OH
HO
OH
OH
S4
OH
O
HO
OH
O
S1
H3C
O
HO
HO
OH
S4
Rutin

2.2 Docking features at the binding pocket of the SARS-CoV-2 M^{pro}

As we reported recently [44], the M^{pro} substrate-binding pocket includes four subsites, S1', S1, S2, and S4. Cys145 is a critical residue located at the space among subsites S1, S1', and S2 (Fig.4a) [39,40]. Several studies have reported that the thoil of the Cys145 residue is crucial for the catalytic activity of M^{pro} and if a compound binds to this residue, it can inhibit the M^{pro} activity [21,39,42]. When ebselen was used as our positive compound for simulation, as we reported recently [44], it bound to this residue featured by three rings facing to the S1 and S1' subsites (Fig. 4 b). The docking simulation results showed that three dihydroflavonols, three flavonols aglycones, and two glycosylated flavonols bound to 2 to 4 subsites via the Cys145 residue. In three dihydroflavonols tested, (+)-DHK and (+)-DHQ showed a difference in their occupation in the binding site. The A and B rings of (+)-DHK and (+)-DHQ dwelled in the S1' and S2 subsites and their heterocycle C ring resided in the space between the S1 and S2 subsites (Fig. 4 c-d). The A and B rings of DHM occupied the S1 and S4 subsites and the heterocycle C ring resided in the space between the S1 and S2 subsites (Fig. 4 e). In three flavonol aglycones tested, the occupation of kaempferol was different from that of quercetin and myricetin. The A-ring, B-ring, and heterocycle C-ring of kaempferol resided in the S1', S2, and the space between S1 and S2, respectively (Fig. 4f). The A-ring, B-ring, and the heterocycle C-ring of quercetin and myricetin dwelled in the S1, S4, and the space between S1 and S2 (Fig. 4 g-h). In comparison, the residing positions of isoquercitrin and rutin were more complicated. The A-ring, B-ring, heterocycle C-ring, and 3-glucose of isoquercitrin occupied the S2, S1', the space between S1 and S2, and S1 (Fig. 2 i). The A-ring, B-ring, heterocycle C-ring, 6-β-glucopyranose, and 1-L-α-rhamnopyranose of rutin occupied S4, S1', the space between S1/S2, S1, and S4 (Fig. 4 j). These

occupations in the binding sites suggested that these compounds might have an inhibitive activity against M^{pro}.

2.3 Ligand-receptor docking of flavonols and dihydroflavonols to the M^{pro} of HCoV-229E

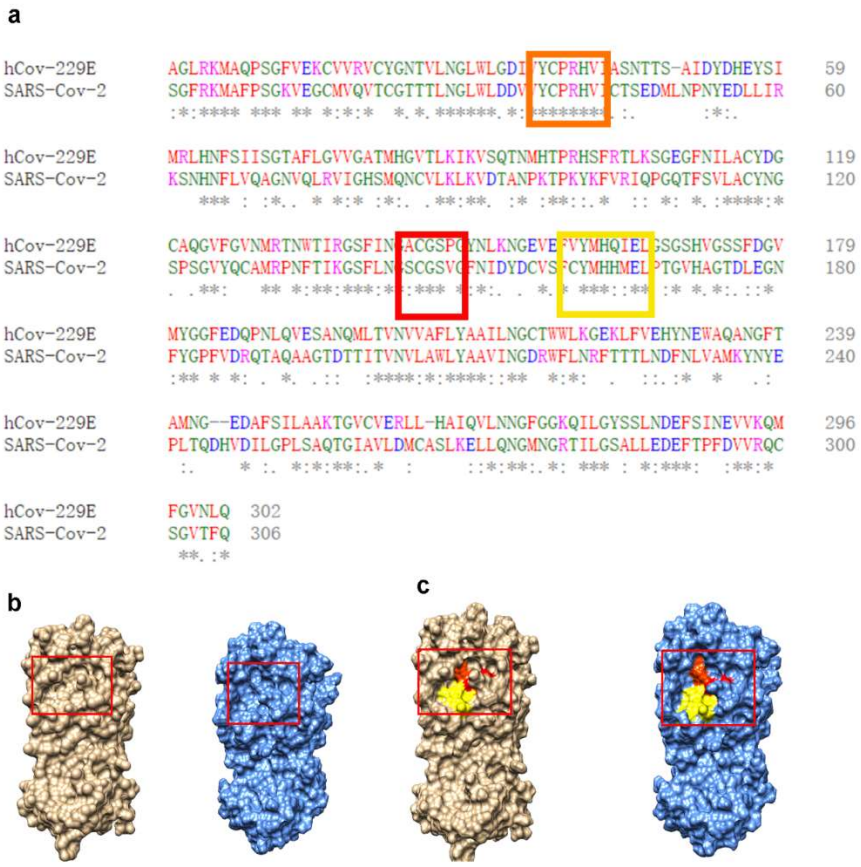


Figure 5 Amino acid sequence alignment of the SARS-CoV-2's and HCoV-229E's M^{pro} homologs and comparison of their three dimensional (3D) models.

a, amino sequence alignment, in which three rectangle frames highlight three conserved domains forming the substrate binding pocket;

b, a comparison of the 3D models of the SARS-CoV-2's (bronze color) and HCoV-229E's (blue color) M^{pro} homologs;

c, yellowish, orange, and reddish colors showing the binding pocket formed from three conserved binding domains highlighted with three rectangle frames in a, in which the reddish and yellowish spaces include Cys-His catalytic dyad.

The M^{pro} of HCoV-229E was also used for docking simulation. A sequence alignment revealed that the identity of between the two M^{pro} homologs of HCoV-229E and SARS-CoV-2 was 42.81% (Figure 5 a). The binding domains were highly conserved. Furthermore, a 3D modeling revealed that the conformation and binding pocket of the HCoV-229E M^{pro} were similar to those of SARS-CoV-2 M^{pro} (Fig. 5 b and c). The simulation results were the same as those of the M^{pro} of SARS-CoV-2 described above. (+)-DHQ, (+)-DHK, (+)-DHM, quercetin, kaempferol, myricetin, quercetin-3-O-glycoside, rutin, (-)-epicatechin, (+)-catechin, and ebselen could bind to the binding pocket of the HCoV-229E M^{pro} (Fig. 6). The affinity scores of these compounds ranged from -7.8 to - 7.1 (Table 1). The scores of rutin and isoquercitrin (two glycosides) binding to the HCoV-229E M^{pro} were -7.8 and -7.5, higher than -8,8 and -8.7, the scores of the two compounds binding to the SARS-CoV-2 M^{pro} (Table 1). This result indicates that compared with the affinity score of quercetin, these two types of glycosylation reduce the affinity scores binding to the SARS-CoV-2 M^{pro}, but do not affect the affinity scores binding to the HCoV-229E M^{pro}.

