ARTICLES
https://doi.org/10.1038/s41557-020-0467-7
nature chemistry

Check for updates

FAD-dependent enzyme-catalysed intermolecular [4+2] cycloaddition in natural product biosynthesis

Lei Gao[1,2,8], Cong Su[1,8], Xiaoxia Du[2,8], Ruishan Wang[3,8], Shuming Chen[4], Yu Zhou[5], Chengwei Liu[6], Xiaojing Liu[2], Runze Tian[2], Liyun Zhang[2], Kebo Xie[1], She Chen[5], Qianqian Guo[2], Lanping Guo[3], Yoshio Hano[7], Manabu Shimazaki[7], Atsushi Minami[6], Hideaki Oikawa[6], Niu Huang[5], K. N. Houk[4], Luqi Huang[3]✉, Jungui Dai[1]✉ and Xiaoguang Lei[2]✉

The Diels–Alder reaction is one of the most powerful and widely used methods in synthetic chemistry for the stereospecific construction of carbon–carbon bonds. Despite the importance of Diels–Alder reactions in the biosynthesis of numerous secondary metabolites, no naturally occurring stand-alone Diels–Alderase has been demonstrated to catalyse intermolecular Diels–Alder transformations. Here we report a flavin adenine dinucleotide-dependent enzyme, *Morus alba* Diels–Alderase (MaDA), from *Morus* cell cultures, that catalyses an intermolecular [4+2] cycloaddition to produce the natural isoprenylated flavonoid chalcomoracin with a high efficiency and enantioselectivity. Density functional theory calculations and preliminary measurements of the kinetic isotope effects establish a concerted but asynchronous pericyclic pathway. Structure-guided mutagenesis and docking studies demonstrate the interactions of MaDA with the diene and dienophile to catalyse the [4+2] cycloaddition. MaDA exhibits a substrate promiscuity towards both dienes and dienophiles, which enables the expedient syntheses of structurally diverse natural products. We also report a biosynthetic intermediate probe (BIP)-based target identification strategy used to discover MaDA.

The Diels–Alder (D–A) reaction, a [4+2] cycloaddition between a conjugated diene and a dienophile to yield a cyclohexene skeleton, is one of the most powerful carbon–carbon bond forming reactions in synthetic chemistry[1]. Although putative Diels–Alderases, or [4+2] cyclases in the broader sense, have been predicted to be involved in the biosyntheses of many secondary metabolites[2,3], only a handful have been identified and characterized[4–7], with the majority being multifunctional enzymes[8–11]. The identification of the first stand-alone [4+2] cyclase, SpnF[12], led to the discovery of other intramolecular [4+2] cyclases from the biosynthetic pathways of spirotetronate[13,14] and/or spirotetramate[15] polyketides, decalin-containing polyketide–amino acid hybrids[16–20] and leporine[21], as well as intramolecular [6+4] cyclases from the biosynthetic pathway of streptoseomycin, a macrocyclic polyketide[22]. However, no stand-alone intermolecular [4+2] cyclases have been identified to date, even though intermolecular [4+2] cycloadditions have been catalysed by artificial enzymes generated via computational design[23,24] and immunological selection[25].

Methylcyclohexene motifs, accessible via D–A reactions, are common in *Moraceae* (mulberry) natural products with a role in Chinese traditional medicine[26,27]. These natural products display antimicrobial as well as diverse biological activities, such as the inhibition of hypoxia-inducible factor-1 and protein tyrosine phosphatase 1B[28–31]. Representative examples that incorporate diverse diene substituents are shown in Fig. 1a. Exogenous substrate feeding[32,33] and ^{13}C-labelling experiments[34] in *M. alba* cell cultures indicated the existence of an oxidase to furnish a reactive diene and a putative [4+2] pericyclase to catalyse the intermolecular [4+2] cycloaddition (Fig. 1b). We developed a chiral boron-mediated asymmetric D–A reaction for the enantioselective total syntheses of several natural products of this family[35]. A few racemic total syntheses are also reported[36,37]. Unfortunately, all the existing syntheses failed to achieve high levels of diastereo- and enantioselectivity, a problem that can be addressed chemoenzymatically if the relevant enzymes are isolated and characterized.

Historically, difficulties associated with identifying plant enzymes have slowed their discovery. Unlike in microbes, many biosynthetic genes for secondary metabolites in plants do not occur in gene clusters[38]. In addition, no stand-alone enzymes capable of catalysing intermolecular D–A reactions are known, which makes it difficult to perform homology-based or genome-context-based searches in modern biological databases. Both traditional activity-based purification and more modern transcriptomics-enabled methods are used to successfully identify elusive enzymes from plant natural product

[1]State Key Laboratory of Bioactive Substance and Function of Natural Medicines; CAMS Key Laboratory of Enzyme and Biocatalysis of Natural Drugs, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [2]Beijing National Laboratory for Molecular Sciences, Key Laboratory of Bioorganic Chemistry and Molecular Engineering of Ministry of Education, Department of Chemical Biology, College of Chemistry and Molecular Engineering, Synthetic and Functional Biomolecules Center, and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [3]State Key Laboratory Breeding Base of Dao-di Herbs, National Resources Center of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China. [4]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. [5]National Institute of Biological Sciences (NIBS), Beijing, China. [6]Division of Chemistry, Graduate School of Science, Hokkaido University, Sapporo, Japan. [7]Faculty of Pharmaceutical Sciences, Teikyo Heisei University, Tokyo, Japan. [8]These authors contributed equally: Lei Gao, Cong Su, Xiaoxia Du, Ruishan Wang. ✉e-mail: huangluqi01@126.com; jgdai@imm.ac.cn; xglei@pku.edu.cn

