Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2025, **23**, 9635

Phylogeny-directed discovery and mutagenesis of a tricyclic gersemiane synthase

Xiaochen Chen,†[a] Bao Chen†[b] and Baofu Xu *[a,b,c]

Received 26th July 2025,
Accepted 1st October 2025

DOI: 10.1039/d5ob01208b

rsc.li/obc

Eunicellane diterpenoids, characterized by a 6,10-bicyclic core, are prone to undergo further cyclization into gersemiane diterpenoids *via* acid- or epoxidation-mediated protonation. However, it remains unclear whether classical type I terpene synthases (TSs), phylogenetically related to reported eunicellane synthases, can catalyze this protonation reaction, a process typically driven by type II TSs. To address this question, we employed a phylogeny-directed discovery approach starting from the eunicellane synthase MicA. This led to the discovery of SsFitTS, a tricyclic gersemiane synthase capable of catalyzing the protonation of a natural eunicellane intermediate within its classical type I TS active pocket. Through mutagenesis and molecular docking studies, we elucidated the mechanistic basis for this type I TS's ability to mediate a typically type II-like reaction while retaining its canonical type I fold. Our study not only demonstrated that the phylogeny-based genome mining method can unveil cryptic TSs but also facilitated the decoding of the structural basis underlying the determination of terpene scaffolds.

Introduction

Terpenoids constitute the most prevalent class of natural products exhibiting a wide array of biological activities, ranging from antiviral, antifungal, antibacterial to cytotoxic effects, which endow them with extensive applications across both medical and industrial sectors, such as paclitaxel[1–3] and artemisinin.[4,5] Among the diverse terpenoid skeletons, eunicellane diterpenoids (Fig. 1A), characterized by a 6,10-bicyclic core, have attracted significant interest. Nearly 400 eunicellanes have been reported, predominantly from soft corals (97.7%), with only a few examples from plants (1.3%) and, more recently, bacteria (1.0%).[6] A prominent example is eleutherobin, which exhibits anticancer activity comparable to that of paclitaxel.[7] In contrast, gersemiane diterpenoids, characterized by a 6,6,6-tricyclic scaffold, remain relatively understudied. The gersemiane scaffold is proposed to form *via* C-2/C-7 carbon–carbon bond formation in a eunicellane precursor.[8] To date, less than ten such compounds have been reported from soft corals and liverworts, including klysimplexin T, which exhibits cytotoxicity against a narrow range of cancer cell lines (Fig. 1A). The biosynthesis of eunicellanes has been extensively studied, beginning with the landmark characterization of the first eunicellane TS Bnd4,[9] followed by *Ec*TPS1,[10] *Ba*TC-2,[11] AlbS,[12] MicA,[6] and CpDTS2.[13] In contrast, research on gersemiane biosynthesis has progressed gradually, with the discovery of the first gersemiane TS, PeuTPS, through structural-model-based genome mining.

In the biosynthesis of terpenoids, TSs, which are classified into type I and type II, are pivotal in generating the diversity of terpenoid structures. The catalytic domain of type I TSs, referred to as the α-domain, harbors the characteristic DDxxD and NSE/DTE motifs, which are essential for metal binding and catalysis. In comparison, class II TSs typically adopt a two-domain (βγ) architecture, where the substrate is positioned within the cleft formed by the β and γ domains. The β-domain contains the catalytically critical DxDD motif, while the γ-domain contributes to structural stability and substrate orientation.[14] TSs facilitate carbocation formation *via* two main pathways. Type I TSs initiate carbocation generation by cleaving the pyrophosphate group, while Type II TSs generate the initial carbocation by protonating the olefin or epoxy group within the substrate.[14] Upon carbocation formation, a sequential cascade of reactions unfolds, encompassing olefin-to-cation cyclization, hydride migrations, proton transfers, alkyl shifts, and terminal deprotonation or quenching events, culminating in the generation of a diverse array of terpene scaffolds. In fact, terpene skeletons are frequently biosynthesized through the synergistic action of type I and type II TSs (Fig. 1B). For example, in hopene biosynthesis, two molecules of farnesyl diphosphate (FPP) are first condensed into squalene by a type I TS and then further catalyzed by a type II TS

[a]*Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan 250117, China*
[b]*Shandong Laboratory of Yantai Drug Discovery, Bohai Rim Advanced Research Institute for Drug Discovery, Yantai, Shandong 264117, China*
[c]*Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China. E-mail: bfxu@simm.ac.cn*
†These authors contributed equally to this work.

