國立臺灣大學生命科學院生化科學研究所

博士論文

Institute of Biochemical Sciences

College of Life Science

National Taiwan University

Doctoral Dissertation

精準結構功能分析應用：臺灣牛樟樹順式異戊二烯轉移酶的演化、多功能纖維素酶/半纖維素酶改造、以及抗病毒/抗菌抑制劑篩選設計

Applications of Precise Structure-Function Analyses: Evolution of cis-Prenyltransferases from Taiwan Stout Camphor Tree (*Cinnamomum kanehirae*), Engineering of Multifunctional Cellulases/Hemicellulases, and Antiviral/Antibacterial Inhibitors Screening and Design

劉嘉晉

Jia-Jin Liu

指導教授：梁博煌 博士

Advisor: Po-Huang Liang, Ph.D.

中華民國 114 年 7 月

July, 2025

doi:10.6342/NTU202502385

國立臺灣大學博士學位論文
口試委員會審定書
DOCTORAL DISSERTATION ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

精準結構功能分析應用：臺灣牛樟樹順式異戊二烯轉移酶的演化、多功能纖維素酶/半纖維素酶改造、以及抗病毒/菌抑制劑篩選設計

Applications of Precise Structure-Function Analyses: Evolution of cis-Prenyltransferases from Taiwan Stout Camphor Tree (*Cinnamomum kanehirae*), Engineering of Multifunctional Cellulases/Hemicellulases, and Antiviral/Antibacterial Inhibitors Screening and Design

本論文係劉嘉晉（F08B46025）在國立臺灣大學生化科學研究所完成之博士學位論文，於民國 114 年 1 月 14 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Institute of Biochemical Sciences on 14/January/2025 have examined a Doctoral Dissertation entitled above presented by Jia-Jin Liu (F08B46025) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

（指導教授 Advisor）		

系（所、學位學程）主管 Director:

CONTENTS

doi:10.6342/NTU202502385

doi:10.6342/NTU202502385

doi:10.6342/NTU202502385

doi:10.6342/NTU202502385

doi:10.6342/NTU202502385

doi:10.6342/NTU202502385

doi:10.6342/NTU202502385

致謝

時光荏苒，長達六年的博士生涯終於走到尾聲。首先，謹向我的指導教授梁博煌老師致上最深的感謝。從一個從未接觸過蛋白質化學的碩士班學生，逐步訓練成一位能夠獨當一面的博士研究員，皆承蒙老師的悉心指導與栽培。學術研究從來不簡單，一路上經歷過無數磨練與失敗，最終方能累積些許研究成果。然而，當實驗結果印證假說，或在長期卡關後迎來突破的那一刻，其喜悅實非筆墨所能形容。梁老師在實驗上給予的邏輯思考訓練，以及論文撰寫上的建議，皆使我受益匪淺，也因為老師對研究經費方面的支持，讓我無後顧之憂地嘗試各種實驗，於此再次表達我最真摯的謝意。同時，謝謝實驗室同仁的陪伴與相互扶持，透過與各位的討論以及一同奮鬥的過程中，總會腦力激盪出新的想法，進而突破瓶頸。在繁忙的博士生活中，與大家共同度過的點滴，也為我的求學歲月增添無數色彩。另外，感謝台大生化所提供的資源，以及各位所辦人員在行政上的協助，讓我得以順利完成博士班資格考、口試以及畢業。感謝我的家人，謝謝你們始終支持我全心投入學術研究，成為我最強大的後盾。

最後，衷心感謝我的女友芃君，在這六年內歷經過無數風雨，謝謝妳一路上的精神支持與陪伴，即使我在追求博士學位時遭遇許多挫折，仍然相信終能撥雲見日。遇見妳始終是我最大的幸運，願往後的人生旅途，能執子之手繼續前行。

劉嘉晉　　謹誌於

臺灣大學生化科學研究所

中華民國一百一十四年七月

doi:10.6342/NTU202502385

Chapter I

Complexation and evolution of cis-prenyltransferases in *Cinnamomum kanehirae* deduced from kinetic and functional characterizations

中文摘要

真核生物的去氫多萜二磷酸合成酶（Dehydrodolichyl diphosphate synthases, DHDDSs）是一類順式異戊二烯轉移酶（cis-prenyltransferases, cis-PTs），負責合成多萜類（dolichols）的前驅物，從而調控蛋白質醣基化。這些酶需要合作夥伴，例如酵母中的 Nus1 和動物中的 NgBR，它們是缺乏催化結構域的 cis-PT 同源蛋白，但能夠增強 DHDDS 活性。與動物不同，植物擁有多種 cis-PTs，這些蛋白可以配對或單獨行動，生成具有不同鏈長的產物，但其生理功能尚不清楚。我們選擇研究台灣牛樟（*Cinnamomum kanehirae*），基因組分析顯示該物種含有兩個 DHDDS 類蛋白與三個 NgBR 類蛋白。我們發現，一個 DHDDS 類蛋白作為同源二聚體的 cis-PT，可生成中鏈長的 C55 產物，而另一個則與兩個 NgBR 類蛋白中的任何一個形成二聚體複合物，以產生更長鏈的產物。此外，這些複合物在較高溫度下能補充酵母 rer2 缺失菌株的生長缺陷。基於聚異戊二烯醇和多萜類的生物合成功能以及序列特徵，我們比較來自多個物種的 cis-PTs 同源蛋白，並揭示植物的 cis-PTs 可能的演化路徑。

關鍵詞：順式異戊二烯轉移酶；多萜；蛋白質醣基化；DHDDS；NgBR

doi:10.6342/NTU202502385

ABSTRACT

Eukaryotic dehydrodolichyl diphosphate synthases (DHDDSs) are cis-prenyltransferases (cis-PTs) that synthesize dolichol precursors essential for protein glycosylation. These enzymes require non-catalytic partners, such as Nus1 in yeast and NgBR in animals, which are cis-PT homologues devoid of the catalytic domain but enhancing DHDDS activity. In contrast to animals, plants possess multiple cis-PT homologues that can function independently or in pairs to produce products of varying chain lengths, with their physiological roles remaining largely unexplored. In this study, we investigated cis-PTs from *Cinnamomum kanehirae*, a tree identified through genomic analysis to express two DHDDS-like and three NgBR-like proteins. One DHDDS-like protein functioned as a homodimeric cis-PT, producing medium-chain C55 products, whereas the other formed heterodimeric complexes with either of two NgBR homologues to synthesize longer-chain products. Both complexes effectively complemented the growth defect of a yeast rer2 deletion strain at elevated temperatures. By examining the roles of these cis-PTs in polyprenol and dolichol biosynthesis and comparing their sequence motifs across species, we inferred the potential evolutionary paths for these cis-PTs.

KEYWORDS: cis-prenyltransferase; dolichol; protein glycosylation; DHDDS; NgBR

doi:10.6342/NTU202502385

ABBREVIATIONS

DHDDS, dehydrodolichyl diphosphate synthase

DMAPP, methylallyl diphosphate

Dol-P, dolichol-phosphate

ER, endoplasmic reticulum

FPP, farnesyl diphosphate

GGPP, geranylgeranyl diphosphate

GPP, geranyl diphosphate

HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPLC, high-performance liquid chromatography

IPP, isopentenyl diphosphate

LB, Luria-Bertani

MESG, 2-amino-6-mercapto-7-methyl-purine ribonucleoside

Ni-NTA, nickel nitrilotriacetic acid

NgBR, Nogo-B receptor

PCR, polymerase chain reaction

PDB, Protein Data Bank

PNP, purine ribonucleoside phosphorylase

PPA, inorganic pyrophosphatase

doi:10.6342/NTU202502385

TCEP, tris(2-carboxyethyl)phosphine

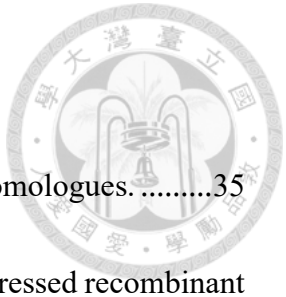
Tris, tris(hydroxymethyl)aminomethane

UPPS, undecaprenyl diphosphate synthase

doi:10.6342/NTU202502385

LIST OF TABLES

doi:10.6342/NTU202502385



LIST OF FIGURES

doi:10.6342/NTU202502385

doi:10.6342/NTU202502385

1. INTRODUCTION

1.1 Cis-Prenyltransferases

Prenyltransferases catalyze the synthesis of linear polyprenyl diphosphates through consecutive condensation reactions of the homoallylic substrate isopentenyl diphosphate (IPP, C5) with an allylic oligoprenyl diphosphate, such as dimethylallyl diphosphate (DMAPP, C5), geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15), or geranylgeranyl diphosphate (GGPP, C20), producing varying chain-length products [1-3]. Based on the stereochemistry of the double bonds formed during the IPP condensation reactions, prenyltransferases are categorized as trans-prenyltransferases (trans-PTs) or cis-prenyltransferases (cis-PTs) [4]. Trans-PTs typically synthesize products with chain lengths up to C50, whereas cis-PTs, despite sharing the ability to catalyze IPP condensation, possess distinct three-dimensional structures and generate significantly longer products that function as lipid carriers (see Figure S1a for the cis-PTs catalyzed reactions) [5]. For instance, a bacterial cis-PT undecaprenyl diphosphate synthase (UPPS) synthesizes a C55 product via eight IPP condensations with FPP, serving as a lipid carrier for mediating peptidoglycan biosynthesis [6]. On the other hand, the longer C70–C120 polyprenyl diphosphates, also known as dehydrodolichyl diphosphates (DHDDs), are synthesized by the eukaryotic cis-PTs named DHDD synthases (DHDDSs) in animals. These intermediates are subsequently dephosphorylated and reduced to dolichols, which

doi:10.6342/NTU202502385

are further converted into dolichyl phosphate (Dol-P), serving as glycosyl carriers across the endoplasmic reticulum (ER) membrane for glycoprotein biosynthesis (see Figure S1b for dolichol biosynthesis) [7]. Despite their distinct biological functions, these cis-PTs that catalyze identical IPP condensation reactions but produce polyprenyl products of varying chain lengths exhibit significant sequence homology.

1.2 The structures of cis-PTs

The earliest solved crystal structures of cis-PTs are from the bacterial UPPS, which has a homodimeric conformation with each subunit features an active-site cleft encompassed by 7 α-helices and 6 β-strands, arranged to accommodate the elongated C55 product, with the bottom sealed by the hydrophobic residue Leu137 [8, 9]. Based on the apo and ligand-bound structures of *E. coli* UPPS, it was determined that Asp26 interacts with FPP to facilitate its dephosphorylation, while deprotonated IPP attacks the partial carbocation intermediate of FPP, resulting in a concerted mechanism [10-12]. In human, there is a membrane protein called Nogo-B receptor (NgBR), which shares amino acid sequence homology with DHDDS but lacks essential catalytic residues (Figure 1), enhances DHDDS protein stability and promotes Dol-P production through forming a complex with DHDDS to provide the C-terminal RXG motif [13, 14]. The crystal structure of the recombinant human $NgBR^{79\text{-}293}$/DHDDS was solved to reveal a heterodimeric conformation [15], while a heterotetrameric assembly, formed by the

doi:10.6342/NTU202502385

dimerization of the heterodimer through the extended DHDDS C-terminus, was observed in solution, significantly enhancing the DHDDS activity by 400-fold [16, 17]. Nus1, a homologue of NgBR from *Saccharomyces cerevisiae* that has been structurally characterized as a homodimer, was also computationally modeled to form a heterodimeric complex with Rer2, one of the two DHDDS homologues, Rer2 and Srt1, in *S. cerevisiae* [18, 19]. However, the structural details of the heterodimeric complex and the potential role of Nus1 in enhancing the activity of either Rer2 or Srt1 remain unexplored.

1.3 Specific aim of this study

Protein glycosylation is a critical post-translational modification that plays essential roles in protein folding, oligomerization, quality control, sorting, and transportation [20]. Mutations in DHDDS or NgBR have been implicated in human diseases resulting from disruptions in glycosylation pathways [21]. Unlike animals that have a single pair of DHDDS/NgBR complex, and yeast possesses two DHDDS-like enzymes Rer2 and Srt1 along with one NgBR-like protein Nus1, plants have multiple DHDDS-like and NgBR-like proteins with sequence homology. In this study, we selected *C. kanehirae*, a tree that contains two DHDDS-like proteins (RWR89830.1 and RWR78504.1) and three NgBR-like proteins (RWR97124.1, RWR71884.1, and RWR93565.1) based on genomic sequences [22] as a model system, which are hereafter shorten as RWR89, RWR78, RWR97, RWR71, and RWR93, respectively, to investigate their enzymatic activities and

doi:10.6342/NTU202502385

quaternary compositions. We aimed to characterize these cis-PT homologues in *C. kanehirae* to understand the mechanisms underlying dolichol and polyprenol biosyntheses in plants. Additionally, the comparative analysis of their motif sequences and phylogenetic tree analysis offer insights into the evolutionary pathways of these cis-PTs.

2. MATERIALS AND METHODS

2.1 Materials

IPP, FPP, and potato acid phosphatase were procured from Sigma (St. Louis, MO, USA). PrimeSTAR® GXL DNA polymerase, utilized for polymerase chain reaction (PCR), In-Fusion® HD Cloning Kit for molecular cloning, and TALON® Metal Affinity Resin for protein purification were obtained from Takara Bio (Shiga, Japan). The plasmid mini-prep and DNA gel extraction kits were purchased from Qiagen (Hilden, Germany). Nickel nitrilotriacetic acid (Ni-NTA) resin was purchased from EBL (Taipei, Taiwan), and StrepTrap XT column and Superdex 200 Increase 10/300 GL column were obtained from Cytiva (Marlborough, USA). Polyprenol mixture standard (comprising 13–21 prenyl units) was acquired from Avanti Polar Lipids (Alabama, USA). ZORBAX XDB-C18 (4.6 × 75 mm, 3.5 μm) reversed-phase column for high-performance liquid chromatography (HPLC) analysis was purchased from Agilent (Santa Clara, CA, USA). All commercial buffers and reagents were of the highest quality.

doi:10.6342/NTU202502385

2.2 Expression and purification of the recombinant RWR89

The gene encoding the full-length of RWR89 synthesized by MORREBIO (Taiwan) with optimized codons for *E. coli* expression was ligated with the pET32a vector between the *Eco*RI and *Xho*I restriction sites for expressing the recombinant protein with N-terminal Thioredoxin (TRX), hexa-His tag (His_6), and TEV protease cutting site (TEVp). The *E. coli* BL21 competent cells transformed with the TRX-His_6-TEVp-RWR89 construct were plated on LB-agar under 100 mg/L ampicillin selection. The selected colonies were inoculated into 20 mL of LB medium with constant shaking at 200–250 rpm overnight. The culture was used to inoculate 2L of 2xYT medium (16 g/L tryptone, 10 g/L yeast extract, and 5 g/L NaCl) containing 100 mg/L ampicillin with constant shaking at 200–250 rpm at 37°C until the cell density reached OD_{600} of 0.6–0.8. The protein production was induced by adding 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) with constant shaking at 180 rpm at 16°C for 16–20 h. The cells were harvested by centrifugation (6500 rpm for 20 min), resuspended in buffer A (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1mM $MgCl_2$, 0.1% Triton X-100, and 2 mM TCEP), and disrupted by a French-press instrument (Constant Cell Disruption System) at 20,000 psi. Cell lysate was centrifuged at 40,000xg for 45 min at 4°C to remove debris, and the supernatant was loaded onto an open column with 5 mL TALON® Metal Affinity Resin. After washing with 100 mL of buffer A with 10 mM imidazole, the target protein was eluted with 25 mL

doi:10.6342/NTU202502385

of buffer A containing 250 mM imidazole, followed by dialysis against 5 L buffer A to remove imidazole and subjected to TEV protease to cleavage the TRX-His_6-tag at 4°C overnight. The mixture was subjected to another open column of 5 mL TALON® Metal Affinity Resin, and the untagged RWR89 was eluted in the imidazole-free buffer A.

2.3 Expression and purification of the recombinant RWR78, sRWR71, sRWR93, RWR78/sRWR71, RWR78/sRWR93 and RWR89 using yeast

Genes of RWR78, sRWR71 (aa 115–329), and sRWR93 (aa 61–275) (s preceding RWR indicates a shorter form without the membrane-binding fragment) were codon-optimized and synthesized by Genomics Co. (Taiwan). Genes of full-length RWR78 and sRWR71 or sRWR93 were amplified by PCR and constructed into pESC-URA yeast bicistronic vector by using In-Fusion® HD Cloning Kit (Takara Bio, Japan). The PCR product of RWR78 or RWR89 with N-terminal His_6-tag encoding sequence was subcloned into EcoRI/NotI restriction sites, while the PCR product of sRWR71 or sRWR93 with N-terminal Strep-tag encoding sequence were respectively subcloned into BamHI/SalI sites. Specific primers used in all constructs are listed in Table S1. Yeast INVSc1 strain for protein overexpression was transformed with either pESC-RWR78, pESC-sRWR71, pESC-sRWR93, pESC-RWR78/sRWR71, pESC-RWR78/sRWR93 or pESC-RWR89 plasmid, and the cells were plated on solid SC minimal medium without uracil and grown at 30°C for two days. Colonies were grown in 20 mL SC minimal

doi:10.6342/NTU202502385

medium containing 2% glucose at 30°C overnight, followed by inoculation into fresh 2 L SC minimal medium with 2% glucose until OD_{600} = 0.4. Cultures were centrifuged at 1500xg for 5 min, resuspended in fresh SC minimal medium containing 2% galactose, and incubated at 30°C for 1–4 days for induction of protein expression. Stationary-phase of yeast cells were harvested by centrifuged at 3300xg for 5 min, followed by washing with sterile water once, and dissolved in buffer A. The cells were disrupted by a French-press instrument (Constant Cell Disruption System) at 22,000 psi, and then the cell lysate was centrifuged at ~40,000xg for 45 min at 4°C to remove debris. From the supernatant, His_6-tagged RWR89, His_6-tagged RWR78 and heteromeric complexes were respectively purified by Ni-NTA resin using a 10–250 mM imidazole gradient. His_6-tagged RWR89 was eluted at 30–50 mM imidazole, His_6-tagged RWR78 was eluted at 20–40 mM imidazole, and complexes were eluted at ~70 mM imidazole. For purifying the Strep-tagged sRWR71 or sRWR93, StrepTrap XT column (Cytiva, Marlborough, USA) was used. The supernatant was applied onto a StrepTrap XT column, washed with 50 mL of buffer A to remove unbound proteins, and the Strep-tagged sRWR71 or sRWR93 was eluted with 50 mM biotin in buffer A. The affinity-purified protein was further purified using a size-exclusion chromatography and the purity was judged by SDS-PAGE.

2.4 Size-exclusion chromatography analysis

Superdex 200 Increase 10/300 GL column (Cytiva) was used to estimate the protein

doi:10.6342/NTU202502385

molecular weights. Proteins were centrifuged at 21000xg for 10 min, loaded onto the column and analyzed at a flow rate of 0.75 mL/min by using buffer A. The chromatography profiles were compared with that of the standard markers containing 1 mg/mL ribonuclease A (13.7 kDa), conalbumin (75 kDa), aldolase (158 kDa) and thyroglobulin (669 kDa) to determine protein molecular weights.

2.5 Cis-PTs activity assay

EnzChek Pyrophosphate assay kit (Thermofisher, MA, USA) was used to detect the continuously PPi release by the enzymatic reactions. The activity was assayed in buffer containing 100 mM Hepes, pH 7.5, 50 mM KCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 0.2 mM MESG, 1 U/mL PNP, and 0.03 U/mL pyrophosphatase, with either 1.95 μM of RWR89, 5.89 μM of RWR78, 2.27 μM of RWR78/sRWR71, or 1.26 μM of RWR78/sRWR93. To determine the K_m values of GPP, FPP, and GGPP, 50 μM IPP was utilized, and K_m of IPP was measured in the presence of 10 μM FPP. The plots of initial rates versus substrate concentrations were used to determined K_m and k_{cat} by fitting the data with Michaelis-Menten equation. All measurements were performed in triplicate at 30°C.

2.6 Yeast complementary assay

The rer2Δ mutant yeast strain was transformed with the empty vector pYES2, pYES2-DHDDS, pYES2-RWR89, the empty vector pESC-URA, pESC-RWR78, pESC-RWR71,

doi:10.6342/NTU202502385

pESC-RWR93, pESC-RWR78/RWR71, and pESC-RWR78/RWR93, respectively. The transformants were grown at 30°C in SC minimal medium containing 2% glucose until OD_{600} = 0.4, followed by centrifugation at 1500xg for 5 min and resuspension in fresh SC minimal medium containing 2% galactose. After growing at 30°C for 1–4 days to induce protein expression, yeast cultures were diluted to OD_{600} = 0.4 and streaked onto solid YP-2% galactose medium and incubated at 23 or 37°C for 5 days. The *S. cerevisiae* wild type strain INVsc1, which typically grew at 37°C, was utilized as the positive control, and non-transformed rer2Δ mutant strain was used as the negative control.

2.7 Polyisoprenoids extraction from yeast rer2Δ strain

Yeast rer2Δ cells were transformed with the empty vector pESC-URA, pYES2-RWR89, pESC-RWR78, pESC-RWR78/RWR71, and pESC-RWR78/RWR93, respectively. Cells were plated on solid SC minimal medium without uracil and grown at 30°C for two days. Colonies were grown in 20 mL SC minimal medium at 30°C overnight, followed by inoculation into fresh 1 L SC minimal medium until OD_{600} = 0.4. Cells were centrifuged at 1500xg for 5 min and resuspended in fresh SC minimal medium containing 2% galactose. Cultures were grown at 30°C for 1–4 days to induce protein expression. Stationary-phase of yeast cells were harvested by centrifuged at 3300xg for 5 min, followed by washing with sterile water once. Cell pellets were dissolved in 10 mL hydrolytic solution (2.5 g KOH was dissolved in 3.5 mL sterile distilled water, and

doi:10.6342/NTU202502385

brought to 10 mL with 99.8% ethanol), vortexed for 1 min, and incubated at 95°C for 1 h. Non-saponifiable lipids were extracted with hexane three times. The pooled extracts were further purified on a silica gel 60 column using isocratic elution with 10% diethyl ether in hexane, evaporated to dryness, and dissolved in isopropanol. Purified polyisoprenoids were examined by HPLC/UV using standards of different chain-length polyprenols for comparison by following the procedure [15].

2.8 HPLC product analysis

For HPLC analysis of the products extracted from rer2Δ yeast expressing RWR89, RWR78/RWR71 or RWR78/RWR93, an established protocol was used [15]. A dual-pump HPLC device (Jasco, Japan) with a ZORBAX XDB-C18 (4.6 × 75 mm, 3.5 μm) reversed-phase column (Agilent, Santa Clara, CA, USA), a Detector (spectrum range: 210–400 nm), methanol/water at 9:1 (v/v) as solvent A, methanol/isopropanol/hexane at 2:1:1 (v/v) as solvent B, and a flow rate of 1.5 mL/min controlled by a gradient programmer were used. The chain lengths of the extracted polyisoprenoids were confirmed by comparison with a polyprenols-containing mixture (13–21 prenyl groups) standard (Avanti Polar Lipids, USA).

2.9 Protein multiple sequence alignment, phylogenetic analysis, and motif elicitation of cis-PTs

The amino acid sequences of cis-PTs from microbes, yeast, plants, and animals were

doi:10.6342/NTU202502385

obtained from GenBank and aligned by using Clustal Omega and Jalview program, and the phylogenetic tree was constructed by using iTOL (Interactive Tree Of Life) website (https://itol.embl.de/). The MEME online tool (https://meme-suite.org/meme/tools/meme, Version 5.5.2), with the motif number set to 30 and the width of each motif ranging from 3 to 50 amino acids, was used to identify the motifs of cis-PTs.

2.10 Homology Modeling of homodimeric RWR89, DHDDS, RWR78 and heteromeric RWR78/RWR71 and RWR78/RWR93

The SWISS-MODEL web service (https://swissmodel.expasy.org) was utilized to perform homology modelling. The closest homology to RWR89 was *S. aureus* UPPS (PDB: 4H8E), and the closest homology of heteromeric complexes is DHDDS/NgBR complex (PDB: 6W2L). Therefore, 4H8E and 6W2L were used as templates to generate structural models. To generate the structural model of homodimeric DHDDS or RWR78, ColabFold v1.5.2 online software using AlphaFold2 algorithm (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) was used. For multiple sequence alignment and template detection, MMseqs2 method and pdb70 template_mode were selected. After structural prediction finished, the rank_1 model was chosen for further comparison.

2.11 RNA-seq analysis of five *C. kanehirae* cis-PTs

The transcriptome dataset of *C. kanehirae* from various tissues, including buds, buds

doi:10.6342/NTU202502385

of flowers, flowers, buds of leaves, young leaves, old leaves, and fruits, was obtained from previous study [22]. The fastq files were downloaded and aligned against the genome of *C. kanehirae* (GCA_003546025.1) using HISAT2 program [23]. The read counts were calculated with FeatureCounts and normalized using upper quantile normalization with edgeR [24, 25].

3. RESULTS

3.1 Multiple sequence alignment of cis-PT homologues

The amino acid sequences of five *C. kanehirae* cis-PT homologues, including RWR89, RWR78, RWR97, RWR71, and RWR93, were retrieved from GenBank. From their phylogenic tree and domain analysis as shown in Figure S2, RWR89 and RWR78 are bacterial UPPS-like protein and human DHDDS-like protein, respectively, while RWR97, RWR71 and RWR93 are homologous to NgBR. Alignment with other homomeric and heteromeric cis-PT homologues from bacteria, yeast, and humans (Figure 1) reveals that these proteins possess highly conserved amino acid sequences among the five key regions. However, crucial residues essential for catalysis and substrate binding in region I and III are notably absent in NgBR homologues. Furthermore, the RXG motif at the C-terminus of human NgBR, which is also conserved in homomeric bacterial cis-PTs and plays a critical role in enzymatic activity by interacting with the pyrophosphates of IPP and FPP via Mg^{2+} ion [14], is replaced by NXG in yeast Nus1 and the two NgBR-like proteins in

doi:10.6342/NTU202502385

C. kanehirae, RWR71 and RWR93. On the other hand, RWR97 thoroughly lacks region IV and V, which have been characterized as the dimeric interface of homomeric and heteromeric cis-PT homologues, indicating its incapacity to interact with DHDDS homologues, and therefore, it was excluded from this study.

3.2 Expression, purification, and characterization of recombinant RWR89

To characterize two *C. kanehirae* cis-PTs, their expression was initially attempted in *E. coli*. Despite the human DHDDS was successfully expressed and purified using *E. coli* in previous studies [26, 27], only RWR89 was soluble when expressed in *E. coli*. The recombinant RWR89 protein, featuring N-terminal Thioredoxin, His_6-tag and a TEV protease cutting site, was purified using TALON® Metal Affinity Resin, followed by TEV protease cleavage to remove the tags and size-exclusion chromatography to yield purified, untagged RWR89. From 2-L culture, the protein yields of RWR89 after the affinity chromatography and the size-exclusion chromatography were 16.9 and 2.4 mg, respectively. RWR89 remained stable without precipitation, and its activity was maintained under experimental conditions conducted at 4°C and room temperature, as described below. In contrast, other proteins expressed in *E. coli* exhibited visible precipitation. Therefore, yeast INVSc1 strain was chosen as the expression system for these proteins, as detailed in section 2.3. The SDS-PAGE analysis of the *E. coli*-expressed RWR89 after size-exclusion chromatography is presented in the left panel of Figure 2a

doi:10.6342/NTU202502385

and it was characterized as a homodimer (75.4 kDa) consisting of two 41.4 kDa monomers, based on the elution volume in comparison with those of molecular weight standards (Figure S3).

The purified RWR89 was subjected to kinetic characterization by using different allylic substrate, including DMAPP, GPP, FPP, or GGPP, with the homoallylic substrate IPP. A coupling assay kit was employed to convert the inorganic diphosphate (PPi) generated by RWR89 catalysis into a product that absorbs light at 360 nm, as described in the MATERIALS AND METHODS section. Through monitoring the absorbance changes at 360 nm using a microplate reader, we found that only GPP, FPP, and GGPP were identified as active allylic substrates for RWR89, and the K_m values for using GPP, FPP, and GGPP were measured to be 3.38 ± 0.45, 2.17 ± 0.20, and 0.80 ± 0.10 μM, and the k_{cat} values were 0.11 ± 0.01, 0.33 ± 0.01, and 0.57 ± 0.01 s^{-1}, respectively (Figure 2b and Table 1). Indeed, GGPP served as the most effective allylic substrate, displaying a smaller K_m and a larger k_{cat}, unlike that FPP was the preferred substrate for bacterial UPPS. The K_m and k_{cat} with respect to IPP while using GGPP were 47.24 ± 4.21 μM and 0.46 ± 0.02 s^{-1}, respectively, indicating the weaker binding affinity for IPP (Figure 2b and Table 1). The product was synthesized by incubating FPP and IPP with RWR89, followed by removal of its inorganic pyrophosphate (PPi) group using phosphatase, and the dephosphorylated product was then subjected to HPLC analysis. By spiking the HPLC

doi:10.6342/NTU202502385

analysis with the C55 product generated by *S. aureus* UPPS (SaUPPS), the chain length of the product was confirmed to be C55, eluted at the same position as that of SaUPPS (Figure 2c top two panels).

For comparison, His_6-tagged RWR89 was also expressed in *S. cerevisiae* INVSc1 cells and purified by Ni-NTA resin. The elution volume of the purified RWR89 from the size-exclusion chromatography was 24.45 mL, corresponding to the molecular weight of 75.1 kDa, consistent with a homodimer (Figure 2a, right panel). The Michaelis-Menten plot using the best substrate GGPP is also presented in Figure 2b (the right panel) to yield K_m of 1.36 ± 0.15 μM and k_{cat} of 0.53 ± 0.01 s^{-1}, similar to the results obtained from *E. coli*-expressed RWR89. The product derived from the lipid extract of the pYES2-RWR89-transformed rer2-deletion (rer2Δ) yeast cells was identical to that of the *E. coli*-expressed RWR89 (C55 as shown in Figure 2c bottom two panels).

3.3 Expression, purification, and characterization of recombinant RWR78, sRWR71, sRWR93 and their heteromeric complexes

Both NgBR homologues, RWR71 and RWR93, were predicted to possess a transmembrane domain (TM) at the N-terminus (Figure S2b). To obtain soluble proteins, RWR71 was expressed without its TM region (residues 1–127), resulting in sRWR71, while sRWR93 lacked the TM residues 1 to 73. The full-length RWR78 along with sRWR71 or sRWR93 were initially respectively co-expressed in *E. coli*. However, the

doi:10.6342/NTU202502385

purified complexes did not exhibit enzymatic activity and were readily precipitated. To obtain the functional complexes, the codon-optimized gene of N-terminal His_6-tagged RWR78 and the gene of N-terminal Strep-tagged sRWR71 or sRWR93 were individually constructed into pESC-URA bicistronic vector and co-overexpressed in *S. cerevisiae* INVSc1 cells under galactose induction. The Strep-tagged sRWR71 or sRWR93 was co-purified with the His_6-tagged RWR78 by Ni-NTA resin, suggesting their interactions and the formation of complexes. From the subsequent size-exclusion chromatography, both RWR78/sRWR71 and RWR78/sRWR93 complexes were eluted as monodisperse peaks, with the corresponding molecular weights of 75.4 and 74.8 kDa, respectively. These complexes were identified as heterodimers composed of one His_6-tagged RWR78 (40.7 kDa) and one Strep-tagged sRWR71 (24.2 kDa) or sRWR93 (24.3 kDa), in contrast with the *E. coli* co-expressed and co-purified complex of human DHDDS and the cytosolic part of NgBR (residues 73-293, named sNgBR) forming a heterotetrameric complex in solution, as indicated by the elution volume from the size-exclusion column (Figure S4a). SDS-PAGE analysis of the complexes demonstrated a near 1:1 ratio of the two subunit components (Figure 3a). His_6-tagged RWR78, alone expressed in *S. cerevisiae* and purified via Ni-NTA resin and size-exclusion chromatography, was characterized as a homodimer (83.6 kDa) (Figure 3a), consistent with the reported conformation of human DHDDS [28] (we also confirmed that *E. coli*-expressed human DHDDS formed a

doi:10.6342/NTU202502385

homodimer as shown in Figure S4b). Besides, Strep-tagged sRWR71 and sRWR93 that expressed in *S. cerevisiae* were purified using StrepTrap XT column and size-exclusion chromatography, and both were identified as homodimers with molecular weights of 67.3 kDa and 61.4 kDa, respectively. These conformations were similar to yeast Nus1 [18], but distinct from the monomeric human sNgBR [29]. Notably, when combining the purified homodimeric RWR78 with homodimeric sRWR71 or sRWR93, they were unable to form a complex, indicating that co-expression was necessary for complexation, and it appears that the individual homodimeric conformations cannot be dissociated and reformed into a heterodimer.

The enzymatic activities of purified homodimeric His_6-tagged RWR78 and its heterodimeric complex with sRWR71 or sRWR93 were subsequently characterized using pyrophosphate assay kit. The K_m value for FPP were determined with 50 μM IPP, while the K_m value for IPP were measured with 10 μM FPP. The Michaelis–Menten plots are shown in Figure 3b, and the kinetic constants are summarized in Table 1. Compared to the activities of His_6-tagged RWR78 and the complexes, the K_m values of FPP and IPP were similar to those of the complexes, but the k_{cat} values of the complexes were 14–50-fold higher than that of RWR78, indicating a significant stimulation of RWR78 activity by sRWR71 or sRWR93, with a higher stimulating effect by sRWR93. The Michaelis–Menten plots for using GPP and GGPP as the alternative trans-prenyl substrates for

doi:10.6342/NTU202502385

RWR78, RWR78/sRWR71, and RWR78/sRWR93 were also measured (Figure 3b and Table 1), but the enzyme and its complexes preferred FPP as the predominant trans-prenyl substrate.

To investigate whether the bacterial UPPS-like RWR89 could also form a heteromeric complex with sRWR71 or sRWR93, the yeast-expressed sRWR71 or sRWR93, was added to the above *E. coli*-expressed RWR89 or the yeast-expressed RWR89, but the activity of RWR89 was not changed (data not shown). Additionally, co-expression of His_6-tagged RWR89 and Strep-tagged sRWR71 or sRWR93 in yeast, similar to the approach used for co-expressing His_6-tagged RWR78 and Strep-tagged sRWR71 or sRWR93, did not result in co-purification from the Ni-NTA column (Figure not shown), indicating the absence of complex formation, consistent with no activity enhancement of RWR89 by sRWR71 or sRWR93. Collectively, RWR89 consistently formed a homodimer, regardless of the expression system (*E. coli* or yeast).

3.4 Functional analysis of two complexes in rer2-deletion (rer2Δ) *S. cerevisiae*

To investigate the function of two complexes, each was co-expressed in a rer2Δ mutant yeast strain lacking the Rer2 cis-PT. While the rer2Δ yeast exhibited impaired growth at elevated temperatures (37°C), rer2Δ yeast expressing either complex (RWR78/RWR71 or RWR78/RWR93) or human DHDDS restored growth at 37°C (Figure 4a). As previously reported, using human DHDDS alone was successfully

doi:10.6342/NTU202502385

complement Rer2 function [30, 31]. In contrast, individually transformation of rer2Δ yeast with RWR78, RWR71, or RWR93 failed to rescue growth at 37°C, indicating that neither subunit functioned independently, and the complex formation was strictly required for dolichol biosynthesis. Furthermore, HPLC analysis of lipid extracts from rer2Δ yeast co-expressing either complex revealed a comparable composition of long-chain lipids. Conversely, no long-chain products were detected in lipid extracts from rer2Δ yeast expressing only RWR78 (Figure 4b).

3.5 Homology modeling of RWR89

To investigate the catalytic mechanism of RWR89, which is the UPPS-like cis-PT that produced C55 product (Figure 2c), the homology-based structural modeling of RWR89 was generated by using the apo structure of *S. aureus* UPPS (PDB: 4H8E), the closest homologue based on the conservation of amino acid sequences, as the template. Consistent with the experimental data confirming RWR89 as a homodimer (Figure 2a), the 3D structure was constructed to reveal the canonical catalytic domain (Figure 5a). To further interpret the roles of active-site residues, the modeled structure of RWR89 was superimposed with the crystal structure of *E. coli* UPPS (EcUPPS), which has been extensively characterized in our previous studies.

With IPP serving as the building blocks for consecutive cis-condensations, RWR89 utilized all three trans-prenyl diphosphates including GPP, FPP, and GGPP as substrates

doi:10.6342/NTU202502385

but did not accept DMAPP as either a primary allylic substrate or as condensational units, indicating RWR89 is unlikely to synthesize short-prenyl diphosphates, such as neryl diphosphate (NPP) or nerylneryl diphosphate (NNPP). In contrast, SlcPT1, a neryl diphosphate synthase in *S. lycopersicum*, prefers DMAPP as the allylic substrate and produces NPP [32]. Structural analysis revealed that Gly144 and Trp322 of RWR89 in Figure 5a are aligned with Gly46 and Trp221 of EcUPPS, but not with Ile106 and Phe276 of SlcPT1, respectively, supporting the similarity of RWR89 to EcUPPS in its preference for longer trans-prenyl diphosphates, particularly GGPP. Additionally, Leu235 of RWR89 corresponds to Leu137 of EcUPPS, a residue known to limit chain length extension and mediate the product chain length [9].

Regarding the inability of RWR89 to form heteromeric complexes with sRWR71 or sRWR93, as demonstrated above, it can be attributed to its inherent structural features. Specifically, RWR89 contains essential catalytic residues, including the C-terminal RXG (RFG) motif, which are sufficient to establish an independent active site. Consequently, it does not require additional amino acids from sRWR71 or sRWR93 to achieve catalytic functionality. In contrast, RWR78 depends on the supplementary amino acids provided by sRWR71 or sRWR93, as described below.

3.6 Homology modeling of RWR78/RWR71 and RWR78/RWR93 heterodimeric complexes

doi:10.6342/NTU202502385

To elucidate the catalytic roles of key amino acids in the complexes, homology models of the RWR78/RWR71 and RWR78/RWR93 complexes were generated using the crystal structure of the human DHDDS/NgBR complex (PDB: 6W2L) as the template. Both complexes were constructed as heterodimers (Figure 5b) based on the experimental observations (Figure 3a). The expanded view of the active sites on the right-hand side reveals a high degree of conservation among residues involved in canonical cis-PT catalysis. The S1 substrate-binding site in RWR78 encompasses Arg98, Arg99, and Arg146, while the S2 site comprises Arg289 and Arg295. The missing catalytic residues necessary for activity are contributed by the C-terminal NXG motif (NYG) from RWR71 or RWR93. Specifically, the Asn326 side-chain nitrogen and Gly328 backbone nitrogen from RWR71, or the Asn272 side-chain nitrogen and Gly274 backbone nitrogen from RWR93, interact with the β-phosphate of IPP at the S2 site, in contrast to the RXG motif (RLG) of NgBR, which facilitates substrate binding and enhances DHDDS catalytic activity. On the other hand, the homodimeric RWR89 contains the complete active site, including the C-terminal RXG (RFG) motif, rendering it self-sufficient and obviating the need for a partner to complete the catalytic machinery.

Several distinctions exist between the human DHDDS/NgBR complex and the two complexes. In NgBR, the conserved C-terminal RLG motif, particularly Arg290, plays a pivotal role in enhancing the catalytic activity of the heteromeric DHDDS/NgBR

doi:10.6342/NTU202502385

complex [14]. Notably, the R290H mutation of NgBR has been implicated in congenital glycosylation disorder [33]. The surface loop location of Arg290 provides structural flexibility, enabling interactions with the catalytic pocket and facilitating allosteric activation of DHDDS. In contrast, the NYG motif of RWR71 and RWR93 features a shorter Asn side chain, which is necessary to maintain the close contact with the diphosphate moiety of IPP. Furthermore, a comparison of Figures 5b and 5c highlights that Arg98, Ile103, and Arg295 of RWR78 align with Arg37, Lys42, and Arg211 of human DHDDS, respectively. These residues are critical for activity, as the R37H, K42E, and R211Q mutations in DHDDS disrupt the substrate binding and thereby significantly reduce cis-PT activity [15, 16]. Arg37 and Arg211, which are positively charged and located near the active site of human DHDDS, are associated with developmental epileptic encephalopathies [34]. Lys42 in DHDDS forms a salt bridge with Glu234, but in the disease-associated K42E mutant linked to retinitis pigmentosa [35, 36], it interacts with Arg38 instead, destabilizing the active site [16]. However, in RWR78, Ile103 could not form a salt bridge with any residue and thus does not contribute to the active-site stabilization.

To further elucidate the mechanism of forming RWR78/RWR71 and RWR78/RWR93 heterodimers, as opposed to the heterotetramer of DHDDS/NgBR in solution, their interfaces were analyzed. In the RWR78/RWR71 complex (Figure 6a), hydrogen-bond

doi:10.6342/NTU202502385

pairs were identified at the interface, including T220/E257, Q228/E252, Q228/F255, Y231/A233, T294/L299, R295/E298, S297/R295, N298/R295, W302/Y226, T305/L287, T305/G288, and L309/R284. Similarly, polar interactions at the interface of the RWR78/RWR93 complex (Figure 6b) included T220/E203, Q228/F201, Y231/A179, K234/C183, Q235/N197, E264/S180, T294/L245, R295/E244, S297/R241, N298/R241, T305/L233, T305/G234, and L309/R230. The high sequence identity (87%) between RWR71 (residues 141–329) and RWR93 (residues 92–275) and the structural similarity of their models (RMSD = 0.282 Å) further support that either RWR71 or RWR93 can form a complex with RWR78. In contrast, the heterotetrameric assembly of DHDDS/NgBR (PDB: 6Z1N) features an extended C-terminal helix-turn-helix motif in DHDDS (highlighted in orange in Figure S5), which facilitates additional interactions between the two DHDDS, promoting heterotetramerization with NgBR [16]. The absence of this additional C-terminal motif in RWR78 prevents the formation of a heterotetramer with RWR71 or RWR93, consistent with observations shown in Figure 3a.

To elucidate why a DHDDS-like protein forms a heteromeric complex with a NgBR-like protein rather than existing as a homodimer, the homodimeric structures of DHDDS and RWR78 were predicted using the AlphaFold2 algorithm (Figure S6), and their homodimeric interfaces were compared to their respective heterodimeric interfaces. In the predicted DHDDS homodimer, the interface in region VI and V exhibits only a few

doi:10.6342/NTU202502385

hydrogen bonds, including interactions of Arg159 with Trp218, Gln219, and Ser221 of the opposite subunit in region VI, as well as hydrogen bonds between V210/L225, S213/S221, and D214/S221 in region V. In contrast, the heterodimeric interface of the DHDDS/NgBR complex features a greater number of hydrogen bonds, including Arg159 of DHDDS with Trp256 and Asp237 of NgBR in region VI, as well as interactions such as V210/I263, R211/E262, S213/R259, D214/R259, S221/L251, S221/G252, S221/K297, W218/K297, and L225/D248 in region V. Similarly, at the predicted RWR78 homodimeric interface, region IV lacks hydrogen bonds entirely, while region V has only a few hydrogen-bond pairs, including T294/L309, S297/T305, and N298/T305. In comparison, the heterodimeric interfaces of RWR78/RWR71 and RWR78/RWR93 exhibit more extensive interactions, such as Arg295/Glu298 and Arg295/Glu244, respectively, than the RWR78 homodimeric interface. These findings suggest that NgBR-like proteins act as the regulatory partners for the catalytic activity of DHDDS-like proteins by forming more stable and extensive heterodimeric interactions. However, unlike monomeric NgBR, RWR71 and RWR93 are dimeric and thus cannot dissociate into monomers to form a heterotetramer with the homodimeric RWR78 in solution.

3.7 Possible evolutionary paths for *C. kanehirae* cis-PTs and homologues

Prokaryotes exclusively possess homodimeric cis-PTs, whereas eukaryotes and archaea feature heteromeric cis-PTs, with plants displaying a notable abundance of

doi:10.6342/NTU202502385

multiple DHDDS-like and NgBR-like proteins. By characterizing the functions of *C. kanehirae* cis-PT homologues and leveraging their amino acid sequences for comparative analyses with other known cis-PT homologues, we have traced the potential evolutionary paths of these enzymes. According to the phylogenetic tree analysis of cis-PT homologues from archaea, bacteria, and eukaryotes (Figure S2), we identified their conserved motifs, which were analyzed to illustrate their distribution and diversity as shown by color-coded boxes (Figure 7). Based on these analyses, cis-PT homologues were classified into three subgroups, including homomeric UPPS orthologs, heteromeric DHDDS orthologs, and heteromeric NgBR orthologs. For clarity, we refer to these as UPPS (homodimeric cis-PTs), DHDDS (the active cis-PTs in heteromeric complexes), and NgBR (the inactive cis-PTs in heteromeric complexes). Remarkably, the plantae, including green algae, hornworts, ferns, pines, monocots and dicots, harbor all three types of orthologs. The plant-specific homomeric UPPS orthologs likely evolved from prokaryotic homomeric UPPS orthologs due to their shared motif architectures (UPPS orthologs in Figure 7), with the addition of N-terminal sequences serving as signal peptides for transporting toward specific organelles such as chloroplasts and mitochondria. For instance, SlCPT1-7 from *S. lycopersicum* excluding SlCPT3 have been experimentally identified to be localized in plastids [37], while ChrUPPS from *Chlamydomonas reinhardtii*, a well-studied

doi:10.6342/NTU202502385

unicellular green alga, was predicted to locate in mitochondria due to the presence of these directed signal peptides.

In contrast to plants and fungi, animals possess a single DHDDS and a single NgBR for dolichol biosynthesis. DHDDS orthologs in animals specifically exhibit the extended C-terminal motif compared to those in plants. On the other hand, animal NgBR orthologs preferentially contain the C-terminal RXG motif, a feature also observed in homomeric prokaryotic UPPS orthologs and plant UPPS-like cis-PTs. However, DHDDS orthologs in both animal and plant lack RXG or NXG motif (Figure 7). On the contrary, plant NgBR orthologs consistently exhibit an NXG motif, but not RXG, at the C-termini (Figure 7). The presence of NXG and RXG motifs might originate from distinct evolutionary trajectories. For instance, *Candidatus megaira*, a bacterium regarded as an endosymbiont of *Mesostigma viride*, a single-celled charophyte algae considered one of the earliest diverging members of green plant lineages, exhibits a homomeric UPPS gene (GenBank: UCM94414.1) featuring the C-terminal NXG (NFG) motif instead of RXG motif. Additionally, some NgBR orthologs in fungi also possess the NXG motif, with fungi often engaging in symbiotic relationships with plants. These suggest that the NXG motif in plant NgBR orthologs may have been acquired from certain bacteria or fungi through symbiosis and horizontal gene transfer, whereas the RXG motif in animal NgBR orthologs might be derived from typical prokaryotic homomeric UPPS orthologs.

doi:10.6342/NTU202502385

In archaea, cis-PTs could be categorized into UPPS, DHDDS, and NgBR orthologs as well, despite archaea lack organelles such as the ER and chloroplast. Among the three subgroups in the phylogenetic tree, archaeal cis-PT homologues occupy the basal clades, representing some of the earliest diverging branches with distinct phylogenetic separation from the sequences of other cis-PT homologues. These archaeal cis-PT homologues may exhibit unique structural or functional features. For instance, MA1831 from *Methanosarcina acetivorans* and MM0014 from *Methanosarcina mazei* are sequentially belonged to UPPS orthologs, yet they differ from typical UPPS functions and lack the common C-terminal RXG motif (Figure 7). MA1831 catalyzes non-canonical head-to-middle condensations to produce C35–C50 products, while MM0014 exhibits O-prenyltransferase activity, yielding C25–C40 products [38, 39]. Furthermore, the heteromeric cis-PT homologues from *Archaeoglobus fulgidus*, a thermophilic archaeon, have recently been biochemically characterized [40]. Unlike eukaryotic NgBR orthologs, Af0707, the NgBR ortholog from *Archaeoglobus fulgidus*, lacks the transmembrane domain, and its DHDDS ortholog Af1219 does not contain the extended C-terminal motif. Despite this, the complex still forms a heterotetrameric assembly, which contrasts with their structurally predicted heterodimeric configuration. To explore the potential origins of heteromeric cis-PTs from archaea, cis-PT homologues from *Candidatus Lokiarchaeota archaeon*, one of the closest relatives of eukaryotes, were investigated as well. We

doi:10.6342/NTU202502385

identified CLaUPPS3 (GenBank: JAGXOB010000163.1) belonging to DHDDS subgroup, while no NgBR ortholog was detected. Interestingly, a UPPS-like cis-PT from *Candidatus Lokiarchaeota archaeon*, named CLaUPPS4 (GenBank: JAGXNS010000099.1), possesses the RXG motif but lacks the conserved catalytic residues, suggesting that NgBR orthologs might be originated from these archaeal UPPS-like cis-PTs with altered catalytic functions. However, the functions of archaeal cis-PTs remain largely unexplored, and further studies are essential to elucidate their evolutionary relationships.

4. DISCUSSION

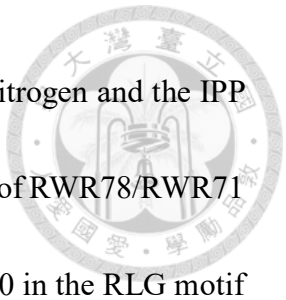
In eukaryotes, dolichols with chain lengths of C70–C120 play essential roles as glycosyl carrier lipids for the biosynthesis of glycoproteins in ER. In human cells, NgBR is crucial for forming a complex that enhances DHDDS activity, while in yeast, Nus1 forms complexes with Rer2 and potentially with Srt1. In plants, the complexation of DHDDS-like and NgBR-like cis-PT homologues is also critical for natural rubber biosynthesis in lettuce and dandelion, as well as dolichol biosynthesis in tomato [41-44]. Conversely, bacteria exclusively possess homomeric cis-PTs, demonstrating a significant difference in the composition of active cis-PTs between prokaryotes and eukaryotes. However, heteromeric cis-PTs have also been identified in methanogenic archae. For example, *Methanosarcina acetivorans* contains a UPPS (MA3723) and a NgBR/Nus1-

doi:10.6342/NTU202502385

like protein (MA4402) [14, 38], while *Methanosarcina mazei* employs a heteromeric cis-PT for the biosynthesis of glycosyl carrier lipids [45]. Here, we present an example of multiple unpaired and paired cis-PT homologues in the plant *C. kanehirae*, elucidating their kinetics and functional roles.

Two NgBR homologues RWR71 and RWR93 in *C. kanehirae* utilize a C-terminal NYG motif to compensate for the absence of this motif in the DHDDS homologue RWR78, enabling complex formation for dolichol biosynthesis. On the contrary, a C-terminal RFG motif is present in the homodimeric RWR89, bacterial SaUPPS and EcUPPS, and the plant AtcPT1, one of the cis-PTs in *Arabidopsis thaliana* [46], which catalyzes the synthesis of medium-chain polyisoprenoids without requiring pairing with an NgBR homologue. The crystal structure of EcUPPS have demonstrated that Arg242 of the RFG motif plays a critical role in binding the diphosphate group of IPP [14]. Additionally, structural evidence from the decaprenyl diphosphate synthase of *Mycobacterium tuberculosis* [47] and the short-chain FPP synthase (cis-FPPS) of *Solanum habrochaites* [48] further supports the significance of the C-terminal motif in IPP binding and catalysis. NgBR, on the other hand, contains the RLG motif as its C-terminal RXG sequence (Figure 1). Arg290 of the RLG motif in NgBR is likely involved in IPP binding, and its mutation to His has been implicated in disease [33]. In RWR71 (Asn272) and RWR93 (Asn326), the NYG motif features an Asn with a shorter side chain

doi:10.6342/NTU202502385



than Arg, allowing an optimal distance between the Asn side-chain nitrogen and the IPP β-phosphate for hydrogen bond formation (see the modeled structures of RWR78/RWR71 and RWR78/RWR93 in Figure 5a and 5b). Furthermore, while Arg290 in the RLG motif of NgBR forms a salt bridge with Glu288, the equivalent position in RWR71 or RWR93 is occupied by Arg270 or Arg324, respectively. If Asn272 or Asn326 were replaced by Arg, the proximity of two positively charged Arg residues would result in electrostatic repulsion. Therefore, the presence of Asn in the NYG motif of RWR71 and RWR93 is functionally advantageous and structurally ideal.

Another notable difference between human DHDDS/NgBR complex and the RWR78/RWR71 and RWR78/RWR93 complexes is their oligomeric states in solution. While the human DHDDS/NgBR complex forms a heterotetramer, the RWR78/RWR71 and RWR78/RWR93 complexes are heterodimers. This structural divergence is likely attributed to the extended C-terminal motif of DHDDS, which facilitates additional interactions between two DHDDS protomers, promoting the dimer-of-heterodimers arrangement.

Moreover, compared to bacterial UPPS, RWR89 possesses an N-terminal putative chloroplast transit peptide (TP), likely serving as a chloroplast-targeting signal to direct this cis-PT to a compartment (chloroplast) where its C55 product is required. According to the RNA-seq analysis [22] as shown in Figure S7, RWR89 had a significantly higher

doi:10.6342/NTU202502385

expression level in old leaves compared to other tissues, suggesting its involvement in plant secondary metabolism within chloroplasts. Similarly, a homodimeric cis-PT from *Lilium longiflorum* anthers producing C40/C45 polyisoprenoids was previously characterized by our lab [31]. Contrarily, the NgBR homologues RWR71 and RWR93 contain a TM domain that anchor them onto the cell membrane, akin to NgBR, a known membrane protein. The role of NgBR or Nus1 as either an accessory subunit facilitating ER membrane docking or as a structural component contributing to catalysis and substrate binding remains unresolved. Experiments indicates that the loss of NgBR does not alter the association of DHDDS with ER membrane fractions [13], and mutations in the C-terminal region of NgBR result in a congenital disorder of glycosylation syndrome due to defective dolichol synthesis [33], underscoring the critical role of NgBR in complementing DHDDS activity. Besides, the expression profiles of *C. kanehirae* cis-PTs highlight their functional distinctions: RWR78 was highly expressed in flower and leaf buds, RWR93 exhibited pronounced expression in fruits, and the expression of RWR71 was slightly elevated in fruits, flower buds, and leaf buds. Flowers and fruits are distinctive organs exclusive to angiosperms, with proper protein glycosylation being fundamental to their developmental processes [49]. Aberrant glycosylation of specific proteins has been implicated in growth deficiencies [50], underscoring the pivotal role of

doi:10.6342/NTU202502385

protein glycosylation in ensuring the normal development and functionality of these organs.

Lastly, the phylogenetic analysis and motif structure of cis-PTs provide insights into the potential evolutionary paths of DHDDS and NgBR orthologs. Plant NgBR orthologs containing the NXG motif may have originated from certain fungi or bacteria through processes such as symbiosis and horizontal gene transfer, while animal NgBR orthologs likely evolved directly from UPPS-like bacterial cis-PTs with the RXG motif. In conclusion, this study presents the kinetic and functional characterizations of both homomeric and heteromeric cis-PT homologues in *C. kanehirae*. These findings may be broadly applicable to other plant species with multiple cis-PT homologues, including those with either catalytic activity or roles in supporting enzymatic function.

doi:10.6342/NTU202502385

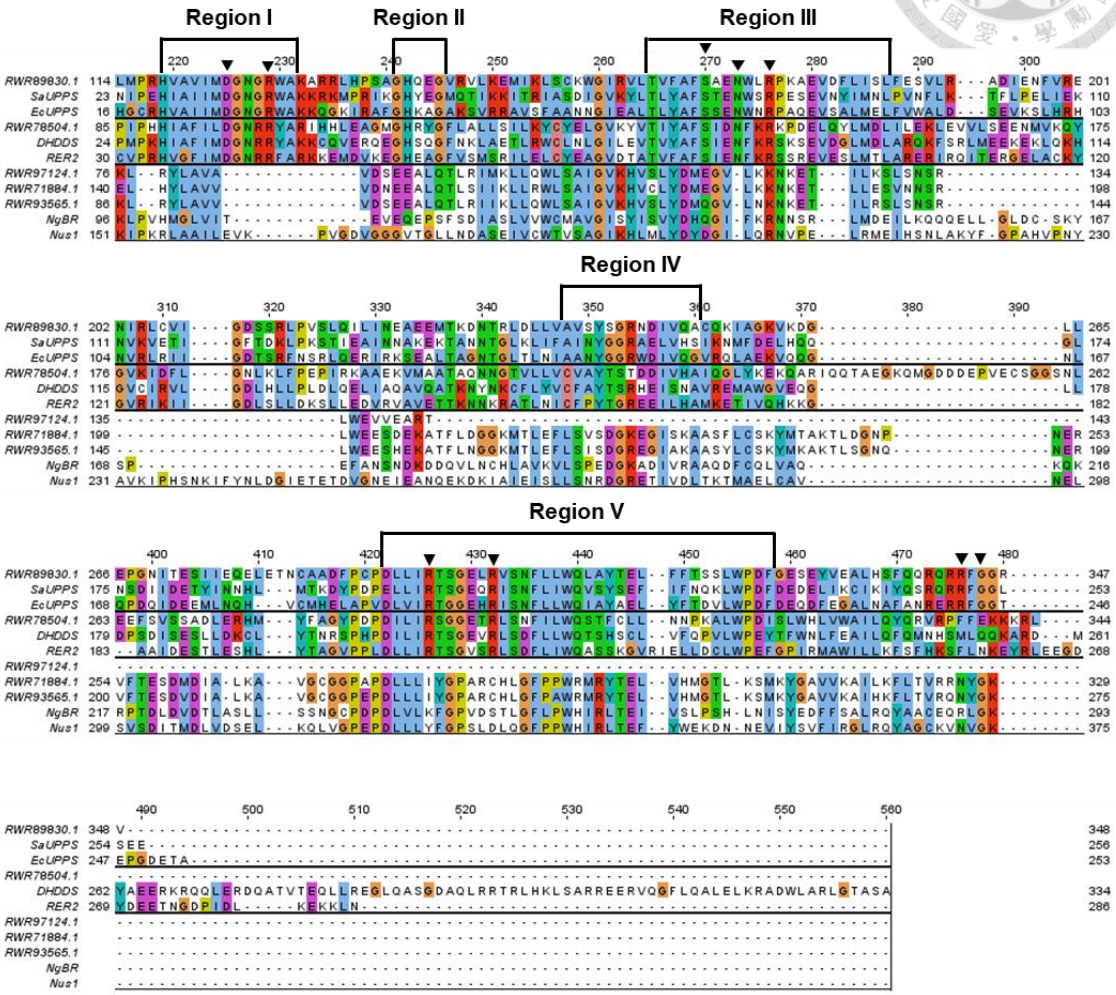
TABLE

Table 1. Kinetic parameters of the purified RWR89, RWR78, RWR78/sRWR71 and RWR78/sRWR93 with respect to GPP, FPP, or GGPP and IPP substrates.

Enzymes and substrates	K_m (μM)	k_{cat} (s^{-1})
***E. coli*-expressed RWR89**		
IPP	47.24 ± 4.21	0.46 ± 0.02
GPP	3.38 ± 0.45	0.11 ± 0.01
FPP	2.17 ± 0.2	0.33 ± 0.01
GGPP	0.80 ± 0.10	0.57 ± 0.01
***S. cerevisiae*-expressed RWR89**		
GGPP	1.36 ± 0.15	0.53 ± 0.01
RWR78		
IPP	10.64 ± 4.73	0.004 ± 0.0005
GPP	4.85 ± 1.31	0.0005 ± 0.00005
FPP	0.50 ± 0.21	0.002 ± 0.0002
GGPP	2.06 ± 0.62	0.0007 ± 0.00006
RWR78/sRWR71		
IPP	9.52 ± 1.29	0.0628 ± 0.0021
GPP	1.25 ± 0.23	0.0206 ± 0.0008
FPP	0.66 ± 0.12	0.0562 ± 0.0018
GGPP	0.74 ± 0.18	0.0496 ± 0.0025
RWR78/sRWR93		
IPP	16.78 ± 3.72	0.1112 ± 0.0072
GPP	1.78 ± 0.17	0.0383 ± 0.0009
FPP	0.36 ± 0.10	0.1083 ± 0.004
GGPP	0.74 ± 0.15	0.1004 ± 0.0039

doi:10.6342/NTU202502385

FIGURES



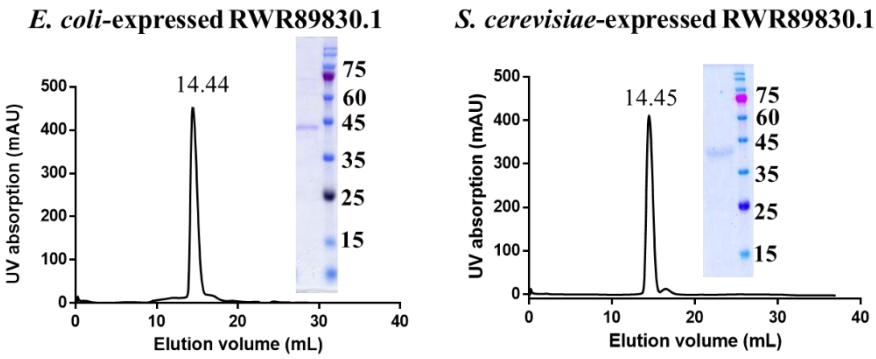
doi:10.6342/NTU202502385

Figure 1. Multiple sequence alignment of five *C. kanehirae* cis-PT homologues.

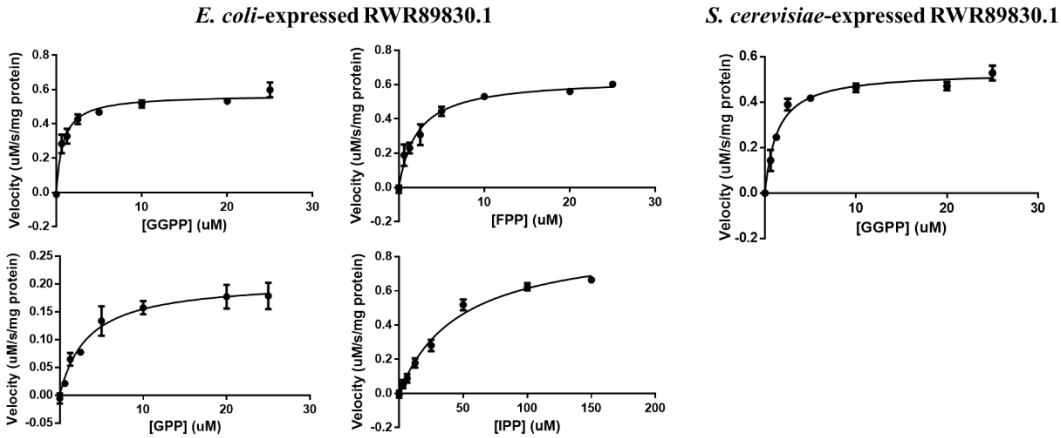
Amino acid sequence alignment of five *C. kanehirae* cis-PT homologues against the conserved cis-PTs from bacteria, yeast, and human. Proteins involved in this figure are prokaryotic cis-PTs: *E. coli* UPPS (EcUPPS) and *S. aureus* UPPS (SaUPPS), orthologs of heteromeric DHDDS subunit: human DHDDS and yeast Rer2, and orthologs of heteromeric NgBR subunit: human NgBR and yeast Nus1. Black triangles indicate the key residues involved in the catalytic reactions and substrates binding.

doi:10.6342/NTU202502385

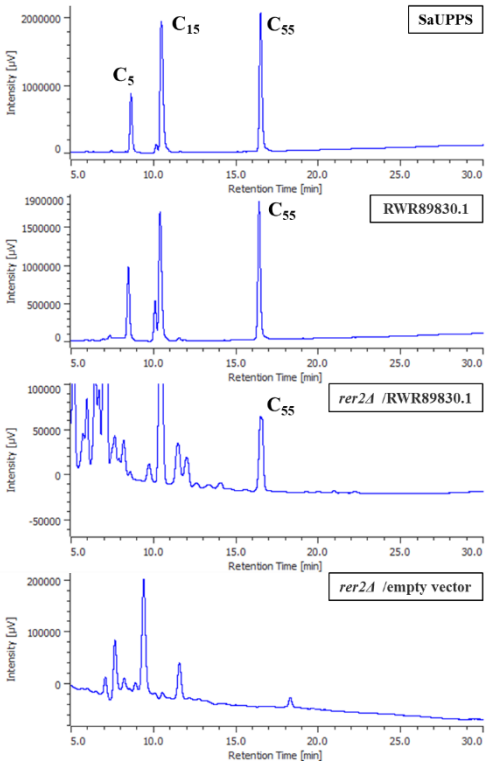
(a)



(b)



(c)



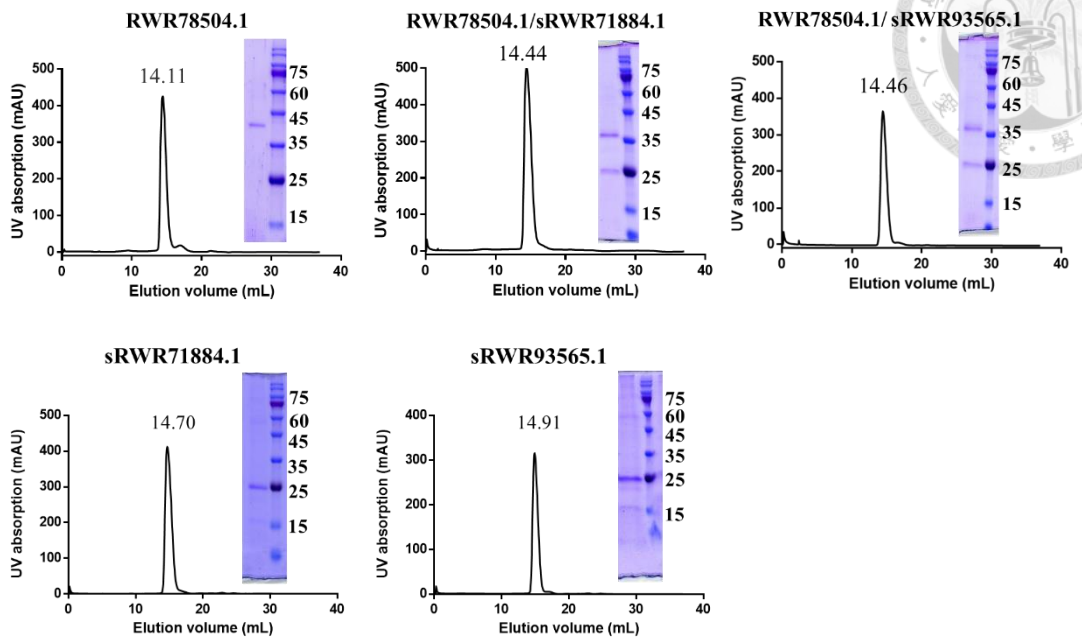
doi:10.6342/NTU202502385

Figure 2. Purification and characterization of the *E. coli* and yeast-expressed recombinant RWR89.

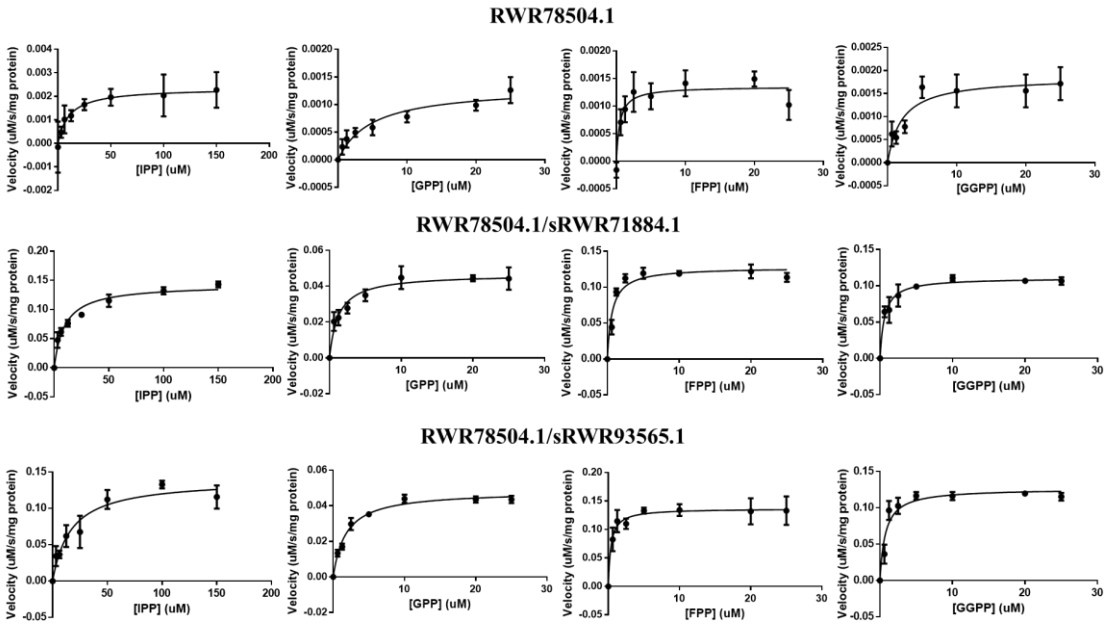
(a) SDS-PAGE analysis of the *E. coli* and yeast-expressed RWR89, purified from size-exclusion chromatography as a homodimer (75.4 kDa and 75.1 kDa, respectively, based on the elution volume). (b) Michaelis-Menten plots of *E. coli*-expressed RWR89 with different concentrations of FPP, GGPP, or GPP and 50 μM IPP to determine the K_m and k_{cat} values (K_m for IPP was measured with 10 μM GGPP). For comparison, the Michaelis-Menten plot of yeast-expressed RWR89 using GGPP is also shown in the right bottom panel. All measurements were performed in triplicate at 30°C. (c) Polyisoprenoid product analysis of RWR89 incubated with IPP and FPP at 30ºC overnight. From top to bottom, the product of the purified SaUPPS, the purified *E. coli*-expressed RWR89, and the polyisoprenoid products in the pYES-RWR89-transformed rer2Δ yeast cells were extracted and analyzed by HPLC to be C55, respectively, as compared to the absence of C55 in the rer2Δ cells with empty vector.

doi:10.6342/NTU202502385

(a)



(b)



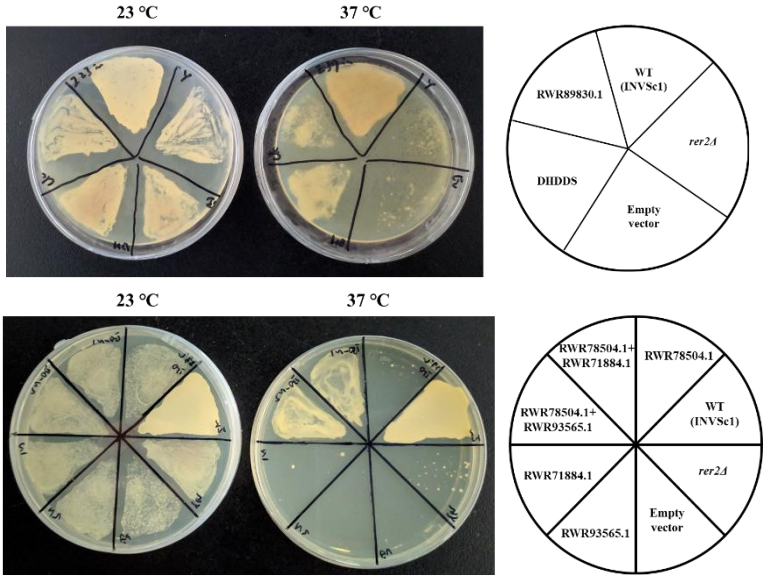
doi:10.6342/NTU202502385

Figure 3. Purification and characterization of the recombinant RWR78, RWR78/sRWR71, RWR78/sRWR93, sRWR71, and sRWR93.

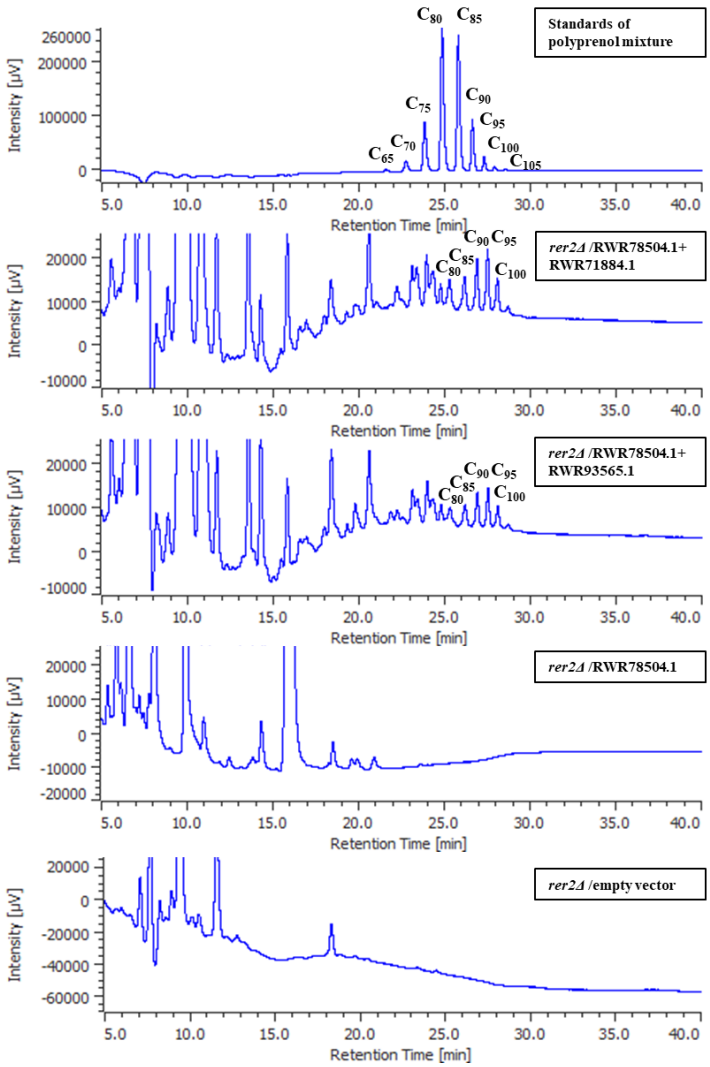
(a) SDS-PAGE analysis of RWR78, RWR78/sRWR71, RWR78/sRWR93, sRWR71, and sRWR93, purified from the size-exclusion chromatography as homodimer (83.6 kDa), heterodimer (75.4 kDa), heterodimer (74.8 kDa), homodimer (67.3 kDa), and homodimer (61.4 kDa), respectively, based on their elution volumes. (b) Michaelis-Menten plots of RWR78, RWR78/sRWR71 and RWR78/sRWR93 using IPP and GPP, FPP, or GGPP as substrates (from left to right panels) to determine the K_m and k_{cat} values as listed in Table 1. All measurements were performed in triplicate at 30°C.

doi:10.6342/NTU202502385

(a)



(b)



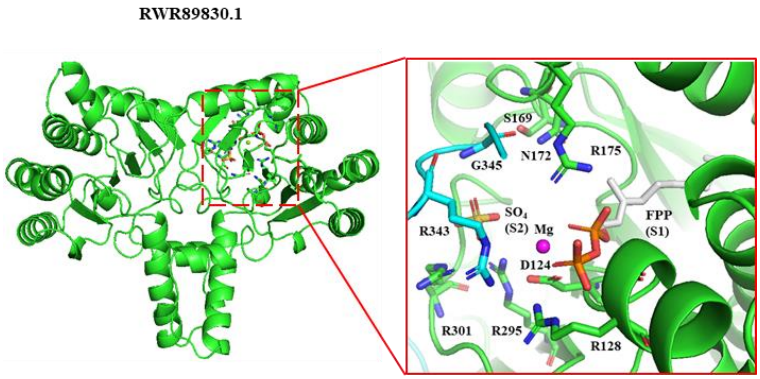
doi:10.6342/NTU202502385

Figure 4. Analysis of complementation of RWR89, RWR78/RWR71 and RWR78/RWR93 to yeast rer2.

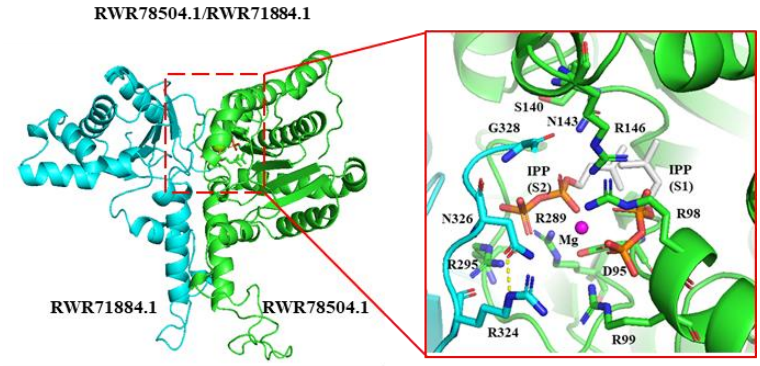
(a) The rer2Δ yeast and that transformed with the empty vector pYES2, but not the wild type (WT) yeast strain INVSc1 showed growth defect at higher temperature 37ºC. Expression of RWR78/RWR71, RWR78/RWR93, or human DHDDS in the rer2Δ yeast rescued the growth defect at 37ºC, while expression of individual RWR78, RWR71 or RWR93 did not complement the Rer2 function. (b) Polyisoprenoid products in the rer2Δ yeast cells transformed with empty vector, pESC-RWR78, pESC-RWR78/RWR71 and pESC-RWR78/RWR93, respectively, analyzed by HPLC, indicating that only RWR78/RWR71 and RWR78/RWR93 produced longer C80–C100 products.

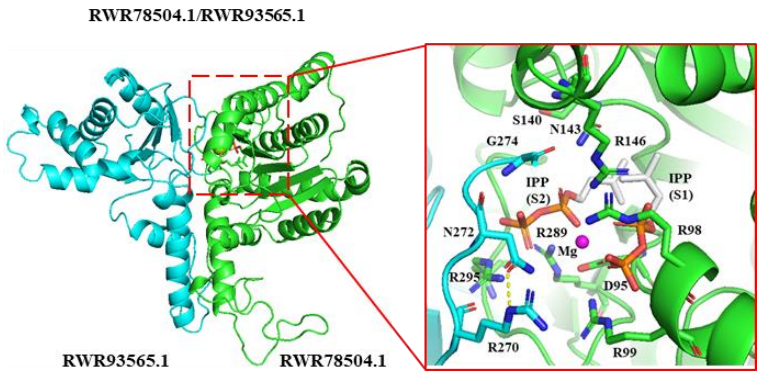
doi:10.6342/NTU202502385

(a)

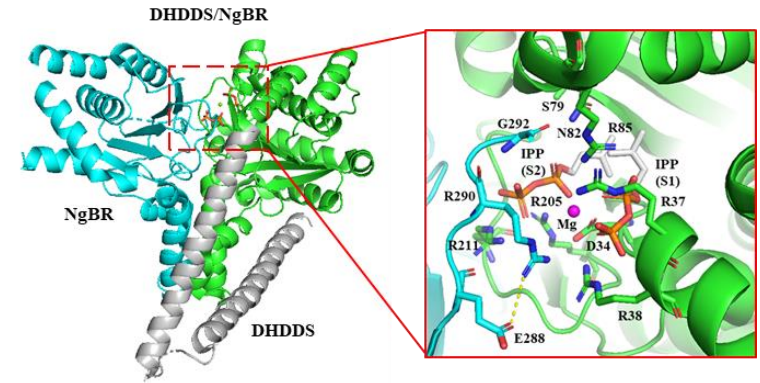


(b)





(c)



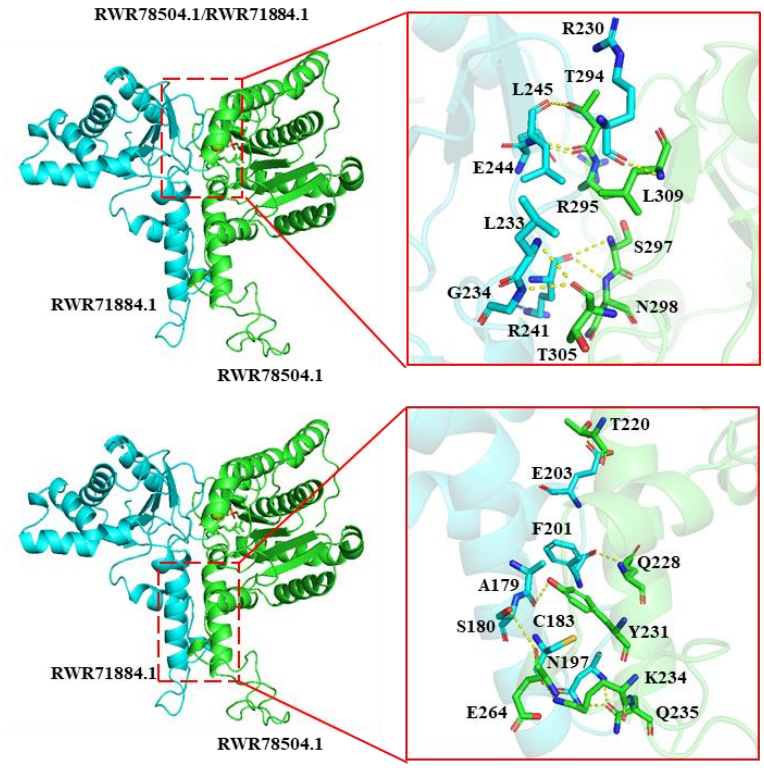
doi:10.6342/NTU202502385

Figure 5. Homology modeling of RWR89, RWR78/RWR71 and RWR78/RWR93 structures.