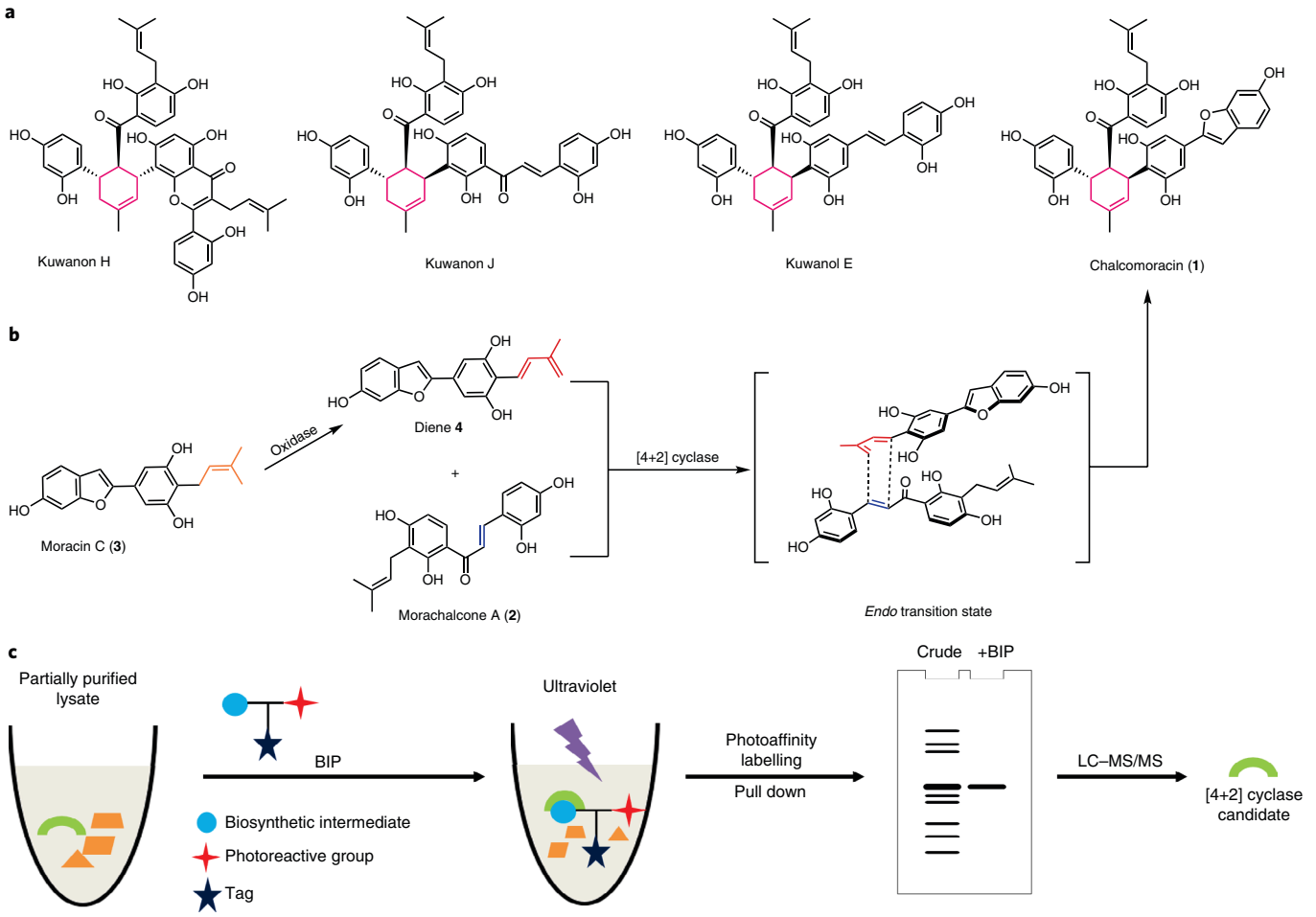


Fig. 1 | Proposed biosynthetic pathway for plant-derived D–A-type natural products and our strategy to identify the hypothetical intermolecular [4+2] pericyclase. a, Representative D–A-type natural products isolated from *Morus* plants. **b**, Proposed biosynthetic pathway for D–A-type *Morus* natural products. **c**, BIP-based target identification method to identify the putative [4+2] pericyclase. Using activity-based protein purification, we hoped to enrich the target enzyme(s), which can be more easily identified by using the BIP-based target identification strategy.

pathways[39,40]. However, transcriptomics-enabled methods often require characterized enzymes as bait, as well as gene prediction based on the specialized metabolism for the co-expression analysis[41], whereas activity-based protein purification entails tedious and lengthy purification processes and often suffers from false negative results. To overcome these challenges, we developed a complementary approach termed 'BIP-based target identification', which reveals unknown biosynthetic enzymes based on substrate binding (Fig. 1c). In our proposal, activity-based protein purification is first used to enrich the target protein(s), which are then subjected to liquid chromatography–mass spectroscopy/mass spectroscopy (LC–MS/MS) proteomics analysis. With this method, possible off-target proteins can be removed, and the abundance of the targeted protein(s) increased to reduce both false positive and false negative results during the photoaffinity labelling and pull-down steps. Hits identified via BIP-based target identification are therefore highly reliable. Transcriptional analysis is then used to narrow down the candidate gene pool for further biochemical and function analysis.

Using the BIP-based method, we identified and functionally characterized the first stand-alone intermolecular Diels-Alderase, *M. alba* Diels–Alderase (MaDA), from *M. alba* cell cultures. MaDA is a flavin adenine dinucleotide (FAD)-dependent oxidase-like Diels–Alderase that demonstrates exceptional catalytic efficiency and broad substrate scope. Density functional theory (DFT) calculations support a concerted but asynchronous D–A mechanism, and preliminary kinetic isotope effect (KIE) studies also show the reaction to be asynchronous in the enzyme. We obtained an X-ray crystal structure of MaDA with FAD covalently bound in (at 2.3 Å resolution), and performed extensive structure-guided mutagenesis to show possible modes of interaction between MaDA and its substrates. *M. alba* moracin C oxidase (MaMO), a FAD-dependent oxidase, was also identified and functionally characterized to catalyse the oxidation of the isoprenyl moiety in moracin C (**3**) into the reactive diene **4**.