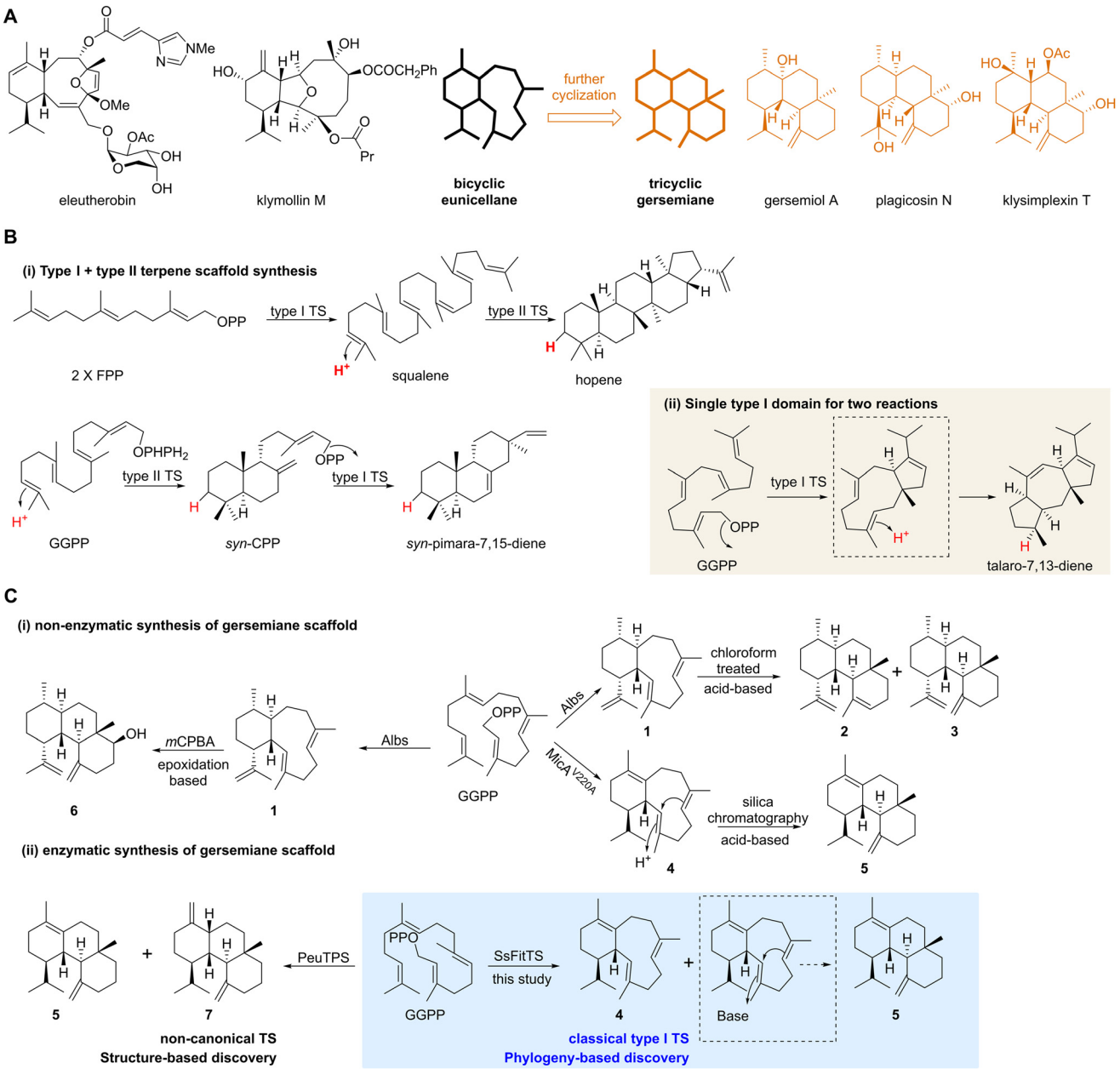

Fig. 1 Diterpenoids with eunicellane or gersemiane scaffolds. (A) Natural examples of eunicellane and gersemiane compounds; (B) terpene scaffold biosynthesis involving both type I and II carbon cation generation reactions; (C) enzymatic and non-enzymatic synthesis of the gersemiane scaffold.

(Fig. 1B).[14] In addition, *syn*-pimara-7,15-diene is produced when geranylgeranyl pyrophosphate (GGPP) is initially cyclized to copalyl diphosphate (CPP) by a type II TS before undergoing additional cyclization by a type I TS (Fig. 1B).[15] Notably, in some cases, type I TSs can simultaneously mediate carbocation-triggered reactions involving both type I and type II TSs. A representative example is talaro-7,13-diene biosynthesis, where GGPP first undergoes pyrophosphate cleavage to generate (3a*S*,5*E*,9*E*,12a*R*)-3,3a,4,7,8,11,12,12a-octahydro-3a,6,10-trimethyl-1-(1-methylethyl)cyclopentacycloundecene, followed by protonation of this neutral intermediate to yield the final product (Fig. 1B).[16] Although some reports describe type I TSs catalyzing double-bond protonation, it remains unclear how to efficiently discover such enzymes and how type I TSs facilitate this reaction.

Interestingly, certain catalytic products of eunicellane synthases have been shown to readily undergo further cyclization into gersemiane diterpenes (Fig. 1C). For example, the product of MicA^{V220A} spontaneously converts into a gersemiane diterpene during silica chromatography,[6] a phenomenon also observed with chloroform-treated albireticulene generated by AlbS.[12] This process resembles the re-protonation-induced carbocation formation catalyzed by the type I TS TadA.[16] Beyond acid-mediated cyclization of eunicellanes,

epoxidation-based oxidation can also drive gersemiane diterpene formation (Fig. 1C). These examples illustrate that the gersemiane scaffold can be easily accessed through non-enzymatic chemical catalysis, suggesting that enzymatic catalysis of this reaction should likewise be feasible as long as TSs can protonate neutral eunicellane intermediates to generate the final gersemiane structure. This hypothesis was recently validated by the discovery of the non-canonical TS PeuTPS by structural-model-based genome mining, which produces gersemiane compounds.

However, prior to our study, both the engineering potential of eunicellane synthases to produce extended cyclization gersemiane products and the inherent capacity of wild-type classical type I TSs to generate such products through re-protonation-induced carbocation formation remained unexplored (Fig. 1C). Furthermore, the structural factors and potential protonation environment governing double-bond protonation by type I TSs remain poorly understood. Herein, we conducted evolutionary-guided mining of NCBI databases to identify gersemiane TSs using our previously reported microeunicellane synthase MicA[6] as a blast query. Our approach successfully identified a classical type I TS that predominantly protonates neutral eunicellane intermediates to form the final gersemiane scaffold. Through extensive phylogeny-guided mutagenesis studies, we have deepened our understanding of the critical amino acid residues and the active site environment responsible for double-bond protonation in classical type I TSs.