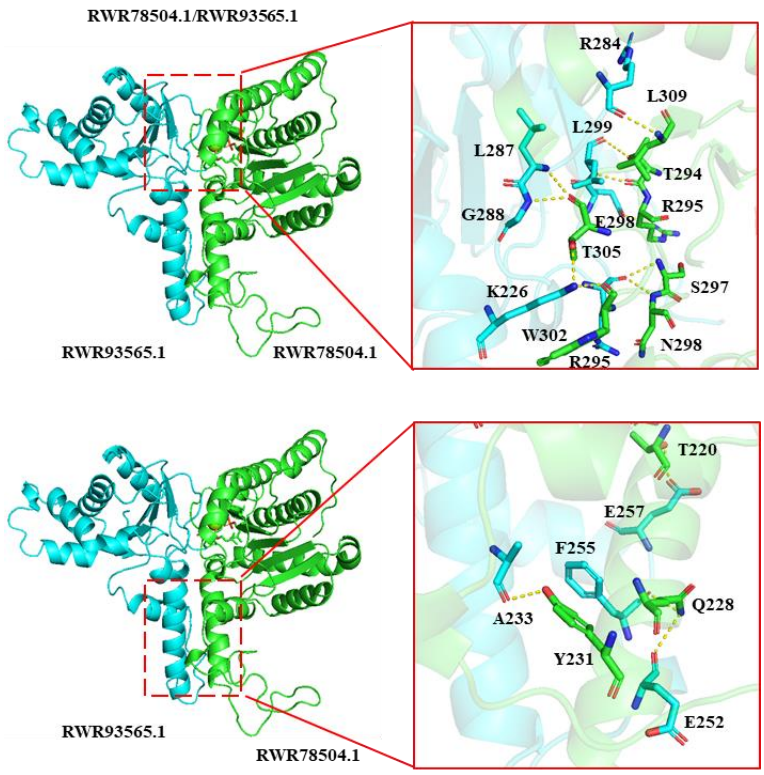
(a) The structural model and the detailed active-site architectures of RWR89. The ligand and the metal ion Mg^{2+} are originated from the superimposed crystal structure of *E. coli* UPPS. (b) The structural models and the active-site architectures of RWR78/RWR71 and RWR78/RWR93 using human DHDDS/NgBR complex (PDB: 6W2L) as template. RWR78 is colored in green, while RWR71 and RWR93 are colored in cyan. The ligands 2 IPP and the metal ion Mg^{2+} are originated from the superimposed crystal structure of human DHDDS/NgBR. (c) For comparison, the DHDDS structure and the active-site architecture are shown. NgBR is colored in cyan, DHDDS is colored in green, and the C-terminal helix-turn-helix motif of DHDDS is colored in gray.

doi:10.6342/NTU202502385

(a)



(b)

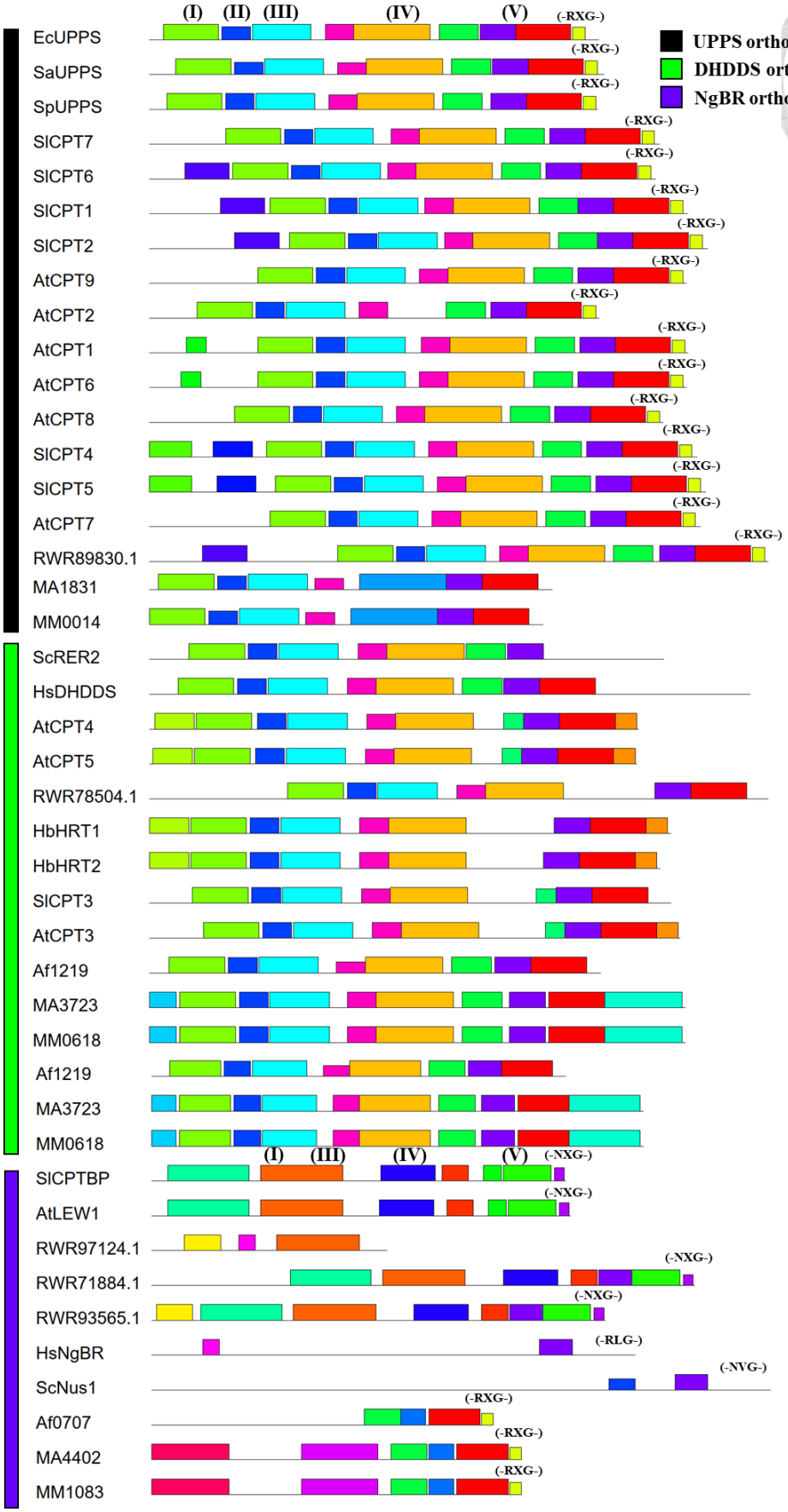


doi:10.6342/NTU202502385

Figure 6. The subunit interfaces of heterodimeric RWR78/RWR71 and RWR78/RWR93 in comparison with that of heterotetrameric human DHDDS/NgBR.

(a) Key residues located at the interface of RWR78 and RWR71. (b) Key residues located at the interface of RWR78 and RWR93. Residues from RWR78 are colored in green, while residues from RWR71 and RWR93 are colored in cyan.

doi:10.6342/NTU202502385



doi:10.6342/NTU202502385

Figure 7. Motif analysis of cis-PTs and homologues.

Among UPPS orthologs, DHDDS orthologs, and NgBR orthologs, thirty motifs expressed in various colored boxes in five conserved regions (I to V) were identified based on the conserved amino acid seuqneces, analyzed by using the MEME online tool (https://meme-suite.org/meme/tools/meme, Version 5.5.2). RXG motif was found at the C-termini of prokaryotic cis-PTs, plant UPPS-like cis-PTs, archaeal NgBR orthologs and *Homo sapien* NgBR (HsNgBR), but not in DHDDS orthologs. On the other hand, NXG motif was found in plant NgBR orthologs. The abbreviations for the species followed that in Figure S2A.

doi:10.6342/NTU202502385

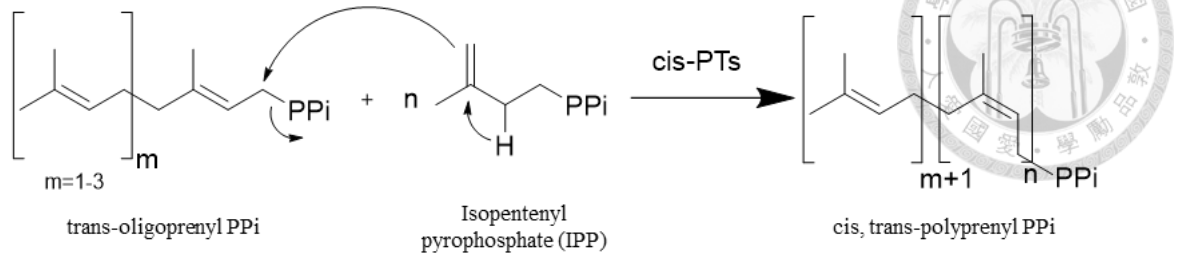
SUPPLEMENTARY INFORMATION

Table S1. Synthetic oligonucleotides as PCR primers for constructing the expression plasmids.

Primer name	Sequence (5'-3')
pESC-RWR78-F	TTTTGAAAATTCGAATTCATGCAAAAGAAG
pESC-RWR78-R	CGATACTAGTGCGGCCGCTCATAA
His_6-RWR78-F	CATCATGGTTCTGGTATGCAAAAGAAGTCCTG GAAC
His_6-RWR78-R	ATGATGATGATGAGACATGAATTCGAATTTTC AAAAATTCTT
Strep-RWR93-F	CCACAATTTGAAAAAGGTGCTGGTATGGGTG CCAATAGAAATGGT
Strep-RWR93/RWR71-R	ATGAGACCAAGAAGCAGACATGGATCCGGGG TTTTTTCTCCT
Strep-RWR71-F	CCACAATTTGAAAAAGGTGCTGGTATGAGAG GTGAAGAATGGGTAAC
sRWR93-F	GGTTTGTTGGATGAATACAAAGAT
sRWR71-F	GGTTTATTGGATAAGTACAAAGATTTGC
sRWR93/sRWR71-R	ACCAGCACCTTTTTCAAATTG
pYES2-RWR89-F	ATAGGGAATATTAAGCTTAACACAATGCCGTG CTCTTCTTCTCTG
pYES2-RWR89-R	TCTAGATGCATGCTCGAGTTAAACGCGGCCAC CGAA
pESC-RWR89-F	TTTTGAAAATTCGAATTCATGCCGTGCTCTTC TTCTCTGCAACTGCTG
pESC-RWR89-R	TACTAGTGCGGCCGCTTAAACGCGGCCACCG AA
His_6-RWR89-F	GGTTCTGGTATGCCGTGCTCTTCTTCT
His_6-RWR89-R	ATGATGATGATGATGATGAGACATGAATTCGA ATTTTCAAAAATTCTT

doi:10.6342/NTU202502385

(a)

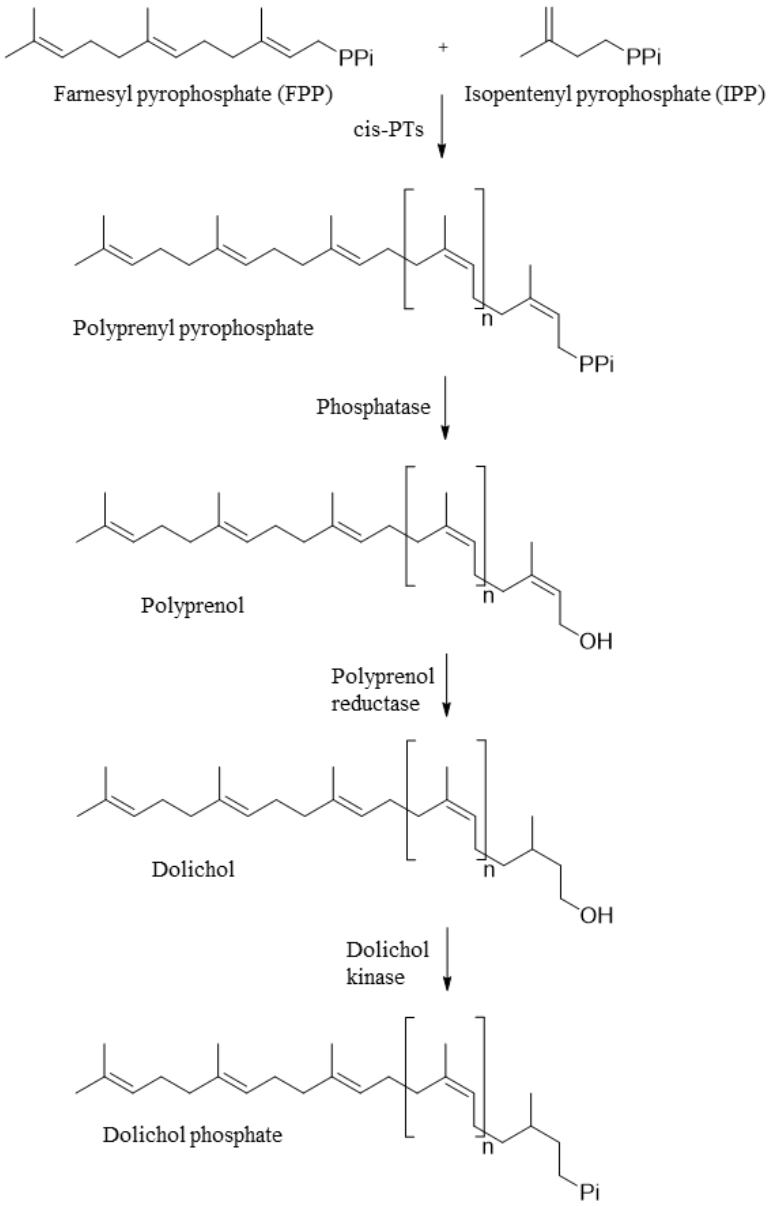


trans-oligoprenyl PPi

Isopentenyl pyrophosphate (IPP)

cis, trans-polyprenyl PPi

(b)



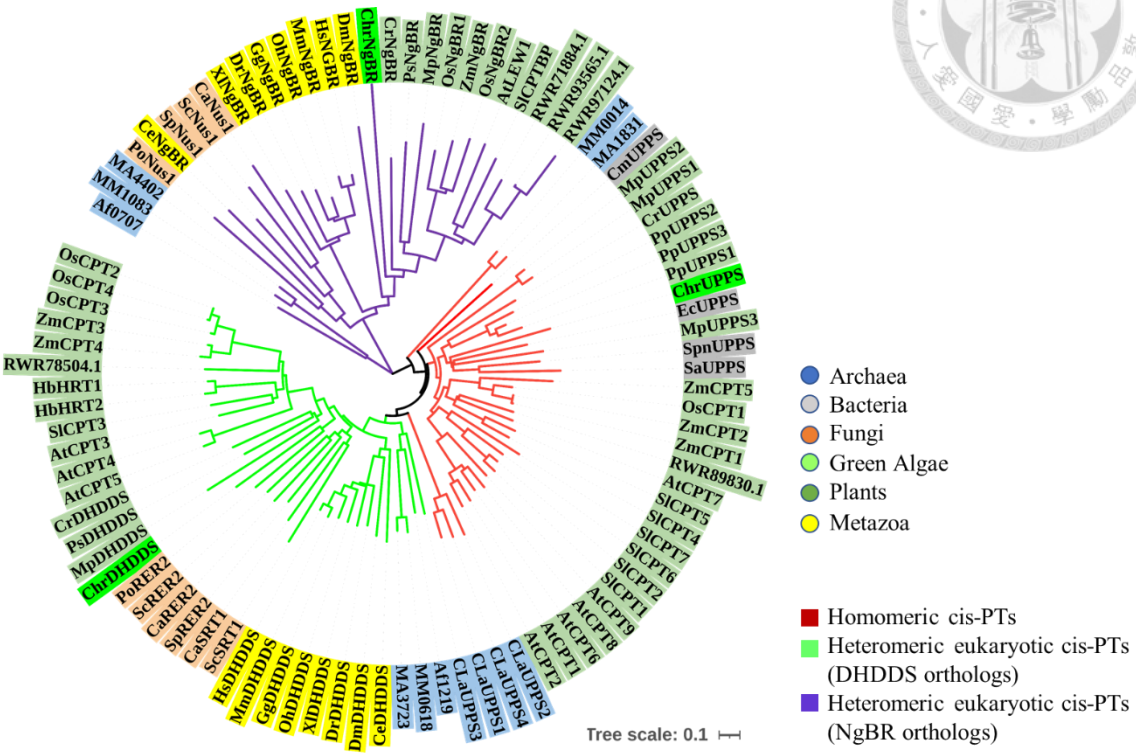
doi:10.6342/NTU202502385

Figure S1. The general reactions catalyzed by cis-prenyltransferases and the reactions leading to dolichol phosphate.

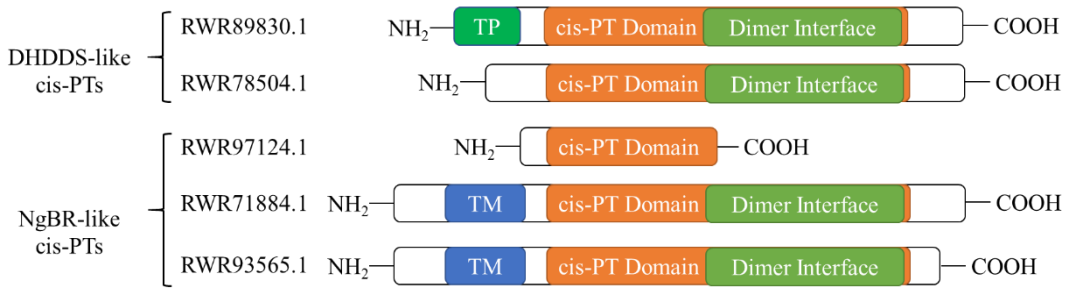
(a) Multiple IPP condensation reactions with a trans-oligoprenyl diphsophate form cis, trans-polprenyl diphsopahte with cis-double bonds catalyzed by cis-PTs. (b) The biosynthesis pathway of dolichol phosphate, the lipidic carrier of glycosyl moieties utilized for protein glycosylation in ER.

doi:10.6342/NTU202502385

(a)



(b)



doi:10.6342/NTU202502385

Figure S2. Phylogenetic analysis and domain architectures of five identified *Cinnamomum kanehirae* cis-PT homologues.

(a) Neighbor-joining phylogenetic tree of cis-PT homologues from archaea, bacteria, fungi, green algae, plants and animals. The cis-PT homologues are divided into three groups, the homomeric cis-PTs (red), the eukaryotic DHDDS orthologs (green) that are involved in dolichol biosynthesis, and NgBR orthologs (purple) that lack several conserved cis-PTs catalytic residues and form complexes with DHDDS orthologs. From the analysis, RWR89 is a plant-specific cis-PT, while RWR78 is a DHDDS ortholog, and the others are NgBR orthologs. Species abbreviations: Af, *Archaeoglobus fulgidus*; At, *Arabidopsis thaliana*; Ca, *Candida albicans*; Ce, *Caenorhabditis elegans*; CLa, *Candidatus Lokiarchaeota archaeon*; Chr, *Chlamydomonas reinhardtii*; Cr, *Ceratopteris richardii*; Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Ec, *Escherichia coli*; Gg, *Gallus gallus*; Hb, *Hevea brasiliensis*; Ma, *Methanosarcina acetivorans*; MM, *Methanosarcina mazei*; Mm, *Mus musculus*; Mp, *Marchantia polymorpha subsp. Ruderalis*; Oh, *Ophiophagus hannah*; Os, *Oryza sativa subsp. japonica*; Po, *Pleurotus ostreatus*; Pp, *Physcomitrella patens*; Ps, *Picea sitchensis*; Sa, *Staphylococcus aureus*; Sc, *Saccharomyces cerevisiae*; Sl, *Solanum lycopersicum*; Sp, *Schizosaccharomyces pombe*; Spn, *Streptococcus pneumoniae*; Xl, *Xenopus laevis*; Zm, *Zea mays*. (b) Compositions of five *Cinnamomum kanehirae* cis-PT homologues with conserved core domains and dimer

doi:10.6342/NTU202502385

interfaces highlighted in orange and green, respectively, except RWR97 that lacks dimer interface. Bacterial UPPS-like RWR89 cotains a putative chloroplast transit peptide (TP) and NgBR-like RWR71 and RWR93 have a transmembrane domain (TM).

doi:10.6342/NTU202502385

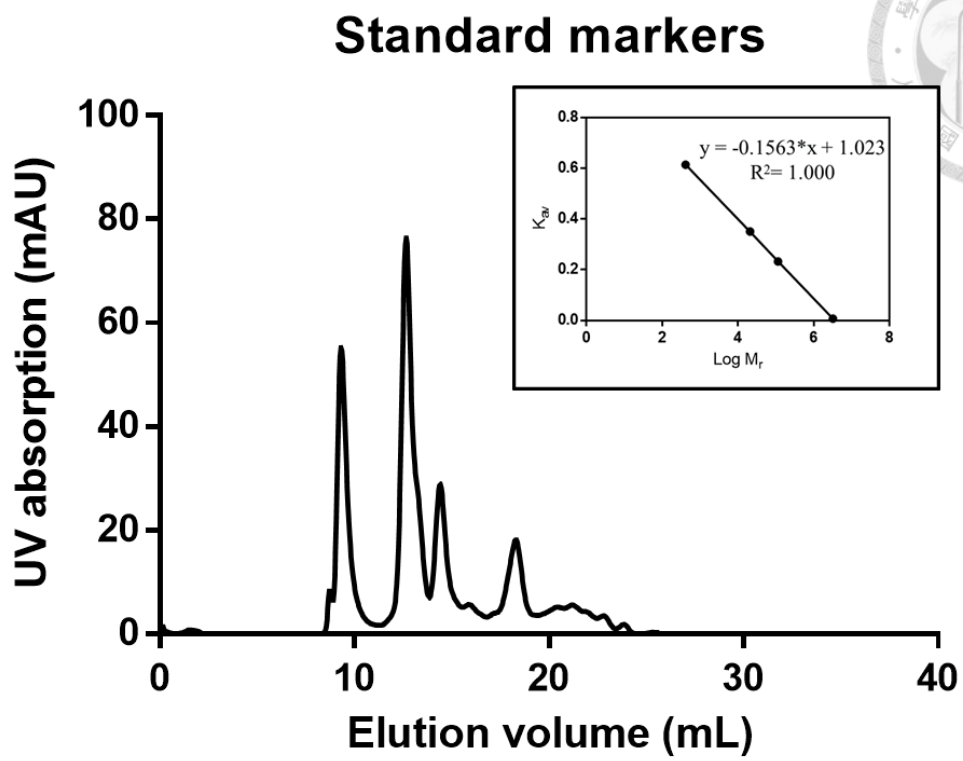
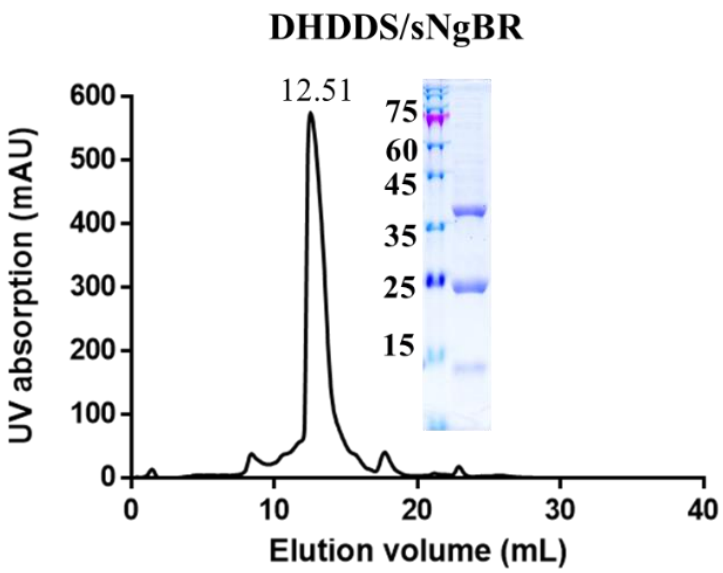


Figure S3. Size-exclusion chromatography of the molecular weight (MW) standards.

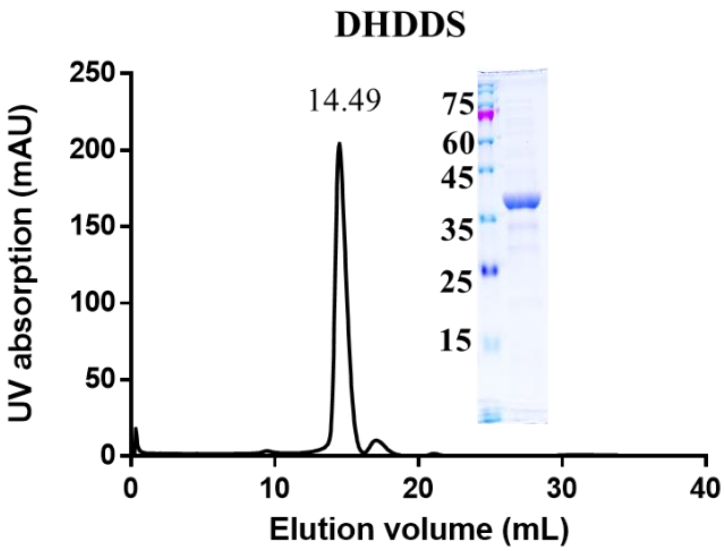
The quaternary compositions of *C. kanehirae* cis-PT homologues and complexes were determined from their elution volumes from a size-exclusion chromatogarphy using the linear calibration curve that correlated the distribution coefficient (K_{av}) calculated from the elution volumes (K_{av} = Ve-Vo/Vt-Vo, where Ve = the elution volume of target protein, Vo = void volume of the column, and Vt = total volume of the column) and the logarithm of molecular weights (Log M_r). The standard markers include Thyroglobulin (669 kDa), Aldolase (158 kDa), Conalbumin (75 kDa) and Ribonuclease A (13.7 kDa).

doi:10.6342/NTU202502385

(a)



(b)



doi:10.6342/NTU202502385

Figure S4. Purification of the *E. coli* co-expressed DHDDS/sNgBR and singly expressed DHDDS and determination of their quaternary compositions.

(a) DHDDS and TRX-Strep-TEVp-sNgBR were co-purified using StrepTrap XT column, followed by TEV protease treatment overnight, and the mixure was loaded onto StrepTrap XT and Ni-NTA column to remove the tag and TEV protease. After the size-exclusion chromatography (the elution volume was 12.51 mL corresponding to 176.1 kDa), the heterotetrameric DHDDS/sNgBR complex contained 1:1 DHDDS and sNgBR as revealed by SDS-PAGE analysis. (b) TRX-His_6-TEVp-DHDDS was purified using the same protocol as the purification of RWR89. After the size-exclusion chromatography (the elution volume was 14.49 mL corresponding to 73.8 kDa), the homodimer contained DHDDS as revealed by SDS-PAGE analysis.

doi:10.6342/NTU202502385

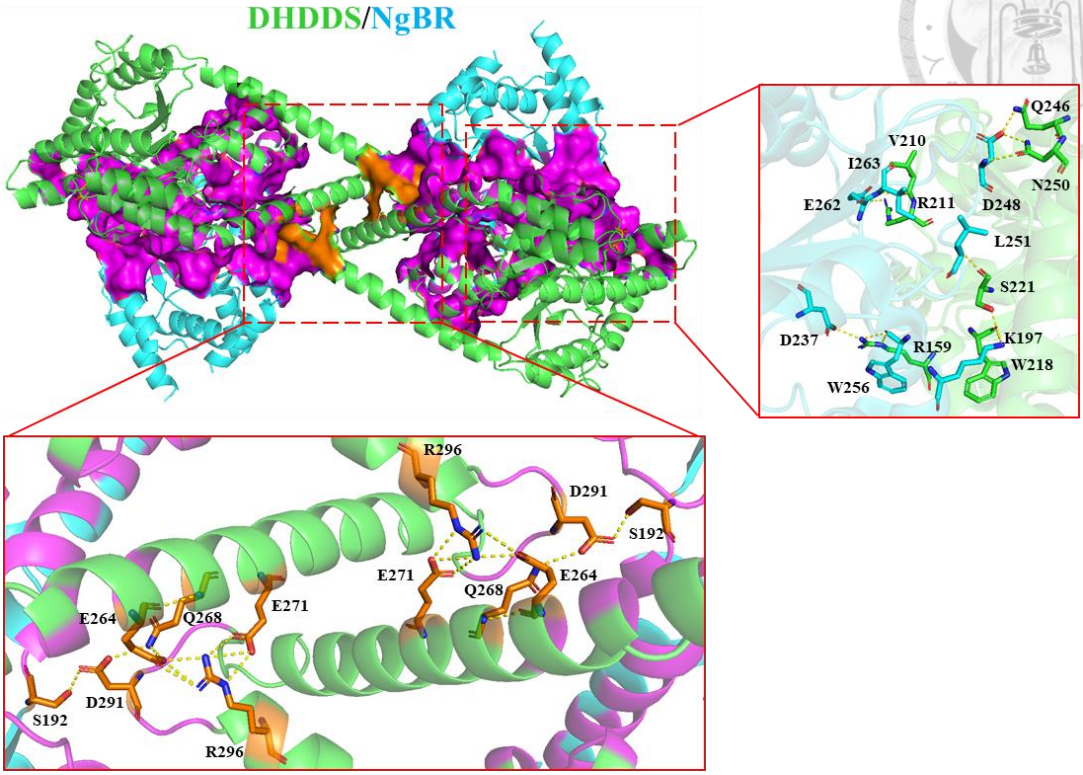
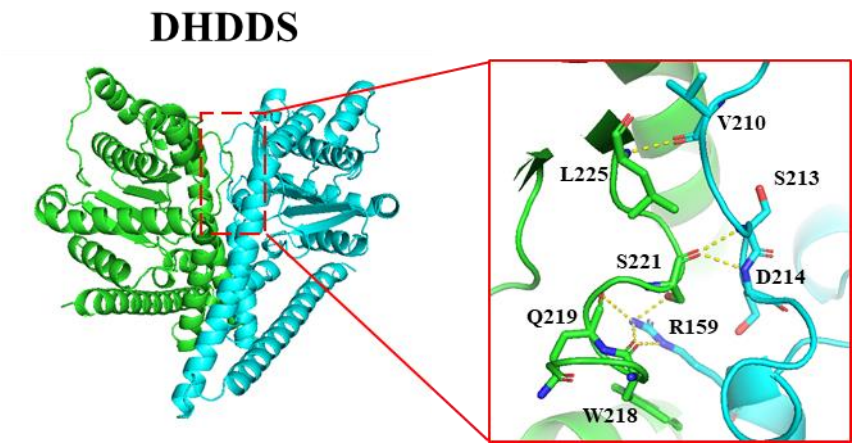


Figure S5. For comparison, key residues in the interfaces of heterotetrameric DHDDS/NgBR complex (PDB: 6Z1N) are shown.

The interface between DHDDS and NgBR is colored in magenta, while the interface between the C-terminal motifs of two DHDDS is colored in orange.

doi:10.6342/NTU202502385

(a)



(b)

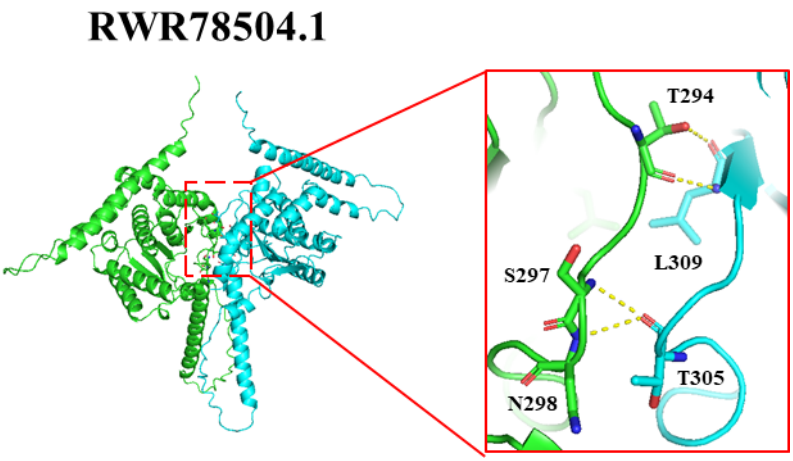


Figure S6. Homology modeling of homodimeric DHDDS and RWR78.

The homodimeric structural models of DHDDS (a) and RWR78 (b) were generated by Alphafold2 algorithm.

doi:10.6342/NTU202502385

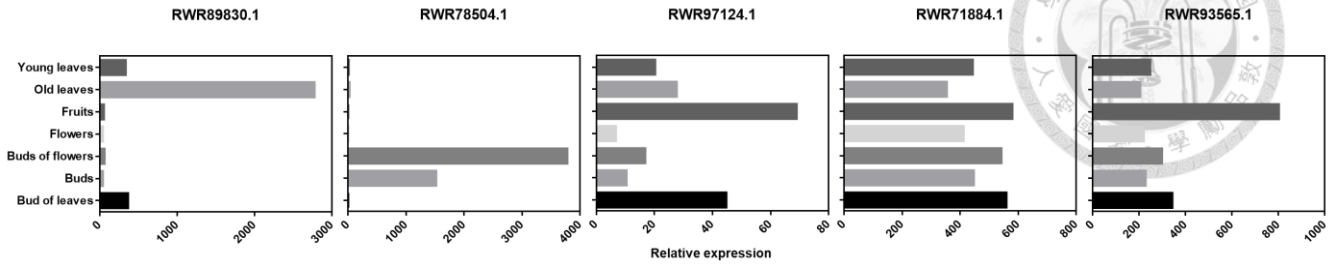


Figure S7. RNA expression levels of five *C. kanehirae* cis-PTs in various tissues, including buds, buds of flowers, flowers, buds of leaves, young leaves, old leaves, and fruits.

doi:10.6342/NTU202502385

REFERENCES

1. Sacchettini JC, Poulter CD. Creating isoprenoid diversity. *Science*. 1997; **277**: 1788–9. doi: 10.1126/science.277.5333.1788.
2. Ogura K, Koyama T. Enzymatic aspects of isoprenoid chain elongation. *Chem Rev*. 1998; **98**: 1263–76. doi: 10.1021/cr9600464.
3. Liang PH, Ko TP, Wang AH. Structure, mechanism and function of prenyltransferases. *Eur J Biochem*. 2002; **269**: 3339–54. doi: 10.1046/j.1432-1033.2002.03014.x.
4. Liang PH. Reaction kinetics, catalytic mechanisms, conformational changes, and inhibitor design for prenyltransferases. *Biochemistry*. 2009; **48**: 6562–70. doi: 10.1021/bi900371p.
5. Takahashi S, Koyama T. Structure and function of cis-prenyl chain elongating enzymes. *Chem Rec*. 2006; **6**: 194–205. doi: 10.1002/tcr.20083.
6. Teng KH, Liang PH. Structures, mechanisms and inhibitors of undecaprenyl diphosphate synthase: a cis-prenyltransferase for bacterial peptidoglycan biosynthesis. *Bioorg Chem*. 2012; **43**: 51–7. doi: 10.1016/j.bioorg.2011.09.004.
7. Cantagrel V, Lefeber DJ, Ng BG, Guan Z, Silhavy JL, Bielas SL, et al. SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder. *Cell*. 2010; 142: 203–17. doi: 10.1016/j.cell.2010.06.001.

doi:10.6342/NTU202502385

8. Fujihashi M, Zhang YW, Higuchi Y, Li XY, Koyama T, Miki K. Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. *Proc Natl Acad Sci U S A*. 2001; **98**: 4337–42. doi: 10.1073/pnas.071514398.
9. Ko TP, Chen YK, Robinson H, Tsai PC, Gao YG, Chen AP, et al. Mechanism of product chain length determination and the role of a flexible loop in *Escherichia coli* undecaprenyl-pyrophosphate synthase catalysis. *J Biol Chem*. 2001; **276**: 47474–82. doi: 10.1074/jbc.M106747200.
10. Chen YH, Chen AP, Chen CT, Wang AH, Liang PH. Probing the conformational change of Escherichia coli undecaprenyl pyrophosphate synthase during catalysis using an inhibitor and tryptophan mutants. *J Biol Chem*. 2002; 277: 7369–76. doi: 10.1074/jbc.M110014200.
11. Chang SY, Ko TP, Liang PH, Wang AH. Catalytic mechanism revealed by the crystal structure of undecaprenyl pyrophosphate synthase in complex with sulfate, magnesium, and triton. *J Biol Chem*. 2003; **278**: 29298–307. doi: 10.1074/jbc.M302687200.
12. Guo RT, Ko TP, Chen AP, Kuo CJ, Wang AH, Liang PH. Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved residues in catalysis. *J Biol Chem*. 2005; **280**: 20762–74. doi:

doi:10.6342/NTU202502385

10.1074/jbc.M502121200.

13. Harrison KD, Park EJ, Gao N, Kuo A, Rush JS, Waechter CJ, et al. Nogo-B receptor is necessary for cellular dolichol biosynthesis and protein N-glycosylation. *EMBO J*. 2011; **30**: 2490–500. doi: 10.1038/emboj.2011.147.

14. Grabińska KA, Edani BH, Park EJ, Kraehling JR, Sessa WC. A conserved C-terminal RXG motif in the NgBR subunit of cis-prenyltransferase is critical for prenyltransferase activity. *J Biol Chem*. 2017; **292**: 17351–61. doi: 10.1074/jbc.M117.806034.

15. Edani BH, Grabińska KA, Zhang R, Park EJ, Siciliano B, Surmacz L, et al. Structural elucidation of the cis-prenyltransferase NgBR/DHDDS complex reveals insights in regulation of protein glycosylation. *Proc Natl Acad Sci U S A*. 2020; **117**: 20794–802. doi: 10.1073/pnas.2008381117.

16. Bar-El ML, Vaňková P, Yeheskel A, Simhaev L, Engel H, Man P, et al. Structural basis of heterotetrameric assembly and disease mutations in the human cis-prenyltransferase complex. *Nat Commun*. 2020; **11**: 5273. doi: 10.1038/s41467-020-18970-z.

17. Giladi M, Lisnyansky Bar-El M, Vaňková P, Ferofontov A, Melvin E, et al. Structural basis for long-chain isoprenoid synthesis by cis-prenyltransferases. *Sci Adv*. 2022; **8**: eabn1171. doi: 10.1126/sciadv.abn1171.

doi:10.6342/NTU202502385

18. Ma J, Ko TP, Yu X, Zhang L, Ma L, Zhai C, et al. Structural insights to heterodimeric cis-prenyltransferases through yeast dehydrodolichyl diphosphate synthase subunit Nus1. *Biochem Biophys Res Commun*. 2019; **515**: 621–26. doi: 10.1016/j.bbrc.2019.05.135.

19. Sato M, Fujisaki S, Sato K, Nishimura Y, Nakano A. Yeast Saccharomyces cerevisiae has two cis-prenyltransferases with different properties and localizations. Implication for their distinct physiological roles in dolichol synthesis. *Genes Cells*. 2001; **6**: 495-506. doi: 10.1046/j.1365-2443.2001.00438.x.

20. Helenius A, Aebi M. Intracellular functions of N-linked glycans. *Science*. 2001; **291**: 2364–9. doi: 10.1126/science.291.5512.2364.

21. Ng BG, Freeze HH. Perspectives on glycosylation and its congenital disorders. *Trends Genet*. 2018; **34**: 466–76. doi: 10.1016/j.tig.2018.03.002.

22. Chaw SM, Liu YC, Wu YW, Wang HY, Lin CI, Wu CS, et al. Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. *Nat Plants*. 2019; **5**: 63–73. doi: 10.1038/s41477-018-0337-0.

23. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015; **12**: 357-60. doi: 10.1038/nmeth.3317

doi:10.6342/NTU202502385

24. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014; **30**: 923-30. doi: 10.1093/bioinformatics/btt656

25. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; **26**: 139-40. doi: 10.1093/bioinformatics/btp616

26. Edri I, Goldenberg M, Lisnyansky M, Strulovich R, Newman H, Loewenstein A, et al. Overexpression and purification of human cis-prenyltransferase in *Escherichia coli*. *J Vis Exp*. 2017; **126**: 56430. doi: 10.3791/56430.

27. Giladi M, Edri I, Goldenberg M, Newman H, Strulovich R, Khananshvili D, et al. Purification and characterization of human dehydrodolychil diphosphate synthase (DHDDS) overexpressed in *E. coli*. *Protein Expr Purif*. 2017; **132**: 138–42. doi: 10.1016/j.pep.2017.02.001.

28. Lisnyansky Bar-El M, Lee SY, Ki AY, Kapelushnik N, Loewenstein A, Chung KY, et al. Structural characterization of full-length human dehydrodolichyl diphosphate synthase using an integrative computational and experimental approach. *Biomolecules*. 2019; **9**: 660. doi: 10.3390/biom9110660.

29. Holcomb J, Doughan M, Spellmon N, Lewis B, Perry E, Zhang Y, et al. SAXS analysis of a soluble cytosolic NgBR construct including extracellular and

doi:10.6342/NTU202502385

transmembrane domains. *PLoS One*. 2018; 1**3**: e0191371. doi: 10.1371/journal.pone.0191371.

30. Sabry S, Vuillaumier-Barrot S, Mintet E, Fasseu M, Valayannopoulos V, Héron D, et al. A case of fatal Type I congenital disorders of glycosylation (CDG I) associated with low dehydrodolichol diphosphate synthase (DHDDS) activity. *Orphanet journal of rare diseases*. 2016; **11**: 1-14. doi: 10.1186/s13023-016-0468-1
31. Yao JY, Teng KH, Liu MC, Wang CS, Liang PH. Characterization of a *Cis*-Prenyltransferase from *Lilium longiflorum* Anther. *Molecules*. 2019; **24**: 2728 doi: 10.3390/molecules24152728.
32. Kang JH, Gonzales-Vigil E, Matsuba Y, Pichersky E, Barry CS. Determination of residues responsible for substrate and product specificity of Solanum habrochaites short-chain cis-prenyltransferases. *Plant Physiol*. 2014; **164**: 80–91. doi: 10.1104/pp.113.230466.
33. Park EJ, Grabińska KA, Guan Z, Stránecký V, Hartmannová H, Hodaňová K, et al. Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation. *Cell Metab*. 2014; **20**: 448–57. doi: 10.1016/j.cmet.2014.06.016.
34. Hamdan FF, Myers CT, Cossette P, Lemay P, Spiegelman D, Laporte AD, et al. High rate of recurrent de novo mutations in developmental and epileptic

doi:10.6342/NTU202502385

encephalopathies. *Am J Hum Genet*. 2017; **101**: 664–85. doi: 10.1016/j.ajhg.2017.09.008.

35. Züchner S, Dallman J, Wen R, Beecham G, Naj A, Farooq A, et al. Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa. *Am J Hum Genet*. 2011; **88**: 201–6. doi: 10.1016/j.ajhg.2011.01.001.
36. Zelinger L, Banin E, Obolensky A, Mizrahi-Meissonnier L, Beryozkin A, Bandah-Rozenfeld D, et al. A missense mutation in DHDDS, encoding dehydrodolichyl diphosphate synthase, is associated with autosomal-recessive retinitis pigmentosa in Ashkenazi Jews. *Am J Hum Genet*. 2011; **88**: 207–15. doi: 10.1016/j.ajhg.2011.01.002.
37. Akhtar TA, Matsuba Y, Schauvinhold I, Yu G, Lees HA, Klein SE, et al. The tomato cis-prenyltransferase gene family. *Plant J*. 2013; **73**: 640–52. doi: 10.1111/tpj.12063.
38. Ogawa T, Emi K, Koga K, Yoshimura T, Hemmi H. A cis-prenyltransferase from Methanosarcina acetivorans catalyzes both head-to-tail and nonhead-to-tail prenyl condensation. *FEBS J*. 2016; **283**: 2369–83. doi: 10.1111/febs.13749.
39. Okada M, Unno H, Emi KI, Matsumoto M, Hemmi H. A versatile cis-prenyltransferase from Methanosarcina mazei catalyzes both C- and O-prenylations. *J Biol Chem*. 2021; **296**: 100679. doi: 10.1016/j.jbc.2021.100679.

doi:10.6342/NTU202502385

40. Sompiyachoke K, Nagasaka A, Ito T, Hemmi H. Identification and biochemical characterization of a heteromeric cis-prenyltransferase from the thermophilic archaeon Archaeoglobus fulgidus. *J Biochem*. 2022; **171**: 641–51. doi: 10.1093/jb/mvac022.

41. Surmacz L, Plochocka D, Kania M, Danikiewicz W, Swiezewska E. cis-Prenyltransferase atCPT6 produces a family of very short-chain polyisoprenoids in planta. *Biochim Biophys Acta*. 2014; **1841**: 240–50. doi: 10.1016/j.bbalip.2013.11.011.

42. Qu Y, Chakrabarty R, Tran HT, Kwon EJ, Kwon M, Nguyen TD, et al. A lettuce (Lactuca sativa) homolog of human Nogo-B receptor interacts with cis-prenyltransferase and is necessary for natural rubber biosynthesis. *J Biol Chem*. 2015; **290**: 1898–914. doi: 10.1074/jbc.M114.616920.

43. Epping J, van Deenen N, Niephaus E, Stolze A, Fricke J, Huber C, et al. A rubber transferase activator is necessary for natural rubber biosynthesis in dandelion. *Nat Plants*. 2015; **1**: 1–9. doi: 10.1038/nplants.2015.48.

44. Brasher MI, Surmacz L, Leong B, Pitcher J, Swiezewska E, Pichersky E, et al. A two-component enzyme complex is required for dolichol biosynthesis in tomato. *Plant J*. 2015; **82**: 903–14. doi: 10.1111/tpj.12859.

doi:10.6342/NTU202502385

45. Emi KI, Sompiyachoke K, Okada M, Hemmi H. A heteromeric cis-prenyltransferase is responsible for the biosynthesis of glycosyl carrier lipids in Methanosarcina mazei. *Biochem Biophys Res Commun*. 2019; **520**: 291–6. doi: 10.1016/j.bbrc.2019.09.143.

46. Surowiecki P, Onysk A, Manko K, Swiezewska E, Surmacz L. Long-chain polyisoprenoids are synthesized by AtCPT1 in *Arabidopsis thaliana*. *Molecules*. 2019; **24**: 2789. doi: 10.3390/molecules24152789.

47. Wang W, Dong C, McNeil M, Kaur D, Mahapatra S, Crick DC, et al. The structural basis of chain length control in Rv1086. *J Mol Biol*. 2008; **381**: 129–40. doi: 10.1016/j.jmb.2008.05.060.

48. Chan YT, Ko TP, Yao SH, Chen YW, Lee CC, Wang AH. Crystal structure and potential head-to-middle condensation function of a Z,Z-farnesyl diphosphate synthase. *ACS Omega*. 2017; **2**: 930–6. doi: 10.1021/acsomega.6b00562.

49. Strasser R. Plant protein glycosylation. *Glycobiology*. 2016; **26**:926-939. doi: 10.1093/glycob/cww023.

50. De Coninck T, Gistelinck K, Janse van Rensburg HC, Van den Ende W, Van Damme EJM. Sweet modifications modulate plant development. *Biomolecules*. 2021; **11**: 756. doi: 10.3390/biom11050756.

doi:10.6342/NTU202502385

Chapter II

Identification of a structural mechanism for substrates discrimination and prediction of multifunctional GH5 cellulases/hemicellulases

中文摘要

具有降解纖維素與半纖維素活性的多功能且耐高溫之酵素，對於將植物性生物質轉化為生物燃料的過程具有重要應用價值。過去的研究中，我們已鑑定來自 *Clostridium thermocellum* 的雙功能纖維素酶/木聚糖酶 Cel5E（*Ct*Cel5E）以及來自 *Thermotoga maritima* 的雙功能纖維素酶/甘露聚糖酶 Cel5A（*Tm*Cel5A）。儘管這兩種酶在胺基酸序列與三維結構上高度保守，卻展現出明顯不同的半纖維素受質特異性。在本研究中，我們進一步分析另一種 GH5 家族的酶 *Ct*Cel5T，其被註釋為三功能纖維素酶/木聚糖酶/甘露聚糖酶。透過結構比較並結合點突變實驗，我們鑑定出 *Ct*Cel5T 的 Met277 及 Glu360 分別佔據與 *Tm*Cel5A 的 His205 和 Trp210 相似的空間位置，並對辨識半纖維素受質發揮關鍵作用。因此，透過親緣關係樹與多重胺基酸序列比對，可以鑑定具結構功能一致性的纖維素酶/半纖維素酶或預測未知酵素之功能。本研究深化了對多功能纖維素酶/半纖維素酶受質特異性機制的理解，並為其在工業應用中的蛋白質工程設計提供了理論支持。

關鍵詞: 生物燃料；纖維素酶；半纖維素酶；受質特異性；點突變

doi:10.6342/NTU202502385

ABSTRACT

Multifunctional and heat-resistant enzymes with both cellulose and hemicellulose degrading activities are particularly useful to convert plant biomass into biofuel. In the past, we have characterized the bifunctional cellulase/xylanase Cel5E from *Clostridium thermocellum* (*Ct*Cel5E) and the bifunctional cellulase/mannanase Cel5A from *Thermotoga maritima* (*Tm*Cel5A). Despite highly conserved amino acid sequences and structures, they exhibit distinct hemicellulase substrate specificities. In the current study, we further characterized another GH5 enzyme, *Ct*Cel5T, annotated as a trifunctional cellulase/xylanase/mannanase. Through structural comparison coupled with site-directed mutagenesis, we identified Glu360 in loop 8 and Met277 in loop 6 of *Ct*Cel5T occupying the overlapped spatial positions with that of Trp210 and His205 in loop 6 of *Tm*Cel5A, respectively, play critical roles in substrate recognition. Therefore, phylogenetic tree analysis and multiple sequence alignment are useful to find characterized cellulases/hemicellulases with functions consistent with the structural features or uncharacterized enzymes for predicting their functions. This study thus enhances our understanding on the mechanisms for substrate specificities and engineering of multifunctional cellulases/hemicellulases for industrial applications.

KEYWORDS: biofuel; cellulase; hemicellulase; substrate specificity; site-directed mutagenesis

doi:10.6342/NTU202502385

ABBREVIATIONS

BWX, beechwood xylan

CMC, carboxymethyl cellulose

*Ct*Cel5E, *Clostridium thermocellum* Cel5E

*Ct*Cel5T, *Hungateiclostridium thermocellum* Cel5T

DNS, 3,5-dinitrosalicylic acid

GH, glycoside hydrolase

HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

IPTG, isopropyl β-thiogalactopyranoside

LB, Luria-Bertani

Ni-NTA, nickel nitrilotriacetic acid

PCR, polymerase chain reaction

PDB, Protein Data Bank

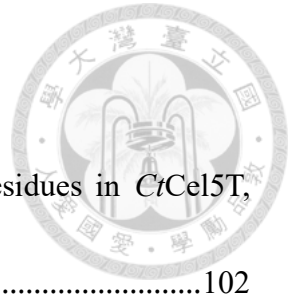
*Tm*Cel5A, *Thermotoga maritima* Cel5A

Tris, tris(hydroxymethyl)aminomethane

doi:10.6342/NTU202502385

LIST OF TABLES

doi:10.6342/NTU202502385



LIST OF FIGURES

doi:10.6342/NTU202502385

1. INTRODUCTION

1.1 Sustainable energy production from agriculture biomass

Fossil has been utilized as the primary energy source in our contemporary society. According to the International Energy Agency (https://www.iea.org/), the global CO_2 emissions from fuel combustion increased by 1.3% in 2022, and the fossil fuels remained the dominant contributors to the global total energy supply, accounting for 81% of the total, with oil constituted nearly 30%, followed by coal at 28% and natural gas at 23%. Besides, global emissions from fuel combustion were primarily driven by coal, which accounted for 45%, with oil and natural gas contributing 33% and 22%, respectively. However, fossil fuel reserves are neither infinite nor environmentally sustainable. Consequently, the exploration of alternative and sustainable energy resources is crucial, not only to address the limitation of finite fossil fuel supplies but also to promote environmentally responsible substitutes. Converting agriculture waste into the renewable energy production is one of the feasible processes. Agriculture biomass represents a versatile energy resource, with the potential to be transformed into combustible fuels, biogas, bioethanol, and electrical energy [1]. Compared to the direct combustion of agricultural waste for thermal energy, manufacturing bioethanol or biodiesel from biomass through enzymatic hydrolysis and microbial fermentation presents a more environmentally sustainable approach by significantly reducing greenhouse gas

doi:10.6342/NTU202502385

emissions and mitigating global warming. Therefore, the effective utilization of agricultural biomass not only provides an alternative energy source but also minimizes the negative environmental impact on our planet.

1.2 Composition of agriculture biomass

The primary agricultural wastes globally are wheat and rice straw, with agriculture biomass predominantly consisting of lignocellulosic materials, including polysaccharides (30-40% celluloses and 30-50% hemicelluloses) and 8-21% lignin [2, 3, 4]. Celluloses are glucose polymers linked via β-1,4-glycosidic bonds, whereas hemicelluloses consist of other monosaccharides such as arabinose, xylose, galactose, mannose etc. For conversion of lignocellulosic feedstock into biofuel, polysaccharides need to be hydrolyzed to monosaccharides prior to fermentation to bioethanol by yeast or engineered microbes. However, the structural diversity of polysaccharides in plant cell wall is a major challenge for cellulosic biofuel production due to rate-limiting and high-cost enzymatic hydrolysis to release fermentable monosaccharides [5, 6]. For example, the polysaccharides such as homo-1,4-β-D-glucans, hetero-1,4-β-D-xylanes, and hetero-1,4-β-D-mannans need to be hydrolyzed by the endo/exo cellulases, xylanases, and mannanases into cellobiose, xylobiose, and mannobiose, respectively [7]. Subsequently, β-glucosidases, β-xylosidases, and β-mannosidases convert the disaccharides into monosaccharides glucose, xylose, and mannose, respectively. We demonstrated that these

doi:10.6342/NTU202502385

enzymes could be combined to display synergistic activities on biomass degradation in vitro or in yeast system [8, 9].

1.3 *Clostridium thermocellum* cellulosome for plant biomass degradation

In nature, *Clostridium thermocellum*, a thermophilic, anaerobic, chemo-organotrophic and efficient plant cell wall-degrading bacterium, has a comprehensive enzymatic system "cellulosome" on the cell surface, containing cellulases, xylanases, mannanases, chitosanases, and so on to synergistically degrade plant cell wall heterogeneous polysaccharides [10, 11, 12]. A cellulosome is built of a scaffoldin backbone that contains several cohesion domains and each cohesion binds with a dockerin domain-carrying enzyme [13]. Most of these enzymes are classified as the glycoside hydrolases 5 (GH5) family [14, 15], the largest among 131 GH families from CAZy database(http://www.cazy.org) [16].

1.4 Substrate specificities of *Ct*Cel5E, *Ct*Cel5T and *Tm*Cel5A

We have biochemically and structurally characterized a GH5 bifunctional cellulase/xylanase within the cellulosomal *celH* gene of *Clostridium thermocellum*, named *Ct*Cel5E [17]. The rest of *celH* gene encodes Lic26A, a GH26 hydrolase containing β-1,3-1,4-mixed linked endoglucanase activity, a family-11 carbohydrate binding module (CBM11), and two C-terminal type I dockerin domain [18, 19, 20]. On the other hand, Cel5A from *Thermotoga maritima* (*Tm*Cel5A), another GH5 enzyme, was

doi:10.6342/NTU202502385

characterized as a bifunctional cellulase/mannanase [21, 22, 23]. Although *Tm*Cel5A and *Ct*Cel5E share significant sequence homology (41% identity and over 55% similarity) and structural homology (RMSD of 1.4Å of 248 Cα atoms), two enzymes display different hemicellulose specificities. Bioinformatics in combination with mutagenesis analyses have shown that the conserved NEPHE motif in *Tm*Cel5A is critical for its mannanase activity [24]. We further identified a loop in *Tm*Cel5A (*Tm*loop) that represents the most different region when comparing with the sequence and structure of *Ct*Cel5E, in which particularly Tyr198, His205, and Trp210, are involved in mannan binding [25]. By incorporating this mannan-recognizing loop, we were able to engineer *Ct*Cel5E, a mannanase-free enzyme to gain mannanase activity, although compromising some xylanase activity [25].

1.5 Specific aim of this study

*Ct*Cel5E and *Tm*Cel5A share cellulase activity but differ in hemicellulase activities (xylanase vs. mannanase), although both enzymes share significant sequence similarity and structural homology. These previous results seem to indicate that both mannanase and xylanase activities could not coexist in multifunctional cellulases/hemicellulases. Therefore, in the study reported herein, we chose another GH5 enzyme from *Hungateiclostridium thermocellum* [26], named *Ct*Cel5T, which was annotated as a trifunctional cellulase/xylanase/mannanase with apo-form crystal structure available

doi:10.6342/NTU202502385

(PDB: 4IM4) [27] to perform structure-based site-directed mutagenesis for understanding the roles of key residues in discriminating different substrates as compared to *Ct*Cel5E and *Tm*Cel5A. This enzyme is encoded by the N-terminal module of another cellulosomal gene *celE*. *Ct*Cel5T was followed by a central type I dockerin module for integrating the enzyme into cellulosome, and a C-terminal esterase, *Ct*CE2, which has dual functions to enhance cellulase catalytic module activity and deacetylate plant polysaccharides. We aim to examine the functions of loops and key residues important to dictate the substrate specificity toward xylan or mannan, while sharing the common substrate cellulose. As cellulose and xylan are two of the most abundant polysaccharides which account for > 47–67% of sugar content in rice straw with hemicelluloses such as arabinan and mannan accounting for ~10% [28], understanding the mechanisms for determining the substrate specificities may pave the way to engineer multifunctional GH5s for biofuel production. Our results thus enhance understanding of how these enzymes discriminate their hemicellulose substrates and the information helps to predict/engineer the functions of multifunctional cellulases/hemicellulases.

2. MATERIALS AND METHODS

2.1 Materials

Carboxymethyl cellulose (CMC), xylose (X1), glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), cellopentaose (G5), cellohexaose (G6), Mannose

doi:10.6342/NTU202502385

(M1), Mannobiose (M2), Mannotriose (M3), 3,5-dinitrosalicylic acid (DNS) were obtained from Sigma-Aldrich (St. Louis, USA). Xylobiose (X2), xylotriose (X3), mannan and beechwood xylan (BWX) were purchased from Megazyme (Wicklow, Ireland). *Ct*Cel5E, *Ct*Cel5E-*Tm*loop/F267A/F323I, and *Tm*Cel5A were previously expressed and purified [17, 25]. For protein purification, nickel nitrilotriacetic acid (Ni-NTA) resin was obtained from EBL (Taipei, Taiwan), and Thrombin protease was retrieved from Cytiva (Marlborough, MA, USA).

2.2 Construction of expression plasmids

The gene of *Ct*Cel5T was synthesized by Bio Basic Inc. (Markham Ontario, Canada) and constructed into pHTTP13 expression vector using *EcoRI* and *XhoI* restriction sites to express the protein with an N-terminal His_6 tag [29]. Site-directed mutants of *Ct*Cel5T were generated using the QuikChange site-directed mutagenesis kit (Agilent, USA). All the synthetic oligonucleotides used for mutagenesis are listed in Table 4 and 5. Mutagenesis was performed by polymerase chain reaction (PCR) amplification of the gene using the designed primers, followed by digestion of the PCR products with DpnI for 3 h at 37°C to cleave the methylated template DNA. The resulting mutated constructs were transformed into *E. coli* DH5α and cultured in Luria-Bertani (LB) medium, and the correct clones were confirmed by DNA sequencing.

2.3 Expression and purification of recombinant proteins

doi:10.6342/NTU202502385

Each verified gene construct was transformed into *E. coli* BL21 (DE3) for protein expression. An overnight culture (20 mL) from a single colony was used to inoculate 1 L of LB broth supplemented with 50 μg/mL kanamycin. Cultures were incubated at 37°C until reaching OD_{600} = 0.6, followed by induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 20 h at 16 °C. Cells were harvested by centrifugation at 6000 rpm for 20 min, and the resulting pellets were resuspended in 50 mL lysis buffer (25 mM Tris-HCl, pH 7.5, 10 mM imidazole, and 500 mM NaCl). Cell disruption was performed using a French press (Constant Cell Disruption System) at 22,000 psi. The lysate was then centrifuged at 40,000xg for 45 min to remove debris. The supernatant was loaded onto a 10 mL Ni-NTA affinity column pre-equilibrated with lysis buffer. The column was washed with 20 column volumes of wash buffer (25 mM Tris-HCl, pH 7.5, 20 mM imidazole, and 500 mM NaCl) and the recombinant protein was eluted with 20 mL elution buffer (25 mM Tris-HCl, pH 7.5, 250 mM imidazole, and 500 mM NaCl). Fractions containing the target protein were collected, treated with thrombin for His_6-tag cleavage, and dialyzed overnight in dialysis buffer (25 mM Tris-HCl, pH 7.5, and 150 mM NaCl). Following dialysis, the protein solution was passed through a second Ni-NTA column to remove the cleaved His_6 tag, and the flow-through containing the tag-free target protein was collected and concentrated. Protein purity was assessed by sodium dodecyl

doi:10.6342/NTU202502385

sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), as shown in Figure S1, and protein concentrations were determined using the Bradford assay [30].

2.4 Enzyme activity assays

Reducing sugars produced by enzymatic hydrolysis were quantified using DNS reagent, with D-glucose, D-xylose, and D-mannose serving as standards [31]. The corresponding standard curves are shown in Figure S2. CMC, BWX, and mannan were used as substrates for measuring cellulase, xylanase, and mannanase activities, respectively. To determine the optimal pH, enzyme activities were assayed using 1 μM enzyme and 1% substrate in CGH buffer (10 mM citric acid, 10 mM glycine, and 10 mM HEPES) across a pH range of 2 to 11 at 50°C for 10 min. Optimal temperature was subsequently determined by measuring enzyme activity at temperatures ranging from 30°C to 90°C under the previously determined optimal pH using the same enzyme and substrate concentrations. Following enzymatic reactions, DNS reagent was added to each reaction mixture, which was then heated at 100°C for 5 min. Absorbance was measured at 540 nm by the microplate reader (BioTeK Synergy H1). Activity unit (IU) is determined as the amount of products released by 1 μmole of the enzyme per minute under the optimal assay conditions.

Kinetic parameters were determined by measuring the initial rates of reducing sugars formation at varying substrate concentrations using a fixed enzyme concentration.

doi:10.6342/NTU202502385

Substrate concentrations (CMC, BWX, and mannan) ranged from 2 to 30 mg/mL, and enzyme concentrations between 0.2 and 1 μM were used under the optimal conditions. Reactions were carried out under optimal pH and temperature conditions for each enzyme activity. The kinetic parameters K_m and V_{max} were calculated by fitting the data to the Michaelis–Menten equation using GraphPad Prism. The turnover number (k_{cat}) was calculated by dividing V_{max} by the concentration of the enzyme used in the assay. All experiments were performed in triplicate.

2.5 End-product analysis

End-product analysis was performed by incubating 1 μM *Ct*Cel5T with 1% CMC, BWX, or mannan in 10 mM CGH buffer to a final volume of 500 μL under optimal conditions. Reactions were examined at various time periods (2, 10, 30, and 60 minutes), and at each time point, 20 μL of the reaction mixture was withdrawn and mixed with 20 μL of 99% ethanol, followed by centrifugation at 13,000 rpm for 1 min to precipitate insoluble materials. 2 μL of supernatant in each sample was spotted onto a thin-layer chromatography (TLC) silica gel 60 plate (Merck, Germany) for analysis. Glucose, cellobiose, cellotriose, cellotetraose, cellopentaose (G1 to G5), xylose, xylobiose, xylotriose (X1 to X3), mannose, mannobiose, and mannotriose (M1 to M3) were used as standards. Separation of products was performed using TLC with a mobile phase consisting of acetic acid, 1-butanol, and distilled deionized water in a 1:2:1 (v/v/v) ratio

doi:10.6342/NTU202502385

at room temperature. Sugars were detected by spraying the TLC plate with diphenylamine-aniline-phosphoric acid reagent (1 mL of 37.5% HCl, 2 mL of aniline, 10 mL of 85% H_3PO_4, 100 mL of ethyl acetate, and 2g of diphenylamine). After spraying, the TLC plate was heated at 50 to 60 ℃ until the dye became light green, followed by additional heating at 150 ℃ for 30 sec to complete visualization.

2.6 Multiple sequence alignment and phylogenetic tree analysis

The amino acid sequences of *Ct*Cel5T, *Ct*Cel5E, and *Tm*Cel5A were aligned and analyzed using Clustal Omega website (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) [32] and Jalview program [33]. The phylogenetic tree was generated using iTOL (Interactive Tree Of Life) website (https://itol.embl.de/) [34], and analysis of the active site architecture for the GH5_4 and GH5_25 subfamilies was conducted using WebLogo website (https://weblogo.berkeley.edu/logo.cgi) [35]. The structures of *Ct*Cel5T (PDB code 4IM4) [27], *Ct*Cel5E-xylobiose (PDB ID: 4U5I) [17], *Tm*Cel5A-mannotriose (PDB ID: 3AZS) [23], and structures of other GH5 enzymes were downloaded from Protein Data Bank (https://www.rcsb.org/) and compared using the PyMOL program (https://pymol.org/).

3. RESULTS

3.1 Structural and sequence comparison of *Ct*Cel5T, *Ct*Cel5E and *Tm*Cel5A

To identify the crucial residues that might interact with substrates in the active site of

doi:10.6342/NTU202502385

*Ct*Cel5T, multiple sequence alignment of *Ct*Cel5T with *Ct*Cel5E and *Tm*Cel5A (Figure 1a) as well as superimposition of the apo-form structure of *Ct*Cel5T (PDB: 4IM4) with the xylobiose-bound *Ct*Cel5E (PDB: 4U5I) and the mannotriose-bound *Tm*Cel5A (PDB: 3AZS) were performed (Figure 1b). Based on the structural comparison, *Ct*Cel5T, like *Ct*Cel5E and *Tm*Cel5A, possesses a typical $(\beta/\alpha)_8$ TIM-barrel fold structure, which is the characteristic of GH5 enzymes [36, 37]. In general, *Ct*Cel5T exhibits high structural similarities with *Ct*Cel5E and *Tm*Cel5A (RMSD = 0.869 Å of 213 Cα atoms and 1.126 Å of 199 Cα atoms, respectively). The amino acid residues involved in catalysis and substrate binding of *Ct*Cel5T, *Ct*Cel5E, and *Tm*Cel5A are highly conserved. Glu193 and Glu316 (indicated by red triangles in Figure 1a), the predicted general acid/base and nucleophile of *Ct*Cel5T, are conserved with Glu209 and Glu314 in *Ct*Cel5E as well as Glu136 and Glu253 in *Tm*Cel5A, respectively. The residues including Asn72, Trp82, His148, His149, Asn192, Tyr270, and Trp349 in *Ct*Cel5T as indicated by black triangles in Figure 1a are strictly conserved as well. However, there are major differences among the three sequences: loop 8 in *Ct*Cel5T contains additional residues, while loop 6 in *Tm*Cel5A exhibits excessive amino acids. Within the loop 6 of *Tm*Cel5A (*Tm*loop), Tyr198, His205, and Trp210 have been shown to be critical for mannanase activity in our previous study [25]. *Ct*Cel5T contains Tyr270 and Met277, which correspond to Tyr198 and His205 in *Tm*Cel5A, but lacks a residue conserved with Trp210 in *Tm*Cel5A (Figure

doi:10.6342/NTU202502385

1a). However, the superimposed structures of *Ct*Cel5T, *Tm*Cel5A-mannotriose, and *Ct*Cel5E-xylobiose (Figure 1b) show that Glu360 and Met277 in *Ct*Cel5T occupy positions similar to that of Trp210 and His205 in *Tm*Cel5A, which may be critical for discriminating xylan and mannan substrates. As shown in Figure 1c, Trp210 and His205 of *Tm*Cel5A form H-bonds with the bound mannotriose (right panel), but there is no amino acid in *Ct*Cel5E corresponding to Glu360 and Met277 in *Ct*Cel5T to interact with the bound xylobiose (left panel). To explain the different substrate specificities of these three multifunctional enzymes, *Ct*Cel5T needed to be expressed, purified, and characterized to compare with our previously characterized *Ct*Cel5E and *Tm*Cel5A.

3.2 Expression, purification, and characterization of *Ct*Cel5T (performed by the previous master student Yung-Yeh Chen in our laboratory)

*Ct*Cel5T was expressed using *E. coli*, as we previously expressed *Ct*Cel5E and *Tm*Cel5T. *Ct*Cel5T-encoding DNA was synthesized and cloned into a vector that expressed a protein with hexa-his tag for purification using Ni-NTA. As annotated, the purified *Ct*Cel5T showed activities toward carboxyl cellulose (CMC), beechwood xylan (BWX), and mannan, the substrates for cellulase, xylanase, and mannanase, respectively. The optimal reaction temperatures for its cellulase, xylanase, and mannanase activities were determined to be 60, 70, and 50° C, respectively (Figure 2a–c). Besides, at pH 4, the enzyme displayed the highest cellulase and mannanase activities, whereas the highest

doi:10.6342/NTU202502385

xylanase activity was observed at pH 5. Under the optimal conditions, K_m and k_{cat} values of *Ct*Cel5T were measured to be 20.4 ± 2.8 mg/mL and 8733 ± 574 min^{-1} for CMC, 9.5 ± 0.7 mg/mL and 832.6 ± 31.6 min^{-1} for BWX, and 5.4 ± 1.6 mg/mL and 47.5 ± 5.7 min^{-1} for mannan, respectively (Table 1). TLC analysis also confirmed that *Ct*Cel5T produced the expected di- and tri-saccharide major products from CMC, BWX, and mannan (Figure 2d).

Compared to other GH5 family members previously studied in our laboratory, including *Tm*Cel5A (a cellulase/mannanase), *Ct*Cel5E (cellulase/xylanase) and *Ct*Cel5E-*Tm*loop/F267A/F323A (a *Ct*Cel5E mutant gaining mannanase activity by incorporating the *Tm*loop), the cellulase activity of *Ct*Cel5T was the highest (Figure 3a), while still maintaining a level of xylanase activity similar to that of *Ct*Cel5E (Figure 3b), but its mannanase activity was significantly lower than that of *Tm*Cel5A (Figure 3c).

3.3 Functional characterization of the crucial residues in CtCel5T by site-directed mutagenesis (performed by the previous master student Yung-Yeh Chen in our laboratory)

To investigate the roles of the identified key residues of *Ct*Cel5T in three enzymatic activities, we first generated twelve *Ct*Cel5T mutants, including N72A, W82A, H148A, H149A, N192A, E193A, Y270A, M277A, E316A, W349A, N351A, and E360A. The k_{cat} values of these mutants toward CMC, BWX, and mannan under the optimal conditions

doi:10.6342/NTU202502385

measured by 3,5-dinitrosalicylic acid (DNS) assay are shown in Figure 4, and their K_m and k_{cat} parameters on three different substrates (CMC, BWX, and mannan) are summarized in Table 1. As expected, the mutations of Glu193 and Glu316, which are supposed to be the general acid/base and the nucleophile for glycosidic bond cleavage, respectively, to alanine resulted in complete loss of activities on all three substrates. Y270A and W349A also caused loss of three activities, demonstrating that these two amino acids are essential for catalysis. In *Tm*Cel5A, Tyr198 forms a hydrogen-bonding interaction with (-1) mannose of the bound mannotriose (Figure 1c right panel). Therefore, the corresponding Tyr270 in *Ct*Cel5T might be responsible for binding with (-1) sugar moiety of the substrates, thus explaining the loss of three activities for Y270A. Furthermore, W349A was inactive against all three substrates, suggesting that this residue is essential for catalysis, likely due to its involvement in providing the hydrophobic interaction with the (-1) sugar ring, as shown in Figure 1b.

Next, N72A exhibited a 9-fold decreased activity, 16-fold decreased k_{cat} and 2-fold decreased K_m on CMC, and lack of both xylanase and mannanase activities. These changes may be attributed to the disruption of hydrogen bonds on (-2) sugar for ligand stabilization (see Figure 1c). Moreover, the corresponding residues of *Ct*Cel5T His148 and His149 in *Ct*Cel5E and *Tm*Cel5A exhibit hydrogen bonds with three ligands (cellobiose and xylobiose or mannotriose) (Figure 1c). However, H148A mutant of

doi:10.6342/NTU202502385

*Ct*Cel5T displayed a more significant reduction of cellulase activity in k_{cat} (15-fold), along with a 4-fold decreased K_m of CMC, in comparison with H149A, which showed a slightly increase in K_m and a similar k_{cat} value, but both mutants showed abolished xylanase and mannanase activities. Moreover, Asn192 might form H-bonds with C2-OH of the (-1) sugar as revealed by Figure 1c, so that N192A showed 15-fold decreased k_{cat}. The W82A mutation had a relatively minor impact, resulting in approximately a 3-fold decrease in k_{cat} and similar K_m values for cellulase activity, while completely abolished xylanase and mannanase activity. Trp82 might also contribute the stacking force with (-3) sugar moiety to stabilize the ligand binding (Figure 1b), but not as important as Trp349 that extremely lost all three catalytic activities upon mutation to Ala. On the other hand, for the mutation on the non-conserved residue, N351A exhibited decreased k_{cat} on BWX, the xylanase substrate, by 7-fold, but less affected cellulase and mannanase activities (Table 1). Asn351 likely forms one and two hydrogen bonds with (-2) xylose of xylobiose, respectively (Figure 1c left panel).

3.4 Pivotal roles of Met277 and Glu360 in *Ct*Cel5T on substrate discrimination

From the superimposed structures (Figure 1b), we found that Glu360 and Met277 in *Ct*Cel5T occupied the overlapped spatial positions of Trp210 and His205 in *Tm*Cel5A, respectively, but mannanase-free *Ct*Cel5E did not have amino acids at the positions, while all three enzymes shared conserved residues in other positions. We suspected that the

doi:10.6342/NTU202502385

amino acids at these two positions determine their different hemicellulase substrate specificities. M277A mutation caused mild decrease in both cellulase and xylanase activities, but led to the loss of mannanase (Table 2) activity, indicating that Met277 is solely essential for mannanase activity. In fact, the corresponding residue of Met277 is His205 in *Tm*Cel5A, which forms a hydrogen bond and stacking interaction with C6-OH of (-1) mannose (Figure 1c right panel) and was demonstrated to be indispensable for mannanase activity of *Tm*Cel5A [25]. We then mutated Met277 to His and the M277H mutant showed a significant decrease in the k_{cat} values of both cellulase and xylanase, whereas mannanase activity remained detectable.

Moreover, Glu360 of *Ct*Cel5T also occupies the position of Trp210 in *Tm*Cel5A (Figure 1b), so that we generated E360A. E360A exhibited decreased k_{cat} on BWX, the xylanase substrate, by 1.9-fold and less affected cellulase and mannanase activities. Mutating Glu360 to Trp, E360W *Ct*Cel5T completely lost the xylanase activity, but the mannanase activity was still maintained, although not enhanced (Table 2). Although E360A did not significantly affected the k_{cat} against BWX, E360W indeed abolished the xylanase activity (Table 2), suggesting that Trp at this position of the *Ct*Cel5T mutant posed a steric hindrance on binding of BWX. We then mutated Glu360 in *Ct*Cel5T to His, Gln, Lys, Tyr, or Arg to explore the effect of substituting different amino acids at this residue. Nevertheless, only E360H mutant preserved mannanase activity, better than that

doi:10.6342/NTU202502385

of E360W, but its xylanase activity was lost, while the other mutants failed to degrade mannan, and all these mutants displayed no or mere cellulase and xylanase activities (Table 2).

Furthermore, mutating both Met277 and Glu360 of *Ct*Cel5T simultaneously to the corresponding His and Trp or His in *Tm*Cel5A, M277H/E360W and M277H/E360H, indeed increased the mannanase k_{cat} by 5-fold, while reducing xylanase k_{cat} by 5-fold as compared to that of the wild-type *Ct*Cel5T (Table 2 and Figure 3). We thus demonstrated conversion of *Ct*Cel5T's xylanase activity toward mannan substrate by mutating both Met277 and Glu360, confirming that both amino acids play significant roles in mannanase activity. On the other hand, we also constructed *Tm*Cel5A mutants H205M, W210E and H205M/W210E to investigate whether the substrate specificities would be interchanged. As demonstrated in Table 3, the W210E mutant completely lost all three activities, while the H205M mutant displayed a 36-fold decrease in the cellulase k_{cat} and a 5-fold decrease in the mannanase k_{cat}. Similarly, the H205M/W210E mutant entirely abolished cellulase activity and reduced mannanase k_{cat} by 5-fold. However, both H205M and H205M/W210E mutants demonstrated a modest (~1.5-fold) increase in the xylanase k_{cat} compared to the wild-type *Tm*Cel5A. These results indicate that, while significantly diminishing cellulase and mannanase activities, H205M and H205M/W210E mutants not only restore but also slightly enhance the xylanase activity.

doi:10.6342/NTU202502385

3.5 Phylogenetic tree analysis of GH5 families and prediction of substrate binding modes of hemicellulases with substrate preference to xylan or mannan

As suggested by the above structural analysis and site-directed mutagenesis studies, *Tm*Cel5A has strong mannanase activity because of His205 and Trp210 in loop 6 (*Tm*loop). *Ct*Cel5T had strong xylanase activity but significantly lower mannanase activity than *Tm*Cel5A likely because of Glu360 in loop 8 and Met277 in loop 6 at the positions of Trp210 and His205 in loop 6 of *Tm*Cel5A. Glu360 may not be essential for xylanase activity as E360A still retained most of the BWX degradation k_{cat} and *Ct*Cel5E that has strong xylanase activity did not contain an amino acid at the position. However, *Ct*Cel5E has no mannanase activity at all, indicating that Glu360 and Met277 together keep both of xylanase and a lower level of mannanase activities in *Ct*Cel5T. The xylanase activity of *Ct*Cel5E and *Ct*Cel5T could still largely depended on the residues of Asn72, Trp82, His148, His149, Asn192, Trp349, and Tyr270 as shown in Figure 1b left panel.

To validate our hypothesis that Met277/Glu360 make *Ct*Cel5T a trifunctional cellulase/xylanase/mannanase, His205/Trp210 endow *Tm*Cel5A a bifunctional cellulase/mannanase, while no amino acid at these position leads *Ct*Cel5E a bifunctional cellulase/xylanase, we analyzed the amino acid sequences and crystal structures of several other cellulases/hemicellulases with or without known substrate preferences to xylan or mannan, to yield the phylogenetic tree for these enzymes (Figure 5a). Despite the

doi:10.6342/NTU202502385

substantial sequence diversity among the GH5 subfamilies, GH5_4, which includes *Ct*Cel5T, forms the closest clade to GH5_25, the subfamily containing *Ct*Cel5E and *Tm*Cel5A. When 124 GH5_4 enzymes and 17 GH5_25 enzymes from the CAZy database were aligned, the relative appearing frequency of each amino acid surrounding the substrates is represented by the size of its alphabetic symbol (Figure 5b). In GH5_4 subfamily, *Ct*Cel5T's Asn72, Trp82, His148, His149, Asn192, Glu193, Tyr270, Glu316, and Trp349 are extremely conserved, whereas the aligned positions of Met277, Asn351, and Glu360 exhibit sequence variabilities. Specifically, Met residue at the position of Met277 is conserved in only 10.5% of the sequences, with Leu being the predominant residue, appearing in 36.3% of the sequences. Totally, 80.8% of the sequences have nonpolar amino acids at this site, for providing hydrophobic interactions with substrates. Besides, Asn at the position of Asn351 in *Ct*Cel5T is the predominant residue in 70.2% of the sequences, assuming its role in forming hydrogen bonds with the (-2) sugar of xylobiose. Although the residues at the position of Glu360 show variability within the GH5_4 subfamily, Glu still appears in 41.1% of the sequences, with the second most common residue being Asp (4.8%), suggesting that negatively charged residues are important at this position. In contrast, key residues His205 and Trp210 involved in substrate binding and catalysis in *Tm*Cel5A are highly conserved within the GH5_25 subfamily, in 94% and 71% of the sequences, respectively, underscoring their essential

doi:10.6342/NTU202502385

roles.

To assess the residues in determining substrate specificities, the structures of several known mannanases and xylanases were superimposed, and residues corresponding to His205/Trp210 in *Tm*Cel5A and Met277/Glu360 in *Ct*Cel5T were highlighted for comparative analysis (Figure 5c). In previous studies, various GH5_4 multifunctional hemicellulases with xylanase activity, including *Cl*CelA from *Clostridium longisporum* (PDB: 6Q1I), RUM_05520 from *Ruminococcus champanellensis* (PDB: 6WQP), *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 (PDB: 6XSU), CAC0826 from *Clostridium acetobutylicum* (PDB: 6PZ7), *Cc*EngD from *Clostridium cellulovorans* (PDB: 3NDY), and *Pr*EglA from *Piromyces rhizinflata* (PDB: 3AYR), were also characterized to possess limited mannanase activity [27, 38, 39], just like *Ct*Cel5T characterized here. Through superimposition of these GH5_4 enzymes, we observed that they either have Met or Leu at the equivalent position of Met277 in *Ct*Cel5T, and most exhibit Glu at the corresponding position of Glu360 in *Ct*Cel5T, except for CAC286, which has Asp (also with a side-chain COOH) at this site. This suggests that these two residues are highly conserved among the identified *Ct*Cel5T-like enzymes. Met/Leu could possibly provide hydrophobic interactions with the (-1) sugar moiety, and Glu/Asp could form hydrogen bonds with hydroxyl group on (-2) sugar ring. Additionally, another characterized xylanase, *Cr*Cel5A from *Cellulosilyticum ruminicola* [40], lacks a solved

doi:10.6342/NTU202502385

structure but also contains the conserved Met602/Asp686 residue pair. These two residues at the specified positions are highly conserved among GH5_4 xylanases.