Results and discussion

Identification of the oxidase and [4+2] pericyclase. Using *M. alba* cell cultures generated from leaves[42], we identified the major D–A type natural product as chalcomoracin (**1**) (Supplementary Fig. 1). Lysates from the cultures were assayed by incubating with both the putative dienophile morachalone A (**2**) and the diene precursor moracin C (**3**). High-performance liquid chromatography (HPLC) analysis showed that chalcomoracin (**1**) was produced as the major product, alongside an unstable compound subsequently confirmed as diene **4** (Fig. 2a). These observations indicate that an oxidase and an intermolecular [4+2] pericyclase may be involved in the biosynthesis of **1**. Subsequent activity-based protein purification, which consisted of ammonium sulfate (AS) fractionation, hydrophobic interaction chromatography (HIC), ion exchange chromatography (IEC) and size exclusion chromatography (SEC),

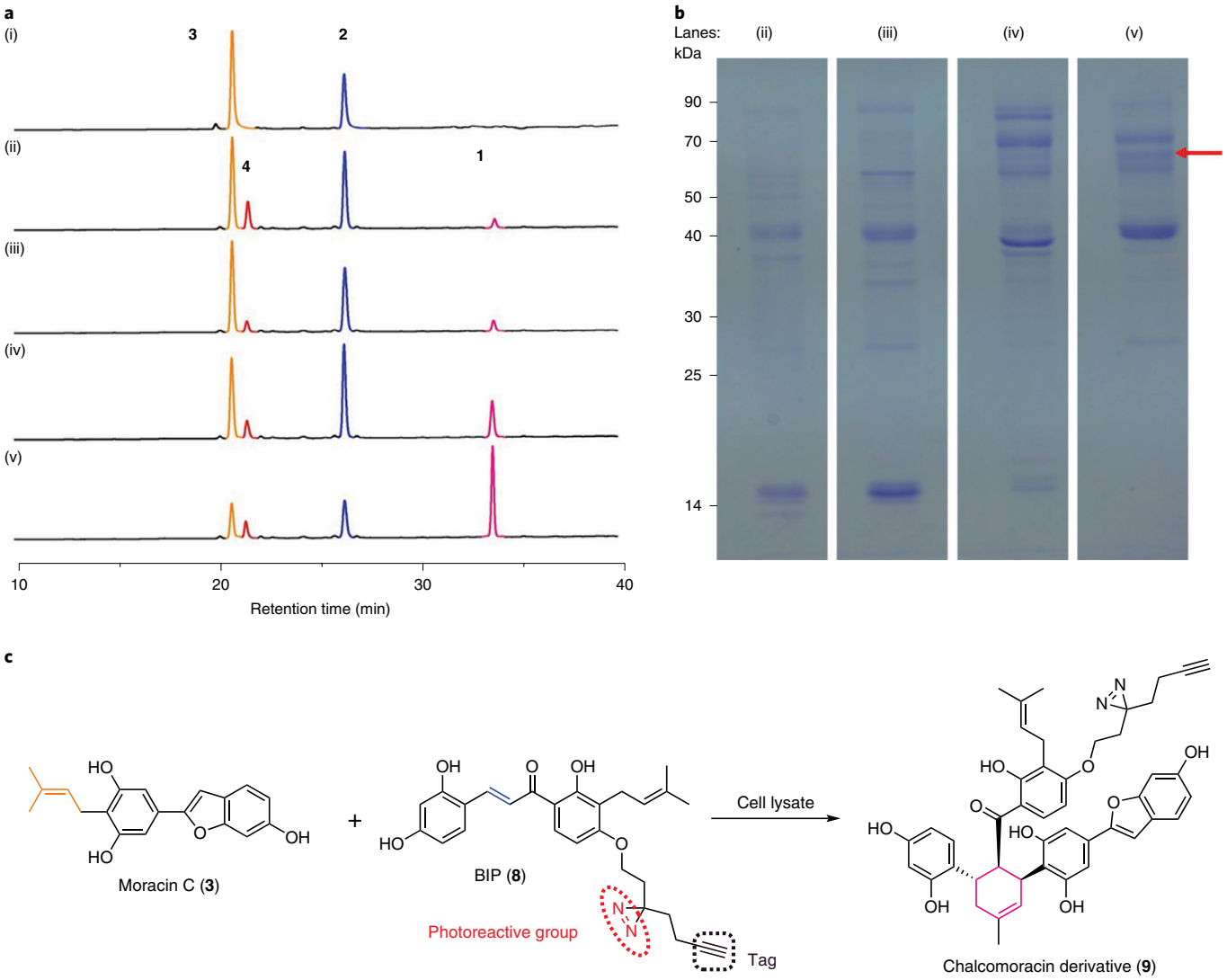


Fig. 2 | Identification of the intermolecular [4+2] pericyclase using activity-based protein purification and BIP-based target identification. a, In vitro analysis ($\lambda = 340$ nm) of **2** (100 μM) and **3** (100 μM) with 0.1 mg ml^{-1} in different protein fractions with the buffer (i), AS fraction (ii), HIC fraction (iii), IEC fraction (iv) and SEC fraction (v). Three experiments were repeated independently with similar results. **b**, Approximately 4 μg of protein was loaded in each lane, and the resolution was achieved using 12% (w/v) acrylamide gel: AS fraction (ii), HIC fraction (iii), IEC fraction (iv) and SEC fraction (v). Three experiments were repeated independently with similar results. **c**, Structures of BIP probe **8** and the corresponding chalcomoracin derivative **9**. Probe **8** can be recognized by the putative [4+2] pericyclase and transformed into the [4+2] adduct **9**. After being activated by ultraviolet light (365 nm), the diazirine forms a reactive carbene species that can covalently crosslink with the [4+2] pericyclase in the enzymatic reaction.