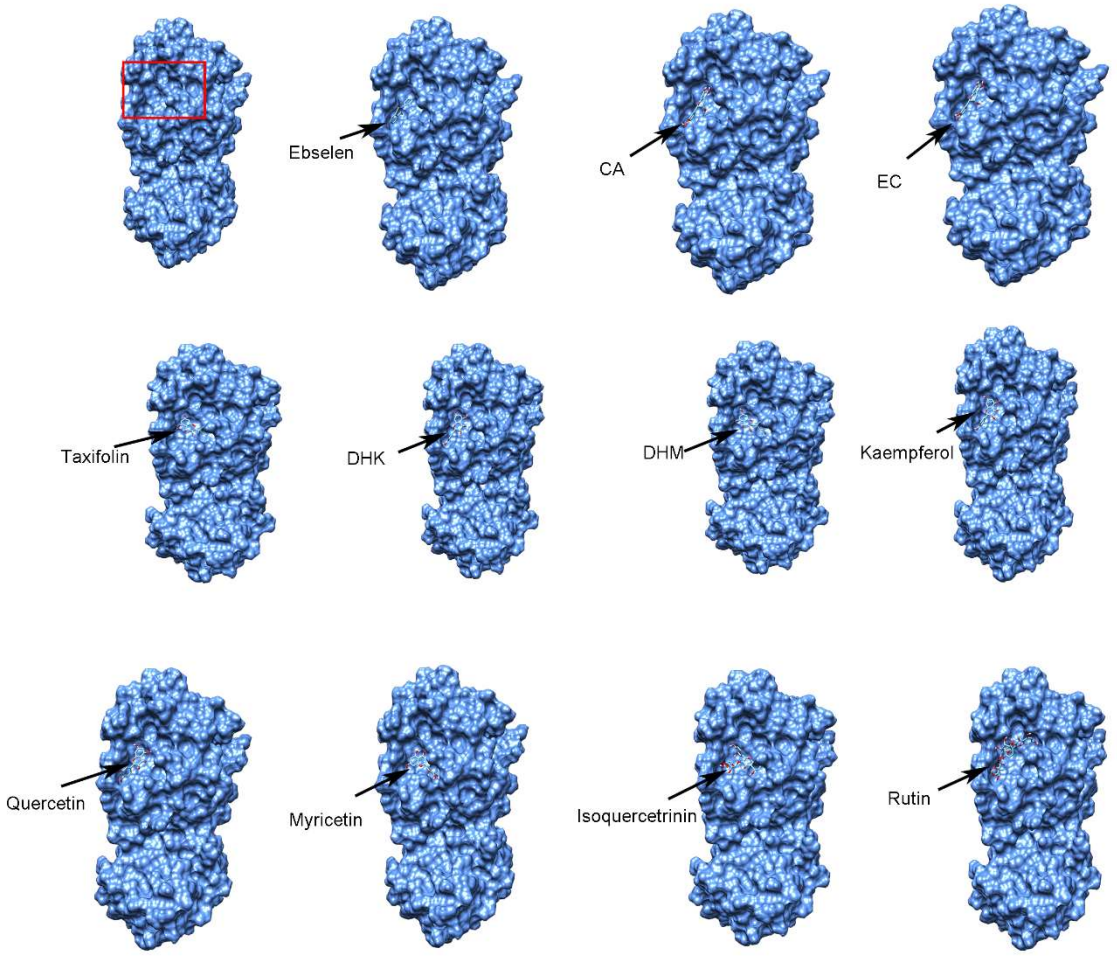


Figure 6 Ligand-receptor docking modeling showing the binding of eleven compounds to the substrate pocket of the HCoV-229E main protease. The first image shows the 3D surface view of the HCoV-229E M^{pro}, on which the red rectangular frame indicates the substrate-binding pocket. Eleven flavonoids and ebselen bind to this pocket. Two flavan-3-ols: (+)-catechin (CA) and (-)-epicatechin (EC); three dihydroflavonol aglycones: (+)-dihydroquercetin (DHQ), (+)-dihydrokaempferol (DHK), and (+)-dihydroquercetin (DHM); three flavonols aglycones, kaempferol, quercetin, and myricetin; two glycosylated flavonols: quercetin-3-O-glycoside (isoquercitrin), and rutin.

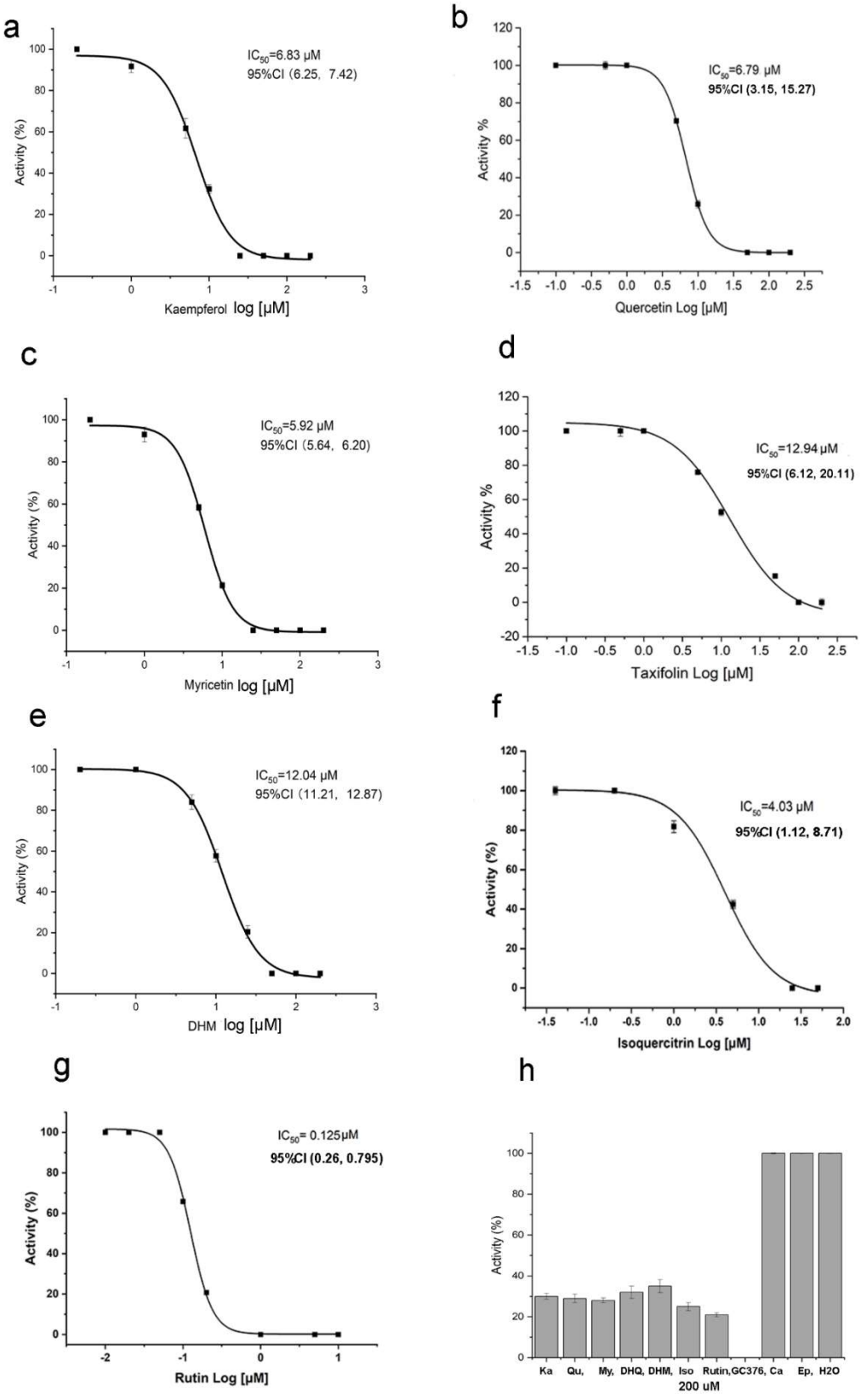
2.4 In vitro inhibitory effects of five flavonols and two dihydroflavonols on the SARS-CoV-2 M^{pro} activity

(+)-DHQ, (+)-DHM, quercetin, kaempferol, myricetin, isoquercetin (quercetin-3-O-glycoside), and rutin were used to test their inhibitory effects on the M^{pro} activity. In addition, based on our

recent report [44], (-)-epicatechin and (+)-catechin were used as negative controls. The resulting data showed that (+)-DHQ, (+)-DHK, (+)-DHM, quercetin, kaempferol, myricetin, isoquercetin (quercetin-3-O-glycoside), and rutin inhibited the SARS-CoV-2 M^{pro} activity. The half maximum inhibitory concentrations (IC50) were 0.125-12.94 μM (Fig. 7). Among the tested seven compounds, rutin had the lowest IC5 value with the most effectiveness to inhibit the M^{pro} activity (Fig. 7 g), while (+)-DHQ had the highest IC50 value with the lowest inhibitive activity (Fig. 7 d). One hundred μM was used to further compare the inhibitive effects of these compounds on the M^{pro} activity in a given time. The resulting data showed the most effectiveness of rutin (Fig. 7 h). In addition, as we reported previously, (+)-catechin and (-)-epicatechin did not show an inhibitory effect on the M^{pro} activity in the range of concentrations from 0-200 μM. For example, the two compounds did not inhibit the catalytic activity of the M^{pro} at 100 μM (Fig. 7 h).