On the contrary, characterized mannanases including *Cp*Man5B from *Caldanaerobius polysaccharolyticus* (PDB: 3W0K), *Tm*Cel5B from *Thermotoga maritima* (PDB: 7EC9), and *Bs*Amn5A from *Bacillus* sp. strain JAMB-602 (PDB: 1WKY) possess conserved Trp and His residues at the same positions of His205 and Trp210 in *Tm*Cel5A (Figure 5c). Although *Cp*Man5B and *Tm*Cel5B are classified within the GH5_36 subfamily [41, 42], phylogenetically distant from GH5_25 subfamily, they display significant structural homology to *Tm*Cel5A (RMSD = 1.189 Å of 181 Cα atoms and 1.166 Å of 202 Cα atoms, respectively), and the His/Trp pair in these enzymes appears to form hydrogen bonds with the 6-hydroxyl group of the (-1) sugar moiety, while Trp likely engages in stacking interactions with the sugar at the (−2) subsite. Besides, *Bs*Amn5A, a mannanase from the GH5_8 subfamily [43], sequentially distinct from both GH5_4 and GH5_25 subfamilies, exhibits less sequence and structural similarity to *Tm*Cel5A (RMSD = 2.37 Å of 151 Cα atoms), but still has a conserved His258/Trp291 pair, occupying the same positions as His205/Trp210 in *Tm*Cel5A. Interestingly, while His205 and Trp210 are located in the loop 6 (*Tm*loop) in *Tm*Cel5A, His258 in *Bs*Amn5A resides in loop 7, and Trp291 in loop 8 with a structural arrangement akin to Glu360 in loop 8 of *Ct*Cel5T. In addition, other GH5_8 mannanases including MANPN11 from *Niallia nealsonii* PN-11, *Bs*Man5A from

doi:10.6342/NTU202502385

Bacillus sp. N16-5 (PDB: 3JUG), and *Ba*Man5A from *Salipaludibacillus agaradhaerens* (PDB: 2WHJ) [44-46], share the same His/Trp feature as well. Furthermore, several identified cellulases, including BcelFp from *Fervidobacterium pennivorans* DSM9078 (PDB: 6KDD), *Mt*Glu5 from *Meiothermus taiwanensis* WR-220 (PDB: 7VT4), and *Ct*CelC from *Clostridium thermocellum* (PDB: 1CEC) [47-49], also exhibit the His/Trp feature, suggesting their potential mannanase activity.

On the other hand, the cellulases/xylanases possessing neither the His/Trp nor the Met/Glu pairs at the corresponding positions were identified, including *Ct*Cel5E and *Cel*CCD from *Clostridium cellulolyticum* [58]. *Cel*CCD shares high sequence similarity with *Ct*Cel5E (69.74%), and its AlphaFold3-predicted structure was superimposed well with the *Ct*Cel5E-xylobiose crystal structure (PDB: 4U5I) with an RMSD of 0.381 Å over 233 Cα atoms, preserving the active-site architecture (Figure S3). The absence of the critical pair correlates with an inability to hydrolyze mannan and reinforces their specialization toward cellulose and xylan as cellulase and xylanase.

3.6 Confirmation of the prediction of *Mt*Glu5 as a multifunctional cellulase/mannanase

Previously, *Mt*Glu5 from *Meiothermus taiwanensis* WR-220 were characterized as a GH5 endoglucanase, but no mannanase or xylanase activity was reported [48]. Since *Mt*Glu5 also exhibits the His/Trp feature as *Tm*Cel5A, we hypothesized that it may

doi:10.6342/NTU202502385

possess mannanase activity. To validate our prediction, the recombinant *Mt*Glu5 was expressed and purified following the protocol described in the previous study [48]. We found it indeed possessed cellulase/mannanase activities but lacked xylanase activity. The optimal pH and temperature conditions for *Mt*Glu5 in the hydrolysis of CMC and mannan were determined as shown in Figure S4. Under the optimal condition of pH 5 and 80°C, its k_{cat} values of cellulase and mannanase were measured to be 1818 ± 246.3 and 832.3 ± 69.1 min^{-1} and the K_m values with respect to CMC and mannan were 20.5 ± 5.9 and 9.8 ± 1.6 mg/mL, respectively (Table S1). This supports the notion that the His/Trp structural feature predicts the cellulase/mannanase activities.

4. DISCUSSION

GH5 represents a large and functionally diverse enzyme family characterized by broad substrate specificities and multifunctional catalytic activities. These attributes render GH5 enzymes particularly valuable for applications such as plant cell wall degradation in biofuel production and other industrial applications. Within this family, *Ct*Cel5E, *Tm*Cel5A, and *Ct*Cel5T and others are of significant interest due to their distinct enzymatic activities and substrate specificities, as demonstrated in previous studies [14, 17, 23, 27, 50-57]. These enzymes exhibit activity against xylan and/or mannan, as well as other polysaccharides, including lichenan and xyloglucan (XG) [37]. As demonstrated in this study, we focus on studying the three representative enzymes *Ct*Cel5E, *Tm*Cel5A,

doi:10.6342/NTU202502385

and *Ct*Cel5T, which were reported to have cellulase/xylanase, cellulase/mannanase (although we also identified a very weak xylanase activity), and cellulase/xylanase/mannanase activities against the different linear oligosaccharides CMC, BWX, and mannan, respectively. Notably, *Ct*Cel5T was confirmed in this study as a trifunctional enzyme capable of degrading CMC, BWX, and mannan, although its mannanase activity is lower than that of *Tm*Cel5A. Based on our measurements, the optimal reaction conditions for polysaccharides depolymerization were identified as follows: 60°C, pH 4 for cellulase activity; 70°C, pH 5 for xylanase activity; and 50°C, pH 4 for mannanase activity. The primary hydrolytic products under these conditions were cellobiose and cellotriose from CMC, xylotriose from BWX, and mannobiose, mannotriose, and mannotetraose from mannan.

At molecular scale, *Ct*Cel5T adopts the canonical $(\beta/\alpha)_8$ TIM-barrel fold [27], a common structural feature of the GH5 family. Sequence and structural alignments of *Ct*Cel5T with *Ct*Cel5E and *Tm*Cel5A identified Glu193 and Glu316 as the general acid/base and nucleophile, respectively. Consistent with this prediction, alanine substitution mutants of these residues resulted in a complete loss of enzymatic activity (Table 1 and Figure 4). Additionally, several residues highly conserved across *Ct*Cel5E, *Tm*Cel5A, and *Ct*Cel5T, including Asn72, Trp82, His148, His149, Asn192, Tyr270, and Trp349, were confirmed as catalytically essential through site-directed mutagenesis

doi:10.6342/NTU202502385

(Table 1 and Figure 4). In contrast, the substrate specificities of these enzymes appear to be dictated by less conserved residues, such as Met277/Glu360 of *Ct*Cel5T and His205/Trp210 of *Tm*Cel5A. Met277 and Glu360 are provided by loop 6 and 8 of *Ct*Cel5T, respectively, while both His205 and Trp210 are from loop 6 (*Tm*loop) of *Tm*Cel5A. These two positions occupied by the residues apparently determine the substrate preferences, mediating a trade-off behavior between xylanase and mannanase activities (when one activity is high, the other activity is low).

From our previous study [17], which elucidated the crystal structures of *Ct*Cel5E (a cellulase/xylanase) and *Tm*Cel5A (a cellulase/mannanase) and their differing substrate specificities, we successfully engineered a trifunctional enzyme, *Ct*Cel5E-*Tm*loop, by incorporating the *Tm*loop into *Ct*Cel5E, and further optimization led to the mutant *Ct*Cel5E-*Tm*loop/F267A/F323I [25]. In this study, we investigated another structural homologue, *Ct*Cel5T, to expand our understanding of substrate specificity and activity modulation in GH5s. Based on the biochemical characterization of wild-type and mutant *Ct*Cel5T, we propose a mechanism wherein two key amino acids from the non-conserved loops 8 and 6 are essential for determining whether the GH5 multifunctional cellulases/hemicellulases exhibit higher xylanase or mannanase activity, while maintaining cellulase activity. Unlike *Ct*Cel5E, which exhibits only cellulase and xylanase activities, *Ct*Cel5T contains Met277 and Glu360, enabling it to retain a low level

doi:10.6342/NTU202502385

of mannanase activity. While *Ct*Cel5E also possesses His277, corresponding to His205 of *Tm*Cel5A and Met277 of *Ct*Cel5T, this residue is positioned away from the active site, even in the cellobiose- and xylobiose-bound structures. Therefore, *Ct*Cel5E is absolutely without mannanase activity. Additionally, although *Tm*Cel5A lacks loop 8, it retains Trp286, corresponding to Trp349 in *Ct*Cel5T, enabling very weak xylanase activity (see Figure 3b).

Structural superimposition of cellulases/hemicellulases with known substrate specificities as shown in Figure 5c reveals distinct patterns of residue conservation in loops 6 and 8, correlating with substrate preferences. Hemicellulases that preferentially degrade xylan while retaining some mannanase activity feature a conserved negatively charged residue (Glu or Asp) in loop 8 and a nonpolar residue (Met or Leu) in loop 6. In contrast, hemicellulases with specificity toward mannan possess conserved Trp and His residues in loop 6, while those with the substrate specificity solely toward xylan lack amino acids at these positions. Key residues at these positions of Glu360/Trp210 and Met277/His205 appear to play a critical role in determining the substrate specificities for the hemicellulases which are functioned predominantly as xylanases or mannanases. This analysis provides a framework for predicting the substrate specificity of GH5 enzymes with previously unannotated functions, offering valuable predictions for functional annotation and enzyme engineering.

doi:10.6342/NTU202502385

Our findings elucidate the structure-activity relationships of GH5 multifunctional cellulases/hemicellulases, which enhance our understanding on the substrate-recognition mechanisms of these useful enzymes. Considering the importance of GH5 enzymes in biomass conversion and other applications, our findings are interesting from the viewpoint of fundamental research, engineering, and development.

doi:10.6342/NTU202502385

TABLES

Table 1. Kinetic parameters of wild-type (WT) and mutant *Ct*Cel5T.

(performed by the previous master student Yung-Yeh Chen in our laboratory)

	CMC			BWX			Mannan		
*Ct*Cel5T	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})
WT	4/60	20.4 ± 2.8	8733.0 ± 574.0	5/70	9.5 ± 0.7	832.6 ± 31.6	4/50	5.4 ± 1.6	47.5 ± 5.7
E193A	4/60	N.D.	N.D.	5/70	N.D.	N.D.	4/50	N.D.	N.D.
E316A	4/60	N.D.	N.D.	5/70	N.D.	N.D.	4/50	N.D.	N.D.
N72A	4/60	11.0 ± 2.3	537.6 ± 63.6	5/70	N.D.	N.D.	4/50	N.D.	N.D.
W82A	4/60	13.3 ± 4.9	2602.0 ± 387.7	5/70	N.D.	N.D.	4/50	N.D.	N.D.
H148A	4/60	4.5 ± 1.2	581.5 ± 60.2	5/70	N.D.	N.D.	4/50	N.D.	N.D.
H149A	4/60	37.2 ± 7.2	6576.0 ± 768.3	5/70	N.D.	N.D.	4/50	N.D.	N.D.
N192A	4/60	8.9 ± 2.6	563.6 ± 86.8	5/70	N.D.	N.D.	4/50	N.D.	N.D.
Y270A	4/60	N.D.	N.D.	5/70	N.D.	N.D.	4/50	N.D.	N.D.
W349A	4/60	N.D.	N.D.	5/70	N.D.	N.D.	4/50	N.D.	N.D.
N351A	4/60	31.5 ± 8.9	7323.0 ± 746.0	5/70	10.9 ± 2.0	118.7 ± 12.6	4/50	11.9 ± 2.4	65.7 ± 7.1

N.D. means non-detectable (activity was below the detection limit).

doi:10.6342/NTU202502385

Table 2. Kinetic parameters of *Ct*Cel5T Met277 and Glu360 mutants.

	CMC			BWX			Mannan		
*Ct*Cel5T	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})
WT	4/60	20.4 ± 2.8	8733.0 ± 574.0	5/70	9.5 ± 0.7	832.6 ± 31.6	4/50	5.4 ± 1.6	47.5 ± 5.7
M277A	4/60	8.9 ± 3.2	4650.0 ± 546.0	5/70	18.6 ± 3.0	775.7 ± 79.1	4/50	N.D.	N.D.
M277H	4/60	22.2 ± 3.9	1054.0 ± 99.4	5/70	11.01 ± 2.332	134.7 ± 13.68	4/50	3.36 ± 0.64	20.75 ± 1.76
E360A	4/60	9.1 ± 2.2	1966.0 ± 160.8	5/70	27.6 ± 8.0	448.0 ± 98.3	4/50	8.3 ± 1.8	66.5 ± 6.8
E360W	4/60	31.6 ± 5.6	3153.0 ± 340.5	5/70	N.D.	N.D.	4/50	6.3 ± 2.5	6.7 ± 1.4
E360H	4/60	20.1 ± 3.9	1076.0 ± 115.3	5/70	N.D.	N.D.	4/50	14.8 ± 4.3	41.1 ± 7.1
E360Q	4/60	2.4 ± 0.5	322.3 ± 16.5	5/70	4.9 ± 0.6	4 ± 0.2	5/50	N.D.	N.D.
E360K	4/60	18.9 ± 2.5	879.0 ± 63.0	5/70	N.D.	N.D.	4/50	N.D.	N.D.
E360Y	4/60	6.0 ± 1.5	222.6 ± 19.5	5/70	N.D.	N.D.	4/50	N.D.	N.D.
E360R	4/60	N.D.	N.D.	5/70	N.D.	N.D.	4/50	N.D.	N.D.
M277H/E360W	4/60	25.2 ± 4.6	316.4 ± 33.3	5/70	5.0 ± 1.1	50.0 ± 4.3	4/50	14.5 ± 3.9	216.7 ± 32.3
M277H/E360H	4/60	7.7 ± 0.9	362.1 ± 16.4	5/70	3.2 ± 0.4	73.9 ± 2.5	4/50	18.8 ± 4.5	227.9 ± 34.6

N.D. means non-detectable (activity was below the detection limit).

doi:10.6342/NTU202502385

Table 3. Kinetic parameters of wild-type (WT) and mutant *Tm*Cel5A.

	CMC			BWX			Mannan		
*Tm*Cel5A	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/mL)	k_{cat} (min^{-1})
WT	4/80	26.4 ± 8.37	10609 ± 1674	5/80	7.85 ± 2.85	27.88 ± 4.88	5/80	2.86 ± 0.56	1280 ± 84.49
H205M	4/80	78.76 ± 57.63	857.0 ± 514.7	5/80	9.704 ± 4.269	39.20 ± 8.828	5/80	9.778 ± 3.651	267.4 ± 47.20
W210E	4/80	N.D.	N.D.	5/80	N.D.	N.D.	5/80	N.D.	N.D.
H205M/W210E	4/80	N.D.	N.D.	5/80	7.355 ± 1.786	44.24 ± 4.654	5/80	12.24 ± 3.924	252.2 ± 41.41

N.D. means non-detectable (activity was below the detection limit).

doi:10.6342/NTU202502385

Table 4. Forward (F) and reverse (R) primers used in generating mutant *Ct*Cel5T.

Primer	Nucleotide sequence (5'-3')
E193A-F	CTACTTCGCGCGGCGCATTCATGGTTTCAAATAACAGG
E193A-R	CCTGTTATTTGAAACCATGAATGCGCCGCGCGAAGTAG
E316A-F	TATCGATGGTACCGAACGCACCAATGATAACCGCA
E316A-R	TGCGGTTATCATTGGTGCGTTCGGTACCATCGATA
N72A-F	GAGCGTCCAGGGTAGCGCCCAGGTTCCAGC
N72A-R	GCTGGAACCTGGGCGCTACCCTGGACGCTC
W82A-F	GCGCGGGTTACCCGCCGCAGTTTCGGTC
W82A-R	GACCGAAACTGCGGCGGGTAACCCGCGC
H148A-F	CCACGTGTTGTCGTGGGCCAGGTTGATGATCGCA
H148A-R	TGCGATCATCAACCTGGCCCACGACAACACGTGG
H149A-F	TGATCCACGTGTTGTCGGCGTGCAGGTTGATGATCG
H149A-R	CGATCATCAACCTGCACGCCGACAACACGTGGATCA
N192A-F	TACTTCGCGCGGCTCAGCCATGGTTTCAAATAACAGGTGG
N192A-R	CCACCTGTTATTTGAAACCATGGCTGAGCCGCGCGAAGTA
Y270A-F	GCGAAGAAGTACGGGGAGGCAGCGTGGATAGAAACGAT
Y270A-R	ATCGTTTCTATCCACGCTGCCTCCCCGTACTTCTTCGC
M277A-F	GTACCGTTAACGTCCGCAGCGAAGAAGTACGGGGAGTAA
M277A-R	TTACTCCCCGTACTTCTTCGCTGCGGACGTTAACGGTAC
W349A-F	GTTGTAGTAACCGTTATCCGCCCAGAAAACCGCGATACCA
W349A-R	TGGTATCGCGGTTTTCTGGGCGGATAACGGTTACTACAAC
N351A-F	TCGCGGTTTTCTGGTGGGATGCCGGTTACTACAACCC
N351A-R	GGGTTGTAGTAACCGGCATCCCACCAGAAAACCGCGA
E360A-F	CAGAGCGTAGGTTGCCGCGTCACCTGG
E360A-R	CCAGGTGACGCGGCAACCTACGCTCTG
M277H-F	CATGACGTTAACGGTACCTCT
M277H-R	AGCGAAGAAGTACGGGGAGTA
E360W-F	TGGACCTACGCTCTGCTGAACCGT
E360H-F	CATACCTACGCTCTGCTGAACCGT
E360Q-F	CAGACCTACGCTCTGCTGAACCGT
E360K-F	AAAACCTACGCTCTGCTGAACCGT
E360Y-F	TACACCTACGCTCTGCTGAACCGT
E360R-F	CGTACCTACGCTCTGCTGAACCGT
E360-R	CGCGTCACCTGGGTTGTAGTAACC

doi:10.6342/NTU202502385

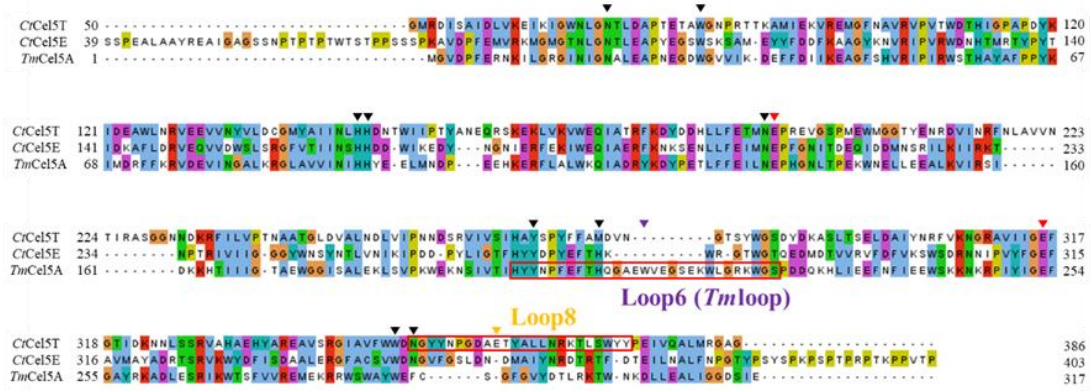
Table 5. Forward (F) and reverse (R) primers used in generating mutant *Tm*Cel5A.

Primer	Nucleotide sequence (5'-3')
H205M-F	GAATTTACCATGCAAGGAGCTGAG
H205M-R	GAAAGGATTGTAGTAGTGAATTGTAACTAT
W210E-F	CCCATCAAGGAGCTGAGGAGGTGGAAGGATCTGAGA
W210E-R	TCTCAGATCCTTCCACCTCCTCAGCTCCTTGATGGG

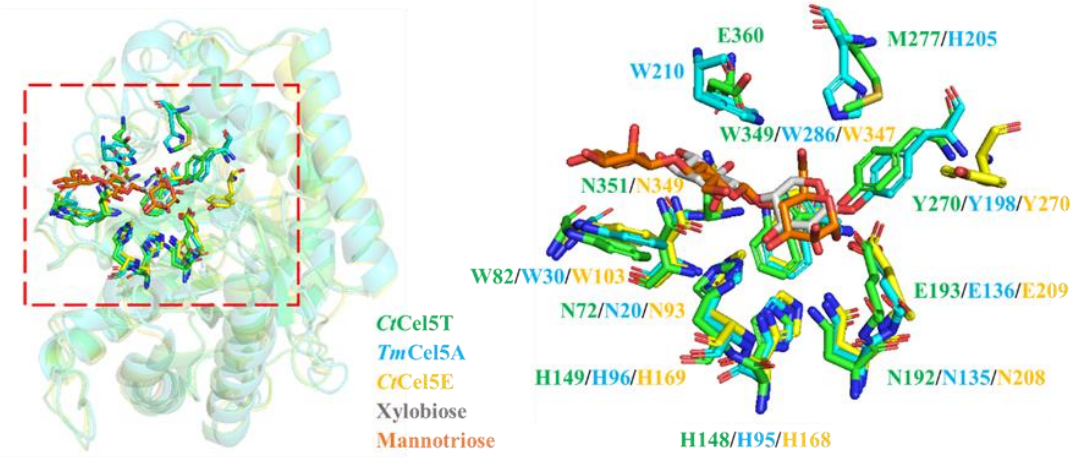
doi:10.6342/NTU202502385

FIGURES

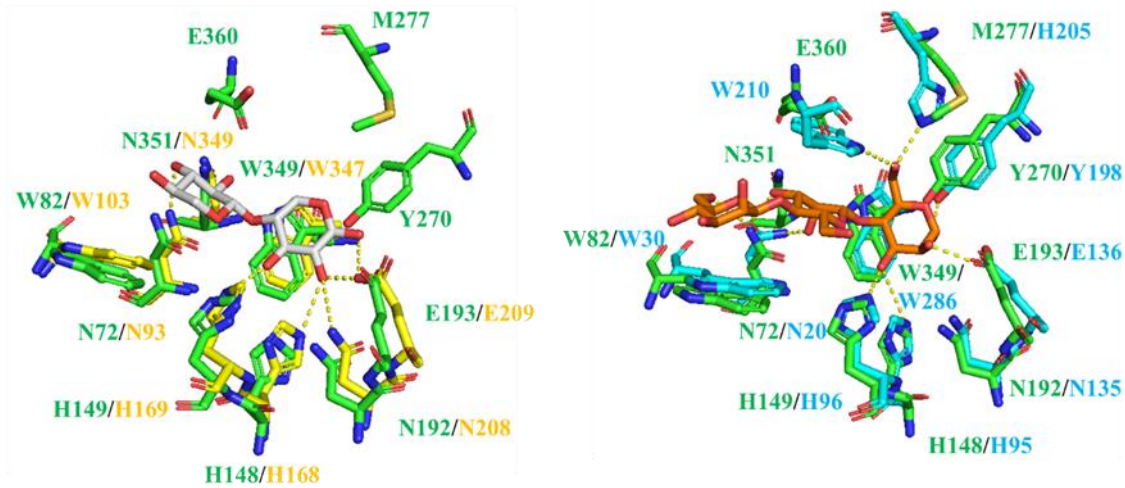
(a)



(b)



(c)

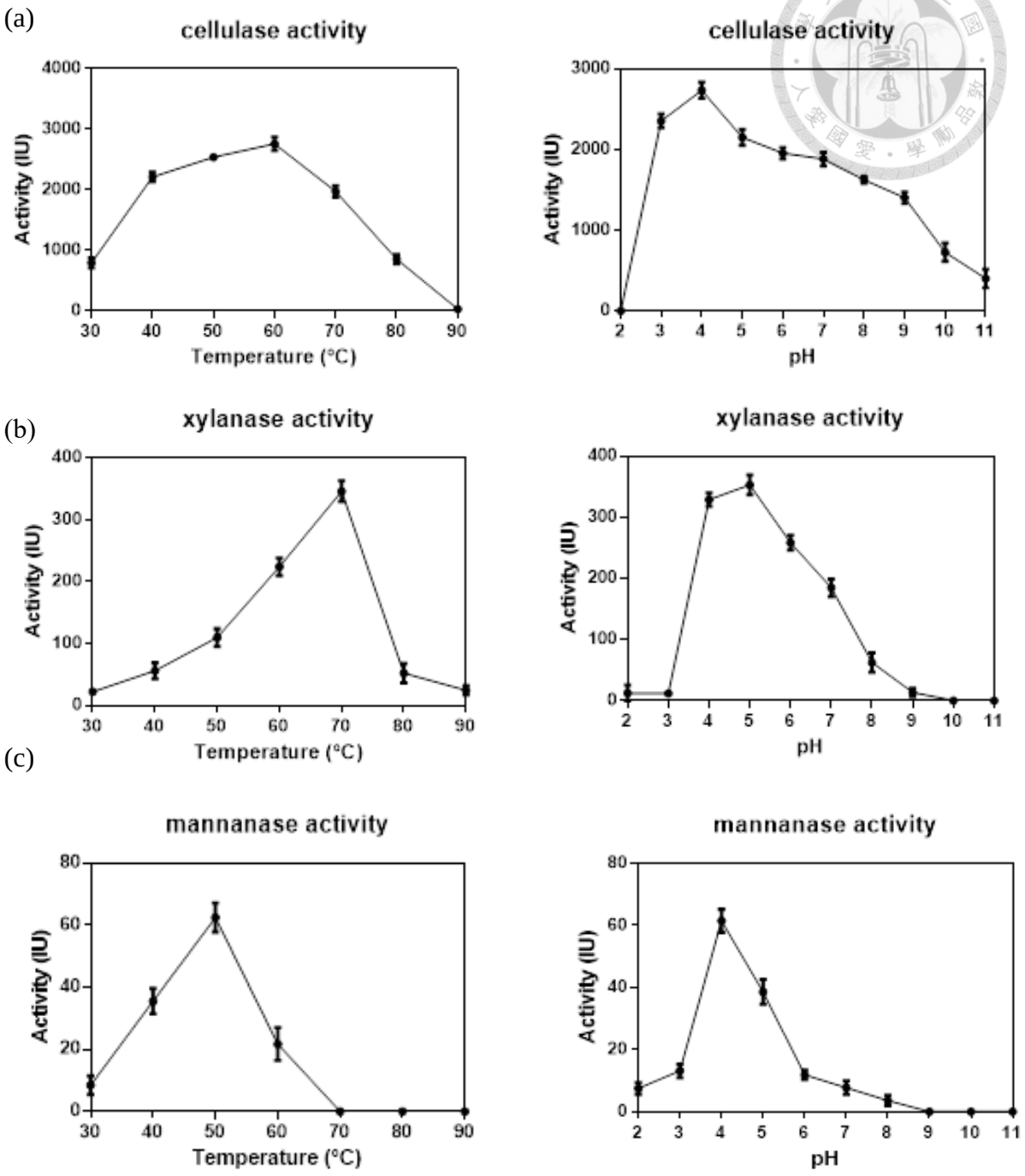


doi:10.6342/NTU202502385

Figure 1. Structural comparison and sequence alignment of key residues in *Ct*Cel5T, *Ct*Cel5E, and *Tm*Cel5A.

(a) Multiple sequence alignment of *Ct*Cel5T, *Ct*Cel5E, and *Tm*Cel5A. Red triangles indicate the Glu as the general bases and the nucleophiles and black triangles indicate the key residues involved in the catalytic reactions and the substrate binding, which were mutated in the study. Trp210 in loop 6 of *Tm*Cel5A and Glu360 in loop 8 of *Ct*Cel5T are indicated in purple and yellow triangle, respectively, while the Loops are highlighted in red hollow rectangles. (b) Superimposition of the active-site residues of *Ct*Cel5T (PDB:4IM4), *Tm*Cel5A-mannotriose (PDB: 3AZS), and *Ct*Cel5E-xylobiose (PDB: 4U5I) structures, colored in green, cyan, and yellow, respectively. Xylobiose is colored in gray and mannotriose in orange. (c) Separation of the above superimposed three structures into two, superimposition of *Ct*Cel5T and *Ct*Cel5E-xylobiose structures (left panel) and superimposition of *Ct*Cel5T and *Tm*Cel5A-mannotriose structures (right panel). The yellow dash lines indicate hydrogen bonds.

doi:10.6342/NTU202502385



doi:10.6342/NTU202502385

Figure 2. Optimal conditions and product analysis for *Ct*Cel5T reactions.

(performed by the previous master student Yung-Yeh Chen in our laboratory)

Temperature- and pH-dependent profiles of *Ct*Cel5T for cellulase (a), xylanase (b), and mannanase (c) activities, respectively, and digested products from 1% CMC, BWX and mannan treated with 1 μM *Ct*Cel5T under optimal conditions (d), are shown. All experiments were performed in three replicates.

doi:10.6342/NTU202502385

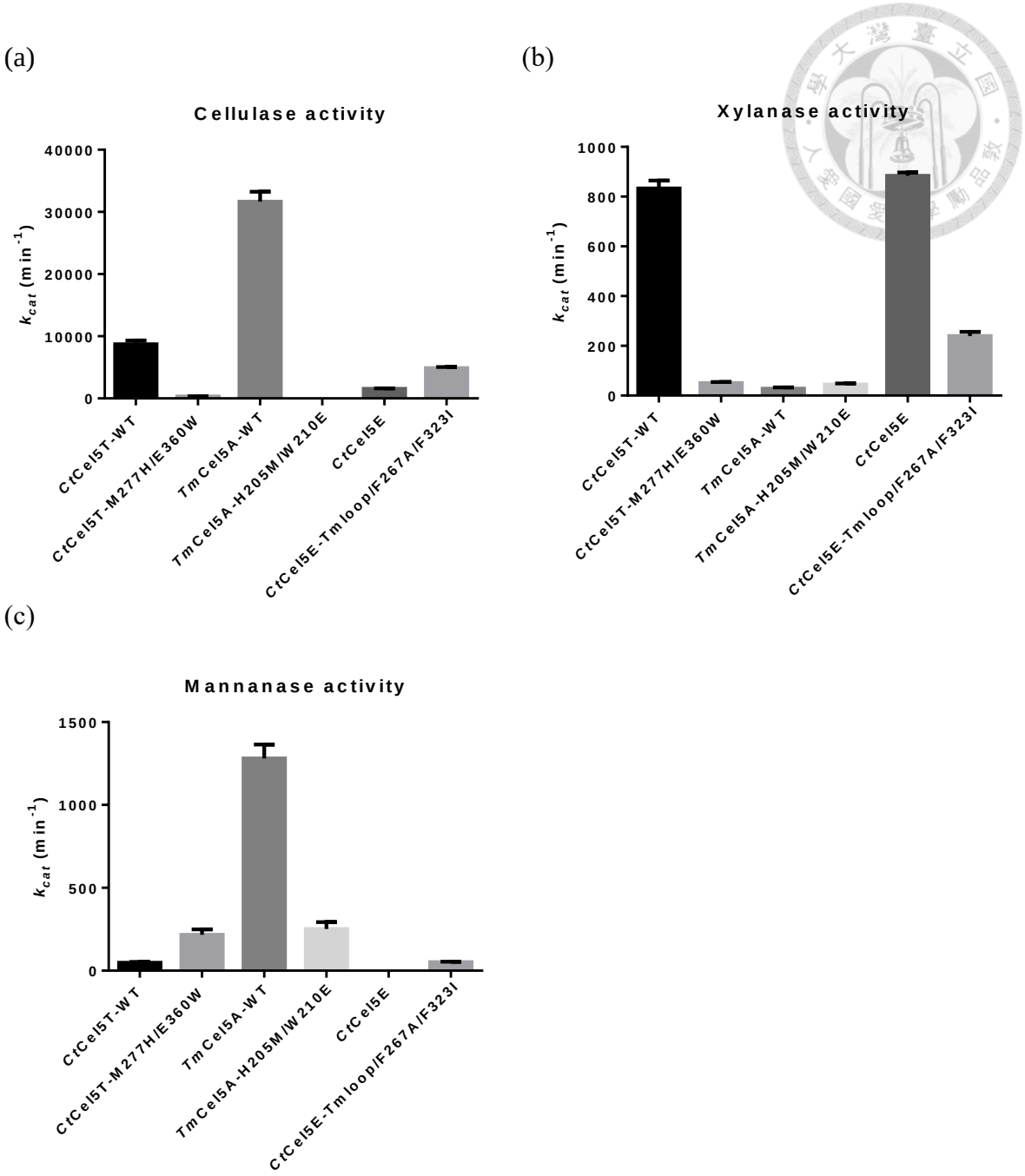
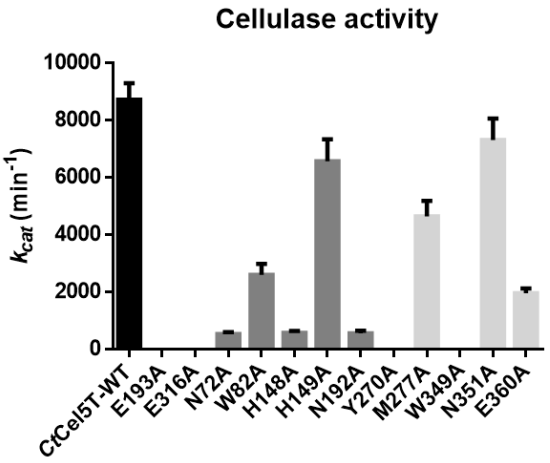


Figure 3. The k_{cat} values of *Ct*Cel5T, *Tm*Cel5A, *Ct*Cel5E, and their engineered enzymes toward different substrates.

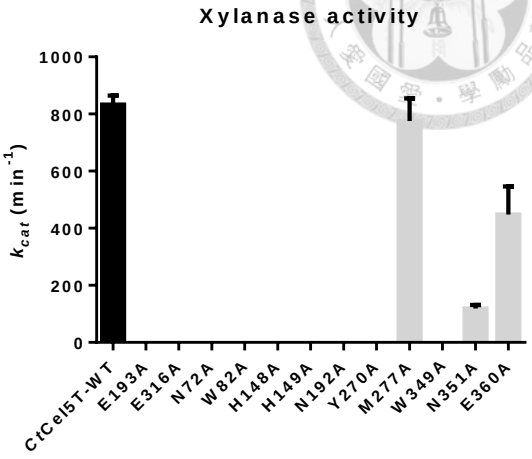
The k_{cat} values under the optimal conditions toward CMC (a), BWX (b), and mannan (c) were measured by DNS assay. Reactions were performed in triplicate.

doi:10.6342/NTU202502385

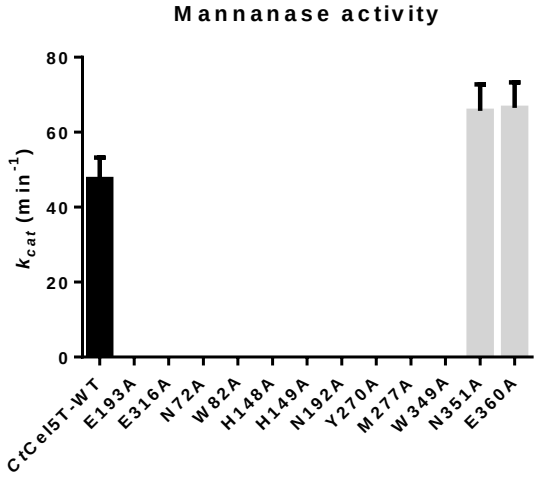
(a)



(b)



(c)



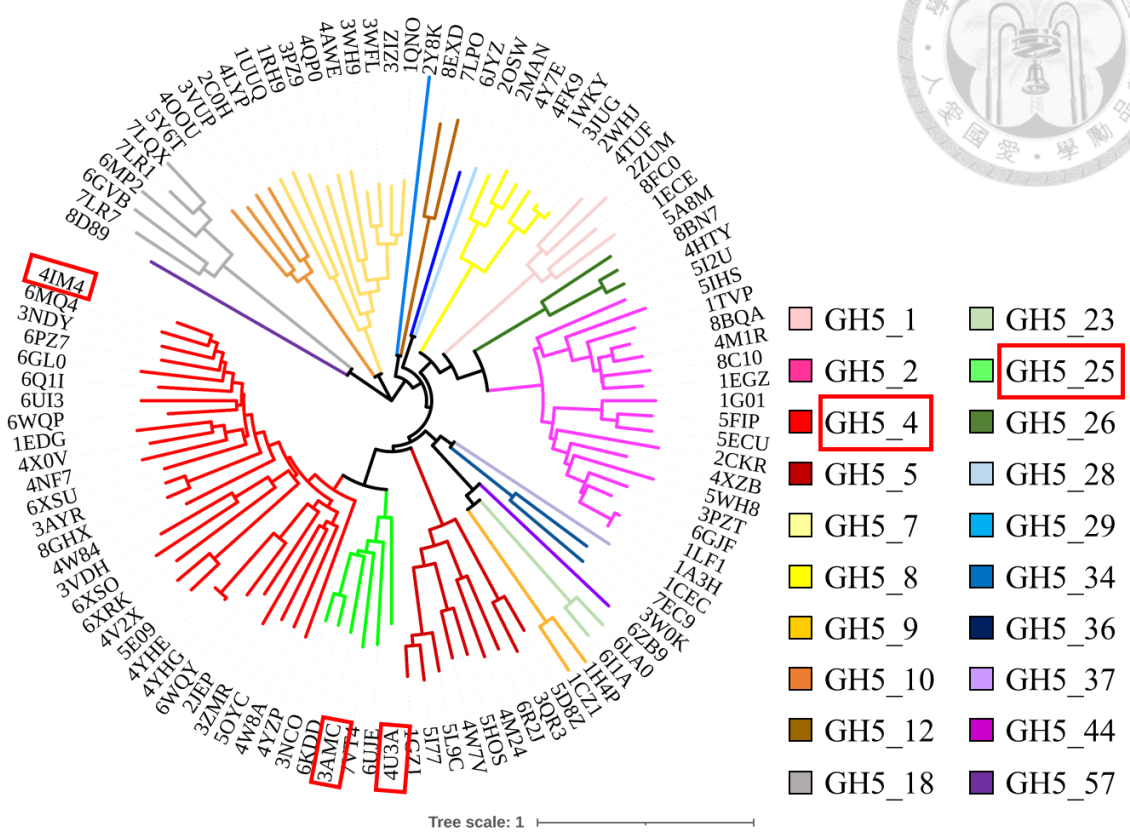
doi:10.6342/NTU202502385

Figure 4. The k_{cat} values of *Ct*Cel5T mutants toward different substrates.

The k_{cat} values of the mutant *Ct*Cel5T toward CMC (a), BWX (b), and mannan (c) under the optimal conditions measured by DNS assay are shown. Black bar represents the *Ct*Cel5T wild type, dark gray bars represent *Ct*Cel5T with mutation in the highly conserved residues (E193A, E316A, N72A, W82A, H148A, H149A, N192A, Y270A, and W349A), and light gray bars represent *Ct*Cel5T mutated in the non-conserved residues (M277A, N351A, and E360A). Each reaction was performed in triplicate and standard deviations are indicated.

doi:10.6342/NTU202502385

(a)



(b)

doi:10.6342/NTU202502385

(c)

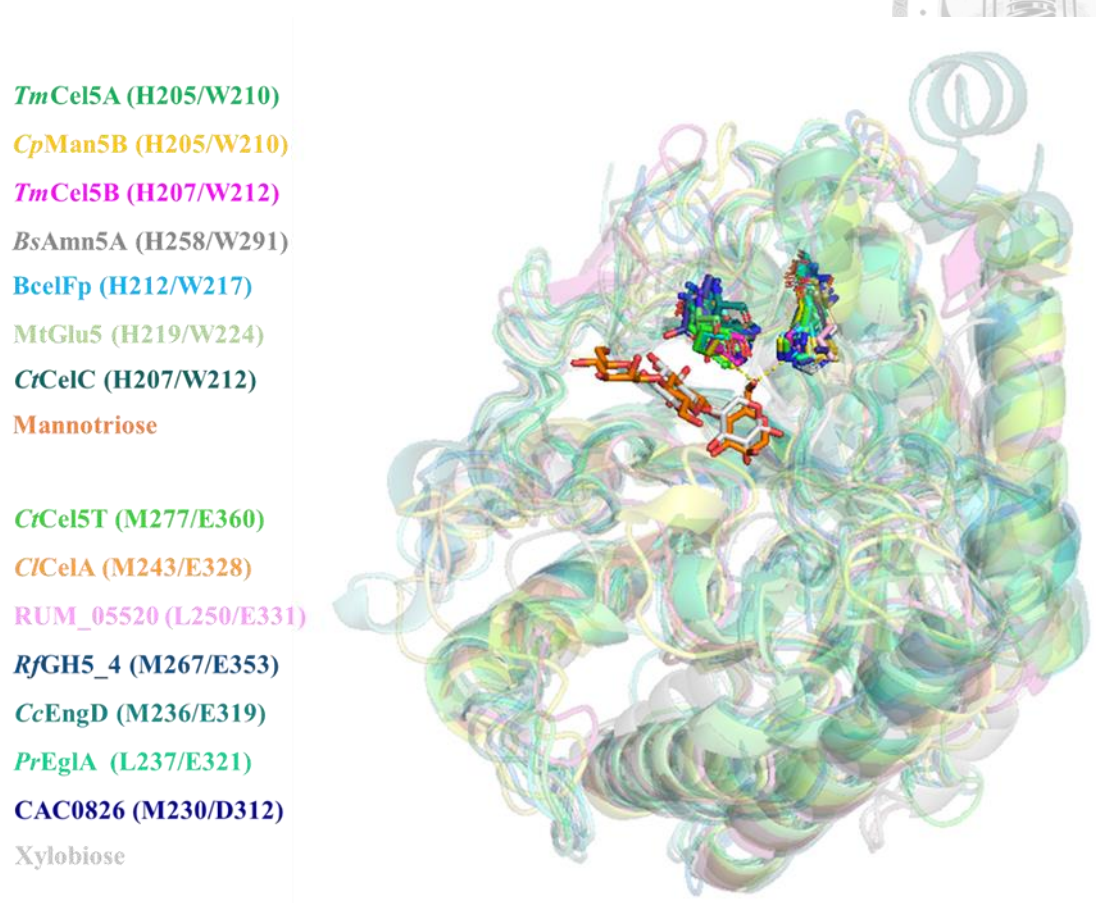


Figure 5. Phylogenetic analysis of GH5 families.

(a) Neighbor-joining phylogenetic tree of structural characterized GH5 families. Sequences of total 106 GH5 enzymes were obtained from CAZy database (http://www.cazy.org). GH5 subfamilies are labeled in different colors, and *Ct*Cel5T (PDB: 4IM4), *Ct*Cel5E (PDB: 4U3A) and *Tm*Cel5A (PDB: 3AMC) are highlighted in red hollow rectangles. (b) Sequence analysis of the active-site architecture in GH5_4 and GH5_25 enzymes. Residues surrounding the substrates were selected for sequence analyses. The size of each letter reflects the conservation level of the residue across all

doi:10.6342/NTU202502385

sequences within the subfamilies. (c) The classification of multifunctional cellulases/xylanases, cellulases/mannanases, and cellulases/xylanases/mannanases based on their structural features in loop 6 and 8. *Tm*Cel5A (PDB: 3AMC; colored green) was superimposed with *Cp*Man5B from *Caldanaerobius polysaccharolyticus* (PDB: 3W0K; colored yellow), *Tm*Cel5B from *Thermotoga maritima* (PDB: 7EC9; colored magneta), *Bs*Amn5A from *Bacillus* sp. strain JAMB-602 (PDB: 1WKY; colored gray), BcelFp from *Fervidobacterium pennivorans* DSM9078 (PDB: 6KDD; colored cyan), *Mt*Glu5 from *Meiothermus taiwanensis* WR-220 (PDB: 7VT4; colored lime), *Ct*CelC from *Clostridium thermocellum* (PDB: 1CEC; colored deepteal), *Ct*Cel5T from *Clostridium thermocellum* (PDB: 4IM4; colored limegreen), *Cl*CelA from *Clostridium longisporum* (PDB: 6Q1I; colored yelloworange), RUM_05520 from *Ruminococcus champanellensis* (PDB: 6WQP; colored lightpink), *Rf*GH5_4 from *Ruminococcus flavefaciens* FD-1 (PDB: 6XSU; colored marine), *Cc*EngD from *Clostridium cellulovorans* (PDB: 3NDY, colored teal), *Pr*EglA from *Piromyces rhizinflata* (PDB: 3AYR; colored aquamarine), and CAC0826 from *Clostridium acetobutylicum* (PDB: 6PZ7; colored deepblue). Mannotriose from *Tm*Cel5A and xylobiose, which is extracted from *Ct*Cel5E (PDB: 4U5I), are colored in orange and white, respectively.

doi:10.6342/NTU202502385

SUPPLEMENTARY INFORMATION

Table S1. Kinetic parameters of wild-type (WT) *Mt*Glu5.

	CMC			BWX			Mannan		
*Mt*Glu5	pH/Temp	K_m (mg/ml)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/ml)	k_{cat} (min^{-1})	pH/Temp	K_m (mg/ml)	k_{cat} (min^{-1})
WT	2/70	20.5 ± 5.9	1818 ± 246.3	N.D.	N.D.	N.D.	4/80	9.8 ± 1.6	832.3 ± 69.1

N.D. means non-detectable (activity was below the detection limit).

doi:10.6342/NTU202502385

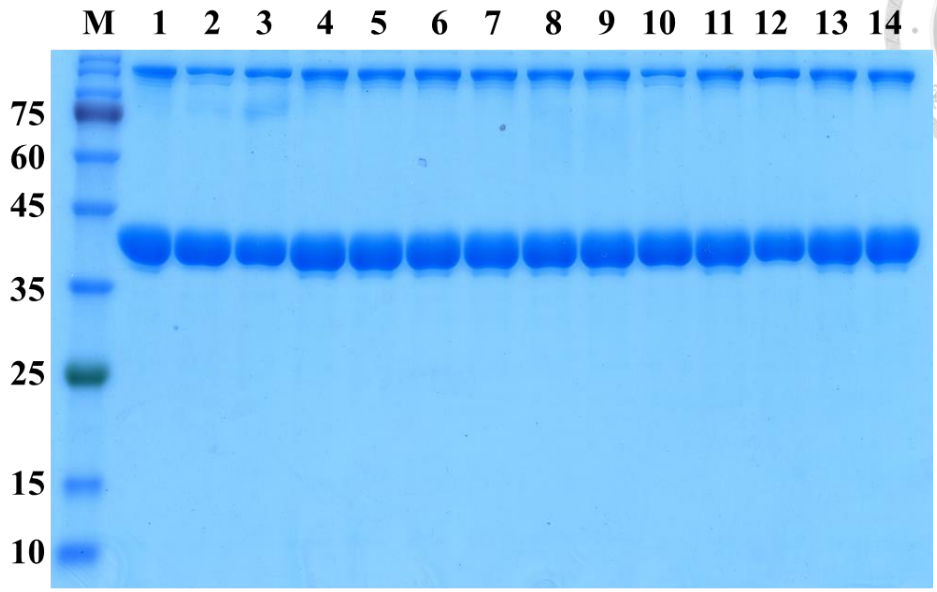


Figure S1. SDS-PAGE analysis of purified wild-type (WT) and mutant *Ct*Cel5T (42.5 kDa) and *Tm*Cel5A (41.0 kDa).

M, marker; 1, *Ct*Cel5T-WT; 2, *Ct*Cel5T-M277H; 3, *Ct*Cel5T-E360W; 4, *Ct*Cel5T-E360H; 5, *Ct*Cel5T-E360Q; 6, *Ct*Cel5T-E360K; 7, *Ct*Cel5T-E360Y; 8, *Ct*Cel5T-E360R; 9, *Ct*Cel5T-M277H-E360W; 10, *Ct*Cel5T-M277H-E60H; 11, *Tm*Cel5A-WT; 12, *Tm*Cel5A-H205M; 13, *Tm*Cel5A-W210E; 14, *Tm*Cel5A-H205M-W210E

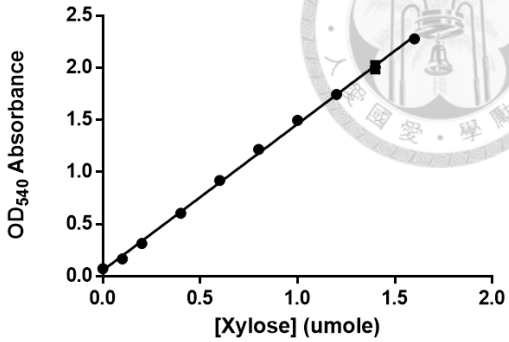
doi:10.6342/NTU202502385

(a)

Equation: Y = 1.835*X + 0.05684

R^2 = 0.998

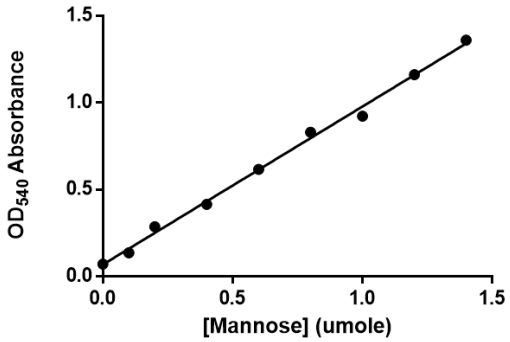
(b)



Equation: Y = 1.404*X + 0.05737

R^2 = 0.999

(c)



Equation: Y = 0.9091*X + 0.06799

R^2 = 0.996

Figure S2. The linear standard curves of DNS assay using glucose (a), xylose (b) and mannose (c) as the products.

The absorbances at 540 nm of all reactions were converted into the amount of product generated by GHs according to the standard equations.

doi:10.6342/NTU202502385

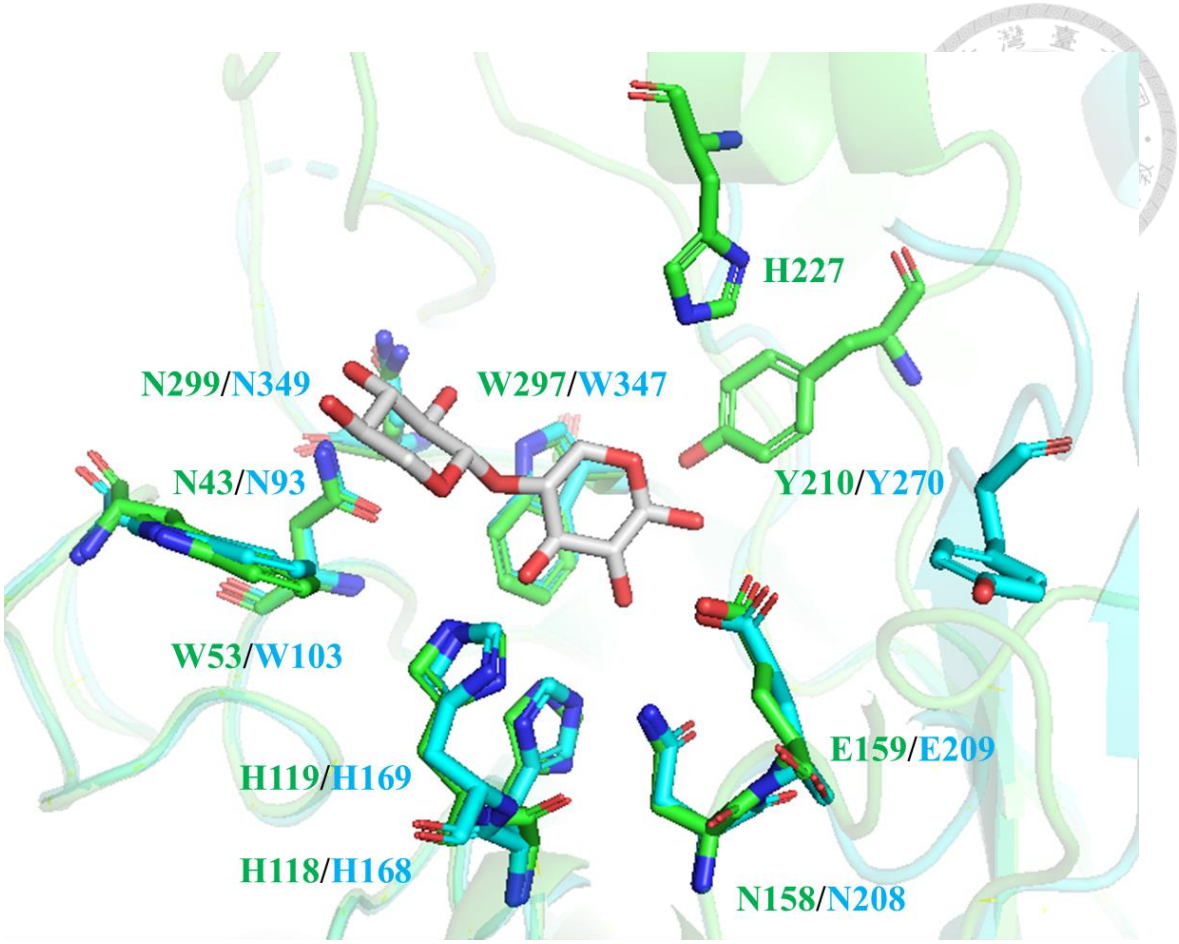
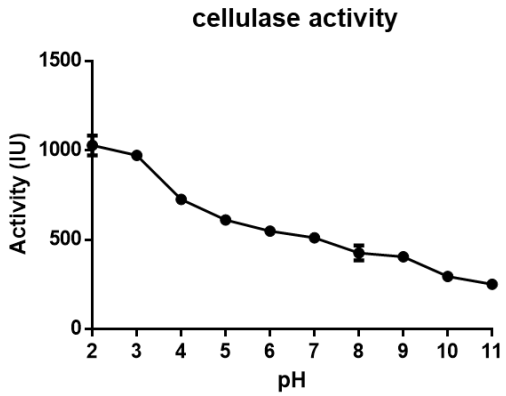


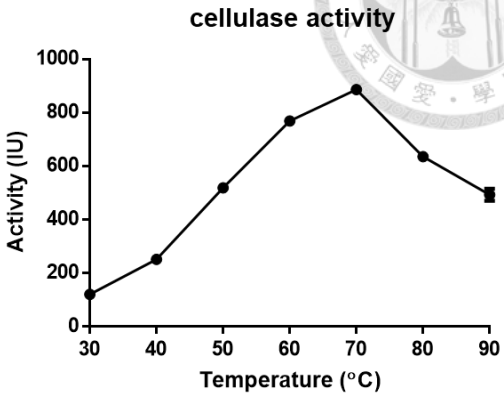
Figure S3. Superimposition of the Alphafold3-predicted *Cel*CCD structure with the *Ct*Cel5E-xylobiose crystal structure (PDB: 4U5I).

Residues from *Cel*CCD and *Ct*Cel5E are labeled in green and cyan, respectively, with xylobiose depicted in gray.

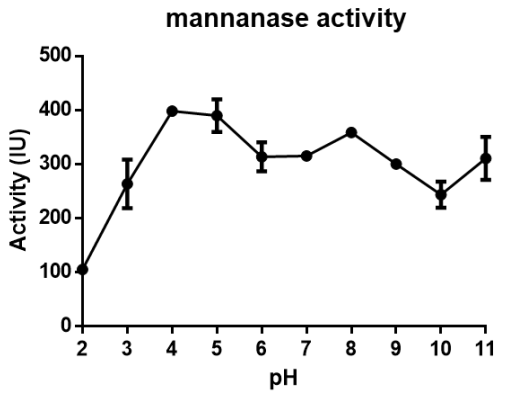
doi:10.6342/NTU202502385

(a)





(b)



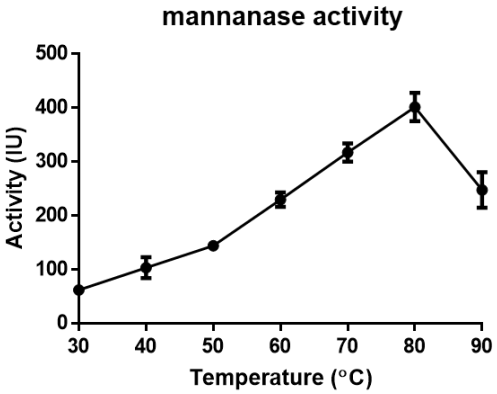


Figure S4. Optimal conditions for *Mt*Glu5 enzymatic activities.

Temperature- and pH-dependent profiles of *Mt*Glu5 cellulase (a) and mannanase (b) activities (IU is defined as the amount of product generated by 1 μmole of the enzyme per minute), respectively, are presented. All experiments were performed in triplicate.

doi:10.6342/NTU202502385

REFERENCES

1. Saleem M. Possibility of utilizing agriculture biomass as a renewable and sustainable future energy source. *Heliyon.* 2022; **8**: e08905. doi: 10.1016/j.heliyon.2022.e08905.
2. Ho DP, Ngo HH, Guo W. A mini review on renewable sources for biofuel. *Bioresour Technol.* 2014; **169**: 742–749. doi: 10.1016/j.biortech.2014.07.022.
3. Zhu JY, Agarwal UP, Ciesielski PN, Himmel ME, Gao R, Deng Y, Morits M, Österberg M. Towards sustainable production and utilization of plant-biomass-based nanomaterials: a review and analysis of recent developments. *Biotechnol Biofuels.* 2021; **14**: 114. doi: 10.1186/s13068-021-01963-5.
4. Babu S, Singh Rathore S, Singh R, Kumar S, Singh VK, Yadav SK, Yadav V, Raj R, Yadav D, Shekhawat K, Ali Wani O. Exploring agricultural waste biomass for energy, food and feed production and pollution mitigation: A review. *Bioresour Technol.* 2022; **360**: 127566. doi: 10.1016/j.biortech.2022.127566.
5. Chandel AK, Garlapati VK, Singh AK, Antunes FAF, da Silva SS. The path forward for lignocellulose biorefineries: bottlenecks, solutions, and perspective on commercialization. *Bioresource Technol.* 2018; **264**: 370–381. doi: 10.1016/j.biortech.2018.06.004.

doi:10.6342/NTU202502385

6. Igwebuike CM, Awad S, Andrès Y. Renewable energy potential: Second-generation biomass as feedstock for bioethanol production. *Molecules.* 2024; **29**: 1619. doi: 10.3390/molecules29071619.

7. Álvarez G, Reyes-Sosa FM, Díez B, Enzymatic hydrolysis of biomass from wood. *Microb Biotechnol.* 2016; **9**: 149–56. doi: 10.1111/1751-7915.12346.

8. Lee HL, Chang CK, Teng KH, Liang PH. Construction and characterization of different fusion proteins between cellulases and β-glucosidase to improve glucose production and thermal stability. *Bioresource Technol.* 2011; **102**: 3973–3976. doi: 10.1016/j.biortech.2010.11.114.

9. Chen CH, Yao Y, Yang B, Lee HL, Yuan SF, Hsieh HY, Liang PH. Engineer multi-functional cellulase/xylanase/β-glucosidase with improved efficacy to degrade rice straw. *Bioresource Technol Rep.* 2019; **5**: 170–177. doi: 10.1016/j.biteb.2019.01.008.

10. Bayer EA, Setter E, Lamed R. Organization and distribution of the cellulosome in Clostridium thermocellum. *J Bacteriol.* 1985; **163**: 552–559. doi: 10.1128/jb.163.2.552-559.

11. Felix CR, Ljungdahl LG. Annu Rev Microbiol. The cellulosome: the exocellular organelle of Clostridium. 1993; **47**: 791–819. doi: 10.1146/annurev.mi.47.100193.004043.

doi:10.6342/NTU202502385

12. Hirano K, Kurosaki M, Nihei S, Hasegawa H, Shinoda S, Haruki M, Hirano N. Enzymatic diversity of the Clostridium thermocellum cellulosome is crucial for the degradation of crystalline cellulose and plant biomass. *Sci Rep.* 2016; **6**: 35709. doi: 10.1038/srep35709.

13. Garvey M, Klose H, Fischer R, Lambertz C, Commandeur U. Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. *Trends Biotechnol.* 2013; **31**: 581–593. doi: 10.1016/j.tibtech.2013.06.006.

14. Lo Leggio L, Larsen S. The 1.62 Å structure of Thermoascus aurantiacus endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Lett.* 2002; **523**: 103–108. doi: 10.1016/s0014-5793(02)02954-x.

15. Aspeborg H, Coutinho PM, Wang Y, Brumer 3rd H, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol.* 2012; **12**: 186. doi: 10.1186/1471-2148-12-186.

16. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol.* 1997; **7**: 637–644. doi: 10.1016/s0959-440x(97)80072-3.

17. Yuan SF, Wu TH, Lee HL, Hsieh HY, Lin WL, Yang B, Chang CK, Li Q, Gao J, Huang CH, Ho MC, Guo RT, Liang PH. Biochemical characterization and structural

doi:10.6342/NTU202502385

analysis of a bifunctional cellulase/xylanase from Clostridium thermocellum. *J Biol Chem.* 2015; **290**: 5739–5748. doi: 10.1074/jbc.M114.604454.

18. Yague E, Beguin P, Aubert JP. Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of Clostridium thermocellum. *Gene* 1990; **89**: 61–67. doi: 10.1016/0378-1119(90)90206-7.

19. Carvalho AL, Goyal A, Prates JA, Bolam DN, Gilbert HJ, Pires VM, Ferreira LM, Planas A, Romao MJ, Fontes CM. The family 11 carbohydrate-binding module of Clostridium thermocellum Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. *J Biol Chem.* 2004; **279**: 34785–34793. doi: 10.1074/jbc.M405867200.

20. Taylor EJ, Goyal A, Guerreiro CI, Prates JA, Money VA, Ferry N, Morland C, Planas A, Macdonald JA, Stick RV, Gilbert HJ, Fontes CM, Davies GJ. How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides: structure and activity of a Clostridium thermocellum lichenase, CtLic26A. *J Biol Chem.* 2005; **280**: 32761–32767. doi: 10.1074/jbc.M506580200.

21. Chhabra SR, Shockley KR, Ward DE, Kelly RM. Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium Thermotoga maritima grown on glucan- and mannan-based polysaccharides. *Appl Environ Microbiol.* 2002; **68**: 545–554. doi: 10.1128/AEM.68.2.545-554.2002.

doi:10.6342/NTU202502385

22. Pereira JH, Chen Z, McAndrew RP, Sapra R, Chhabra SR, Sale KL, Simmons BA, Adams PD. Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic Thermotoga maritima. *J Struct Biol.* 2010; **172**: 372–379. doi: 10.1016/j.jsb.2010.06.018.

23. Wu TH, Huang CH, Ko TP, Lai HL, Ma Y, Chen CC, Cheng YS, Liu JR, Guo RT. Diverse substrate recognition mechanism revealed by Thermotoga maritima Cel5A structures in complex with cellotetraose, cellobiose and mannotriose. *Biochim Biophys Acta.* 2011; **1814**: 1832–1840. doi: 10.1016/j.bbapap.2011.07.020.

24. Chen Z, Friedland GD, Pereira JH, Reveco SA, Chan R, Park JI, Thelen MP, Adams PD, Arkin AP, Keasling JD, Blanch HW, Simmons BA, Sale KL, Chivian D, Chhabra SR. Tracing determinants of dual substrate specificity in glycoside hydrolase family 5. *J Biol Chem.* 2012; **287**: 25335–25343. doi: 10.1074/jbc.M112.362640.

25. Liang PH, Lin WL, Hsieh HY, Lin TY, Chen CH, Tewary SK, Lee HL, Yuan SF, Yang B, Yao JY, Ho MC. A flexible loop for mannan recognition and activity enhancement in a bifunctional glycoside hydrolase family 5. *Biochim Biophys Acta Gen Subj.* 2018; **1862**: 513–521. doi: 10.1016/j.bbagen.2017.11.004.

26. Freier D, Mothershed CP, Wiegel J. Characterization of Clostridium thermocellum JW20. *Appl Environ Microbiol.* 1988; **54**: 204–211. doi: 10.1128/aem.54.1.204-

doi:10.6342/NTU202502385

211.1988.

27. Glasgow EM, Kemna EI, Bingman CA, Ing N, Deng K, Bianchetti CM, Takasuka TE, Northen TR, Fox BG. A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *J Biol Chem.* 2020; **295**: 17752–17769. doi: 10.1074/jbc.RA120.015328.

28. Roberto IC, Mussatto SI, Rodrigues RCLB. Dilute-acid hydrolysis for optimization of xylose recovery from rice straw in a semi-pilot reactor. *Industrial Crops and Products.* 2003; **17**: 171–176.

29. Wang HM, Shih YP, Hu SM, Lo WT, Lin HM, Ding SS, Liao HC, Liang PH. Parallel gene cloning and protein production in multiple expression systems. *Biotechnol Prog.* 2009; **25**: 1582–1586. doi: 10.1002/btpr.274.

30. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 1976; **72**: 248–254. doi: 10.1006/abio.1976.9999.

31. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem.* 1959; **31**: 426–428.

32. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S. The EMBL-EBI Job Dispatcher sequence analysis tools framework in

doi:10.6342/NTU202502385

2024. *Nucleic Acids Res.* 2024; **52**: W521–W525. doi: 10.1093/nar/gkae241.

33. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009; **25**:1189–1191. doi: 10.1093/bioinformatics/btp033.

34. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021; **49**: W293–W296. doi: 10.1093/nar/gkab301.

35. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; **14**: 1188–1190. doi: 10.1101/gr.849004.

36. Goldman AD, Beatty JT, Landweber LF. The TIM barrel architecture facilitated the early evolution of protein-mediated metabolism. *J Mol Evol.* 2016; **82**: 17–26. doi: 10.1007/s00239-015-9722-8.

37. Glasgow EM, Vander Meulen KA, Takasuka TE, Bianchetti CM, Bergeman LF, Deutsch S, Fox BG. Extent and origins of functional diversity in a subfamily of glycoside hydrolases. *J Mol Biol.* 2019; **431**: 1217–1233. doi: 10.1016/j.jmb.2019.01.024.

38. Bianchetti CM, Brum, P, Smith RW, Dyer K, Hura GL, Rutkoski TJ, Phillips Jr GN. Structure, dynamics, and specificity of endoglucanase D from Clostridium cellulovorans. *J Mol Biol.* 2013; **425**: 4267–4285. doi: 10.1016/j.jmb.2013.05.030.

doi:10.6342/NTU202502385

39. Liu J, Tsai C, Liu J, Cheng K, Cheng C. The catalytic domain of a Piromyces rhizinflata cellulase expressed in Escherichia coli was stabilized by the linker peptide of the enzyme. *Enzyme Microb Technol.* 2001; **28**: 582–589. doi: 10.1016/s0141-0229(00)00349-5. PMID: 11339938.

40. Cai S, Li J, Hu FZ, Zhang K, Luo Y, Janto B, Boissy R, Ehrlich G, Dong X. Cellulosilyticum ruminicola, a newly described rumen bacterium that possesses redundant fibrolytic-protein-encoding genes and degrades lignocellulose with multiple carbohydrate-borne fibrolytic enzymes. *Appl Environ Microbiol.* 2010; **76**: 3818–3824. doi: 10.1128/AEM.03124-09.

41. Oyama T, Schmitz GE, Dodd D, Han Y, Burnett A, Nagasawa N, Mackie RI, Nakamura H, Morikawa K, Cann I. Mutational and structural analyses of Caldanaerobius polysaccharolyticus Man5B reveal novel active site residues for family 5 glycoside hydrolases. *PLoS One.* 2013; **8**: e80448. doi: 10.1371/journal.pone.0080448.

42. Garg P, Manoj N. Structure of an iminosugar complex of a glycoside hydrolase family 5 lichenase provides insights into the active site. *Biochimie.* 2023; **204**: 69-77. doi: 10.1016/j.biochi.2022.09.001.

43. Takeda N, Hirasawa K, Uchimura K, Nogi Y, Hatada Y, Akita M, Usami R, Yoshida Y, Grant WD, Ito S, Horikoshi K. Alkaline mannanase from a novel

doi:10.6342/NTU202502385

species of alkaliphilic Bacillus. *J Appl Glycosci.* 2004; **51**: 229–236. doi: 10.5458/jag.51.229.

44. Chauhan, PS, Sharma P, Puri N, Gupta N. (2014). Purification and characterization of an alkali-thermostable β-mannanase from Bacillus nealsonii PN-11 and its application in mannooligosaccharides preparation having prebiotic potential. *Eur Food Res Technol.*, 2014; **238**: 927–936. doi: 10.1007/s00217-014-2170-7.

45. Ma Y, Xue Y, Dou Y, Xu Z, Tao W, Zhou P. Characterization and gene cloning of a novel beta-mannanase from alkaliphilic Bacillus sp. N16-5. *Extremophiles.* 2004; **8**: 447–454. doi: 10.1007/s00792-004-0405-4.

46. Tailford LE, Ducros VM, Flint JE, Roberts SM, Morland C, Zechel DL, Smith N, Bjørnvad ME, Borchert TV, Wilson KS, Davies GJ, Gilbert HJ. Understanding how diverse beta-mannanases recognize heterogeneous substrates. *Biochemistry.* 2009; **48**: 7009-7018. doi: 10.1021/bi900515d.

47. Zheng F, Zhao H, Wang N, Zhong P, Zhou K, Yu S. Cloning and characterization of thermophilic endoglucanase and its application in the transformation of ginsenosides. *AMB Express.* 2022; **12**: 136. doi: 10.1186/s13568-022-01473-z.

48. Ye TJ, Huang KF, Ko TP, Wu SH. Synergic action of an inserted carbohydrate-binding module in a glycoside hydrolase family 5 endoglucanase. *Acta Crystallogr D Struct Biol.* 2022; **78**: 633–646. doi: 10.1107/S2059798322002601.

doi:10.6342/NTU202502385

49. Domínguez R, Souchon H, Lascombe M, Alzari PM. The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. *J Mol Biol.* 1996; **257**: 1042–1051. doi: 10.1006/jmbi.1996.0222.

50. Gloster TM, Ibatullin FM, Macauley K, Eklöf JM, Roberts S, Turkenburg JP, Bjørnvad ME, Jørgensen PL, Danielsen S, Johansen KS, Borchert TV, Wilson KS, Brumer H, Davies GJ. Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *J Biol Chem.* 2007; **282**: 19177–19189. doi: 10.1074/jbc.M700224200.

51. Tseng CW, Ko TP, Guo RT, Huang JW, Wang HC, Huang CH, Cheng YS, Wang AHJ, Liu JR. Substrate binding of a GH5 endoglucanase from the ruminal fungus Piromyces rhizinflata. *Acta Crystallogr.. Sect. F Struct.. Biol. Cryst. Commun.* 2011; **67:** 1189–1194. doi: 10.1107/S1744309111032428.

52. Badieyan S, Bevan DR, Zhang C. Study and design of stability in GH5 cellulases. *Biotechnol Bioeng.* 2012; **109**: 31–34. doi: 10.1002/bit.23280.

53. Dos Santos CR, Cordeiro RL, Wong DW, Murakami MT. Structural basis for xyloglucan specificity and β-d-Xylp(1→6)-D-Glcp recognition at the -1 subsite within the GH5 family. *Biochemistry* 2015; **54**: 1930–1942. doi: 10.1021/acs.biochem.5b00011.

doi:10.6342/NTU202502385

54. Venditto I, Najmudin S, Luis AS, Ferreira LM, Sakka K, Knox JP, Gilbert HJ, Fontes CM. Family 46 carbohydrate-binding modules contribute to the enzymatic hydrolysis of xyloglucan and β-1,3-1,4-glucans through distinct mechanisms. *J Biol Chem.* 2015; **290**: 10572–10586. doi: 10.1074/jbc.M115.637827.

55. McGregor N, Morar M, Fenger TH, Stogios P, Lenfant N, Yin V, Xu X, Evdokimova E, Cui H, Henrissat B, Savchenko A, Brumer H. Structure-function analysis of a mixed-linkage β-glucanase/xyloglucanase from the key ruminal Bacteroidetes Prevotella bryantii B14. *J Biol Chem.* 2016; **291**: 1175–1197. doi: 10.1074/jbc.M115.691659.

56. Meng DD, Liu X, Dong S, Wang YF, Ma XQ, Zhou H, Wang X, Yao LS, Feng Y, Li FL. Structural insights into the substrate specificity of a glycoside hydrolase family 5 lichenase from Caldicellulosiruptorsp. F32. *Biochem J.* 2017; **474**: 3373–338. doi: 10.1042/BCJ20170328.

57. Attia MA, Nelson CE, Offen WA, Jain N, Davies GJ, Gardner JG, Brumer H. In vitro and in vivo characterization of three Cellvibrio japonicus glycoside hydrolase family 5 members reveals potent xyloglucan backbone-cleaving functions. *Biotechnol Biofuels* 2018; **11**: 45. doi: 10.1186/s13068-018-1039-6.

58. Shima S, Igarashi Y & Kodama T (1993) Purification and properties of two truncated endoglucanases produced in *Escherichia coli* harbouring *Clostridium cellulolyticum*

doi:10.6342/NTU202502385

endoglucanase gene celCCD. *Appl Microbiol Biotechnol* **38**, 750–754.

doi:10.6342/NTU202502385

Chapter III

Other studies for identification of inhibitors as potent antiviral or antibacterial agents

中文摘要

2019 年新型冠狀病毒肺炎 （COVID-19）於 2019 年 12 月首次被確認，隨後演變為全球大流行，已導致超過 700 萬人死亡，致死率約為 1%。我們報導並鑑定了兩類化合物，1-（4-（芳乙基羰基）苯基）-4-羧基-2-吡咯烷酮和四氫吡嗪[2，1-a：5,4-a']二異喹啉，作為嚴重急性呼吸綜合症冠狀病毒 2 型 （SARS-CoV-2）的抑制劑，該病毒是 COVID-19 的病原體。我參與的研究包含提供電腦模擬和藥物動力學預測技術，以合理化該系列中最有效的抑製劑與其目標蛋白之結合模式，並評估其作為藥物之特徵。此外，具抗生素耐藥性之細菌的出現，如抗甲氧苯青黴素金黃色葡萄球菌 （MRSA），主要歸因於長期濫用和過度依賴抗生素，已成為當代醫療保健的一個關鍵挑戰。我們設計、合成和評估了一類 MRSA 抑製劑。作為初步研究，我分析了這類 MRSA 抑製劑針對潛在的目標蛋白，催化十一異戊基二烯焦磷酸（UPP）生成以作為細胞壁肽聚糖生合成之脂質載體前驅物的十一異戊基二烯焦磷酸合成酶 （UPPS）之抑制效果，並利用電腦模擬合理化最佳抑制劑的結合模式，以及對查耳酮衍生物進行針對 UPPS 的電腦輔助藥物篩選，為開發潛在抗菌劑提供新的見解。

關鍵詞：新型冠狀病毒肺炎；抗生素；耐藥性；MRSA；查爾酮

doi:10.6342/NTU202502385

Abstract

The coronavirus disease 2019 (COVID-19), initially identified in December 2019, has evolved into a global pandemic, resulting in over 7 million deaths with approximately 1% mortality rate. We identified and characterized two classes of compounds, 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones and tetrahydropyrazino[2,1-a:5,4-a']diisoquinolines, as inhibitors against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the etiological agent of COVID-19. I have engaged in the studies for providing computer modeling and pharmacokinetic prediction technology to rationalize the binding mechanisms and assess the drug-likeness properties of the most potent candidates within these series. Furthermore, the emergence of antibiotic-resistant bacteria, such as methicillin-resistant *Staphylococcus aureus* (MRSA), largely attributed to the prolonged misuse and excessive reliance on antibiotics, has become a critical challenge in contemporary healthcare. We designed, synthesized, and evaluated a class of MRSA inhibitors. As a preliminary study, I assayed the inhibitory activities of the active compounds against the possible target, undecaprenyl diphosphate synthase (UPPS) that catalyzes C55 UPP as a lipid carrier for biosynthesis of the cell wall peptidoglycan, and rationalize the binding mode using computer modeling. Moreover, virtual screening of chalcone derivatives against UPPSs provides valuable insights into developing potential anti-bacterial agents.

doi:10.6342/NTU202502385

KEYWORDS: COVID-19; antibiotics; drug resistance; MRSA; chalcone

doi:10.6342/NTU202502385

ABBREVIATIONS

ADMET, absorption, distribution, metabolism, excretion, and toxicity

BBB, blood brain barrier

Caco2, colorectal adenocarcinoma cell line

CC_{50}, 50% cytotoxicity concentration

cis-PTs, cis-prenyltransferase

CNS, central nervous system

COVID-19, coronavirus disease 2019

CYP, Cytochrome P450

ELISA, enzyme-linked immunosorbent assay

FPP, farnesyl diphosphate

HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIA, human intestinal absorption

IC_{50}, half inhibition concentration

IPP, isopentenyl diphosphate

LB, Luria-Bertani

MESG, 2-amino-6-mercapto-7-methyl-purine ribonucleoside

MIC, minimum inhibitory concentration;

MRSA, Methicillin-resistant *Staphylococcus aureus*

doi:10.6342/NTU202502385

MSSA, Methicillin-sensitive *Staphylococcus aureus*

OCT2, organic cation transporter-2

OPPS, octaprenyl diphosphate synthase

PCR, polymerase chain reaction

PDB, Protein Data Bank

PNP, purine ribonucleoside phosphorylase

PPA, inorganic pyrophosphatase

PRSP, penicillin-resistant *Streptococcus pneumoniae*

*Sa*UPPS, *Staphylococcus aureus* UPPS

TCEP, tris(2-carboxyethyl)phosphine

Tris, tris(hydroxymethyl)aminomethane

UPPS, undecaprenyl diphosphate synthase

VDss, volume of distribution at steady-state

VRE, vancomycin-resistant *Enterococci*

doi:10.6342/NTU202502385

LIST OF TABLES

doi:10.6342/NTU202502385

LIST OF FIGURES

doi:10.6342/NTU202502385

1. INTRODUCTION

1.1 Entry of SARS-CoV-2 into host cells

The coronavirus disease 2019 (COVID-19), initially reported in December 2019 [1, 2, 3], has been declared a global pandemic by the World Health Organization in March 2020 and remains ongoing, with an estimated case mortality rate of approximately 1%. The causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), shares significant genomic similarity (79.6%) with SARS-CoV, which emerged in 2002–2003 [4, 5]. For infection, both viruses utilize human angiotensin-converting enzyme 2 (ACE2) as the receptor for virus entry through recognizing the receptor-binding domain (RBD) located in the S1 domain of the viral Spike glycoprotein [6]. After ACE2 binding, the human protease Furin cleaves the Spike protein at the S1/S2 site (P**R**RA**R**$_{685}$↓S), separating it into S1 and S2 subunits. Subsequently, the human transmembrane protease serine 2 (TMPRSS2) cleaves the S2 at KPS**KR**$_{815}$↓S, exposing the membrane fusion peptide and facilitating the fusion of virus membrane with the host cell membrane for endocytosis, enabling intracellular viral replication [7, 8]. Alternatively, SARS-CoV-2 could enter cells via ACE2-mediated endocytosis, wherein the Spike protein could be cleaved by an endosomal protease Cathepsin L to release the viral RNA into cytosol [9]. Thus, the RBD:ACE2 interaction, as well as the proteases TMPRSS2, Furin, and Cathepsin L, represent critical targets for therapeutic strategies for inhibiting SARS-CoV-

doi:10.6342/NTU202502385

2 entry and infection.

1.2 Discovery of 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones as SARS-CoV-2 entry inhibitors

A series of 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones was rationally designed and synthesized using Michael addition, cyclization, aldol condensation, and deprotonation to target and inhibit TMPRSS2 and Furin, both of which are important in priming the SARS-CoV-2 Spike protein for virus entry. The most potent compound, **2f**, demonstrated efficient inhibition of SARS-CoV-2 Delta and Omicron variants replication in VeroE6 and Calu-3 cells, with EC_{50} values ranging from 0.001 to 0.026 μM when pre-incubated with the virus to block the entry. Further investigation revealed that the antiviral activities of the synthesized compounds exceeded their proteases inhibitory activities, identifying the primary mechanism as inhibition of the Spike RBD interaction with ACE2, with their antiviral efficacy synergistically enhanced by concurrent inhibition of TMPRSS2 and/or Furin. In a pseudovirus entry assay, compound **2f** effectively blocked SARS-CoV-2 Spike pseudovirus entry in an ACE2-dependent manner, primarily through interrupting RBD:ACE2 interaction and inhibiting TMPRSS2 in Calu-3 cells. Lastly, in the in vivo hamster model of SARS-CoV-2 infection, the oral administration of 25 mg/kg of compound **2f** significantly reduced body weight loss and achieved a 5-fold reduction in viral RNA levels in nasal turbinate three days post-

doi:10.6342/NTU202502385

infection. These findings demonstrated the potential of **2f** as a lead compound for further preclinical development as a therapeutic agent against SARS-CoV-2.

1.3 Discovery of tetrahydropyrazino[2,1-a:5,4-a']diisoquinolines as SARS-CoV-2 inhibitors

A class of tetrahydropyrazino[2,1-a:5,4-a']diisoquinoline derivatives was synthesized under an environmentally friendly condition using water as the reaction solvent. The three-dimensional (3D) structures of several synthesized compounds were elucidated via X-ray diffraction analysis. Inspired by the significant antiviral activities of naturally occurring isoquinoline alkaloids against a broad range of viruses, including coronaviruses, the synthesized derivatives were evaluated for their inhibitory activity against SARS-CoV-2. Our results demonstrated that compounds **50** exhibited potent inhibition of the Delta SARS-CoV-2 entry into VeroE6 cells, with EC_{50} values of 26.5 ± 6.9 μM, and CC_{50} values exceeding 100 μM. The antiviral mechanism involved disrupting the interaction between the SARS-CoV-2 Spike RBD and the human ACE2. This study introduces a sustainable synthesis approach for tetrahydropyrazinodiisoquinoline derivatives, highlighting their potential for antiviral and other biomedical applications.