led to an 11-fold improvement in chalcomoracin (**1**) production (Fig. 2a), suggesting the enrichment of the putative oxidase and [4+2] pericyclase. Incubation of **2** and **3** in the SEC fraction for longer times resulted in increased production of **1**, but not diene **4** (Extended Data Fig. 1), which indicates that the [4+2] pericyclase is possibly a different enzyme from the oxidase that has a higher reaction rate. Active fractions obtained by AS precipitation and HIC, IEC and SEC were also resolved by SDS–polyacrylamide gel electrophoresis (SDS–PAGE). We observed one clearly enriched band in the SEC fraction with a mobility that corresponded to a molecular mass of ~60 kDa (indicated by the red arrow in Fig. 2b), which possibly contains the oxidase and/or the [4+2] pericyclase. This band was later analysed by a LC–MS/MS proteomics analysis, which suggested that the proteins enriched by activity-based protein purification were berberine bridge enzyme (BBE)-like enzymes (also known as reticuline oxidase-like enzymes) (Supplementary Table 1).

BBE-like enzymes are widely involved in the biosynthesis of bacterial, fungal and plant natural products. PenH, a BBE homologue, is reported to be responsible for the formation of structurally similar dienes in the biosynthesis of fungal natural products[43]. We propose that the BBE-like enzymes in our cell cultures oxidize moracin C (**3**) into diene **4** (Extended Data Fig. 2a) through a similar mechanism, which utilizes O_2 and generates H_2O_2. Indeed, enzymatic activity was lost in the absence of O_2 (Extended Data Fig. 2b), whereas in the presence of O_2 (using the SEC fraction as the crude enzyme), H_2O_2 was detected as an end product (Extended Data Fig. 2c).

We next used our BIP-based method to reveal the intermolecular [4+2] pericyclase. As diene **4** is unstable under our assay conditions and can also bind to the oxidase, we designed the BIP molecule based on the structure of dienophile **2**. To determine which hydroxyl group in **2** would be suitable for the introduction of a side chain that contained a photoaffinity group and an alkyne tag, we synthesized three derivatives of **2** (**5**–**7**; Extended Data Fig. 3a)