Results and discussion

Discovery of SsFitTS that catalyzes further cyclization of the eunicellane skeleton

The sequence of MicA was searched against the NCBI database, resulting in 26 sequences with over 35% identity, including 8 sequences exceeding 83% identity. These 26 sequences were subjected to phylogenetic analysis using the maximum likelihood method (Fig. 2A and B). Among these sequences, SsFitTS was selected for functional characterization due to its naturally occurring V220A mutation, as we previously reported that mutant MicA^{V220A} can facilitate 2*E* double-bond formation that is prone to further cyclize into the gersemiane scaffold under acidic conditions (Fig. 1C),[6] as well as its relatively low sequence identity with other genes in the same clade that naturally possess this V220A mutation (Fig. 2A). The gene was subsequently commercially synthesized in the bacterial expression vector pET28a for activity assessment. Activity evaluations were conducted by co-expressing the constructed vectors with an *in vivo* genetic system that enhances the efficient production of GGPP in *Escherichia coli*. After fermenting the recombinant bacteria, the resulting products were analyzed using HPLC (Fig. 2C). Compounds corresponding to novel peaks observed in the HPLC chromatogram compared to the pET28a empty vector control were purified through large-scale fermentations and their structures were elucidated using GC-MS, 1D, and 2D NMR analyses. Detailed structural elucidation can be found in the SI under "Compound Structure Elucidation". Fortunately, the TS we identified not only generated the same eunicellane derivative (compound **4**) as MicA^{V220A}, but predominantly produced a further cyclized product (compound **5**), representing an advanced cyclization state of the eunicellane skeleton. This compound features a gersemiane skeleton, belonging to a rare chemical family and previously isolated from the Arctic soft coral *Gersemia fruticose*.[17]

Compound **5** features four consecutive chiral centers, and the ^{1}H NMR chemical shifts for the two methyl groups, CH_3-16 (δ_H 0.90) and CH_3-19 (δ_H 0.91), exhibit considerable similarity, presenting a significant challenge in determining its absolute configuration. To address this issue, attempts were made to crystallize compound **5** under various conditions. Unfortunately, the compound's oily nature prevented the formation of crystals. Following our attempt to introduce heteroatoms into compound **5** through chemical reactions to promote crystallization, we successfully oxidized compound **5** with one equivalent of *meta*-chloroperoxybenzoic acid (*m*CPBA) to yield the oxidation product: (10*S*,11*R*)-epoxy-**5** (**5a**) (see the Experimental section and Fig. 1D). However, despite various efforts, no crystals of compound **5a** formed. A pivotal small molecule crystallization breakthrough occurred with the utilization of a highly fluorinated cyclic trinuclear silver Ag(I) pyrazolate complex (Ag_3Pz_3), known for its strong π-acidity, as a crystallizing agent.[18] This approach, developed by the Dan Li group, proved instrumental in structurally elucidating compounds that are challenging to crystallize independently. We readily prepared a eutectic mixture of compound **5a** and Ag_3Pz_3, where two molecules of **5a** were bound to one molecule of Ag_3Pz_3 in a single-site binding mode (Fig. S2). This architecture enabled the determination of the absolute stereochemistry of **5a** by X-ray diffraction analysis, which was established as 1*S*,2*R*,7*S*,10*S*,11*R*,14*R*.

Interestingly, the ^{1}H and ^{13}C NMR spectra of compound **5** were identical to those of the artificial acid-catalyzed products derived from MicA^{V220A} during silica gel chromatography (Fig. 1B).[6] These findings indicate that the catalytic pockets of both proteins likely maintain highly similar structural environments, where specific amino acid substitutions enable MicA^{V220A}'s eunicellane products to undergo additional cyclization into gersemiane scaffolds. The identification of these critical residues provided a foundation for investigating how classical type I TSs facilitate protonation to drive subsequent type II cyclization reactions.

Identification of potential critical amino acid residues mediating further cyclization in eunicellane TS-catalyzed products through evolutionary insights

TSs with over 85% sequence identity to MicA can theoretically generate the same eunicellane compound. Their sequences were aligned to delineate the conserved motifs of eunicellane TSs (Fig. S1) using Mega11 software.[19] This analysis revealed 156 amino acid residues that deviate from the conserved resi-