1.4 Threats of multiple drug-resistant pathogens

Mortality resulting from the failure of treatment for infectious diseases caused by drug-resistant bacterial strains has been escalating [10, 11]. Prominent multidrug-resistant

doi:10.6342/NTU202502385

pathogens include methicillin-resistant *Staphylococcus aureus* (MRSA), penicillin-resistant *Streptococcus pneumoniae* (PRSP), and vancomycin-resistant *Enterococci* (VRE), among others. The dearth of novel and effective antibiotics has exacerbated mortality rates associated with these pathogens, presenting a significant global public health crisis [12]. Despite intensive research efforts, no new major classes of antibiotics were discovered between 1962 and 2000, underscoring the stagnation in antimicrobial innovation during this period [13]. Since 2017, 12 novel antibacterial agents have received approval globally, yet vaborbactam is the sole representative of a novel antibacterial class [14]. This highlights the urgent necessity for the development of effective antibiotics targeting drug-resistant bacteria. MRSA has emerged as a particularly critical threat due to its resistance to key therapeutic antibiotics, including methicillin and vancomycin, which are traditionally used in clinical treatments [15]. As an opportunistic pathogen, MRSA is frequently associated with nosocomial infections and can precipitate severe complications such as septicemias. Reflecting its public health significance, the World Health Organization (WHO) classified MRSA as one of the high-priority pathogens in 2017, categorizing it among the most critical bacterial threats to human health [16].

1.5 Role of UPPS for bacterial cell wall biosynthesis

In bacteria, undecaprenyl diphosphate synthase (UPPS) is a cis-prenyltransferase (cis-

doi:10.6342/NTU202502385

PTs) that catalyzes the consecutive condensation of eight C5 isopentenyl diphosphate (IPP) units with one C15 farnesyl diphosphate (FPP) to synthesize C55 undecaprenyl diphosphate (UPP), a polyisoprenoid lipid that plays an essential role as a glycosyl carrier in the biosynthesis of peptidoglycan [17], a fundamental component of bacterial cell wall structure and integrity. Therefore, the inhibition of UPPS disrupts cell wall biosynthesis, effectively impeding the growth of both Gram-positive and Gram-negative bacteria, highlighting its potential as a broad-spectrum antibacterial strategy. Notably, several UPPS inhibitors have been previously identified as effective agents for inhibiting bacterial growth [18-23].

1.6 Specific aims of this study

We previously identified a class of 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones as multi-targeting inhibitors of the RBD:ACE2 interaction for anchoring the virus, as well as the human proteases Furin and TMPRSS2, which play crucial roles in priming the Spike protein of SARS-CoV-2 for viral entry, thereby demonstrating antiviral activity [24]. In addition, we previously identified a class of tetrahydropyrazino[2,1-a:5,4-a']diisoquinolines as inhibitors of the RBD:ACE2 interaction to block the virus entry [25]. We aimed to provide plausible binding modes and pharmacokinetics of the most potent inhibitors to support them as drug candidates. As reported in this thesis, I conducted computer modeling to rationalize the binding

doi:10.6342/NTU202502385

modes of the most potent inhibitors as well as analyze their pharmacokinetics properties. Furthermore, a class of anti-MRSA compounds have been synthesized and evaluated. We aimed to understand the mechanism for their antibacterial activity and identify more antibacterial compounds by virtual screening. I therefore performed assays to evaluate the inhibitory activities of the potent compounds against a possible target UPPS and rationalized the binding mode of the most potent inhibitor by using computer modeling. Furthermore, I virtually screened out chalcone derivatives as potential UPPS inhibitors.

2. MATERIALS AND METHODS

2.1 Molecular docking for 1-(4-(3-(1H-indol-4-yl)acryloyl)phenyl)-4-carboxy-2-pyrrolidinone (2f) and a tetrahydropyrazino[2,1-a:5,4-a']diisoquinoline (50)

The molecular docking was performed using the iGEMDOCK software [26] to predict how compound **2f** binds with TMPRSS2, Furin, and RBD of SARS-CoV-2 delta variant Spike protein. The 3D structures of TMPRSS2 (PDB: 7MEQ) [27], Furin (PDB: 5JXG) [28], and RBD (PDB: 7W92) [29] were retrieved from the RCSB protein data bank (PDB, https://www.rcsb.org/). All water molecules and bound ligands were removed from these structures prior to docking. The binding cavity of TMPRSS2 were extracted by selecting the 12 Å residues around the catalytic triad (Ser441-His296-Asp345), and the binding cavity of Furin was extracted by selecting the 12 Å residues around the catalytic residues, Asp153, His194, Asn295, and Ser368. RBD domain (residues 319 to

doi:10.6342/NTU202502385

541) was extracted from the open state of SARS-CoV-2 delta variant Spike protein. The cavities were then prepared by defining the residue atom types and charges assignment via iGEMDOCK method. The 3D structures of compounds **2f** and **50** were generated from Molview website (https://molview.org/), and the structural information was transformed into mol2 format by using Open Babel GUI software [30].

For molecular docking process, docking accuracy settings (GA parameters including population size: 800, generations: 80, and number of solutions: 10) were chosen, and iGEMDOCK software was used to generate protein-ligand interaction profiles of Electrostatic (E), Hydrogen-Bonding (H), and Van der Waals interaction (V). Once docking was finished, iGEMDOCK was used to analyze and rank all docked poses based on the estimated binding energy. The scores of the docked poses were estimated as the total energy of Electrostatic (E) + Hydrogen-Bonding (H) + Van der Waals (V) in the docking site. Each docked pose with the lowest energy was considered as the best pose for compound **2f** or **50** against the target cavity.

2.2 Drug-likeness analysis

The 3D structures of the potent inhibitors including compounds **2f**, **50**, **87**, and **94** were converted into SMILES format using Open Babel GUI software [30]. To evaluate the drug-likeness of these compounds, the Lipinski Rule of Five parameters were assessed using the SCFBio website (http://www.scfbio-iitd.res.in/software/drugdesign/lipinski.jsp)

doi:10.6342/NTU202502385

[31, 32]. Additionally, the ADMET profile of these compounds was predicted using the pkCSM platform (http://biosig.unimelb.edu.au/pkcsm/) [33].

2.3 Expression and purification of recombinant EcUPPS and SaUPPS

The genes encoding the full-length of SaUPPS and EcUPPS were synthesized by Bio Basic Inc. (Canada), and subsequently respectively subcloned into pET32a vector for expressing the recombinant protein. pET-32a-SaUPPS features N-terminal Thioredoxin (TRX), hexa-His tag (His_6), and TEV protease cutting site (TEVp), while pET-32a-EcUPPS contains N-terminal TRX, His_6, and Factor Xa (FXa) cutting site. The *E. coli* BL21 competent cells were transformed with either TRX-His_6-FXa-EcUPPS or TRX-His_6-TEVp-SaUPPS construct and plated on Luria-Bertani (LB)-agar containing 100 mg/L ampicillin for selection. A single colony was selected and inoculated into 20 mL of LB medium with constant shaking at 200–250 rpm overnight. The overnight culture was used to inoculate 1 liter of LB medium supplemented with 100 mg/L ampicillin with constant shaking at 200–250 rpm at 37°C until the cell density reached OD_{600} of 0.6–0.8, followed by adding 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) with constant shaking at 180 rpm at 16°C for 16–20 h to induce the protein expression. The cells were harvested by centrifugation (6500 rpm for 20 min), resuspended in buffer A (10 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 2 mM 2-mercaptoethanol), and disrupted by a French-press instrument (Constant Cell Disruption System) at 20,000 psi. Cell lysate was

doi:10.6342/NTU202502385

centrifuged at 16000 rpm for 30 min at 4°C to remove debris, and the supernatant was loaded onto an open column with 5 mL Ni-NTA resin. After washing with 100 mL of buffer A with 25 mM imidazole, the target protein was eluted with 25 mL of buffer A containg 250 mM imidazole, followed by dialysis against 5 L buffer A to remove imidazole, and the TRX-His_6-tag was cleaved by TEV protease or Factor Xa at 4℃ overnight. The overnight mixture was subsequently applied into another open column containing 5 mL Ni-NTA resin, and the untagged SaUPPS or EcUPPS was eluted in the buffer A without imidazole.

2.4 Inhibition Assay against EcUPPS and SaUPPS

The dose-dependent inhibition against EcUPPS or SaUPPS was assayed using EnzChek™ Pyrophosphate assay kit (Thermofisher, MA, USA). The enzymatic activities were assayed in buffer containing 100 mM Hepes, pH 7.5, 50 mM KCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 0.2 mM 2-amino-6-mercapto-7-methylpurine ribonucleoside, 1 U/mL purine ribonucleoside phosphorylase, 0.03 U/mL inorganic pyrophosphatase, 0.1 μM of purified EcUPPS or 0.037 μM of purified SaUPPS, 0.5 μM FPP, 10 μM IPP, and various concentrations of inhibitors at 25 ℃. IC_{50} value was determined by fitting the concentration-dependent SaUPPS inhibition curves with the equation A(I) = A(0) x {1-[I/(I + IC_{50})]} using GraphPad Prism software (v.9.4.0). In this equation, A(I) represents the enzyme activity with inhibitor concentration I, I is the concentration of the inhibitor,

doi:10.6342/NTU202502385

and A(0) is the enzyme activity in the absence of the inhibitor. All measurements were performed in triplicate to generate an averaged IC_{50} and a standard deviation for each inhibitor.

2.5 Molecular docking of the UPPS inhibitors

To predict the binding interactions of chalcone derivatives against UPPSs, molecular docking analysis was carried out using iGEMDOCK [26] and Discovery Studio software [34]. The 3D structures of EcUPPS (PDB: 1V7U) [35] and SaUPPS (PDB: 4H8E) [36] were obtained from the RCSB Protein Data Bank (PDB, https://www.rcsb.org/). Prior to docking, water molecules and bound ligands were removed from the structure. The binding cavity was defined as residues within an 8 Å radius around the FPP binding site, followed by specifying residue atom types and assigning charges using the iGEMDOCK methods. The 3D structures of 1130 chalcone-like compounds retrieved from PubChem database (https://pubchem.ncbi.nlm.nih.gov/), along with in-house synthesized chalcones were generated using the Molview website (https://molview.org/), and the structural data were subsequently converted into mol2 format utilizing the Open Babel GUI software [30].

For the molecular docking process using iGEMDOCK, docking accuracy settings were selected with generic evolutionary method (GA) parameters set to a population size of 800, 80 generations, and 10 solutions. The iGEMDOCK software was employed to

doi:10.6342/NTU202502385

generate protein-ligand interaction profiles, including Electrostatic (E), Hydrogen-Bonding (H), and Van der Waals (V) interactions. Following docking, the software analyzed and ranked all docked poses based on estimated binding energies. The docking scores were calculated as the sum of the total energy contributions from Electrostatic (E), Hydrogen-Bonding (H), and Van der Waals (V) interactions within the binding site. The pose with the lowest energy was identified as the optimal binding conformation for the compound against EcUPPS or SaUPPS.

2.6 Virtual screening of chalcones against UPPSs

To identify potential chalcone inhibitors of UPPSs, Discovery Studio, which employs an alternative algorithm for molecular docking and virtual screening, was utilized. The binding cavities of EcUPPS and SaUPPS, along with the structures of 1,130 chalcone compounds, were energy-minimized and prepared using the standard protocols of Discovery Studio. The LibDock program was initially used to screen for potential compounds, selecting the top 10% (113 compounds) based on the LibDock scores. These selected compounds were subsequently subjected to a second round of docking using the CDOCKER program to generate more precise binding poses. Finally, the top 10 chalcones with the highest CDOCKER scores were identified as the most promising candidates.

2.7 Pharmacophore anchors generation

doi:10.6342/NTU202502385

After molecular docking with the iGEMDOCK program, all docked poses were analyzed to elucidate the interactions between protein residues and compounds. To generate site-moiety maps, the SiMMap platform (http://simfam.life.nctu.edu.tw/) was employed [37]. Site-moiety mapping identified several pharmacophore anchors, representing key physical-chemical properties of the binding environment elements critical for biological functions. These anchors reflect specific geometric and chemical features of the binding pocket that facilitate interactions with preferred moieties. Site-moiety maps of 1,130 chalcones derived from docking poses were constructed, and pharmacophore anchors were analyzed to characterize patterns of protein-ligand interactions.

3. RESULTS

3.1 Binding modes of 2f with TMPRSS2, Furin, and RBD

To rationalize the structure-activity relationship of the antivirals on inhibiting TMPRSS2, a docking study was conducted with the clues that the carboxylate on the pyrrolidinone ring and the indole ring on the other side are important for **2f**-mediated inhibition. As shown in Figure 1a, the best inhibitor **2f** was docked into the active site of TMPRSS2. There are three major binding interactions contributed by the carboxylate of pyrrolidinone, the carbonyl group, and the indole ring, respectively. The carboxylate on the pyrrolidinone ring forms two hydrogen bonds with the hydroxyl group of Thr393.

doi:10.6342/NTU202502385

This explains why the compound showed a significantly weaker inhibitory activity against TMPRSS2 after removing the carboxylate. The carbonyl group forms a hydrogen bond with the side-chain N atom of His296. Moreover, the indole ring forms two hydrogen bonds with the backbone carbonyl oxygen of Gly462 and side-chain oxygen of Ser436.

Docking of **2f** into Furin also revealed importance of the carboxylate on inhibition. As shown in Figure 1b, **2f** was docked against the Furin protease (PDB: 5JXG). According to the docking result, the terminal carboxyl group of **2f** has strong interactions with Furin, including an electrostatic interaction with the guanidine group of Arg197 and a hydrogen bonding with Arg193 and His364. Besides, the nitrogen of the indole ring also forms a hydrogen bond with Pro256, and the oxygen of the ketone group forms a hydrogen bond with the nitrogen of the main chain of Ser368.

Since our thermal shift experiments showed RBD as the target, computer docking was performed to understand the binding mode of compound **2f** in RBD. As shown in Figure 1c, **2f** was docked against the RBD of SARS-CoV-2 delta variant Spike protein (PDB: 7W92). Compound **2f** has both hydrogen bonding and van der Waal's interaction with Gln493 of RBD, which plays an important role on contacting with Lys31 of ACE2. Besides, the oxygen on the pyrrolidinone group also interacts with Gly496 of RBD through hydrogen bonding. In the delta RBD:ACE2 complex, Lys353 of ACE2 forms a

doi:10.6342/NTU202502385

hydrogen bond with Gly496 of RBD. Therefore, by forming interactions with the key residues of RBD, compound **2f** might disrupt the contact between ACE2 and RBD.

3.2 Drug-likeness analysis of 2f as judged from Lipinski rule of five and ADMET properties

Lipinski's rule was developed to determine the drug-ability of compounds for pharmaceutical development. For preparation of compound **2f**'s 3D structure, Molview website (https://molview.org/) was utilized, and the structure information was converted into a SMILES format by Open Babel GUI software [30]. To assess the drug likeness of compound **2f**, the online tool (http://www.scfbio-iitd.res.in/software/drugdesign/lipinski.jsp) was used for Lipinski rule of five estimation [31, 32]. As shown in Table 1, compound **2f** was accepted based on Lipinski rule of five in parameters including molecular weight (less than 500 Da), logP (less than 5), hydrogen bond acceptors (less than 10) and donors (less than 5), and molar refractivity (between 40 to 130), indicating that compound **2f** is likely to be a candidate for drug development.

ADMET properties predicted in silico could be used to estimate the possibility of compounds to be used as therapeutic agents. There are 5 parameters involved in the ADMET prediction, including absorption, distribution, metabolism, excretion and toxicity. The ADMET profile of compound **2f** was estimated by pkcsm website (http://biosig.unimelb.edu.au/pkcsm/) [33] and listed in Table 2. Its absorption was

doi:10.6342/NTU202502385

portrayed by human colorectal adenocarcinoma cell line (Caco2) permeability and human intestinal absorption (HIA). The Caco2 permeability and the HIA scores are relatively high, indicating compound **2f** could be absorbed into the human intestine. Analysis of distribution progress, the blood brain barrier (BBB) and central nervous system (CNS) permeability of compound **2f** indicates that it is unable to penetrate BBB and CNS, and the volume of distribution at steady-state (VDss) implicates that compound **2f** has a higher possibility to distribute in plasma rather than in tissue. In human body, Cytochrome P450 (CYP) enzymes are responsible for the metabolism of drugs, and inhibitors of these CYP enzymes can significantly influence the effect of pharmaceuticals. The predicted results show that compound **2f** might be a substrate of CYP3A4 (an inhibitor of CYP3A4 might be co-administrated with **2f** to increase its half-life), but not other CYP enzymes. For excretion prediction, compound **2f** is not likely to be a renal organic cation transporter-2 (OCT2) substrate, showing there is no potential contraindication. In the toxicity estimation, compound **2f** is not a mutagenic drug and a hERG I/II inhibitor, and it does not have acute toxicity and skin sensitization. However, compound **2f** is predicted to possess hepatotoxicity.

3.3 Binding Mode of the tetrahydropyrazino[2,1-a:5,4-a']diisoquinoline Inhibitor 50 with RBD

doi:10.6342/NTU202502385

To elucidate the inhibition mechanism of compound **50** in virus entry, a docking study was conducted. **50** was docked into RBD of the SARS-CoV-2 delta variant Spike protein (PDB: 7W92), as shown in Figure 2. The binding energy of compound **50** with RBD of the delta-strain Spike protein was estimated to be -93.3 kcal/mol. The analysis revealed that compound **50** forms van der Waals interactions with Tyr448, Tyr453, Gln493 and Tyr505 of the RBD. In the delta RBD:ACE2 complex, Tyr453, Gln493 and Tyr505 of RBD form hydrogen bonds with His34, Lys31, and Glu37 of ACE2, respectively. Additionally, an oxygen atom in one of the acetophenone groups forms a hydrogen bond with Gly496 of the RBD, which also forms a hydrogen bond with Lys353 of ACE2 in the delta RBD:ACE2 complex. Consequently, through interactions with these pivotal residues on the RBD, compound **50** potentially disrupt the RBD:ACE2.

3.4 Drug-likeness analysis of compound 50 as judged from Lipinski rule of five and ADMET properties

According to the Lipinski rule of five, compound **50** fits several key criteria: its log P value is 5, the number of hydrogen bond acceptors is 4, fewer than 10, and the number of hydrogen bond donors is 0, fewer than 5. The molecular weight of compound 50 is 494 Da, which is smaller than the ideal limit of 500 Da. However, its molar refractivity is 146.607559, which is relatively surpassing the standard (between 40 and 130). These

doi:10.6342/NTU202502385

parameters suggested that compound **50** could be a promising candidate for drug development (Table 3).

Moreover, the ADMET properties predicted in silico provided valuable insights into the potential of compounds for therapeutic use. These properties encompass absorption, distribution, metabolism, excretion, and toxicity, as summarized in Table 4 for compound **50**. For absorption, the Caco-2 permeability and human intestinal absorption (HIA) scores for compound **50** are relatively high, suggesting effective absorption in the human intestine. Regarding distribution, the ability of compound **50** to cross the blood-brain barrier (BBB) and penetrate the central nervous system (CNS), was supported by the predicted permeability values. Additionally, the volume of distribution at steady state (VDss) indicated a higher likelihood of tissue distribution. Metabolism prediction revealed that compound **50** may act as a substrate for CYP2D6 or CYP3A4. However, it was not predicted to inhibit CYP1A2, CYP2C19, CYP2C9, CYP2D6, or CYP3A4. In terms of excretion, compound **50** was unlikely to be a substrate for renal organic cation transporter-2 (OCT2), suggesting minimal risk of contraindications related to renal transport. Toxicity predictions indicated that compound **50** is not mutagenic, does not inhibit the hERG I channel, and lacks a skin sensitization potential. Furthermore, the predicted maximum recommended tolerated dose (MRTD) of compound **50** exceeds the standard values, suggesting a favorable safety profile with no unacceptable side effects.

doi:10.6342/NTU202502385

Nonetheless, there was a potential concern for hepatotoxicity associated with compound **50**.

3.5 Evaluations of the inhibitors against UPPSs

Several compounds were synthesized by Hu-Jia Chang in our laboratory for anti-bacterial drug development. To investigate the potential inhibition of these compounds against UPPSs, EcUPPS and SaUPPS were expressed and purified, evaluating these chalcone-like compounds against UPPSs utilizing EnzChek^{TM} Pyrophosphate assay kit. The standard curve for the enzymatic assay is presented in Figure 3a, while SDS-PAGE analysis of UPPSs along with their respective Michaelis-Menten plots are demonstrated in Figure 3b-d. The kinetic parameters of EcUPPS and SaUPPS, summarized in Table 5, demonstrate comparable enzymatic activities to those reported in the literature [38].

The compounds are classified into three groups according to the number of aromatic moieties they contain. As demonstrated in Table 6 and Figure 4, compounds featuring three aromatic rings in the first category including **34** with a bromobenzene, **35** with a chlorobenzene, **48** with a tert-butyl benzene, and **58** containing a benzene ring exhibited no inhibition activity against both EcUPPS and SaUPPS, while **72** with a isopropyl benzene demonstrated weak inhibition against both SaUPPS (IC_{50} = 15.34 ± 1.026 μM) and EcUPPS (IC_{50} = 37.92 ± 0.737 μM), and **98** featuring a di-bromo benzene displayed no inhibition against EcUPPS but had minor inhibition against SaUPPS (IC_{50} = 20.79 ±

doi:10.6342/NTU202502385

1.380 μM). In the second category with four aromatic rings, compounds **65** incorporating a 1,3-benzothiazole and **83** featuring a 2-naphthalene demonstrated no inhibition activity against UPPSs. On the contrary, compounds including **52** featuring a diphenyl structure and **84** containing a 1-naphthalene showed slight inhibition against SaUPPS (IC_{50} = 21.99 ± 1.821 and 19.24 ± 2.730 μM, respectively) and EcUPPS (IC_{50} = 17.68 ± 0.982 and 23.89 ± 2.464 μM, respectively), whereas compound **79** featuring a di-chloro phenyl benzene structure demonstrated an excellent inhibition against SaUPPS (IC_{50} = 3.18 ± 0.182 μM) and a weak activity against EcUPPS (IC_{50} = 13.94 ± 1.426 μM). Specifically, compound **97** with a benzyl phenyl ether group features great MIC values did not inhibit UPPSs, indicating it might target other proteins in vivo. Compounds in the third category overall exhibited great inhibitions against UPPSs. Compound **85** with an anthracene showed inhibition activities against SaUPPS (IC_{50} = 12.77 ± 1.773 μM) and EcUPPS (IC_{50} = 29.33 ± 1.107 μM), while compounds including **87** with a benzyl carbazole, **88** containing a benzyl indole group, **94** featuring di-benzyl phenyl ether groups, and **96** with a different benzyl carbazole, exhibited potent inhibitions against SaUPPS (IC_{50} = 5.43 ± 0.521, 8.89 ± 1.466, 6.10 ± 0.652, and 4.92 ± 0.278 μM, respectively) and EcUPPS (IC_{50} = 5.24 ± 0.706, 17.25 ± 0.878, 10.43 ± 0.911, 9.48 ± 0.578 μM, respectively).

To predict the inhibition mechanism of these compounds against UPPSs, molecular docking analysis was performed. 3D structures of these compounds were generated via

doi:10.6342/NTU202502385

Molview website (https://molview.org/), and the structural data were subsequently converted into mol2 format by the Open Babel GUI software [30]. The 3D structures of EcUPPS (PDB: 1V7U) [35] and SaUPPS (PDB: 4H8E) [36] were obtained from the RCSB Protein Data Bank (PDB, https://www.rcsb.org/), and the binding pockets were extracted from 8 Å of FPP binding site. Molecular docking was performed using iGEMDOCK software under accurate docking parameters. As shown in Table 7, the binding energies of these compounds in SaUPPS are lower than those in EcUPPS, reflecting a stronger binding preference for SaUPPS, consistent with the results of the inhibition assay. Among these derivatives, the most potent compounds **87** and **94** exhibit great binding affinities to the FPP binding site of SaUPPS (binding energy = **-**184.46 and **-**230.743 kcal/mol, respectively), aligning well with their excellent IC_{50} values.

3.6 Drug-Likeness of 87 and 94 as Judged from Lipinski Rule of Five and ADMET Properties

Lipinski Rule of Five provides a framework for evaluating the potential of a compound as an orally bioavailable drug candidate. The 3D structures of compound **87** and **94**, identified as the most potent anti-MRSA candidates, were prepared using the MolView platform (https://molview.org/). Structural data were converted into SMILES format using Open Babel GUI software [30]. The drug-likeness of compounds **87** and **94** was assessed using an online application (http://www.scfbio-

doi:10.6342/NTU202502385

iitd.res.in/software/drugdesign/lipinski.jsp), in accordance with the Lipinski Rule of Five criteria, which include molecular weight (<500 Da), logP (<5), hydrogen bond acceptors (<10), hydrogen bond donors (<5), and molar refractivity (40–130) [31, 32]. As summarized in Table 8, compound **87** slightly exceeded standard values for molecular weight and molar refractivity, while compound **94** marginally surpassed thresholds for molecular weight, LogP, and molar refractivity. These deviations, however, remain within acceptable limits, supporting their potential as drug development candidates.

The therapeutic potential of chemicals can be evaluated through five key parameters: absorption, distribution, metabolism, excretion, and toxicity (ADMET), predicted *in silico*. The ADMET profiles of compounds **87** and **94** were estimated using the pkCSM platform (http://biosig.unimelb.edu.au/pkcsm/) [33], and the results are summarized in Table 9. Absorption was assessed through human epithelial colorectal adenocarcinoma cell line (Caco-2) permeability, human intestinal absorption (HIA), skin permeability, and the likelihood of being a substrate or inhibitor of P-glycoprotein. Both compounds showed relatively low skin permeability values, indicating their possibilities of skin absorption, and high HIA scores, suggesting potential intestinal absorption. Besides, P-glycoprotein, an ATP-binding cassette transporter, plays a crucial role in recognizing a broad spectrum of xenobiotics and facilitating their efflux from cells. Compound **94** is neither a substrate nor an inhibitor of P-glycoprotein, implying it would not be effluxed

doi:10.6342/NTU202502385

by this transporter. On the contrary, compound **87** might be the substrate of P-glycoprotein. For distribution, both compounds were predicted to penetrate the blood-brain barrier (BBB) and central nervous system (CNS), and the steady-state volume of distribution (VDss) values indicate that both compounds are more likely to distribute in tissue than in plasma. Furthermore, metabolism predictions indicate that both compounds may act as substrates for CYP3A4 and inhibitors of CYP2C9, suggesting potential interactions with co-administered CYP inhibitors to enhance stability. Moreover, pharmaceuticals are metabolized in human body by Cytochrome P450 (CYP) enzymes, with inhibitors of these enzymes potentially altering drug efficacy significantly. Predictions indicate that compounds **87** and **94** may act as substrates for CYP3A4, suggesting that co-administration with a CYP3A4 inhibitor could enhance their stability, and they are unlikely to be metabolized by other CYP enzymes. Additionally, compound **94** is a potential inhibitor of CYP2C9. Excretion predictions indicate that both **87** and **94** are unlikely to act as a substrate for renal organic cation transporter-2 (OCT2), minimizing potential contraindications. Lastly, toxicity assessments reveal that neither compound **87** nor **94** is mutagenic, a hERG I/II inhibitor, or associated with skin sensitization. However, compound **94** might possess hepatotoxicity, but both compounds exhibit relatively high maximum tolerated doses.

3.7 Virtual screening of chalcone-like compounds against EcUPPS and SaUPPS

doi:10.6342/NTU202502385

Chalcones (IUPAC name: (E)-1,3-diphenylprop-2-en-1-one) are a class of α, β-unsaturated ketones involved in plant flavonoids and isoflavonoids biosynthesis [39, 40]. Due to their characteristics of convenient synthesis and diverse bioactivities, chalcone and its analogs were historically deemed excellent candidates for the pharmaceutical developments [41]. Numerous synthetic chalcone derivatives have been developed in chemical laboratories, exhibiting diverse biological activities [42]. These chalcones, particularly those with specific functional groups, demonstrate a broad spectrum of activities, including anti-microbial, [43-47] anti-malarial [48-50], anti-cancer [51-53], anti-inflammatory [54, 55], anti-protozoal [56, 57], anti-HIV [58], anti-oxidant [59], and anti-ulcer [60] properties.

To further explore the potential of chalcone-like compounds against EcUPPS and SaUPPS, 3D structures of 1,130 chalcone derivatives were retrieved from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/). These structures were converted into the mol2 format using the Open Babel GUI software [30] and subsequently docked into the binding cavities of EcUPPS and SaUPPS. The workflow of the virtual screening process is illustrated in Figure 5a. Molecular docking was performed using iGEMDOCK software, and binding preferences were statistically analyzed using SiMMap website (http://simfam.life.nctu.edu.tw/) [37]. Pharmacophore anchor models were constructed

doi:10.6342/NTU202502385

based on electrostatic (E), hydrogen-bonding (H), and van der Waals (V) interactions between binding site residues and the ligands.

The pharmacophore anchor model of EcUPPS identifies two H anchors and four V anchors near the FPP binding site. As shown in Figure 5b, H1 anchor is composed of Gly27 and Asn28, while H2 anchor is formed by Asn74. The V anchors are distributed throughout the binding pocket, with V1 involving Met65, Tyr68, and Ala69; V2 associated with Asp25, Asn28, and Arg39; V3 comprising Ala69 and Phe89; and V4 including Asn28, Val50, and Leu88. In the EcUPPS structure, key residues such as Asp26, Gly27, Asn28, Arg30, and Arg39 form critical hydrogen bonds with FPP at the S1 binding site. At the S2 binding site, Arg194, Glu198, and Arg200 establish hydrogen bonds with the pyrophosphate moiety of IPP. Docking analyses indicate that chalcone derivatives predominantly occupy the S1 site, interacting with residues including Gly27, Asn28, Arg30, and Arg39 through hydrogen bonding and hydrophobic interactions. Furthermore, the analysis of functional moieties preferred by these anchors provides valuable insights for designing improved inhibitors. In the H anchors, the H1 anchor favors ketone groups attached to benzyl rings of the chalcone scaffold, while the H2 anchor exhibits a preference for hydroxyl group. Across all V anchors, aromatic moieties are strongly preferred, suggesting that chalcone derivatives engage these spatially diverse positions, often interacting with two out of the four V anchors through two benzene rings. To

doi:10.6342/NTU202502385

enhance binding affinity against EcUPPS, structural modifications to the chalcone scaffold, such as adding benzyl functional groups capable of engaging all four V anchors, are proposed. These modifications could optimize interactions within the binding pocket, thereby improving the inhibitory potency of the compounds.

In contrast, the model of SaUPPS reveals the presence of one E anchor, one H anchor, and two V anchors, as illustrated in Figure 5c. The E1 anchor is formed by Asp33 and Arg84, while the H1 anchor comprises Asn35 and His50. The V anchors are distributed as follows: V1 includes Asn35, His50, and Ala76, while V2 consists of Pro96, Phe99, and Phe148. Within the SaUPPS structure, residues such as Asp33, Gly34, Gly36, Arg37, Arg46, and Arg84 play crucial roles in forming hydrogen bonds with FPP at the S1 binding site. At the S2 binding site, Arg201, Arg207, and Ser229 establish hydrogen bonds with the pyrophosphate moiety of IPP. Docking analyses predict that chalcone derivatives primarily occupy the S1 site, engaging in electrostatic interactions with Asp33 and Arg84, hydrogen bonding with Asn35 and His50, and hydrophobic interactions with residues involved in FPP binding. Specific moiety preferences within the binding site anchors were also identified. The carboxyl group of chalcone derivatives is favored in the E1 anchor, the ketone group on one of the benzyl rings of the chalcone scaffold is preferred in the H1 anchor, and the two aromatic moieties are highly favorable for hydrophobic interactions in both V1 and V2 anchors.

doi:10.6342/NTU202502385

Based on binding affinities and SiMMap analyses, a subset of chalcone derivatives has been identified with specific inhibitory potential (Figure 5d). Nine chalcones were predicted to selectively inhibit SaUPPS, while seven were proposed as potential inhibitors for EcUPPS. Additionally, five chalcones were predicted to exhibit potential dual inhibitory activity against both SaUPPS and EcUPPS. These findings provide an insight for further optimization and development of chalcone derivatives as inhibitors against UPPSs.

4. DISCUSSION

We originally designed (4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones as TMPRSS2/Furin dual reversible inhibitors, and as expected the compounds were active antivirals only if pretreating SARS-CoV-2 to block the virus entry, but not added after infection. Because some analogues that did not inhibit TMPRSS2 and Furin still inhibited the virus entry and production [24], we then identified them as also RBD:ACE2 inhibitors. Our results thus indicate that the compounds exert their antiviral activities by inhibiting RBD:ACE2 interaction as a major factor, and inhibiting TMPRSS2 and/or Furin enhances their antiviral activities. The best inhibitor **2f** displays great inhibition against TMPRSS2, Furin, and RBD:ACE2 interaction, respectively, with IC_{50} values of 1.27, 3.31, and 3.76 μM. I performed computer modeling to provide the binding modes for the best inhibitor **2f** on its targets, TMPRSS2, Furin, and RBD. It

doi:10.6342/NTU202502385

achieves EC_{50} values of 0.023 and 0.002 μM against SARS-CoV-2 delta and omicron variants, respectively, while infecting VeroE6 cells and CC_{50} >100 μM, giving the largest therapeutic index of >50000. By using lung Calu-3 cells, the EC_{50} were measured to be 0.026 and 0.001 μM, respectively, against SARS-CoV-2 delta and omicron variants. The preliminary animal study also demonstrates the effectiveness in reducing the viral load by 5−5.5-fold in nasal turbinate and safety without causing the animal weight loss while orally given a dose of 25 mg/kg into hamsters. Pharmacokinetics study shows higher concentrations of **2f** in serum samples by IV injection than oral treatment, and both gave the concentrations above the antiviral EC_{50}. Further animal studies through different doses and ways of administration, and even in the presence of cytochrome P-450 inhibitor Ritonavir can be conducted to optimize the treatment effect in vivo.

To understand how the compound **50** in the series of tetrahydropyrazino[2,1-a:5,4-a']diisoquinolines binds RBD:ACE2 interface, computational modeling that plays a pivotal role in investigating the interactions between potential drug candidates and target proteins was utilized. Understanding these interactions is crucial as it expedites the drug discovery process by identifying promising candidates with higher efficacy and lower side effects. Using compound **50** as an example, Figure 2 reveals the interactions between the compound with Spike's RBD of SARS-CoV-2 delta strain. Based on X-ray analysis, all our synthesized compounds have a shape with a central tetrahydropyrazino ring and 4

doi:10.6342/NTU202502385

extended edges containing different groups. The extended shape helps to block the interface of RBD and ACE2 as modeled.

Computational drug discovery, through methods such as molecular docking, is vital for investigating how molecules interact with protein targets. In this study, we employ iGEMDOCK software to facilitate steps such as design of inhibitors and post-screening analysis. iGEMDOCK is especially useful for post-screening analysis and inferring pharmacological interactions from the active inhibitors [26]. Moreover, by examining the computation-based study of drug-likeliness and pharmacokinetics, compound **50** obeys the Lipinski rule of five and shows favorable ADMET profiles. However, further in vitro and in vivo studies are needed to validate its effectiveness and safety.

Furthermore, we designed, synthesized, and evaluated a series of *MRSA* inhibitors. Generally, the compounds with highly hydrophobic groups at phenyl ring display better antibacterial activities. The two lead compounds **87** and **94** showed the respective MIC of 3.68 and 3.21 ug/mL towards MRSA without apparent toxicity. We first predicted UPPS as their target. Inhibition assay against UPPSs indicates compound **87** and **94** exhibit potent activity against SaUPPS, featuring IC_{50} = 5.43 ± 0.521 and 6.10 ± 0.652, respectively. Although the actual target in vivo of this series of compounds remains unidentified, their strong inhibitory activity combined with straightforward synthesis presents significant potential for further derivatization. Given the critical challenge posed

doi:10.6342/NTU202502385

by antibiotic resistance, the development of novel antibiotics is an urgent priority. These chalcone derivatives represent a noteworthy advancement in addressing this pressing issue. Besides, virtual screening of chalcones against UPPSs and pharmacophore anchor analysis of UPPSs identified chalones potentially inhibit EcUPPS and SaUPPS for future investment and developing anti-bacteria agents. Consequently, we have successfully identified promising anti-MRSA lead compounds that warrant further investigation in pre-clinical studies, alongside efforts to confirm their actual molecular target in future study.

doi:10.6342/NTU202502385

Tables

Table 1. The properties of 2f analyzed using Lipinski rule of five.

Properties	Predicted value
Molecular weight	374 Da
Hydrogen bond donors	2
Hydrogen bond acceptors	5
LogP	3.501499
Molar refractivity	106.200974

doi:10.6342/NTU202502385

Table 2. The ADMET profile of compound 2f.

Properties	Predicted value
Absorption	
Water solubility (log mol/L)	-4.836
Caco2 permeability (log Papp in 10^{-6} cm/s)	1.029
Intestinal absorption (human) (% Absorbed)	97.014
Skin Permeability (log Kp)	-2.722
P-glycoprotein substrate	Yes
P-glycoprotein I inhibitor	No
P-glycoprotein II inhibitor	No
Distribution	
VDss (human) (log L/kg)	-0.674
Fraction unbound (human) (Fu)	0
BBB permeability (log BB)	-0.395
CNS permeability (log PS)	-2.301
Metabolism	
CYP2D6 substrate	No
CYP3A4 substrate	Yes

doi:10.6342/NTU202502385

CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	No
CYP3A4 inhibitor	No
Excretion	
Total Clearance (log ml/min/kg)	0.477
Renal OCT2 substrate	No
Toxicity	
AMES toxicity	No
Max. tolerated dose (human) (log mg/kg/day)	0.703
hERG I inhibitor	No
hERG II inhibitor	No
Oral Rat Acute Toxicity (LD_{50}) (mol/kg)	3.186
Oral Rat Chronic Toxicity (LOAEL) (log mg/kg_bw/day)	1.879
Hepatotoxicity	Yes
Skin Sensitization	No

doi:10.6342/NTU202502385

T.Pyriformis toxicity (log ug/L)	0.415
Minnow toxicity (log mM)	0.443

doi:10.6342/NTU202502385

Table 3. The properties of compound 50 analyzed using Lipinski rule of five.

Properties	Predicted value
Molecular weight	494 Da
Hydrogen bond donors	0
Hydrogen bond acceptors	2
LogP	4.946290
Molar refractivity	153.388077

doi:10.6342/NTU202502385

Table 4. The ADMET parameters of compound 50.

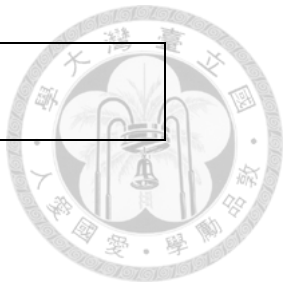
Properties	Predicted value
Absorption	
Water solubility (log mol/L)	-3.176
Caco2 permeability (log Papp in 10^{-6} cm/s)	1.051
Intestinal absorption (human) (% Absorbed)	95.817
Skin Permeability (log Kp)	-2.735
P-glycoprotein substrate	Yes
P-glycoprotein I inhibitor	Yes
P-glycoprotein II inhibitor	Yes
Distribution	
VDss (human) (log L/kg)	0.475
Fraction unbound (human) (Fu)	0.198
BBB permeability (log BB)	0.099
CNS permeability (log PS)	1.074
Metabolism	
CYP2D6 substrate	Yes
CYP3A4 substrate	Yes

doi:10.6342/NTU202502385

CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	No
CYP3A4 inhibitor	No
Excretion	
Total Clearance (log ml/min/kg)	0.305
Renal OCT2 substrate	No
Toxicity	
AMES toxicity	No
Max. tolerated dose (human) (log mg/kg/day)	0.787
hERG I inhibitor	No
hERG II inhibitor	Yes
Oral Rat Acute Toxicity (LD_{50}) (mol/kg)	2.474
Oral Rat Chronic Toxicity (LOAEL) (log mg/kg_bw/day)	1.81
Hepatotoxicity	Yes
Skin Sensitization	No
T.Pyriformis toxicity (log ug/L)	0.285

doi:10.6342/NTU202502385

Minnow toxicity (log mM)	-1.532



doi:10.6342/NTU202502385

Table 5. Kinetic parameters of the purified EcUPPS and SaUPPS

Enzymes and substrates	K_m (μM)	k_{cat} (s^{-1})
EcUPPS		
IPP	11.3 ± 1.69	1.71 ± 0.01
FPP	1.67 ± 0.23	1.36 ± 0.01
SaUPPS		
IPP	9.39 ± 2.11	2.52 ± 0.26
FPP	0.62 ± 0.16	2.08 ± 0.22

doi:10.6342/NTU202502385

Table 6. The IC_{50} values of compounds against SaUPPS and EcUPPS.

Compounds	IC_{50} against SaUPPS (μM)	IC_{50} against EcUPPS (μM)
34[a]	> 50	> 50
35[a]	> 50	> 50
48[a]	> 50	> 50
52[b]	21.99 ± 1.821	17.68 ± 0.982
58[a]	> 50	37.54 ± 0.778
65[b]	> 50	> 50
72[a]	15.34 ± 1.026	37.92 ± 0.737
79[b]	3.18 ± 0.182	13.94 ± 1.426
83[b]	> 50	> 50
84[b]	19.24 ± 2.730	23.89 ± 2.464
85[c]	12.77 ± 1.773	29.33 ± 1.107
87[c]	5.43 ± 0.521	5.24 ± 0.706
88[c]	8.89 ± 1.466	17.25 ± 0.878
94[c]	6.10 ± 0.652	10.43 ± 0.911
96[c]	4.92 ± 0.278	9.48 ± 0.578
97[b]	> 50	> 50
98[a]	20.79 ± 1.380	> 50

[a]Compounds with three aromatic rings

[b]Compounds with four aromatic moieties

[c]Compounds containing over four aromatic moieties

doi:10.6342/NTU202502385

Table 7. The binding energies of compounds against SaUPPS or EcUPPS.

Compounds	Binding Energy (kcal/mol)	
	SaUPPS	**EcUPPS**
52[b]	-154.759	-118.329
58[a]	-132.019	-105.069
72[a]	-143.672	-120.847
79[b]	-157.662	-122.035
84[b]	-140.016	-125.462
85[c]	-188.309	-161.115
87[c]	-184.46	-159.804
88[c]	-149.724	-142.198
94[c]	-230.743	-186.794
96[c]	-126.943	-133.768
98[a]	-143.03	-109.131

[a]Compounds with three aromatic rings

[b]Compounds with four aromatic moieties

[c]Compounds containing over four aromatic moieties

doi:10.6342/NTU202502385

Table 8. The properties of compound 87 and 94 analyzed by Lipinski's Rule of Five.

	87	**94**
Properties	Predicted values	
Molecular weight (Da)	504	547
Hydrogen bond donors	1	1
Hydrogen bond acceptors	4	6
LogP	4.574	5.140
Molar refractivity	138.167	144.288

doi:10.6342/NTU202502385

Table 9. The ADMET Profile of Compound 87 and 94.

Properties	**87**	**94**
	Predicted values	
Absorption		
Water solubility (log mol/L)	-4.913	-4.438
Caco2 permeability (log Papp in 10^{-6} cm/s)	0.779	0.829
Intestinal absorption (human) (% Absorbed)	98.49	87.21
Skin Permeability (log Kp)	-2.735	-2.735
P-glycoprotein substrate	Yes	No
P-glycoprotein I inhibitor	No	No
P-glycoprotein II inhibitor	No	No
Distribution		
VDss (human) (log L/kg)	-0.341	-1.677
Fraction unbound (human) (Fu)	0	0.144
BBB permeability (log BB)	-0.349	-0.841
CNS permeability (log PS)	-1.914	-2.718
Metabolism		
CYP2D6 substrate	No	No

doi:10.6342/NTU202502385

CYP3A4 substrate	Yes	Yes
CYP1A2 inhibitor	No	No
CYP2C19 inhibitor	No	No
CYP2C9 inhibitor	No	Yes
CYP2D6 inhibitor	No	No
CYP3A4 inhibitor	No	No
Excretion		
Total Clearance (log ml/min/kg)	0.354	0.32
Renal OCT2 substrate	No	No
Toxicity		
AMES toxicity	No	No
Max. tolerated dose (human) (log mg/kg/day)	0.793	0.498
hERG I inhibitor	No	No
hERG II inhibitor	No	No
Oral Rat Acute Toxicity (LD_{50}) (mol/kg)	2.981	3.079
Oral Rat Chronic Toxicity (LOAEL) (log mg/kg_bw/day)	2.33	0.615
Hepatotoxicity	No	Yes
Skin Sensitization	No	No

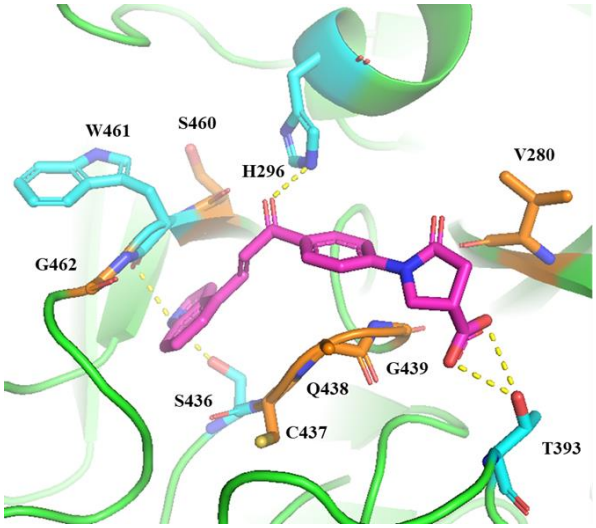
doi:10.6342/NTU202502385

T.Pyriformis toxicity (log ug/L)	0.292	0.285
Minnow toxicity (log mM)	-2.323	-3.982

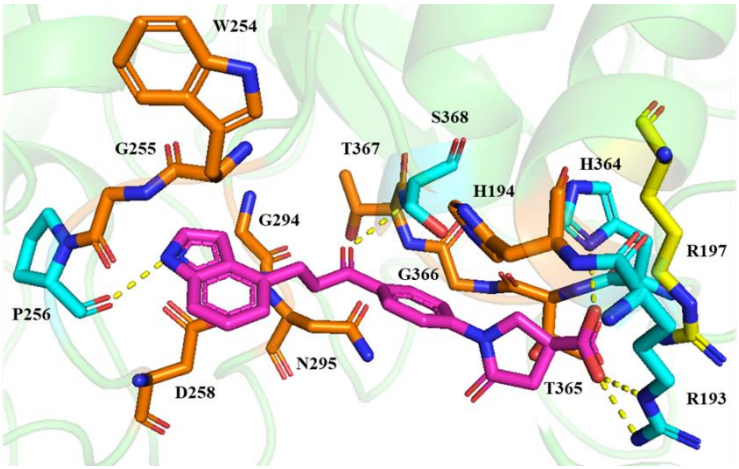
doi:10.6342/NTU202502385

FIGURES

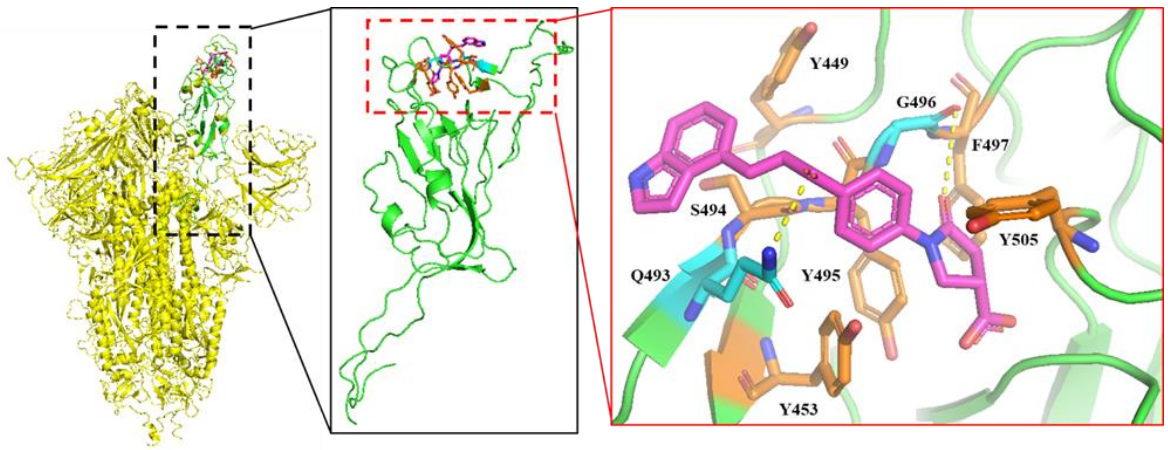
(a)



(b)



(c)



doi:10.6342/NTU202502385

Figure 1. Compound 2f was docked into (a) the TMPRSS2 ectodomain (PDB: 7MEQ), (b) Furin (PDB: 5JXG), and (c) RBD of the delta-strain Spike protein (PDB: 7W92).

The estimated binding energies of **2f** with TMPRSS2, Furin, and RBD are -115.8, -118.9, and -115.1 kcal/mol, respectively. Compound **2f** is colored in magenta, and the residues colored in cyan, orange, and yellow are involved in hydrogen-bonding, van der Waals, and electrostatic interactions, respectively. In (c), colored in yellow is trimeric Spike protein and in green is RBD. **2f** binds with the residues of RBD in the interface to interfere with the RBD binding with ACE2.

doi:10.6342/NTU202502385

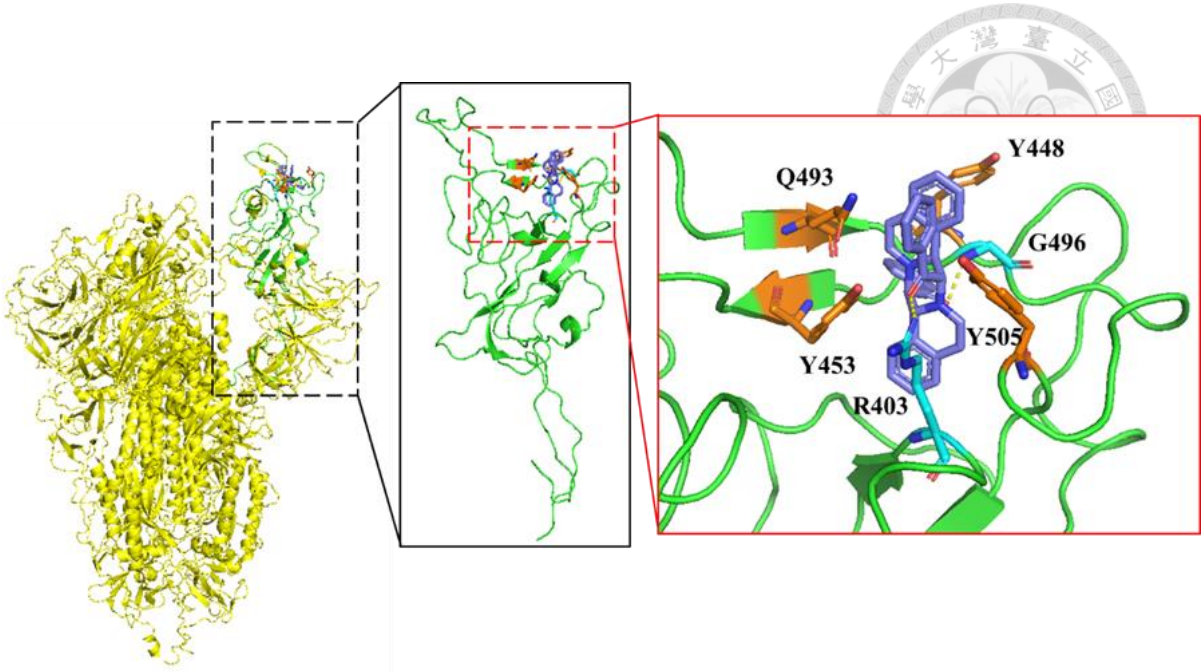
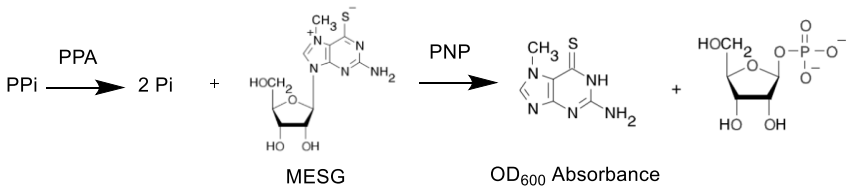


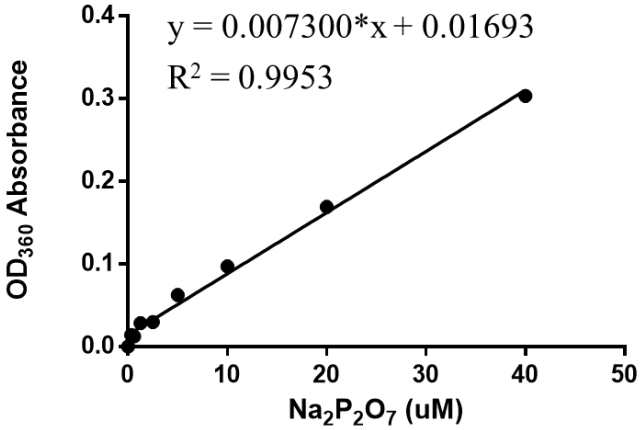
Figure 2. Binding mode of an active antiviral compound 50 with RBD.

Compound **50** was docked into the RBD of the delta-strain Spike protein (PDB: 7W92). Compound **50** is colored in purple, and the residues involved in hydrogen bonding are highlighted in cyan, while those involved in van der Waals interactions are in orange. The trimeric Spike protein is depicted in yellow and the RBD is in green. In this model, compound **50** interacts with residues at the RBD:ACE2 interface, potentially interfering with the binding of RBD to ACE2.

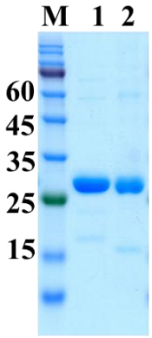
doi:10.6342/NTU202502385

(a)



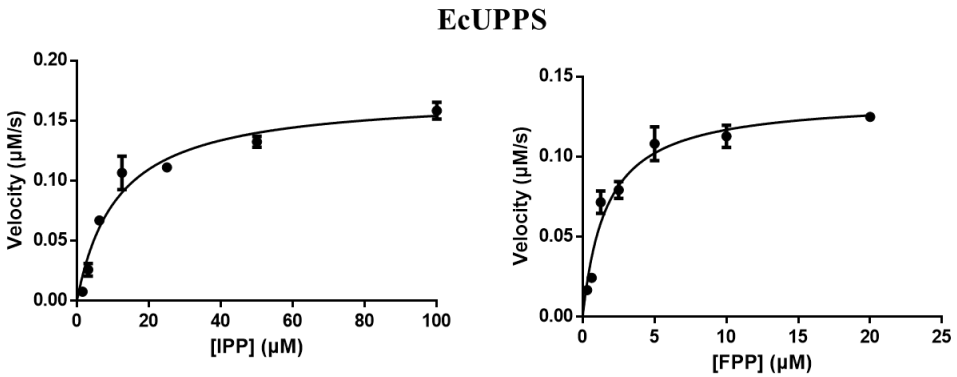


(b)



M: Marker
1: SaUPPS
(30.4 kDa)
2: EcUPPS
(28.3 kDa)

(c)



doi:10.6342/NTU202502385

(d)

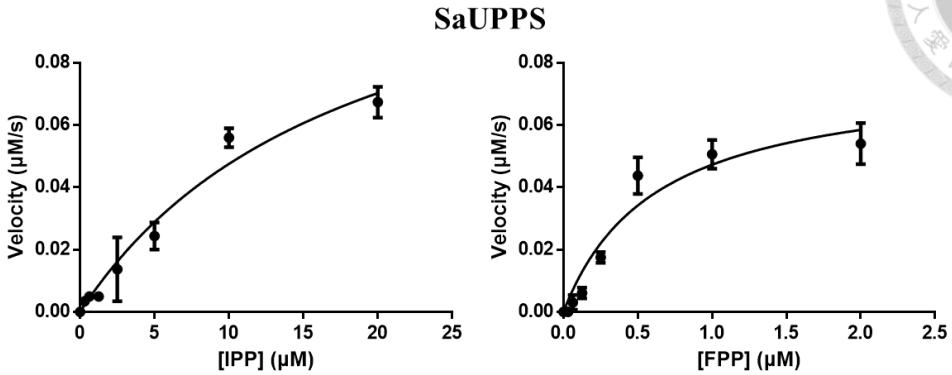


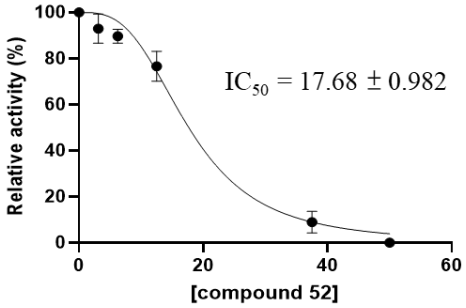
Figure 3. In vitro enzymatic activity assay of EcUPPS and SaUPPS.

(a) Schematic representation of the enzymatic activity measurement using the EnzChek™ Pyrophosphate Assay Kit, along with the standard curve generated using disodium diphosphate ($Na_2P_2O_7$) as the inorganic pyrophosphate standard. (b) SDS-PAGE analysis of the purified EcUPPS and SaUPPS. (c-d) Michaelis-Menten plots of EcUPPS (c) and SaUPPS (d) with different concentrations of FPP and 50 μM IPP to determine the K_m and k_{cat} values. For K_m determination with respect to IPP, 10 μM FPP was used. All measurements were conducted in triplicate at 30°C.

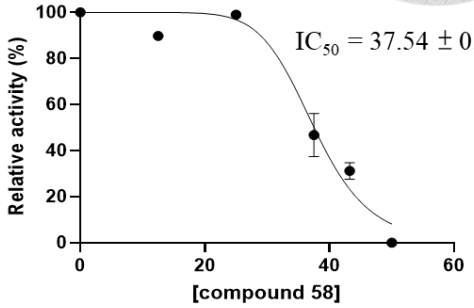
doi:10.6342/NTU202502385

EcUPPS

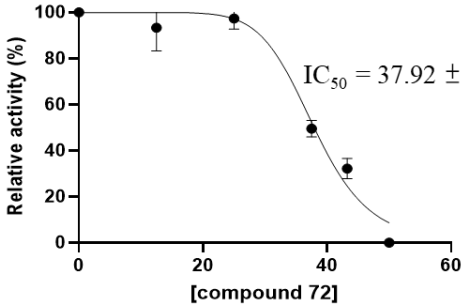
(a)



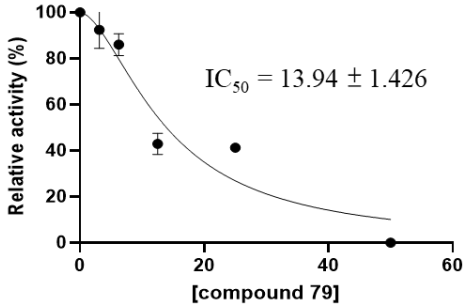
(b)



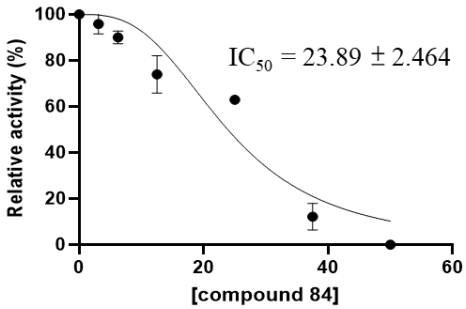
(c)



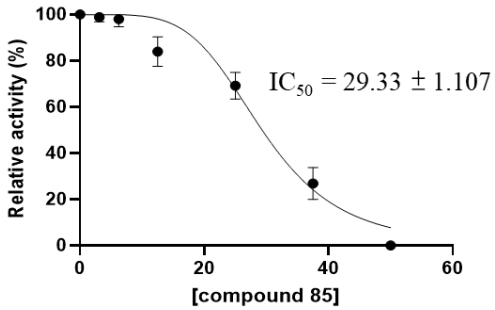
(d)



(e)

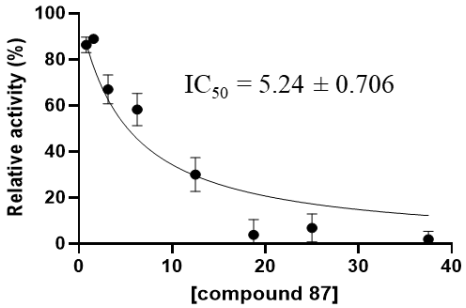


(f)

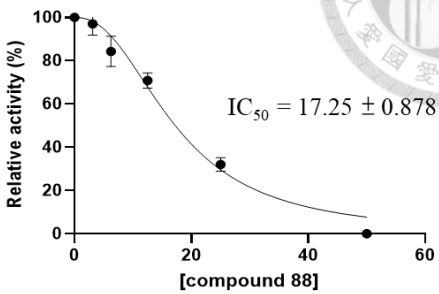


doi:10.6342/NTU202502385

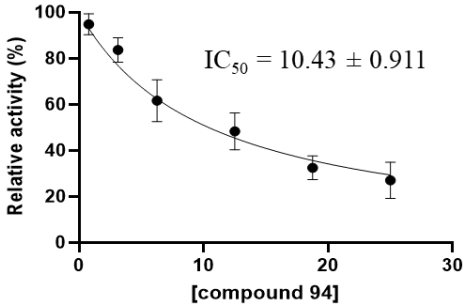
(g)



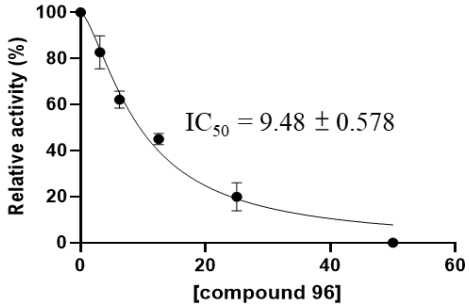
(h)



(i)



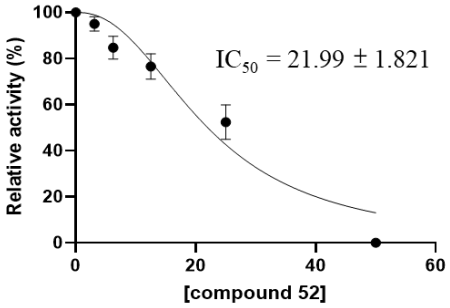
(j)



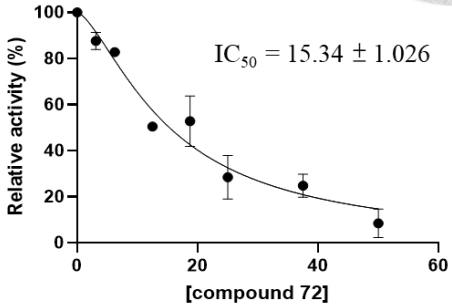
doi:10.6342/NTU202502385

SaUPPS

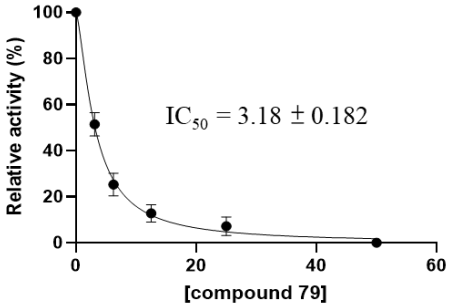
(k)



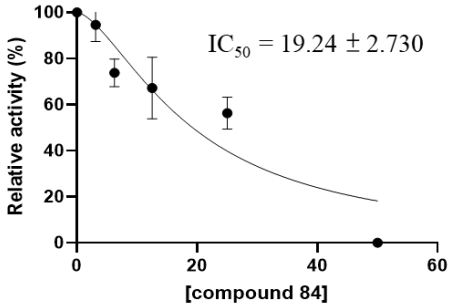
(l)



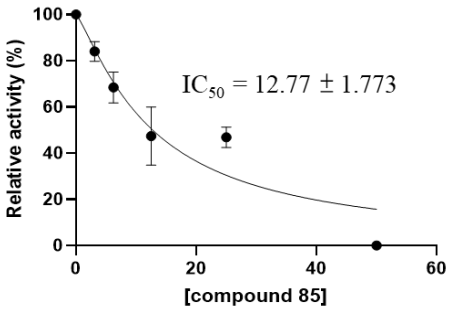
(m)



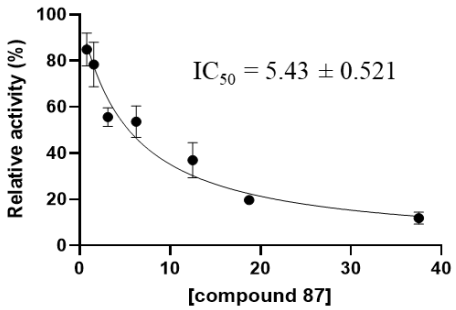
(n)



(o)



(p)



doi:10.6342/NTU202502385

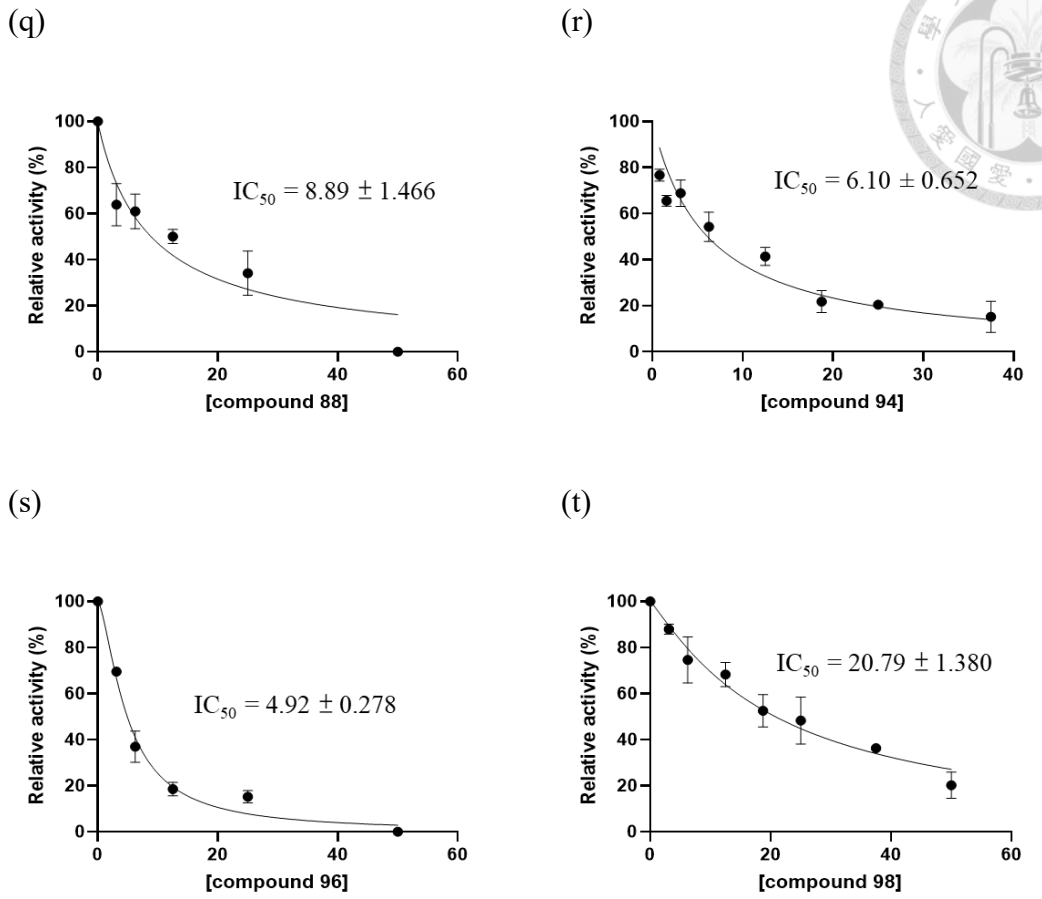
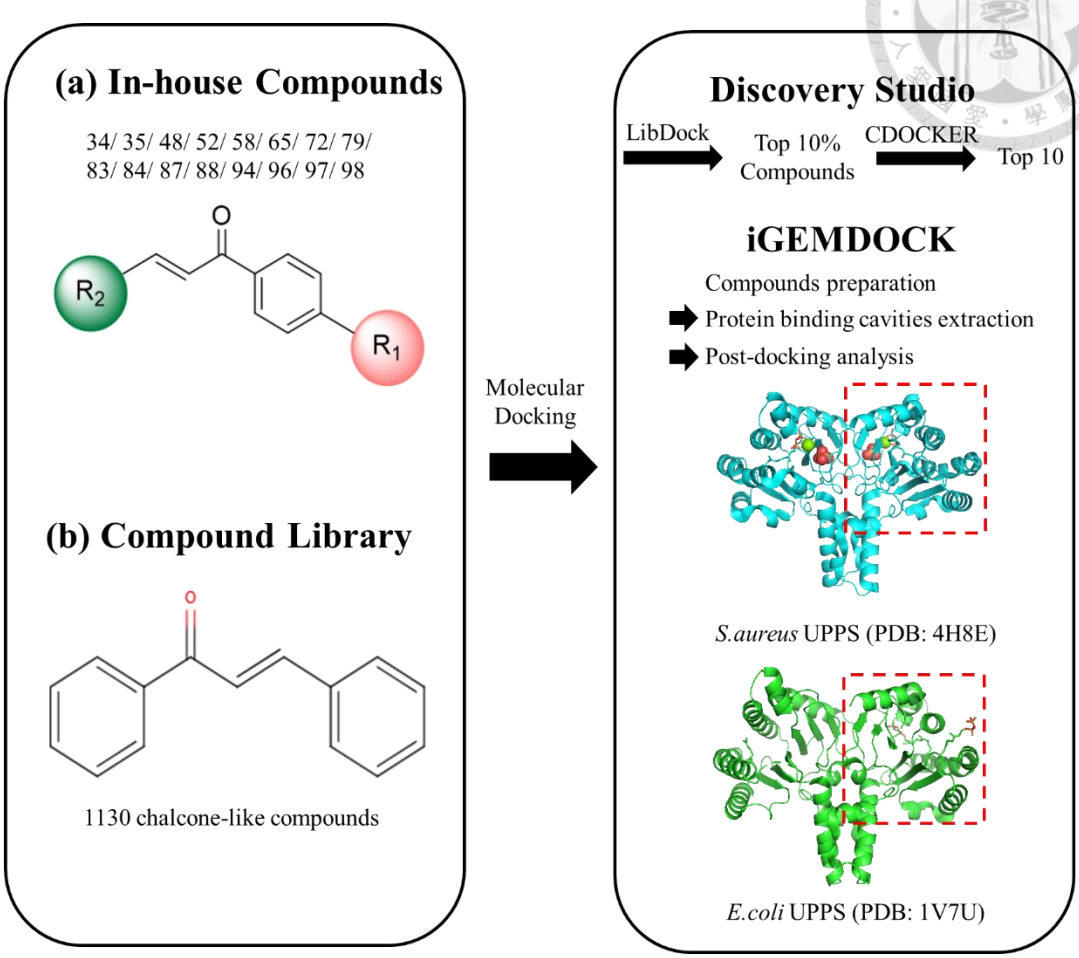


Figure 4. Dose-dependent inhibition curves of the active compound against EcUPPS (a-j) and SaUPPS (k-t).

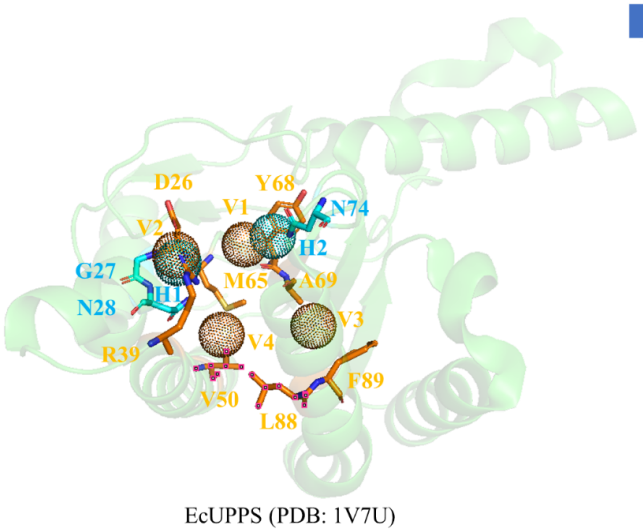
The plots of decreasing EcUPPS or SaUPPS reaction initial rates in the presence of increasing concentrations of the compounds were fitted with an equation to yield the IC_{50} values.

doi:10.6342/NTU202502385

(a)



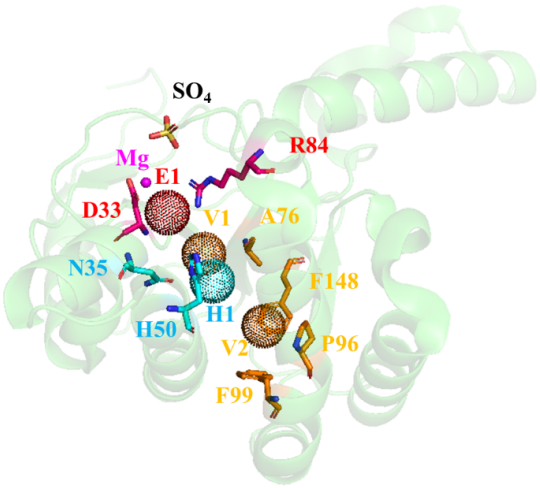
(b)



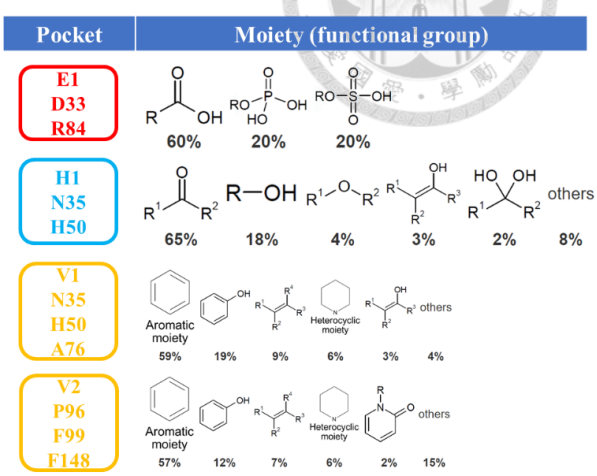
Pocket	Moiety (functional group)
H1 G27 N28	R−OH 33% · 18% · 15% · 12% · 5% · 17% others
H2 N74	R−OH 31% · 23% · 17% · 7% · 6% · 16% others
V1 M65 Y68 A69	Aromatic moiety 57% · 21% · 5% · 5% · 2% · 10% others
V2 D26 N28 R39	Aromatic moiety 53% · 27% · 5% · 5% · 2% · 8% others
V3 A69 F89	Aromatic moiety 65% · 7% · 7% · 3% · 1% · 17% others
V4 N28 V50 L88	Aromatic moiety 57% · 15% · 10% · 6% · 1% · 12% others

doi:10.6342/NTU202502385

(c)

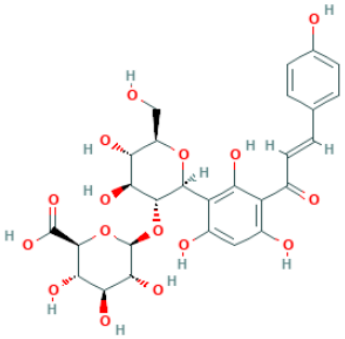


SaUPPS (PDB: 4H8E)

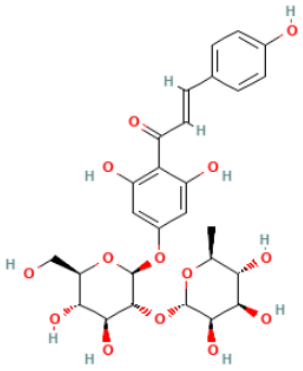


(d)

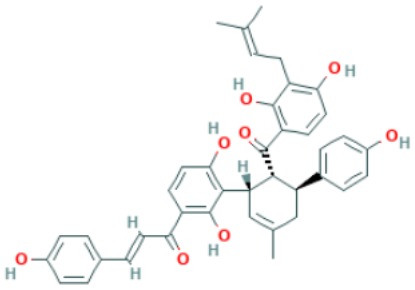
Potential EcUPPS inhibitors

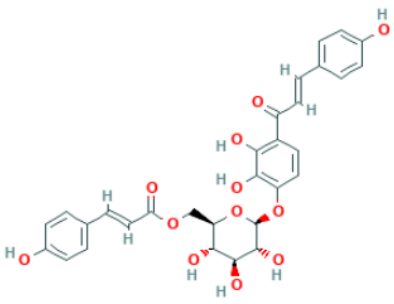


Compound CID: 102032633



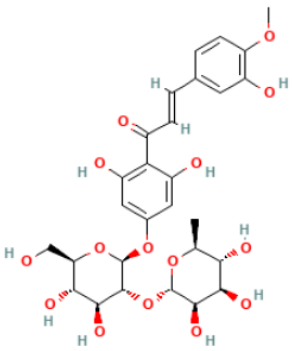
Compound CID: 23724747



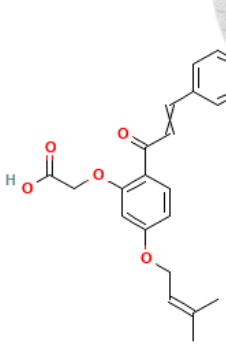


doi:10.6342/NTU202502385

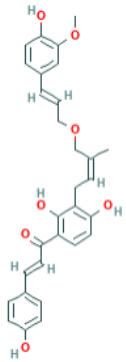
Compound CID: 45375891



Compound CID: 10460858



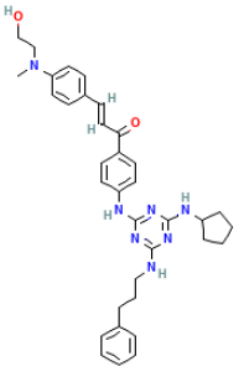
Compound CID: 5276742



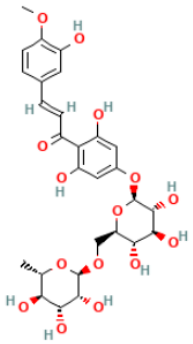
Compound CID: 54147825

Compound CID: 102309228

Potential SaUPPS inhibitors

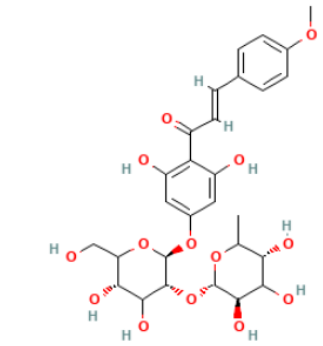


Compound CID: 102202410

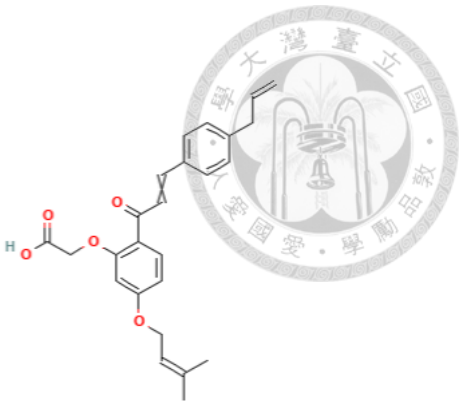


Compound CID: 89258670

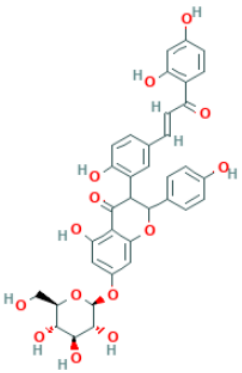
doi:10.6342/NTU202502385



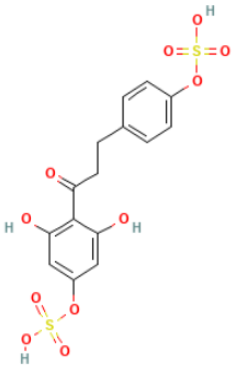
Compound CID: 91746158



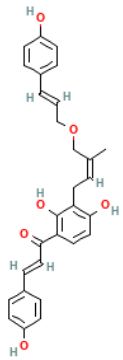
Compound CID: 54147825



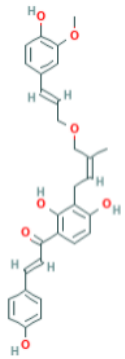
Compound CID: 102271454



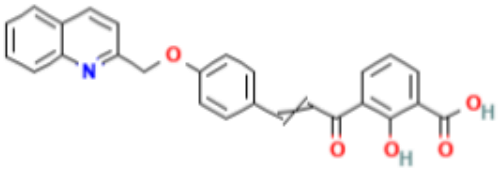
Compound CID: 122231134



Compound CID: 102309227



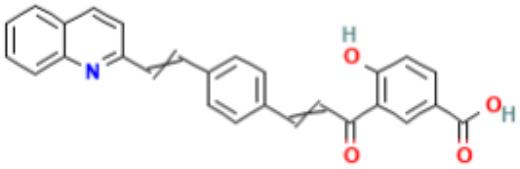
Compound CID: 102309228



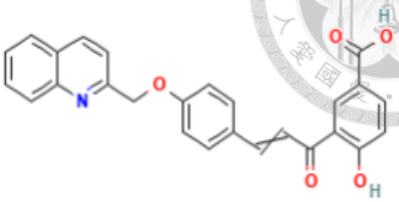
Compound CID: 53822053

doi:10.6342/NTU202502385

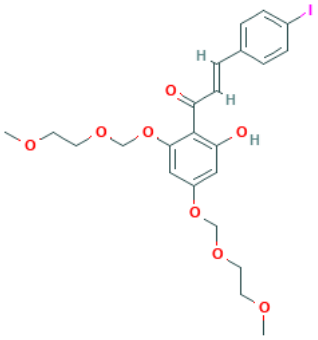
Potential dual EcUPPS and SaUPPS inhibitors



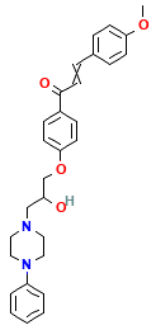
Compound CID: 53867644



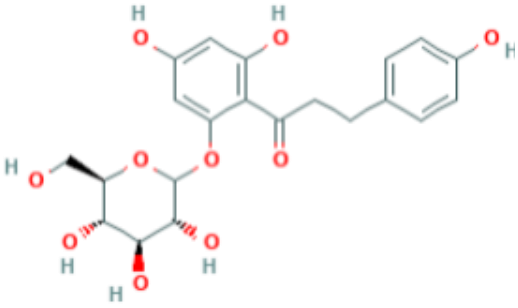
Compound CID: 54200226



Compound CID: 10602905



Compound CID:



Compound CID: 12358713

doi:10.6342/NTU202502385

Figure 5. Virtual Screening of chalcone derivatives targeting EcUPPS and SaUPPS. (a) Overview of the molecular docking process and pharmacophore anchor generation to identify natural chalcones as potential inhibitors of UPPSs. (b-c) The pharmacophore anchor models for EcUPPS (b) and SaUPPS (c). Cyan spheres represent hydrogen-bonding (H) anchors, while orange spheres indicate van der Waals (V) anchors. Key interacting residues within each anchor are highlighted: cyan-colored residues are specific to H interactions, and orange-colored residues are specific to V interactions. The interacting residues of anchors and the preferred binding moieties of chalcone derivatives within UPPSs are summarized at the right panel, providing insight into structure-activity relationships. (d) Potential chalcone derivatives inhibitors against UPPSs. Top-scored compounds are selected based on the binding scores of compounds against EcUPPS or SaUPPS analyzed by Discovery Studio and iGEMDOCK molecular docking software. 2D structures of these selected compounds were downloaded from PubChem website (https://pubchem.ncbi.nlm.nih.gov/).

doi:10.6342/NTU202502385

REFERENCES

1. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* **2020**, *395* (10223), 497-506, DOI: 10.1016/S0140-6736(20)30183-5.
2. Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; Yuan, M. L.; Zhang, Y. L.; Dai, F. H.; Liu, Y.; Wang, Q. M.; Zheng, J. J.; Xu, L.; Holmes, E. C.; Zhang, Y. Z. A new coronavirus associated with human respiratory disease in China. *Nature.* **2020**, *579* (7798), 265-269, DOI: 10.1038/s41586-020-2008-3.
3. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W.; China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* **2020**, *382* (8), 727-733. DOI: 10.1056/NEJMoa2001017.
4. Zhou, P.; Yang, X. L.; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. D.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.;

doi:10.6342/NTU202502385

Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* **2020**, *579* (7798), 270-273. DOI: 10.1038/s41586-020-2012-7.

5. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; Bi, Y.; Ma, X.; Zhan, F.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z.; Zhou, W.; Zhao, L.; Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J.; Xie, Z.; Ma, J.; Liu, W. J.; Wang, D.; Xu, W.; Holmes, E. C.; Gao, G. F.; Wu, G.; Chen, W.; Shi, W.; Tan, W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* **2020**, *395* (10224), 565-574. DOI: 10.1016/S0140-6736(20)30251-8.
6. Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol.* **2022**, *23* (1), 3-20. DOI: 10.1038/s41580-021-00418-x.
7. Bestle, D.; Heindl, M. R.; Limburg, H.; Van Lam van, T.; Pilgram, O.; Moulton, H.; Stein, D. A.; Hardes, K.; Eickmann, M.; Dolnik, O.; Rohde, C.; Klenk, H. D.; Garten, W.; Steinmetzer, T.; Böttcher-Friebertshäuser, E. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Alliance.* **2020**, *3* (9), e202000786. DOI: 10.26508/lsa.202000786.

doi:10.6342/NTU202502385

8. Essalmani, R.; Jain, J.; Susan-Resiga, D.; Andréo, U.; Evagelidis, A.; Derbali, R. M.; Huynh, D. N.; Dallaire, F.; Laporte, M.; Delpal, A.; Sutto-Ortiz, P.; Coutard, B.; Mapa, C.; Wilcoxen, K.; Decroly, E.; Nq Pham, T.; Cohen, É. A.; Seidah, N. G. Distinctive Roles of Furin and TMPRSS2 in SARS-CoV-2 Infectivity. *J Virol.* **2022**, *96* (8), e0012822. DOI: 10.1128/jvi.00128-22.

9. Mondal, S.; Chen, Y.; Lockbaum, G. J.; Sen, S.; Chaudhuri, S.; Reyes, A. C.; Lee, J. M.; Kaur, A. N.; Sultana, N.; Cameron, M. D.; Shaffer, S. A.; Schiffer, C. A.; Fitzgerald, K. A.; Thompson, P. R. Dual Inhibitors of Main Protease (MPro) and Cathepsin L as Potent Antivirals against SARS-CoV2. *J Am Chem Soc.* **2022**, *144* (46), 21035-21045. DOI: 10.1021/jacs.2c04626.

10. Levy, S. B.; Marshall, B. Antibacterial Resistance Worldwide: Causes, Challenges and Responses. *Nat. Med.* **2004**, *10* (12 Suppl), S122–S129, DOI: 10.1038/nm1145

11. Ghosh, C.; Haldar, J. Membrane-Active Small Molecules: Designs Inspired by Antimicrobial Peptides. *ChemMedChem.* **2015**, *10* (10), 1606–1624, DOI: 10.1002/cmdc.201500299

12. Lim, S. M.; Webb, S. A. R. Nosocomial Bacterial Infections in Intensive Care Units. I: Organisms and Mechanisms of Antibiotic Resistance. *Anaesthesia* **2005**, *60* (9), 887–902, DOI: 10.1111/j.1365-2044.2005.04220.x

doi:10.6342/NTU202502385

13. Fischbach, M. A.; Walsh. C. T. Antibiotics for Emerging Pathogen. *Science* **2009**, *325* (5944), 1089–1093, DOI: 10.1126/science.1176667

14. Bulter, M. S.; Gigante V.; Sati, H.; Paulin, S.; Al-Sulaiman, L.; Rex, J. H.; Fernandes, P.; Arias, C. A.; Paul, M.; Thwaites, G. E.; Czaplewski, L.; Alm, R. A.; Lienhardt, C.; Spigelman, M.; Silver, L. L.; Ohmagari, N.; Kozlov, R.; Harbarth, S.; Beyer, P. Analysis of the Clinical Pipeline of Treatments for Drug-Resistant Bacterial Infections: Despite Progress, More Action Is Needed. *Antimicrob. Agents Chemother.* **2022**, *66* (3), e0199121, DOI: 10.1128/AAC.01991-21.

15. Grundmann, H.; Aires-de-Sousa, M.; Boyce, J.; Tiemersma, E. Emergence and Resurgence of Meticillin-Resistant Staphylococcus aureus as a Public-Health Threat. *Lancet* **2006**, *368* (9538), 874–885, DOI: 10.1016/S0140-6736(06)68853-3

16. Tacconelli, E.; Carrara, E.; Savoldi, A.; Harbarth, S.; Mendelson, M.; Monnet, D. L.; Pulcini, C.; Kahlmeter, G.; Kluytmans, J.; Carmeli, Y.; Ouellette, M.; Outterson, K.; Patel, J.; Cavaleri, M.; Cox, E. M.; Houchens, C. R.; Grayson, M. L.; Hansen, P.; Singh, N.; Theuretzbacher, U.; Magrini, N.; WHO Pathogens Priority List Working Group. Discovery, Research, and Development of New Antibiotics: the WHO Priority List of Antibiotic-Resistant Bacteria and Tuberculosis. *Lancet Infect. Dis.* **2018**, *18* (3), 318–327, DOI: 10.1016/S1473-3099(17)30753-3

doi:10.6342/NTU202502385

17. Teng, K. H.; Liang, P. H. Structures, Mechanisms and Inhibitors of Undecaprenyl Diphosphate Synthase: A cis-Prenyltransferase for Bacterial Peptidoglycan Biosynthesis. *Bioorg. Chem.* **2012**, *43*, 51–57, DOI: 10.1016/j.bioorg.2011.09.004

18. Sinko, W.; Wang, Y.; Zhu, W.; Zhang, Y.; Feixas, F.; Cox, C. L.; Mitchell, D. A.; Oldfield, E.; McCammon, J. A. Undecaprenyl Diphosphate Synthase Inhibitors: Antibacterial Drug Leads. *J. Med. Chem.* **2014**, *57* (13), 5693–5701, DOI: 10.1021/jm5004649.

19. Fatha, M. A.; Czarny, T. L.; Myers, C. L.; Worrall, L. J.; French, S.; Conrady, D. G.; Wang, Y.; Oldfield, E.; Strynadka, N. C.; Brown, E. D. Antagonism Screen for Inhibitors of Bacterial Cell Wall Biogenesis Uncovers an Inhibitor of Undecaprenyl Diphosphate Synthase. *Proc. Natl. Acad. Sci., USA* **2015**, *112* (35), 11048–11053, DOI: 10.1073/pnas.1511751112

20. Wang, Y.; Desai, J.; Zhang, Y.; Malwal, S. R.; Shin, C. J.; Feng, X.; Sun, H.; Liu, G.; Guo, R. T.; Oldfield, E. Bacterial Cell Growth Inhibitors Targeting Undecaprenyl Diphosphate Synthase and Undecaprenyl Diphosphate Phosphatase. *ChemMedchem.* **2016**, *11* (20), 2311–219, DOI: 10.1002/cmdc.201600342

21. Malwal, S. R.; Chen, L.; Hicks, H.; Qu, F.; Liu, W.; Shillo, A.; Law, W. X.; Zhang, J.; Chandnani, N.; Han, X.; Zheng, Y.; Chen, C. C.; Guo, R. T.; AbdelKhalek, A.; Seleem, M. N.; Oldfield, E. Discovery of Lipophilic Bisphosphonates that Target

doi:10.6342/NTU202502385

Bacterial Cell Wall and Quinone Biosynthesis. *J. Med. Chem.* **2019**, *62* (5), 2564–2581, DOI: 10.1021/acs.jmedchem.8b01878.

22. Song, J.; Malwal, S. R.; Baig, N.; Schurig-Briccio, L. A.; Gao, Z.; Vaidya, G. S.; Yang, K.; Abutaleb, N. S.; Seleem, M. N.; Gennis, R. B.; Pogorelov, T. V.; Oldfield, E.; Feng, X. Discovery of Prenyltransferase Inhibitors with In Vitro and In Vivo Antibacterial Activity. *ACS Infect. Dis.* **2020**, *6* (11), 2979–2993, DOI: 10.1021/acsinfecdis.0c00472.

23. Kuo, C. J.; Guo, R. T.; Lu, I. L.; Liu, H. G.; Wu, S. Y.; Ko, T. P.; Wang, A. H.-J.; Liang, P. H. Structure-Based Inhibitors Exhibited Differential Activities against H. pylori and E. coli Undecaprenyl Pyrophosphate Synthases. *J. Biomed. Biotechnol.* **2008**, *2008*, 841312, DOI: 10.1155/2008/841312.

24. Palla. S. R.; Li, C. W.; Chao, T. L.; Lo, H. L. V.; Liu, J. J.; Pan, M. Y. C.; Chiu, Y. T.; Lin, W. C.; Hu, C. W.; Yang, C. M.; Chen, Y. Y.; Fang, J. T.; Lin, S. W.; Lin, Y. T.; Lin, H. C.; Kuo, C.J.; Wang, L. H. C.; Chang, S. Y.; Liang, P. H. Synthesis, Evaluation, and Mechanism of 1-(4-(Arylethylenylcarbonyl)phenyl)-4-Carboxy-2-Pyrrolidinones as Potent Reversible SARS-CoV-2 Entry Inhibitors. Antiviral Res. **2023**, *219*, 105735, DOI: 10.1016/j.antiviral.2023.105735

25. Palla, S.; Palla, S. R.; Liu, J. J.; Chao, T. L.; Lee, T. H.; Kavala, V.; Liu, I. C.; Wang, L. H. C.; Chang, S. Y.; Yao, C. F.; Liang, P. H. Green Synthesis of

doi:10.6342/NTU202502385

Tetrahydropyrazino[2,1-a:5,4-a']diisoquinolines as SARS-CoV-2 Entry Inhibitors. *ACS Omega.* **2024**. DOI: 10.1021/acsomega.4c08640

26. Hsu, K. C.; Chen, Y. F.; Lin, S. R.; Yang, J. M. iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *BMC bioinformatics.* **2011**, *12* Suppl 1 (Suppl 1), S33. DOI: 10.1186/1471-2105-12-S1-S33.
27. Fraser, B. J.; Beldar, S.; Seitova, A.; Hutchinson, A.; Mannar, D.; Li, Y.; Kwon, D.; Tan, R.; Wilson, R. P.; Leopold, K.; Subramaniam, S.; Halabelian, L.; Arrowsmith, C. H.; Bénard, F. Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. *Nat Chem Biol.* **2022**, *18* (9), 963-971. DOI: 10.1038/s41589-022-01059-7.
28. Dahms, S. O.; Arciniega, M.; Steinmetzer, T.; Huber, R.; Than, M. E. Structure of the unliganded form of the proprotein convertase furin suggests activation by a substrate-induced mechanism. *Proc Natl Acad Sci U S A.* **2016**, *113* (40), 11196-11201. DOI: 10.1073/pnas.1613630113.
29. Wang, Y.; Liu, C.; Zhang, C.; Wang, Y.; Hong, Q.; Xu, S.; Li, Z.; Yang, Y.; Huang, Z.; Cong, Y. Structural basis for SARS-CoV-2 Delta variant recognition of ACE2 receptor and broadly neutralizing antibodies. *Nat Commun.* **2022**, *13* (1), 871. DOI: 10.1038/s41467-022-28528-w.

doi:10.6342/NTU202502385

30. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R.; Open Babel: An Open Chemical Toolbox 2011. *J. Cheminform.* **2011**, *3*, 33, DOI: 10.1186/1758-2946-3-33.

31. Lipinski, C. A. Lead-and Drug-Like Compounds: the Rule-of-Five Revolution. *Drug Discov.* Today Technol. **2004**, *1* (4), 337–341, DOI: 10.1016/j.ddtec.2004.11.007.

32. Jayaram, B.; Singh, T.; Mukherjee, G.; Mathur, A.; Shekhar, S.; Shekhar, V. A. Freely Accessible Web-Server for Target Directed Lead Molecule Discovery. *BMC bioinformatics* **2012**, *13* (Suppl 17), 1−13, DOI: 10.1186/1471-2105-13-S17-S7.

33. Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **2015**, *58*, 4066–4072, DOI: 10.1021/acs.jmedchem.5b00104.

34. Studio, D. Discovery studio. **2008**. Accelrys [2.1], 420.

35. Chang, S. Y.; Ko, T. P.; Chen, A. P.; Wang, A. H; Liang, P. H. Substrate binding mode and reaction mechanism of undecaprenyl pyrophosphate synthase deduced from crystallographic studies. *Protein Sci.* **2004**, *13* (4), 971-8. DOI: 10.1110/ps.03519904.

36. Zhu, W.; Zhang, Y.; Sinko, W.; Hensler, M. E.; Olson, J.; Molohon, K. J.; Lindert, S.; Cao, R.; Li, K.; Wang, K.; Wang, Y.; Liu, Y. L.; Sankovsky, A.; de Oliveira, C.

doi:10.6342/NTU202502385

A.; Mitchell, D. A.; Nizet, V.; McCammon, J. A.; Oldfield, E. Antibacterial drug leads targeting isoprenoid biosynthesis. *Proc Natl Acad Sci U S A.* **2013**, *110* (1), 123-8. DOI: 10.1073/pnas.1219899110.

37. Chen, Y. F.; Hsu, K. C.; Lin, S. R.; Wang, W. C.; Huang, Y. C.; Yang, J. M. SiMMap: a web server for inferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties. *Nucleic Acids Res.* 2010, *38* (suppl_2): W424-30, DOI: 10.1093/nar/gkq480.
38. Chang, S. Y.; Ko, T. P.; Liang, P. H.; Wang, A.H. Catalytic mechanism revealed by the crystal structure of undecaprenyl pyrophosphate synthase in complex with sulfate, magnesium, and triton. *J Biol Chem.* **2003**, 278(31), 29298-307, DOI: 10.1074/jbc.M302687200.
39. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol*. **2001**, *126* (2), 485-93, DOI: 10.1104/pp.126.2.485.
40. Katsori AM, Hadjipavlou-Litina D. Recent progress in therapeutic applications of chalcones. *Expert Opin Ther Pat.* **2011**, *21* (10), 1575-96, DOI: 10.1517/13543776.2011.596529.

doi:10.6342/NTU202502385

41. Zhuang C, Zhang W, Sheng C, Zhang W, Xing C, Miao Z. Chalcone: A Privileged Structure in Medicinal Chemistry. *Chem Rev.* **2017**, *117* (12), 7762-7810, DOI: 10.1021/acs.chemrev.7b00020.

42. Xu, M.; Wu, P.; Shen, F.; Ji, J.; Rakesh, K. P. Chalcone Derivatives and Their Antibacterial Activities: Current Development. *Bioorg. Chem.* **2019**, *91*, 103133, DOI: 10.1016/j.bioorg.2019.103133

43. Bhale, P. S.; Dongare, S. B.; Chanshetti, U. B. Synthesis and antimicrobial screening of chalcones containing imidazo[1,2-a] pyridine nucleus. *Res. J. Chem. Sci.* **2013**, *3*, 38–42.

44. Mohammad, F.; Rahaman, S. A.; Moinuddin, M. D. Synthesis and Antimicrobial Activity of 1-(2',4'-Dichlorophenyl)-3-(Substituted aryl)-2-Propene-1-Ones. *Int. J. Life Sci. Pharm. Res.* **2009**, *2*, 82-87.

45. Asiri, A. M.; Khan, S. A. Synthesis, Characterization, and In Vitro Antibacterial Activities of Macromolecules Derived from Bis-Chalcone. *J. Heter. Chem.*, **2012**, *49* (6), 1434–1438, DOI: https://doi.org/10.1002/jhet.942

46. Asiri, A. M.; Khan, S. A. Green Synthesis, Characterization and Biological Evaluation of Novel Chalcones as Anti Bacterial Agents. *Ara. J. Chem.* **2017**, *10* (2), S2890–S2895, DOI: 10.1016/j.arabjc.2013.11.018

doi:10.6342/NTU202502385

47. Ghani, S. B. A.; Weaver, L.; Zidan, Z. H.; Ali, H. M.; Keevil, C. W.; Brown, R. Microwave Assisted Synthesis and Antimicrobial Activities of Flavonoid Derivatives. *Bioorg. Med. Chem. Lett.* **2008**, *18* (2), 518–522, DOI: 10.1016/j.bmcl.2007.11.081

48. Yadav, N.; Dixit, S. K..; Bhattacharya, A.; Mishra, L. C.; Sharma, M.; Awasthi, S. K.; Bhasin, V. K. Antimalarial Activity of Newly Synthesized Chalcone Derivatives in Vitro. *Chem. Biol. Drug Des.* **2012**, *80* (2), 340–347, DOI: 10.1111/j.1747-0285.2012.01383.x

49. Insuasty, B.; Ramírez, J.; Becerra, D.; Echeverry, C.; Quiroga, J.; Abonia, R.; Robledo, S. M.; Velez, I. D.; Upegui, Y.; Munoz, J. A.; Ospina, V.; Nogueras, M.; Cobo, J. An Efficient Synthesis of New Caffeine-Based Chalcones, Pyrazolines and Pyrazolo[3,4-b][1,4]diazepines as Potential Antimalarial, Antitrypanosomal and Antileishmanial Agents. *Eur. J. Med. Chem.* **2015**, *93*, 401–413, DOI: 10.1016/j.ejmech.2015.02.040

50. Auffret, G.; Labaied, M.; Frappier, F.; Rasoanaivo, P.; Grellier, P.; Lewin, G. Synthesis and Antimalarial Evaluation of a Series of Piperazinyl Flavones. *Bioorg. Med. Chem. Lett.* **2007**, *17* (4), 959–963, DOI: 10.1016/j.bmcl.2006.11.051

51. Salum, L. B.; Altei, W. F.; Chiaradia, L. D.; Cordeiro, M. N. S.; Canevarolo, R. R.; Melo, C. P. S.; Winter, E.; Mattei, B.; Daghestani, H. N.; Santos-Silva, M. C.;

doi:10.6342/NTU202502385

Creczynski-Pasa, T. B.; Yunes, R. A.; Yunes, J. A.; Andricopulo, A. D.; Day, B. W.; Nunes, R. J.; Vogt, A. Cytotoxic 3,4,5-Trimethoxychalcones as Mitotic Arresters and Cell Migration Inhibitors. *Eur. J. Med. Chem.* **2013**, *63*, 501–510, DOI: 10.1016/j.ejmech.2013.02.037

52. Murthy, Y. L. N.; Suhasini, K. P.; Pathania, A. S.; Bhushan, S.; Sastry. Y. N. Synthesis, Structure Activity Relationship and Mode of Action of 3-Substitutedphenyl-1-(2,2,8,8-Tetramethyl-3,4,9,10-Tetrahydro-2*H*, 8*H*-Pyrano[2,3-*f*]chromen-6-yl)-Propenones as Novel Anticancer Agents in Human Leukaemia HL-60 Cells. *Eur. J. Med. Chem.* **2013**, *62*, 545–555, DOI: 10.1016/j.ejmech.2013.01.027

53. Yuan, Q.; Liu, Z.; Xiong, C.; Wu, L.; Wang, J.; Ruan, J. A Novel, Broad-Spectrum Antitumor Compound Containing the 1-Hydroxycyclohexa-2,5-Dien-4-One group: the Disclosure of a New Antitumor Pharmacophore in Protoapigenone 1. *Bioorg. Med. Chem. Lett.* **2011**, *21* (11), 3427–3430, DOI: 10.1016/j.bmcl.2011.03.108

54. Liu, Z.; Tang, L.; Zou, P.; Zhang, Y.; Wang, Z.; Fang, Q.; Jiang, L.; Chen, G.; Xu, Z.; Zhang, H.; Liang, G. Synthesis and Biological Evaluation of Allylated and Prenylated Mono-Carbonyl Analogs of Curcumin as Anti-Inflammatory Agents. *Eur. J. Med. Chem.* **2014**, *74*, 671–682, DOI: 10.1016/j.ejmech.2013.10.061

doi:10.6342/NTU202502385

55. Nakatsuka, T.; Tomimori, Y.; Fukuda, Y.; Nukaya, H. First Total Synthesis of Structurally Unique Flavonoids and Their Strong Anti-Inflammatory Effect. *Bioorg. Med. Chem. Lett.* **2004**, *14* (12), 3201–3203, DOI: 10.1016/j.bmcl.2004.03.108

56. Reichwald, C.; Shimony, O.; Dunkel, U.; Sacerdoti-Sierra, N.; Jaffe, C. L.; Kunick, C. 2-(3-Aryl-3-Oxopropen-1-yl)-9-Tert-Butyl-Paullones: a New Antileishmanial Chemotype. *J. Med. Chem.* **2008**, *51* (3), 659–665, DOI: 10.1021/jm7012166

57. Carvalho, S. A.; Feitosa, L. O.; Soares, M.; Costa, T. E. M. M.; Henriques, M. G.; Salomao, K.; Castro, S. L. D.; Kaiser, M.; Brun, R.; Wardell, J. L.; Wardell, S. M. S. V.; Trossini, G. H. G.; Andricopulo, A. D.; Silva, E. F. D.; Fraga, C. A. M. Design and Synthesis of New (E)-Cinnamic N-Acylhydrazones as Potent Antitrypanosomal Agents. *Eur. J. Med. Chem.* **2012**, *54*, 512–521, DOI: 10.1016/j.ejmech.2012.05.041

58. Sharma, H.; Patil, S.; Sanchez, T. W.; Neamati, N.; Schinazi, R. F.; Buolamwini, J. K. Synthesis, Biological Evaluation and 3D-QSAR Studies of 3-Keto Salicylic Acid Chalcones and Related Amides as Novel HIV-1 Integrase Inhibitors. *Bioorg. Med. Chem.* **2011**, *19* (6), 2030–2045, DOI: 10.1016/j.bmc.2011.01.047

59. Sivakumar, P. M.; Prabhakar, P. K.; Doble, M. Synthesis, Antioxidant Evaluation, and Quantitative Structure-Activity Relationship Studies of Chalcones. *Med. Chem. Res.* **2011**, *20* (4), 482–492, DOI: 10.1007/s00044-010-9342-1

doi:10.6342/NTU202502385

60. Sashidhara, K. V.; Avula, S. R.; Mishra, V.; Palnati, G. R.; Singh, L. R.; Singh, N.; Chhonker, Y. S.; Swami, P.; Bhatta, R. S.; Palit, G. Identification of Quinoline-Chalcone Hybrids as Potential Antiulcer Agents. *Eur. J. Med. Chem.* **2015**, *89*, 638–653, DOI: 10.1016/j.ejmech.2014.10.068

doi:10.6342/NTU202502385

APPENDIX

LIST OF PUBLICATIONS

doi:10.6342/NTU202502385

STATEMENT OF ORIGINALITY

Most of the results presented in this thesis have been published, with myself listed as either the first author or a co-author on these publications. Chapter I has been published as “Complexation and evolution of cis-prenyltransferase homologues in *Cinnamomum kanehirae* deduced from kinetic and functional characterizations”, while part of Chapter III has been published in “Synthesis, Evaluation, and Mechanism of 1-(4-(Arylethylenylcarbonyl)phenyl)-4-Carboxy-2-Pyrrolidinones as Potent Reversible SARS-CoV-2 Entry Inhibitors” and “Green Synthesis of Tetrahydropyrazino[2,1-a:5,4-a']diisoquinolines as SARS-CoV-2 Entry Inhibitors”.

doi:10.6342/NTU202502385

Received: 29 June 2023 | Revised: 20 October 2023 | Accepted: 27 October 2023
DOI: 10.1002/pro.4828

RESEARCH ARTICLE

THE PROTEIN SOCIETY WILEY

Complexation and evolution of cis-prenyltransferase homologues in *Cinnamomum kanehirae* deduced from kinetic and functional characterizations

Jia-Jin Liu[1] | **Po-Huang Liang**[1,2]

[1]Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan
[2]Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan

Correspondence
Po-Huang Liang, Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan.
Email: phliang@gate.sinica.edu.tw

Funding information
National Council of Science and Technology, Taiwan, Grant/Award Number: 102-2113-M-001-005-MY3

Review Editor: John Kuriyan

Abstract

Eukaryotic dehydrodolichyl diphosphate synthases (DHDDSs), cis-prenyltransferases (cis-PTs) synthesizing precursors of dolichols to mediate glycoprotein biosynthesis require partners, for eample Nus1 in yeast and NgBR in animals, which are cis-PTs homologues without activity but to boost the DHDDSs activity. Unlike animals, plants have multiple cis-PT homologues to pair or stand alone to produce various chain-length products with less known physiological roles. We chose *Cinnamomum kanehirae*, a tree that contains two DHDDS-like and three NgBR-like proteins from genome analysis, and found that one DHDDS-like protein acted as a homodimeric cis-PT to make a medium-chain C55 product, while the other formed heterodimeric complexes with either one of two NgBR homologues to produce longer-chain products. Both complexes were functional to complement the growth defect of the yeast rer2 deficient strain at a higher temperature. From the roles for the polyprenol and dolichol biosynthesis and sequence motifs, their homologues in various species were compared to reveal their possible evolutionary paths.

KEYWORDS
cis-prenyltransferase, DHDDS, dolichol, NgBR, protein glycosylation

1 | INTRODUCTION

A group of prenyltransferases synthesize linear polyprenyl diphosphates through consecutive condensation reactions of homoallylic C5 isopentenyl diphosphate (IPP), with an allylic oligoprenyl diphosphate, C5 methylallyl diphosphate (DMAPP), C10 genanyl diphosphate (GPP), C15-farnesyl diphosphate (FPP), or C20-geranylgeranyl diphosphate (GGPP), to form different chain-length products (Liang et al., 2002; Ogura & Koyama, 1998; Sacchettini & Poulter, 1997). Based on the stereochemistry of the double bonds formed from the IPP condensation reactions, these enzymes are classified as trans- (trans-PTs) or cis-prenyltransferases (cis-PTs) (Liang, 2009). While trans-PTs usually synthesize up to C50 products, cis-PTs with different 3-D structures from those of trans-PTs, although both catalyze IPP condensation, produce much longer products as lipid carriers (see Figure S1a for the cis-PTs catalyzed reactions) (Takahashi & Koyama, 2006). For example, a bacterial cis-PT undecaprenyl diphosphate synthase (UPPS) synthesizes C55 product via eight IPP condensation reactions with FPP, as a lipid carrier to mediate biosynthesis of peptidoglycan (Teng & Liang, 2012). On the other hand, longer C70–120 polyprenyl diphosphates, named dehydrodolichyl diphosphate synthases (DHDDs), catalyzed by the cis-PTs called DHDDSs in eukaryotes need to be dephosphorylated and reduced to dolichols and finally dolichyl phosphate

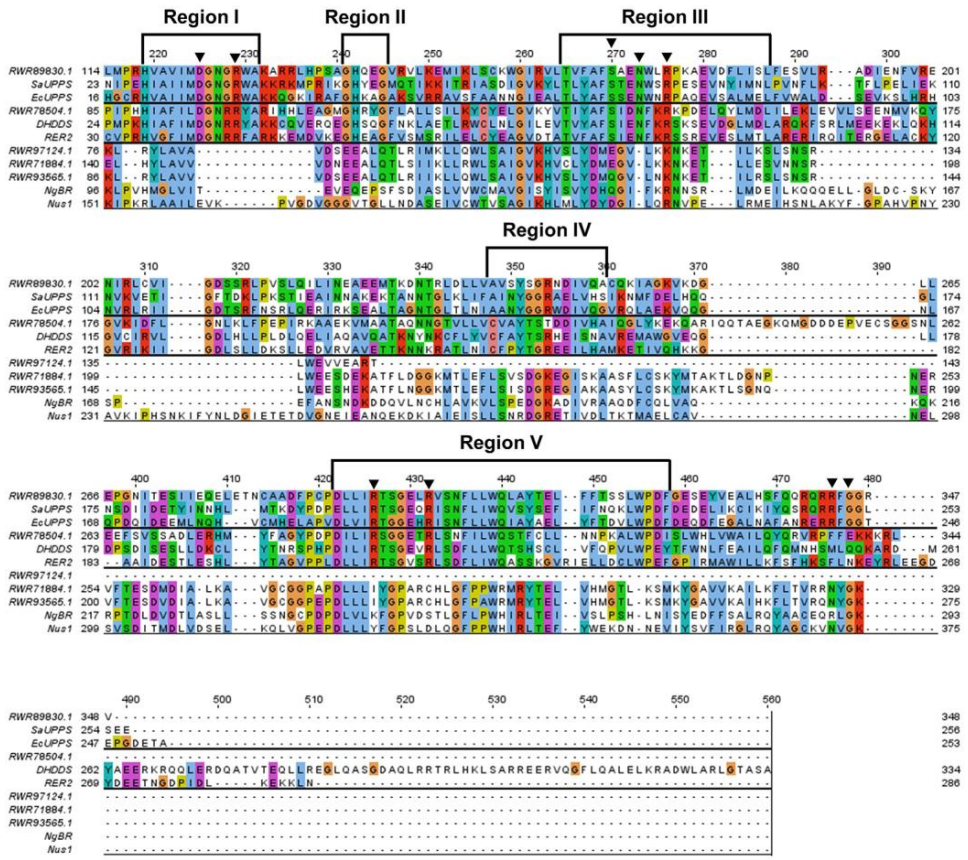


FIGURE 1 Multiple sequence alignment of five *C. kanehirae* cis-PT homologues. (a) Amino acid sequence alignment of five *C. kanehirae* cis-PT homologues against the conserved cis-PTs from bacteria, yeast, and human. Proteins involved in this figure are prokaryotic cis-PTs: *E. coli* UPPS (EcUPPS) and *S. aureus* UPPS (SaUPPS), orthologs of heteromeric DHDDS subunit: human DHDDS and yeast Rer2, and orthologs of heteromeric NgBR subunit: human NgBR and yeast Nus1. Black triangles indicate the key residues involved in the catalytic reactions and substrates binding.

(Dol-P) as a glycosyl carrier across the endoplasmic reticulum (ER) membrane for making glycoproteins (Cantagrel et al., 2010) (see Figure S1b for dolichol biosynthesis). Although playing different biological functions, these cis-PTs catalyzing the same IPP condensation reactions but leading to different long chain-length products share significant sequence homology (Figure 1).

The earliest solved crystal structures of cis-PTs came from the bacterial dimeric UPPS with each subunit containing an active-site crevice surrounded by 7 α-helices and 6 β-strands, for accommodating elongated C55 product with the bottom sealed by Leu137 (Fujihashi et al., 2001; Ko et al., 2001). From several apo and liganded *Escherichia coli* UPPS structures, we concluded that FPP is bound with Asp26 to facilitate its dephosphorylation while deprotonated IPP attacks the partial carbocation intermediate of FPP, a concerted mechanism (Chang et al., 2003; Chen et al., 2002; Guo et al., 2005). In human, Nogo-B receptor (NgBR) was found to interact with DHDDS to enhance its protein stability and promote Dol-P production (Harrison et al., 2011). Catalytically inactive NgBR having amino acid sequence homology

doi:10.6342/NTU202502385

with DHDDS (Figure 1), but without the catalytically essential residues, can form a complex with DHDDS to complete the active site by providing the C-terminal RXG motif (Grabińska et al., 2017). Later, the complexed crystal structure of recombinant human NgBR$^{79-293}$/DHDDS was solved to reveal a heterodimer (Edani et al., 2020), but a heterotetrameric (dimer of the heterodimer via the extended DHDDS C-terminus) assembly was found in solution to enhance the DHDDS activity by 400-fold (Bar-El et al., 2020; Giladi et al., 2022). Yeast Nus1, a NgBR homologue, forms a homodimer as revealed by its crystal structure (Ma et al., 2019). This Nus1 protein was modeled for binding with Rer2 to form a heterodimer (Lisnyansky et al., 2019), which is one of the two DHDDS-like enzymes Rer2 and Srt1 in yeast (Shimizu et al., 1998), but the complexed structure and whether Nus1 could stimulate Rer2 activity were not investigated.

Protein glycosylation is a critical post-translational modification, which plays essential roles in protein folding, oligomerization, quality control, sorting, and transportation (Helenius & Aebi, 2001). Mutations on DHDDS or NgBR have been found associated with the human diseases resulted from glycosylation disorders (Ng & Freeze, 2018). Unlike animals that have only one pair of DHDDS/NgBR, and yeast has two DHDDS-like enzymes Rer2 and Srt1 as well as a NgBR-like protein Nus1, plants have multiple DHDDS-like and NgBR-like proteins with sequence homology. As demonstrated in this study, we chose *Cinnamomum kanehirae*, a tree that contains two cis-PTs and three NgBR-like proteins based on genomic sequences, to study how they form complexes, the paired and unpaired enzyme activities, and quaternary compositions. Characterization of these cis-PT homologues in *C. kanehirae* reveals the mechanisms for biosynthesis of dolichol and polyprenol in the plant. Moreover, by comparing their motif sequences, we propose the evolutionary paths for these cis-PT homologues.

2 | RESULTS

2.1 | Multiple sequence alignment of cis-PT homologues

The amino acid sequences of five *C. kanehirae* cis-PT homologues, including RWR89830.1, RWR78504.1, RWR97124.1, RWR71884.1, and RWR93565.1, hereafter shorten as RWR89, RWR78, RWR97, RWR71, and RWR93, respectively, were obtained from GenBank. From their phylogenic tree and domain analysis as shown in Figure S2, RWR89 and RWR78 are bacterial UPPS-like protein and human DHDDS-like protein, respectively, but RWR97, RWR71, and RWR93 are homologous to NgBR. While aligning with other homomeric and heteromeric cis-PT homologues from bacteria, yeast, and human (Figure 1), these proteins have well-conserved amino acid sequences in the five regions. However, several key residues that are involved in catalysis and substrate binding in region I and III are absent in NgBR homologues. Moreover, the RXG motif in the C-terminus of human NgBR, which is also conserved in homomeric cis-PTs, important for activity by interacting with the pyrophosphates of IPP and FPP via Mg^{2+} ion (Grabińska et al., 2017), is replaced by NXG in yeast Nus1 and the two NgBR-like proteins in *C. kanehirae*. On the other hand, RWR97 thoroughly lacks region IV and V, which had been characterized as dimeric interface of homomeric and heteromeric cis-PT homologues, indicating its disability to interact with DHDDS homologues, so that it was ignored in this study.

2.2 | Expression, purification, and characterization of recombinant RWR89

To characterize two *C. kanehirae* cis-PTs, they were initially tried for expression in *E. coli*. Although human DHDDS was previously successfully expressed in *E. coli* (Edri et al., 2017; Giladi et al., 2017), only RWR89 was soluble when expressed in *E. coli*. Since the recombinant protein contained *N*-terminal thioredoxin and His-tag with a TEV protease site, it was purified by Talon Metal Affinity Resin, followed by TEV tag cleavage to remove the tags, and size-exclusion chromatography to yield untagged purified RWR89. The protein yields for RWR89 from 2-L culture after the affinity chromatography and the size-exclusion chromatography were 16.9 and 2.4 mg, respectively. Through the experiments conducted at 4°C or room temperature as described below, the protein was stable without precipitation and the activity was not decreased. On the contrary, other proteins when expressed in *E. coli*, showed visible precipitation, so we decided to use yeast to express these proteins as described below. The SDS-PAGE analysis of the *E. coli*-expressed RWR89 after size-exclusion chromatography is shown in Figure 2a left panel and it was judged as a homodimer (75.4 kDa) composed of two 41.4 kDa monomers, based on the elution volume compared to those of molecular weight standards (Figure S3).

The purified RWR89 was subjected to kinetic characterization by trying different allylic substrate DMAPP, GPP, FPP, or GGPP, with the homoallylic substrate IPP. A coupling assay kit to convert the RWR89-catalyzed inorganic diphosphate (PPi) into a product that absorbs light at 360 nm (see Section 4) was used. We found that only GPP, FPP, and GGPP served as active allylic

doi:10.6342/NTU202502385

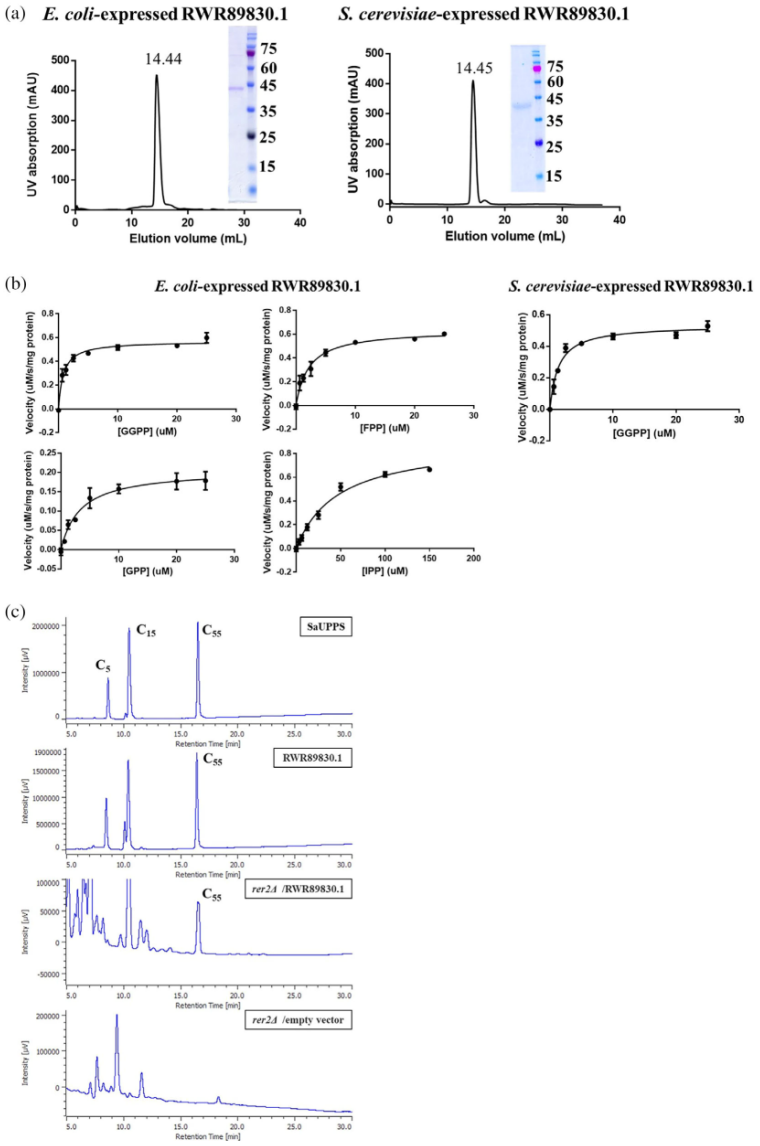


FIGURE 2 Purification and characterization of the *E. coli* and yeast-expressed recombinant RWR89. (a) SDS-PAGE analysis of the *E. coli* and yeast-expressed RWR89, purified from size-exclusion chromatography as a homodimer (75.4 and 75.1 kDa, respectively, based on the elution volume). (b) Michaelis–Menten plots of *E. coli*-expressed RWR89 with different concentrations of FPP, GGPP, or GPP and 50 μM IPP to determine the K_m and k_{cat} values (K_m for IPP was measured with 10 μM GGPP). For comparison, the Michaelis–Menten plot of yeast-expressed RWR89 using GGPP is also shown in the right panel. All measurements were performed in triplicate at 30°C. (c) Polyisoprenoid product analysis of RWR89 incubated with IPP and FPP at 30°C overnight. From top to bottom, the product of the purified SaUPPS, the purified *E. coli*-expressed RWR89, and the polyisoprenoid products in the pYES-RWR89-transformed rer2Δ yeast cells were extracted and analyzed by HPLC to be C55, respectively, as compared to the absence of C55 in the rer2Δ cells with empty vector.

doi:10.6342/NTU202502385

TABLE 1 Kinetic parameters of the purified RWR89, RWR78, RWR78/sRWR71, and RWR78/sRWR93 with respect to GPP, FPP, or GGPP and IPP substrates.

Enzymes and substrates	K_m (μM)	k_{cat} (s^{-1})
E. coli-expressed RWR89		
IPP	47.24 ± 4.21	0.46 ± 0.02
GPP	3.38 ± 0.45	0.11 ± 0.01
FPP	2.17 ± 0.2	0.33 ± 0.01
GGPP	0.80 ± 0.10	0.57 ± 0.01
S. cerevisiae-expressed RWR89		
GGPP	1.36 ± 0.15	0.53 ± 0.01
RWR78		
IPP	10.64 ± 4.73	0.004 ± 0.0005
GPP	4.85 ± 1.31	0.0005 ± 0.00005
FPP	0.50 ± 0.21	0.002 ± 0.0002
GGPP	2.06 ± 0.62	0.0007 ± 0.00006
RWR78/sRWR71		
IPP	9.52 ± 1.29	0.0628 ± 0.0021
GPP	1.25 ± 0.23	0.0206 ± 0.0008
FPP	0.66 ± 0.12	0.0562 ± 0.0018
GGPP	0.74 ± 0.18	0.0496 ± 0.0025
RWR78/sRWR93		
IPP	16.78 ± 3.72	0.1112 ± 0.0072
GPP	1.78 ± 0.17	0.0383 ± 0.0009
FPP	0.36 ± 0.10	0.1083 ± 0.004
GGPP	0.74 ± 0.15	0.1004 ± 0.0039

Abbreviations: FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; GPP, geranyl diphosphate; IPP, isopentenyl diphosphate.

substrates. By monitoring the absorbance changes at 360 nm using a microplate reader, the K_m for using GPP, FPP, and GGPP were measured to be 3.38 ± 0.45, 2.17 ± 0.2, and 0.80 ± 0.10 μM, and the k_{cat} 0.11 ± 0.01, 0.33 ± 0.01, and 0.57 ± 0.01 s^{-1}, respectively (Figure 2b; Table 1). Apparently, GGPP was the best allylic substrate with a smaller K_m and a larger k_{cat}, unlike that FPP was the best substrate for bacterial UPPS. The K_m and k_{cat} with respect to IPP while using GGPP were 47.24 ± 4.21 μM and 0.46 ± 0.02 s^{-1}, respectively, indicating a poor binding for IPP (Figure 2b; Table 1). The product was formed by incubating FPP and IPP with RWR89 and its PPi group was removed by phosphatase and then subjected to HPLC analysis. By spiking the purchased polyprenols with known chain lengths, chain length of the product was determined to be C55, eluted at the same position as that of the C55 product generated by *Staphylococcus aureus* UPPS (SaUPPS) (Figure 2c top two panels).

For comparison, His-tagged RWR89 was also expressed in *Saccharomyces cerevisiae* INVSc1 cells and purified using Ni-NTA resin. The elution volume of the purified RWR89 from the size-exclusion chromatography was 24.45 mL, indicating the molecular weight of 75.1 kDa, which was also judged as a homodimer (Figure 2a, right panel). The Michaelis–Menten plot of using the best substrate GGPP is also shown in Figure 2b (the right panel) to yield K_m of 1.36 ± 0.15 μM and k_{cat} of 0.53 ± 0.01 s^{-1}, similar to that of *E. coli*-expressed RWR89. The product obtained from the lipid extract of the pYES2-RWR89-transformed rer2-deletion (rer2Δ) yeast cells was the same as the *E. coli*-expressed RWR89 (C55 as shown in Figure 2c bottom two panels).

2.3 | Expression, purification, and characterization of recombinant RWR78, sRWR71, sRWR93 and their heteromeric complexes

Two NgBR homologues RWR71 and RWR93 were predicted to have a transmembrane domain (TM) (Figure S2b). To obtain soluble proteins, RWR71 was expressed without the transmembrane residues 1–127 (named sRWR71), while sRWR93 lacked the transmembrane residues 1–73. The full-length RWR78 along with sRWR71 or sRWR93 were first respectively co-expressed in *E. coli*. However, the purified complexes failed to show enzymatic activity and easily precipitated after freezing and thawing. To obtain the functional complexes, the codon-optimized gene of N-terminal His-tagged RWR78 and the gene of N-terminal Strep-tagged sRWR71 or sRWR93 were respectively constructed into pESC-URA bicistronic vector and co-overexpressed in *S. cerevisiae* INVSc1 cells through galactose induction. The Strep-tagged sRWR71 or sRWR93 was co-purified with the His-tagged RWR78 by Ni-NTA resin, indicating that they formed complexes. From the following size-exclusion chromatography, both RWR78/sRWR71 and RWR78/sRWR93 complexes were eluted in a monodisperse peak, and the corresponding molecular weights were 75.4 and 74.8 kDa, respectively, as a heterodimer composed of a His-tagged RWR78 (40.7 kDa) and a Strep-tagged sRWR71 (24.2 kDa) or sRWR93 (24.3 kDa), unlike the *E. coli* co-expressed and co-purified complex of human DHDDS and the cytosolic part of NgBR (residues 73–293, named sNgBR) forming a heterotetrameric complex in solution, as judged from the elution volume from the size-exclusion column (Figure S4a). SDS-PAGE analysis of the complexes revealed the two protein components in roughly 1:1 ratio (Figure 3a). His-tagged RWR78 alone

doi:10.6342/NTU202502385

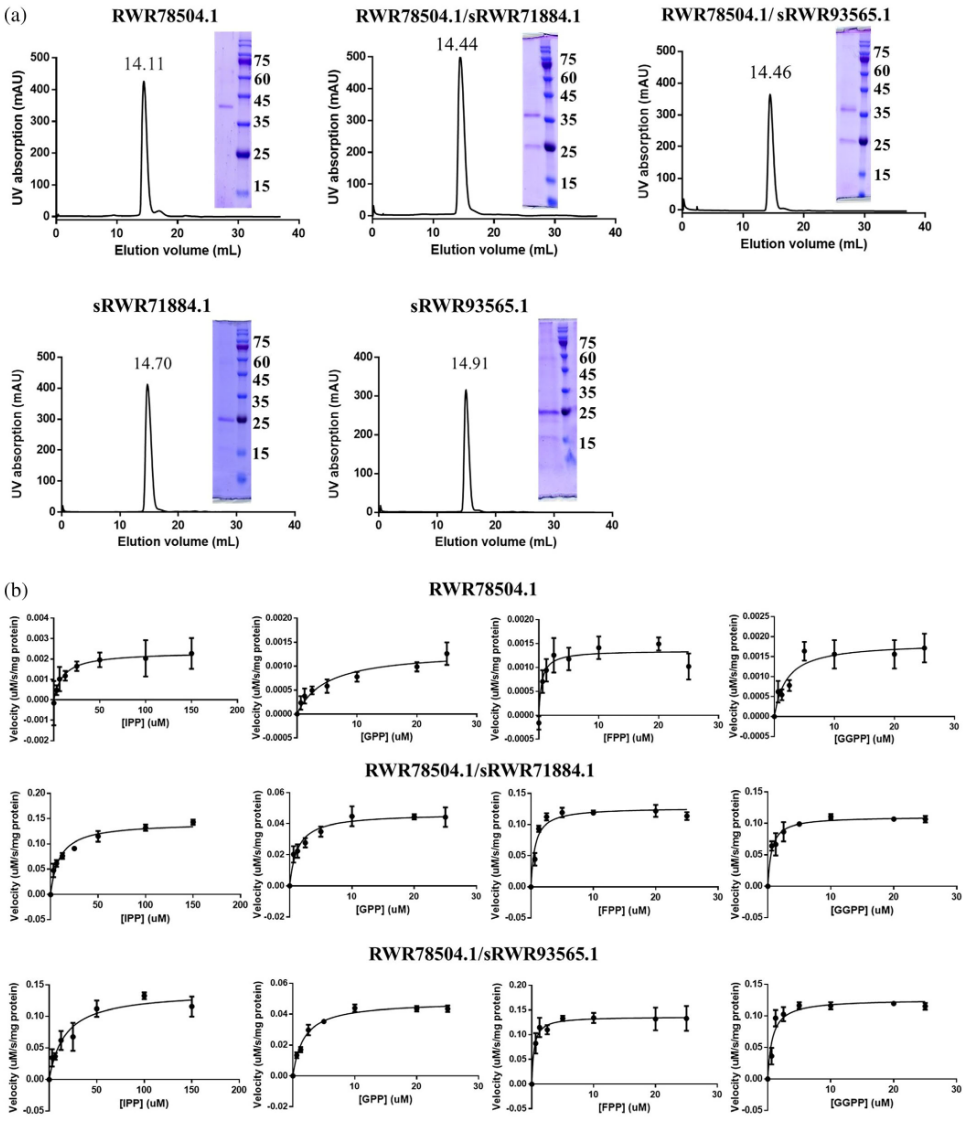


FIGURE 3 Purification and characterization of the recombinant RWR78, RWR78/sRWR71, RWR78/sRWR93, sRWR71, and sRWR93. (a) SDS-PAGE analysis of RWR78, RWR78/sRWR71, RWR78/sRWR93, sRWR71, and sRWR93, purified from the size-exclusion chromatography as homodimer (83.6 kDa), heterodimer (75.4 kDa), heterodimer (74.8 kDa), homodimer (67.3 kDa), and homodimer (61.4 kDa), respectively, based on their elution volumes. (b) Michaelis–Menten plots of RWR78, RWR78/sRWR71, and RWR78/sRWR93 using IPP and GPP, FPP, or GGPP as substrates (from left to right panels) to determine the K_m and k_{cat} values as listed in Table 1. All measurements were performed in triplicate at 30°C.

doi:10.6342/NTU202502385

expressed in *S. cerevisiae* and purified via Ni-NTA resin and size-exclusion chromatography was a homodimer (83.6 kDa) (Figure 3a), the same as the reported human DHDDS conformation (Lisnyansky et al., 2019) (we also confirmed that *E. coli*-expressed human DHDDS formed a homodimer as shown in Figure S4b). Besides, while using StrepTrap XT column and size-exclusion chromatography to purify the *S. cerevisiae*-expressed sRWR71 or sRWR93, they formed a homodimer (67.3 and 61.4 kDa, respectively), similar to yeast Nus1 (Ma et al., 2019), but unlike the human sNgBR that forms a monomer (Holcomb et al., 2018). However, by mixing the purified homodimeric RWR78 with homodimeric sRWR71 or sRWR93, they failed to form a complex.

The enzymatic activities of purified homodimeric RWR78 and its heterodimeric complex with sRWR71 or sRWR93 were subsequently characterized via pyrophosphate assay kit. The K_m values for FPP were measured with 50 μM IPP, while the K_m values for IPP were measured with 10 μM FPP. The Michaelis–Menten plots are shown in Figure 3b, and the kinetic constants are summarized in Table 1. Compared to the activities of His-tagged RWR78 and the complexes, the K_m of FPP and IPP were similar, but the k_{cat} values of the complexes were 14–50-fold higher than that of RWR78, indicating stimulation of RWR78 activity by sRWR71 or sRWR93, with a higher stimulating effect by sRWR93. The Michaelis–Menten plots for using GPP and GGPP as alternative substrates for RWR78, RWR78/sRWR71, and RWR78/sRWR93 were also measured (Figure 3b and Table 1). We found that the enzyme and its complexes preferred FPP as substrate.

To test whether the bacterial UPPS-like RWR89 could also form a heteromeric complex with sRWR71 or sRWR93, we added the yeast-expressed sRWR71 or sRWR93, to the above *E. coli*-expressed RWR89 or the yeast-expressed RWR89, but the activity of RWR89 was not changed (data not shown). We then co-expressed His-tagged RWR89 and Strep-tagged sRWR71 or sRWR93 in yeast, just like what we did for co-expressing His-tagged RWR78 and Strep-tagged sRWR71 or sRWR93, but they were not co-purified from Ni-NTA column (figure not shown), indicating no complexation, consistent with no activity stimulation of RWR89 by sRWR71 or sRWR93. Taken together, RWR89 only formed a homodimer no matter using *E. coli* or yeast to express it.

2.4 | Functional analysis of two complexes in rer2-deletion *S. cerevisiae*

To determine the function of two complexes, either complex was co-expressed in a rer2Δ mutant yeast lacking the Rer2 cis-PT. Compared to the rer2Δ yeast that did not grow well at a higher temperature (37°C), rer2Δ yeast co-expressing either complex (RWR78/RWR71 or RWR78/RWR93) or human DHDDS regained the growth ability at 37°C (Figure 4a). Using human DHDDS alone was able to complement Rer2 function as previously observed (Sabry et al., 2016; Yao et al., 2019). However, when independently transforming only RWR78, RWR71, or RWR93 into rer2Δ strain, cells failed to grow at 37°C, indicating that either single subunit did not have the function and complexation was strictly required for producing dolichols. By examining the extracted lipid mixtures from the rer2Δ yeast co-expressing either complex, the HPLC analysis revealed a similar composition of long-chain lipids, while long-chain products were not observed in the lipids extracted from RWR78-transformed rer2Δ yeast (Figure 4b).

2.5 | Homology modeling of RWR89

To elucidate the catalytic mechanism of RWR89 that produced C55 product (Figure 2c), the homology-based structure prediction of RWR89 was generated by using the *S. aureus* UPPS structure (PDB: 4H8E), the closest homologue in term of conserved amino acids, as the template. Based on the experimental data showing that RWR89 was a homodimer (Figure 2a), its 3-D structure was constructed by modeling to reveal the canonical catalytic domain (Figure 5a). Because *E. coli* UPPS was well studied by us, we chose its crystal structure to superimpose with the RWR89 modeled structure to interpret the roles of active-site amino acids.

With IPP as the building blocks in consecutive cis-condensations, all three trans-prenyl diphosphates (GPP, FPP, and GGPP) served as substrates for RWR89. However, RWR89 could not accept DMAPP as either a primary allylic substrate or condensational units, indicating RWR89 is unlikely to produce short-prenyl diphosphates such as neryl diphosphate (NPP) or nerylneryl diphosphate (NNPP) as SlcPT1, a neryl diphosphate synthase in *Solanum lycopersicum*, prefers DMAPP as the allylic substrate and produces NPP (Kang et al., 2014). Gly144 and Trp322 of RWR89 in Figure 5a are aligned with Gly46 and Trp221 of EcUPPS, but not Ile106 and Phe276 in SlcPT1, respectively, indicating RWR89 is more similar to EcUPPS to accept longer trans-prenyl diphosphates as substrates, with GGPP the most suitable. Leu235 of RWR89 is at an identical position of the aligned Leu137 of EcUPPS, which was characterized to block further chain length extension and regulate the product chain length (Ko et al., 2001).

doi:10.6342/NTU202502385

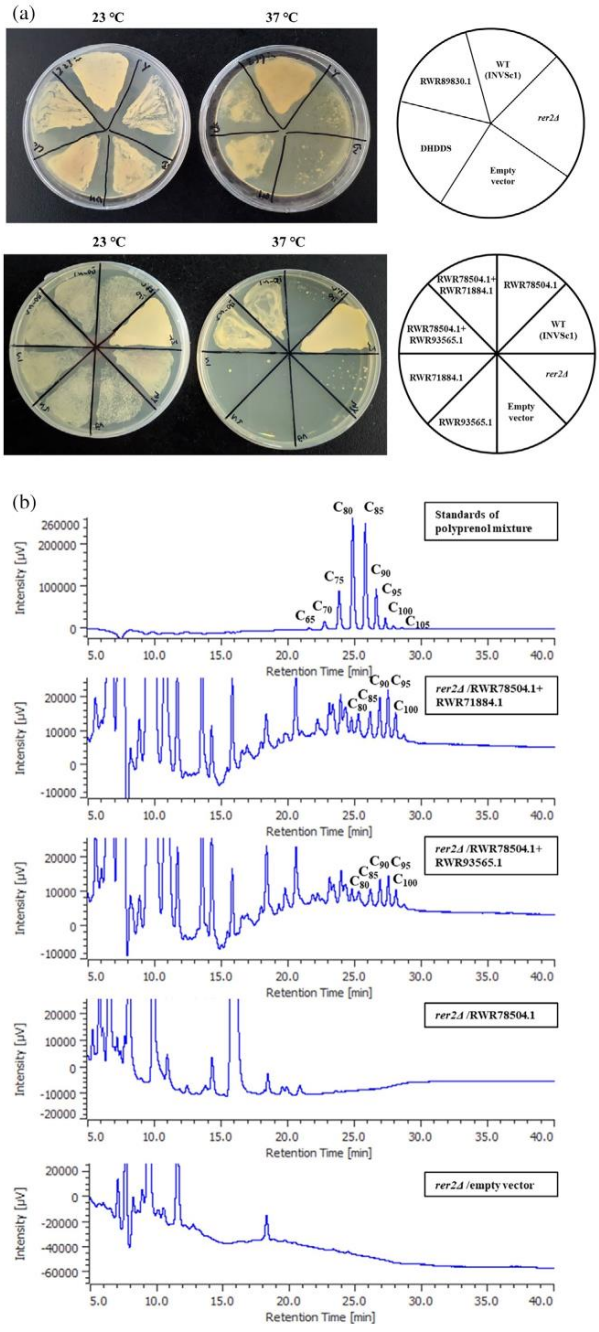


FIGURE 4 Legend on next page.

doi:10.6342/NTU202502385

Regarding why RWR89 not forming heteromeric complex with sRWR71 or sRWR93 as showed above, RWR89 already has the amino acids such as the C-terminal RXG (RFG) self-sufficient to form an independent active site, without need of the amino acids contributed by sRWR71 or sRWR93, but RWR78 required the extra amino acids from sRWR71 or sRWR93 as shown below.

2.6 | Homology modeling of RWR78/RWR71 and RWR78/RWR93 heterodimeric complexes

To understand the catalytic roles of amino acids, homology modeling of RWR78/RWR71 and RWR78/RWR93 were built by using the crystal structure of human DHDDS/NgBR complex (PDB: 6W2L) as the template. Both RWR78/RWR71 and RWR78/RWR93 were constructed as heterodimer (Figure 5b) based on the experimental observation (Figure 3a). According to the expanded view of the active sites on the right-hand side, the residues involved in the canonical cis-PT catalysis are quite conserved. The S1 substrate binding site of RWR78 encompasses Arg98, Arg99, and Arg146, while S2 site contains Arg289 and Arg295. The missing catalytic residues are provided by the C-terminal NXG motif (NYG), that is Asn326 sidechain N and Gly328 backbone N or Asn272 sidechain N and Gly274 backbone N from RWR71 or RWR93, to interact with the β-phosphate of IPP at the S2 site, in contrast to the RXG motif (RLG) of NgBR for binding substrate and accelerating DHDDS catalytic activity. On the other hand, the homodimeric RWR89 has already the C-terminal RXG (RFG) so that it does not require a partner to complete the active site.

Some differences can be found between the human DHDDS/NgBR and the two complexes. In NgBR, Arg290 of the conserved C-terminal RLG motif was previously shown to play a critical role to enhance the catalytic activity of the heteromeric DHDDS/NgBR complex (Grabińska et al., 2017), and a mutation R290H is related to congenital glycosylation disorder (Park et al., 2014). Indeed, the location of Arg290 on a surface loop may provide NgBR with the structural flexibility needed for interacting with the catalytic pocket and, in turn, to allosterically activate DHDDS. In comparison with the RLG motif in NgBR, the shorter side chain of Asn in NYG motif of RWR71 and RWR93 is needed to maintain the close contact with diphosphate moiety of IPP. Moreover, as revealed by comparing Figure 5b,c, Arg98, Ile103, and Arg295 of RWR78 are aligned with Arg37, Lys42, and Arg211 of DHDDS, respectively, which are all important because the R37H, K42E, and R211Q mutations cause disruption of substrate binding, thereby significantly decreasing the cis-PT activity (Bar-El et al., 2020; Edani et al., 2020). Arg37 and Arg211, related to developmental epileptic encephalopathies (Hamdan et al., 2017), are all positively charged residues in close proximity to the active site of human DHDDS. Lys42 forms a salt bridge with Glu234, but in the retinitis pigmentosa disease-associated K42E mutant DHDDS (Zelinger et al., 2011; Züchner et al., 2011), it was predicted to interact with Arg38 instead, causing instability of the active site (Bar-El et al., 2020). However, in RWR78, Ile103 could not form a salt bridge with any residue, making no contribution on the active-site stabilization.

To further understand the mechanism of forming RWR78/RWR71 and RWR78/RWR93 heterodimers in contrast to heterotetramer of DHDDS/NgBR in solution, their interfaces were investigated. As shown in Figure 6a, hydrogen-bond pairs T220/E257, Q228/E252, Q228/F255, Y231/A233, T294/L299, R295/E298, S297/R295, N298/R295, W302/Y226, T305/L287, T305/G288, and L309/R284 are formed in the interface of RWR78/RWR71; while in the interface of RWR78/RWR93 complex (Figure 6b), there are polar interactions including T220/E203, Q228/F201, Y231/A179, K234/C183, Q235/N197, E264/S180, T294/L245, R295/E244, S297/R241, N298/R241, T305/L233, T305/G234, and L309/R230. Not only sharing high sequence identity of 87% between RWR71 (residues 141–329) and RWR93 (residues 92–275), their structural models also show high similarity (RMSD = 0.282 Å), supporting that either RWR71 or RWR93 could form a complex with RWR78. In contrast, the extended C-terminal helix-turn-helix motif (shown in orange in Figure S5) of DHDDS in the heterotetrameric DHDDS/NgBR complex (PDB: 6Z1N) forms additional interactions with that of the other DHDDS molecule to promote heterotetramerization with NgBR (Bar-El

FIGURE 4 Analysis of complementation of RWR89, RWR78/RWR71, and RWR78/RWR93 to yeast rer2. (a) The rer2Δ yeast and that transformed with the empty vector pYES2, but not the wild type (WT) yeast strain INVSc1 showed growth defect at higher temperature 37°C. Expression of RWR78/RWR71, RWR78/RWR93, or human DHDDS in the rer2Δ yeast rescued the growth defect at 37°C, while expression of individual RWR78, RWR71, or RWR93 did not complement the Rer2 function. (b) Polyisoprenoid products in the rer2Δ yeast cells transformed with empty vector, pESC-RWR78, pESC-RWR78/RWR71, and pESC-RWR78/RWR93, respectively, analyzed by HPLC, indicating that only RWR78/RWR71 and RWR78/RWR93 produced longer C80–C100 products.

doi:10.6342/NTU202502385

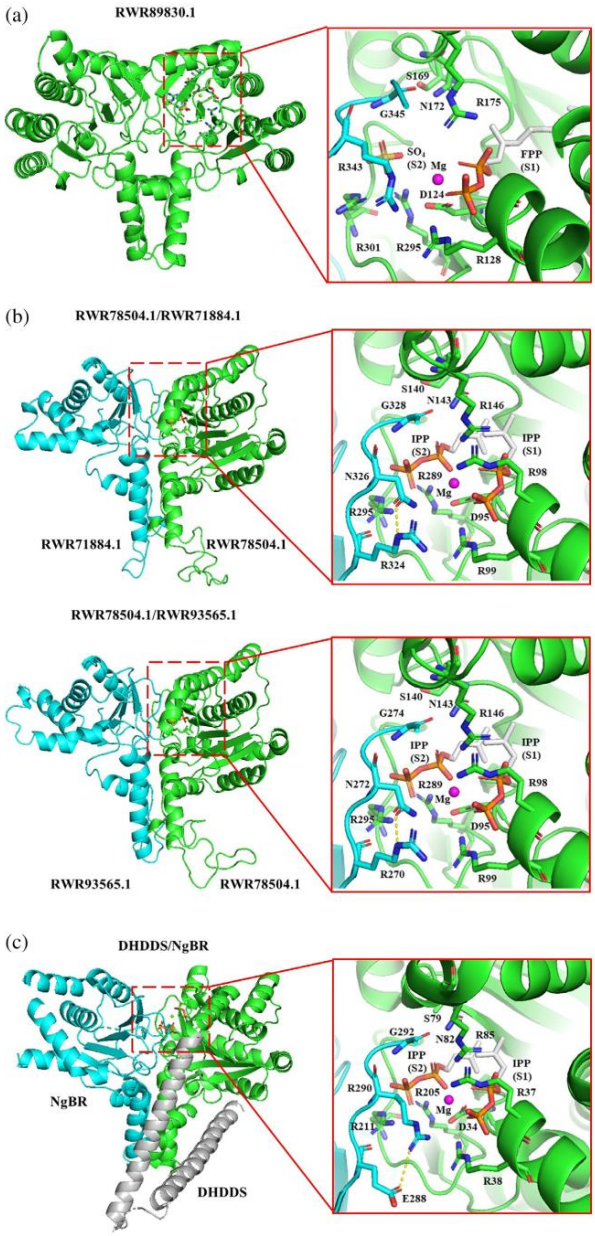


FIGURE 5 Homology modeling of RWR89, RWR78/RWR71 and RWR78/RWR93 structures. (a) The structural model and the detailed active-site architectures of RWR89. The ligand and the metal ion Mg^{2+} are originated from the superimposed crystal structure of *E. coli* UPPS. (b) The structural models and the active-site architectures of RWR78/RWR71 and RWR78/RWR93 using human DHDDS/NgBR complex (PDB: 6W2L) as template. RWR78 is colored in green, while RWR71 and RWR93 are colored in cyan. The ligands 2 IPP and the metal ion Mg^{2+} are originated from the superimposed crystal structure of human DHDDS/NgBR. (C) For comparison, the DHDDS structure and the active-site architecture are shown. NgBR is colored in cyan, DHDDS is colored in green, and the C-terminal helix-turn-helix motif of DHDDS is colored in gray.

doi:10.6342/NTU202502385

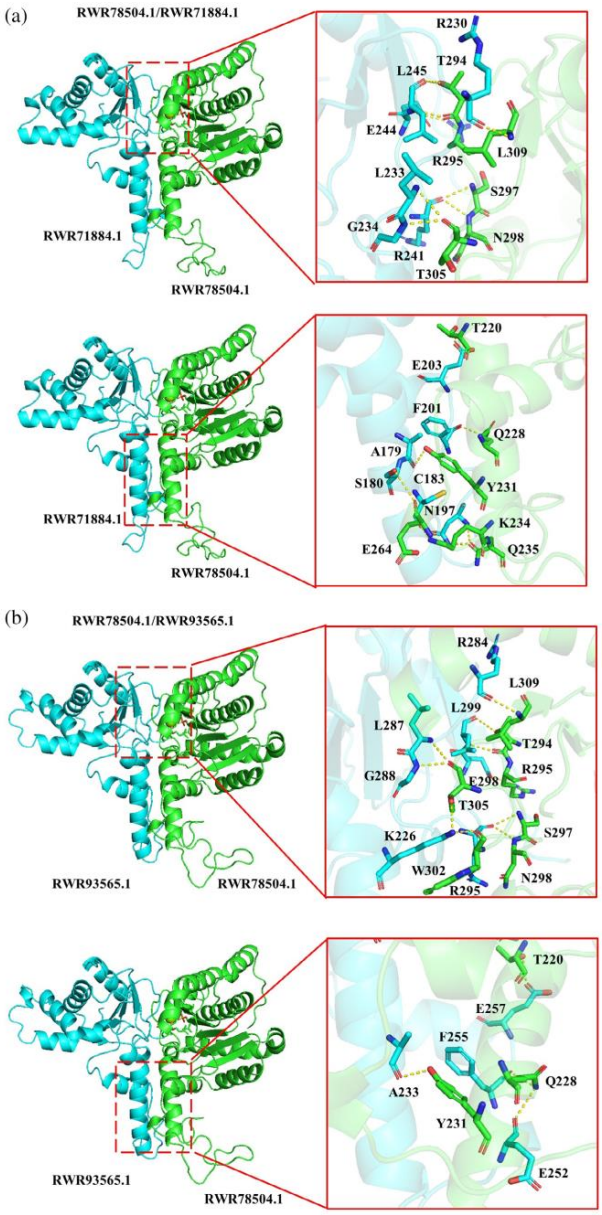


FIGURE 6 The subunit interfaces of heterodimeric RWR78/RWR71 and RWR78/RWR93 in comparison with that of heterotetrameric human DHDDS/NgBR. (a) Key residues located at the interface of RWR78 and RWR71. (b) Key residues located at the interface of RWR78 and RWR93. Residues from RWR78 are colored in green, while residues from RWR71 and RWR93 are colored in cyan.

et al., 2020). Due to lack of the additional C-terminal motif, RWR78 could not form a heterotetramer with RWR71 or RWR93 (see Figure 3a).

Regarding why a DHDDS-like protein rather forming heteromeric complex with a NgBR-like protein than forming a homodimer itself, the homodimeric structures

doi:10.6342/NTU202502385

of DHDDS and RWR78 were predicted using Alphafold2 algorithm (Figure S6), and the homodimeric interfaces were compared to the heterodimeric interfaces. In regions VI and V of the predicted DHDDS homodimer, there are only few hydrogen bonds at the interface, including Arg159 with Trp218, Gln219, and Ser221 of the other subunit in region VI, as well as V210/L225, S213/S221, and D214/S221 hydrogen bonds in region V. In comparison, the heterodimeric interface of DHDDS/NgBR complex has more hydrogen bonds, including Arg159 of DHDDS with Trp256 and Asp237 of NgBR in region VI, as well as V210/I263, R211/E262, S213/R259, D214/R259, S221/L251, S221/G252, S221/K297, W218/K297, and L225/D248 in region V. At the predicted RWR78 homodimeric interface, there is no hydrogen bonding in region IV and only few H-bond pairs in region V, including T294/L309, S297/T305, and N298/T305. In comparison, either the heterodimeric interface of RWR78/RWR71 or RWR78/RWR93 has more interactions such as Arg295/Glu298 or Arg295/Glu244 than homodimeric RWR78. Therefore, NgBR-like proteins are regulators of catalytic activity of DHDDS-like proteins. However, unlike monomeric NgBR, RWR71 and RWR93 were both dimeric, so could not dissociate into monomeric to form heterotetramer with homodimeric RWR78 in solution.

2.7 | Possible evolutionary paths for *C. kanehirae* cis-PTs and homologues

Prokaryotes only contain homodimeric cis-PTs, while eukaryotes and archaea have heteromeric cis-PTs, but plural DHDDS-like and NgBR-like proteins are particularly prevalent in plants. Since we have characterized the functions of *C. kanehirae* cis-PT homologues as shown here and their amino acid sequences are known for comparison with other known cis-PT homologues, we could trace their possible evolutionary paths. Based on the phylogenetic tree of the cis-PT homologues from archaea, bacteria, and eukaryotes (Figure S2), we identified their common motifs and analyzed the distribution and diversity of these motifs as shown by various colored boxes (Figure 7). Accordingly, these cis-PT homologues from archaea, bacteria, human, plant, and yeast could be divided into three subgroups, including homomeric UPPS orthologs, heteromeric DHDDS orthologs, and heteromeric NgBR orthologs (for convenience, we use here UPPS, DHDDS, and NgBR to represent the homodimeric cis-PTs, active cis-PTs of the heteromeric complexes, and the inactive cis-PT homologues of the heteromeric complexes, respectively). The plantae, including green algae, hornworts, ferns, pines, monocots, and dicots, contain all three types of orthologs. The plant-specific homomeric UPPS orthologs could evolve from prokaryotic homomeric UPPS orthologs because they share the same motif architectures (see UPPS orthologs in Figure 7), but with additional N-terminal sequences as signal peptides for transporting toward specific organelles including chloroplast and mitochondria. For example, SlCPT1-7 except SlCPT3 were examined to be localized in plastids (Akhtar et al., 2013), and ChrUPPS from *Chlamydomonas reinhardtii*, a well-studied unicellular green alga, was predicted to be localized in mitochondria, because they possess the directed signal peptides.

In contrast to plants and fungi, animals have one DHDDS and one NgBR for only dolichol biosynthesis. DHDDS orthologs from animals specifically have the extended C-terminal motif compared with DHDDS orthologs from plants. However, NgBR orthologs from animals prefer to have a C-terminal RXG motif, also observed in homomeric prokaryotic UPPS orthologs and plant UPPS-like cis-PTs, but both animal and plant DHDDS orthologs lack RXG or NXG motif (Figure 7). On the contrary, plant NgBR orthologs have NXG but not RXG motif at C-termini (Figure 7). The NXG and RXG motifs might originate from different evolutionary paths. We found that *Candidatus megaira*, a bacterium regarded as an endosymbiont of *Mesostigma viride*, a single-celled charophyte algae considered as one of the earliest diverging members of green plant lineages, contains a homomeric UPPS gene (GenBank: UCM94414.1), which has the C-terminal NXG (NFG) motif instead of RXG motif. Moreover, some NgBR orthologs from fungi also possess NXG motif, and fungi usually have symbiotic relationships with plants. Therefore, the NXG motif of plant NgBR orthologs was probably obtained from certain bacteria or fungi through symbiosis and horizontal gene transfer, while the RXG motif of animal NgBR orthologs might be derived from those typical prokaryotic homomeric UPPS orthologs.

In archaea, their cis-PT homologues could be also divided into UPPS, DHDDS, and NgBR orthologs, although archaea cells do not possess organelles such as ER and chloroplast. Among the three subgroups in the phylogenetic tree, archaeal cis-PT homologues all belong to the basal clades, the earliest diverging branches with their sequences phylogenetically apart from the other cis-PT homologues. These archaeal cis-PT homologues may have unique structural or functional characteristics. For example, MA1831 from *Methanosarcina acetivorans* and MM0014 from *Methanosarcina mazei* are UPPS orthologs, but their functions differ from UPPS, and both of them do not possess the common C-terminal RXG motif (Figure 7). MA1831 could catalyze non-canonical head-to-middle condensation and produce C35–C50

doi:10.6342/NTU202502385

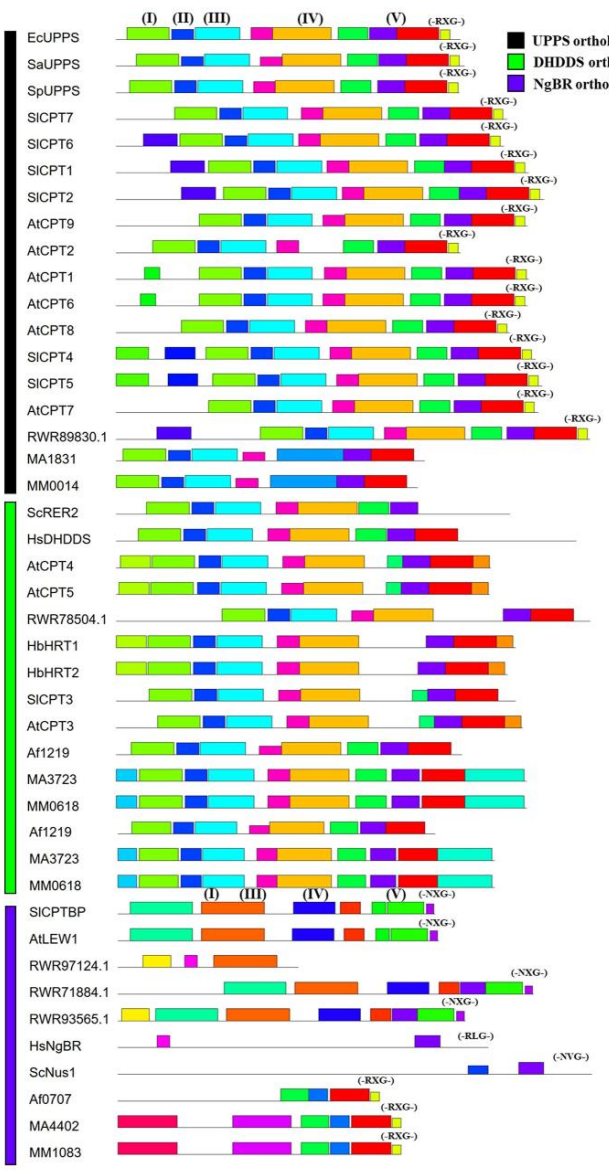


FIGURE 7 Motif analysis of cis-PTs and homologues. Among UPPS orthologs, DHDDS orthologs, and NgBR orthologs, 30 motifs expressed in various colored boxes in five conserved regions (I–V) were identified based on the conserved amino acid seuqneces, analyzed by using the MEME online tool (https://meme-suite.org/meme/tools/meme, version 5.5.2). RXG motif was found at the C-termini of prokaryotic cis-PTs, plant UPPS-like cis-PTs, archaeal NgBR orthologs, and *Homo sapien* NgBR (HsNgBR), but not in DHDDS orthologs. On the other hand, NXG motif was found in plant NgBR orthologs. The abbreviations for the species followed that in Figure S2a.

products, while MM0014 could yield C25–C40 products and have O-prenyltransferase activity (Ogawa et al., 2016; Okada et al., 2021). Besides, the heteromeric cis-PT homologues from *Archaeoglobus fulgidus*, a thermophilic archaeon, were recently biochemically characterized (Sompiyachoke et al., 2022). Unlike eukaryotic NgBR orthologs, Af0707, the NgBR ortholog from *A. fulgidus*, does not have the transmembrane domain, and the

doi:10.6342/NTU202502385

DHDDS ortholog Af1219 lacks the C-terminal extended motif. However, the complex still has a heterotetrameric assembly, which is contrary to their structural heterodimeric complex prediction. To investigate the possible origin of heteromeric cis-PTs from archaea, cis-PT homologues from *Candidatus Lokiarchaeota archaeon*, an archaeal species from Lokiarchaeota and one of the nearest relatives of eukaryotes, were also investigated. We found one cis-PT, called CLaUPPS3 (GenBank: JAGXOB010000163.1), belonging to DHDDS subgroup, while no NgBR ortholog was detected. However, a UPPS-like cis-PT from *C. Lokiarchaeota archaeon*, called CLaUPPS4 (GenBank: JAGXNS010000099.1), lacks the conserved catalytic residues of cis-PTs but owns the RXG motif. We suspect that NgBR orthologs might be originated from these UPPS-like cis-PTs that lack the conserved catalytic residues. However, functions of cis-PTs from archaea remain large unknown and further studies are needed to elucidate their evolutionary relationships.