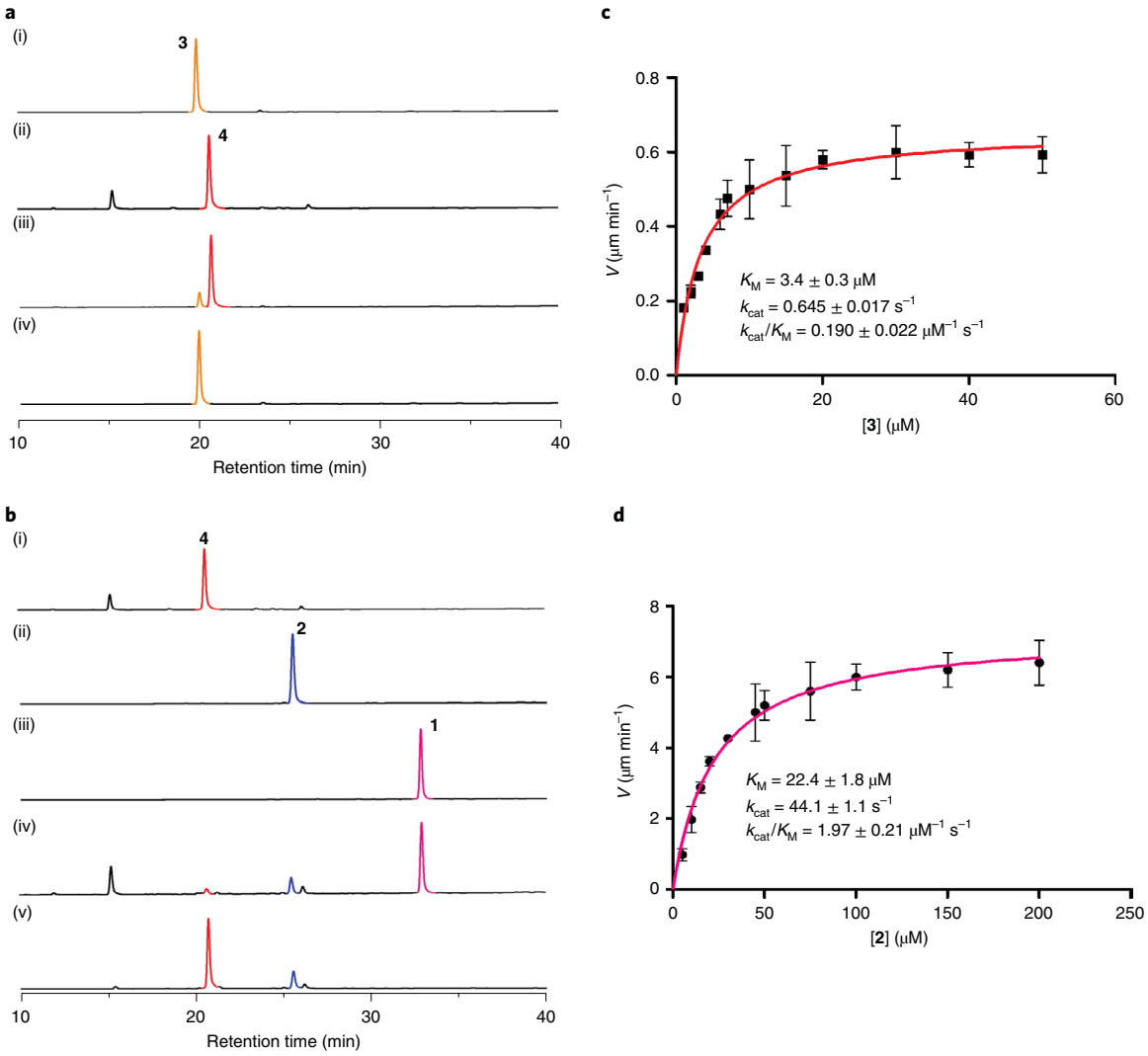


Fig. 3 | Functional and biochemical characterization of MaMO and MaDA. a, In vitro analysis ($\lambda = 340$ nm) of **3** with 17 nM MaMO for 5 min for standard compound **3** alone (i), standard compound **4** alone (ii), compound **3** with MaMO (iii) and compound **3** with buffer (iv). Three experiments were repeated independently with similar results. **b**, In vitro analysis ($\lambda = 340$ nm) of **2** and **4** with 2.7 nM MaDA for 5 min: standard compound **4** alone (i), standard compound **2** alone (ii), standard compound **1** alone (iii), compounds **2** and **4** with MaDA (iv) and compounds **2** and **4** with buffer (v). Three experiments were repeated independently with similar results. **c**, Kinetic analysis of MaMO for moracin C (**3**). The reactions were performed with **3** from 1 to 50 μM at pH 8.0 and 50 °C in a total volume of 100 μl that contained 17 nM MaMO. **d**, Kinetic analysis of MaDA for morachalcone A (**2**). The reactions were performed with **2** from 5 to 200 μM with the saturating diene **4** (1,000 μM) at pH 8.0 and 50 °C in a total volume of 100 μl that contained 2.7 nM MaDA. K_M, k_{cat} and k_{cat}/K_M values represent the mean ± s.d. of three independent replicates.

and found that **7** showed the best [4+2] activity, comparable to that of **2** (Extended Data Fig. 3b). Based on this result, we synthesized BIP **8**, which gave D–A adduct **9** (Fig. 2c and Extended Data Fig. 3b). We then used **8** in photoaffinity labelling in the SEC fraction, followed by a pull-down assay. One distinctive band with a mobility that corresponded to a molecular mass of ~60 kDa, similar in mass to the aforementioned oxidase, was specifically precipitated by the probe in the presence of ultraviolet light (Extended Data Fig. 3c). This band was further excised and subjected to LC–MS/MS proteomics analysis. Interestingly, BBE-like enzymes were identified in this band as among the top 20 possible candidate interacting proteins (Supplementary Table 2), which indicates that the putative [4+2] pericyclase in the *M. alba* cell cultures may also be of this protein family.

We investigated the expression profile of these BBE-like enzymes in our cell cultures. Transcriptome analysis revealed that 14 different transcripts were annotated as BBE-like enzymes or related proteins with the fragments per kilobase of exon per million reads mapped greater than 10 (Supplementary Table 3). We successfully cloned the full-length complementary DNAs (cDNAs) that encoded the top two BBE-like enzymes with a sequence identity of 61.6%, henceforth named MaMO (CL5271.Contig4_ MaL_022) and MaDA (CL1657.Contig4_MaL_022). The deduced protein sequences showed a high sequence similarity with the BBE-type tetrahydrocannabinolic acid (THCA) synthase[44] (47% identity for MaMO and 45% identity for MaDA). Both MaMO and MaDA share the conserved His residue (His115/116) previously shown to undergo covalent binding with FAD[44]. A notable difference, however, was that the other previously reported Cys residue of BBE-like enzymes also known to bind covalently to FAD was present in MaMO (Cys177), but absent in MaDA (Ala178) (Supplementary Fig. 2).