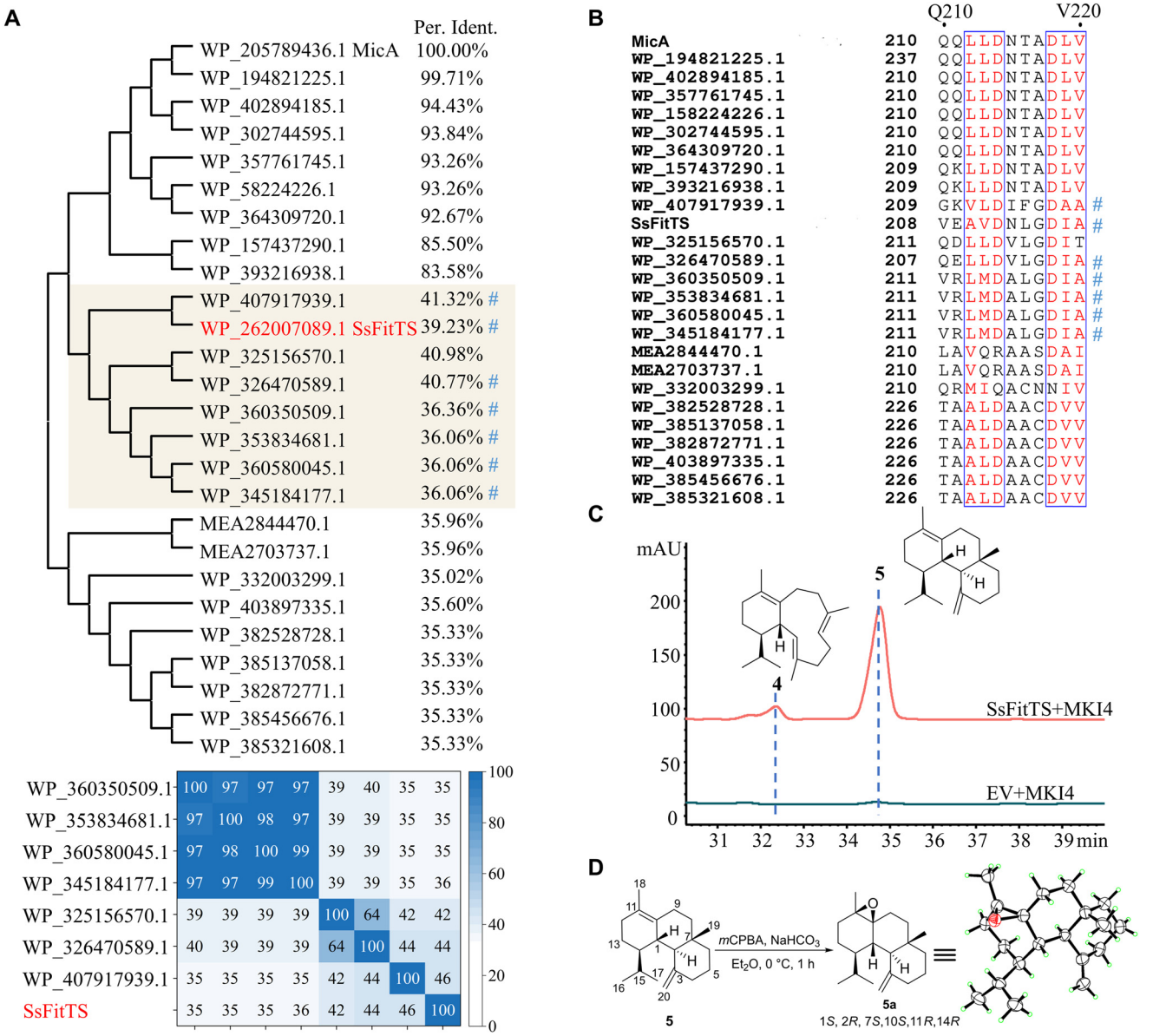

Fig. 2 Mining and functional characterization of gersemiane synthase. (A) Phylogenetic analysis of sequences (>35% identity to MicA) using the maximum likelihood method. The matrix highlights sequences containing the catalytically essential V220 residue (# tag). The TS marked in red was selected for activity assays. (B) Alignment of essential V220 residues of sequences with >35% identity to MicA. (C) HPLC analysis of SsFitTS. "EV" denotes the "empty vector pET28a". (D) Synthesis and structural elucidation of the epoxidized derivative of compound **5**. Epoxide **5a** was generated *via m*CPBA-mediated epoxidation, and its absolute configuration was determined by X-ray crystallography.

dues in eunicellane TSs (Fig. S3). Subsequently, SsFitTS was included in the sequence alignment to identify variations within the conserved amino acid residues of eunicellane TSs (Fig. S4). By comparing these residues with the 23 key residues in MicA (denoted by the "#" symbol, as described in our recent study[6]), we identified six residues (E53, M74, M184, S187, V220, and E232 in MicA) that differ and are likely responsible for promoting further cyclization in eunicellane TS products (Fig. 3A). To test this hypothesis, we performed reciprocal mutations: we replaced these residues in SsFitTS with their counterparts in MicA and *vice versa*, assessing whether these mutations can alter the products. Interestingly, we found that mutants of SsFitTS, including Q51E, I72M, I182M, and A218V, promoted the relatively increased production of bicyclic eunicellane compound **4** over wild-type SsFitTS (Fig. 3B). Further investigations were conducted on double and triple mutants of these four amino acid residues (Fig. S6). Notably, the triple mutant SsFitTS$^{I72M/I182M/A218V}$ produced a higher yield of compound **4** relative to the tricyclic gersemiane compound **5**. These results indicated that these four residues are critical for the re-protonation process in SsFitTS. As for the mutants of MicA, including MicA^{M184I}, MicA^{V220A} and MicA^{E232D}, they