3 | DISCUSSION

In eukaryotes, dolichols of C70–120 chain lengths play indispensable roles as glycosyl carrier lipids for the biosynthesis of glycoproteins in ER. In human cells, NgBR is required to form a complex and boost the DHDDS activity, and in yeast Nus1 complexes with Rer2 and probably also Srt1. In plants, complexation between DHDDS-like and NgBR-like cis-PT homologues is also required for natural rubber biosynthesis in Lettuce and dandelion, and dolichol biosynthesis in tomato (Brasher et al., 2015; Epping et al., 2015; Qu et al., 2015; Surmacz et al., 2014), but in bacteria only homomeric cis-PTs exist, demonstrating a major difference in the compositions of active cis-PTs in prokaryotes and eukaryotes. However, heteromeric cis-PTs were also found in methanogenic archaeon, such as *M. acetivorans* that contains a UPPS (MA3723) and a NgBR/Nus1-like protein (MA4402) (Grabińska et al., 2017; Ogawa et al., 2016) as well as *M. mazei* that utilizes a heteromeric cis-prenyltransferase for the biosynthesis of glycosyl carrier lipids (Emi et al., 2019). Here, we provide an example of multiple unpaired and paired cis-PT homologues in a plant *C. kanehirae* and demonstrate their kinetics and functions.

Two NgBR homologues RWR71 and RWR93 in *C. kanehirae* utilize a C-terminal NYG motif to compensate the missing motif in the DHDDS homologue RWR78 by complexation for making dolichols. However, a C-terminal RFG motif was found in homodimeric RWR89, bacterial SaUPPS and EcUPPS, and a plant AtcPT1 (Surowiecki et al., 2019), one of the cis-PTs in *Arabidopsis thaliana*, for synthesizing medium-chain polyisoprenoids without need of pairing with a NgBR homologue. Based on the crystal structure of *E. coli* UPPS, Arg242 of RFG motif is involved in the binding of the diphosphate group of IPP (Grabińska et al., 2017). Furthermore, the role of C-terminus in IPP binding and catalysis is supported by structural information obtained from a decaprenyl diphosphate synthase of *Mycobacterium tuberculosis* (Wang et al., 2008) and even a short FPP producing enzyme cis-FPPS of *Solanum habrochaites* (Chan et al., 2017). On the other hand, NgBR has RLG as the C-terminal RXG motif (Figure 1). It is likely that Arg290 of NgBR RLG is involved in IPP binding and its mutation to His was found in the patients (Park et al., 2014). The Asn of NYG in RWR71 (Asn272) or RWR83 (Asn326) has a shorter side chain than Arg to ensure a suitable distance between Asn side-chain N and IPP β-phosphate for a H-bond (see the modeling structures of RWR78/RWR71 and RWR78/RWR93 in Figure 5a,b). Moreover, the Arg290 of RLG in NgBR forms a salt bridge with Glu288 of NgBR, but at the position of Glu288, there is Arg270 in RWR71 or Arg324 in RWR93, so if the Asn272 or Asn326 was an Arg, there will be two positively charged Arg residues in close contact, causing electrostatic repulsion. Therefore, an Asn at NYG of RWR71 or RWR83 is ideal.

Another difference between human DHDDS/NgBR complex and RWR78/RWR71 and RWR78/RWR93 complexes is that the former is a heterotetramer but latter are heterodimers in solution. This is likely due to the extended C-terminus of DHDDS, which provides additional interaction between two DHDDS protomers (see Figure S5).

Furthermore, comparing to DHDDS, RWR89 contains a TP, a putative chloroplast transit peptide, at the N-terminus, which may be used to direct this cis-PT into a compartment (chloroplast) where C55 product is needed. Based on the RNA-seq analysis as shown in Figure S7 (Chaw et al., 2019), RWR89 had a significantly higher expression level in old leaves than in other tissues, indicating it might be involved in the plant secondary metabolism in chloroplasts. We also previously characterized a homodimeric C40/C45 producing cis-PT from the plant *Lilium longiflorum* Anther (Yao et al., 2019). On the other hand, two NgBR homologues RWR71 and RWR93 contain a TM domain to anchor them onto a cell membrane, like NgBR, a membrane protein. Whether NgBR or Nus1 serves as an accessory subunit for docking the complex to the ER membrane or is a structural component contributing to catalysis and substrate binding is still a matter of debate. Experiments show that (1) the loss of NgBR does not lead to changes in the relative amount of DHDDS associated with ER membrane

doi:10.6342/NTU202502385

fractions (Harrison et al., 2011) and (2) a mutation in the C terminus of NgBR causes a congenital disorder of glycosylation syndrome because of defective dolichol synthesis (Park et al., 2014), supporting the role of NgBR as a complement to DHDDS catalysis. Moreover, RWR78 had a highly expression level in buds of flowers and buds, RWR93 was highly expressed in fruits, and RWR71 was slightly higher expressed in fruits, buds of flowers, and buds of leaves. Flowers and fruits are the specific organs that only exist in angiosperms, and the proper protein glycosylation is indispensable for developmental processes including flowers and fruits, and inappropriate glycosylated regulation of certain proteins resulted in growth deficiencies (De Coninck et al., 2021).

Finally, the phylogenetic tree and motif structure analysis of cis-PTs illustrate the possible evolutionary paths of DHDDS and NgBR orthologs. Plant NgBR orthologs with NXG motif might be derived from that form certain fungi or bacteria through symbiosis and horizontal gene transfer, and NgBR orthologs from animals might be directly derived from UPPS-like bacterial cis-PTs with RXG motif. In summary, we provide here the information of kinetics and functions of the homomeric and heteromeric cis-PT homologues in the tree *C. kanehirae*, which may apply to other plants with multiple cis-PT homologues, either active or supporting activity.

4 | MATERIALS AND METHODS

4.1 | Materials

IPP, FPP, and potato acid phosphatase were obtained from Sigma (St. Louis, MO, USA). PrimeSTAR® GXL DNA polymerase for PCR and TALON® Metal Affinity Resin were obtained from Takara Bio (Shiga, Japan). The plasmid mini-prep kit and DNA gel extraction kit were purchased from Qiagen (Hilden, Germany). Polyprenol mixture (13–21 prenyl groups) standard was obtained from Avanti Polar Lipids (Alabama, USA). HPLC ZORBAX XDB-C18 (4.6 × 75 mm, 3.5 μm) reversed-phase column was purchased from Agilent (Santa Clara, CA, USA). All commercial buffers and reagents were of the highest grade.

4.2 | Expression and purification of the recombinant RWR89

The gene encoding the full-length of RWR89 synthesized by MORREBIO (Taiwan) with optimized codons for *E. coli* expression was ligated with the pET32a vector between the *Eco*RI and *Xho*I restriction sites for expressing the recombinant protein with N-terminal thioredoxin, hexa-His tag, and TEV protease cutting site. The *E. coli* BL21 competent cells transformed with the TRX-His_6-TEVp-RWR89 construct were plated on LB-agar under 100 mg/L ampicillin selection. The selected colonies were inoculated into 20 mL of LB medium with constant shaking at 200–250 rpm overnight. The culture was used to inoculate 2 L of 2×YT medium (16 g/L tryptone, 10 g/L yeast extract, and 5 g/L NaCl) containing 100 mg/L ampicillin with constant shaking at 200–250 rpm at 37°C until the cell density reached OD_{600} of 0.6–0.8. The protein production was induced by adding 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) with constant shaking at 180 rpm at 16°C for 16–20 h. The cells were harvested by centrifugation (6500 rpm for 20 min), resuspended in buffer A (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1 mM $MgCl_2$, 0.1% Triton X-100, and 2 mM TCEP), and disrupted by a French-press instrument (Constant Cell Disruption System) at 20,000 psi. Cell lysate was centrifuged at 40,000×*g* for 45 min at 4°C to remove debris, and the supernatant was loaded onto an open column with 5 mL TALON Metal Affinity Resin. After washing with 100 mL of buffer A with 10 mM imidazole, the target protein was eluted with 25 mL of buffer A containg 250 mM imidazole, followed by dialysis against 5 L buffer A to remove imidazole and subjected to TEV protease to cleavage the TRX-His_6-tag at 4°C overnight. The mixture was subjected to another open column of 5 mL TALON Metal Affinity Resin, and the untagged RWR89 was eluted in the imidazole-free buffer A.

4.3 | Expression and purification of the recombinant RWR78, sRWR71, sRWR93, RWR78/sRWR71, RWR78/sRWR93, and RWR89 using yeast

Genes of RWR78, sRWR71 (aa 115–329), and sRWR93 (aa 61–275) were codon-optimized and synthesized by Genomics Co. (Taiwan). Genes of full-length RWR78 and sRWR71 or sRWR93 were amplified by PCR and constructed into pESC-URA yeast bicistronic vector by using In-Fusion® HD Cloning Kit (Takara Bio, Japan). The PCR product of RWR78 or RWR89 with N-terminal His-tag encoding sequence was subcloned into EcoRI/NotI restriction sites, while the PCR product of sRWR71 or sRWR93 with N-terminal Strep-tag encoding sequence were, respectively subcloned into BamHI/SalI sites. Specific primers used in all constructs are listed in Table S1.

Yeast INVSc1 strain for protein overexpression was transformed with either pESC-RWR78, pESC-sRWR71,

doi:10.6342/NTU202502385

pESC-sRWR93, pESC-RWR78/sRWR71, pESC-RWR78/sRWR93, or pESC-RWR89 plasmid, and the cells were plated on solid SC minimal medium without uracil and grown at 30°C for 2 days. Colonies were grown in 20 mL SC minimal medium at 30°C overnight, followed by inoculation into fresh 2 L SC minimal medium until $OD_{600} = 0.4$. Cells were centrifuged at 1500 × *g* for 5 min and resuspended in fresh SC minimal medium containing 2% galactose. Cultures were grown at 30°C for 1–4 days to induce protein expression. Stationary-phase of yeast cells were harvested by centrifuged at 3300 × *g* for 5 min, followed by washing with sterile water once, and dissolved in buffer A. The cells were disrupted by a French-press instrument (Constant Cell Disruption System) at 22,000 psi, and then the cell lysate was centrifuged at ~40,000 × *g* for 45 min at 4°C to remove debris. From the supernatant, His-tagged RWR89, His-tagged RWR78, and heteromeric complexes were respectively purified by Ni-NTA resin using a 10–250 mM imidazole gradient. His-tagged RWR78 was eluted at 30–50 mM imidazole, His-tagged RWR78 was eluted at 20–40 mM imidazole, and complexes were eluted at ~70 mM imidazole. For purifying the Strep-tagged sRWR71 or sRWR93, StrepTrap XT column (Cytiva, Marlborough, USA) was used. The supernatant was applied onto a StrepTrap XT column, washed with 50 mL of buffer A to remove unbound proteins, and the Strep-tagged sRWR71 or sRWR93 was eluted with 50 mM biotin in buffer A. The affinity-purified protein was further purified using a size-exclusion chromatography and the purity was judged by SDS-PAGE.

4.4 | Size-exclusion chromatography analysis

Superdex 200 Increase 10/300 GL column (Cytiva) was used to estimate the protein molecular weights. Proteins were centrifuged at 21,000 × *g* for 10 min, loaded onto the column and analyzed at a flow rate of 0.75 mL/min by using buffer A. The chromatography profiles were compared with that of the standard markers containing 1 mg/mL ribonuclease A (13.7 kDa), conalbumin (75 kDa), aldolase (158 kDa), and thyroglobulin (669 kDa) to determine protein molecular weights.

4.5 | Cis-PTs activity assay

EnzChek Pyrophosphate assay kit (Thermofisher, MA, USA) was used to detect the continuously PPi release by the enzymatic reactions. The activity was assayed in buffer containing 100 mM Hepes, pH 7.5, 50 mM KCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 0.2 mM MESG, 1 U/mL PNP, and 0.03 U/mL pyrophosphatase, with either 1.95 μM of RWR89, 5.89 μM of RWR78, 2.27 μM of RWR78/sRWR71, or 1.26 μM of RWR78/sRWR93. To determine the K_m values of GPP, FPP, and GGPP, 50 μM IPP was utilized, and K_m of IPP was measured in the presence of 10 μM FPP. The plots of initial rates versus substrate concentrations were used to determined K_m and k_{cat} by fitting the data with Michaelis–Menten equation. All measurements were performed in triplicate at 30°C.

4.6 | Yeast complementary assay

The rer2Δ mutant yeast strain was transformed with the empty vector pYES2, pYES2-DHDDS, pYES2-RWR89, the empty vector pESC-URA, pESC-RWR78, pESC-RWR71, pESC-RWR93, pESC-RWR78/RWR71, and pESC-RWR78/RWR93, respectively. The transformants were grown at 30°C in SC minimal medium containing 2% glucose until $OD_{600} = 0.4$, followed by centrifugation at 1500 × *g* for 5 min and resuspension in fresh SC minimal medium containing 2% galactose. After growing at 30°C for 1–4 days to induce protein expression, yeast cultures were diluted to $OD_{600} = 0.4$ and streaked onto solid YP-2% galactose medium and incubated at 23 or 37°C for 5 days. The *S. cerevisiae* wild-type strain INVsc1, which typically grew at 37°C, was utilized as the positive control, and non-transformed rer2Δ mutant strain was used as the negative control.

4.7 | Polyisoprenoids extraction from yeast rer2Δ strain

Yeast rer2Δ cells were transformed with the empty vector pESC-URA, pYES2-RWR89, pESC-RWR78, pESC-RWR78/RWR71, and pESC-RWR78/RWR93, respectively. Cells were plated on solid SC minimal medium without uracil and grown at 30°C for 2 days. Colonies were grown in 20 mL SC minimal medium at 30°C overnight, followed by inoculation into fresh 1 L SC minimal medium until $OD_{600} = 0.4$. Cells were centrifuged at 1500 × *g* for 5 min and resuspended in fresh SC minimal medium containing 2% galactose. Cultures were grown at 30°C for 1–4 days to induce protein expression. Stationary-phase of yeast cells were harvested by centrifuged at 3300 × *g* for 5 min, followed by washing with sterile water once. Cell pellets were dissolved in 10 mL hydrolytic solution (2.5 g KOH was dissolved in 3.5 mL sterile distilled water, and brought to 10 mL with 99.8% ethanol), vortexed for 1 min, and incubated at 95°C for

doi:10.6342/NTU202502385

1 h. Non-saponifiable lipids were extracted with hexane three times. The pooled extracts were further purified on a silica gel 60 column using isocratic elution with 10% diethyl ether in hexane, evaporated to dryness, and dissolved in isopropanol. Purified polyisoprenoids were examined by HPLC/UV using standards of different chain-length polyprenols for comparison by following the procedure (Edani et al., 2020).

4.8 | HPLC product analysis

For HPLC analysis of the products extracted from rer2Δ yeast expressing RWR78/RWR71 or RWR78/RWR93, an established protocol was used (Edani et al., 2020). A dual-pump HPLC device (Jasco, Japan) with a ZORBAX XDB-C18 (4.6 × 75 mm, 3.5 μm) reversed-phase column (Agilent, Santa Clara, CA, USA), a Detector (spectrum range: 210–400 nm), methanol/water at 9:1 (v/v) as solvent A, methanol/isopropanol/hexane at 2:1:1 (v/v) as solvent B, and a flow rate of 1.5 mL/min controlled by a gradient programmer were used. The chain lengths of the extracted polyisoprenoids were confirmed by comparison with a polyprenols-containing mixture (13–21 prenyl groups) standard (Avanti Polar Lipids, USA).

4.9 | Protein multiple sequence alignment, phylogenetic analysis, and motif elicitation of cis-PTs

The amino acid sequences of cis-PTs from microbes, yeast, plants, and animals were downloaded from GenBank and aligned by using Clustal Omega and Jalview program, and the phylogenetic tree was constructed by using iTOL (Interactive Tree Of Life) website (https://itol.embl.de/). The MEME online tool (https://meme-suite.org/meme/tools/meme, Version 5.5.2), with the motif number set to 30 and the width of each motif ranging from 3 to 50 amino acids, was used to identify the motifs of cis-PTs.

4.10 | Homology modeling of homodimeric RWR89, DHDDS, RWR78 and heteromeric RWR78/RWR71, and RWR78/RWR93

The SWISS-MODEL web service (https://swissmodel.expasy.org) was utilized to perform homology modeling. The closest homology to RWR89 was *S. aureus* UPPS (PDB: 4H8E), and the closest homology of heteromeric complexes is DHDDS/NgBR complex (PDB: 6W2L). Therefore, 4H8E and 6W2L were used as templates to generate structural models. To generate the structural model of homodimeric DHDDS or RWR78, ColabFold v1.5.2 online software using AlphaFold2 algorithm (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) was used. For multiple sequence alignment and template detection, MMseqs2 method and pdb70 template_mode were selected. After structural prediction finished, the rank_1 model was chosen for further comparison.

4.11 | RNA-seq analysis of five *C. kanehirae* cis-PTs

The transcriptome dataset of *C. kanehirae* from various tissues, including buds, buds of flowers, flowers, buds of leaves, young leaves, old leaves, and fruits, was obtained from previous study (Chaw et al., 2019). The fastq files were downloaded and aligned against the genome of *C. kanehirae* (GCA_003546025.1) using HISAT2 program (Kim et al., 2015). The read counts were calculated with FeatureCounts and normalized using upper quantile normalization with edgeR (Liao et al., 2014; Robinson et al., 2010).

AUTHOR CONTRIBUTIONS

J.-J. L. performed all the experiments. P.-H. L. designed the experiments, supervised all aspects of research, and wrote the manuscript.

ACKNOWLEDGMENTS

The authors thank Dr Suh-Yuen Liang (IBC, Academia Sinica) for her assistance with RNA-seq analysis.

FUNDING INFORMATION

This study was supported by the Ministry of Science and Technology, Taiwan (102-2113-M-001-005-MY3).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Po-Huang Liang https://orcid.org/0000-0003-1207-5256

REFERENCES

Akhtar TA, Matsuba Y, Schauvinhold I, Yu G, Lees HA, Klein SE, et al. The tomato cis-prenyltransferase gene family. Plant J. 2013;73:640–52. https://doi.org/10.1111/tpj.12063

Bar-El ML, Vaňková P, Yeheskel A, Simhaev L, Engel H, Man P, et al. Structural basis of heterotetrameric assembly and disease mutations in the human cis-prenyltransferase complex. Nat Commun. 2020;11:5273. https://doi.org/10.1038/s41467-020-18970-z

doi:10.6342/NTU202502385

Brasher MI, Surmacz L, Leong B, Pitcher J, Swiezewska E, Pichersky E, et al. A two-component enzyme complex is required for dolichol biosynthesis in tomato. Plant J. 2015;82:903–14. https://doi.org/10.1111/tpj.12859

Cantagrel V, Lefeber DJ, Ng BG, Guan Z, Silhavy JL, Bielas SL, et al. SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder. Cell. 2010;142:203–17. https://doi.org/10.1016/j.cell.2010.06.001

Chan YT, Ko TP, Yao SH, Chen YW, Lee CC, Wang AH. Crystal structure and potential head-to-middle condensation function of a Z,Z-Farnesyl Diphosphate Synthase. ACS Omega. 2017;2:930–6. https://doi.org/10.1021/acsomega.6b00562

Chang SY, Ko TP, Liang PH, Wang AH. Catalytic mechanism revealed by the crystal structure of undecaprenyl pyrophosphate synthase in complex with sulfate, magnesium, and triton. J Biol Chem. 2003;278:29298–307. https://doi.org/10.1074/jbc.M302687200

Chaw SM, Liu YC, Wu YW, Wang HY, Lin CI, Wu CS, et al. Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. Nat Plants. 2019;5:63–73. https://doi.org/10.1038/s41477-018-0337-0

Chen YH, Chen AP, Chen CT, Wang AH, Liang PH. Probing the conformational change of *Escherichia coli* undecaprenyl pyrophosphate synthase during catalysis using an inhibitor and tryptophan mutants. J Biol Chem. 2002;277:7369–76. https://doi.org/10.1074/jbc.M110014200

De Coninck T, Gistelinck K, Janse van Rensburg HC, Van den Ende W, Van Damme EJM. Sweet modifications modulate plant development. Biomolecules. 2021;11:756. https://doi.org/10.3390/biom11050756

Edani BH, Grabińska KA, Zhang R, Park EJ, Siciliano B, Surmacz L, et al. Structural elucidation of the cis-prenyltransferase NgBR/DHDDS complex reveals insights in regulation of protein glycosylation. Proc Natl Acad Sci U S A. 2020;117:20794–802. https://doi.org/10.1073/pnas.2008381117

Edri I, Goldenberg M, Lisnyansky M, Strulovich R, Newman H, Loewenstein A, et al. Overexpression and purification of human cis-prenyltransferase in *Escherichia coli*. J Vis Exp. 2017;126:56430. https://doi.org/10.3791/56430

Emi KI, Sompiyachoke K, Okada M, Hemmi H. A heteromeric cis-prenyltransferase is responsible for the biosynthesis of glycosyl carrier lipids in *Methanosarcina mazei*. Biochem Biophys Res Commun. 2019;520:291–6. https://doi.org/10.1016/j.bbrc.2019.09.143

Epping J, van Deenen N, Niephaus E, Stolze A, Fricke J, Huber C, et al. A rubber transferase activator is necessary for natural rubber biosynthesis in dandelion. Nat Plants. 2015;1:1–9. https://doi.org/10.1038/nplants.2015.48

Fujihashi M, Zhang YW, Higuchi Y, Li XY, Koyama T, Miki K. Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. Proc Natl Acad Sci U S A. 2001;98:4337–42. https://doi.org/10.1073/pnas.071514398

Giladi M, Edri I, Goldenberg M, Newman H, Strulovich R, Khananshvili D, et al. Purification and characterization of human dehydrodolychil diphosphate synthase (DHDDS) overexpressed in *E. coli*. Protein Expr Purif. 2017;132:138–42. https://doi.org/10.1016/j.pep.2017.02.001

Giladi M, Lisnyansky Bar-El M, Vaňková P, Ferofontov A, Melvin E, Alkaderi S, et al. Structural basis for long-chain isoprenoid synthesis by *cis*-prenyltransferases. Sci Adv. 2022;8:eabn1171. https://doi.org/10.1126/sciadv.abn1171

Grabińska KA, Edani BH, Park EJ, Kraehling JR, Sessa WC. A conserved C-terminal RXG motif in the NgBR subunit of cis-prenyltransferase is critical for prenyltransferase activity. J Biol Chem. 2017;292:17351–61. https://doi.org/10.1074/jbc.M117.806034

Guo RT, Ko TP, Chen AP, Kuo CJ, Wang AH, Liang PH. Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved residues in catalysis. J Biol Chem. 2005;280:20762–74. https://doi.org/10.1074/jbc.M502121200

Hamdan FF, Myers CT, Cossette P, Lemay P, Spiegelman D, Laporte AD, et al. High rate of recurrent de novo mutations in developmental and epileptic encephalopathies. Am J Hum Genet. 2017;101:664–85. https://doi.org/10.1016/j.ajhg.2017.09.008

Harrison KD, Park EJ, Gao N, Kuo A, Rush JS, Waechter CJ, et al. Nogo-B receptor is necessary for cellular dolichol biosynthesis and protein N-glycosylation. EMBO J. 2011;30:2490–500. https://doi.org/10.1038/emboj.2011.147

Helenius A, Aebi M. Intracellular functions of N-linked glycans. Science. 2001;291:2364–9. https://doi.org/10.1126/science.291.5512.2364

Holcomb J, Doughan M, Spellmon N, Lewis B, Perry E, Zhang Y, et al. SAXS analysis of a soluble cytosolic NgBR construct including extracellular and transmembrane domains. PloS One. 2018;13:e0191371. https://doi.org/10.1371/journal.pone.0191371

Kang JH, Gonzales-Vigil E, Matsuba Y, Pichersky E, Barry CS. Determination of residues responsible for substrate and product specificity of *Solanum habrochaites* short-chain cis-prenyltransferases. Plant Physiol. 2014;164:80–91. https://doi.org/10.1104/pp.113.230466

Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60. https://doi.org/10.1038/nmeth.3317

Ko TP, Chen YK, Robinson H, Tsai PC, Gao YG, Chen AP, et al. Mechanism of product chain length determination and the role of a flexible loop in *Escherichia coli* undecaprenyl-pyrophosphate synthase catalysis. J Biol Chem. 2001;276:47474–82. https://doi.org/10.1074/jbc.M106747200

Liang PH. Reaction kinetics, catalytic mechanisms, conformational changes, and inhibitor design for prenyltransferases. Biochemistry. 2009;48:6562–70. https://doi.org/10.1021/bi900371p

Liang PH, Ko TP, Wang AH. Structure, mechanism and function of prenyltransferases. Eur J Biochem. 2002;269:3339–54. https://doi.org/10.1046/j.1432-1033.2002.03014.x

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656

Lisnyansky Bar-El M, Lee SY, Ki AY, Kapelushnik N, Loewenstein A, Chung KY, et al. Structural characterization of full-length human dehydrodolichyl diphosphate synthase using an integrative computational and experimental approach. Biomolecules. 2019;9:660. https://doi.org/10.3390/biom9110660

Ma J, Ko TP, Yu X, Zhang L, Ma L, Zhai C, et al. Structural insights to heterodimeric cis-prenyltransferases through yeast dehydrodolichyl diphosphate synthase subunit Nus1. Biochem Biophys

doi:10.6342/NTU202502385

Res Commun. 2019;515:621–6. https://doi.org/10.1016/j.bbrc.2019.05.135
Ng BG, Freeze HH. Perspectives on glycosylation and its congenital disorders. Trends Genet. 2018;34:466–76. https://doi.org/10.1016/j.tig.2018.03.002
Ogawa T, Emi K, Koga K, Yoshimura T, Hemmi H. A cis-prenyltransferase from *Methanosarcina acetivorans* catalyzes both head-to-tail and nonhead-to-tail prenyl condensation. FEBS J. 2016;283:2369–83. https://doi.org/10.1111/febs.13749
Ogura K, Koyama T. Enzymatic aspects of isoprenoid chain elongation. Chem Rev. 1998;98:1263–76. https://doi.org/10.1021/cr9600464
Okada M, Unno H, Emi KI, Matsumoto M, Hemmi H. A versatile cis-prenyltransferase from *Methanosarcina mazei* catalyzes both C- and O-prenylations. J Biol Chem. 2021;296:100679. https://doi.org/10.1016/j.jbc.2021.100679
Park EJ, Grabińska KA, Guan Z, Stránecký V, Hartmannová H, Hodaňová K, et al. Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation. Cell Metab. 2014;20:448–57. https://doi.org/10.1016/j.cmet.2014.06.016
Qu Y, Chakrabarty R, Tran HT, Kwon EJ, Kwon M, Nguyen TD, et al. A lettuce (*Lactuca sativa*) homolog of human Nogo-B receptor interacts with cis-prenyltransferase and is necessary for natural rubber biosynthesis. J Biol Chem. 2015;290:1898–914. https://doi.org/10.1074/jbc.M114.616920
Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40. https://doi.org/10.1093/bioinformatics/btp616
Sabry S, Vuillaumier-Barrot S, Mintet E, Fasseu M, Valayannopoulos V, Héron D, et al. A case of fatal type I congenital disorders of glycosylation (CDG I) associated with low dehydrodolichol diphosphate synthase (DHDDS) activity. Orphanet J Rare Dis. 2016;11:1–14. https://doi.org/10.1186/s13023-016-0468-1
Sacchettini JC, Poulter CD. Creating isoprenoid diversity. Science. 1997;277:1788–9. https://doi.org/10.1126/science.277.5333.1788
Shimizu N, Koyama T, Ogura K. Molecular cloning, expression, and purification of undecaprenyl diphosphate synthase. No sequence similarity between E- and Z-prenyl diphosphate synthases. J Biol Chem. 1998;273:19476–81. https://doi.org/10.1074/jbc.273.31.19476
Sompiyachoke K, Nagasaka A, Ito T, Hemmi H. Identification and biochemical characterization of a heteromeric cis-prenyltransferase from the thermophilic archaeon *Archaeoglobus fulgidus*. J Biochem. 2022;171:641–51. https://doi.org/10.1093/jb/mvac022
Surmacz L, Plochocka D, Kania M, Danikiewicz W, Swiezewska E. cis-Prenyltransferase atCPT6 produces a family of very short-chain polyisoprenoids in planta. Biochim Biophys Acta. 2014;1841:240–50. https://doi.org/10.1016/j.bbalip.2013.11.011
Surowiecki P, Onysk A, Manko K, Swiezewska E, Surmacz L. Long-chain polyisoprenoids are synthesized by AtCPT1 in *Arabidopsis thaliana*. Molecules. 2019;24:2789. https://doi.org/10.3390/molecules24152789
Takahashi S, Koyama T. Structure and function of cis-prenyl chain elongating enzymes. Chem Rec. 2006;6:194–205. https://doi.org/10.1002/tcr.20083
Teng KH, Liang PH. Structures, mechanisms and inhibitors of undecaprenyl diphosphate synthase: a cis-prenyltransferase for bacterial peptidoglycan biosynthesis. Bioorg Chem. 2012;43:51–7. https://doi.org/10.1016/j.bioorg.2011.09.004
Wang W, Dong C, McNeil M, Kaur D, Mahapatra S, Crick DC, et al. The structural basis of chain length control in Rv1086. J Mol Biol. 2008;381:129–40. https://doi.org/10.1016/j.jmb.2008.05.060
Yao JY, Teng KH, Liu MC, Wang CS, Liang PH. Characterization of a *cis*-prenyltransferase from *Lilium longiflorum* Anther. Molecules. 2019;24:2728. https://doi.org/10.3390/molecules24152728
Zelinger L, Banin E, Obolensky A, Mizrahi-Meissonnier L, Beryozkin A, Bandah-Rozenfeld D, et al. A missense mutation in DHDDS, encoding dehydrodolichyl diphosphate synthase, is associated with autosomal-recessive retinitis pigmentosa in Ashkenazi Jews. Am J Hum Genet. 2011;88:207–15. https://doi.org/10.1016/j.ajhg.2011.01.002
Züchner S, Dallman J, Wen R, Beecham G, Naj A, Farooq A, et al. Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa. Am J Hum Genet. 2011;88:201–6. https://doi.org/10.1016/j.ajhg.2011.01.001

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu J-J, Liang P-H. Complexation and evolution of cis-prenyltransferase homologues in *Cinnamomum kanehirae* deduced from kinetic and functional characterizations. Protein Science. 2023;32(12): e4828. https://doi.org/10.1002/pro.4828

doi:10.6342/NTU202502385

Antiviral Research 219 (2023) 105735

Contents lists available at ScienceDirect

Antiviral Research

journal homepage: www.elsevier.com/locate/antiviral

Synthesis, evaluation, and mechanism of 1-(4-(arylethylenylcarbonyl) phenyl)-4-carboxy-2-pyrrolidinones as potent reversible SARS-CoV-2 entry inhibitors

Srinivasa Rao Palla [b,c], Chen-Wei Li [b], Tai-Ling Chao [d], Hoi-Ling Vienn Lo [b], Jia-Jin Liu [b], Max Yu-Chen Pan [e], Yu-Ting Chiu [e], Wen-Chin Lin [f,g], Chih-Wei Hu [f,g,h], Chuen-Mi Yang [f], Yi-Ying Chen [f,h], Jun-Tung Fang [d], Sheng-Wei Lin [a], Yi-Tzu Lin [a], Hsiao-Ching Lin [a,b,c], Chih-Jung Kuo [h,**], Lily Hui-Ching Wang [e,i,***], Sui-Yuan Chang [d,j,****], Po-Huang Liang [a,b,c,*]

[a] Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan
[b] Institute of Biochemical Sciences, National Taiwan University, 10617, Taiwan
[c] Taiwan International Graduate Program, Academia Sinica, Taipei 11529, Taiwan
[d] Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University, Taipei 10048, Taiwan
[e] Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu 30013, Taiwan
[f] Institute of Preventive Medicine, National Defense Medical Center, New Taipei City 23742, Taiwan
[g] Graduate Institute of Pathology and Parasitology, National Defense Medical Center, Taipei 11490, Taiwan
[h] Department of Veterinary Medicine, College of Veterinary Medicine, National Chung Hsing University, Taichung 40227, Taiwan
[i] School of Medicine, National Tsing Hua University, Hsinchu 30013, Taiwan
[j] Department of Laboratory Medicine, National Taiwan University Hospital, Taipei 10002, Taiwan

ARTICLE INFO

Keywords:
COVID-19
SARS-CoV-2
Pyrrolidinones
TMPRSS2
Furin
RBD:ACE2

ABSTRACT

A class of 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones were designed and synthesized via Michael addition, cyclization, aldol condensation, and deprotonation to inhibit the human transmembrane protease serine 2 (TMPRSS2) and Furin, which are involved in priming the SARS-CoV-2 Spike for virus entry. The most potent inhibitor **2f** (**81**) was found to efficiently inhibit the replication of various SARS-CoV-2 delta and omicron variants in VeroE6 and Calu-3 cells, with EC_{50} range of 0.001–0.026 μM by pre-incubation with the virus to avoid the virus entry. The more potent antiviral activities than the proteases inhibitory activities led to discovery that the synthesized compounds also inhibited Spike's receptor binding domain (RBD):angiotensin converting enzyme 2 (ACE2) interaction as a main target, and their antiviral activities were enhanced by inhibiting TMPRSS2 and/or Furin. To further confirm the blocking effect of **2f** (**81**) on virus entry, SARS-CoV-2 Spike pseudovirus was used in the entry assay and the results showed that the compound inhibited the pseudovirus entry in a ACE2-dependent pathway, via mainly inhibiting RBD:ACE2 interaction and TMPRSS2 activity in Calu-3 cells. Finally, in the *in vivo* animal model of SARS-CoV-2 infection, the oral administration of 25 mg/kg **2f** (**81**) in hamsters resulted in reduced bodyweight loss and 5-fold lower viral RNA levels in nasal turbinate three days post-infection. Our findings demonstrated the potential of the lead compound for further preclinical investigation as a potential treatment for SARS-CoV-2.

* Corresponding author. Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan.
** Corresponding author. Department of Veterinary Medicine, National Chung Hsing University, Taichung 40227, Taiwan.
*** Corresponding author. Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu 30013, Taiwan.
**** Corresponding author. Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University, Taipei 10048, Taiwan.
E-mail addresses: ck476@dragon.nchu.edu.tw (C.-J. Kuo), lilywang@life.nthu.edu.tw (L.H.-C. Wang), sychang@ntu.edu.tw (S.-Y. Chang), phliang@gate.sinica.edu.tw (P.-H. Liang).

https://doi.org/10.1016/j.antiviral.2023.105735
Received 1 May 2023; Received in revised form 13 October 2023; Accepted 16 October 2023
Available online 18 October 2023

Abbreviations

COVID-19	Coronavirus Disease 2019
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus-2
ACE2	angiotensin converting enzyme 2
RBD	receptor binding domain
TMPRSS2	transmembrane protease serine 2
NSP	non-structural protein
$3CL^{pro}$	3C-like protease
PL^{pro}	papain-like protease
RdRp	RNA-dependent RNA polymerase
PBS	phosphate buffered saline
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal Bovine Serum
pfu	plaque forming unit
IV	intravenous

1. Introduction

The coronavirus disease 2019 (COVID-19) first reported in December of 2019 (Huang et al., 2020; Wu et al., 2020; Zhu et al., 2020) has been declared a worldwide pandemic by the World Health Organization in March 2020 and is still ongoing causing an approximately 1% case mortality rate. The pathogen is the human severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) that is highly similar (79.6% identity in genome) to the previously (2002–2003) emerged SARS-CoV (Zhou et al., 2020; Lu et al., 2020). For infection, both SARS-CoV and SARS-CoV-2 utilize human angiotensin-converting enzyme 2 (ACE2) as the receptor to bind with the virus surface Spike glycoprotein through its receptor binding domain (RBD) located in the S1 domain (Jackson et al., 2022). After binding, the human protease Furin cleaves Spike at S1/S2 ($\mathbf{PRRAR}_{685}$↓S) into two fragments S1 and S2. Then the human Type II transmembrane protease serine 2 (TMPRSS2) further cleaves S2 at $\mathbf{KPSKR}_{815}$↓S, exposing the membrane fusion peptide and allowing the virus to fuse with the cell membrane for entry into the cell, where the virus replicates (Bestle et al., 2020; Essalmani et al., 2022). Alternatively, SARS-CoV-2 could enter the cell through the ACE2-mediated endocytosis and then the Spike protein could be cleaved by an endosomal human protease Cathepsin L to release the viral RNA into cytosol (Mondal et al., 2022). Therefore, RBD:ACE2 interaction, TMPRSS2, Furin, and Cathepsin L could serve as targets for inhibiting SARS-CoV-2 entry and infection.

Like SARS-CoV, SARS-CoV-2 is an enveloped, positive single-stranded RNA virus, belonging to β-CoVs. Their relatively large 29.7 kb genomes encode two polyproteins pp1a and pp1ab that can be processed into mature non-structural proteins (NSPs) to become functional by the virus-encoded proteases, 3C-like protease ($3CL^{pro}$) as homologous to the picornavirus 3C protease (also called main protease, M^{pro}) and papain-like protease (PL^{pro}). Both viral proteases are chymotrypsin-like but utilize Cys as a nucleophile for proteolysis. Mature NSPs, including NSP7, 8, 10, 12, 13, and 14 then form a replication-transcription complex with RNA-dependent RNA polymerase (RdRp), helicase, exonuclease etc. activities for making sub-genomic RNAs of the structural proteins, Envelop (E), Nucleocapsid (N), Membrane (M), and Spike (S) proteins, to be assembled with the (+)RNA into new virus particles (Kuo and Liang, 2015; de Wit et al., 2016). $3CL^{pro}$ and PL^{pro} cleave 11 and 3 sites on the polyproteins for maturation of NSP1−16 (Hsu et al., 2005; Kuo and Liang, 2022). Therefore, these two virus-encoded proteases are attractive drug targets and many drug discovery efforts have been made for developing covalent and non-covalent inhibitors (e.g. Kuo et al., 2021; Lee et al., 2021). Covalent peptidomimetic inhibitors against $3CL^{pro}$ contain different warheads such as α-ketoamide (Zhang et al., 2020), α,β-unsaturated ester (Jin et al., 2020), aldehyde (Dai et al., 2020), bisulfite (Kim et al., 2016; Vuong et al., 2020), keto (Hoffman et al., 2020), and ester (Ghosh et al., 2021) for forming a covalent bond with the catalytic Cys145, where the P1 residue is strictly a cyclic lactam mimicking Gln to ensure tight binding. To date, Paxlovid developed by Pflizer Co., a nitrile-based covalent $3CL^{pro}$ peptidomimetic inhibitor combined with ritonavir for inhibiting CYP3A4, was approved for treatment. After maturation by $3CL^{pro}$, the viral RdRp essential for gene replication is the antiviral target of Remdesivir (Gordon et al., 2020; Wang et al., 2020) developed by Gilead Co., which showed a marginal efficacy in clinical trials (Grein et al., 2020). Another RdRp inhibitor, Molnupiravir developed by Merck Co., a prodrug of the nucleoside analogue N4-hydroxycytidine (Sheahan et al., 2020), has also been approved.

While using $3CL^{pro}$ and RdRp as targets has yielded three small-molecule drugs for combating COVID-19, only antibodies such as Bebtelovimab, which bind Spike to block RBD:ACE2 interaction and prevent virus entry into cells have been approved for therapy. Nevertheless, these neutralizing antibodies are largely evaded by BA.2 and BA.4/BA.5 owing to Spike D405N and F486V mutations, and react weakly to pre-Omicron variants, exhibiting narrow neutralization breadths (Cao et al., 2022). Targeting human Furin using a covalent inhibitor decanoyl-RVKR-chloromethylketone at 5 μM effectively inhibited virus production in SARS-CoV-2-infected VeroE6 cells by blocking virus entry (Cheng et al., 2020). By irreversibly inhibiting human TMPRSS2, Camostat and Nafamostat used a common warhead to block SARS-CoV-2 infection *in vitro* (Hoffmann et al., 2020a, 2020b; Yamamoto et al., 2020). Moreover, covalent small-molecule ketobenzothiazole TMPRSS2 inhibitors (MM3122 as the lead) were shown to block SARS-CoV-2 viral entry and protect human epithelial lung cells (Mahoney et al., 2021). Specific TMPRSS2 ketobenzothiazole covalent peptidomimetic inhibitor N-0385 was effective against SARS-CoV-2 *in vitro* and *in vivo* using human ACE2 transgenic mouse model of severe COVID-19 (Shapira et al., 2022). However, their keto-reduced hydroxyl analogues not able to link covalently with the active-site essential Ser residue could not inhibit TMPRSS2 and the virus. Aiming to develop potent reversible inhibitors of SARS-CoV-2 entry, but a reversible inhibitor of TMPRSS2 or Furin alone was expected not to achieve the desired potency, we thus designed small molecules which might reversibly inhibit TMPRSS2 and Furin simultaneously. As reported herein, guided by computer modeling on a class of chalcones, we synthesized a series of 1-(4-(arylethylenylcarbonyl)phenyl)-4--carboxy-2-pyrrolidinones inhibiting TMPRSS2 and Furin, but later found they also inhibited RBD:ACE2 interaction, and studied their structure-activity relationship. The best inhibitor thus achieved potent half maximal effective concentrations (EC_{50}) of 0.023 and 0.002 μM as well as 0.026 and 0.001 μM against the SARS-CoV-2 delta and omicron variants infecting VeroE6 and Calu-3 cells, respectively, and $CC_{50} > 100$ μM, representing a potential lead compound for further preclinical study.

2. Materials and methods

2.1. General methods

Reactions for synthesizing compounds were monitored by using thin-layer chromatography (TLC) on silica gel. Flash chromatography was performed on silica gel of 60–200 μm particle size for compound purification. Yields were reported for spectroscopically pure compounds. Melting points were recorded on the Fargo MP-1D Melting Point Apparatus. NMR spectra were recorded on Bruker AVIIIHD 400 MHz FT-NMR in the Department of Chemistry, National Taiwan University. Chemical shifts were given in δ values relative to tetramethylsilane (TMS); coupling constants *J* were given in Hz. Internal standards were $CDCl_3$ ($\delta_H = 7.24$) or DMSO-d_6 ($\delta_H = 2.49$) for ^{1}H-NMR spectra, and $CDCl_3$ ($\delta_C = 77.0$) or DMSO-d_6 ($\delta_C = 39.5$) for ^{13}C-NMR spectra. The splitting patterns were reported as s (singlet), d (doublet), t (triplet), q

doi:10.6342/NTU202502385

(quartet), m (multiplet), br (broad) and dd (double of doublets). High resolution mass spectra were measurement by Bruker UPLC-MS in the TechComm core facility, College of Life Science, National Taiwan University.

2.2. Chemicals

All the reagents were of the highest commercially available grade and used without further purification. 4-aminoacetophenone, itaconic acid, 3-nitrobenzaldehyde, 4-nitrobenzaldehyde, furan-2-carboxaldehyde, 5-nitrothiophene-2-carboxaldehyde and indole-4-carboxaldehyde were purchased from AK Scientific (Union City, CA, USA). Benzaldehyde and indole-3-carboxaldehyde were purchased from Acros Organics (NJ, USA).

2.3. Synthesis of compounds

2.3.1. 1-(4-acetylphenyl)-4-carboxy-2-pyrrolidinone (1)

A mixture of 1 g (7.40 mmol) 4-aminoacetophenone and 1.2 g (8.89 mmol) itaconic acid was stirred and heated (110−130 °C) under reflux for 18 h. After cooling to room temperature, 10 mL methanol was added to the reaction mixture. The reaction mixture was sonicated and heated to dissolve the product in methanol, and then cooled to room temperature for recrystallization. Crystallized solid was filtered and washed with ethyl acetate to yield the pure product. White solid, yield 49.8%, mp 180−181 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.807 (s, 1H), 7.991 (d, j = 8.8 Hz, 2H), 7.833 (d, j = 8.88 Hz, 2H), 4.138−4.009 (m, 2H), 3.408−3.322 (m, 1H), 2.876−2.723 (m, 2H), 2.556 (s, 3H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 196.64, 174.02, 172.57, 143.10, 132.09, 129.15, 118.41, 49.78, 35.32, 35.01, 26.49. MS m/z [M+H]$^+$ calcd. for $C_{13}H_{13}NO_4$ 247.0845, found 248.0929.

2.3.2. 1-(4-styrylcarbonylphenyl)-4-carboxy-2-pyrrolidinone (2a)

A mixture of 2.47 g (0.01 mol) 1-(4-acetylphenyl)-4-carboxy-2-pyrrolidinone (compound **1**), 0.015 mol benzaldehyde, and 15 mL of 10% NaOH in 10 mL of ethanol was refluxed for 4 h and cooled. Then, 15 mL water was added and the mixture was acidified with aq. HCl to pH 1−2. The precipitate was filtered and then washed with water. Light yellow solid, yield 70.0%, mp 209−210 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.866 (s, 1H), 8.219 (d, J = 8.68 Hz, 2H), 7.9840−7.8699 (m, 5H), 7.7642 (d, J = 15.48 Hz, 1H), 7.4806 (t, J = 5.28 Hz, 3H), 4.1723−4.0453 (m, 2H), 3.4300−3.3105 (m, 1H), 2.8883−2.7406 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 188.18, 174.51, 173.11, 144.08, 143.69, 135.22, 133.21, 131.02, 130.08, 129.38, 129.33, 122.44, 199.04, 50.30, 35.85, 35.52. MS m/z [M+H]$^+$ calcd. for $C_{20}H_{17}NO_4$ 335.1158, found 336.1246.

2.3.3. 1-(4-(4-nitrostyrylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinone (2b)

Compound **2b** was prepared by following the procedure for preparing **2a**, except using 4-nitrobenzaldehyde. Light yellow solid, yield 30.2%, mp 224−225 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.903 (s, 1H), 8.302 (q, j = 8.8 Hz, 4H), 8.193 (q, j = 1.9 Hz, 3H), 7.903 (q, j = 8.9 Hz, 3H), 4.174−4.045 (m, 2H), 3.442−3.366 (m, 1H), 2.892−2.743 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 187.95, 174.50, 173.20, 148.51, 143.98, 141.74, 141.20, 132.79, 130.32, 126.47, 124.39, 119.03, 50.31, 35.87, 35.53. MS m/z [M+H]$^+$ calcd. for $C_{20}H_{16}N_2O_6$ 380.1008, found 381.1075.

2.3.4. 1-(4-(3-nitrostyrylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinone (2c)

Compound **2c** was prepared by following the procedure for preparing **2a**, except using 3-nitrobenzaldehyde. Light yellow solid, yield 63.9%, mp 234−235 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.771 (s, 1H), 8.343 (d, j = 7.64 Hz, 1H), 8.2684−2.2268 (m, 3H), 8.194 (d, j = 15.64 Hz, 1H), 7.883 (t, j = 12.32 Hz, 2H), 7.813 (s, 1H), 7.760 (t, j = 15.8 Hz, 1H). 4.1411−4.0285 (m, 2H), 3.3733−3.2973 (m, 1H), 2.8543−2.7233 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 187.48, 174.15, 172.93, 148.39, 143.49, 140.99, 136.64, 135.04, 132.30, 130.30, 129.80, 124.68, 124.55, 122.98, 118.45, 50.07, 35.60, 35.35. MS m/z [M-H]$^+$ calcd. for $C_{20}H_{16}N_2O_6$ 380.1008, found 379.0902.

2.3.5. 1-(4-(3-(furan-2-yl)acryloyl)phenyl)-4-carboxy-2-pyrrolidinone (2d)

Compound **2d** was prepared by following the procedure for preparing **2a**, except using furan-2-carboxaldehyde. Brown solid, yield 49.7%, mp 205−206 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.820 (s, 1H), 8.126 (d, j = 8.7 Hz, 2H), 7.924 (t, j = 8.6, Hz, 3H), 7.571 (s, 2H), 7.118 (d, j = 3.3 Hz, 1H), 6.709−6.697 (m, 1H), 4.1641−4.033 (m, 2H), 3.435−3.349 (m, 1H), 2.890−2.735 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 187.10, 174.02, 172.61, 151.17, 146.12, 143.12, 132.64, 130.12, 129.34, 118.58, 116.90, 113.10, 49.79, 35.34, 35.02. MS m/z [M+H]$^+$ calcd. for $C_{18}H_{15}NO_5$ 325.0950, found 326.0970.

2.3.6. 1-(4-(3-(5-nitrothiophene-2-yl)acryloyl)phenyl)-4-carboxy-2-pyrrolidinone (2e)

Compound **2e** was prepared by following the procedure for preparing **2a**, except using 5-nitrothiophene-2-carboxaldehyde. Light yellow solid, yield 34.6%, mp 270−271 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.835 (s, 1H), 9.940 (s, 1H), 8.288 (d, j = 3.16, 1H), 7.986 (d, j = 8.84, 2H), 7.828 (d, j = 8.84, 2H), 7.52−504 (m, 1H), 7.285−7.199 (m, 1H), 4.131−4.008 (m, 2H), 3.410−3.364 (m, 1H), 2.875−2.727 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 197.10, 185.40, 174.49, 173.04, 143.58, 138.87, 137.51, 129.63, 124.59, 123.91, 122.57, 121.28, 118.89, 112.87, 50.26, 35.81, 35.50, 26.95. MS m/z [M+H]$^+$ calcd. for $C_{18}H_{14}N_2O_6S$: 386.0573, found 386.0581.

2.3.7. 1-(4-(3-(1H-indol-4-yl)acryloyl)phenyl)-4-carboxy-2-pyrrolidinone (2f)

Compound **2f** was prepared by following the procedure for preparing **2a**, except using indole-4-carboxaldehyde. Dark brown solid, yield 47.0%, mp 232−233 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.904 (s, 1H), 11.458 (s, 1H), 8.222 (d, j = 8.8 Hz, 2H), 8.137 (d, j = 15.6 Hz, 1H), 7.983 (d, j = 15.5 Hz, 1H), 7.902 (d, j = 8.8 Hz, 2H), 7.671 (d, j = 7.4 Hz, 1H), 7.556 (t, j = 4.2 Hz, 2H), 7.220 (t, j = 7.7 Hz, 1H), 6.886 (s, 1H), 4.175−4.047 (m, 2H), 3.441−3.387 (m, 1H), 2.894−2.743 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 188.21, 174.56, 173.11, 143.52, 143.25, 136.93, 133.56, 129.92, 127.71, 127.60, 127.36, 121.54, 121.51, 121.12, 119.10, 114.77, 100.39, 50.34, 35.88, 35.58. MS m/z [M+H]$^+$ calcd. for $C_{22}H_{18}N_2O_4$ 374.1267, found 375.1326.

2.3.8. 1-(4-(3-(1H-indol-3-yl)acryloyl)phenyl)-4-carboxy-2-pyrrolidinone (2g)

Compound **2g** was prepared by following the procedure for preparing **2a**, except using indole-3-carboxaldehyde. Dark yellow solid, yield 52.2%, mp 230−231 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.849 (s, 1H), 11.450 (s, 1H), 8.220−8.097 (m, 2H), 7.993−7.971 (m, 2H), 7.902 (d, j = 8.88 Hz, 1H), 7.835 (d, J = 8.84 Hz, 1H), 7.668 (d, J = 7.74 Hz, 1H) 7.560−7.537 (m, 1H), 7.220 (t, j = 15.44 Hz, 1H) 6.885 (s, 1H), 6.639 (d, j = 8.8 Hz, 1H), 4.177−4.009 (m, 2H), 3.442−3.366 m, 1H), 2.894−2.723 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 174.55, 173.12, 167.32, 143.62, 143.33, 130.61, 129.93, 121.57, 121.48, 121.01, 119.91, 114.80, 100.40, 100.10, 50.31, 35.80, 35.52. MS m/z [M+H]$^+$ calcd. for $C_{22}H_{18}N_2O_4$ 374.1267, found 374.1238.

2.3.9. 1-(4-acetylphenyl)-2-pyrrolidinone (3)

For making compound **3**, a reported procedure for Copper/N,N-Dimethylglycine catalyzed Goldberg reactions was followed (Jiang, 2014). A Schlenk tube was charged with amide (1.2 mmol), aryl halide (1 mmol), CuI (0.05 or 0.1 mmol), *N,N*-dimethylglycine (0.1 or 0.2 mmol), and potassium carbonate (2 mmol). The tube was evacuated and backfilled with argon at room temperature. DMF (0.5 mL) was added

doi:10.6342/NTU202502385

under argon via syringe. The Schlenk tube was immersed in a preheated oil bath and the reaction mixture was stirred for the specified time at the indicated temperature. The cooled mixture was partitioned between water and ethyl acetate. The organic layer was separated and the aqueous layer was extracted with ethyl acetate. The combined organic layers were washed with brine, dried over Na_2SO_4, and concentrated *in vacuo*. The residue was purified by column chromatography on silica gel (eluting with 1:8 to 1:2 ethyl acetate/petroleum ether) to give the product **3**, a non-carboxylate analogue of **1**. Brown solid, yield 52.3%, mp 172−173 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.988 (d, j = 8.8 Hz, 2H), 7.831 (d, j = 8.8 Hz, 2H), 3.906 (t, j = 7.0 Hz, 2H), 3.342 (d, j = 1.5 Hz, 2H), 2.553 (s, 3H), 2.123−2.0475 (m, 2H), ^{13}C NMR (100 MHz, DMSO-d_6) δ 197.1, 175.1, 144.1, 132.4, 129.6, 118.8, 48.4, 32.9, 27.0, 17.8. MS m/z $[M+H]^+$ calcd. for $C_{12}H_{13}NO_2$ 203.0946, found 204.1021.

2.3.10. 1-(4-styrylcarbonylphenyl)-2-pyrrolidinone (**4a**)

A mixture of 2.47 g (0.01 mol) 1-(4-Acetylphenyl)-2-pyrrolidinone **3**, 0.015 mol benzaldehyde, and 15 mL 10% NaOH in 10 mL of ethanol was refluxed for 4 h and cooled. Then, 15 mL water was added and the mixture was acidified with aq. HCl to pH 1−2. The precipitate was filtered and then washed with water. The residue was purified by column chromatography on silica gel (eluting with 2:8 hexane/ethyl acetate) to give the desired product. Yellow solid, yield 63.7%, mp 280−281 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.988 (d, J = 8.84 Hz, 2H), 7.9037−7.8334 (m, 3H), 7.6419 (d, J = 15.56 Hz, 1H), 7.4751−7.4221 (m, 2H), 6.7823 (t, J = 10.76 Hz, 1H), 6.6592 (d, J = 8.92 Hz, 2H), 3.1746−3.1257 (m, 2H), 2.364 (t, J = 14.48 Hz, 2H), 1.8288−1.7568 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 186.31, 174.75, 153.60, 141.84, 135.66, 131.51, 129.31, 128.97, 125.70, 122.93, 111.40, 42.03, 31.57, 24.42. MS m/z $[M+H]^+$ calcd. for $C_{19}H_{17}NO_2$ 291.1259, found 292.1346.

2.3.11. 1-(4-(3-(1H-indol-4-yl)acryloyl)phenyl)-2-pyrrolidinone (**4b**)

Compound **4b** was prepared by following the procedure for preparing **4a**, except using indole-4-carboxaldehyde. Dark yellow solid, yield 62.8%, mp 228−231 °C, ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.459 (s, 1H), 8.223−8.099 (m, 3H), 7.983−7.881 (m, 3H), 7.672 (d, j = 7.6 Hz 1H), 7.556 (t, j = 7.5 Hz 2H), 7.221 (t, J = 7.2 Hz 1H), 6.886 (s, 1H), 4.175−4.047 (m, 2H), 3.442−3.387 (m, 2H), 2.894−2.743 (m, 2H).^{13}C NMR (100 MHz, DMSO-d_6) δ 193.84, 188.44, 174.64, 173.23, 143.40, 136.90, 133.57, 129.93, 127.75, 127.57, 126.26, 121.64, 121.42, 121.17, 199.22, 144.87, 122.66, 100.39, 50.37, 35.84, 35.53. MS m/z $[M+H]^+$ calcd. for $C_{21}H_{18}N_2O_2$ 330.1368, found 330.1394.

2.4. TMPRSS2 preparation and IC_{50} measurements

Human TMPRSS2 is composed of an intracellular domain, a single-pass transmembrane domain, and a biologically active ectodomain with three subdomains, a low-density lipoprotein receptor type-A (LDLR-A) domain, a Class A Scavenger Receptor Cysteine-Rich (SRCR) domain, and a C-terminal trypsin-like serine protease (SP) domain with a canonical Ser441-His296-Asp345 catalytic triad. For expression of the TMPRSS2 ectodomain (residues 109–492) lacking the N-terminal transmembrane domain, a published procedure was followed (Fraser et al., 2022). The gene synthesized by Toolsbiotech Co. (Taiwan) was constructed into a pFastbac1 vector with a N-terminal honeybee melittin signal sequence and a C-terminal His_8-tag. The activation sequence SSQSR255↓IVGGE (arrow indicates the cleavage site) on the target protein was replaced with an enteropeptidase cleavable DDDDK255 graft by PCR with a forward primer 5′-GATGATGACGACAAGATTGTGGGCGGCGAGAGC-3′ and a reverse primer 5′-GTTCAAGTTGACCCCGCAGGC-3′ for generating S251D/R252D/Q253D/S254D/R255K mutations. The plasmid for TMPRSS2 expression was transformed into *E. coli* DH10Bac cells (catalog no. 10361012, Thermo Fisher Scientific, MA, USA) to generate recombinant viral Bacmid DNA. ExpiSf Baculovirus Expression System (catalog no. A38841, A39111, A39112, Thermo Fisher Scientific) was used for high-level protein expression in insect host cells. ExpiSf9 cells were transfected with Bacmid DNA using ExpiFectamine Sf Transfection Reagent to produce the recombinant baculovirus particles which were then amplified from P0 to P1 viral stocks. Recombinant P1 viruses were used to generate suspension culture of baculovirus infected ExpiSf9 insect cells in ExpiSf CD Medium to a density of 5 × 10^6 cells/ml and infected with the suspension culture of baculovirus infected insect cells before incubation on an orbital shaker (125 r.p.m., 27 °C).

After 4 days of infection, culture supernatant containing the secreted TMPRSS2(109–492)-EFVEHHHHHHH with S251D/R252D/Q253D/S254D/R255K mutations was harvested by centrifugation at 6,000g for 20 min. The pH of 1L cell-free medium was adjusted to pH 7.4 by addition of concentrated 10× phosphate buffered saline (PBS) stock, and loaded onto a HisTrap column (cytiva, MA, USA) to capture the target protein. The HisTrap column was washed separately with 10-column volume of PBS buffer (pH 7.4) containing 10 and 25 mM imidazole, and then eluted with 5-column volume of PBS buffer (pH 7.4) containing 50, 100, and 250 mM imidazole each. The eluted protein solution was pooled and concentrated to 5 mg/mL with a 30 kDa MWCO Amicon filter (Merck, Germany). The target protein was activated by adding 13 units of enteropeptidase (New England Biolabs, MA, USA) per mg of TMPRSS2, and dialyzed against assay buffer (25 mM Tris, pH 8.0, 75 mM NaCl, and 2 mM $CaCl_2$) at room temperature overnight. Activated TMPRSS2 ectodomain (referred as TMPRSS2 here) solution was exchanged to the size-exclusion chromatography buffer (50 mM Tris, pH 7.4, and 250 mM NaCl), spun down at 21,000g, and then loaded to a Superdex 75 10/300 GL column (GE Healthcare, Chicago, USA). The fractions spanning the dominant peak eluted at 9.74 mL was confirmed to be TMPRSS2 by reducing SDS-PAGE, and were subsequently pooled and concentrated. Aliquots of concentrated (0.19 mg/mL~4.3 μM) stock in the buffer (50 mM Tris pH 7.4, 250 mM NaCl, and 25% glycerol) were flash-frozen by liquid nitrogen and stored at −80 °C.

TMPRSS2 activity was assayed using a fluorogenic substrate Boc-Gln-Ala-Arg-AMC (catalog no. 4017019.0025, Bachem, Bubendorf, Switzerland), and the signal was detected by the BioTek Synergy H1 microplate reader (Agilent Technologies, Inc., CA, USA). In the 96-well plate, the 100-μL reaction mixtures per well containing 1.3 nM TMPRSS2, 10 μM substrate, and different concentrations of inhibitors in the buffer of 20 mM Tris-HCl pH 7.4 with 1% DMSO for dissolving the compounds. The fluorescence was monitored for 5 min at excitation/emission of 355/460 nm. IC_{50} values, the concentrations of inhibitors required to reduce half of the enzymatic activities, were determined by pre-incubating TMPRSS2 with inhibitors at concentrations ranging from approximately half to 5-fold IC_{50} values for 5 min and fitting the concentration-dependent TMPRSS2 inhibition curves with equation A (I) = A (0) × {1-[I/(I + IC_{50})]} using GraphPad Prism software (v.9.4.0). In this equation, A(I) is the enzyme activity with inhibitor concentration I, A (0) is the enzyme activity without inhibitor, and I is the inhibitor concentration. For each inhibitor, the measurements were repeated three times to yield the averaged IC_{50} and the standard deviation.

2.5. Furin IC_{50} measurements

Furin activity was assayed using a commercial Furin Protease Assay Kit (catalog # 78040, BPS Bioscience, CA, USA). In 20-μL reaction mixture on a 384-well plate, it contained 0.25 ng/μL Furin, 2 μM fluorogenic substrate, and different concentrations of inhibitors in assay buffer with 1% DMSO for dissolving compounds. The fluorescence change was monitored for 30 min at excitation/emission of 380/460 nm. IC_{50} values were determined by pre-incubating Furin with inhibitors at concentrations ranging from approximately half to 5-fold IC_{50} values for 5 min and the concentration-dependent Furin activity inhibition curves were fitted with equation A(I) = A (0) × {1-[I/(I + IC_{50})]} using GraphPad Prism software (v.9.4.0). In this equation, A(I) is the enzyme activity with inhibitor concentration I, A (0) is the enzyme activity

doi:10.6342/NTU202502385

without inhibitor, and I is the inhibitor concentration. For each inhibitor, the measurements were repeated three times to yield the averaged IC_{50} and the standard deviation.

2.6. Antivirus EC_{50} and cytotoxicity CC_{50} measurements

Two methods, the plaque reduction assay and the yield reduction assay, were used to determine the EC_{50} of compounds against SARS-CoV-2. All experiments involving the SARS-CoV-2 virus were performed in the Biosafety Level-3 Laboratory of the First Core Laboratory, National Taiwan University College of Medicine. For the viral plaque reduction assay, Vero E6 cells were seeded into a 24-well culture plate in Dulbecco's Modified Eagle Medium (DMEM) with 10% Fetal Bovine Serum (FBS) and antibiotics one day before the infection. VeroE6 cells were infected by SARS-CoV-2 delta virus (NTU92) or omicron BA.1 (NTU128) at 50−100 pfu (plaque forming unit) for 1 h at 37 °C. After removal of virus inoculum, the cells were washed once with PBS and overlaid with 1 mL overlay medium containing 1% methylcellulose for 5 days at 37 °C. After 5 days, the cells were fixed with 10% formalin overnight. After removal of the overlay medium, the cells were stained with 0.5% crystal violet and the plaques were counted. The percentage of inhibition was calculated as $[1-(V_D/V_C)] \times 100\%$, where V_D and V_C refer to the virus titer in the presence and absence of the inhibitors, respectively. The minimal concentrations of the compounds required to reduce plaque numbers by 50% (EC_{50}) were calculated by regression analysis of the dose-response curves.

To characterize the EC_{50} of the compound against SARS-CoV-2 infecting human Calu-3 cells, yield reduction assays were performed because the virus could not form plaques on the Calu-3 cells. In brief, Calu-3 cells were seeded into 24-well culture plates and cultured for 4 days. About 6000 pfu of viruses were pretreated with the compound or DMSO for 1 h before infection. After 1 h of infection, the virus-containing medium was removed, and cells were incubated in DMEM supplemented with 2% FBS and 1% antibiotics for 24 h. Afterward, culture supernatant was harvested for virus titration. Subsequently, the infectious virus titer in the culture supernatant was quantified by plaque assay in Vero E6 cells. The procedures of the plaque assay were described below. Vero E6 cells (2×10^5 cells/well) were seeded in 24-well plates in DMEM supplemented with 10% FBS and antibiotics one day before infection. The supernatant was 10-fold serially diluted and added to the cell monolayer for 1 h at 37 °C. Subsequently, the virus was removed, and the cell monolayer was washed once with PBS before overlay medium containing 1% methylcellulose was added. The cells were then incubated for 5 days. Afterward, the cells were fixed with 10% formaldehyde overnight. The overlay medium was removed, and the cells were stained with 0.5% crystal violet, and the plaques were counted. The inhibition percentage of compound **81** against viruses was calculated as $[1-(V_D/V_C)] \times 100\%$, where V_D and V_C refer to the virus titer in the presence and absence of the compound, respectively. The EC_{50} values were calculated by regression analysis of the dose-response curves.

Cytotoxicity of the inhibitors was determined by using the acid phosphatase assay. Briefly, Vero E6 or Calu-3 cells were seeded onto a 96-well culture plate at a concentration of 2×10^4 cells per well. Next day, the medium was removed and each well was washed once with PBS before addition of DMEM containing 2% FBS and different concentrations of inhibitors. After 3 days of incubation at 37 °C, medium was removed and each well was washed once with PBS. Next, a buffer containing 0.1 M sodium acetate (pH = 5.0), 0.1% Triton X-100, and 5 mM p-nitrophenyl phosphate (P5994, Sigma-Aldrich/Merck, Germany) was added. After incubating at 37 °C for 30 min, 1 N NaOH was added to stop the reaction. The absorbance was determined by ELISA reader (VERSAmax, Molecular Devices, Sunnyvale, CA, USA) at a wavelength of 405 nm. The percentage of cytotoxicity was calculated using the following formula: cytotoxicity % = [(At/As) × 100]%, where At and As refer to the absorbance of a tested substance and solvent control, respectively. The 50% cytotoxicity concentration (CC_{50}) was defined as the concentration reducing 50% of cell viability. For each data point, the measurements were repeated three times to yield the averaged number and the standard deviation. The CC_{50} values were calculated by non-linear regression analysis of the dose-response curves.

2.7. RBD:ACE2 interaction IC_{50} measurements and pseudovirus assays

Inhibition of RBD:ACE2 binding by the synthesized compounds were measured by using the commercial kit of NanoBiT technology (Promega, WI, USA). A stable cell line expressing SmBiT-tagged human ACE2 on HeLa cells was established and the recombinant RBD-LgBiT protein (amino acid 330–521 of Spike) was produced (Lee et al., 2022). To monitor the interaction between RBD and ACE2, SmBiT-ACE2 expressing cells were seeded onto a white 96-well plate at a density of 1×10^4 cells per well. For each binding assay, cells were washed once with PBS and pretreated with 50 μL of an indicated compound per well for 10 min. Next, a 50 μL reaction mixture containing 10 ng RBD-LgBiT, 0.5 μL of Nano-Glo luciferase assay substrate, and 9.5 μL of luciferase assay diluent (Promega) was added into each well. Luminescence signal was recorded every 2 min and continuously for 1 h by microplate reader (BioTek Synergy HTX, VT, USA) at 37 °C with time-lapsed kinetics program. For calculating RBD inhibition of all agents, luminescent data from the time point showing highest signal in negative control sample was chosen for downstream calculation. Inhibition (%) = [1- (luminescence signal of test sample)/(luminescence signal of negative control sample)] × 100%. The compound concentration for inhibiting 50% of the interaction was defined as IC_{50}. The measurements were performed in triplicate.

2.8. Pseudovirus assays

For producing lentivirus-based pseudoviruses (Lee et al., 2022), 5 μg transfer plasmid (pLAS2w.Nluc-T2A-RFP-C.Ppuro), 4 μg packaging plasmid (pCMVdeltaR8.91 from the RNAicore, Academia Sinica, Taipei, Taiwan), and 1 μg of the envelope glycoprotein of vesicular stomatitis virus (the VSV-G protein) or spike-expressing plasmid (delta derivatives of pcDNA3.1-2019-nCoV-S-d18, from the RNAicore, Academia Sinica, Taiwan) were co-transfected with Lenti-X 293T cells. Virus-containing supernatants were collected and quantitated by real-time PCR. For pseudovirus infection assays, HeLa cell line stably expressed ACE2-SmBiT (Hela-ACE2) or Calu-3 cell line was seeded into the 96-well white plate at a density of 1×10^4 cells per well in the presence of 100 μL DMEM. After 16 h, cells were treated with 2×10^5 copies of VSV-G or spike pseudovirus containing delta Spike and NanoLuc reporter (Lee et al., 2022) along with a two-fold serial dilution compound **81**. After 6 h of incubation, cells were washed with PBS three times and then changed with complete DMEM for additional 24 h. To measure the NanoLuc activity of the pseudovirus particles inside the cells, a mixture of 50 μL OPTI-MEM containing 8 μM substrate of Nanoluc, furimazine (TargetMol) was added to each well. The luminescence signal was immediately recorded using a luminescent microplate reader (BioTek Synergy HTX, Agilent Technologies) at 37 °C with a time-lapsed kinetics program of 2-min intervals for 30 min. To calculate the percentage of infectivity, the luminescent data from the time point showing the highest signal in the wild-type ACE2 infected with indicated pseudoviruses sample was chosen. The inhibition (%) was calculated as follows: 1-(luminescence signal of the test sample)/(luminescence signal of DMSO vehicle control) × 100%. The compound concentration for inhibiting 50% of the pseudovirus entry was defined as EC_{50}. The measurements were performed in triplicate.

2.9. Expression and purification of the recombinant SARS-CoV-2 delta variant RBD and human ACE2

Expression of RBD followed a reported procedure (Lan et al., 2020).

doi:10.6342/NTU202502385

The gene encoding delta strain RBD was PCR amplified with a ligation-independent cloning-compatible primer pair (forward primer 5′-TCTTACATCTATGCGAGAGTCCAACCAACAGAA-3′ and reverse primer 5′-GTGCTCTACGAACTCGAAATTGACACATTTGTTTTT-3′). The PCR product containing sequences encoding an N-terminal melittin signal sequence for secretion and a C-terminal His_6-tag for purification, was cloned into pFastBac-1 vector. The sequencing-verified plasmid was transformed into DH10Bac competent cells to generate a bacmid encoding RBD (residues Arg319–Phe541). The ExpiSf Baculovirus Expression System with a chemically defined medium for suspension culture of high-density ExpiSf9 cells was used for virus production and protein overexpression. After 48–72 h of culturing, ExpiSf9 cells were collected by centrifugation and resuspended in HBS buffer (10 mM HEPES, pH 7.2, and 150 mM NaCl). The recombinant delta SARS-CoV-2 RBD was first captured by Ni-NTA resin (Agarose Bead Technologies, USA) and eluted with 500 mM imidazole in HBS buffer. Then, it was purified by gel filtration chromatography using a pre-equilibrated Superdex 75 10/300 GL column (GE Healthcare) and eluted with HBS buffer.

Expression of ACE2 was performed by a previously reported procedure (Lan et al., 2020). The gene was provided by Dr. Shang-Te Danny Hsu in our institute and PCR amplified with a ligation-independent cloning-compatible primer pair (foreard primer 5′-TCTTACATCTATGCGTCCACCATTGAGGAACAG-3′ and reverse primer 5′-GTGCTCTACGAACTCGTCTGCATATGGACTCCA-3′). The PCR product containing sequences encoding an N-terminal melittin signal sequence for secretion and a C-terminal His_6-tag for purification, was cloned into pFastBac-1 vector. The sequencing-verified plasmid was transformed into DH10Bac competent cells to generate a bacmid encoding the N-terminal peptidase domain of human ACE2 (residues Ser19–Asp615). The ExpiSf Baculovirus Expression System with a chemically defined medium for suspension culture of high-density ExpiSf9 cells was used for virus production and protein overexpression. After 48–72 h, ExpiSf9 cells were collected by centrifugation and resuspended in HBS buffer (10 mM HEPES, pH 7.2, 150 mM NaCl). The N-terminal peptidase domain of human ACE2 was subsequently purified using the same protocol for RBD purification described above.

2.10. Thermal shift experiments

The experiments were performed as described (Bojadzic et al., 2021) to determine which of RBD or ACE2 bound **2f** (**81**). A RT-PCR machine (CFX Real-Time PCR System, Bio-Rad Laboratories, CA, USA) was programmed to equilibrate each sample at 25 °C for 90 s and then increase temperature to 95 °C by 0.5 °C every 10 s for taking a fluorescence reading of the denatured portion of protein stained with SYPRO Orange dye (Merck KGaA, Darmstadt, Germany). Melting point of each protein was the lowest point of the first derivative plot, as calculated by the software (Bio-Rad CFX Manager ver.3.0) included with the RT-PCR machine. After performing a series of preliminary scans at various concentrations of proteins, compound, and dye, optimal concentrations were determined to be 0.05 mg/mL RBD, 0.05 mg/mL ACE2, 50 × SYPRO Orange, 1× PBS, and 10 μM compound **2f**. Using this optimal condition, the plots of -d (RFU)/dT (RFU is relative fluorescence units) vs. temperatures for RBD or ACE2 in the absence or presence of **81** were compared.

2.11. Molecular docking

The molecular docking was performed using the iGEMDOCK software to predict how compound **2f** binds with TMPRSS2, Furin, and RBD of SARS-CoV-2 delta variant Spike protein. The three-dimensional (3D) structures of TMPRSS2 (PDB: 7meq), Furin (PDB: 5jxg), and RBD (PDB: 7w92) were retrieved from the RCSB protein data bank (PDB, https://www.rcsb.org/). All water molecules and bound ligands were removed from these structures prior to docking. The binding cavity of TMPRSS2 were extracted by selecting the 12 Å residues around the catalytic triad (Ser441-His296-Asp345), and the binding cavity of Furin was extracted by selecting the 12 Å residues around the catalytic residues, Asp153, His194, Asn295, and Ser368. RBD domain (residues 319 to 541) was extracted from the open state of SARS-CoV-2 delta variant Spike protein. The cavities were then prepared by defining the residue atom types and charges assignment via iGEMDOCK method. Compound **2f**'s 3D structure was generated from Molview website (https://molview.org/), and the structural information was transformed into mol2 format by using Open Babel GUI software.

For molecular docking process, docking accuracy settings (GA parameters including population size: 800, generations: 80, and number of solutions: 10) were chosen, and iGEMDOCK software was used to generate protein-ligand interaction profiles of Electrostatic (E), Hydrogen-Bonding (H), and van der Waals interaction (V). Once docking was finished, iGEMDOCK was used to analyze and rank all docked poses based on the estimated binding energy. The scores of the docked poses were estimated as the total energy of Electrostatic (E) + Hydrogen-Bonding (H) + van der Waals (V) in the docking site. Each docked pose with the lowest energy was considered as the best pose for compound **2f** against the target cavity.

2.12. Drug likeness analysis

The 3D structure of **2f** was converted into a SMILES format by Open Babel GUI software (O'Boyle et al., 2011). To assess the drug likeness of compound **2f**, the online tool (http://www.scfbio-iitd.res.in/software/drugdesign/lipinski.jsp) was used for Lipinski rule of five estimation (Lipinski 2004; Jayaram et al., 2012). The ADMET profile of **2f** was estimated by pkcsm website (http://biosig.unimelb.edu.au/pkcsm/) (Pires et al., 2015).

2.13. Animal studies

The prototype SARS-CoV-2 strain CGU4 (Taiwan/CGMH-CGU-01/2020; GISAID identifier: EPI_ISL_411915; NCBI accession number: MT192759) was kindly provided by the Taiwan CDC and propagated in Vero E6 cells (Tang et al., 2021). All animal procedures complied with the ARRIVE guidelines (www.arriveguidelines/org/) and were reviewed and approved by the National Defense Medical Center Animal Care and Use Committee (approval numbers: AN-111-24). Syrian hamsters were purchased from the National Laboratory Animal Center (Taipei, Taiwan) and fed a standard low-fat polysaccharide-rich chow diet (LabDiet 5053; Purina, St Louis, MO, USA). Intranasal inoculation of SARS-CoV-2 (strain CGU4) was performed under anesthesia (Zoletil® 40 mg/kg plus xylazine 4 mg/kg). The antiviral activity of compound **81** was assessed through the following parameters: daily changes in body weight and virus quantification in the nasal turbinates and lungs. Compound **81** was dissolved in DMSO and further diluted with sterile PBS. Syrian hamsters, ranging in age from 5 to 6 weeks, were divided into two groups. Treatment group (n = 6) received **81** by oral administration (25 mg/kg once per day) and vehicle group (n = 6) received PBS. The hamsters were challenged intranasally with SARS-CoV-2 (5.0×10^4 pfu in 50 μL PBS). The body weight of each hamster was recorded daily. Hamsters were euthanized on day 3 or day 6 post viral challenge by Zoletil® overdosing, and nasal turbinate and lung samples were collected for analyses.

For the RNA extraction and qRT-PCR analysis, nasal turbinate and lung tissues were placed in RNAlater (Thermo Fisher Scientific, MA, USA) overnight at 4 °C and homogenized by bead-beating technology. Total RNA was extracted with TRIzol reagent (Thermo Fisher Scientific) and genomic DNA was removed by DNase (Promega, Madison, WI, USA) according to the manufacturer's instructions. RNA samples were further purified by phenol/chloroform method and ethanol precipitation. Equal amounts of total extracted RNA were subsequently reverse transcribed with the M-MLV reverse transcriptase system (Thermo Fisher Scientific)

doi:10.6342/NTU202502385

using random primers. Subsequently, qRT-PCR was performed on a QuantStudio3 qRT-PCR system (Applied Biosystems, Foster City, CA, USA) using the following the specific primers targeting E gene (forward primer 5′-ACAGGTACGTTAATAGTTAATAGCGT-3′ and reverse primer 5′-ATATTGCAGTACGCACACA-3′). The hamster β-actin (internal control) primers were forward primer 5′-ACTGCCGCATCCTCTTCCT-3′ and reverse 5′-TCGTTGCCAATGGTGATGAC-3′. Probe 5′-[6FAM]ACACTAGCCATCCTTACTGCGCTTCG [BHQ1]-3′ was used for E gene subgenomic mRNA quantitation. The relative mRNA expression was calculated with the $2^{-\triangle Ct}$ method.

For pharmacokinetic studies, a total of eight K18-hACE2 transgenic mice (n = 4 for oral administration and n = 4 for intravenous injection via tail vein) at 12 weeks of age and 20.3–24.2 g body weight were used. Compound **81** was dissolved in DMSO and further diluted with sterile PBS to 1 mg/mL and then administered orally or intravenously via the tail vein as a single injection at a volume of 25 mL/kg. Both oral and intravenous (IV) administration groups were split in half to allow further repeated blood sampling at 0, 15, 30, 60, 120, and 240 min as follows: n = 2 were bled at 0, 15, and 30 min post injection from the facial vein, and at 60 min via cardiac puncture under anesthesia; the other n = 2 were bled at 0, and 120 min post injection from the facial vein, and at 240 min via cardiac puncture under anesthesia. Blood samples were collected in a collection tube (365992, BD Microtainer BD bioscience, USA) and centrifuged at 3000 g for 5 min at 4 °C to separate the serum. The serum samples were stored at −20 °C until LC-MS analysis. Pharmacokinetic analysis of serum compound **81** was performed by a non-compartmental analysis using PKSolver 2.0 software (China Pharmaceutical University, Nanjing, China).

For LC-MS analysis, serum samples were mixed with 100% MeOH (1:10, v/v) and sonicated for 30 min. After centrifuging at 20,000 g for 10 min at 4 °C, the supernatant was evaporated to dryness at 35 °C and then dissolved with 100 μL of MeOH and filtrated. The resulting supernatant was injected into the LC-MS system for analysis. All LC-MS analyses were carried out using the mass spectrometer Shimadzu LCMS-2020 with both positive and negative mode electrospray ionization using a 10 min linear gradient of 5–95% MeCN–H_2O with 0.5 % formic acid followed by 95% MeCN–H_2O for 4 min at a flow rate of 0.5 mL/min (Kinetex® 2.6 μm Polar C18 100 Å, 100 × 2.1 mm column). Calibration curves used 7 concentration points prepared in the concentration ranges 0.25−30 μM of compound **81**. Analyte concentrations were determined using the internal standard method.

3. Results

3.1. Synthesis of 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones 2a–g

The preliminary product 1-(4-acetylphenyl)-4-carboxy-2-pyrrolidinone **1** was synthesized by heating of 4-aminoacetophenone with itaconic acid without any solvent (Scheme 1). Then by following a similar procedure (Voskiene et al., 2007), the compounds **2a–g** were synthesized with good yields by base (NaOH)-catalyzed aldol condensation of **1** and the unsubstituted or substituted benzaldehydes or heterocyclic aldehydes in ethanol (Scheme 1).

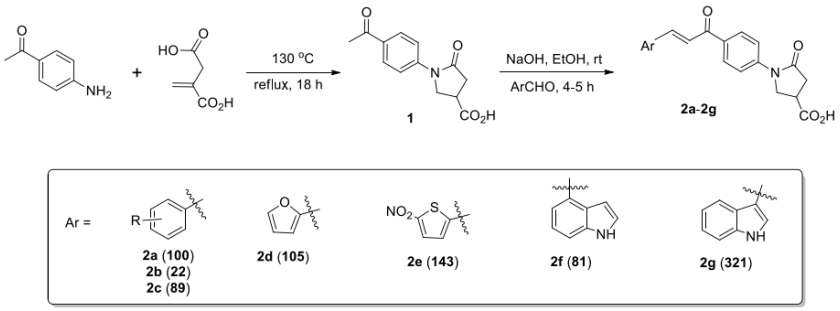
The chemical structures of synthesized compounds were confirmed by NMR and MS. The NMR assignment was made on the substituents additivity rules, spectral characteristics of structurally related compounds, signal intensities and multiplicities. ^{13}C NMR spectra were used to prove the interpretation of the carbon resonances in some cases. The ^{1}H and ^{13}C NMR spectra for structural characterization are presented in Fig. S1.

3.2. Synthesis of 4a and 4b without the carboxylate

The negatively charged carboxylate on the pyrrolidinone may impede the passage of compounds across cellular membrane. Therefore, the carboxylate group was removed from the two selected antiviral compounds to examine whether it could facilitate their cellular uptake. For synthesis of compound **3**, the non-carboxylate analogue of **1**, Scheme 2 was adopted by using a similar procedure (Jiang, 2014). Then, by following Scheme 2, compounds **4a** and **4b**, the analogues of **2a** and **2f**, respectively, without the carboxylate group on the pyrrolidinone moiety were synthesized.