Figure 7 Inhibitory effects of nine compounds on the M^{pro} activity of SARS-CoV-2.
a-g, seven plots show the inhibitory curves of seven compounds against the M^{pro} activity. All dots in each plot are an average value calculated from five replicates. IC50 value for each compound is inserted in each plot. "95% Cl" means 95% confidence internal. "(value 1, value 2)" means values in the range with 95% Cl.

h, a comparison shows the inhibitory effects of 11 compounds at 100 μM on the M^{pro} activity. GC376 is an inhibitor used as positive control. (+)-catechin, (-)-epicatechin, and water are used as negative controls. Two flavan-3-ols: (+)-catechin (Ca) and (-)-epicatechin (Ep); three dihydroflavonol aglycones: (+)-dihydroquercetin (DHQ), (+)-dihydrokaempferol (DHK), and (+)-dihydroquercetin (DHM); three flavonols aglycones, kaempferol (Ka), quercetin (Qu), and myricetin (My); two glycosylated flavonols: quercetin-3-O-glycoside (isoquercitrin, Iso), and rutin.


a
IC50=6.83 μM
95%CI (6.25, 7.42)
Activity (%)
Kaempferol log [μM]
b
IC50=6.79 μM
95%CI (3.15, 15.27)
Activity %
Quercetin Log [μM]
c
IC50=5.92 μM
95%CI (5.64, 6.20)
Activity (%)
Myricetin log [μM]
d
IC50=12.94 μM
95%CI (6.12, 20.11)
Activity %
Taxifolin Log [μM]
e
IC50=12.04 μM
95%CI (11.21, 12.87)
Activity (%)
DHM log [μM]
f
IC50=4.03 μM
95%CI (1.12, 8.71)
Activity (%)
Isoquercitrin Log [μM]
g
IC50= 0.125μM
95%CI (0.26, 0.795)
Activity (%)
Rutin Log [μM]
h
Activity (%)
Ka Qu, My, DHQ, DHM, Iso Rutin, GC376, Ca Ep, H2O
200 uM

2.5 Inhibitory effects of five compounds on the replication of HCoV-229E in Huh-7 cells

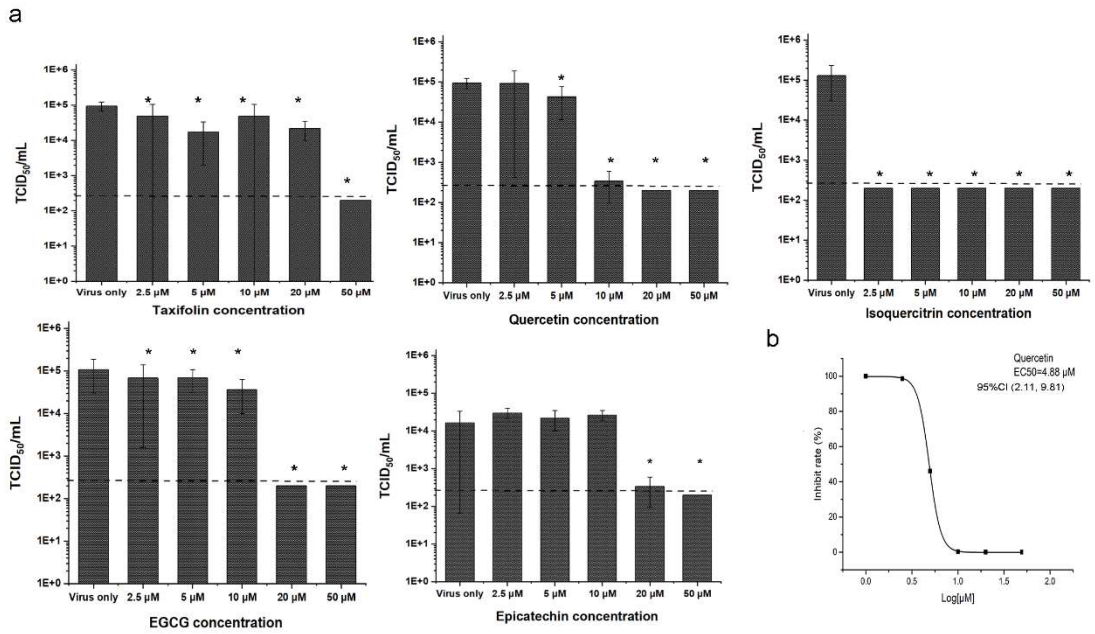


Figure 8 Inhibition of five compounds on the replication of HCoV-229E in Huh-7 cells.

a, plots were built with TCID50/mL versus concentrations of each compound.

b, this plot was built with the inhibition rate (%) versus log [μM] values to estimate the EC50 of quercetin. "95% Cl" means 95% confidence internal. "(value 1, value 2)" means values in the range with 95% Cl. Bars labeled with "*" means were significant difference compared with control without adding compounds (P-value less than 0.05).

Quercetin, isoquercetin, taxifolin, were tested their inhibitory effects on the replication of HCoV-229E in Huh-7 cells. In addition, epigallocatechin gallate (EGCG), and epicatechin two examples of flavan-3-ols, were tested. The reason was that we recently reported that EGCG effectively inhibited the SARS-CoV-2 M^{pro} activity, while epicatechin could not [44], however whether they could inhibit coronavirus replication in the host cells was untested. It was essential to test them. The resulting data indicated that all five compounds showed an inhibition against the replication of HCoV-229E in Huh-7 cells (Figure 8). Based on TCID50/ml values, taxifolin started to show its inhibition at 2.5 μM and its inhibitory activity increased as its concentration was increased. Quercetin started to have inhibition at 5 μM. As its concentrations were increased, its inhibitive activities were more effective. At a concentration tested higher 10 μM, quercetin could strongly inhibit the replication of the virus. Its EC50 value was estimated to be 4.88 μM (Figure 8 b). Isoquercitrin strongly inhibited the replication starting with 2.5 μM. EGCG started to show its inhibition against the replication of the virus at 2.5 μM and its inhibition became stronger as its concentrations were increased. It was interesting that epicatechin could strongly inhibit the replication starting at 20 μM.

3. Discussion

The development of medicines is necessary to complement the use of vaccines to control COVID-19. The SARS-CoV-2 M^{pro} is one of the targets to screen, repurpose, or develop drugs to treat or prevent SARS-CoV-2 [21,22,39]. One strategy is to inhibit the M^{pro} activity via delivering a compound to the Cys145 residue at the space across the region of S1' and S1 subsites [39]. Ebselen is a small molecule candidate that has been found to inhibit the M^{pro} activity with an

IC_{50} 0.46 μM [40]. Its structure featured with three rings was revealed to be an effective vessel to deliver its carbonyl group to the CYS145 residue (Fig. 4 b). We recently reported another strategy. We have found that epicatechin gallate, epigallocatechin gallate, gallocatechin gallate, catechin gallate, and procyanidin B2 could effectively inhibited the activity of M^{pro} via the formation of hydrogen bonds with different amino acids in the binding pocket [44]. Our findings indicated that the formation of peptide bonds was effective to screen more flavonoids to intervene COVID-19. Quercetin and other flavonols are common nutraceuticals with antiviral activities, such as influenza virus, hepatitis B virus, Zika virus, and Ebola viruses [59,81-83]. In this study, we took advantage of our recent strategy to perform this docking simulation of flavonols and dihydroflavonols. These two groups of compounds (Figs. 2 and 4) have C4 keto and 3-OH structures in the heterocycle C-ring. Like flavan-3-ol gallates, the structures of these two groups might have a potential to reside in the space S1 and S2 subsites. In the present study, our ligand-docking simulation showed that these two groups of compounds could bind to the substrate-binding pocket of M^{pro} and occupied their heterocyclic C ring in the crossing region between S1 and S2. Furthermore, the docking results predicted the A-ring and B-ring of two, three, two, and one compounds could bind to S1' and S2, S1 and S4, S2 and S1', and S4 and S1', respectively (Fig. 4). The docking results further showed that a glycosylation of quercetin increased the dwelling capacity in the binding site. Rutin was predicted to occupy all four subsites (Fig. 4 j). The increase of binding subsites was also reflected by the affinity scores of M^{pro}-ligands. Rutin and isoquercitrin had the lowest and second lowest score values (Table 1). These data indicated that not only might these compounds have an inhibitive activity but also a lower and better affinity score might indicate a strong inhibition against the M^{pro} activity. Further *in vitro* assays substantiated the prediction of docking simulation. Seven available compounds

inhibited the activity of M^{pro} with IC5 values from 0.125 to 12.9 μM. These data imply that these compounds might be potential therapeutics.