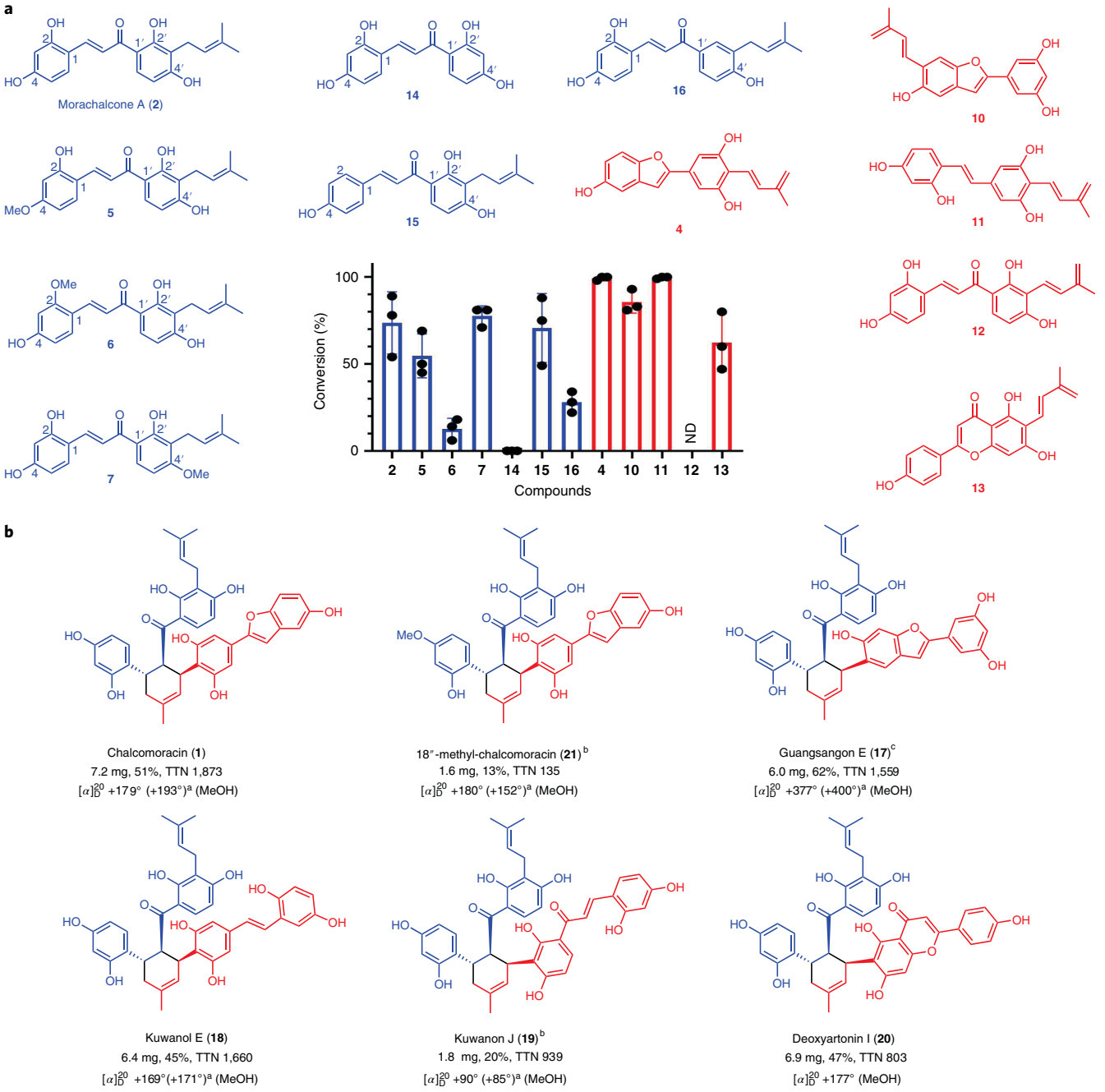


Fig. 4 | Substrate scope of MaDA and the chemoenzymatic synthesis of D–A natural products. a, Substrate scope of MaDA. Although the D–A adduct kuwanon J could be detected in a HPLC analysis, diene **12** was too unstable to be detected in a HPLC analysis after incubation at 50 °C for 5 min. Accordingly, the conversion rate of **12** was not determined (ND). Conversion values represent mean ± s.d. of three independent replicates. **b**, Chemoenzymatic synthesis of various D–A-type natural products by MaDA. Under standard conditions, 0.35 mg of MaDA was used to promote the [4+2] cycloaddition in 100 ml of Tris-HCl buffer (20 mM, pH = 8.0). [a]Specific rotation value reported in the literature. [b]The enzymatic reaction was conducted in 150 ml of Tris-HCl buffer with 1.05 mg of MaDA. [c]The enzymatic reaction was conducted in 50 ml of Tris-HCl buffer with 0.35 mg of MaDA. $[\alpha]_D^{20}$, specific rotation at 589 nm (sodium D line) and 20 °C.

Functional and biochemical characterization of MaMO and MaDA. Next, we successfully expressed MaMO and MaDA in *Komagataella phaffii* and Hi5 insect cells, respectively, and purified them using nickel-charged nitrilotriacetic acid (Ni-NTA) resin chromatography. Solutions with both recombinant MaMO and MaDA appeared yellow in colour, and displayed ultraviolet–visible absorption spectra consistent with those of typical flavoproteins (Supplementary Fig. 3). When MaDA was incubated with diene precursor **3**, no product was detected (Extended Data Fig. 4). In contrast, when MaMO was assayed in the same conditions, diene **4** was produced (Fig. 3a), which indicates that MaMO is the oxidase responsible for diene formation. Incubating dienophile **2** and diene **4** together with MaDA generated chalcomoracin (**1**) (see Supplementary Information for the high-resolution MS, ^{1}H NMR, ^{13}C NMR and polarimetry characterizations), which shows that MaDA is an intermolecular [4+2] pericyclase (Fig. 3b).