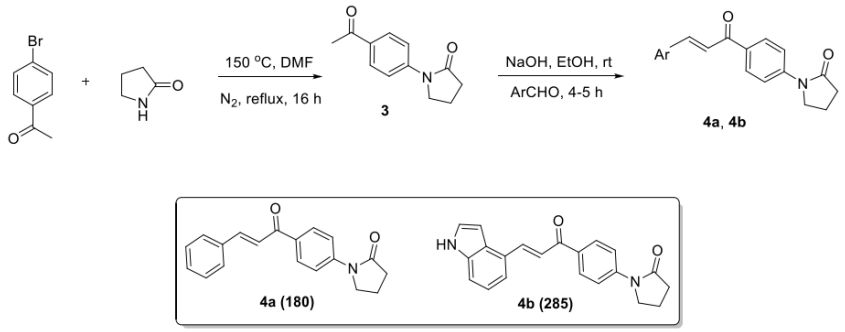
3.3. Evaluation of 2a–g and 4a–b against TMPRSS2 and furin

TMPRSS2 and Furin are involved in SARS-CoV-2 entry by processing of the viral Spike protein. Since they both cleave the Arg-Ser amide bond, it may be possible to design common inhibitors for TMPRSS2 and Furin. Based on our computer modeling on a group of chalcones, we first synthesized **2a–g** and evaluated their IC_{50} against the recombinant human protease TMPRSS2 ectodomain that was expressed using the baculovirus system and activated by the enteropeptidase. The purity of recombinant TMPRSS2 was determined by SDS-PAGE as shown in Fig. S2A. The IC_{50} values of three representative inhibitors with the Ar groups, **2b (22)** (Ar = 4-nitrobenzene), **2e (143)** (Ar = 5-nitrothiophene-2-yl), and **2f (81)** (Ar = indol-4-yl), were 12.46, 3.52, and 1.27 μM, respectively (Fig. 1A–C). Other compounds exhibited less potent inhibitory effects against TMPRSS2. The does-dependent curves



Scheme 1. Synthesis of 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones.

doi:10.6342/NTU202502385



Scheme 2. Synthesis of 1-(4-(arylethylenylcarbonyl)phenyl)-2-pyrrolidinones.

of **2a (100)** (Ar = benzene) and **2c (89)** are shown in Fig. S2B. **2g (321)** with Ar = indol-3-yl linked via different atom from that of **2f (81)**, and **2d (105)** with Ar = furan-2-yl showed no inhibitory activity against TMPRSS2. As summarized in Table 1, while **2a (100)** (Ar = benzene) showed a poor IC_{50} of 28.35 μM, compounds **2b (22)** with a larger 4-NO_2 substituent on the benzene ring increased more than 2-fold of potency against TMPRSS2 (IC_{50} = 12.46 μM). Nevertheless, the 3-NO_2 substituent in compound **2c (89)** did not significantly increase the potency.

The inhibitory activities of the synthesized compounds against Furin and another human protease Cathepsin L were examined. Based on the does-dependent curves of **2b (22)**, **2e (143)**, and **2f (81)** against Furin (Fig. 1D–F), the IC_{50} values were 10.14, 4.55, and 3.31 μM, respectively. and other plots are shown in Fig. S3. All the data are summarized in Table 1. The compounds at 100 μM did not inhibit Cathepsin L (data not shown). From Table 1, the inhibitors of TMPRSS2 showed a similar trend on inhibiting Furin, indicating the two proteases may provide similar environments for binding with the compounds. Therefore, **2d** and **2g** which were inactive against TMPRSS2 did not inhibit Furin as well (Table 1). These indicated similar structure-activity relationship of the synthesized compounds against TMPRSS2 and Furin, consistent with our modeling prediction.

To better understand the importance of carboxylate on the inhibition of TMPRSS2 and Furin, the carboxylate was removed from compound **2a** and **2f** to synthesize **4a** and **4b,** respectively. As shown in Figs. 1G, **4b (285)** inhibited TMPRSS2 with an IC_{50} of 31.64 μM, which was significantly higher than that of **2f**. Similarly, **4a** failed to inhibit TMPRSS2 and Furin (see Table 2). These data indicated that the carboxylate of **2a** and **2f** participated the binding with TMPRSS2 and Furin.

3.4. Evaluation of *2a–g* and *4a–b* against SARS-CoV-2

As mentioned above, the inhibitory activity of compounds on TMPRSS2 and Furin could endow them with antiviral activity because these proteases are essential for viral entry by trimming Spike. We then measured their EC_{50} against SARS-CoV-2 infection of VeroE6 cells under two different ways of treatment. Each compound was either pretreated with the virus and included during the virus infection of the VeroE6 cells (entry treatment), or added only after virus infection (post-entry treatment) to examine at which stage did these compounds block virus infection of the target cells. As expectedly, we found that the compounds suppressed the viral plaque formation by entry treatment, but not by the post-entry treatment. The EC_{50} values measured under "entry treatment" were obtained from the dose-dependent virus (SARS-CoV-2 delta strain) plaque reduction curves by using different concentrations of the compounds as shown in Fig. 2A–C for **2b (22)**, **2e (143)**, and **2f (81)**, respectively. Among them, the best inhibitor **81** showed potent antiviral activities against two different variants, delta and Omicron BA.1 strains, with EC_{50} values of 0.023 and 0.002 μM, respectively, and CC_{50} > 100 μM (Fig. 2C and D). The plots for other compounds against delta SARS-CoV-2 infecting VeroE6 cells are shown in Fig. S4. Furthermore, the antiviral activity of the most potent inhibitor **2f (81)** in human Calu-3 lung cells was determined, and the EC_{50} values against the delta and omicron SARS-CoV-2 infection were 0.026 and 0.001 μM, respectively, based on the dose-dependent SARS-CoV-2 yield reduction and CC_{50} > 100 μM (Fig. 2E and F). All the measured EC_{50} and CC_{50} values were summarized in Table 1. To our surprise, **2d** showing no TMPRSS2 and Furin inhibitory activity at 50 μM still inhibited the virus with a less potent EC_{50} of 1.06 μM. This suggested other target for the active antivirals in this series of compounds. We then chose RBD:ACE2 interaction as a possible target to investigate the inhibitory power of all the active antivirals.

3.5. IC_{50} of the compounds against RBD:ACE2 interaction

A cell-based commercial assay kit (NanoLuc® Binary Technology, NanoBiT), composed of a Large BiT (LgBiT; 18 kDa) subunit conjugated with delta SARS-CoV-2 RBD and SmBiT-ACE2 stably expressing HeLa cells (HeLa-ACE2), was used to measure IC_{50} of test compounds on inhibiting RBD:ACE2 interaction. In the asssy, binding of RBD with ACE2 would bring Large BiT and small SmBiT together to form an active enzyme to generate a bright luminescent signal in the presence of substrate. The presence of inhibitor would weaken the signal, and the dose-dependent inhibition curves against delta virus RBD:ACE2 interaction by the potent antivirals **2b (22)**, **2e (143)**, and **2f (81)** are shown in Fig. 3A, B, and C, respectively. The plots for other compounds are shown in Fig. S5. As summarized in Table 1, all of the synthesized compounds, including the two inactive TMPRSS2 and Furin inhibitors **2d (105)** and **2g (321)**, could inhibit RBD:ACE2 interaction with IC_{50} values between 28.14 and 33.99 μM. The analogues of **2a** and **2f**, **4a (180)** and **4b (285)**, also inhibited RBD:ACE2 interaction (shown in Fig. 3D for **4b** and Fig. S5 for **4a**). From these *in vitro* inhibition studies, the most potent antiviral **2f** actually displayed potent inhibition on enzyme activities of TMPRSS2 and Furin, and RBD:ACE2 interaction, which are concordant with its most potent antiviral activity against SARS-CoV-2. Inhibition of RBD:ACE2 interaction provided the basic antiviral activity for **2d**, **2g**, **4a** (Fig. S5), and **4b** (Fig. 3D), which showed none or minimal inhibition on TMPRSS2 and Furin enzyme activities. For compounds which can inhibit TMPRSS2 and Furin activities, their antiviral activities could be further enhanced, such as **2b**, **2e**, and **2f**.

doi:10.6342/NTU202502385

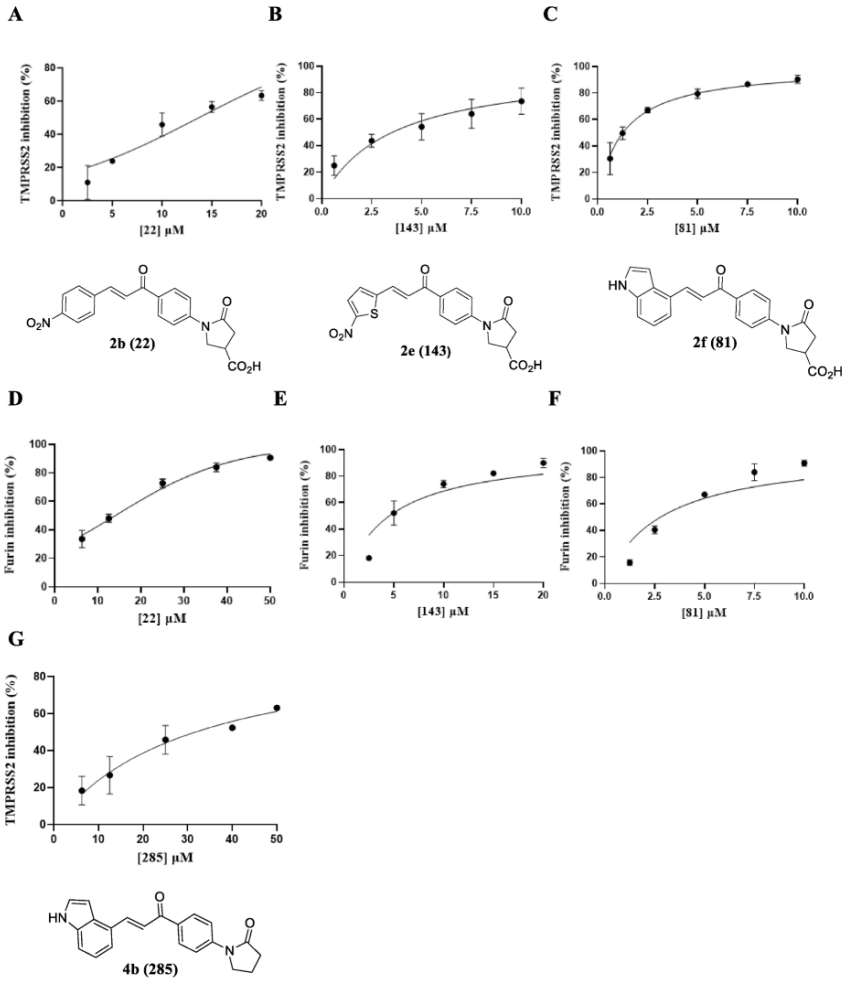


Fig. 1. Dose-dependent inhibition curves of the representative inhibitors **2b** (**22**), **2e** (**143**), and **2f** (**81**) against TMPRSS2 (A–C) and Furin (D–F). In (A–C), the increasing inhibition percentages of TMPRSS2 activity in the presence of increasing concentrations of **2b**, **2e**, and **2f** were fitted with the equation to yield the IC_{50} of 12.46 ± 0.95, 3.52 ± 0.55, and 1.27 ± 0.08 μM, respectively. In (D–F), the increasing inhibition percentages of Furin activity in the presence of increasing concentrations of **2b**, **2e**, and **2f** were fitted with the equation to yield the IC_{50} of 10.14 ± 1.12, 4.55 ± 0.69, and 3.31 ± 0.52 μM, respectively. (G) The increasing inhibition percentages of TMPRSS2 activity in the presence of increasing concentrations of **4b** (**285**) were fitted with the equation to yield the IC_{50} of 31.64 ± 2.93 μM. All the measurements were performed in triplicate to yield the averaged IC_{50} values and the standard deviations.

3.6. Pseudovirus assays

To clarify the contribution of inhibiting RBD:ACE2, TMPRSS2, and/or Furin on antiviral activities of our compounds, we examined the **81**-mediated inhibition of SARS-CoV-2 pseudoviral entry by using a pseudovirus containing delta Spike or VSV-G on the surface and NanoLuc reporter inside to infect the chosen cells. As shown in Fig. 3E, **2f** (**81**) effectively inhibited the pseudovirus entry into Calu-3 lung cells with an EC_{50} of 8.3 μM. As a control, entry of VSV-G bearing pseudovirus that is independent from RBD:ACE2, TMPRSS2, and/or Furin into Calu-3 cells was not inhibited by our compound **81** (Fig. 3F). We showed that the SARS-CoV-2 pseudovirus infected Hela-ACE2 stable cell line with an EC_{50} of 14.7 μM (Fig. 3G), but not its parental HeLa cells (Fig. not shown), indicating the requirement of inhibiting RBD:ACE2 for antiviral activity. We further showed that pseudoviral entry into Calu-3 cells that express TMPRSS2 could be potently inhibited by the specific TMPRSS2 inhibitor Camostat with an EC_{50} of 0.04 μM (Fig. 3H), but not by the specific Furin inhibitor decanoyl-RVKR-chloromethylketone (Fig. 3I). As

doi:10.6342/NTU202502385

Table 1
Compounds **2a–g** inhibiting TMPRSS2, Furin, RBD:ACE2 and SARS-CoV-2.

Entry	Compd	Ar	IC_{50} (μM)			EC_{50} (μM)	CC_{50} (μM)
			RBD:ACE2	Furin	TMPRSS2	SARS-CoV-2	VeroE6
1	**2a (100)**		16.55 ± 3.92	23.12 ± 2.35	28.35 ± 2.46	0.26 ± 0.06	>100
2	**2b (22)**		7.15 ± 1.52	10.14 ± 1.12	12.46 ± 0.95	0.15 ± 0.10	>100
3	**2c (89)**		16.97 ± 4.92	18.89 ± 2.30	18.20 ± 1.45	0.70 ± 0.18	>100
4	**2d (105)**		28.14 ± 8.17	>50	>50	1.06 ± 0.54	>100
5	**2e (143)**		2.81 ± 0.49	4.55 ± 0.69	3.52 ± 0.55	0.087 ± 0.010	59.9 ± 0.99
6	**2f (81)**		3.76 ± 0.84	3.31 ± 0.52	1.27 ± 0.08	0.023 ± 0.005/0.002 ± 0.002[a] 0.026 ± 0.012/0.001 ± 0.0007[b]	>100
7	**2g (321)**		33.99 ± 11.42	>50	>50	3.78 ± 0.60	>100

[a] The EC_{50} of **81** for inhibiting delta/omicron BA.1 SARS-CoV-2 infecting VeroE6 cells.
[b] The EC_{50} of **81** for inhibiting delta/omicron BA.1 SARS-CoV-2 infecting Calu-3 cells.

Table 2
Compounds **4a** and **4b** inhibiting TMPRSS2, RBD:ACE2, and SARS-CoV-2.

Entry	Compd	Ar	IC_{50} (μM)			EC_{50} (μM)	CC_{50} (μM)
			RBD:ACE2	Furin	TMPRSS2	SARS-CoV-2	VeroE6
8	**4a (180)**		25.04 ± 9.26	>50	>50	2.09 ± 0.06	>100
9	**4b (285)**		51.72 ± 51.84	>50	31.64 ± 2.93	0.67.± 0.38	>100

expected, Camostat could not inhibit pseudoviral entry into HeLa-ACE2 cells that lack TMPRSS2 (Fig. 3J). Taken together, compound **81** inhibited mainly RBD:ACE2 and additionally by TMPRSS2, but not necessarily by Furin, for anti-pseudovirus activity on Calu-3 cells.

3.7. Thermal shift experiments to distinguish the target for RBD:ACE2 inhibitors

To determine whether the RBD:ACE2 inhibitors targeted RBD and/or ACE2, we expressed and purified recombinant RBD and ACE2 proteins (Fig. S6A). To prove the recombinant RBD and ACE2 were functional, RDB was coated on the CM5 chip and ACE2 was flown through to measure the ACE2 concentration-dependent sensograms using Biacore apparatus (Fig. S6B), which allowed the determination of RBD:ACE2 K_d of 81.28 nM, a tight binding. Nevertheless, **2f** (**81**) did not influence the concentration-dependent binding curves with either ACE2 or RBD coating on a chip, probably due to not sufficient mass change to be detected. Thermal shift experiments were conducted instead. When adding increasing concentrations of **2f**, no changes of the melting temperature (T_m) of ACE2 was observed (Fig. 4A), whereas T_m of RBD was shifted (Fig. 4B), indicating binding of **2f** with RBD only.

3.8. Binding modes of the inhibitor with TMPRSS2, furin, and RBD

To rationalize the structure-activity relationship of the antivirals on inhibiting TMPRSS2, a docking study was conducted with the clues that the carboxylate on the pyrrolidinone ring and the indole ring on the other side are important for **81**-mediated inhibition. As shown in Fig. 5A, the best inhibitor **2f** (**81**) was docked into the active site of TMPRSS2. There are three major binding interactions contributed by the carboxylate of pyrrolidinone, the carbonyl group, and the indole ring, respectively. The carboxylate on the pyrrolidinone ring forms two hydrogen bonds with the hydroxyl group of Thr393. This explains why the compound showed a significantly weaker inhibitory activity against TMPRSS2 after removing the carboxylate. The carbonyl group forms a hydrogen bond with the side-chain N atom of His296. Moreover, the indole ring forms two hydrogen bonds with the backbone carbonyl oxygen of Gly462 and side-chain oxygen of Ser436.

Docking of **2f** (**81**) into Furin also revealed importance of the carboxylate on inhibition. As shown in Fig. 5B, **2f** was docked against the Furin protease (PDB: 5jxg). According to the docking result, the terminal carboxyl group of **2f** has strong interactions with Furin, including an electrostatic interaction with the guanidine group of Arg197 and a hydrogen bonding with Arg193 and His364. Besides, the nitrogen of the indole ring also forms a hydrogen bond with Pro256, and

doi:10.6342/NTU202502385

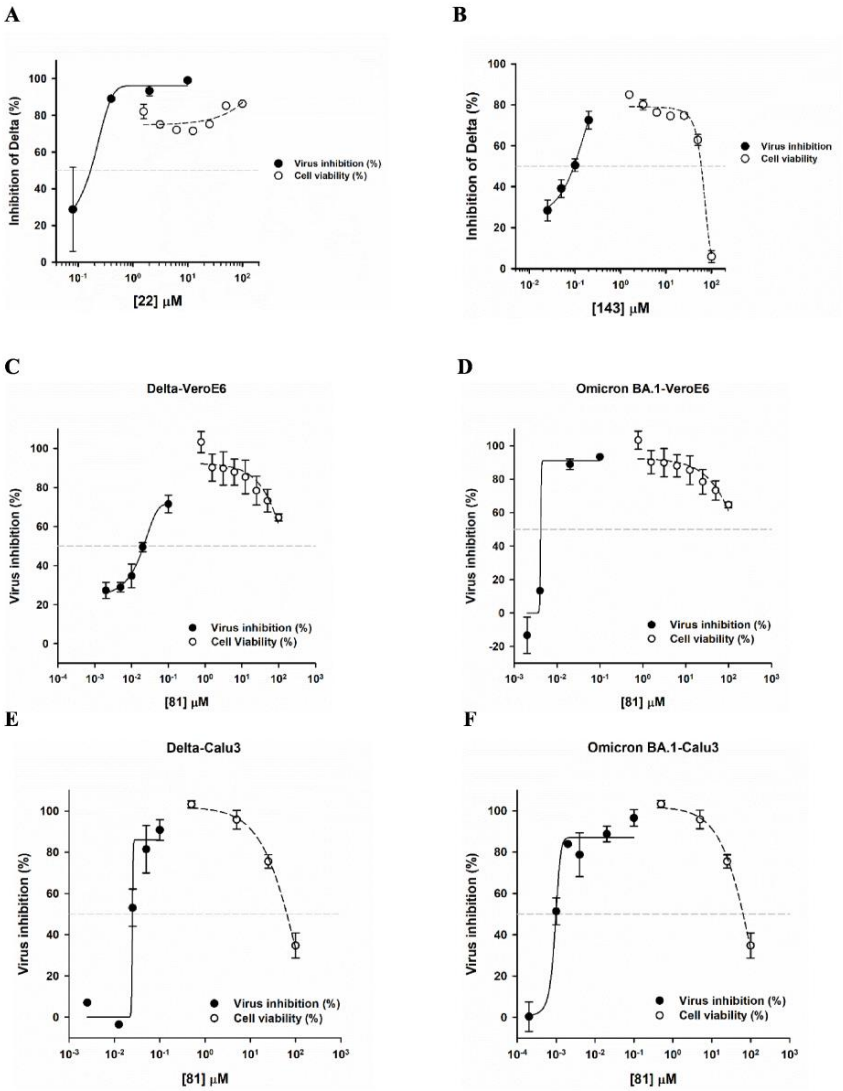
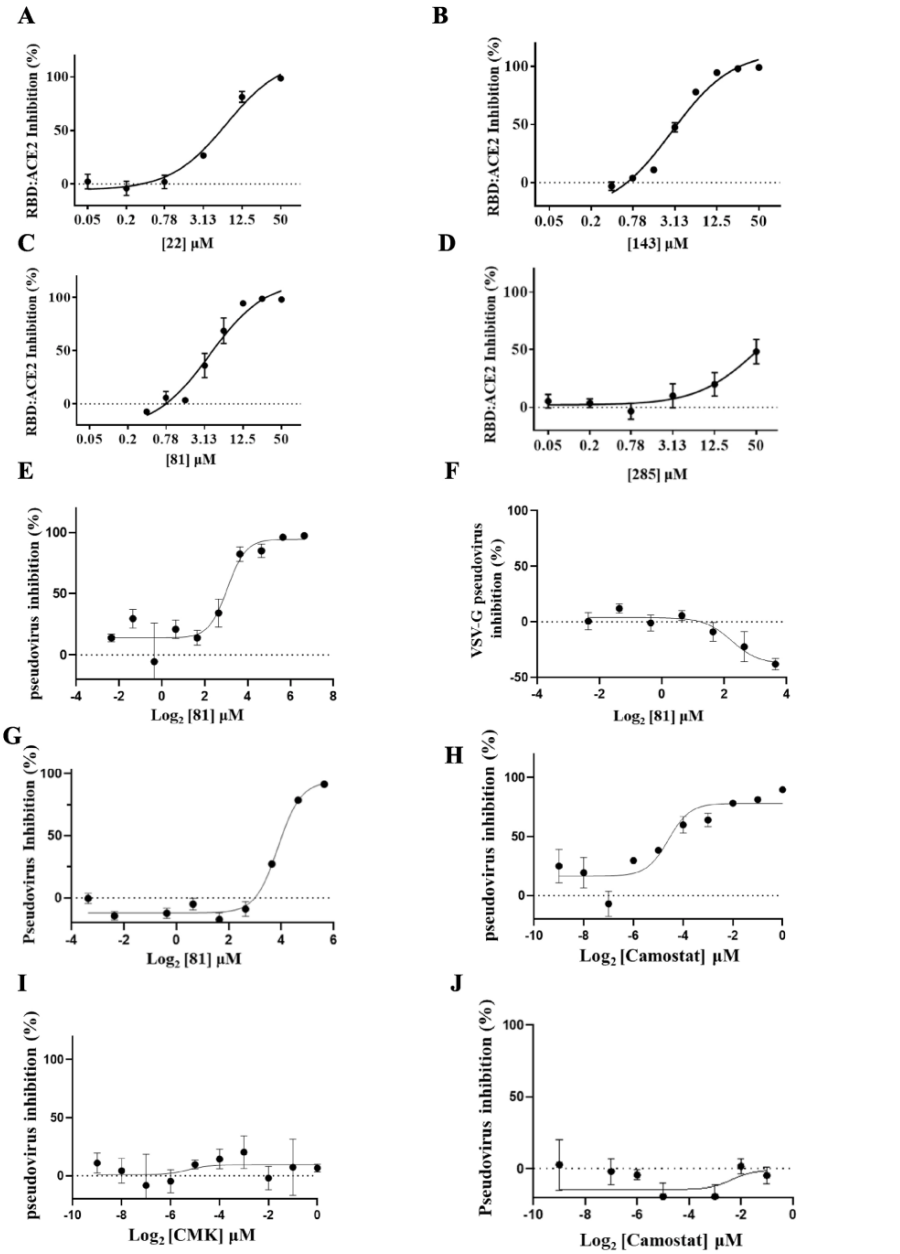


Fig. 2. Dose-dependent inhibition curves of the representative inhibitors **2b** (**22**), **2e** (**143**), and **2f** (**81**) against SARS-CoV-2. In (A–C), the plaque reduction assay was performed to determine the EC_{50} of **2b**, **2e**, and **2f** to be 0.15 ± 0.10, 0.087 ± 0.010, and 0.023 ± 0.005 μM, respectively, against the delta strain of SARS-CoV-2 (NTU92) infecting VeroE6 cells at 50−100 pfu for 1 h at 37 °C. The X-axis concentrations adopt the log base 10 scale. (D) The EC_{50} of **81** was determined to be 0.002 ± 0.002 μM against the omicron BA.1 strain (NTU128) infecting VeroE6 cells at 50−100 pfu for 1 h at 37 °C based on **81**-caused virus plaque reduction. (E, F) The dose-dependent inhibition on the virus yield by pretreating **81** with delta or omicron strain was used to determine the EC_{50} of **81** to be 0.026 ± 0.012 and 0.001 ± 0.0007 μM, respectively, against the delta and omicron BA.1 SARS-CoV-2 infecting Calu-3 cells at 50−100 pfu for 1 h at 37 °C based on **81**-caused virus plaque reduction. The drop of Calu-3 cell viability was due to higher concentrations of DMSO at higher concentrations of **81**. All the measurements were performed in triplicate to yield the averaged EC_{50} values and the standard deviations.

doi:10.6342/NTU202502385



(caption on next page)

doi:10.6342/NTU202502385

Fig. 3. The dose-dependent inhibition curves of the compounds on delta strain Spike's RBD:ACE2 interaction and pseudovirus. (A–C) The increasing inhibition percentages of RBD:ACE2 interaction in the presence of increasing concentrations of **2b** (**22**), **2e** (**143**), and **2f** (**81**) were fitted with the equation to yield the IC_{50} of 7.15 ± 1.52, 2.81 ± 0.49, and 3.76 ± 0.84 μM, respectively. The X-axis concentrations adopt the log base 2 scale. (D) The inhibition percentages on RBD:ACE2 binding were increased with increased concentrations of **4b** (**285**). The plot gave EC_{50} of 51.72 ± 51.84 μM. In the above measurements, HeLa cells expressing SmBiT-tagged human ACE2 and the recombinant RBD-LgBiT protein were mixed to form decreasing levels of an active NanoLuc complex, a luciferase, to be assayed using the provided substrate in the presence of increasing concentrations of synthesized inhibitors to yield IC_{50}. (E) The does-dependent curve for inhibiting delta Spike pseudovirus entry into Calu-3 cells by **2f** (**81**) to yield EC_{50} of 8.3 ± 1.2 μM. (F) Infection of Calu-3 cells by VSV-G Pseudovirus was not inhibited by **81**. (G) Infection of Hela-ACE2 cells by pseudovirus was inhibited by compounds **81** with EC_{50} of 14.7 ± 1.1 μM. (H) Specific TMPRSS2 covalent inhibitor Camostat inhibited pseudovirus infection on Calu-3 cells with EC_{50} of 0.04 ± 0.001 μM. (I) Specific Furin covalent inhibitor decanoyl-RVKR-chloromethylketone (CMK) showed no effect on pseudovirus infection on Calu-3 cells. (J) Camostat showed no effect on pseudovirus infection on HeLa-ACE2 cells. All the measurements were performed in triplicate to yield the averaged IC_{50} or EC_{50} values and the standard deviations.

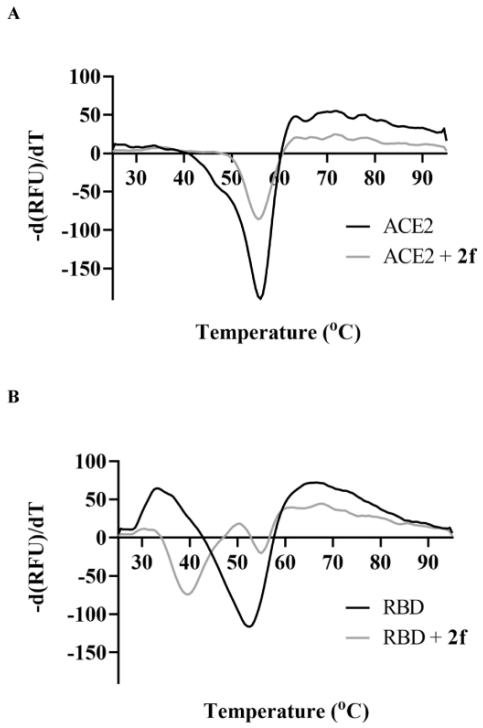


Fig. 4. Thermal shift experiments to measure the binding partner of **2f** (**81**). (A) Melting temperature change was not observed for adding **2f** to ACE2 (A), but obvious to RBD (B), based on the differential scanning fluorimetry assay, indicating RBD as the binding partner of **2f**. In this assay, a fluorescent dye was bound to the exposed surface on the heat-induced unfolded protein RBD or ACE2. The temperatures causing the maximal unfolding of the proteins (the peaks) were defined as melting temperatures. The curves were obtained from a single run of experiments with repeatable results.

the oxygen of the ketone group forms a hydrogen bond with the nitrogen of the main chain of Ser368.

Since our thermal shift experiments showed RBD as the target, computer docking was performed to understand the binding mode of compound **2f** in RBD. As shown in Figs. 5C, **2f** (**81**) was docked against the RBD of SARS-CoV-2 delta variant Spike protein (PDB: 7w92). Compound **2f** has both hydrogen bonding and van der Waal's interaction with Gln493 of RBD, which plays an important role on contacting with Lys31 of ACE2. Besides, the oxygen on the pyrrolidinone group also interacts with Gly496 of RBD through hydrogen bonding. In the delta RBD:ACE2 complex, Lys353 of ACE2 forms a hydrogen bond with Gly496 of RBD. Therefore, by forming interactions with the key residues of RBD, compound **2f** might disrupt the contact between ACE2 and RBD.

3.9. Drug-likeness of the inhibitor as judged from lipinski rule of five

Lipinski's rule was developed to determine the druggability of compounds for pharmaceutical development. For preparation of compound **2f**'s 3D structure, Molview website (https://molview.org/) was utilized, and the structure information was converted into a SMILES format by Open Babel (O'Boyle et al., 2011). To assess the drug likeness of compound **2f** (**81**), the online tool (http://www.scfbio-iitd.res.in/software/drugdesign/lipinski.jsp) was used for Lipinski rule of five estimation (Lipinski, 2004; Jayaram et al., 2012). Based on Lipinski rule of five, compound **2f** was accepted in parameters including molecular weight (less than 500 Da), logP (less than 5), hydrogen bond acceptors (less than 10) and donors (less than 5), and molar refractivity (between 40 and 130), indicating that the compound is likely to be a candidate for drug development.

3.10. Drug-likeness of the inhibitor as judged from ADMET properties

ADMET properties predicted in silico could be used to estimate the possibility of compounds to be used as therapeutic agents. There are 5 parameters involved in the ADMET prediction, including absorption, distribution, metabolism, excretion and toxicity. The ADMET profile of compound **2f** (**81**) was estimated by pkcsm website (http://biosig.unimelb.edu.au/pkcsm/) (Pires et al., 2015). Its absorption was portrayed by human colorectal adenocarcinoma cell line (Caco2) permeability and human intestinal absorption (HIA). The Caco2 permeability and the HIA scores are relatively high, indicating compound **2f** could be absorbed into the human intestine. Analysis of distribution progress, the blood brain barrier (BBB) and central nervous system (CNS) permeability of compound **2f** indicates that it is unable to penetrate BBB and CNS, and the volume of distribution at steady-state (VDss) implicates that compound **2f** has a higher possibility to distribute in plasma rather than in tissue. In human body, Cytochrome P450 (CYP) enzymes are responsible for the metabolism of drugs, and inhibitors of these CYP enzymes can significantly influence the effect of pharmaceuticals. The predicted results show that compound **2f** might be a substrate of CYP3A4 (an inhibitor of CYP3A4 might be co-administrated with **2f** to increase its half-life), but not other CYP enzymes. For excretion prediction, compound **2f** is not likely to be a renal organic cation transporter-2 (OCT2) substrate, showing there is no potential contraindication. In the toxicity estimation, compound **2f** is not a mutagenic drug and a hERG I/II inhibitor, and it does not have acute toxicity and skin sensitization. However, compound **2f** is predicted to possess hepatotoxicity.

*3.11. Preliminary animal study of **2f***

As predicted above, compound **2f** (**81**) could be a drug-like molecule

doi:10.6342/NTU202502385

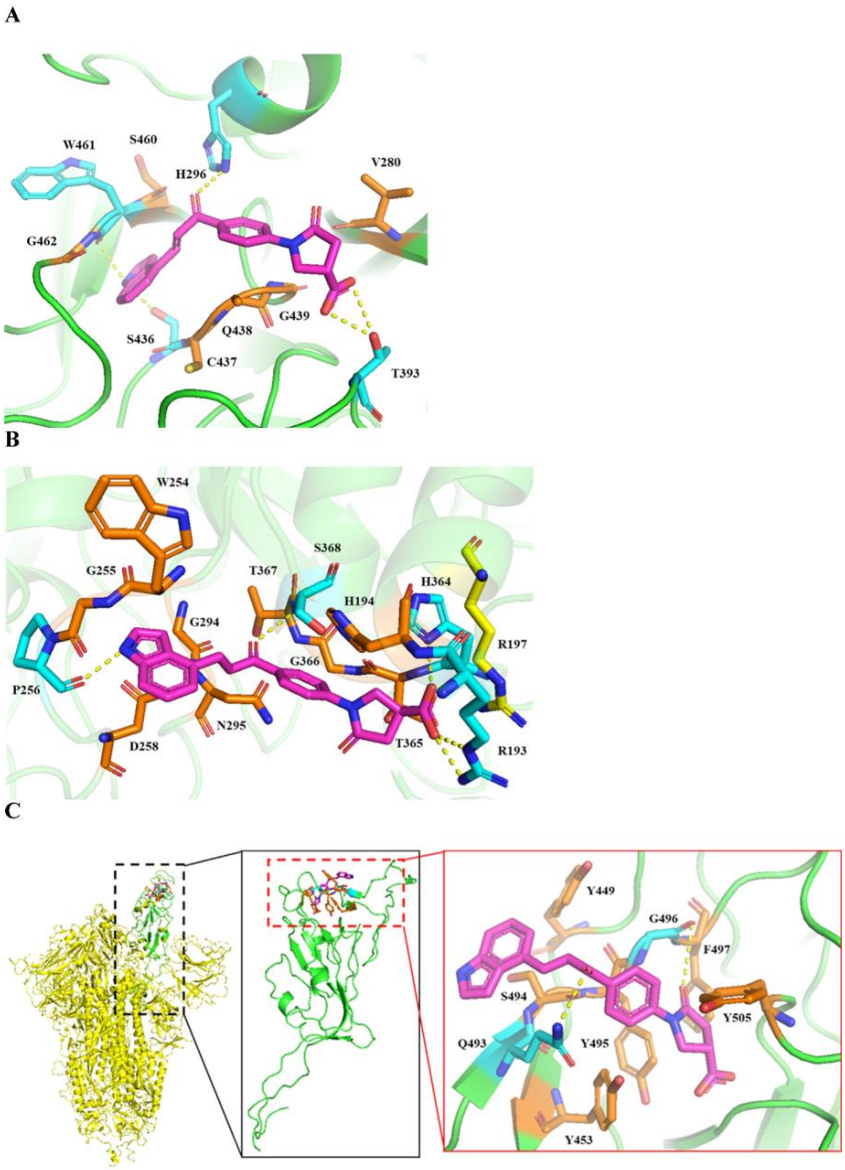


Fig. 5. Compound **2f** (**81**) was docked into (A) the TMPRSS2 ectodomain (PDB: 7meq), (B) Furin (PDB: 5jxg), and (C) RBD of the delta-strain Spike protein (PDB: 7w92). The estimated binding energies of **2f** with TMPRSS2, Furin, and RBD are −115.8, −118.9, and −115.1 kcal/mol, respectively. Compound **2f** is colored in magenta, and the residues colored in cyan, orange, and yellow are involved in hydrogen-bonding, van der Waals, and electrostatic interactions, respectively. In (C), colored in yellow is trimeric Spike protein and in green is RBD. **2f** binds with the residues of RBD in the interface to interfere with the RBD binding with ACE2.

doi:10.6342/NTU202502385

obeying Lipinski rule of five and with suitable ADMET properties. To verify *in vivo* effectiveness and safety of **2f**, we used hamster, which has been shown to be a satisfactory animal model for SARS-CoV-2 infection (Imai et al., 2020; Sia et al., 2020). Oral administration with **2f** or vehicle began on the same day as infection with 5.0×10^4 pfu SARS-CoV-2 by intranasal inoculation. Body weights were monitored daily for each animal. All hamsters lost weight after inoculation with SARS-CoV-2. However, compared to the vehicle group, hamsters treated with 25 mg/kg of **2f** showed a slightly less weight loss with statistically significance ($P = 0.014$; Fig. 6A). Moreover, in **2f** treatment group, viral genomic copy assessment assays of homogenized hamster nasal turbinates revealed a 5.0- and 5.5-fold reduction in RNA copies at 3 and 6 days post-infection, respectively, in comparison with the vehicle group (Fig. 6B, left panel). However, viral RNA copies detected in lung tissue showed less reduction by the treatment at 3 and 6 days post-infection (Fig. 6B, right panel).

The pharmacokinetics of **81** was also accessed by analyzing the serum samples withdrawn from ACE2 transgenic mice via oral administration or intravenous injection of 25 mg/kg of compound. The serum samples before treatment and after 15, 30, 60, 120, and 240 min of treatment were extracted with methanol and injected into LC-MS system to analyze the concentrations of **81** in the samples based on a linear standard curve generated by using purified **81** at the comparable concentrations. Following the intravenous injection, the serum concentration-time profile of compound **81** was depleted in a biphasic manner (Fig. 6C, left panel for linear and right panel for semi-logarithmic plots, respectively). The PK parameters are shown in Table 3. Through IV injection, compound **81** level was rapidly drop from 2.97 μM at 15 min to 0.78 μM at 60 min, then it was eliminated from the body with the terminal half-life of 1.59 h. Regarding the oral administration, compound **81** reached the peak of 1.13 μM at 15 min and declined to 0.067 μM at 4 h with the terminal half-life of 1.35 h. The oral bioavailability was determined to be 35.19%. Within the experimental period of 4 h, the serum concentrations of **81** were above the anti-SARS-CoV-2 EC_{50} of 0.001–0.026 μM. This could explain the effectiveness in reducing the viral load in tested animal nasal turbinates. The methanol extract of mice lung tissues taken at the end of experiments contained several overlapped peaks in HPLC profile around the MW of compound **81**, making it difficult to determine the real concentration of the compound (data not shown).

4. Discussion

While two RdRp inhibitors and a $3CL^{pro}$ inhibitor which inhibit virus replication have been approved by FDA for treatment of COVID-19 patients, no compound targeting other essential components for SARS-CoV-2 infection has been approved. Although many $3CL^{pro}$ and PL^{pro} covalent and non-covalent inhibitors have already been identified, the small-molecule reversible inhibitors targeting virus entry, such as TMPRSS2, Furin, and RBD:ACE2 interaction, are relatively limited. Previously, Camostat and Nafamostat, have been identified to irreversibly inhibit TMPRSS2 and thus SARS-CoV-2 infection by blocking viral entry into cells (Hoffmann et al., 2020a, 2020b). Some covalent peptidomimetic ketobenzothiazole inhibitors, which were developed based on the substrate specificity of TMPRSS2, showed anti-SARS-CoV-2 activity (Mahoney et al., 2021; Shapira et al., 2022). Animal study demonstrated the effectiveness of the TMPRSS2 irreversible inhibitor *in vivo* via intranasal treatment (7.2 mg/kg) 1 day before virus inoculation It remained half-effective when treated (14.4 mg/kg) 12 h after virus inoculation. These data demonstrated that the covalent TMPRSS2 inhibitors can be used for effective anti-SARS-CoV-2 treatment *in vitro* and *in vivo* by blocking the virus entry. However, the covalent inhibitors with the keto warhead but not the reduced hydroxyl analogue showed inhibition on TMPRSS2 and SARS-CoV-2, indicating the difficulty to develop reversible TMPRSS2 inhibitors for anti-SARS-CoV-2. Furin peptidomimetic covalent inhibitor decanoyl-RVKR-chloromethylketone has also been shown to inhibit SARS-CoV-2 entry into VeroE6 cells and thus virus production (Cheng et al., 2020). Several dyes (Congo red, direct violet 1, Evans blue) and drug-like compounds (DRI-C23041, DRI-C91005) inhibited the interaction of ACE2 with the spike proteins of SARS-CoV-2 with IC_{50} of 0.2–3.0 μM in the cell-free ELISA-type assays (Bojadzic et al., 2021). Here, we demonstrate the reversible inhibition of RBD:ACE2, TMPRSS2, and/or Furin as a novel and workable strategy for preventing SARS-CoV-2 entry and production.

We originally designed TMPRSS2/Furin dual reversible inhibitors, and as expected the compounds were active antivirals only if pretreating SARS-CoV-2 to block the virus entry, but not added after infection. Because some analogues that did not inhibit TMPRSS2 and Furin still inhibited the virus entry and production (Table 1), we then identified them as also RBD:ACE2 inhibitors. Our results thus indicate that the compounds exert their antiviral activities by inhibiting RBD:ACE2 interaction as a major factor, and inhibiting TMPRSS2 and/or Furin enhances their antiviral activities. The best inhibitor **2f** (**81**) displays great inhibition against TMPRSS2, Furin, and RBD:ACE2 interaction, respectively, with IC_{50} values of 1.27, 3.31, and 3.76 μM. It achieves EC_{50} values of 0.023 and 0.002 μM against SARS-CoV-2 delta and omicron variants, respectively, while infecting VeroE6 cells and $CC_{50} >$ 100 μM, giving the largest therapeutic index of >50000. By using lung Calu-3 cells, the EC_{50} were measured to be 0.026 and 0.001 μM, respectively, against SARS-CoV-2 delta and omicron variants. The preliminary animal study also demonstrates the effectiveness in reducing the viral load by 5–5.5-fold in nasal turbinate and safety without causing the animal weight loss while orally given a dose of 25 mg/kg into hamsters. Pharmacokinetics study shows higher concentrations of **81** in serum samples by IV injection than oral treatment, and both gave the concentrations above the antiviral EC_{50}. Further animal studies through different doses and ways of administration, and even in the presence of cytochrome P-450 inhibitor Ritonavir can be conducted to optimize the treatment effect *in vivo*.

For the antiviral mechanism, we show that compound **81** could inhibit the entry of SARS-CoV-2 pseudovirus into Calu-3 cells with an EC_{50} of 8.3 μM (Fig. 3E), but did not inhibit VSV-G pseudovirus into Calu-3 cells (Fig. 3F) as a control experiment because VSV-G pseudovirus entry is not dependent on RBD:ACE2, TMPRSS2, and Furin. We then show that the SARS-CoV-2 pseudovirus could only infect the HeLa-ACE2 stable cell line with enforcedly expressed ACE2 (Fig. 3G), but not the parental HeLa cells. These indicate that compound **81** only blocks ACE2-dependent entry of SARS-CoV-2 pseudoviruses by blocking RBD:ACE2 interaction. Moreover, we confirm that viral entry of Calu-3 could be inhibited by the specific TMPRSS2 inhibitor Camostat with an EC_{50} of 0.04 μM (Fig. 3H), but not by the Furin inhibitor decanoyl-RVKR-chloromethylketone (Fig. 3I). This is probably because Calu-3 cells express TMPRSS2 to replace the function of Furin in cleaving Spike for virus entry. As expected, Camostate did not inhibit pseudovirus entry into Hela-ACE2 cells that lack TMPRSS2 (Fig. 3J). However, as ACE2 and TMPRSS2 are both required for the viral entry into Calu-3 cells, to what extent blockade of ACE2 and/or TMPRSS2 by **81** accounted for inhibition of pseudoviral entry into Calu-3 cells is unclear. On the other hand, the anti-SARS-CoV-2 entry effect of decanoyl-RVKR-chloromethylketone was observed in TMPRSS2-limited VeroE6 cells where Furin is likely essential for virus entry (Cheng et al., 2020). Taken together, these data indicate that the entry of SARS-CoV-2 pseudovirus on Calu-3 cells is dependent on RBD:ACE2 interaction and TMPRSS2, but probably not Furin protease, although our compound **81** can inhibit all of these targets. Therefore, besides inhibiting RBD:ACE2 as the main factor for antiviral activity, the TMPRSS2 inhibitory activity of **81** could enhance its antiviral activity in Calu-3 cells, but inhibiting RBD:ACE2 and Furin could account for the antiviral activity of **81** in VeroE6 cells.

Our compounds contain a 4-styrylcarbonylphenyl, chalcone basic skeleton (chemical structure shown in Fig. 7A). Different chalcone derivatives have been shown to have anti-bacterial and anti-cancer activities by *in vitro* and *in vivo* experiments (Xu et al., 2019; Ouyang et al.,

doi:10.6342/NTU202502385

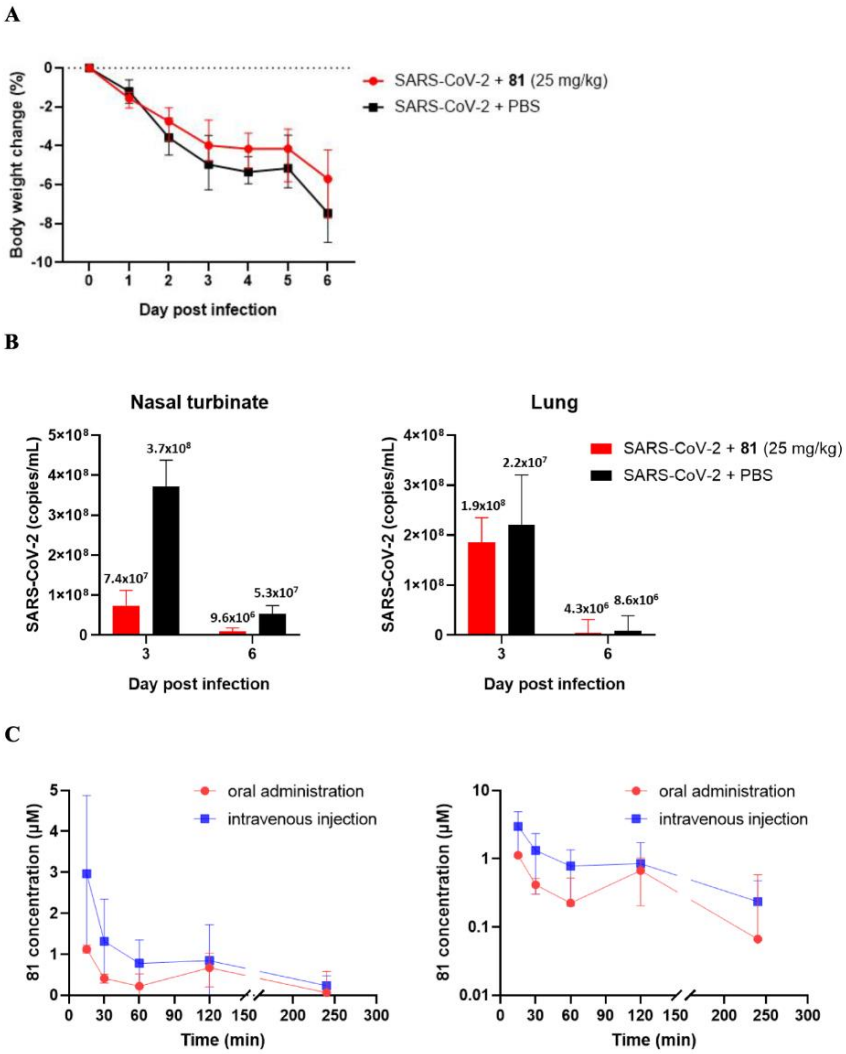


Fig. 6. Body weight changes and viral loads of Syrian hamsters infected intranasally with SARS-CoV-2 with or without **2f** (**81**) oral treatment. (A) Time courses of body weight percent changes from the averages of three hamsters with treatment of 25 mg/kg shown in red and the averages of three without treatment (only PBS) shown in black. (B) Corresponding virus genome copy numbers as determined by qRT-PCR in nasal turbinate (left) and lung (right) from the averages of three hamsters with treatment of 25 mg/kg shown in red and the averages of three without treatment (only PBS) shown in black. (C) Linear (left) and semi-logarithmic (right) plots of the serum concentration-time profile of compound **81** following oral administration or intravenous injection (25 mg/kg). The serum samples obtained from the mice before treatment and 15, 30, 60, 120, and 240 min after oral administration or IV injection of 25 mg/kg compound **81** were subjected to LC-MS to determine the concentrations shown in red and blue dots for oral and injection treatment, respectively. Data represented are mean ± SD. Apparently, IV injection yielded higher serum concentrations of **81** than oral treatment, but both gave the concentrations above the antiviral EC_{50}.

doi:10.6342/NTU202502385

Table 3
Pharmacokinetic parameters of compound **81** in K18-hACE2 transgenic mice following oral administration or intravenous (IV) injection (25 mg/kg).

Parameters	Unit	oral administration	IV injection
λ	1/h	0.51	0.43
$t_{1/2\,\lambda}$	h	1.35	1.59
AUC	μM·h	1.58	4.48
MRT	h	2.00	1.79

Note: The pharmacokinetic parameters of compound **81** were determined by a non-compartmental model. λ, terminal rate constant; $t_{1/2\lambda}$, terminal half-life; AUC, area under the serum concentration-time curve; MRT, mean residence time.

2021). Recently, chalcone derivatives (a chemical structure shown in Fig. 7B) were shown to inhibit SARS-CoV-2 replication in Vero E6 cells, and the targets were predicted to be RdRp, $3CL^{pro}$, and Spike's RBD by computer docking (Duran et al., 2021). A 7-(4-(N-substituted-carbamoyl-methyl)piperazin-1yl) ciprofloxacin-chalcone (chemical structure shown in Fig. 7C) was shown to inhibit $3CL^{pro}$ and SARS-CoV-2 (Alaaeldin et al., 2022). Besides chalcone, our best inhibitor **2f** also contains an indole moiety. There are few reports of similar indole chalcones (chemical structures shown in Fig. 7D–F) showing anti-inflammatory (Özdemir et al., 2015), anti-tumor (Yan et al., 2016), and anti-*Mycobacterium tuberculosis* (Mtb) (Ramesh et al., 2020) activities, respectively. Some indole chalcones were predicted by docking to

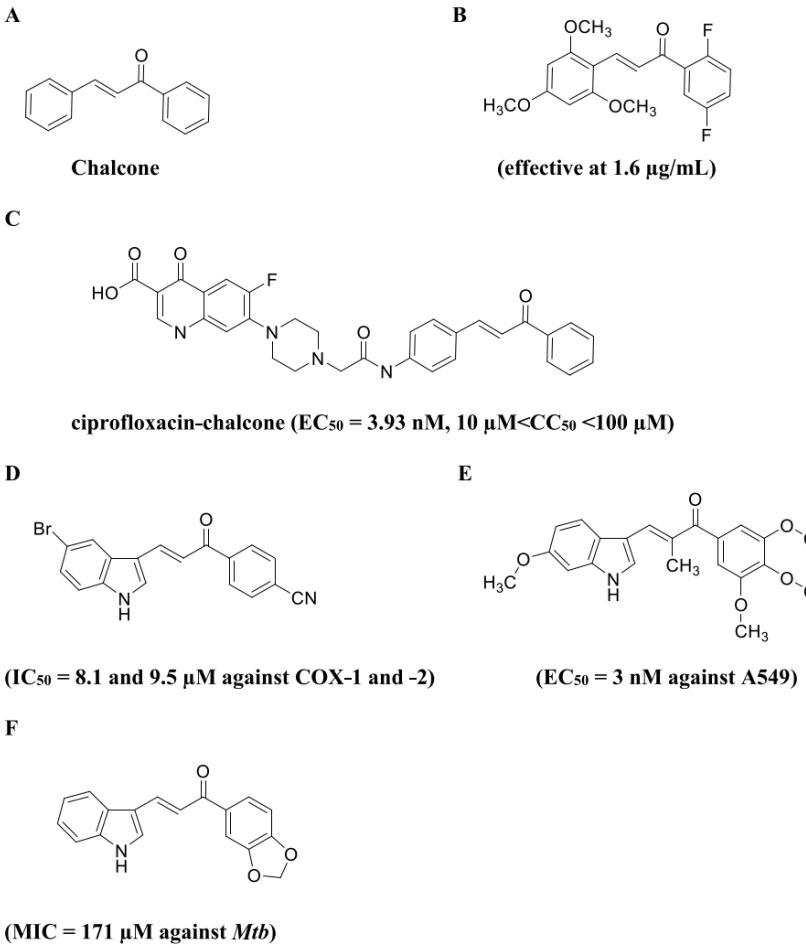


Fig. 7. Chemical structures of some biologically active chalcones. These include the chemical structures of (A) chalcone, (B) a chalcone derivative shown to inhibit SARS-CoV-2 viral growth in Vero E6 cell (effective at 1.6 μg/mL) by targeting RdRp, $3CL^{pro}$, and Spike's RBD as predicted by computer docking, (C) 7-(4-(N-substituted-carbamoyl-methyl)piperazin-1yl) ciprofloxacin-chalcone shown to inhibit $3CL^{pro}$ and SARS-CoV-2, and (D–F) indole chalcones possessing anti-inflammatory, anti-tumor, and anti-*Mycobacterium tuberculosis* activities, respectively.

doi:10.6342/NTU202502385

target SARS-CoV-2 RdRp, $3CL^{pro}$, and/or Spike's RBD (Vijayakumar et al., 2020), but further experiments are required to confirm the prediction. Indole seems to increase the compound affinity with the SARS-CoV-2 $3CL^{pro}$ and a compound with indole ring was shown to reversibly bind with the protease (Hattori et al., 2021). Ghosh et al. also reported indole chloropyridinyl ester-derived SARS-CoV-2 $3CL^{pro}$ inhibitors (Ghosh et al., 2021). On the other hand, indole derivatives were shown by silicon screening to bind Spike protein of SARS-CoV-2 (Gobinath et al., 2021).

In summary, compound **2f** (**81**) containing chalcone and indole moieties obeys the Lipinski's and Veber's rules and hence displays drug-like molecular nature. Not only the lung Calu-3 cell-based assay, our preliminary *in vivo* animal study also shows the promising effectiveness and safety of the lead compound. While the therapeutic antibody drugs used to prevent virus entry into cells are expensive and are prone to reduce treatment sensitivity once encountering the mutations on Spike, our small-molecule potent antiviral effective among different variants of SARS-CoV-2 could be a useful alternative for treatment. In conclusion, we have successfully identified promising inhibitors of SARS-CoV-2 by reversibly blocking the essential targets at the virus entry.

Author contribution

SPR synthesized all the compounds, CWL prepared recombinant TMPRSS2 ectodomain and assayed IC_{50} of TMPRSS2 and Furin, HLVL prepared RBD and ACE2 and determined the binding partner of **2f** by performing thermal shift experiments, SWL performed Biacore experiments. JJL performed structural analysis and computer modeling, TLC and JTF did virus plaque reduction assay and yield reduction assay to measure EC_{50} and CC_{50}, MYCP and YTC assayed IC_{50} of inhibiting RBD: ACE2 interaction and pseudovirus entry, LHCW supervised the research of RBD:ACE2 interaction and pseudovirus assay, WCL, CWH, CMY, and YYC performed animal studies, YTL and HCL performed and supervised LC-MS analysis, respectively, and CJK provided clones and did initial TMPRSS2 assay as well as coordinated animal studies, SYC supervised antiviral assay, PHL designed the experiments, supervise all aspects of research, and wrote the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Authors thank the financial support (the grants 109-2745-B-001-001, 109-2327-B-007-002, 110-2320-B-007-004-MY3, 110-2740-B-002-006, 111-2740-B-002-006, and 112-2740-B-002-004) and technical supports of National Core Facility for Biopharmaceuticals (including C6/RNAi core, P3-3/BSL-3 Lab, and N2/NGNAD platform) from Ministry of Science and Technology, Taiwan and an infectious disease research grant AS-IDR-112-12 and AS-KPQ-110-EIMD from Academia Sinica, Taiwan, and a grant support from National Tsing Hua University (NTHU-112QI033E1). We would also like to acknowledge the service provided by the Biosafety Level-3 Laboratory of the First Core Laboratory, National Taiwan University College of Medicine, and the Biosafety Level-3 Laboratory from National Taiwan University Hospital. Authors would like to thank Dr. Tirawat Rairat for the discussion of pharmacokinetic analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.antiviral.2023.105735.

References

Alaaeldin, R., Mustafa, M., Abuo-Rahma, G.E.-D.A., Fathy, M., 2022. In vitro inhibition and molecular docking of a new ciprofloxacin-chalcone against SARS-CoV-2 main protease. Fundam. Clin. Pharmacol. 36, 160–170. https://doi.org/10.1111/fcp.12708.

Bestle, D., Heindl, M.R., Limburg, H., Van Lam van, T., Pilgram, O., Moulton, H., Stein, D. A., Hardes, K., Eickmann, M., Dolnik, O., Rohde, C., Klenk, H.D., Garten, W., Steinmetzer, T., Böttcher-Friebertshäuser, E., 2020. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. Life Sci. Alliance 3, e202000786. https://doi.org/10.26508/lsa.202000786.

Bojadzic, D., Alcazar, O., Chen, J., Chuang, S.T., Condor Capcha, J.M., Shehadeh, L.A., Buchwald, P., 2021. Small-molecule inhibitors of the coronavirus Spike: ACE2 protein-protein interaction as blockers of viral attachment and entry for SARS-CoV-2. ACS Infect. Dis. 7, 1519–1534. https://doi.org/10.1021/acsinfecdis.1c00070.

Cao, Y., Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., Du, S., Wang, J., Li, Q., Chen, X., Yu, Y., Wang, P., Zhang, Z., Liu, P., An, R., Hao, X., Wang, Y., Wang, J., Feng, R., Sun, H., Zhao, L., Zhang, W., Zhao, D., Zheng, J., Yu, L., Li, C., Zhang, N., Wang, R., Niu, X., Yang, S., Song, X., Chai, Y., Hu, Y., Shi, Y., Zheng, L., Li, Z., Gu, Q., Shao, F., Huang, W., Jin, R., Shen, Z., Wang, Y., Wang, X., Xiao, J., Xie, X.S., 2022. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. Nature 608, 593–602. https://doi.org/10.1038/s41586-022-04980-y.

Cheng, Y.W., Chao, T.L., Li, C.L., Chiu, M.F., Kao, H.C., Wang, S.H., Pang, Y.H., Lin, C.H., Tsai, Y.M., Lee, W.H., Tao, M.H., Ho, T.C., Wu, P.Y., Jang, L.T., Chen, P.J., Chang, S. Y., Yeh, S.H., 2020. Furin inhibitors block SARS-CoV-2 Spike protein cleavage to suppress virus production and cytopathic effects. Cell Rep. 33, 108254 https://doi.org/10.1016/j.celrep.2020.108254.

Dai, W., Zhang, B., Jiang, X.M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., Li, C., Li, Y., Bai, F., Wang, H., Cheng, X., Cen, X., Hu, S., Yang, X., Wang, J., Liu, X., Xiao, G., Jiang, H., Rao, Z., Zhang, L.K., Xu, Y., Yang, H., Liu, H., 2020. Structure-based design of antiviral drug candidates targeting SARS-CoV-2 main protease. Science 368, 1331–1335. https://doi.org/10.1126/science.abb4489.

de Wit, E., van Doremalen, N., Falzarano, D., Munster, V.J., 2016. SARS and MERS: recent insights into emerging coronaviruses. Nat. Rev. Microbiol. 14, 523–534. https://doi.org/10.1038/nrmicro.2016.81.

Duran, N., Polat, M.F., Aktas, D.A., Alagoz, M.A., Ay, E., Cimen, F., Tek, E., Anil, B., Burmaoglu, S., Algul, O., 2021. New chalcone derivatives as effective against SARS-CoV-2 Agent. Int. J. Clin. Pract. 75, e14846 https://doi.org/10.1111/ijcp.14846.

Essalmani, R., Jain, J., Susan-Resiga, D., Andréo, U., Evagelidis, A., Derbali, R.M., Huynh, D.N., Dallaire, F., Laporte, M., Delpal, A., Sutto-Ortiz, P., Coutard, B., Mapa, C., Wilcoxen, K., Decroly, E., Nq Pham, T., Cohen, É.A., Seidah, N.G., 2022. Distinctive roles of Furin and TMPRSS2 in SARS-CoV-2 infectivity. J. Virol. 96, e0012822 https://doi.org/10.1128/jvi.00128-22.

Fraser, B.J., Beldar, S., Seitova, A., Hutchinson, A., Mannar, D., Li, Y., Kwon, D., Tan, R., Wilson, R.P., Leopold, K., Subramaniam, S., Halabelian, L., Arrowsmith, C.H., Bénard, F., 2022. Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. Nat. Chem. Biol. 18, 963–971. https://doi.org/10.1038/s41589-022-01059-7.

Ghosh, A.K., Raghavaiah, J., Shahabi, D., Yadav, M., Anson, B.J., Lendy, E.K., Hattori, S. I., Higashi-Kuwata, N., Mitsuya, H., Mesecar, A.D., 2021. Indole chloropyridinyl ester-derived SARS-CoV-2 3CLpro Inhibitors: enzyme inhibition, antiviral Efficacy, structure-activity relationship, and X-ray structural studies. J. Med. Chem. 64, 14702–14714. https://doi.org/10.1021/acs.jmedchem.1c01214.

Gobinath, P., Packialakshmi, P., Ali, D., Alarifi, S., Alotaibi, A., Idhayadhulla, A., Surendrakumar, R., 2021. Synthesis and cytotoxic activity of novel indole derivatives and their in Silico screening on Spike glycoprotein of SARS-CoV-2. Front. Mol. Biosci. 8, 637989 https://doi.org/10.3389/fmolb.2021.637989.

Gordon, C.J., Tchesnokov, E.P., Woolner, E., Perry, J.K., Feng, J.Y., Porter, D.P., Götte, M., 2020. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from Severe Acute Respiratory Syndrome Coronavirus 2 with high potency. J. Biol. Chem. 295, 6785–6797. https://doi.org/10.1074/jbc.RA120.013679.

Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., Feldt, T., Green, G., Green, M.L., Lescure, F.X., Nicastri, E., Oda, R., Yo, K., Quiros-Roldan, E., Studemeister, A., Redinski, J., Ahmed, S., Bernett, J., Chelliah, D., Chen, D., Chihara, S., Cohen, S.H., Cunningham, J., D'Arminio Monforte, A., Ismail, S., Kato, H., Lapadula, G., L'Her, E., Maeno, T., Majumder, S., Massari, M., Mora-Rillo, M., Mutoh, Y., Nguyen, D., Verweij, E., Zoufaly, A., Osinusi, A.O., DeZure, A., Zhao, Y., Zhong, L., Chokkalingam, A., Elboudwarej, E., Telep, L., Timbs, L., Henne, I., Sellers, S., Cao, H., Tan, S.K., Winterbourne, L., Desai, P., Mera, R., Gaggar, A., Myers, R.P., Brainard, D.M., Childs, R., Flanigan, T., 2020. Compassionate use of Remdesivir for patients with severe Covid-19. N. Engl. J. Med. 382, 2327–2336. https://doi.org/10.1056/NEJMoa2007016.

Hattori, S.I., Higashi-Kuwata, N., Hayashi, H., Allu, S.R., Raghavaiah, J., Bulut, H., Das, D., Anson, B.J., Lendy, E.K., Takamatsu, Y., Takamune, N., Kishimoto, N., Murayama, K., Hasegawa, K., Li, M., Davis, D.A., Kodama, E.N., Yarchoan, R., Wlodawer, A., Misumi, S., Mesecar, A.D., Ghosh, A.K., Mitsuya, H., 2021. A small molecule compound with an indole moiety inhibits the main protease of SARS-CoV-2

doi:10.6342/NTU202502385

and blocks virus replication. Nat. Commun. 12, 668. https://doi.org/10.1038/s41467-021-20900-6.
Hoffman, R.L., Kania, R.S., Brothers, M.A., Davies, J.F., Ferre, R.A., Gajiwala, K.S., He, M., Hogan, R.J., Kozminski, K., Li, L.Y., Lockner, J.W., Lou, J., Marra, M.T., Mitchell Jr., L.J., Murray, B.W., Nieman, J.A., Noell, S., Planken, S.P., Rowe, T., Ryan, K., Smith 3rd, G.J., Solowiej, J.E., Steppan, C.M., Taggart, B., 2020. Discovery of ketone-based covalent inhibitors of Coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. J. Med. Chem. 63, 12725–12747. https://doi.org/10.1021/acs.jmedchem.0c01063.
Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., Müller, M.A., Drosten, C., Pöhlmann, S., 2020a. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181, 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052.
Hoffmann, M., Schroeder, S., Kleine-Weber, H., Müller, M.A., Drosten, C., Pöhlmann, S., 2020b. Nafamostat mesylate blocks activation of SARS-CoV-2: new treatment option for COVID-19. Antimicrob. Agents Chemother. 64, e00754–20 https://doi.org/10.1128/AAC.00754-20.
Hsu, M.F., Kuo, C.J., Chang, K.T., Chang, H.C., Chou, C.C., Ko, T.P., Shr, H.L., Chang, G. G., Wang, A.H., Liang, P.H., 2005. Mechanism of the maturation process of SARS-CoV 3CL protease. J. Biol. Chem. 280, 31257–31266. https://doi.org/10.1074/jbc.M502577200.
Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.
Imai, M., Iwatsuki-Horimoto, K., Hatta, M., Loeber, S., Halfmann, P.J., Nakajima, N., Watanabe, T., Ujie, M., Takahashi, K., Ito, M., Yamada, S., Fan, S., Chiba, S., Kuroda, M., Guan, L., Takada, K., Armbrust, T., Balogh, A., Furusawa, Y., Okuda, M., Ueki, H., Yasuhara, A., Sakai-Tagawa, Y., Lopes, T.J.S., Kiso, M., Yamayoshi, S., Kinoshita, N., Ohmagari, N., Hattori, S.I., Takeda, M., Mitsuya, H., Krammer, F., Suzuki, T., Kawaoka, Y., 2020. Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development. Proc. Natl. Acad. Sci. U.S.A. 117, 16587–16595. https://doi.org/10.1073/pnas.2009799117.
Jackson, C.B., Farzan, M., Chen, B., Choe, H., 2022. Mechanisms of SARS-CoV-2 entry into cells. Nat. Rev. Mol. Cell Biol. 23, 3–20. https://doi.org/10.1038/s41580-021-00418-x.
Jayaram, B., Singh, T., Mukherjee, G., Mathur, A., Shekhar, S., Shekhar, V., 2012. A freely accessible web-server for target directed lead molecule discovery. BMC Bioinf. 13, 1–13. https://doi.org/10.1186/1471-2105-13-S17-S7.
Jiang, L., 2014. Copper/N,N-dimethylglycine catalyzed Goldberg reactions between aryl bromides and amides, aryl iodides and secondary acyclic amides. Molecules 19, 13448–13460. https://doi.org/10.3390/molecules190913448.
Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L.W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., Yang, H., 2020. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. Nature 582, 289–293. https://doi.org/10.1038/s41586-020-2223-y.
Kim, Y., Liu, H., Galasiti Kankanamalage, A.C., Weerasekara, S., Hua, D.H., Groutas, W. C., Chang, K.O., Pedersen, N.C., 2016. Reversal of the progression of fatal coronavirus infection in cats by a broad-spectrum coronavirus protease inhibitor. PLoS Pathog. 12, e1005531 https://doi.org/10.1371/journal.ppat.1005531.
Kuo, C.J., Chao, T.L., Kao, H.C., Tsai, Y.M., Liu, Y.K., Wang, L.H., Hsieh, M.C., Chang, S. Y., Liang, P.H., 2021. Kinetic characterization and inhibitor screening for the proteases leading to identification of drugs against SARS-CoV-2. Antimicrob. Agents Chemother. 65, e02577-20 https://doi.org/10.1128/AAC.02577-20.
Kuo, C.J., Liang, P.H., 2015. Characterization, inhibition, and engineering of the main protease of severe acute respiratory syndrome coronavirus. ChemBioEng Rev. 2, 118–132. https://doi.org/10.1002/cben.201400031.
Kuo, C.J., Liang, P.H., 2022. SARS-CoV-2 3CLpro displays faster self-maturation in vitro than SARS-CoV 3CLpro due to faster C-terminal cleavage. FEBS Lett. 596, 1214–1224. https://doi.org/10.1002/1873-3468.14337.
Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., Wang, X., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220. https://doi.org/10.1038/s41586-020-2180-5.
Lee, J.Y., Kuo, C.J., Shin, J.S., Jung, E., Liang, P.H., Jung, Y.S., 2021. Identification of non-covalent 3C-like protease inhibitors against severe acute respiratory syndrome coronavirus-2 via virtual screening of a Korean compound library. Bioorg. Med. Chem. Lett. 42, 128067 https://doi.org/10.1016/j.bmcl.2021.128067.
Lee, R.K., Li, T.N., Chang, S.Y., Chao, T.L., Kuo, C.H., Pan, M.Y., Chiou, Y.T., Liao, K.J., Yang, Y., Wu, Y.H., Huang, C.H., Juan, H.F., Hsieh, H.P., Wang, L.H., 2022. Identification of entry inhibitors against delta and omicron variants of SARS-CoV-2. Int. J. Mol. Sci. 23, 4050. https://doi.org/10.3390/ijms23074050.
Lipinski, C.A., 2004. Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov. Today Technol. 1, 337–341. https://doi.org/10.1016/j.ddtec.2004.11.007.
Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W.J., Wang, D., Xu, W., Holmes, E.C., Gao, G.F., Wu, G., Chen, W., Shi, W., Tan, W., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implication for virus origins and receptor binding. Lancet 395, 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8.
Mahoney, M., Damalanka, V.C., Tartellm, M.A., Chung, D.H., Lourenço, A.L., Pwee, D., Bridwell, A.E.M., Hoffmann, M., Voss, J., Karmakar, P., Azouz, N.P., Klingler, A.M., Rothlauf, P.W., Thompson, C.E., Lee, M., Klampfer, L., Stallings, C.L., Rothenberg, M.E., Pöhlmann, S., Whelan, S.P.J., O'Donoghue, A.J., Craik, C.S., Janetka, J.W., 2021. A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells. Proc. Natl. Acad. Sci. U.S.A. 118, e2108728118 https://doi.org/10.1073/pnas.2108728118.
Mondal, S., Chen, Y., Lockbaum, G.J., Sen, S., Chaudhuri, S., Reyes, A.C., Lee, J.M., Kaur, A.N., Sultana, N., Cameron, M.D., Shaffer, S.A., Schiffer, C.A., Fitzgerald, K.A., Thompson, P.R., 2022. Dual inhibitors of main protease (MPro) and cathepsin L as potent antivirals against SARS-CoV2. J. Am. Chem. Soc. 144, 21035–21045. https://doi.org/10.1021/jacs.2c04626.
O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open Babel: an open chemical toolbox 2011. J. Cheminf. 3, 33. https://doi.org/10.1186/1758-2946-3-33.
Ouyang, Y., Li, J., Chen, X., Fu, X., Sun, S., Wu, Q., 2021. Chalcone derivatives: role in anticancer therapy. Biomolecules 11, 894. https://doi.org/10.3390/biom11060894.
Özdemir, A., Altintop, M.D., Turan-Zitouni, G., Çiftçi, G.A., Ertorun, İ., Alataş, Ö., Kaplancikli, Z.A., 2015. Synthesis and evaluation of new indole-based chalcones as potential antiinflammatory agents. Eur. J. Med. Chem. 89, 304–309. https://doi.org/10.1016/j.ejmech.2014.10.056.
Pires, D.E.V., Blundell, T.L., Ascher, D.B., 2015. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J. Med. Chem. 58, 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104.
Ramesh, D., Joji, A., Vijayakumar, B.G., Sethumadhavan, A., Mani, M., Kannan, T., 2020. Indole chalcones: design, synthesis, in vitro and in silico evaluation against Mycobacterium tuberculosis. Eur. J. Med. Chem. 198, 112358 https://doi.org/10.1016/j.ejmech.2020.112358.
Sheahan, T.P., Sims, A.C., Zhou, S., Graham, R.L., Pruijssers, A.J., Agostini, M.L., Leist, S. R., Schäfer, A., Dinnon 3rd, K.H., Stevens, L.J., Chappell, J.D., Lu, X., Hughes, T.M., George, A.S., Hill, C.S., Montgomery, S.A., Brown, A.J., Bluemling, G.R., Natchus, M. G., Saindane, M., Kolykhalov, A.A., Painter, G., Harcourt, J., Tamin, A., Thornburg, N.J., Swanstrom, R., Denison, M.R., Baric, R.S., 2020. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. Sci. Transl. Med. 12, eabb5883 https://doi.org/10.1126/scitranslmed.abb5883.
Shapira, T., Monreal, I.A., Dion, S.P., Buchholz, D.W., Imbiakha, B., Olmstead, A.D., Jager, M., Désilets, A., Gao, G., Martins, M., Vandal, T., Thompson, C.A.H., Chin, A., Rees, W.D., Steiner, T., Nabi, I.R., Marsault, E., Sahler, J., Diel, D.G., Van de Walle, G.R., August, A., Whittaker, G.R., Boudreault, P.L., Leduc, R., Aguilar, H.C., Jean, F., 2022. A TMPRSS2 inhibitor acts as a pan-SARS-CoV-2 prophylactic and therapeutic. Nature 605, 340–348. https://doi.org/10.1038/s41586-022-04661-w.
Sia, S.F., Yan, L.M., Chin, A.W.H., Fung, K., Choy, K.T., Wong, A.Y.L., Kaewpreedee, P., Perera, R.A.P.M., Poon, L.L.M., Nicholls, J.M., Peiris, M., Yen, H.L., 2020. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. Nature 583, 834–838. https://doi.org/10.1038/s41586-020-2342-5.
Tang, W.F., Tsai, H.P., Chang, Y.H., Chang, T.Y., Hsieh, C.F., Lin, C.Y., Lin, G.H., Chen, Y. L., Jheng, J.R., Liu, P.C., Yang, C.M., Chin, Y.F., Chen, C.C., Kau, J.H., Hung, Y.J., Hsieh, P.S., Horng, J.T., 2021. Perilla (Perillafrutescens) leaf extract inhibits SARS CoV-2 via direct virus inactivation. Biomed. J. 44, 293–303. https://doi.org/10.1016/j.bj.2021.01.005.
Vijayakumar, B.G., Ramesh, D., Joji, A., Prakasan, J.J., Kannan, T., 2020. In Silico pharmacokinetic and molecular docking studies of natural flavonoids and synthetic indole chalcones against essential proteins of SARS-CoV-2. Eur. J. Pharmacol. 886, 173448 https://doi.org/10.1016/j.ejphar.2020.173448.
Voskiene, A., Mickevicius, V., Mikulskiene, G., 2007. Synthesis and structural characterization of products condensation 4-carboxy-1-(4-styrylcarbonylphenyl)-2-pyrrolidinones with hydrazines. ARKIVOC (Gainesville, FL, U. S.) 303–314. https://doi.org/10.3998/ark.5550190.0008.f29, 2007.
Vuong, W., Khan, M.B., Fischer, C., Arutyunova, E., Lamer, T., Shields, J., Saffran, H.A., McKay, R.T., van Belkum, M.J., Joyce, M.A., Young, H.S., Tyrrell, D.L., Vederas, J.C., Lemieux, M.J., 2020. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. Nat. Commun. 11, 4282. https://doi.org/10.1038/s41467-020-18096-2.
Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W., Xiao, G., 2020. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 30, 269–271. https://doi.org/10.1038/s41422-020-0282-0.
Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3.
Xu, M., Wu, P., Shen, F., Ji, J., Rakesh, K.P., 2019. Chalcone derivatives and their antibacterial activities: current development. Bioorg. Chem. 91, 103133 https://doi.org/10.1016/j.bioorg.2019.103133.
Yamamoto, M., Kiso, M., Sakai-Tagawa, Y., Iwatsuki-Horimoto, K., Imai, M., Takeda, M., Kinoshita, N., Ohmagari, N., Gohda, J., Semba, K., Matsuda, Z., Kawaguchi, Y., Kawaoka, Y., Inoue, J.I., 2020. The anticoagulant nafamostat potently inhibits SARS-CoV-2 S protein-mediated fusion in a cell fusion assay system and viral infection in vitro in a cell-type-dependent manner. Viruses 12, 629. https://doi.org/10.3390/v12060629.
Yan, J., Chen, S., Zhang, Hu, J., Huang, L., Li, X., 2016. Synthesis, evaluation, and mechanism study of novel Indole-Chalcone derivatives exerting effective antitumor activity through microtubule destabilization in vitro and in vivo. J. Med. Chem. 59, 5264–5283. https://doi.org/10.1021/acs.jmedchem.6b00021.

doi:10.6342/NTU202502385

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., Hilgenfeld, R., 2020. X-ray structure of main protease of the novel coronavirus SARS-CoV-2 enables design of α-ketoamide inhibitors. Science 368, 409–412. https://doi.org/10.1126/science.abb3405.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273. https://doi.org/10.1038/s41586-020-2012-7.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., China Novel Coronavirus Investigating and Research Team, 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733. https://doi.org/10.1056/NEJMoa2001017.

doi:10.6342/NTU202502385

ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Green Synthesis of Tetrahydropyrazino[2,1-a:5,4-a′]diisoquinolines as SARS-CoV-2 Entry Inhibitors

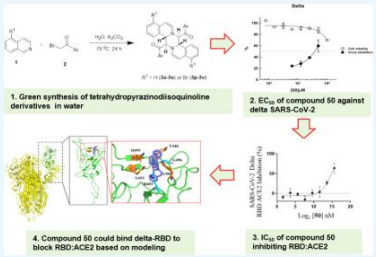
Sowndarya Palla, Srinivasa Rao Palla, Jia-Jin Liu, Tai-Ling Chao, Ting-Hui Lee, Veerababurao Kavala, I-Chen Liu, Lily Hui-Ching Wang, Sui-Yuan Chang, Ching-Fa Yao,* and Po-Huang Liang*

Cite This: https://doi.org/10.1021/acsomega.4c08640

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A class of tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline derivatives were synthesized under environmentally friendly conditions using water as the solvent. The 3-D structures of some synthesized compounds were determined by X-ray diffraction. Since naturally occurring isoquinoline alkaloids have significant antiviral activities against a wide range of viruses, including coronaviruses, the synthesized compounds were assayed for their inhibitory activities against SARS-CoV-2. Our results showed that the active compounds **50** and **96** blocked the delta SARS-CoV-2 entry into VeroE6 cells to display EC_{50} of 26.5 ± 6.9 and 17.0 ± 3.7 μM, respectively, by inhibiting the interaction between SARS-CoV-2 Spike's receptor binding domain (RBD) and human receptor angiotensin-converting enzyme 2 (ACE2), and CC_{50} greater than 100 μM. This study provides a green synthesis method of tetrahydropyrazinodiisoquinoline for antiviral or other applications.