Given that SARS-CoV-2 can be only handled in the BSL-3 laboratories, we cannot access this deadly virus to test the effects of these compounds on its replication in host cells. Instead, we selected the less pathogenic HCoV-229E to test the inhibitory activity of these compounds. We hypothesized that inhibitory compounds screened against this virus might be appropriate for the potential therapy of COVID-19. The reason is that like SARS-CoV-2, the replication of HCoV-229E also depends on its M^{pro} activity in human cells and the active site is highly conserved between HCoV229E and SARS-CoV-2. Accordingly, the resulting data might help design medicines for the therapy of both COVID-19 and HCoV-229E respiratory diseases. An amino sequence alignment revealed that the identity of HCoV-229E and SARS-CoV-2 M^{pro} homologs was approximately 48%. The binding domain of substrates between the two homologs was conserved. Docking simulation and the resulting affinity scores further indicated that these compounds could reside in the binding pocket to potentially inhibit the activity of the M^{pro} of HCoV-229E (Fig. 4 and Table 1). Based on these data, we could test five compounds with HCoV-229E. In tested concentrations, taxifolin and isoquercitrin starting from 2.5 μM showed a significant inhibition of HCoV-229E replication in Huh-7 cells. Quercetin could slightly reduce the replication of HCoV-229E at 2.5 μM and significantly inhibited the replication of this virus at higher concentrations (Fig. 8 a). These positive results not only supported the docking simulation results that these compounds bound to the M^{pro} of HCoV-229E (Table 1), but also substantiated the results of *in vitro* assays that these compounds effectively inhibited the SARS-CoV-2 M^{pro} activity (Fig. 7). We previously demonstrated that EGCG could effectively inhibit

the activity of the SARS-CoV-2 M^{pro} [44]. Herein, we used it as a positive control. The resulting data showed that EGCG starting with 2.5 μM could significantly inhibit the replication of HCoV-229E in Huh-7 cells. This positive control result further supported that taxifolin, isoquercitrin and quercetin inhibited the replication of HCoV-29E via the reduction of the M^{pro} activity. In addition, we tested epicatechin, which was not shown to have an inhibitive activity against the SARS-CoV-2 M^{pro} in our vitro assays. It was interesting that epicatechin starting with 20 μM tested could inhibit the replication of HCoV-229E (Fig. 8 a), which was supported by the results of the docking simulation and its affinity score (Fig. 6 and Table 1). Accordingly, this datum indicates the difference between HCoV-229E and SARS-CoV-2. Taken together, these data indicate that HCoV-229E is appropriate substitute to screen inhibitors of SARS-CoV-2 by targeting the M^{pro} of these two viruses.

Quercetin, isoquercitrin and rutin are three common supplements, given that their nutritional values benefit human health [84-89]. Our data suggest that quercetin, isoquercitrin, and rutin might be helpful to intervene COVID-19. These compounds are plant natural flavonoids that their bio-availability, metabolism, and toxicity have been studied extensively [46]. In general, these compounds are safe nutrients sold as supplements or in food products such as onion and common dinner table fruits [90-95]. More importantly, quercetin can be absorbed into the human body from the intestines. A large number of human health studies have reported the presence of quercetin and its derivatives in the blood plasma and their nutritional benefits after consumption [96-99]. For example, the quercetin concentration in plasma was reported to reach 5.0±1.0 μM after the intake of 150 mg in one hour [100,101]. In addition, these compounds are potent antioxidants [102,103]. The intake of quercetin can inhibit the oxidation of LDL and prevent the

cardiovascular diseases [101,104] [105]. Moreover, quercetin and its derivatives have strong anti-inflammation activity [106-110]. All of these functions can benefit people's health.

4. **Materials and methods**

4.1 Dihydroflavonols, flavonols, cell line, and coronavirus

Flavonols used in this study included kaempferol, quercetin, myricetin, quercetin-3-O-glycoside, and rutin. Dihydroflavonols used were (+)-taxifolin (dihydroquercetin, DHQ), (+)-dihydrokaempferol (DHK), and (+)-dihydromyricetin (DHM). Two flavan-3-ols, (-)-epicatechin and (+)-catechin, were used as compound controls. The ebselen was used as a positive control. These compounds were purchased from Sigma-Aldrich (https://www.sigmaaldrich.com/)

Huh-7 cells, a human hepatocellular carcinoma cell line, are an appropriate to study the replication of different viruses [62-66]. This cell line was used for infection and propagation of virus and for testing antiviral activity of compounds. Human coronavirus 229E (HCoV-229E) is a positive sense and single-stranded RNA virus that infects the human respiratory system [67-71]. HCoV-229E was propagated on Huh-7 cells and tittered by TCID50 assay.

4.2 Docking simulation of the SARS-CoV-2 M^{pro}

We recently reported the docking simulation of flavan-3-ols, such as epicatechin and catechin (Fig. 2) [44]. Herein, we used the same steps for docking simulation of flavonols and dihydroflavonols in this experiment. In brief, three main steps were completed, protein preparation, ligand preparation, and protein-ligand docking. The first step was protein

preparation. The SARS-CoV-2 M^{pro} was used as a receptor to test ligands. Its ID is PDB ID: 6LU7 at Protein Data bank (https://www.rcsb.org/), from which its 3D structure was downloaded to a desktop computer and then was prepared as a receptor of ligand via the Dock Prep tool of UCSF-Chimera (https://www.cgl.ucsf.edu/chimera/). Because M^{pro} contains the inhibitor peptide N3, we removed N3 prior to docking simulation. Hydrogens and charges were added and optimized to allow determining the histidine protonation state. The second step was ligand preparation. The 3D structures of compounds (Fig. 2) were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/) and then used as ligands. All structures were minimized by using the minimize structure tool of UCSF-Chimera. Hydrogens and charges were added to the ligands, which were then saved as mol2 format for the protein-ligand docking simulation. The third step was protein-ligand docking. The modeling of protein-ligand docking was performed via the publically available AutoDock Vina (http://vina.scripps.edu/) software. The protein and ligand files were loaded to the AutoDock Vina through the UCSF-Chimera surface binding analysis tools. A working box was created to contain the whole receptor. The box center was set at x =-27, y = 13, and z = 58. The box size was set as x = 50, y =55, and z = 50, which framed the entire receptor to allow free position changes and ligand binding to the receptor at any potential positions.

4.3 Docking simulation of the HCoV-229E M^{pro}

HCoV-229E needs a M^{pro} (3C-like protease) for its replication in the host cells [72-74]. The M^{pro} is also a target for screening anti-HCoV-229E medicines [75-77]. The sequence of the HCoV-229E M^{pro} was obtained from the GenBank and then used for an alignment and docking simulation. The steps of simulation were the same as described above.

4.4 Inhibition assay of the SARS-CoV-2 M^{pro} activity

(+)-DHQ, (+)-DHK, (+)-DHM, quercetin, kaempferol, myricetin, quercetin-3-O-glycoside, rutin, (-)-epicatechin, and (+)-catechin were dissolved in DMSO to prepare a 1.0 M solution. A SARS-CoV-2 Assay Kit (BPS bioscience, https://bpsbioscience.com/) was used to test the inhibitory activity of these compounds. The steps of *in vitro* assay followed the manufacturer's protocol as performed in our recent report [44]. In brief, each reaction was carried out in a 25 µl volume in 384-well plates. Each reaction solution contains 150 ng recombinant M^{pro} (6 ng/µl), 1 mM DDT, 50 µM fluorogenic substrate, and one compound (0, 0.02, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 150, and 200 µM) in pH 8.0 mM Tris-HCl and 5 µM EDTA buffer. GC376 (50 µM) was used as a positive control, while (-)-epicatechin and (+)-catechin were used as two negative controls. The reaction mixtures were incubated for 2 hrs at room temperature. The fluorescence intensity of each reaction was measured and recorded on a microtiter plate-reading fluorimeter (BioTek's Synergy H4 Plate Reader for detect fluorescent and luminescent signals). The excitation wavelength was 360 nm and the detection emission wavelength was 460 nm. Each concentration of every compound was tested five times. A mean value was calculated with five individual replicates. Plots were built with the percentiles of catalysis versus log [µM] values of concentrations to show the effect of each compound on the M^{pro} activity. Statistical evolution is described below.

4.5 Inhibition assay of Human coronavirus 229E

Huh-7 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (10% FBS) and 1% antibiotics. HCoV-229E was propagated on Huh-7 cells. Virus containing supernatants were harvested 72h post infection and stored at -80°C. The

virus titer was determined by the Median Tissue Culture Infectious Dose 50 (TCID50) assay in Huh-7 cells.

Then, we performed virus inhibition assays. Huh-7 cells were seeded in 96 well plates at a density of 25,000 cells/well and incubated overnight. HCoV 229-E was diluted in MEM with 1% FBS, 1% HEPES buffer, and 1% antibiotic solution (MEM 1+1+1). The cells were inoculated with HCoV-229E at an MOI of 1 in a total volume of 50 µl. The infected plates were incubated at 35°C with 5% CO2 for one hour. Phytochemicals dissolved in DMSO were added in cell culture medium to the following concentrations: 0 µM, 2.5 µM, 5 µM, 10 µM, 20 µM, and 50 µM.