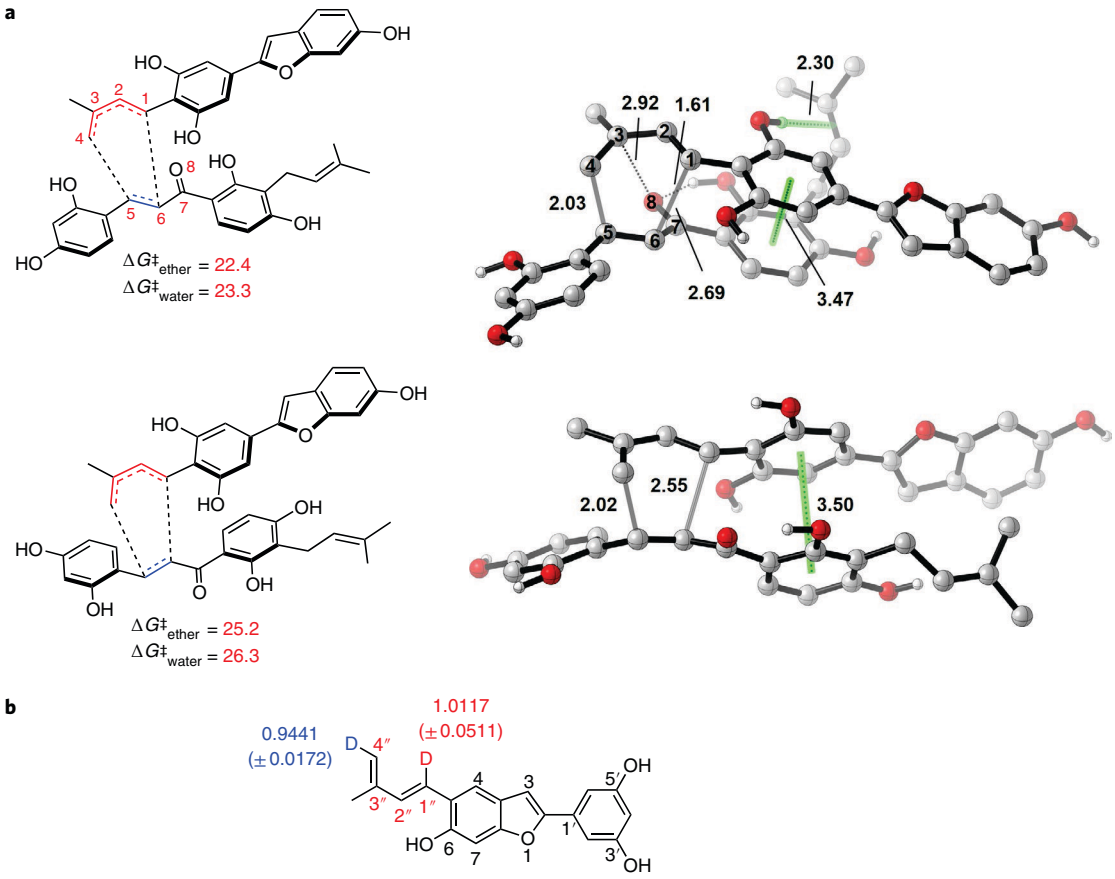


Fig. 5 | DFT calculations and KIE results for the intermolecular [4+2] cycloaddition reaction. a, DFT-calculated D–A transition states at the ωB97X-D/6-311++G(d,p), CPCM//ωB97X-D/6-31G(d) level of theory for *endo* **TS-1** (top) and *exo* **TS-2** (bottom). C–H hydrogens are omitted for clarity. Interatomic distances are in Å. Energies are in kcal mol^{-1}. **b**, KIE values obtained using two deuterium-labelled diene analogues. The α-secondary deuterium KIE at each site of interest was measured using the method of internal competition in three separate experimental trials. The enzymatic reaction that contained deuterated and non-deuterated diene **10** was quenched at three different timepoints and then the remaining diene **10** from these samples was recovered by preparative HPLC and analysed by MS. Peak integrations of $M+1$ and $M+2$ were used to determine the ratios of deuterated/non-deuterated diene **10** in different samples, and these ratios were further used to determine the deuterium KIEs at two different sites of **10**.

Comparison with the chemically synthesized racemic **1** confirmed that the product generated by MaDA is optically pure (100% e.e.) and exclusively *endo* (Supplementary Fig. 4). In contrast, when MaMO was incubated with **2** and **3**, **3** was completely converted into diene **4**, whereas neither the oxidation product of **2** nor the D–A product **1** was observed (Extended Data Fig. 4). These observations further demonstrate that MaMO is an oxidase that selectively oxidizes **2** to give diene **4**, whereas MaDA is a stand-alone intermolecular [4+2] pericyclase. Biochemical characterization of the two purified recombinant enzymes showed that both enzymes were independent of the divalent cation. The optimal temperature was 50 °C for both enzymes, and the optimal pH values were 7.5 and 8.0 for MaMO and MaDA, respectively (Extended Data Fig. 5). We also determined the steady-state kinetic parameters for both MaDA and MaMO (Fig. 3c,d). We found MaDA to be more catalytically efficient ($k_{cat}/K_M = 1.97\,\mu M^{-1}\,s^{-1}$ for dienophile **2**) than MaMO ($k_{cat}/K_M = 0.19\,\mu M^{-1}\,s^{-1}$ for **3**). As diene **4** is highly unstable (Extended Data Fig. 6), we synthesized a more stable diene analogue **10** and measured the kinetic parameters of MaDA with both diene **10** and dienophile **2** (Extended Data Fig. 7), and found that k_{cat}/K_M for diene **10** and dienophile **2** were $4.04 \pm 0.18\,mM^{-1}\,s^{-1}$ and $10.1 \pm 3.8\,mM^{-1}\,s^{-1}$, respectively.

MaDA is a substrate promiscuous Diels–Alderase. To explore the substrate scope and probe the synthetic utility of this novel pericyclase in vitro, three putative diene precursors (**11**–**13**) of D–A products in *M. alba* were synthesized and incubated with dienophile **2**. Six natural and unnatural derivatives (**5**–**7** and **14**–**16**) of **2** as dienophiles were also assayed with diene **4**. LC–MS analysis revealed that MaDA displayed a substrate selectivity for the dienophile, but could react with all of the tested dienes (Fig. 4a), which enabled the incorporation of diverse diene substituents into *Morus* natural products. Efficient chemoenzymatic total syntheses of guangsangon E (**17**), kuwanol E (**18**), kuwanon J (**19**), deoxyartonin I (**20**) and 18″-*O*-methylchalcomoracin (**21**) were accomplished through late-stage enzymatic reactions with a high efficiency (up to 1,873 total turnover number (TTN); Fig. 4b). All the synthesized natural products were characterized by NMR spectroscopy and polarimetry, and the data for synthetic samples fully matched those reported for natural products.

DFT calculation and KIE experiment for the [4+2] cycloaddition reaction. We used DFT calculations to determine the mechanism and intrinsic selectivity of the uncatalysed D–A reaction between **2** and **4**. The lowest-energy D–A transition state (**TS-1**) leads to the *endo* product, and is favoured over the *exo* TS (**TS-2**) by 2.8–3.0 kcal mol^{-1} as well as other regioisomers (Fig. 5a and Extended Data Fig. 8a). The geometry of **TS-1** is characteristic of a concerted but asynchronous reaction, with the two forming C–C bonds at 2.03 and 2.69 Å (Fig. 5a). The intramolecular hydrogen