1. INTRODUCTION

Since December 2019, coronavirus disease 2019 (COVID-19) has caused a worldwide pandemic with an approximately 1% case mortality rate.[1−3] The pathogen is human severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) with a genome highly similar (79.6% identity) to that of the previously (2002−2003) emerged SARS-CoV.[4,5] Unlike another CoV called Middle East Respiratory Syndrome-CoV (MERS-CoV) that employs dipeptidyl peptidase-4 (DPP4) as the human receptor to anchor the virus, SARS-CoV and SARS-CoV-2 utilize human angiotensin-converting enzyme 2 (ACE2) to bind with the virus surface Spike glycoprotein through the receptor binding domain (RBD).[6] The virus entry after RBD:ACE2 interaction is initiated by the cleavage of Spike at S1/S2 and S2′ sites by two human proteases, Furin and transmembrane protease serine 2 (TMPRSS2), respectively. The primed Spike exposes the membrane fusion peptide to fuse with the host cell membrane and the virus's RNA enters into the cell.[7] TMPRSS2 could replace the function of Furin so that the irreversible serine protease inhibitors Camostat and Nafamost antagonize SARS-CoV-2 by inhibiting TMPRSS2 as off-target.[8,9] Alternatively to the above pathway, another human protease Cathepsin L is involved in the endosomal virus entry, mainly found in SARS-CoV-2 omicron strain.[10]

After entry into host cell, the 29.7-kb SARS-CoV-2 positive-sense single-strand RNA is translated by human cell machinery to form two polyproteins pp1a and pp1ab that can be processed by two virus-encoded proteases, 3C-like protease ($3CL^{pro}$) as homologous to the picornavirus 3C protease and papain-like protease (PL^{pro}), into 16 functional nonstructural proteins (NSPs).[11−13] Mature NSPs, including NSP7, 8, 10, 12, and 13 then form a replication-transcription complex with RNA-dependent RNA polymerase (RdRp), helicase, exonuclease, and other activities for making subgenomic RNAs of the structural proteins, Envelop (E), Nucleocapsid (N), Membrane (M), and Spike (S) proteins. The structural proteins are assembled with the RNA into new virus particles.[14] Therefore, these two viral proteases and RdRp are attractive drug targets, and many covalent or noncovalent inhibitors have been discovered. To date, Pfizer Co. has developed Paxlovid, a nitrile-based covalent $3CL^{pro}$ peptidomimetic inhibitor nirmatrelvir combined with ritonavir for inhibiting CYP3A4 as a market drug.[15] A noncovalent nonpeptidomimetic $3CL^{pro}$ inhibitor ensitrelvir has been approved for use in Japan.[16] In China, leritrelvir[17] and simnotrelvir[18,19] have been conditionally approved. Remdesivir, a nucleotide analog prodrug developed by Gilead Co., was found to inhibit SARS-CoV-2 RdRp.[20,21] It showed promising efficacy in clinical trials and thus was approved for COVID-19 treatment.[22,23] Another nucleotide analog, molnupiravir, a prodrug of the nucleoside analog N4-hydroxycytidine developed by Merck Co. has also

Received: September 20, 2024
Revised: December 10, 2024
Accepted: December 16, 2024

doi:10.6342/NTU202502385

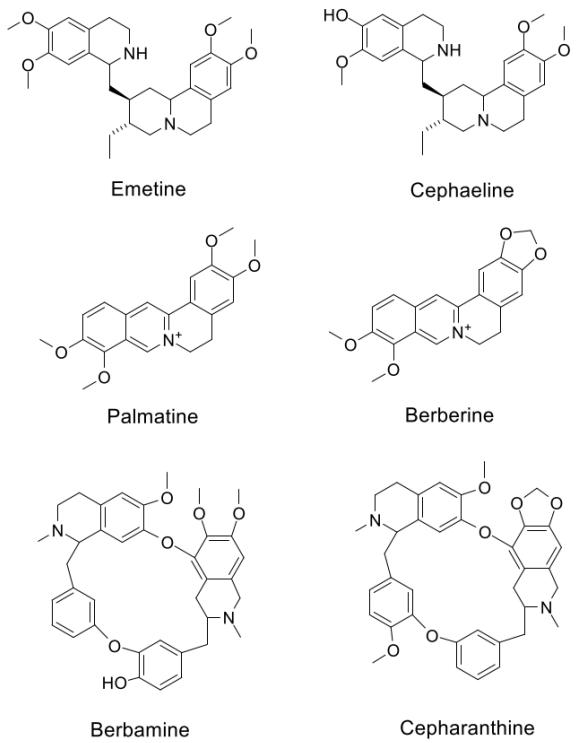


Figure 1. Chemical structures of some biologically active isoquinoline derivatives. These include emetine and its OCH_3 group reduced analogue cephaline that shows lower toxicity, palmatine, and berberine containing a nitrogen cation, as well as berbamine, and cepharanthine in cyclic forms.

been approved for anti-COVID-19.[24,25] For preventing virus entry, several antibodies have been developed as approved drugs.[26] However, no small-molecule drug to inhibit virus entry has been approved, although some small-molecule inhibitors of RBD:ACE2 interaction and/or human proteases have been identified.[27−32] However, resistance problems have occurred on various anti-COVID-19 drugs.[33−37] Therefore, continuing to search for drug candidates is still necessary.

Isoquinoline alkaloids are a class of naturally occurring molecules that have significant antiviral activities against a wide range of viruses, including CoVs.[38] Isoquinoline is a unique nucleus in chemistry, which structurally is composed of two rings of benzene and pyridine in such a way that the nitrogen of pyridine is placed away from the benzene ring. Isoquinoline alkaloids usually include all alkaloids that have various isoquinoline analogs, for example, 1,2,3,4-tetrahydroisoquinoline (containing a saturated pyridine ring with a secondary amine) such as tylophorine, dicentrine, corydaline, discretine, lycorine, emetine, argemonine, etc. (the chemical structures of emetine and cephaline with the OCH_3 being reduced to OH to decrease cytotoxicity, are shown in Figure 1 as examples). In addition, an important class of isoquinoline alkaloids have interconnected rings, in which the nitrogen atom has become a quaternary cation (the chemical structures of palmatine and berberine are shown in Figure 1 as examples). The isoquinoline alkaloids such as emetine, palmatine, and berberine could inhibit SARS-CoV-2 by multiple-targeting S protein, RdRp, eIF4A, N, and/or orf6 proteins.[39] Moreover, a cyclic bis-benzylisoquinoline alkaloid berbamine (structure shown in Figure 1) could inhibit SARS-CoV-2 by binding to the postfusion core of Spike's S2 subunit.[40] From a library containing 188 natural products, cyclic bis-benzylisoquinoline alkaloids such as cepharanthine (structure shown in Figure 1) were identified as SARS-CoV-2 entry inhibitors.[41]

Similar to cyclic bis-benzylisoquinoline alkaloids, our *in silico* prestudy revealed that the extended structures of diisoquinoline derivatives may block RBD:ACE2 interaction, thereby inhibiting SARS-CoV-2 entry. As presented herein, we developed a green synthesis protocol using water as solvent for tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline derivatives from isoquinoline or bromoisoquinoline and various phenacyl bromides in water and optimized their yields. Since regulations are becoming increasingly severe regarding production and use of organic solvents, forcing chemists to find greener and safer alternatives, water as the main if not exclusive reaction medium for organic transformations represents a safe, nontoxic, cheap, and environmentally friendly alternative (42 for a review of the reactions using water as solvent). The synthesized compounds

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

were characterized for their three-dimensional (3D) structures by X-ray diffraction and evaluated for antagonizing SARS-CoV-2 through inhibiting RBD:ACE2 interaction. The green methodology developed here is useful to synthesize tetrahydropyrazinodiisoquinoline for antiviral or other applications.

2. EXPERIMENTAL METHODS

2.1. General Methods. Reactions for synthesizing compounds were monitored by using thin-layer chromatography (TLC) on silica gel. Flash chromatography was performed on silica gel of 60−200 μm particle size for compound purification. Yields were reported for spectroscopically pure compounds. Melting points were recorded on the Fargo MP-1D Melting Point Apparatus. NMR spectra were recorded on a Bruker AVIIIHD 400 MHz FT-NMR instrument in the Department of Chemistry, National Taiwan Normal University. Chemical shifts were given in δ values relative to tetramethylsilane (TMS); coupling constants J were given in Hz. Internal standards were $CDCl_3$ (δ_H = 7.24) or DMSO-d_6 (δ_H = 2.49) for ^{1}H NMR spectra, and $CDCl_3$ (δ_C = 77.0) or DMSO-d_6 (δ_C = 39.5) for ^{13}C NMR spectra. The splitting patterns were reported as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), br (broad), and dd (double of doublets). High-resolution mass spectra were measured by a Bruker UPLC-MS instrument in the TechComm core facility, Department of Chemistry, National Taiwan Normal University.

2.2. Synthesis of Tetrahydropyrazino[2,1-a:5,4-a′]-diisoquinoline Derivatives. A mixture of isoquinoline **1** (1 mmol), 2-bromoacetophenone **2** (1 mmol), K_2CO_3 (1 mmol), and H_2O (0.5 mL) was stirred at 70 °C for 24 h. Upon the consumption of isoquinoline **1** (monitored by TLC), the reaction was cooled to room temperature, filtered, concentrated, and purified by silica gel flash chromatography (Hex/EtOAc) to afford compound **3a**. Other compounds were synthesized from isoquinoline or bromoisoquinoline and various phenacyl bromides under the same condition.

*2.2.1. Compound **3a** (**50**).* Light yellow solid, 86% yield. Mp: 178−179 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.75 (dd, J = 6.84 Hz, 4H), 7.47−7.42 (m, 2H), 7.35−7.31 (m, 4H), 7.06−7.01 (m, 2H), 6.90−6.83 (m, 6H), 6.37 (dd, J = 7.56 Hz, 2H), 5.59 (d, J = 7.28 Hz, 2H), 5.43 (d, J = 7.44 Hz, 2H), 5.01 (d, J = 8.9 Hz, 2H), ^{13}C NMR (100 MHz, $CDCl_3$): δ 194.6, 138.1, 136.2, 133.2, 131.9, 131.6, 128.6, 128.4, 128.3, 127.3, 126.2, 125.6, 123.2, 101.2, 65.1, 50.2. HRMS (ESI): m/z calculated for $C_{34}H_{27}N_2O_2^+$ [M + H]$^+$: 495.2067, found 495.2051.

*2.2.2. Compound **3b** (**105**).* Yellow solid, 53% yield. Mp: 136−137 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.29 (d, J = 6.84 Hz, 2H), 7.25 (s, 5H), 7.20−7.16 (m, 2H), 7.13 (dd, J = 6.24 Hz, 1H), 7.09 (dd, J = 4.72 Hz, 2H), 6.99 (td, J = 7.44 Hz, 2H), 6.89 (d, J = 7.56 Hz, 2H), 6.01 (d, J = 7.28 Hz, 2H), 5.42 (d, J = 7.44 Hz, 2H), 5.26 (d, J = 8.88 Hz, 2H), 4.81 (d, J = 8.84 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 198.1, 152.5, 138.2, 137.8, 131.8, 131.6, 130.9, 130.2, 129.3, 128.1, 126.7, 125.7, 123.5, 120.5, 100.9, 68.7, 49.9. HRMS (ESI): m/z calculated for $C_{34}H_{24}Cl_2{}^{(35)}N_2O_2^+$ [M]$^+$: 564.1180, found 564.1101 and for $C_{34}H_{24}Cl_2{}^{(37)}N_2O_2^+$ [M]$^+$: 566.1150, found 566.1160.

*2.2.3. Compound **3c** (**110**).* Light yellow solid, 61% yield. Mp: 172−173 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.62 (dd, J = 5.96 Hz, 3H), 7.37−7.33 (m, 2H), 7.08−7.04 (m, 2H), 6.96−6.84 (m, 8H), 6.75 (d, J = 8.36 Hz, 2H), 6.19 (d, J = 7.48 Hz, 2H), 5.44 (d, J = 7.44 Hz, 2H), 5.30 (d, J = 9.24 Hz, 2H), 5.16 (d, J = 8.92 Hz, 2H), 3.44 (s, 6H). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 199.0, 158.0, 139.0, 133.5, 132.4, 130.7, 127.5, 127.0, 123.0, 122.9, 120.8, 111.3, 99.2, 68.9, 55.2, 51.1. HRMS (ESI): m/z calculated for $C_{36}H_{31}N_2O_4$ [M + H]$^+$: 555.2278, found 555.2269.

*2.2.4. Compound **3d** (**103**).* Light orange solid, 53% yield. Mp: 177−178 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.23 (d, J = 1.28 Hz, 1H), 7.16−7.06 (m, 9H), 6.96−6.89 (m, 6H), 6.12 (t, J = 7.52 Hz, 2H), 5.45 (d, J = 7.56 Hz, 2H), 5.43 (d, J = 7.56 Hz, 2H), 4.79 (d, J = 9.08 Hz, 2H), 2.38 (s, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 198.4, 138.5, 138.1, 137.6, 132.0, 131.7, 131.0, 128.1, 127.7, 127.5, 126.6, 125.5, 125.3,123.5, 100.6, 67.4, 49.9, 20.7. HRMS (MALDI-TOF): m/z calculated for $C_{36}H_{30}N_2O_2^+$ [M]$^+$: 522.2307, found 522.2301.

*2.2.5. Compound **3e** (**74**).* Whitish yellow solid, 60% yield. Mp: 140−141 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.79 (dd, J = 7.76 Hz, 4H), 7.10−6.86 (m, 12H), 6.37 (d, J = 7.44 Hz, 2H), 5.62 (d, J = 7.4 Hz, 2H), 5.40 (d, J = 8.8 Hz, 2H), 4.96 (d, J = 8.84 Hz, 2H). ^{13}C NMR (100 MHz, DMSO-d_6): δ = 194.4, 166.9, 164.4, 138.6, 132.5, 132.4, 132.3, 128.4, 127.5, 126.3, 125.9, 123.8, 116.3, 116.1, 101.0, 64.3, 50.2. HRMS (ESI): m/z calculated for $C_{34}H_{25}F_2N_2O_2^+$ [M + H]$^+$: 531.1879, found 531.1885.

*2.2.6. Compound **3f** (**96**).* Yellow solid, 54% yield. Mp: 160−161 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.68−7.66 (m, 4H), 7.32−7.28 (m, 4H), 7.09−7.05 (m, 2H), 6.93 (d, J = 7.48 Hz, 2H), 6.89−6.84 (m, 4H), 6.34 (dd, J = 7.48 Hz, 2H), 5.62 (d, J = 7.48 Hz, 2H), 5.36 (dd, J = 8.8 Hz, 2H), 4.92 (d, J = 8.8 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 193.7, 139.7, 137.7, 134.4, 131.7, 130.9, 130.0, 128.8, 128.1, 127.6, 127.2, 126.7, 125.8, 123.7, 101.5, 65.2, 50.2. HRMS (ESI): m/z calculated for $C_{17}H_{13}ClNO^{2+}$ [M/2 + H]$^{2+}$: 282.0680, found 282.0686.

*2.2.7. Compound **3g** (**68**).* Light yellow solid, 30% yield. Mp: 173−174 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.66 (d, J = 7.80 Hz, 4H), 7.14 (d, J = 7.84 Hz, 4H), 7.05 (t, J = 6.68 Hz, 2H), 6.91−6.81 (m, 6H), 6.38 (d, J = 7.36 Hz, 2H), 5.58 (d, J = 7.36 Hz, 2H), 5.44 (d, J = 8.84 Hz, 2H), 4.99 (d, J = 8.84 Hz, 2H), 2.32 (s, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ = 194.3, 144.0 138.1, 133.7, 131.9, 129.1, 128.8, 128.2, 127.9, 127.8, 127.6, 126.8, 125.5, 123.4, 100.8, 65.0, 50.2, 21.5. HRMS (ESI): m/z calculated for $C_{36}H_{31}N_2O_2^+$ [M + H]$^+$: 523.2380, found 523.2383.

*2.2.8. Compound **3h** (**72**).* Light orange solid, 21% yield. Mp: 152−153 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.74 (d, J = 8.84 Hz, 4H), 7.06−7.02 (m, 2H) 6.91 (d, J = 7.4 Hz, 2H), 6.86−6.79 (m, 8H), 6.38 (d, J = 7.4 Hz, 2H), 5.57 (d, J = 7.44 Hz, 2H), 5.47 (d, J = 8.9 Hz, 2H), 4.98 (d, J = 8.9 Hz, 2H), 3.79 (s, 6H). ^{13}C (100 MHz, $CDCl_3$): δ 193.4, 163.5, 138.1, 132.0, 131.0, 129.2, 127.9, 127.8, 126.8, 125.5, 123.3, 113.7, 100.6, 64.9, 60.3, 55.4, 50.4, 14.1. HRMS (ESI): m/z calculated for $C_{36}H_{31}N_2O_4^+$ [M + H]$^+$: 555.2278, found 555.2279.

*2.2.9. Compound **3i** (**115**).* Light yellow solid, 57% yield. Mp: 163−164 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.88 (d, J = 8.4 Hz, 3H), 7.60−7.47 (m, 8H), 7.47−7.39 (m, 8H), 7.11 (td, J = 7.16 Hz, 2H), 6.97−6.87 (m, 5H), 6.47 (d, J = 7.5 Hz, 2H), 5.66 (d, J = 7.44 Hz, 2H), 5.51 (d, J = 8.9 Hz, 2H), 5.08 (d, J = 8.9 Hz, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ: 194.2, 145.9, 139.7, 138.1, 131.9, 130.4, 129.3, 129.0, 128.9, 128.2, 128.0, 127.8, 127.6, 127.2, 127.1, 126.8, 126.6, 126.4, 125.7, 123.5, 101.2, 65.2, 50.2. HRMS (ESI): m/z calculated for $C_{46}H_{35}N_2O_2^+$ [M + H]$^+$: 647.2693, found 647.2660.

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

*2.2.10. Compound **3j (106)**.* Light orange solid, 60% yield. Mp: 167−168 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.10 (s, 2H), 7.87−7.80 (m, 2H), 7.81−7.80 (m, 5H), 7.56−7.51 (m, 2H), 7.49−7.46 (m, 3H), 7.05−6.99 (m, 2H), 6.94 (d, J = 6.84 Hz, 2H), 6.88 (d, J = 7.12 Hz, 2H), 6.79 (td, J = 6.32 Hz, 2H), 6.48 (d, J = 7.52 Hz, 2H), 5.65 (d, J = 7.44 Hz, 2H), 5.57 (d, J = 8.9 Hz, 2H), 5.22 (d, J = 8.9 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): 195.1, 152.4, 142.8, 138.0, 135.5, 133.6, 132.2, 132.0, 130.6, 129.6, 127.6, 127.6, 126.9, 126.6, 125.6, 124.2, 123.5, 120.5, 101.0, 65.3, 50.8. HRMS (ESI): m/z calculated for $C_{21}H_{16}NO^{2+}$ $[M/2 + H]^{2+}$: 298.1226, found 298.1232.

*2.2.11. Compound **3k (111)**.* Light orange solid, 70% yield. Mp: 174−175 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.37 (d, J = 8.08 Hz, 2H), 7.89 (d, J = 8.08 Hz, 2H), 7.82 (d, J = 7.4 Hz, 3H), 7.54−7.46 (m, 7H), 7.34−7.28 (m, 4H), 7.10 (td, J = 6.32 Hz, 2H), 7.01 (t, J = 6.96 Hz, 2H), 6.91−6.85 (m, 3H), 6.12 (t, J = 6.68 Hz, 2H), 5.71 (d, J = 9.4 Hz, 2H), 5.44 (d, J = 7.48 Hz, 2H), 5.04 (d, J = 9.2 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$); δ 198.8, 138.1, 135.4, 133.8, 132.7, 132.1, 130.5, 128.4, 128.2, 128.0, 127.6, 127.4, 126.8, 126.5, 125.6, 125.6, 124.2, 123.6, 100.7, 67.8, 50.7. HRMS (ESI): m/z calculated for $C_{42}H_{31}N_2O_2^+$ $[M + H]^+$: 595.2380, found 595.2386.

*2.2.12. Compound **3l (88)**.* Yellow solid, 22% yield. Mp: 148−149 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.41 (d, J = 1.92 Hz, 2H), 7.26−7.23 (m, 2H), 7.07−7.03 (m, 2H) 6.92 (d, J = 7.5 Hz, 2H), 6.85 (d, J = 4.40 Hz, 4H), 6.72 (d, J = 8.48 Hz, 2H), 6.37 (d, J = 6.92 Hz, 2H), 5.57 (d, J = 7.44 Hz, 2H), 5.48 (d, J = 8.84 Hz, 2H), 4.99 (d, J = 8.96 Hz, 2H), 3.86 (s, 12H). ^{13}C (100 MHz, $CDCl_3$): δ 193.5, 153.4, 149.0, 138.0, 131.9, 129.3, 127.9, 127.8, 126.9, 125.5, 123.4, 123.2, 111.0, 109.9, 100.6, 64.9, 56.0, 55.8, 50.7. HRMS (ESI): m/z calculated for $C_{38}H_{35}N_2O_6^+$ $[M + H]^+$: 615.2490, found 615.2497.

*2.2.13. Compound **3m (98)**.* Yellow solid, 78% yield. Mp: 176−177 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 7.27−7.24 (m, 3H), 7.07−7.02 (m, 2H), 6.92 (d, J = 7.48 Hz, 2H), 6.86−6.82 (m, 5H), 6.69 (d, J = 8.0 Hz, 2H), 6.34 (d, J = 7.52 Hz, 2H), 5.97−5.96 (m, 4H), 5.57 (d, J = 7.48 Hz, 2H), 5.43 (d, J = 8.92 Hz, 2H), 4.92 (d, J = 8.96 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ = 192.9, 151.9, 148.1, 137.9, 131.9, 131.0, 127.9, 127.7, 126.8, 125.5, 125.0, 123.4, 108.4, 107.7, 101.8, 100.8, 64.9, 50.5. HRMS (ESI): m/z calculated for $C_{36}H_{27}N_2O_6^+$ $[M + H]^+$: 583.1864, found 583.1867.

*2.2.14. Compound **3n (114)**.* Light orange solid, 30% yield. Mp: 153−154 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.64−7.61 (m, 4H), 7.51−7.49 (m, 4H), 7.12−7.08 (m, 2H), 6.96 (d, J = 7.48 Hz, 2H), 6.92−6.87 (m, 3H), 6.37 (d, J = 7.52 Hz, 1H), 5.65 (d, J = 7.48 Hz, 2H), 5.37 (t, J = 8.76 Hz, 2H), 4.94 (d, J = 8.84 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$); δ 193.8, 152.3, 142.6, 137.7, 135.8, 134.8, 132.2, 131.8, 130.1, 129.4, 126.5, 125.8, 123.7, 120.5, 101.6, 65.2, 50.1. HRMS (EI): m/z calculated for $C_{17}H_{11}Br_2NOK^+$ $[M/2 + K]^+$: 441.8833, found 441.0895.

*2.2.15. Compound **3o (122)**.* Light orange solid, 74% yield. Mp: 148−149 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.58−7.57 (m, 2H), 7.54−7.53 (m, 2H), 7.13−7.09 (m, 2H), 7.02−6.91 (m, 8H), 6.38 (d, J = 7.24 Hz, 2H), 5.66 (d, J = 7.48 Hz, 2H), 5.40 (d, J = 8.72 Hz, 2H), 4.74 (d, J = 8.8 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 188.3, 152.4, 142.6, 138.0, 134.3, 132.9, 131.7, 128.0, 127.1, 125.7, 123.5, 101.1, 67.6, 50.7, 29.7. HRMS (ESI): m/z calculated for $C_{30}H_{23}N_2O_2S_2^+$ $[M + H]^+$: 507.1195, found 507.1201.

*2.2.16. Compound **3p (50Br)**.* Light yellow solid, 94% yield. Mp: 115−116 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.82 (d, J = 7.52 Hz, 4H), 7.55 (t, J = 7.52 Hz, 2H), 7.42 (t, J = 7.64 Hz, 4H), 7.32 (d, J = 7.88 Hz, 2H), 6.85 (d, J = 7.36 Hz, 2H), 6.74 (t, J = 7.76 Hz, 2H), 6.55 (d, J = 7.56 Hz, 2H), 5.95 (d, J = 7.64 Hz, 2H), 5.44 (d, J = 9 Hz, 2H), 5.07 (d, J = 8.8 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 193.9, 189.5, 172.1, 139.6, 135.8, 133.5, 132.1, 131.4, 129.1, 128.7, 128.7, 126.6, 126.0, 119.3, 100.0, 64.6, 50.1. HRMS (MALDI-TOF): m/z calcd for $C_{34}H_{24}Br_2N_2O_2$ $[M]^+$: 650.0205, found 650.0285.

*2.2.17. Compound **3q (103Br)**.* Yellow solid, 34% yield. Mp: 177−178 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.37−7.29 (m, 4H), 7.21−7.18 (m, 4H), 7.15 (t, J = 7.2 Hz, 2H), 6.92 (d, J = 7.3 Hz, 2H), 6.82−6.77 (m, 2H), 6.29 (d, J = 8.0 Hz, 2H), 5.81 (d, J = 7.6 Hz, 2H), 5.45 (m, J = 8.0 Hz, 2H) 4.87 (d, J = 9 Hz, 2H), 2.43 (s, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 197.7, 139.8, 138.7, 137.1, 132.2, 131.9, 131.6, 131.4, 128.8, 127.8, 126.5, 125.8, 125.5, 119.3, 99.5, 66.9, 50.1, 20.8. HRMS (MALDI-TOF): m/z calculated for $C_{36}H_{28}Br_2N_2O_2^+$ $[M]^+$: 678.0518, found 678.0512.

*2.2.18. Compound **3r (91Br)**.* Yellow solid, 21% yield. Mp: 145−146 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.38 (s, 2H), 7.32 (t, J = 7.4 Hz, 6H), 7.08−7.05 (m, 2H), 6.84 (d, J = 7.36 Hz, 2H), 6.75 (t, J = 7.76 Hz, 2H), 6.53 (d, J = 7.72 Hz, 2H), 5.94 (d, J = 7.64 Hz, 2H), 5.43 (d, J = 8.84 Hz, 2H), 5.05 (d, J = 8.92 Hz, 2H), 3.83 (s, 6H). ^{13}C NMR (100 MHz, $CDCl_3$); δ 193.7, 159.9, 139.6, 137.1, 132.1, 129.6, 126.6, 126.0, 121.2, 120.2, 112.9, 100.0, 64.7, 58.4, 55.4, 50.2, 30.9,18.4. HRMS (ESI): m/z calculated for $C_{36}H_{29}Br_2N_2O_4^+$ $[M + H]^+$: 711.0489, found 711.0494.

*2.2.19. Compound **3s (68Br)**.* Yellow solid, 60% yield. Mp: 169−170 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.71 (d, J = 8.16 Hz, 3H), 7.31 (d, J = 8.0 Hz, 3H), 7.20 (d, J = 8.1 Hz, 4H), 6.84 (d, J = 7.44 Hz, 2H), 6.73 (t, J = 7.7 Hz, 2H), 6.54 (d, J = 7.6 Hz, 2H), 5.92 (d, J = 7.6 Hz, 2H), 5.44 (d, J = 8.8 Hz, 2H), 5.04 (d, J = 8.8 Hz, 2H), 2.36 (s, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 193.5, 144.5, 139.7, 133.2, 132.0, 131.4, 129.3, 129.2, 128.8, 126.5, 126.0, 119.2, 99.8, 64.5, 50.1, 21.6. HRMS (ESI): m/z calculated for $C_{36}H_{29}Br_2N_2O_2^+$ $[M + H]^+$: 679.0590, found 679.0596.

*2.2.20. Compound **3t (70Br)**.* Pale yellow solid, 44% yield. Mp: 181−182 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.79 (d, J = 8.8 Hz, 4H), 7.31 (d, J = 8 Hz, 2H), 6.87−6.81 (m, 6H) 6.73 (t, J = 7.7 Hz, 2H), 6.53 (d, J = 7.68 Hz, 2H), 5.91 (d, J = 7.56 Hz, 2H), 5.46 (d, J = 8.92 Hz, 2H), 5.03 (d, J = 8.8 Hz, 2H), 3.83 (s, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 192.5, 163.8, 148.0, 139.8, 132.0, 131.5, 131.1, 129.2, 128.8, 126.5, 126.1, 119.1, 113.8, 99.0, 64.5, 55.5, 50.4. HRMS (ESI): m/z calculated for $C_{18}H_{15}BrNO_2^{2+}$ $[M/2 + H]^{2+}$: 356.0275, found 356.0294.

*2.2.21. Compound **3u (122Br)**.* Light yellow solid, 44% yield. Mp: 176−177 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.62 (d, J = 4.8 Hz, 2H), 7.58 (d, J = 3.8 Hz, 2H), 7.35 (d, J = 7.9 Hz, 2H) 7.06 (t, J = 4.4 Hz, 2H), 6.96 (d, J = 7.4 Hz, 2H), 6.80 (t, J = 7.7 Hz, 2H), 6.50 (d, J = 7.6 Hz, 2H), 5.98 (d, J = 7.6 Hz, 2H), 5.39 (d, J = 8.7 Hz, 2H), 4.78 (d, J = 8.7 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 187.5, 142.2, 139.6, 134.7, 133.2, 132.2, 131.3, 129.0, 128.2, 126.7, 126.3, 119.3, 100.1, 67.1, 50.7. HRMS (MALDI-TOF): m/z calcd for $C_{30}H_{20}Br_2N_2O_2S_2^+$ $[M]^+$: 661.9333, found 661.9306.

*2.2.22. Compound **3v (115Br)**.* Light orange solid, 56% yield. Mp: 156−157 °C. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.91 (d, J = 8.3 Hz, 3H), 7.63 (d, J = 8.4 Hz, 4H), 7.59 (t, J = 7.1 Hz, 4H), 7.48−7.38 (m, 7H), 7.33 (t, J = 7.3 Hz, 2H), 6.88 (d, J = 7.4 Hz, 2H), 6.76 (t, J = 7.7 Hz, 2H), 6.60 (d, J = 7.6 Hz,

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

2H), 5.98 (d, J = 7.6 Hz, 2H), 5.48 (d, J = 8.8 Hz, 2H), 5.11 (d, J = 8.8 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 193.4, 146.3. 139.7, 139.6, 134.4, 132.1, 131.4, 130.5, 129.3, 129.1, 128.9, 128.3, 128.0, 127.3, 127.2, 127.0, 126.7, 126.1, 119.3, 100.1, 64.8, 50.2. HRMS (EI): m/z calculated for $C_{46}H_{33}Br_2N_2O_2^+$ $[M + H]^+$: 803.0898, found 803.0909.

2.3. X-ray Analysis of the Synthesized Compounds. Single-crystal X-ray data were acquired through a slow evaporation technique utilizing dichloromethane with ethyl acetate as the solvent system at room temperature. X-ray reflections were recorded at 200 K on the single crystals using Mo Kα X-radiation (λ = 0.71073 Å) with a Bruker Kappa APEX-II diffractometer. The crystal structures were solved and refined using SHELX-97. Additional information regarding the data collection and refinement parameters for the crystals can be found in the Supporting Information Tables S2−S6.

2.4. Antivirus EC_{50} and Cytotoxicity CC_{50} Measurements. The antiviral EC_{50} values of the synthesized compounds against SARS-CoV-2 were determined by using the plaque reduction assay as previously described.[26,43] For the assay, VeroE6 cells were seeded into a 24-well culture plate in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS) and antibiotics 1 day before the infection. VeroE6 cells were infected by SARS-CoV-2 delta virus (NTU92) at 50−100 pfu for 1 h at 37 °C. After removal of virus inoculum, the cells were washed once with phosphate-buffered saline (PBS) and overlaid with 1 mL overlay medium containing 1% methylcellulose for 5 days at 37 °C. After 5 days, the cells were fixed with 10% formalin overnight. After removal of the overlay medium, the cells were stained with 0.5% crystal violet, and the plaques were counted. The percentage of inhibition was calculated as $[1 - (V_D/V_C)] \times 100\%$, where V_D and V_C refer to the virus titer in the presence and absence of an inhibitor, respectively. The minimal concentration of an antiviral required to reduce the plaque number by 50% (EC_{50}) was calculated by regression analysis of the dose−response curve generated from the plaque assay.

Cytotoxicity of the inhibitors was determined by using acid phosphatase assay. Briefly, VeroE6 cells were seeded onto a 96-well culture plate at a concentration of 2×10^4 cells per well. Next day, the medium was removed, and each well was washed once with PBS before adding DMEM containing 2% FBS and different concentrations of an inhibitor. Next, DMEM containing 2 μg/mL TPCK-trypsin was added. After 1 h of incubation at 37 °C, the medium was removed, and cells were washed by PBS. Then, DMEM containing 2% FBS and different concentrations of an antiviral was added. After 3 days of incubation at 37 °C, the medium was removed, and each well was washed once with PBS. Next, buffer containing 0.1 M sodium acetate (pH = 5.0), 0.1% Triton X-100, and 5 mM p-nitrophenyl phosphate was added. After the mixture was incubated at 37 °C for 2 h, 1 N NaOH was added to stop the reaction. The absorbance was then determined by an ELISA reader (VERSAmax, Molecular Devices, Sunnyvale, CA) at a wavelength of 405 nm. The percentage of cytotoxicity was calculated using the following formula: cytotoxicity % = $[1 - (A_t/A_s) \times 100]$%, where A_t and A_s refer to the absorbance of a tested substance and solvent control, respectively. The 50% cytotoxicity concentration (CC_{50}) was defined as the concentration reducing 50% of cell viability. For each data point, the measurements were repeated three times to yield the averaged CC_{50} values and the standard deviation.

2.5. Test of the Synthesized Compounds on Inhibiting RBD:ACE2. The active antivirals were tested for inhibiting RBD:ACE2 interaction using the NanoBiT technology commercial kit from Promega (WI, USA) as reported previously.[26] The IC_{50} values of the synthesized compounds against the target were determined from the concentration-dependent inhibition curves fitted with the equation: $A(I) = A(0) \times \{1 - [I/(I + IC_{50})]\}$ using GraphPad Prism software (v.9.4.0).

2.6. Molecular Docking. Molecular docking studies were carried out using the iGEMDOCK software available at http://gemdock.life.nctu.edu.tw/dock/igemdock.php[44] to predict the binding interactions of compound **50** with RBD of the SARS-CoV-2 delta variant Spike protein. The three-dimensional (3D) structure of RBD (PDB ID: 7w92)[45] was obtained from the RCSB Protein Data Bank (PDB, https://www.rcsb.org/). Before docking, all water molecules were removed from the structure. The RBD domain (residues 319 to 541) was isolated from the open state of the SARS-CoV-2 delta variant Spike protein. This binding site was prepared by assigning residue atom types and charges using the iGEMDOCK method. The 3D structure of compound **50** was created using the Molview Web site (https://molview.org/) and its structural data were converted to mol2 format with the Open Babel GUI software.[46]

For the molecular docking process, genetic algorithm (GA) settings were used with population sizes of 800, 80 generations, and 10 solutions. iGEMDOCK software was used to create profiles of protein−ligand interactions based on electrostatic (E), hydrogen-bonding (H), and van der Waals (V) forces. After docking, iGEMDOCK analyzed and ranked all poses based on their estimated binding energies. The binding energy scores were calculated as the sum of the interaction energies, and the pose with the lowest energy was considered the best fit for compound **50** in the target binding site.

2.7. Drug-likeness Analysis. The 3D structure of compound **50** was converted to an SMILES format using Open Babel GUI software. To evaluate the drug-likeness of compound **50**, the Lipinski rule of five was assessed using an online tool (http://www.scfbio-iitd.res.in/software/drugdesign/lipinski.jsp).[47,48] The Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) profile of **50** was predicted through the pkCSM Web site (http://biosig.unimelb.edu.au/pkcsm/).[49]

3. RESULTS

3.1. Synthesis of Tetrahydropyrazino[2,1-a:5,4-a′]-diisoquinoline Derivatives from Isoquinoline and Various Phenacyl Bromides. To synthesize the target molecules, in the beginning, a mixture of isoquinoline **1** (1 mmol) and 2-bromoacetophenone **2** (1 mmol) with different solvents and different bases under different temperatures was tried. EtOH as solvent and K_2CO_3 as base at the temperature of 40 °C gave the yield of 86% for compound **3a**. For further optimization, we found using K_2CO_3 as base and H_2O as solvent, at 70 °C was sufficient to give the product with a higher yield of 96% (Table 1). Therefore, our general synthetic scheme (Scheme 1) became a mixture of isoquinoline **1** (1 mmol) with R^1 substituent H or Br, phenacyl bromide **2** (1 mmol), K_2CO_3 (1 mmol), and H_2O (0.5 mL), and stirred at 70 °C for 24 h. Upon the consumption of isoquinoline **1** (monitored by TLC), the reaction mixtures were cooled to room temperature, filtered, concentrated, and purified by silica

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

Table 1. Yields of 3a Using the Reaction Conditions of 1 (1 mmol), 2 (1 mmol), K_2CO_3 (1 mmol), and H_2O (0.5 mL), at Various Temperatures and Time Periods

entry	solvent	base	TEMP (°C)	*t* (h)	yield (%)
1	H_2O	K_2CO_3	sonication	2	20
2	H_2O	K_2CO_3	rt	12	30
3	H_2O	K_2CO_3	40 °C	28	42
4	H_2O	K_2CO_3	50 °C	28	58
5	H_2O	K_2CO_3	70 °C	24	96
6	H_2O	K_2CO_3	90 °C	24	96
7	H_2O	K_2CO_3	100 °C	24	96

gel flash chromatography (Hex/EtOAc) to afford compounds **3a−3v**, each as a single diastereomer.

The chemical structures of the synthesized compounds were confirmed by NMR and MS spectral data. The NMR assignment was made on the substituent additivity rules, spectral characteristics of structurally related compounds, signal intensities, and multiplicities. ^{13}C NMR spectra were used to prove the interpretation of the carbon resonances. The ^{1}H and ^{13}C NMR spectra for structural characterization are presented in Figure S1. The 3-D structures of five selected compounds were characterized by X-ray diffraction (the statistics are summarized in Tables S2−S6) and are shown in Figure 2.

The chemical structures and the yields of the synthesized compounds are summarized in Tables 2 and 3. The reaction of unsubstituted phenacyl bromide gave an excellent yield (96%) of the product (**3a**). However, *ortho*- or *para*-substituted phenacyl bromide substrates containing a halogen atom or methyl group produced the desired products in moderate yields (50−68%). On the other hand, the substrates containing bulky diphenyl or naphthyl groups afforded the corresponding tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline derivatives in lower yields (40−50%). Further, the substrate bearing methylenedioxy moiety furnished an excellent yield of the product **3m** (98%). The method was efficient, involving the reaction of the phenacyl bromide possessing thiophene group to produce the desired product **3o** in a good yield (74%). Next, we explored the methodology by replacing isoquinoline with bromoisoquinoline and various phenacyl bromides to obtain the corresponding tetrahydropyrazino[2,1-a:5,4-a′]-diisoquinoline derivatives (**3p−3v**) in good to moderate yields. The substrate without a substituent or with a phenyl group on the benzene ring gave the best yield (94%).

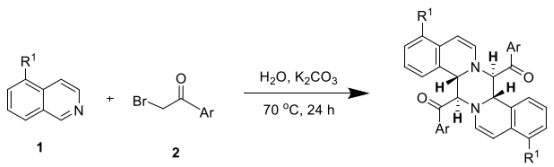
3.2. Antiviral Activities of Some Synthesized Compounds against SARS-CoV-2. The synthesized compounds were screened for inhibiting infection of delta SARS-CoV-2 on VeroE6 host cells. Due to their weak anti-SARS-CoV-2 activities, EC_{50} measurements of some compounds could not be achieved. However, we found that **50** and **96** were repeatedly active under the entry treatment (preincubating with the virus). Their dose-dependent inhibition profiles shown in Figure 3 gave EC_{50} of 26.5 ± 6.9 and 17.0 ± 3.7 μM, respectively, without cytotoxicity (CC_{50} > 100 μM). The compounds were not active when added to the veroE6 cells after the cells were infected with SARS-CoV-2 (the postentry treatment) so that the compounds should inhibit the target(s) outside the cells for virus entry.

3.3. Evaluation of the Active Antivirals against RBD:ACE2 Involved in Virus Entry. To understand the target(s) of the active antivirals, assays of the RBD:ACE2 interaction were carried out. The active compounds **50** and **96** displayed IC_{50} values of 32.7 ± 8.9 and 10.4 ± 2.1 μM, respectively, based on the dose-dependent inhibition profiles shown in Figure 4. These two antivirals showed no inhibition of the human proteases Furin, TMPRSS2, and Cathepsin L, which are also involved in virus entry (data not shown).

3.4. Binding Mode of the Inhibitor with RBD. To elucidate the inhibition mechanism of compound **50** in virus entry, a docking study was conducted. **50** was docked into RBD of the SARS-CoV-2 delta variant Spike protein (PDB: 7w92), as shown in Figure 5. The binding energy of compound **50** with the RBD of the delta-strain Spike protein was estimated to be −93.3 kcal/mol. The analysis revealed that compound **50** forms van der Waals interactions with Tyr448, Tyr453, Gln493, and Tyr505 of the RBD. In the delta RBD:ACE2 complex, Tyr453, Gln493, and Tyr505 of RBD form hydrogen bonds with His34, Lys31, and Glu37 of ACE2, respectively. Additionally, an oxygen atom in one of the acetophenone groups forms a hydrogen bond with Gly496 of the RBD, which also forms a hydrogen bond with Lys353 of ACE2 in the delta RBD:ACE2 complex. Consequently, through interactions with these pivotal residues on the RBD, compound **50** potentially disrupts the RBD:ACE2.

3.5. Drug-likeness of the Inhibitor as Judged from Lipinski Rule of Five and ADMET Properties. According to the Lipinski rule of five, compound **50** fits several key criteria: its log *P* value is 5, the number of hydrogen bond acceptors is 4, fewer than 10, and the number of hydrogen bond donors is 0, fewer than 5. The molecular weight of compound **50** is 494 Da, which is smaller than the ideal limit of 500 Da. However, its molar refractivity is 146.607559, which is relatively surpassing the standard (between 40 and 130). These parameters suggested that compound **50** could be a promising candidate for drug development.

Scheme 1. Synthesis of Tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline Derivatives from Isoquinoline or Bromoisoquinoline and Various Phenacyl Bromides under the Optimized Condition



R^1 = H (**3a-3o**) or Br (**3p-3v**)

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

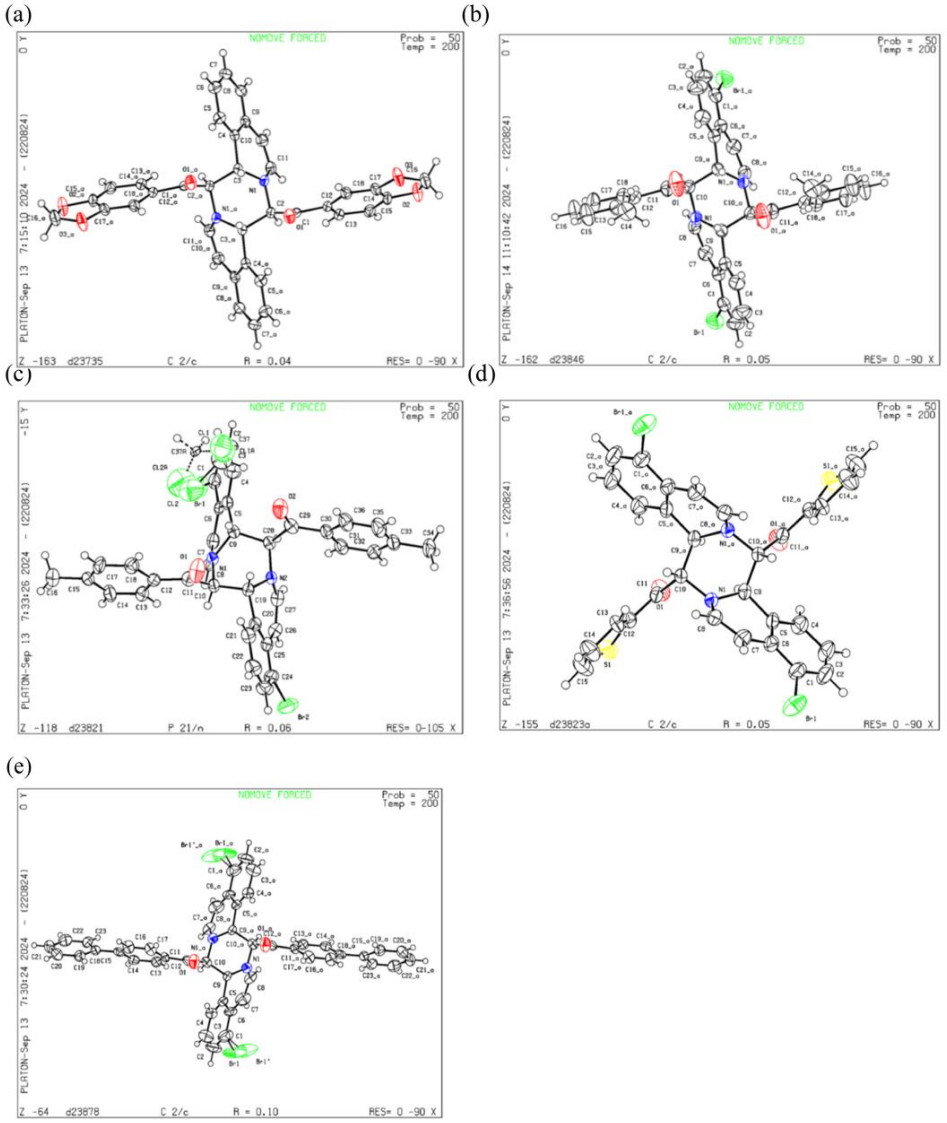


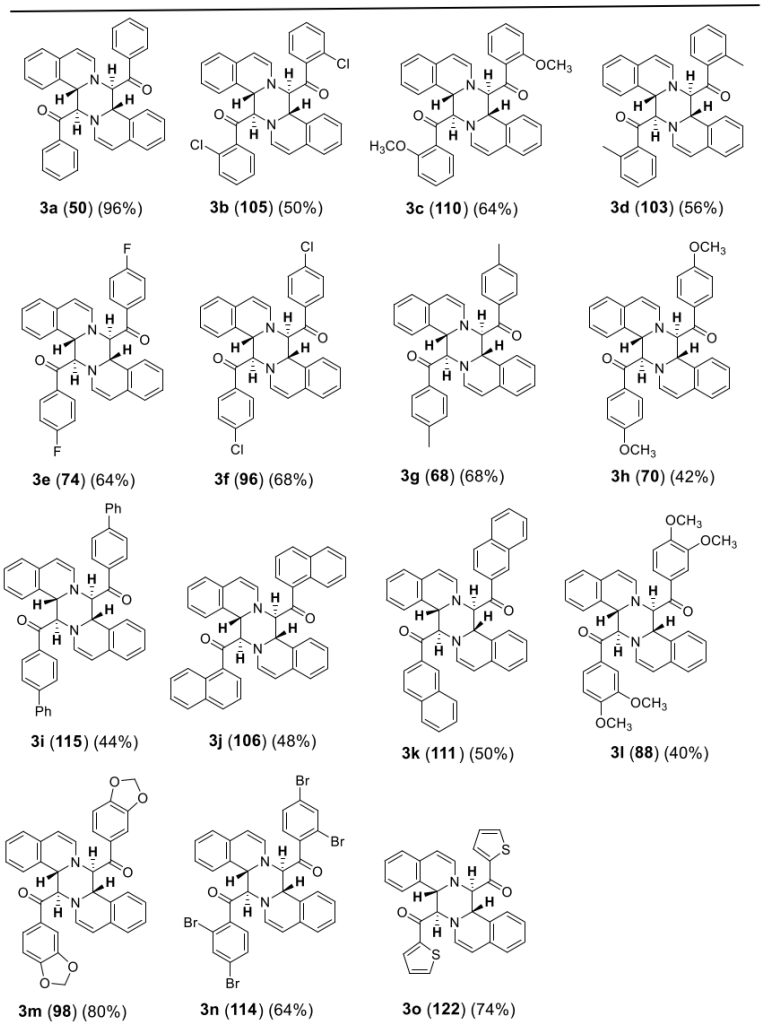
Figure 2. 3-D structures of (A) **98**, (B) **103Br**, (C) **68Br**, (D) **122Br**, and (E) **115Br** determined from the diffraction data are summarized in Tables S2−S6. Their chemical structures are shown in Tables 2 and 3.

Moreover, the ADMET properties predicted *in silico* provided valuable insights into the potential of compounds for therapeutic use. These properties encompass absorption, distribution, metabolism, excretion, and toxicity, as summarized in Table S1 for compound **50**. For absorption, the Caco-2 permeability and human intestinal absorption (HIA) scores for compound **50** are relatively high, suggesting effective absorption in the human intestine. Regarding distribution, the ability of compound **50** to cross the blood-brain barrier (BBB) and penetrate the central nervous system (CNS), was

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

Table 2. Structures and Yields of Tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline Derivatives Synthesized from Isoquinoline and Various Phenacyl Bromides[a,b]

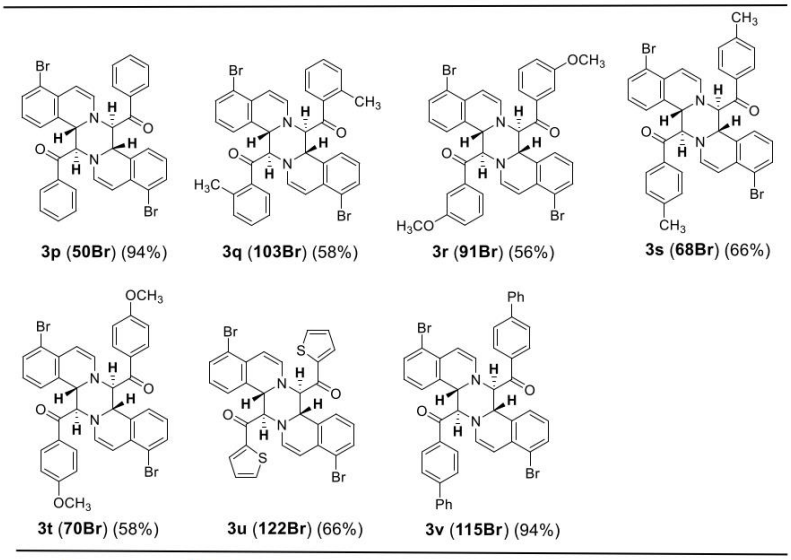


[a]Reactions performed in 1 mmol scale. [b]Isolated yields.

supported by the predicted permeability values. Additionally, the volume of distribution at steady state (VDss) indicated a higher likelihood of tissue distribution. Metabolism prediction revealed that compound **50** may act as a substrate for CYP2D6 or CYP3A4. However, it was not predicted to inhibit CYP1A2, CYP2C19, CYP2C9, CYP2D6, or CYP3A4. In terms of excretion, compound **50** was unlikely to be a substrate for renal organic cation transporter-2 (OCT2), suggesting minimal risk of contraindications related to renal transport. Toxicity predictions indicated that compound **50** is not mutagenic, does not inhibit the hERG I channel, and lacks a skin sensitization potential. Furthermore, the predicted maximum recommended tolerated dose (MRTD) of compound **50** exceeds the standard values, suggesting a favorable safety

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

Table 3. Structures and Yields of Tetrahydropyrazino[2,1-a:5,4-a′]dibromoisoquinoline Derivatives Synthesized from Bromoisoquinoline and Various Phenacyl Bromides[a,b]



[a]Reactions performed in 1 mmol scale. [b]Isolated yields.

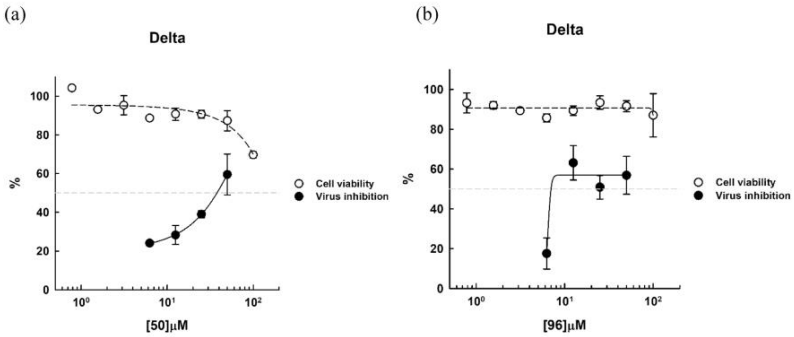


Figure 3. EC_{50} measurements of the antiviral activities of active compounds. (A, B) Inhibitor dose-dependent virus inhibition curves for measurements of EC_{50} values of the inhibitors, **50** and **96**, against the delta variant of SARS-CoV-2 infecting VeroE6 cells, are 26.5 ± 6.9 and 17.0 ± 3.7 μM, respectively. These curves were generated according to the plaque reduction assay data. Their CC_{50} values derived from the plots were >100 μM. All of the measurements were performed in triplicate to yield the averaged EC_{50} and standard deviations.

profile with no unacceptable side effects. Nonetheless, there was a potential concern for hepatotoxicity associated with compound **50**.

4. DISCUSSION

Quinoline is an important building block found in biologically active natural products, pharmaceuticals, and materials science; hence, considerable efforts have been made to develop environmentally benign synthetic protocols, workup, and purification procedures for quinoline derivatives. The synthesis of tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline first appeared in early 1969 when authors observed this compound as a side product.[50] Later, Kutsuma et al. synthesized tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline by the treatment of isoquinoline and phenacyl bromide in the presence of trimethylamine, but only a single compound was synthesized.[51] Recently, Cui et al. have reported the synthesis of the hydrogenated pyrazino[2,1-*a*]isoquinoline derivatives

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

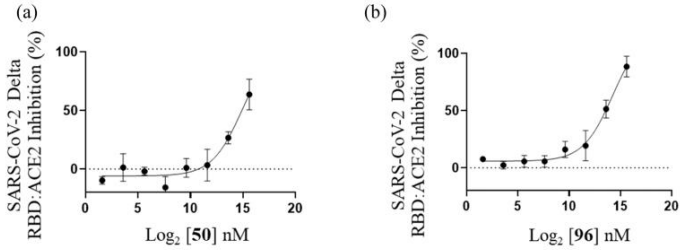


Figure 4. Inhibition of RBD:ACE2 by the compounds. (A, B) Percentages of RBD:ACE2 binding inhibition by increasing compound concentrations were measured with RBD from delta SARS-CoV-2 to yield IC_{50} values of 32.7 ± 8.9 and 10.4 ± 2.1 μM for compounds **50** and **96**, respectively. All of the measurements were performed in triplicate to yield the averaged IC_{50} values and standard deviations.

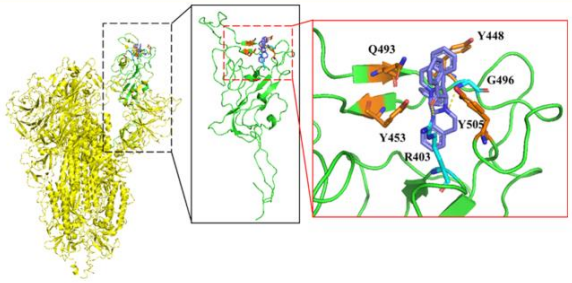


Figure 5. Binding mode of active antiviral compound **50** with RBD. Compound **50** was docked into the RBD of the delta-strain Spike protein (PDB: 7w92). Compound **50** is colored purple, and the residues involved in hydrogen bonding are highlighted in cyan, while those involved in van der Waals interactions are in orange. The trimeric Spike protein is depicted in yellow and the RBD is in green. In this model, compound **50** interacts with residues at the RBD:ACE2 interface, potentially interfering with the binding of the RBD to ACE2.

from dihydroisoquinolines and various phenacyl bromides in the presence of base in acetonitrile solvent.[52] To the best of our knowledge, there is no green methodology developed for accessing these derivatives. Further, the bioactivities of these derivatives have not been investigated. Hence, we developed a green synthetic protocol for tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline derivatives from isoquinoline and phenacyl bromides in water, as demonstrated here. Water is an inexpensive, naturally abundant, safe, nontoxic, and environmentally benign solvent; hence, the development of novel organic reactions in aqueous media has attracted tremendous attention from the chemical community.[42] In addition, since many organic compounds are not soluble in water, the resulting products can be precipitated out from the aqueous reaction medium after the reactions. Consequently, the products can be easily isolated by simple filtration without tedious workup and purification procedures. Our initial studies focused on optimizing the reaction conditions to obtain better yields of the target molecules. In this regard, we chose isoquinoline and phenacyl bromides as model substrates and conducted the reactions at various temperatures in water. After obtaining the optimized conditions, we synthesized a variety of tetrahydropyrazino[2,1-a:5,4-a′]diisoquinoline derivatives from various substituted phenacyl bromides and isoquinoline substrates. Next, we explored the methodology by replacing isoquinoline with bromoisoquinoline and various phenacyl bromides to obtain the corresponding derivatives. In all reactions, the *trans*-diastereomer was obtained as a major product, and the other isomer was formed in trace amounts. However, a pure *trans*-isomer was isolated by column chromatography. This protocol has several additional advantages such as simple operation, broad substrate scope, good functional group tolerance, and easy product isolation.

The synthesized compounds were screened against SARS-CoV-2 targeting the RBD:ACE2 interaction as hinted from modeling. Our results indicate that the compounds **50** and **96** inhibiting RBD:ACE2 but not the human proteases showed anti-SARS-CoV-2 activities under the entry treatment (the compound was preincubated with the virus and included during virus infecting the host to avoid virus entry), but not under the postentry treatment (the compound was added after the host was infected with the virus to prevent viral replication). Therefore, the active compounds exert their antiviral activities by targeting RBD:ACE2 to block virus entry. As compared to berbamine hydrochloride, a cyclic bis-benzylisoquinoline alkaloid that could bind to the postfusion core of SARS-CoV-2 S2 subunit, thereby inhibiting SRAS-CoV-2 entry into VeroE6 host cells with an EC_{50} of 1.73 μM and a CC_{50} of 66.88 μM, respectively,[40] our inhibitors **50** and **96** display lower efficacy (EC_{50} = 26.5 and 17.0 μM), but higher safety (CC_{50} > 100 μM). As isoquinoline and dihydroisoquinoline derivatives may have different reactivities and bioactivities, dihydroisoquinolines can be considered in future studies for the synthesis of natural product-like compounds, and the synthesized molecules could be further tested for drug discovery. Several reports on the synthesis of

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

doi:10.6342/NTU202502385

related hydrogenated products can be considered for future study.[53−55]

To understand how the compounds bind the RBD:ACE2 interface, computational modeling that plays a pivotal role in investigating the interactions between potential drug candidates and target proteins was utilized. Understanding these interactions is crucial, as it expedites the drug discovery process by identifying promising candidates with higher efficacy and lower side effects. Using compound **50** as an example, Figure 5 reveals the interactions between the compound with Spike's RBD of SARS-CoV-2 delta strain. Based on X-ray analysis, all our synthesized compounds have a shape with a central tetrahydropyrazino ring and 4 extended edges containing different groups. The extended shape helps to block the interface of RBD and ACE2 as modeled.

Computational drug discovery, through methods such as molecular docking, is vital for investigating how molecules interact with protein targets. In this study, we employ *i*GEMDOCK software to facilitate steps such as design of inhibitors and postscreening analysis. *i*GEMDOCK is especially useful for postscreening analysis and inferring pharmacological interactions from the active inhibitors.[44] Moreover, by examining the computation-based study of drug-likeliness and pharmacokinetics, compound **50** obeys the Lipinski rule of five and shows favorable ADMET profiles. However, further in vitro and in vivo studies are needed to validate its effectiveness and safety.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c08640.

NMR spectra and X-ray statistics of the compounds as well as the predicted properties of **50** (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Ching-Fa Yao − *Department of Chemistry, National Taiwan Normal University, Taipei 11677, Taiwan;* orcid.org/0000-0002-8692-5156; Email: cheyaocf@ntnu.edu.tw

Po-Huang Liang − *Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan; Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan; Taiwan International Graduate Program, Academia Sinica, Taipei 11529, Taiwan;* orcid.org/0000-0003-1207-5256; Phone: +886-2-3366-4069; Email: phliang@gate.sinica.edu.tw; Fax: +886-2-2363-5038

Authors

Sowndarya Palla − *Department of Chemistry, National Taiwan Normal University, Taipei 11677, Taiwan*

Srinivasa Rao Palla − *Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan; Taiwan International Graduate Program, Academia Sinica, Taipei 11529, Taiwan*

Jia-Jin Liu − *Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan*

Tai-Ling Chao − *Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University, Taipei 10048, Taiwan*

Ting-Hui Lee − *Department of Life Science, National Tsing Hua University, Hsinchu 30013, Taiwan*

Veerababurao Kavala − *Department of Chemistry, National Taiwan Normal University, Taipei 11677, Taiwan*

I-Chen Liu − *Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan*

Lily Hui-Ching Wang − *Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu 30013, Taiwan*

Sui-Yuan Chang − *Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University, Taipei 10048, Taiwan; Department of Laboratory Medicine, National Taiwan University Hospital, Taipei 10002, Taiwan;* orcid.org/0000-0002-6009-6724

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c08640

Author Contributions

S.P. and S.R.P. synthesized all of the compounds, J.-J.L. performed computer modeling and ADMET analysis, T.-L.C. did virus plaque reduction assay to measure the EC_{50} and CC_{50}, T.-H.L. assayed IC_{50} of inhibiting RBD:ACE2 interaction, L.H.-C.W. supervised RBD:ACE2 interaction assay, S.-Y.C. supervised antiviral assay, C.-F.Y. designed and supervised the synthesis and characterization of the compounds, and P.-H.L. designed and supervised the biochemical experiments, analyzed the data, and wrote the manuscript.

Funding

This work was supported by grants (NSTC 113-2320-B-001-004 and 113-2113-M-003-004) from the National Science and Technology Council, Taiwan. The authors also thank the support from Academia Sinica.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the service provided by the Biosafety Level-3 Laboratory of the First Core Laboratory, National Taiwan University College of Medicine.

■ ABBREVIATIONS

COVID-19:Coronavirus Disease 2019; SARS-CoV-2:Severe Acute Respiratory Syndrome Coronavirus-2; ACE2:angiotensin-converting enzyme 2; RBD:receptor binding domain; TMPRSS2:transmembrane protease serine 2; $3CL^{pro}$:3C-like protease; PL^{pro}:papain-like protease; NSP:nonstructural protein; RdRp:RNA-dependent RNA polymerase; TLC:thin-layer chromatography; PBS:phosphate-buffered saline; DMEM:Dulbecco's modified Eagle's medium; FBS:fetal bovine serum

■ REFERENCES

(1) Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395* (10223), 497−506.

(2) Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; Yuan, M. L.; Zhang, Y. L.; Dai, F. H.; Liu, Y.; Wang, Q. M.; Zheng, J. J.; Xu, L.; Holmes, E. C.; Zhang, Y. Z. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579* (7798), 265−269.

(3) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. China novel coronavirus

investigating and research team. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382* (8), 727−733.
(4) Zhou, P.; Yang, X. L.; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. D.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270−273.
(5) Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; Bi, Y.; Ma, X.; Zhan, F.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z.; Zhou, W.; Zhao, L.; Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J.; Xie, Z.; Ma, J.; Liu, W. J.; Wang, D.; Xu, W.; Holmes, E. C.; Gao, G. F.; Wu, G.; Chen, W.; Shi, W.; Tan, W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **2020**, *395* (10224), 565−574.
(6) Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **2022**, *23* (1), 3−20.
(7) Bestle, D.; Heindl, M. R.; Limburg, H.; Van Lam van, T.; Pilgram, O.; Moulton, H.; Stein, D. A.; Hardes, K.; Eickmann, M.; Dolnik, O.; Rohde, C.; Klenk, H. D.; Garten, W.; Steinmetzer, T.; Böttcher-Friebertshäuser, E. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci. Alliance* **2020**, *3* (9), No. e202000786.
(8) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Kruger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N. H.; Nitsche, A.; Muller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **2020**, *181* (2), 271−280.
(9) Yamamoto, M.; Kiso, M.; Sakai-Tagawa, Y.; Iwatsuki-Horimoto, K.; Imai, M.; Takeda, M.; Kinoshita, N.; Ohmagari, N.; Gohda, J.; Semba, K.; Matsuda, Z.; Kawaguchi, Y.; Kawaoka, Y.; Inoue, J. I. The anticoagulant nafamostat potently inhibits SARS-CoV-2 S protein-mediated fusion in a cell fusion assay system and viral infection in vitro in a cell-type-dependent manner. *Viruses* **2020**, *12* (6), 629.
(10) Zhao, M. M.; Yang, W. L.; Yang, F. Y.; Zhang, L.; Huang, W. J.; Hou, W.; Fan, C. F.; Jin, R. H.; Feng, Y. M.; Wang, Y. C.; Yang, J. K. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduction Targeted Ther.* **2021**, *6* (1), No. 134.
(11) Hsu, M. F.; Kuo, C. J.; Chang, K. T.; Chang, H. C.; Chou, C. C.; Ko, T. P.; Shr, H. L.; Chang, G. G.; Wang, A. H.; Liang, P. H. Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **2005**, *280* (35), 31257−31266.
(12) Kuo, C. J.; Liang, P. H. Characterization and Inhibition of the Main Protease of Severe Acute Respiratory Syndrome Coronavirus. *ChemBioEng Rev.* **2015**, *2* (2), 118−132.
(13) Kuo, C. J.; Liang, P. H. SARS-CoV-2 3CLpro displays faster self-maturation in vitro than SARS-CoV 3CLpro due to faster C-terminal cleavage. *FEBS Lett.* **2022**, *596* (9), 1214−1224.
(14) de Wit, E.; van Doremalen, N.; Falzarano, D.; Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **2016**, *14* (8), 523−534.
(15) Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; Dantonio, A.; Di, L.; Eng, H.; Ferre, R.; Gajiwala, K. S.; Gibson, S. A.; Greasley, S. E.; Hurst, B. L.; Kadar, E. P.; Kalgutkar, A. S.; Lee, J. C.; Lee, J.; Liu, W.; Mason, S. W.; Noell, S.; Novak, J. J.; Obach, R. S.; Ogilvie, K.; Patel, N. C.; Pettersson, M.; Rai, D. K.; Reese, M. R.; Sammons, M. F.; Sathish, J. G.; Singh, R. S. P.; Steppan, C. M.; Stewart, A. E.; Tuttle, J. B.; Updyke, L.; Verhoest, P. R.; Wei, L.; Yang, Q.; Zhu, Y. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021**, *374* (6575), 1586−1593.
(16) Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.; Yamamoto, S.; Maruyama, Y.; Taoda, Y.; Kasamatsu, K.; Suto, T.; Kouki, K.; Nakahashi, A.; Kawashima, S.; Sanaki, T.; Toba, S.; Uemura, K.; Mizutare, T.; Ando, S.; Sasaki, M.; Orba, Y.; Sawa, H.; Sato, A.; Sato, T.; Kato, T.; Tachibana, Y. Discovery of S-217622, a noncovalent oral SARS-CoV-2 3CL protease inhibitor clinical candidate for treating COVID-19. *J. Med. Chem.* **2022**, *65* (9), 6499−6512.
(17) Wang, B.; Li, H. J.; Cai, M. M.; Lin, Z. X.; Ou, X. F.; Wu, S. H.; Cai, R. H.; Wei, Y. N.; Yang, F.; Zhu, Y. M.; Yang, Z. F.; Zhong, N. S.; Lin, L. Antiviral efficacy of RAY1216 monotherapy and combination therapy with ritonavir in patients with COVID-19: A phase 2, single centre, randomised, double-blind, placebo-controlled trial. *eClinicalMedicine* **2023**, *63*, No. 102189.
(18) Jiang, X.; Su, H.; Shang, W.; Zhou, F.; Zhang, Y.; Zhao, W.; Zhang, Q.; Xie, H.; Jiang, L.; Nie, T.; Yang, F.; Xiong, M.; Huang, X.; Li, M.; Chen, P.; Peng, S.; Xiao, G.; Jiang, H.; Tang, R.; Zhang, L.; Shen, J.; Xu, Y. Structure-based development and preclinical evaluation of the SARS-CoV-2 3C-like protease inhibitor simnotrelvir. *Nat. Commun.* **2023**, *14* (1), No. 6463.
(19) Cao, B.; Wang, Y.; Lu, H.; Huang, C.; Yang, Y.; Shang, L.; Chen, Z.; Jiang, R.; Liu, Y.; Lin, L.; Peng, P.; Wang, F.; Gong, F.; Hu, H.; Cheng, C.; Yao, X.; Ye, X.; Zhou, H.; Shen, Y.; Liu, C.; Wang, C.; Yi, Z.; Hu, B.; Xu, J.; Gu, X.; Shen, J.; Xu, Y.; Zhang, L.; Fan, J.; Tang, R.; Wang, C. Oral simnotrelvir for adult patients with mild-to-moderate Covid-19. *N. Engl. J. Med.* **2024**, *390* (3), 230−241.
(20) Gordon, C. J.; Tchesnokov, E. P.; Woolner, E.; Perry, J. K.; Feng, J. Y.; Porter, D. P.; Götte, M. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J. Biol. Chem.* **2020**, *295* (20), 6785−6797.
(21) 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* (3), 269−271.
(22) Beigel, J. H.; Tomashek, K. M.; Dodd, L. E.; Mehta, A. K.; Zingman, B. S.; Kalil, A. C.; Hohmann, E.; Chu, H. Y.; Luetkemeyer, A.; Kline, S.; Lopez de Castilla, D.; Finberg, R. W.; Dierberg, K.; Tapson, V.; Hsieh, L.; Patterson, T. F.; Paredes, R.; Sweeney, D. A.; Short, W. R.; Touloumi, G.; Lye, D. C.; Ohmagari, N.; Oh, M. D.; Ruiz-Palacios, G. M.; Benfield, T.; Fätkenheuer, G.; Kortepeter, M. G.; Atmar, R. L.; Creech, C. B.; Lundgren, J.; Babiker, A. G.; Pett, S.; Neaton, J. D.; Burgess, T. H.; Bonnett, T.; Green, M.; Makowski, M.; Osinusi, A.; Nayak, S.; Lane, H. C.; ACTT-1 Study Group Members. Remdesivir for the treatment of COVID-19−final report. *N. Engl. J. Med.* **2020**, *383* (19), 1813−1826.
(23) Gottlieb, R. L.; Vaca, C. E.; Paredes, R.; Mera, J.; Webb, B. J.; Perez, G.; Oguchi, G.; Ryan, P.; Nielsen, B. U.; Brown, M.; Hidalgo, A.; Sachdeva, Y.; Mittal, S.; Osiyemi, O.; Skarbinski, J.; Juneja, K.; Hyland, R. H.; Osinusi, A.; Chen, S.; Camus, G.; Abdelghany, M.; Davies, S.; Behenna-Renton, N.; Duff, F.; Marty, F. M.; Katz, M. J.; Ginde, A. A.; Brown, S. M.; Schiffer, J. T.; Hill, J. A.; GS-US-540−9012 (PINETREE) Investigators. Early remdesivir to prevent progression to severe COVID-19 in outpatients. *N. Engl. J. Med.* **2022**, *386* (4), 305−315.
(24) Sheahan, T. P.; Sims, A. C.; Zhou, S.; Graham, R. L.; Pruijssers, A. J.; Agostini, M. L.; Leist, S. R.; Schäfer, A.; Dinnon, K. H., 3rd; Stevens, L. J.; Chappell, J. D.; Lu, X.; Hughes, T. M.; George, A. S.; Hill, C. S.; Montgomery, S. A.; Brown, A. J.; Bluemling, G. R.; Natchus, M. G.; Saindane, M.; Kolykhalov, A. A.; Painter, G.; Harcourt, J.; Tamin, A.; Thornburg, N. J.; Swanstrom, R.; Denison, M. R.; Baric, R. S. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci. Transl. Med.* **2020**, *12* (541), No. eabb5883.
(25) Jayk Bernal, A.; Gomes da Silva, M. M.; Musungaie, D. B.; Kovalchuk, E.; Gonzalez, A.; Delos Reyes, V.; Martín-Quirós, A.; Caraco, Y.; Williams-Diaz, A.; Brown, M. L.; Du, J.; Pedley, A.; Assaid, C.; Strizki, J.; Grobler, J. A.; Shamsuddin, H. H.; Tipping, R.; Wan, H.; Paschke, A.; Butterton, J. R.; Johnson, M. G.; De Anda, C.; MOVe-OUT Study Group. Molnupiravir for Oral Treatment of

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX

Covid-19 in Nonhospitalized Patients. *N. Engl. J. Med.* **2022**, *386* (6), 509−520.
(26) Barnes, C. O.; Jette, C. A.; Abernathy, M. E.; Dam, K. A.; Esswein, S. R.; Gristick, H. B.; Malyutin, A. G.; Sharaf, N. G.; Huey-Tubman, K. E.; Lee, Y. E.; Robbiani, D. F.; Nussenzweig, M. C.; West, A. P., Jr; Bjorkman, P. J. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **2020**, *588* (7839), 682−687.
(27) Bojadzic, D.; Alcazar, O.; Chen, J.; Chuang, S. T.; Condor Capcha, J. M.; Shehadeh, L. A.; Buchwald, P. Small-molecule inhibitors of the coronavirus spike: ACE2 protein−protein interaction as blockers of viral attachment and entry for SARS-CoV-2. *ACS infect. Dis.* **2021**, *7* (6), 1519−1534.
(28) Yang, L. J.; Chen, R. H.; Hamdoun, S.; Coghi, P.; Ng, J. P. L.; Zhang, D. W.; Guo, X.; Xia, C.; Law, B. Y. K.; Wong, V. K. W. Corilagin prevents SARS-CoV-2 infection by targeting RBD-ACE2 binding. *Phytomedicine* **2021**, *87*, No. 153591.
(29) Omotuyi, O.; Olatunji, O. M.; Nash, O.; Oyinloye, B.; Soremekun, O.; Ijagbuji, A.; Fatumo, S. Benzimidazole compound abrogates SARS-COV-2 receptor-binding domain (RBD)/ACE2 interaction in vitro. *Microb. Pathog.* **2023**, *176*, No. 105994.
(30) Liu, X.; Jiang, L.; Li, L.; Lu, F.; Liu, F. Bionics design of affinity peptide inhibitors for SARS-CoV-2 RBD to block SARS-CoV-2 RBD-ACE2 interactions. *Heliyon* **2023**, *9* (1), No. e12890.
(31) Palla, S. R.; Lee, C. W.; Chao, T. L.; Lo, H. L. V.; Liu, J. J.; Pan, M. Y. C.; Chiu, Y. T.; Lin, W. C.; Hu, C. W.; Yang, C. M.; Chen, Y. Y.; Fang, J. T.; Lin, S. W.; Lin, Y. T.; Lin, H. C.; Kuo, C. J.; Wang, L. H. C.; Chang, S. Y.; Liang, P. H. Synthesis, evaluation, and mechanism of 1-(4-(arylethylenylcarbonyl)phenyl)-4-carboxy-2-pyrrolidinones as potent reversible SARS-CoV-2 entry inhibitors. *Antiviral Res.* **2023**, *219*, No. 105735.
(32) Li, C. W.; Chao, T. L.; Lai, C. L.; Lin, C. C.; Pan, M. Y.; Cheng, C. L.; Kuo, C. J.; Wang, L. H.; Chang, S. Y.; Liang, P. H. Systematic studies on the anti-SARS-CoV-2 mechanisms of tea polyphenol-related natural products. *ACS Omega* **2024**, *9* (22), 23984−23997.
(33) Ip, J. D.; Wing-Ho Chu, A.; Chan, W. M.; Cheuk-Ying Leung, R.; Abdullah, S. M. U.; Sun, Y.; Kai-Wang To, K. Global prevalence of SARS-CoV-2 3CL protease mutations associated with nirmatrelvir or ensitrelvir resistance. *EBioMedicine* **2023**, *91*, No. 104559.
(34) Hu, Y.; Lewandowski, E.; Tan, H.; Zhang, X.; Morgan, R. T.; Zhang, X.; Jacobs, L. M. C.; Butler, S. G.; Gongora, M. V.; Choy, J.; Deng, X.; Chen, Y.; Wang, J. Naturally occurring mutations of SARS-CoV-2 main protease confer drug resistance to nirmatrelvir. *ACS Cent. Sci.* **2023**, *9* (8), 1658−1669.
(35) Gandhi, S.; Klein, J.; Robertson, A. J.; Pena-Hernández, M. A.; Lin, M. J.; Roychoudhury, P.; Lu, P.; Fournier, J.; Ferguson, D.; Bakhash, S. A. K. M.; Muenker, M. C.; Srivathsan, A.; Wunder, E. A., Jr; Kerantzas, N.; Wang, W.; Lindenbach, B.; Pyle, A.; Wilen, C. B.; Ogbuagu, O.; Greninger, A. L.; Iwasaki, A.; Schulz, W. L.; Ko, A. I. De novo emergence of a remdesivir resistance mutation during treatment of persistent SARS-CoV-2 infection in an immunocompromised patient: a case report. *Nat. Commun.* **2022**, *13* (1), No. 1547.
(36) Heyer, A.; Gunther, T.; Robitaille, A.; Lutgehetmann, M.; Addo, M. M.; Jarczak, D.; Kluge, S.; Aepfelbacher, M.; zur Wiesch, J. S.; Fischer, N.; Grundhoff, A. Remdesivir-induced emergence of SARS-CoV2 variants in patients with prolonged infection. *Cell Rep. Med.* **2022**, *3* (9), No. 100735.
(37) Cao, Y.; Yisimayi, A.; Jian, F.; Song, W.; Xiao, T.; Wang, L.; Du, S.; Wang, J.; Li, Q.; Chen, X.; Yu, Y.; Wang, P.; Zhang, Z.; Liu, P.; An, R.; Hao, X.; Wang, Y.; Wang, J.; Feng, R.; Sun, H.; Zhao, L.; Zhang, W.; Zhao, D.; Zheng, J.; Yu, L.; Li, C.; Zhang, N.; Wang, R.; Niu, X.; Yang, S.; Song, X.; Chai, Y.; Hu, Y.; Shi, Y.; Zheng, L.; Li, Z.; Gu, Q.; Shao, F.; Huang, W.; Jin, R.; Shen, Z.; Wang, Y.; Wang, X.; Xiao, J.; Xie, X. S. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* **2022**, *608* (7923), 593−602.
(38) Valipour, M.; Hosseini, A.; Di Sotto, A.; Irannejad, H. Dual action anti-inflammatory/antiviral isoquinoline alkaloids as potent naturally occurring anti-SARS-CoV-2 agents: A combined pharmacological and medicinal chemistry perspective. *Phytother. Res.* **2023**, *37* (5), 2168−2186.
(39) Ren, P.-x.; Shang, W.; Yin, W.; Ge, H.; Wang, L.; Zhang, X.; Li, B.; Li, H.; Xu, Y.; Xu, E. H.; Jiang, H. L.; Zhu, L. L.; Zhang, L. K.; Bai, F. A multi-targeting drug design strategy for identifying potent anti-SARS-CoV-2 inhibitors. *Acta Pharmacol. Sinica* **2022**, *43* (2), 483−493.
(40) Zhang, Z. R.; Zhang, Y. N.; Zhang, H. Q.; Zhang, Q. Y.; Li, N.; Li, Q.; Deng, C. L.; Zhang, B.; Li, X. D.; Ye, H. Q. Berbamine hydrochloride potently inhibits SARS-CoV-2 infection by blocking S protein-mediated membrane fusion. *PLoS Negl. Trop. Dis.* **2022**, *16* (4), No. e0010363.
(41) He, C. L.; Huang, L. Y.; Wang, K.; Gu, C. J.; Hu, J.; Zhang, G. J.; Xu, W.; Xie, Y. H.; Tang, N.; Huang, A. L. Identification of bis-benzylisoquinoline alkaloids as SARS-CoV-2 entry inhibitors from a library of natural products. *Signal Transduction Targeted Ther.* **2021**, *6* (1), No. 131.
(42) Cortes-Clerget, M.; Yu, J.; Kincaid, J. R. A.; Walde, P.; Gallou, F.; Lipshutz, B. H. Water as the reaction medium in organic chemistry: from our worst enemy to our best friend. *Chem. Sci.* **2021**, *12* (12), 4237−4266.
(43) Kuo, C. J.; Chao, T. L.; Kao, H. C.; Tsai, Y. M.; Liu, Y. K.; Wang, L. H.; Hsieh, M. C.; Chang, S. Y.; Liang, P. H. Kinetic characterization and inhibitor screening for the proteases leading to identification of drugs against SARS-CoV-2. *Antimicrob. Agents Chemother.* **2021**, *65* (4), No. e02577-20.
(44) Hsu, K. C.; Chen, Y. F.; Lin, S. R.; Yang, J. M. iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *BMC Bioinf.* **2011**, *12* (Suppl 1), No. S33.
(45) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581* (7807), 215−220.
(46) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, No. 33.
(47) Lipinski, C. A. Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today Technol.* **2004**, *1* (4), 337−341.
(48) Jayaram, B.; Singh, T.; Mukherjee, G.; Mathur, A.; Shekhar, S.; Shekhar, V. Sanjeevini: a freely accessible web-server for target directed lead molecule discovery. *BMC Bioinf.* **2012**, *13* (Suppl 17), No. S7.
(49) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **2015**, *58* (9), 4066−4072.
(50) Ahlbrecht, H.; Fröhlich, J.; Habermalz, U.; Kröhnke, F. Zur struktur der alkaliprodukte aus N-phenacyl-isochinoliniumsalzen. *Tetrahedron Lett.* **1967**, *8* (37), 3649−3652.
(51) Kutsuma, T.; Sekine, Y.; Hujiyama, K.; Kobayashi, Y. Studies on the reactions of heterocyclic compounds. IX. Syntheses of 1, 10b-dihydropyrrolo[2, 1-α]isoquinolines and 2, 3-dihydropyrrolo[2, 1-α]isoquinolines. *Chem. Pharma. Bull.* **1972**, *20* (12), 2701−2706.
(52) Yang, M.-C.; Tang, X.; Liu, S.-W.; Deng, H.-Q.; Lei, J.-J.; Gao, Y.-J.; Han, B.; Cui, H.-L. K_3PO_4 promoted dipolar [3 + 3] cyclization: Direct synthesis of pyrazino[2,1-a]isoquinoline derivatives. *Tetrahedron Lett.* **2018**, *59* (2), 138−142.
(53) Huisgen, R.; Niklas, K. The chemistry of an isolable azomethine ylide. *Heterocycles* **1984**, *22*, 21−26.
(54) Kałuża, Z.; Bielawski, K.; Ćwiek, R.; Niedziejko, P.; Kaliski, P. C2-Symmetric hemiaminal ethers and diamines: new ligands for copper-catalyzed desymmetrization of meso-1,2-diols and asymmetric Henry reactions. *Tetrahedron: Asymmetry* **2013**, *24* (21−22), 1435−1442.
(55) Xia, P.-J.; Sun, Y.-H.; Xiao, J.-A.; Zhou, Z.-F.; Wen, S.-S.; Xiong, Y.; Ou, G.-C.; Chen, X.-Q.; Yang, H. Regioselectivity-tunable self-1,3-dipolar [3 + 3] cyclizations of azomethine ylides to assemble dispirooxindole-piperazines. *J. Org. Chem.* **2015**, *80* (22), 11573−11579.

https://doi.org/10.1021/acsomega.4c08640
ACS Omega XXXX, XXX, XXX−XXX