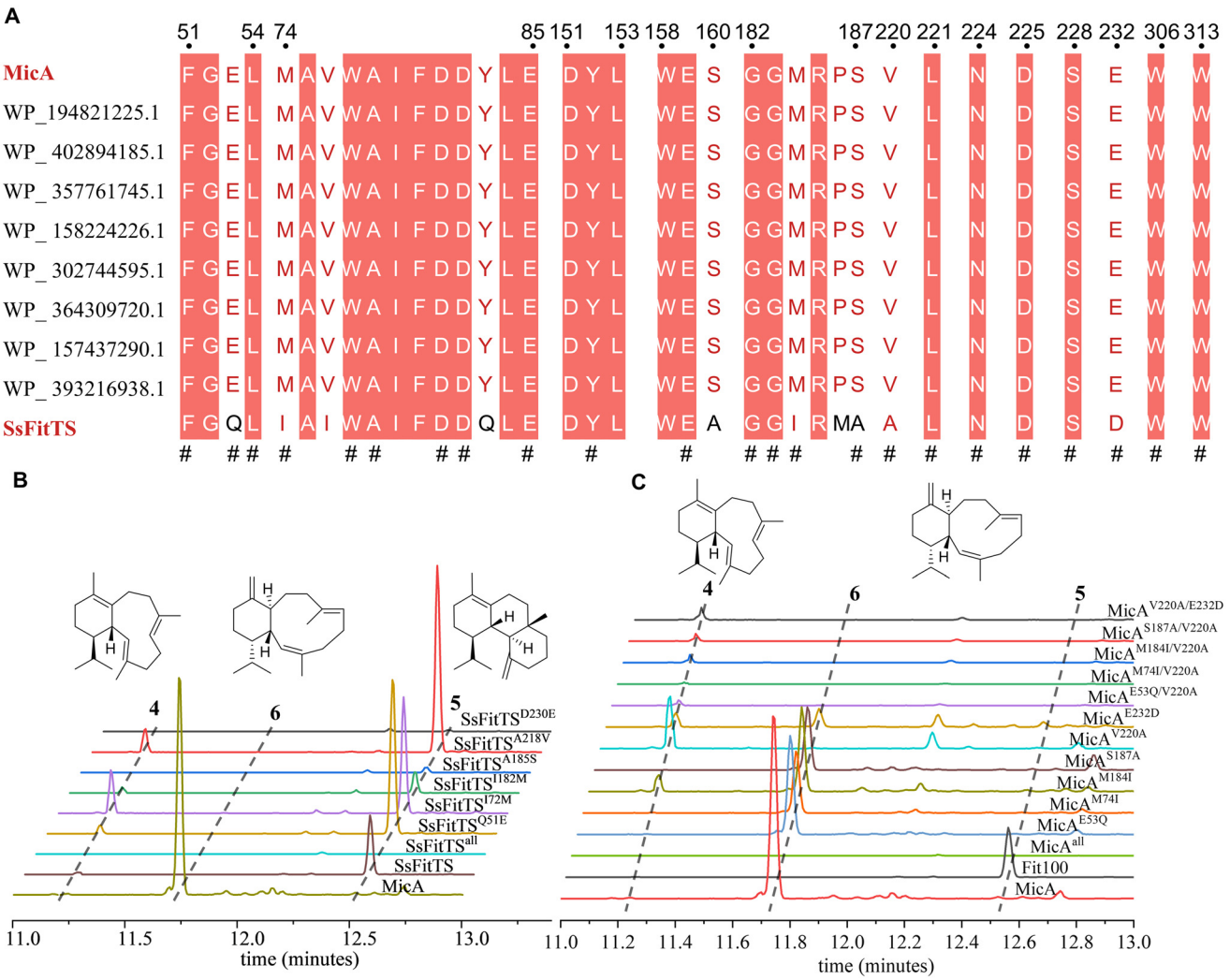

Fig. 3 Evolutionary-guided mutation analysis of SsFitTS and MicA. (A) Sequence alignment of TSs sharing >85% identity with MicA, highlighting conserved eunicellane-TS motifs. Key residues (#) from our previous study[6] were compared with SsFitTS to identify potential determinants of cyclization specificity. (B) GC-MS analysis (EIC, *m/z* 272) of SsFitTS mutants' products. (C) GC-MS chromatography (EIC, *m/z* 272) of MicA mutants' activities.

yielded eunicellane-type compound **4** featuring an *E*-configured C2 double bond, which is essential for promoting the subsequent cyclization of the eunicellane skeleton into the gersemiane backbone. Given that MicA^{V220A} exhibits the highest production yield of compound **4**, we introduced additional mutations at other sites to probe the potential formation of compound **5**. However, these mutations of MicA failed to cause gersemiane scaffold formation (Fig. 3C), demonstrating that targeted residue modifications alone cannot initiate the re-protonation process necessary for generating tricyclic products. Subsequent mutagenesis of all six candidate residues in both SsFitTS and MicA completely abolished activity (Fig. 3), revealing that simple point mutations at key positions are insufficient to confer protonation capability in classical type I TSs. These findings highlight the need for a deeper structural understanding of the protonation mechanism and more sophisticated mutation strategies beyond single-residue alterations.

Structural insights into classical type I TSs driving type II cyclization with a type I-fold

Despite extensive optimization efforts—including truncating the N-terminal region with various lengths and fusing the protein with solubilizing tags (*e.g.*, MBP)—SsFitTS remained insoluble. It was also expressed in multiple competent cell lines, including BL21(DE3), C41(DE3), and Rosetta(DE3), yet remained insoluble. Thus, further efforts to optimize the expression of this enzyme, including strategies such as expressing homologous proteins of SsFitTS or utilizing a *Streptomyces* host, will be necessary to enable its structural characterization. We then performed molecular docking to gain deeper insights into the catalytic mechanism of classical type I TSs driving type II cyclization with a type I-fold. The molecular docking model of SsFitTS with GGPP reveals that while the identified key sites including Q51, I72, I182, and A218 decrease tricyclic gersemiane formation and promote