After one hour, virus and medium were removed from the infected cells and washed once with 200ul of PBS. 100 µl of each compound master mix was added to triplicate wells for each concentration. Virus was allowed to grow in the presence of each compound at 35 °C and 5% CO2 for 24 hours. Supernatants were harvested and virus titers on Huh-7 cells were determined by TCID50 assay [78]. Plates were incubated at 37°C and 5% CO2 for 96 hours, inspected visually for cytopathic effect (CPE) and TCID50/mL was calculated using the Spearman-Kaerber method [79,80]. A mean value was calculated using three replicates. Plots were built with TCID50/mL versus concentrations to show the effect of each compound on the replication of virus in Huh-7 cells. The minimum level of detection in this assay was 632 TCID50/ml.

4.6 Statistical evaluation

One-way analysis of variance (ANOVA) was performed to evaluate the statistical significance. The P-value less than 0.05 means significant differences.

REFERENCES

1. Wang, C.; Horby, P.W.; Hayden, F.G.; Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**, *395*, 470-473, doi:10.1016/S0140-6736(20)30185-9.

2. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R., et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **2020**, *382*, 727-733, doi:10.1056/NEJMoa2001017.

3. Ding, Q.; Lu, P.; Fan, Y.; Xia, Y.; Liu, M. The clinical characteristics of pneumonia patients coinfected with 2019 novel coronavirus and influenza virus in Wuhan, China. *J Med Virol* **2020**, 10.1002/jmv.25781, doi:10.1002/jmv.25781.

4. WHO. *Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)*; World Health Organization: 2020.

5. WHO. Director-General's opening remarks at the media briefing on COVID-19-11 March 2020. **2020**.

6. Kim, J.H.; Marks, F.; Clemens, J.D. Looking beyond COVID-19 vaccine phase 3 trials. *Nature Medicine* **2021**, 10.1038/s41591-021-01230-y, doi:10.1038/s41591-021-01230-y.

7. Knoll, M.D.; Wonodi, C. Oxford-AstraZeneca COVID-19 vaccine efficacy. *The Lancet* **2021**, *397*, 72-74, doi:10.1016/S0140-6736(20)32623-4.

8. Painter, E.M.; Ussery, E.N.; Patel, A.; Hughes, M.M.; Zell, E.R.; Moulia, D.L.; Scharf, L.G.; Lynch, M.; Ritchey, M.D.; Toblin, R.L., et al. Demographic Characteristics of Persons Vaccinated During the First Month of the COVID-19 Vaccination Program —

United States, December 14, 2020–January 14, 2021. *Morbidity and Mortality Weekly Report* **2021**, *70*, 174–177.

9. Dooling, K.; Marin, M.; Wallace, M.; McClung, N.; Chamberland, M.; Lee, G.M.; Talbot, H.K.; Romero, J.R.; Bell, B.P.; Oliver, S.E. The Advisory Committee on Immunization Practices' Updated Interim Recommendation for Allocation of COVID-19 Vaccine — United States, December 2020. *Morbidity and Mortality Weekly Report* **2021**, *69*, 1657-1660.

10. CDC; FDA. Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Moderna COVID-19 Vaccine — United States, December 21, 2020–January 10, 2021. *Morbidity and Mortality Weekly Report* **2021**, *70*, 125–129.

11. Gharpure, R.; Guo, A.; Bishnoi, C.K.; Patel, U.; Gifford, D.; Tippins, A.; Jaffe, A.; Shulman, E.; Stone, N.; Mungai, E., et al. Early COVID-19 First-Dose Vaccination Coverage Among Residents and Staff Members of Skilled Nursing Facilities Participating in the Pharmacy Partnership for Long-Term Care Program — United States, December 2020–January 2021. *Morbidity and Mortality Weekly Report* **2021**, *70()*, 178–182.

12. McCarthy, K.R.; Rennick, L.J.; Nambulli, S.; Robinson-McCarthy, L.R.; Bain, W.G.; Haidar, G.; Duprex, W.P. Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *bioRxiv* **2020**, 10.1101/2020.11.19.389916, 2020.2011.2019.389916, doi:10.1101/2020.11.19.389916.

13. Cohen, J. South Africa suspends use of AstraZeneca's COVID-19 vaccine after it fails to clearly stop virus variant. *Science* 2021.

14. Kenilworth, N.J. Merck discontinues development of SARS-CoV2/COVID-19 vaccine candidates; continues development of two investigational therapeutic candidates. *BUSINESS WIRE* 2021.

15. Wang, M.; Cao, R.; Zhang, L.; Yang, X.; Liu, J.; Xu, M.; Shi, Z.; Hu, Z.; Zhong, W.; Xiao, G. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res* **2020**, *30*, 269-271, doi:10.1038/s41422-020-0282-0.

16. Bull-Otterson, L.; Gray, E.B.; Budnitz, D.S.; Strosnider, H.M.; Schieber, L.Z.; Courtney, J.; García, M.C.; Brooks, J.T.; Kenzie, W.R.M.; Gundlapalli, A.V. Hydroxychloroquine and Chloroquine Prescribing Patterns by Provider Specialty Following Initial Reports of Potential Benefit for COVID-19 Treatment — United States, January–June 2020. *Morbidity and Mortality Weekly Report* **2020**, *69*, 1210–1215.

17. Cao, B.; Wang, Y.; Wen, D.; Liu, W.; Wang, J.; Fan, G.; Ruan, L.; Song, B.; Cai, Y.; Wei, M., et al. A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N Engl J Med* **2020**, *382*, 1787-1799, doi:10.1056/NEJMoa2001282.

18. Holshue, M.L.; DeBolt, C.; Lindquist, S.; Lofy, K.H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A., et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med* **2020**, *382*, 929-936, doi:10.1056/NEJMoa2001191.

19. Chen, Y.; Liu, Q.; Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol* **2020**, 10.1002/jmv.26234, doi:10.1002/jmv.26234.

20. Hussain, S.; Pan, J.; Chen, Y.; Yang, Y.; Xu, J.; Peng, Y.; Wu, Y.; Li, Z.; Zhu, Y.; Tien, P., et al. Identification of novel subgenomic RNAs and noncanonical transcription

initiation signals of severe acute respiratory syndrome coronavirus. *J Virol* **2005**, *79*, 5288-5295, doi:10.1128/JVI.79.9.5288-5295.2005.

21. Ramajayam, R.; Tan, K.P.; Liang, P.H. Recent development of 3C and 3CL protease inhibitors for anti-coronavirus and anti-picornavirus drug discovery. *Biochem Soc Trans* **2011**, *39*, 1371-1375, doi:10.1042/BST0391371.

22. Ren, Z.; Yan, L.; Zhang, N.; Guo, Y.; Yang, C.; Lou, Z.; Rao, Z. The newly emerged SARS-like coronavirus HCoV-EMC also has an "Achilles' heel": current effective inhibitor targeting a 3C-like protease. *Protein Cell* **2013**, *4*, 248-250, doi:10.1007/s13238-013-2841-3.

23. Zhao, P.; Praissman, J.L.; Grant, O.C.; Cai, Y.; Xiao, T.; Rosenbalm, K.E.; Aoki, K.; Kellman, B.P.; Bridger, R.; Barouch, D.H., et al. Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host Microbe* **2020**, *28*, 586-601.e586, doi:10.1016/j.chom.2020.08.004.

24. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T.S.; Herrler, G.; Wu, N.-H.; Nitsche, A., et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **2020**, *181*, 271-280.e278, doi:https://doi.org/10.1016/j.cell.2020.02.052.

25. Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367*, 1444-1448, doi:10.1126/science.abb2762.

26. Abdool Karim, S.S.; de Oliveira, T. New SARS-CoV-2 variants — Clinical, Public Health, and Vaccine Implications. *New England Journal of Medicine* **2021**, *384*, 1866-1868, doi:10.1056/NEJMc2100362.

27. Walensky, R.P.; Walke, H.T.; Fauci, A.S. SARS-CoV-2 Variants of Concern in the United States—Challenges and Opportunities. *JAMA* **2021**, *325*, 1037-1038, doi:10.1001/jama.2021.2294.

28. Fontanet, A.; Autran, B.; Lina, B.; Kieny, M.P.; Karim, S.S.A.; Sridhar, D. SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet* **2021**, *397*, 952-954, doi:10.1016/S0140-6736(21)00370-6.

29. Altmann, D.M.; Boyton, R.J.; Beale, R. Immunity to SARS-CoV-2 variants of concern. *Science* **2021**, *371*, 1103-1104, doi:10.1126/science.abg7404.