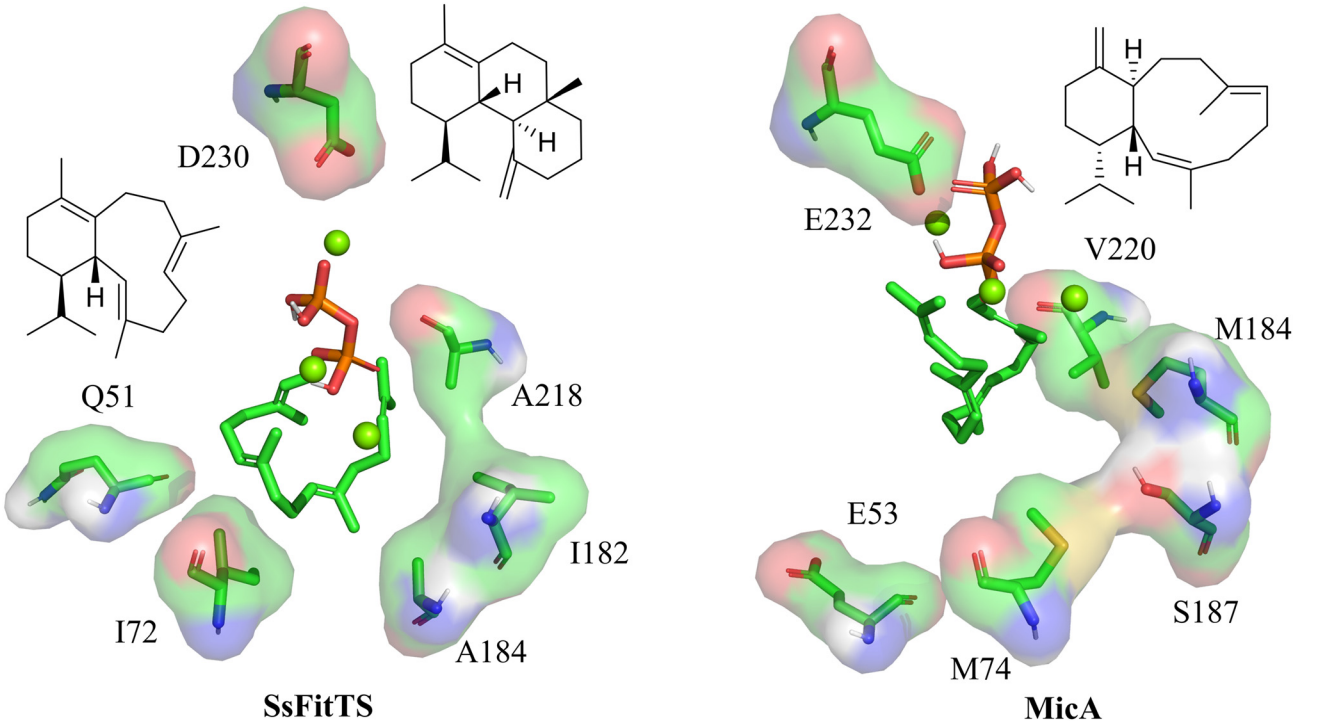

Fig. 4 Structural comparisons of key residues of SsFitTS and MicA in docking models.

bicyclic eunicellane production (Fig. 4), they primarily influence carbocation stability rather than completely disrupting the potential protonation environment within SsFitTS's binding pocket. Consequently, mutants of SsFitTS retain the ability to generate the tricyclic gersemiane scaffold. Regarding eunicellane synthase MicA, mutation studies identified key residues including M184, V220, and E232 near the C2 double bond that facilitate the formation of the 2*E* double bond.

However, these residues alone were insufficient to establish a suitable C2 protonation environment, explaining the absence of tricyclic gersemiane products. Further studies, including substrate–protein co-crystallization and molecular dynamics simulations, are needed so as to gain deeper insight into the mechanism of a type I TS that catalyzes double-bond protonation.

Experimental

Instruments and reagents

HPLC separations were performed using an Agilent 1260 Infinity LC system, with analytical-scale runs conducted on a Zorbax SB-C18 column (4.6 × 150 mm, 5 μm) and preparative purifications carried out using a larger C18 column (21.2 × 250 mm, 7 μm). Structural characterization was achieved through NMR spectroscopy using a Bruker Avance III 600 MHz instrument (Bruker Biospin, Germany). Volatile compounds were analyzed using an Agilent 8890 GC system (Agilent Technologies, Shanghai, China) equipped with a 7650 autosampler (Agilent Technologies, China) and coupled to a 5977C mass spectrometric detector (Agilent Technologies, DE, USA). Separation was achieved using an HP-5 ms capillary column (30 m × 0.25 mm, 0.25 μm film thickness; Agilent Technologies, USA). For compound purification, column chromatography was employed using silica gel (Yantai Xinnuo Chemical Co., China), with analytical-grade solvents from Shanghai Chemical Reagents Co. and chromatographic-grade solvents from Dikma Technologies (USA). Essential molecular biology enzymes were sourced from TransGen Biotech, with key reagents such as IPTG and isoprenol obtained from BBI Life Sciences and Sigma-Aldrich, respectively, and other synthetic chemicals from Shanghai Macklin Biochemical Co.

Mining and functional characterization of terpene synthases

A BLAST search of the eunicellane synthase MicA against the NCBI database identified 26 homologous sequences (>35% identity), with 8 exhibiting >83% identity. Maximum likelihood phylogenetic reconstruction (Fig. 2) demonstrated that seven sequences containing the catalytically essential A220 residue, a residue critical for 2*E* double-bond formation and subsequent cyclization of the eunicellane skeleton into the gersemiane scaffold,[6] formed a distinct clade. Matrix analysis (Fig. 2A) of these seven sequences revealed two phylogenetically divergent candidates (WP_407917939.1 and WP_262007089.1). The latter (designated SsFitTS) was chosen for functional analysis through co-expression with the GGPP-overproduction system in *E. coli*.[6] HPLC profiling of fermentation products (Fig. 2C) detected novel peaks, from which scaled-up cultures yielded metabolites that were purified and

structurally characterized by GC-MS, 1D/2D NMR, and comprehensive spectral analysis (SI).

GC-MS analysis

GC-MS analysis was performed with 1 μL injection (30 : 1 split ratio) at an inlet temperature of 250 °C. Chromatographic separation employed helium carrier gas (1.0 mL min−^{1} constant flow) using a temperature program from 50 °C to 260 °C (10 °C min^{-1} ramp, 21 min total runtime). The mass spectrometer was operated in the EI^{+} mode (70 eV) with source, quadrupole, and transfer line temperatures maintained at 250 °C, 130 °C, and 250 °C, respectively. Full-scan data (50–550 *m*/*z*) were acquired and processed using MassHunter 10.2 (Agilent), with compound identification against the NIST 23 database.

Isolation of compounds 4 and 5

For heterologous expression, recombinant *E. coli* BL21(DE3) harboring both CDF-MKI4 and pET28a-SsFitTS plasmids was cultured in 500 mL of LB medium containing kanamycin and streptomycin (50 mg L^{-1} each). Following overnight growth, 10 mL aliquots were used to inoculate 30 × 1 L fresh LB cultures. Protein expression was induced at OD_{600} = 1.5 with 0.5 mM IPTG and 1.0 mM isoprenol, followed by 18 h of incubation at 28 °C (220 rpm). Cells were harvested by centrifugation (4000 rpm, 15 min), and the pellet was extracted with methanol. The crude extract was partitioned with petroleum ether (PE), concentrated under vacuum, and initially fractionated by silica gel chromatography (200–300 mesh, PE elution) to obtain compounds **4** and **5**. Final purification was achieved by preparative HPLC (Agilent Zorbax SB-C18, 21.2 × 250 mm, 7 μm) using a gradient of 5–95% acetonitrile in water over 60 min (0–3 min, 5% CH_3CN; 3–18 min, 5–95% CH_3CN; 18–60 min, 95% CH_3CN; 20 mL min^{-1} flow rate).

Epoxidation of compound 5

Compound **5** (10.5 mg, 38.5 μmol, 1.0 eq.) was dissolved in anhydrous Et_2O (1.0 mL) and cooled in an ice bath. $NaHCO_3$ (4.2 mg, 50.0 μmol, 1.3 eq.) was added, followed by portionwise addition of *m*CPBA (total 6.6 mg, 38.2 μmol, 1.0 eq. in 2 equal portions at 0.5 h intervals, each dissolved in 0.5 mL of Et_2O). After 1 h of total reaction time, the mixture was extracted with Et_2O (3 × 2 mL). The combined organic layers were concentrated *in vacuo* and purified by silica gel CC (PE/EA 20 : 1) to afford epoxide **5a** (5.6 mg, 50% yield). Structural characterization data are provided in the SI.

X-ray diffraction analysis of 5a

Single-crystal X-ray diffraction data for compound **5a** were collected using a Bruker D8 Venture diffractometer (Cu Kα radiation, λ = 1.54178 Å). The structure was solved by intrinsic phasing using SHELXT and refined by full-matrix least-squares methods with ShelXL.[20–23] Crystals suitable for analysis were obtained by co-crystallization with Ag_3Pz_3 (1 : 1 PE/DCM, 4 °C), yielding colorless plates with a melting point of 136–138 °C. Crystallographic data have been deposited with the Cambridge Crystallographic Data Centre (CCDC 2474762).

$C_{15}H_3Ag_3F_{18}N_6{\cdot}2$ ($C_{20}H_{30}O$), M = 1509.75 g mol^{-1}, triclinic, space group $P1$ (no. 1), a = 12.0210(5) Å, b = 15.3300(6) Å, c = 17.4835(7) Å, α = 84.639(2)°, β = 71.705(2)°, γ = 85.554(2)°, V = 3041.7(2) Å^3, T = 170(2) K, Z = 2, μ(Cu Kα) = 8.561 mm^{-1}, 30 119 reflections measured, 16 776 independent reflections (R_{int} = 0.0703). The final R_1 value was 0.0857 ($I > 2\sigma(I)$). The final w$R(F^2)$ value was 0.2115 ($I > 2\sigma(I)$). The final R_1 value was 0.0932 (all data). The final w$R(F^2)$ value was 0.2266 (all data). The goodness of fit on F^2 was 1.084. Hooft parameter = 0.072 (8).

Site-directed mutation

The pET28a-SsFitTS gene was subjected to site-directed mutagenesis using overlap extension PCR with TransStar FastPfu Fly DNA polymerase. Following gel purification, the PCR products were cloned into a linearized pET28a vector (*BamHI*/*HindIII*-digested) and transformed into *E. coli* Turbo competent cells. Sequence-verified plasmids were subsequently expressed in *E. coli* BL21Gold(DE3), with product analysis performed using GC-MS. Primer sequences are provided in Table S2.

Protein structure prediction and docking analysis

The three-dimensional structure of SsFitTS was predicted using AlphaFold (v2.3.1) with default parameters. Molecular docking simulations were then performed using AutoDock4 (v4.2.6), following established protocols as previously described.[6] Structural visualization and analysis were conducted using PyMOL (v2.5.2).

Conclusions

In conclusion, we have performed a phylogeny-guided strategy for the discovery of SsFitTS, a tricyclic gersemiane synthase that catalyzes the protonation of a eunicellane intermediate within its catalytic pocket. Through mutagenesis and molecular docking studies, we have elucidated the mechanistic basis for this type I TS's ability to mediate a typically type II reaction while maintaining a classical type I fold. However, our findings revealed that precise modulation of double-bond protonation activity in type I TSs cannot be obtained through evolution-guided substitutions of catalytic residues within 5 Å of the GGPP binding pocket in SsFitTS. To fully elucidate the catalytic mechanism, future investigations should focus on determining the apo structure of SsFitTS, obtaining its cocrystal structures with GGPP and Mg^{2+} ions, performing molecular dynamics simulations, and conducting targeted mutagenesis based on these structural insights. It is particularly intriguing to investigate how this classical type I TS mediates olefin protonation to initiate carbocation formation, given its lack of the characteristic Asp-rich motif (DxDD) found in type II TSs, which typically serves as a general acid to protonate the alkene and generate the initial carbocation.[14] Elucidating the auth-

entic proton donor in this system, whether derived from water molecules like MgMS from *Myrothecium gramineum*[24] or specific amino acid residues like Tyr91 in TadA,[16] remains a critical mechanistic question. While this study focused exclusively on residues within the catalytic core, future phylogeny-based analysis of residues within 10 Å of the ligand and surface residues involved in solubility or stability could address both protein insolubility issues and provide a comprehensive understanding of the structural environment required for type I TSs to catalyze alkene protonation. The genome mining strategy implemented here establishes a paradigm for phylogeny-guided discovery of novel TSs. Furthermore, the engineering approaches provided by this study may offer promising solutions to persistent challenges in protein solubility and enable rational engineering of TSs for terpene scaffold expansion.

Author contributions

Investigation: X. C. and B. C.; writing – original draft: X. C., B. C., and B. X.; writing – review & editing: X. C., B. C., and B. X.; and supervision: B. X.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information is available. See DOI: **https://doi.org/10.1039/d5ob01208b**.