30. Petra, M.; Steven, K.; Mahesh Shanker, D.; Guido, P.; Bo, M.; Swapnil, M.; Charlie, W.; Thomas, M.; Isabella, F.; Rawlings, D., et al. SARS-CoV-2 B.1.617.2 Delta variant emergence and vaccine breakthrough. *Nature Portfolio* **2021**, 10.21203/rs.3.rs-637724/v1, doi:10.21203/rs.3.rs-637724/v1.

31. Woo, P.C.Y.; Huang, Y.; Lau, S.K.P.; Yuen, K.-Y. Coronavirus genomics and bioinformatics analysis. *Viruses* **2010**, *2*, 1804-1820.

32. Gaunt, E.R.; Hardie, A.; Claas, E.C.J.; Simmonds, P.; Templeton, K.E. Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J. Clin. Microbiol.* **2010**, *48*, 2940-2947, doi:10.1128/jcm.00636-10.

33. Shirato, K.; Kanou, K.; Kawase, M.; Matsuyama, S. Clinical isolates of human coronavirus 229E bypass the endosome for cell entry. *Journal of Virology* **2017**, *91*, 12, doi:10.1128/jvi.01387-16.

34. Farsani, S.M.J.; Dijkman, R.; Jebbink, M.F.; Goossens, H.; Ieven, M.; Deijs, M.; Molenkamp, R.; van der Hoek, L. The first complete genome sequences of clinical isolates of human coronavirus 229E. *Virus Genes* **2012**, *45*, 433-439, doi:10.1007/s11262-012-0807-9.

35. Fehr, A.R.; Perlman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. In *Coronaviruses: Methods and Protocols*, Maier, H.J., Bickerton, E., Britton, P., Eds. Springer New York: New York, NY, 2015; 10.1007/978-1-4939-2438-7_1pp. 1-23.

36. Shirato, K.; Kawase, M.; Watanabe, O.; Hirokawa, C.; Matsuyama, S.; Nishimura, H.; Taguchi, F. Differences in neutralizing antigenicity between laboratory and clinical isolates of HCoV-229E isolated in Japan in 2004-2008 depend on the S1 region sequence of the spike protein. *J. Gen. Virol.* **2012**, *93*, 1908-1917, doi:10.1099/vir.0.043117-0.

37. Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J., et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol* **2005**, *3*, e324, doi:10.1371/journal.pbio.0030324.

38. Kim, Y.; Liu, H.; Galasiti Kankanamalage, A.C.; Weerasekara, S.; Hua, D.H.; Groutas, W.C.; Chang, K.O.; Pedersen, N.C. Reversal of the Progression of Fatal Coronavirus Infection in Cats by a Broad-Spectrum Coronavirus Protease Inhibitor. *PLoS Pathog* **2016**, *12*, e1005531, doi:10.1371/journal.ppat.1005531.

39. Dai, W.; Zhang, B.; Jiang, X.M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F., et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368*, 1331-1335, doi:10.1126/science.abb4489.

40. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C., et al. Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289-293, doi:10.1038/s41586-020-2223-y.

41. Ye, G.; Wang, X.; Tong, X.; Shi, Y.; Fu, Z.F.; Peng, G. Structural Basis for Inhibiting Porcine Epidemic Diarrhea Virus Replication with the 3C-Like Protease Inhibitor GC376. *Viruses* **2020**, *12*, doi:10.3390/v12020240.

42. Chen, L.; Gui, C.; Luo, X.; Yang, Q.; Gunther, S.; Scandella, E.; Drosten, C.; Bai, D.; He, X.; Ludewig, B., et al. Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. *J Virol* **2005**, *79*, 7095-7103, doi:10.1128/JVI.79.11.7095-7103.2005.

43. Zhang, L.; Liu, Y. Potential interventions for novel coronavirus in China: A systematic review. *J Med Virol* **2020**, *92*, 479-490, doi:10.1002/jmv.25707.

44. Zhu, Y.; Xie, D.-Y. Docking characterization and in vitro inhibitory activity of flavan-3-ols and dimeric proanthocyanidins against the main protease activity of SARS-Cov-2. *Front. Plant Sci.* **2020**, *11*, doi:10.3389/fpls.2020.601316.

45. Fowler, Z.L.; Koffas, M.A. Biosynthesis and biotechnological production of flavanones: current state and perspectives. *Appl Microbiol Biotechnol* **2009**, *83*, 799-808, doi:10.1007/s00253-009-2039-z.

46. Hostetler, G.L.; Ralston, R.A.; Schwartz, S.J. Flavones: Food Sources, Bioavailability, Metabolism, and Bioactivity. *Adv Nutr* **2017**, *8*, 423-435, doi:10.3945/an.116.012948.

47. Xie, D.-Y.; Jackson, L.A.; Cooper, J.D.; Ferreira, D.; Paiva, N.L. Molecular and biochemical analysis of two cDNA clones encoding dihydroflavonol-4-reductase from *Medicago truncatula*. *Plant Physiol* **2004**, *134*, 979-994, doi:10.1104/pp.103.030221.

48. Xie, D.-Y.; Dixon, R.A. Proanthocyanidin biosynthesis - still more questions than answers? *Phytochemistry* **2005**, *66*, 2127-2144.

49. Moon, J.H.; Tsushida, T.; Nakahara, K.; Terao, J. Identification of quercetin 3-O-beta-D-glucuronide as an antioxidative metabolite in rat plasma after oral administration of quercetin. *Free Radic. Biol. Med.* **2001**, *30*, 1274-1285, doi:10.1016/s0891-5849(01)00522-6.

50. Murota, K.; Terao, J. Antioxidative flavonoid quercetin: implication of its intestinal absorption and metabolism. *Archives of Biochemistry an Biophysics* **2003**, *417*, 12-17.

51. Egert, S.; Wolffram, S.; Bosy-Westphal, A.; Boesch-Saadatmandi, C.; Wagner, A.E.; Frank, J.; Rimbach, G.; Mueller, M.J. Daily quercetin supplementation dose-dependently increases plasma quercetin concentrations in healthy humans. *J. Nutr.* **2008**, *138*, 1615-1621.

52. Chopra, M.; Fitzsimons, P.E.E.; Strain, J.J.T.; Thurnham, D.I.; Howard, A.N. Nonalcoholic red wine extract and quercetin inhibit LDL oxidation without affecting plasma antioxidant vitamin and carotenoid concentrations. *Clin. Chem.* **2000**, *46*, 1162-1170.

53. Hertog, M.G.L.; Feskens, E.J.M.; Hollman, P.C.H.; Katan, M.B.; Kromhout, D. Dietary antioxidant flavonols and risk of coronary heart disease- the Zutphen elderly study. *Lancet* **1993**, *342*, 1007-1011.

54. de Vries, J.H.M.; Hollman, P.C.H.; Meyboom, S.; Buysman, M.; Zock, P.L.; van Staveren, W.A.; Katan, M.B. Plasma concentrations and urinary excretion of the antioxidant flavonols quercetin and kaempferol as biomarkers for dietary intake. *Am. J. Clin. Nutr.* **1998**, *68*, 60-65.

55. Kolhir, V.K.; Bykov, V.A.; Baginskaja, A.I.; Sokolov, S.Y.; Glazova, N.G.; Leskova, T.E.; Sakovich, G.S.; Tjukavkina, N.A.; Kolesnik, Y.A.; Rulenko, I.A. Antioxidant activity of a dihydroquercetin isolated from Larix gmelinii (Rupr) Rupr wood. *Phytother. Res.* **1996**, *10*, 478-482.

56. Teselkin, Y.O.; Babenkova, I.V.; Tjukavkina, N.A.; Rulenko, I.A.; Kolesnik, Y.A.; Kolhir, V.K.; Eichholz, A.A. Influence of dihydroquercetin on the lipid peroxidation of mice during post-radiation period. *Phytother. Res.* **1998**, *12*, 517-519, doi:10.1002/(sici)1099-1573(199811)12:7<517::Aid-ptr342>3.0.Co;2-d.

57. Teselkin, Y.O.; Babenkova, I.V.; Kolhir, V.K.; Baginskaya, A.I.; Tjukavkina, N.A.; Kolesnik, Y.A.; Selivanova, I.A.; Eichholz, A.A. Dihydroquercetin as a means of antioxidative defence in rats with tetrachloromethane hepatitis. *Phytother. Res.* **2000**, *14*, 160-162, doi:10.1002/(sici)1099-1573(200005)14:3<160::Aid-ptr555>3.0.Co;2-y.

58. Weidmann, A.E. Dihydroquercetin: More than just an impurity? *Eur. J. Pharmacol.* **2012**, *684*, 19-26, doi:10.1016/j.ejphar.2012.03.035.

59. Mehrbod, P.; Hudy, D.; Shyntum, D.; Markowski, J.; Los, M.J.; Ghavami, S. Quercetin as a natural therapeutic candidate for the treatment of influenza virus. *Biomolecules* **2021**, *11*, 29, doi:10.3390/biom11010010.

60. dos Santos, A.E.; Kuster, R.M.; Yamamoto, K.A.; Salles, T.S.; Campos, R.; de Meneses, M.D.F.; Soares, M.R.; Ferreira, D. Quercetin and quercetin 3-O-glycosides from Bauhinia longifolia (Bong.) Steud. show anti-Mayaro virus activity. *Parasites Vectors* **2014**, *7*, 7, doi:10.1186/1756-3305-7-130.

61. Cheng, Z.K.; Sun, G.; Guo, W.; Huang, Y.Y.; Sun, W.H.; Zhao, F.; Hu, K.H. Inhibition of hepatitis B virus replication by quercetin in human hepatoma cell lines. *Virol. Sin.* **2015**, *30*, 261-268, doi:10.1007/s12250-015-3584-5.

62. Nakabayashi, H.; Taketa, K.; Yamane, T.; Miyazaki, M.; Miyano, K.; Sato, J. Phenotypical stability of a human hepatoma-cell line, Huh-7, in long-term culture with chemically defined medium. *Gann* **1984**, *75*, 151-158.

63. Shih, C.M.; Lo, S.C.J.; Miyamura, T.; Chen, S.Y.; Lee, Y.H.W. Suppression of hepatitis-B virus expression and replication by hepatitis-C virus core protein in Huh-7 cells. *Journal of Virology* **1993**, *67*, 5823-5832, doi:10.1128/jvi.67.10.5823-5832.1993.

64. Dewi, B.E.; Desti, H.; Ratningpoeti, E.; Sudiro, M.; Fithriyah; Angelina, M.; Iop. Effectivity of quercetin as antiviral to dengue virus-2 strain New Guinea C in Huh 7-it 1 cell line. In *3rd International Conference on Natural Products and Bioresource Sciences*, Iop Publishing Ltd: Bristol, 2020; Vol. 462.

65. Thongsri, P.; Sa-ngiamsuntorn, K.; Sithisarn, P.; Chomnawang, M.T.; Thirapanmethee, K. Cladogynos orientalis Zipp. extracts inhibit cell culture-derived hepatitis C virus

genotype 2a replication in Huh-7 cells through NS5B inhibition. *Asian Pac. Trop. Biomed.* **2019**, *9*, 346-352, doi:10.4103/2221-1691.262079.

66. Logue, J.; Licona, W.V.; Cooper, T.K.; Reeder, B.; Byrum, R.; Qin, J.; Murphy, N.D.; Cong, Y.; Bonilla, A.; Sword, J., et al. Ebola virus isolation using Huh-7 cells has methodological advantages and similar sensitivity to isolation using other cell types and suckling BALB/c laboratory mice. *Viruses-Basel* **2019**, *11*, 14, doi:10.3390/v11020161.

67. Bucknall, R.A.; Chanock, R.M.; Kapikian, A.Z.; King, L.M. Studies with human coronaviruses. 2. Some properties of strains 229E and OC43. *Proc. Soc. Exp. Biol. Med.* **1972**, *139*, 722-&.

68. Kennedy, D.A.; Johnsonlussenburg, C.M. Isolation and morphology of internal component of human coronavirus, strain 229E. *Intervirology* **1976**, *6*, 197-206.

69. Hierholzer, J.C. Purification and biophysical properties of human coronavirus 229E. *Virology* **1976**, *75*, 155-165, doi:10.1016/0042-6822(76)90014-3.

70. Macnaughton, M.R.; Madge, M.H. Genome of human coronavirus strain 229E. *J. Gen. Virol.* **1978**, *39*, 497-504, doi:10.1099/0022-1317-39-3-497.

71. Friedman, N.; Jacob-Hirsch, J.; Drori, Y.; Eran, E.; Kol, N.; Nayshool, O.; Mendelson, E.; Rechavi, G.; Mandelboim, M. Transcriptomic profiling and genomic mutational analysis of Human coronavirus (HCoV)-229E-infected human cells. *PLoS One* **2021**, *16*, 15, doi:10.1371/journal.pone.0247128.

72. Ziebuhr, J.; Heusipp, G.; Siddell, S.G. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *Journal of Virology* **1997**, *71*, 3992-3997, doi:10.1128/jvi.71.5.3992-3997.1997.

73. Ziebuhr, J.; Herold, J.; Siddell, S.G. Characterization of a human coronavius (strain 229E) 3C-like proteinase activity. *Journal of Virology* **1995**, *69*, 4331-4338, doi:10.1128/jvi.69.7.4331-4338.1995.

74. Ziebuhr, J.; Heusipp, G.; Seybert, A.; Siddell, S.G. Substrate specificity of the human coronavirus 229E 3C-like proteinase. In *Coronaviruses and Arteriviruses*, Enjuanes, L., Siddell, S.G., Spaan, W., Eds. Plenum Press Div Plenum Publishing Corp: New York, 1998; Vol. 440, pp. 115-120.

75. Prior, A.M.; Kim, Y.J.; Weerasekara, S.; Moroze, M.; Alliston, K.R.; Uy, R.A.Z.; Groutas, W.C.; Chang, K.O.; Hua, D.H. Design, synthesis, and bioevaluation of viral 3C and 3C-like protease inhibitors. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 6317-6320, doi:10.1016/j.bmcl.2013.09.070.

76. Chuck, C.P.; Chen, C.; Ke, Z.H.; Wan, D.C.C.; Chow, H.F.; Wong, K.B. Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases. *Eur. J. Med. Chem.* **2013**, *59*, 1-6, doi:10.1016/j.ejmech.2012.10.053.

77. Lee, C.C.; Kuo, C.J.; Ko, T.P.; Hsu, M.F.; Tsui, Y.C.; Chang, S.C.; Yang, S.; Chen, S.J.; Chen, H.C.; Hsu, M.C., et al. Structural Basis of Inhibition Specificities of 3C and 3C-like Proteases by Zinc-coordinating and Peptidomimetic Compounds. *J. Biol. Chem.* **2009**, *284*, 7646-7655, doi:10.1074/jbc.M807947200.

78. Barrett, P.N.; Meyer, H.; Wachtel, I.; Eibl, J.; Dorner, F. Determination of the inactivation kinetics of hepatitis A virus in human plasma products using a simple TCID50 assay. *J. Med. Virol.* **1996**, *49*, 1-6, doi:10.1002/(sici)1096-9071(199605)49:1<1::Aid-jmv1>3.0.Co;2-a.

79. Kärber, G. Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche. *Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie* **1931**, *162*, 480-483, doi:10.1007/BF01863914.

80. Spearman, C. THe method of 'Right and Wrong Cases' ('Constant stiumuli') without Gauss's formulae. *Br. J. Psychol.* **1908**, *2*, 227-242, doi:10.1111/j.2044-8295.1908.tb00176.x.

81. Parvez, M.K.; Al-Dosari, M.S.; Arbab, A.H.; Al-Rehaily, A.J.; Abdelwahid, M.A.S. Bioassay-guided isolation of anti-hepatitis B virus flavonoid myricetin-3-O-rhamnoside along with quercetin from Guiera senegalensis leaves. *Saudi Pharm. J.* **2020**, *28*, 550-559, doi:10.1016/j.jsps.2020.03.006.

82. Wong, G.; He, S.H.; Siragam, V.; Bi, Y.H.; Mbikay, M.; Chretien, M.; Qiu, X.G. Antiviral activity of quercetin-3-beta-O-D-glucoside against Zika virus infection. *Virol. Sin.* **2017**, *32*, 545-547, doi:10.1007/s12250-017-4057-9.

83. Qiu, X.G.; Kroeker, A.; He, S.H.; Kozak, R.; Audet, J.; Mbikay, M.; Chretienc, M. Prophylactic Efficacy of Quercetin 3-beta-O-D-Glucoside against Ebola Virus Infection. *Antimicrob. Agents Chemother.* **2016**, *60*, 5182-5188, doi:10.1128/aac.00307-16.

84. Xu, Z.; Li, X.Q.; Yang, H.; Poolsawat, L.; Wang, P.; Leng, X.J. Dietary rutin promoted the growth, serum antioxidant response and flesh collagen, free amino acids contents of

grass carp (Ctenopharyngodon idella). *Aquac. Nutr.* **2021**, *27*, 544-555, doi:10.1111/anu.13205.