CCDC 2474762 (**5a**) contains the supplementary crystallographic data for this paper.[25]

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2022YFC2804100 to B. X.), the Taishan Scholars Program (tsqn20230623 to B. X.), the Shandong Laboratory Program (SYS202205 to B. X.), the Shanghai Pujiang Program (23PJ1415300 to B. X.), the Shandong Provincial Natural Science Foundation (ZR2024QB357 to B. X.), the Key R&D Program of Shandong Province, China (2024CXPT029 to B. X.), and the National Natural Science Foundation of China (22407117 to B. C.).

References

1 A. Ahmed Khalil, A. Rauf, F. A. Alhumaydhi, A. S. M. Aljohani, M. S. Javed, M. A. Khan, I. A. Khan, M. A. El-Esawi, S. Bawazeer, A. Bouyahya, M. Rebezov, M. A. Shariati and M. Thiruvengadam, *Curr. Pharm. Des.*, 2022, **28**, 3363–3373.
2 S. M. Swain, S. F. Honig, M. C. Tefft and L. Walton, *Invest. New Drugs*, 1995, **13**, 217–222.
3 M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon and A. T. McPhail, *J. Am. Chem. Soc.*, 1971, **93**, 2325–2327.
4 S. Krishna, L. Bustamante, R. K. Haynes and H. M. Staines, *Trends Pharmacol. Sci.*, 2008, **29**, 520–527.
5 J. Wang, C. Xu, Y. K. Wong, Y. Li, F. Liao, T. Jiang and Y. Tu, *Engineering*, 2019, **5**, 32–39.
6 J. Li, B. Chen, Z. Fu, J. Mao, L. Liu, X. Chen, M. Zheng, C.-Y. Wang, C. Wang, Y.-W. Guo and B. Xu, *Nat. Commun.*, 2024, **15**, 5940.
7 T. P. Hughes, A. H. Baird, D. R. Bellwood, M. Card, S. R. Connolly, C. Folke, R. Grosberg, O. Hoegh-Guldberg, J. B. C. Jackson, J. Kleypas, J. M. Lough, P. Marshall, M. Nyström, S. R. Palumbi, J. M. Pandolfi, B. Rosen and J. Roughgarden, *Science*, 2003, **301**, 929–933.
8 B.-W. Chen, C.-H. Chao, J.-H. Su, C.-W. Tsai, W.-H. Wang, Z.-H. Wen, C.-Y. Huang, P.-J. Sung, Y.-C. Wu and J.-H. Sheu, *Org. Biomol. Chem.*, 2011, **9**, 834–844.
9 B. Xu, D. J. Tantillo and J. D. Rudolf, *Angew. Chem., Int. Ed.*, 2021, **60**, 23159–23163.
10 P. D. Scesa, Z. Lin and E. W. Schmidt, *Nat. Chem. Biol.*, 2022, **18**, 659–663.
11 I. Burkhardt, T. de Rond, P. Y.-T. Chen and B. S. Moore, *Nat. Chem. Biol.*, 2022, **18**, 664–669.
12 Z. Li, B. Xu, V. Kojasoy, T. Ortega, D. A. Adpressa, W. Ning, X. Wei, J. Liu, D. J. Tantillo, S. Loesgen and J. D. Rudolf, *Chem*, 2023, **9**, 698–708.
13 G. Tabekoueng, H. Li, B. Goldfuss, G. Schnakenburg and J. S. Dickschat, *Angew. Chem., Int. Ed.*, 2024, **63**, e202413860.
14 X. Pan, J. D. Rudolf and L.-B. Dong, *Nat. Prod. Rep.*, 2024, **41**, 402–433.
15 A. Cyr, P. R. Wilderman, M. Determan and R. J. Peters, *J. Am. Chem. Soc.*, 2007, **129**, 6684–6685.
16 J.-H. Huang, J.-M. Lv, L.-Y. Xiao, Q. Xu, F.-L. Lin, G.-Q. Wang, G.-D. Chen, S.-Y. Qin, D. Hu and H. Gao, *Beilstein J. Org. Chem.*, 2022, **18**, 1396–1402.
17 C. Angulo-Preckler, G. Genta-Jouve, N. Mahajan, M. de la Cruz, N. de Pedro, F. Reyes, K. Iken, C. Avila and O. P. Thomas, *J. Nat. Prod.*, 2016, **79**, 1132–1136.
18 J.-G. Song, J. Zheng, R.-J. Wei, Y.-L. Huang, J. Jiang, G.-H. Ning, Y. Wang, W. Lu, W.-C. Ye and D. Li, *Chem*, 2024, **10**, 924–937.
19 K. Tamura, G. Stecher and S. Kumar, *Mol. Biol. Evol.*, 2021, **38**, 3022–3027.
20 C. Zhu, B. Xu, D. A. Adpressa, J. D. Rudolf and S. Loesgen, *Angew. Chem., Int. Ed.*, 2021, **60**, 14163–14170.
21 B. Xu, D. J. Tantillo and J. D. Rudolf, *Angew. Chem., Int. Ed.*, 2021, **60**, 23159–23163.

22 Z. Li, B. Xu, V. Kojasoy, T. Ortega, D. A. Adpressa, W. Ning, X. Wei, J. Liu, D. J. Tantillo, S. Loesgen and J. D. Rudolf, *Chem*, 2023, **9**, 698–708.
23 B. Xu, W. Ning, X. Wei and J. D. Rudolf, *Org. Biomol. Chem.*, 2022, **20**, 8833–8837.
24 F.-L. Lin, K. A. Taizoumbe, Y.-X. Wang, J.-H. Huang, G.-Q. Wang, G.-D. Chen, J.-M. Lv, D. Hu, H. Gao and J. S. Dickschat, *Org. Biomol. Chem.*, 2024, **22**, 7971–7975.
25 CCDC 2474762: Experimental Crystal Structure Determination, 2025, DOI: **10.5517/ccdc.csd.cc2p2611**.