85. Ragheb, S.R.; El Wakeel, L.M.; Nasr, M.S.; Sabri, N.A. Impact of rutin and vitamin C combination on oxidative stress and glycemic control in patients with type 2 diabetes. *Clin. Nutr. ESPEN* **2020**, *35*, 128-135, doi:10.1016/j.clnesp.2019.10.015.

86. da Silva, E.R.; Brogi, S.; Lucon, J.F.; Campiani, G.; Gemma, S.; Maquiaveli, C.D. Dietary polyphenols rutin, taxifolin and quercetin related compounds target Leishmania amazonensis arginase. *Food Funct.* **2019**, *10*, 3172-3180, doi:10.1039/c9fo00265k.

87. Kolarevic, A.; Pavlovic, A.; Djordjevic, A.; Lazarevic, J.; Savic, S.; Kocic, G.; Anderluh, M.; Smelcerovic, A. Rutin as Deoxyribonuclease I Inhibitor. *Chem. Biodivers.* **2019**, *16*, 9, doi:10.1002/cbdv.201900069.

88. Seifert, J. *Role of Quercetin in Sports Nutrition*; Elsevier Academic Press Inc: San Diego, 2013; 10.1016/b978-0-12-396454-0.00052-7pp. 497-500.

89. Amanzadeh, E.; Esmaeili, A.; Rahgozar, S.; Nourbakhshnia, M. Application of quercetin in neurological disorders: from nutrition to nanomedicine. *Rev. Neurosci.* **2019**, *30*, 555-572, doi:10.1515/revneuro-2018-0080.

90. Burak, C.; Brull, V.; Langguth, P.; Zimmermann, B.F.; Stoffel-Wagner, B.; Sausen, U.; Stehle, P.; Wolffram, S.; Egert, S. Higher plasma quercetin levels following oral administration of an onion skin extract compared with pure quercetin dihydrate in humans. *Eur. J. Nutr.* **2017**, *56*, 343-353, doi:10.1007/s00394-015-1084-x.

91. Egert, S.; Wolffram, S.; Schulze, B.; Langguth, P.; Hubbermann, E.M.; Schwarz, K.; Adolphi, B.; Bosy-Westphal, A.; Rimbach, G.; Muller, M.J. Enriched cereal bars are more effective in increasing plasma quercetin compared with quercetin from powder-filled hard capsules. *Br. J. Nutr.* **2012**, *107*, 539-546, doi:10.1017/s0007114511003242.

92. Careri, M.; Corradini, C.; Elviri, L.; Nicoletti, I.; Zagnoni, I. Direct HPLC analysis of quercetin and trans-resveratrol in red wine, grape, and winemaking byproducts. *J Agric Food Chem* **2003**, *51*, 5226-5231.

93. Meng, X.F.; Maliakal, P.; Lu, H.; Lee, M.J.; Yang, C.S. Urinary and plasma levels of resveratrol and quercetin in humans, mice, and rats after ingestion of pure compounds and grape juice. *J. Agric. Food Chem.* **2004**, *52*, 935-942, doi:10.1021/jf030582e.

94. Erlund, I.; Silaste, M.L.; Alfthan, G.; Rantala, M.; Kesaniemi, Y.A.; Aro, A. Plasma concentrations of the flavonoids hesperetin, naringenin and quercetin in human subjects following their habitual diets, and diets high or low in fruit and vegetables. *Eur. J. Clin. Nutr.* **2002**, *56*, 891-898, doi:10.1038/sj.ejcn.1601409.

95. Snyder, S.M.; Zhao, B.X.; Luo, T.; Kaiser, C.; Cavender, G.; Hamilton-Reeves, J.; Sullivan, D.K.; Shay, N.F. Consumption of Quercetin and Quercetin-Containing Apple and Cherry Extracts Affects Blood Glucose Concentration, Hepatic Metabolism, and Gene Expression Patterns in Obese C57BL/6J High Fat-Fed Mice. *J. Nutr.* **2016**, *146*, 1001-1007, doi:10.3945/jn.115.228817.

96. Day, A.J.; Williamson, G. Biomarkers for exposure to dietary flavonoids: a review of the current evidence for identification of quercetin glycosides in plasma. *Br. J. Nutr.* **2001**, *86*, S105-S110, doi:10.1079/bjn2001342.

97. Shi, Y.L.; Williamson, G. Quercetin lowers plasma uric acid in pre-hyperuricaemic males: a randomised, double-blinded, placebo-controlled, cross-over trial. *Br. J. Nutr.* **2016**, *115*, 800-806, doi:10.1017/s0007114515005310.

98. Mohammadi-Sartang, M.; Mazloom, Z.; Sherafatmanesh, S.; Ghorbani, M.; Firoozi, D. Effects of supplementation with quercetin on plasma C-reactive protein concentrations: a systematic review and meta-analysis of randomized controlled trials. *Eur. J. Clin. Nutr.* **2017**, *71*, 1033-1039, doi:10.1038/ejcn.2017.55.

99. Huang, H.H.; Liao, D.; Dong, Y.; Pu, R. Effect of quercetin supplementation on plasma lipid profiles, blood pressure, and glucose levels: a systematic review and meta-analysis. *Nutr. Rev.* **2020**, *78*, 615-626, doi:10.1093/nutrit/nuz071.

100. Olthof, M.R.; Hollman, P.C.; Vree, T.B.; Katan, M.B. Bioavailabilities of quercetin-3-glucoside and quercetin-4'-glucoside do not differ in humans. *J Nutr* **2000**, *130*, 1200-1203, doi:10.1093/jn/130.5.1200.

101. de Whalley, C.V.; Rankin, S.M.; Hoult, J.R.; Jessup, W.; Leake, D.S. Flavonoids inhibit the oxidative modification of low density lipoproteins by macrophages. *Biochem Pharmacol* **1990**, *39*, 1743-1750, doi:10.1016/0006-2952(90)90120-a.

102. Justino, G.; Santos, M.; Mira, L. Plasma metabolites of quercetin and their antioxidant potential. *Free Radic. Biol. Med.* **2002**, *33*, S203-S203.

103. Terao, J.; Yamaguchi, S.; Shirai, M.; Miyoshi, M.; Moon, J.H.; Oshima, S.; Inakuma, T.; Tsushida, T.; Kato, Y. Protection by quercetin and quercetin 3-O-beta-glueuronide of peroxynitrite-induced antioxidant consumption in human plasma low-density lipoprotein. *Free Radic. Res.* **2001**, *35*, 925-931, doi:10.1080/10715760100301421.

104. Hertog, M.G.; Feskens, E.J.; Hollman, P.C.; Katan, M.B.; Kromhout, D. Dietary antioxidant flavonoids and risk of coronary heart disease: the Zutphen Elderly Study. *Lancet* **1993**, *342*, 1007-1011, doi:10.1016/0140-6736(93)92876-u.

105. Manach, C.; Morand, C.; Crespy, V.; Demigne, C.; Texier, O.; Regerat, F.; Remesy, C. Quercetin is recovered in human plasma as conjugated derivatives which retain antioxidant properties. *FEBS Lett.* **1998**, *426*, 331-336, doi:10.1016/s0014-5793(98)00367-6.

106. Sato, S.; Mukai, Y. Modulation of chronic inflammation by quercetin: The beneficial effects on obesity. *J. Inflamm. Res.* **2020**, *13*, 421-431, doi:10.2147/jir.S228361.

107. Carullo, G.; Cappello, A.R.; Frattaruolo, L.; Badolato, M.; Armentano, B.; Aiello, F. Quercetin and derivatives: useful tools in inflammation and pain management. *Future Med. Chem.* **2017**, *9*, 79-93, doi:10.4155/fmc-2016-0186.

108. Tejada, S.; Nabavi, S.M.; Capo, X.; Martorell, M.; Bibiloni, M.D.; Tur, J.A.; Pons, A.; Sureda, A. Quercetin effects on exercise induced oxidative stress and inflammation. *Curr. Org. Chem.* **2017**, *21*, 348-356, doi:10.2174/1385272820666161017122202.

109. Li, Y.; Yao, J.Y.; Han, C.Y.; Yang, J.X.; Chaudhry, M.T.; Wang, S.N.; Liu, H.N.; Yin, Y.L. Quercetin, Inflammation and Immunity. *Nutrients* **2016**, *8*, 14, doi:10.3390/nu8030167.

110. Chen, S.; Jiang, H.M.; Wu, X.S.; Fang, J. Therapeutic Effects of Quercetin on Inflammation, Obesity, and Type 2 Diabetes. *Mediat. Inflamm.* **2016**, *2016*, 5, doi:10.1155/2016/9340637.