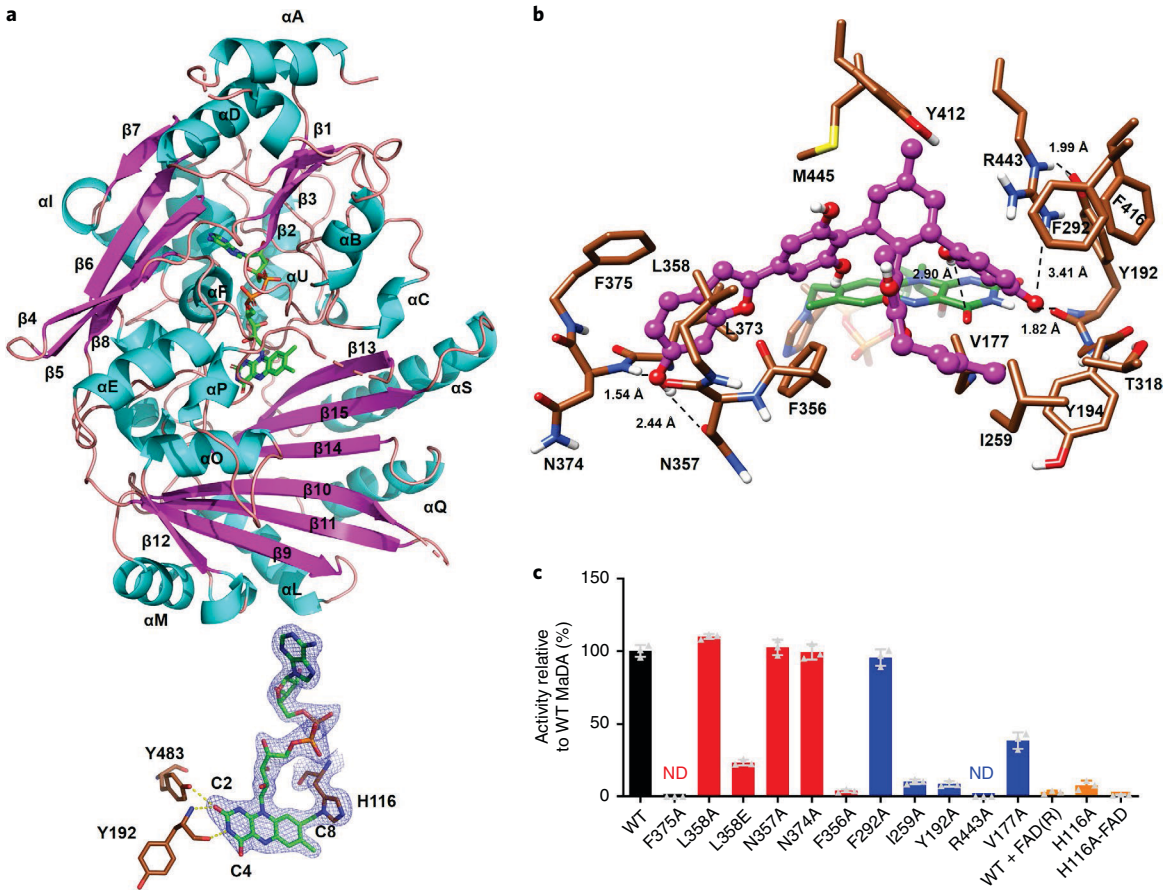


Fig. 6 | Structure, molecular dynamics (MD) simulation and site-directed mutagenesis characterization of MaDA. a, Overall structure of MaDA shown in a ribbon model. α-helices, cyan; β-strands, magenta; loops, light pink. The FAD cofactor (green and shown in a stick representation) forms a covalent bond with MaDA, as shown in the 2Fo−Fc electron density map (blue mesh, contoured at the 1.0σ level) for FAD and the covalently bound His116. Residues that form hydrogen bonds with FAD are shown in brown (Y192 and Y483). For amino acids: nitrogen, blue; oxygen, red; carbon, brown. **b**, Product **1** docked into a substrate cavity and optimized by 500 ns MD. Stereo views of the docking model show the interactions between **1** (purple) and nearby residues (brown). FAD is shown in green and the hydrogen-bonding interactions are shown as dotted lines. **c**, Enzymatic reactions were performed with 2.7 nM MaDA variants and 10 μM of **2** and **4**. Red columns indicate that mutated residues are predicted to interact with **4**, blue columns with **2** and brown columns with FAD. WT+FAD(R), WT with the reduced form of FAD; H116A-FAD, the H116A variant for which non-covalent FAD was removed by adding a 1 M KBr solution. Enzyme activity values represent mean ± s.d. of three independent replicates. ND, activity not detected in the enzymatic assay.

bond between O8 and the *ortho* hydroxyl group presumably promotes the D–A reaction by lowering the lowest unoccupied molecular orbital energy of the dienophile. The short (2.92 Å) O8–C3 distance indicates a strong secondary orbital interaction between the O8 lone pair and the antibonding π orbital of the diene (Fig. 5a and Supplementary Fig. 5)[45].

Although the rate of the uncatalysed D–A reaction between **2** and **4** could not be measured due to the instability of **4**, we performed preliminary in-enzyme KIE measurements using a more stable diene analogue (**10**) to obtain insight into the enzyme mechanism. DFT calculations support a similar concerted but asynchronous pathway for the D–A reaction between diene **10** and **2**, with the two forming C–C bonds at 2.05 and 2.69 Å (Extended Data Fig. 8b). Two deuterated derivatives of **10** (**10**-1″D and **10**-4″D) were reacted with **2** in the presence of MaDA to obtain α-secondary KIE values (k_H/k_D) of 1.0117 (0.0511) and 0.9441 (± 0.0172), respectively, which provided evidence that an asynchronous D–A pathway was also operational in the enzyme (Fig. 5b).

Interactions between substrates and MaDA. To further understand how this intermolecular [4+2] cyclization is catalysed by the enzyme, the crystal structure of MaDA was determined at 2.3 Å resolution (Protein Database ((PDB) ID 6JQH; Extended Data Fig. 10) using molecular replacement based on the THCA synthase structure (PDB ID 3VTE)[44]. Two monomers of MaDA were identified in the asymmetric unit within space group $P2_1$. The overall topology of MaDA is similar to that of THCA synthase (Supplementary Fig. 6), and comprises 26 α-helixes (αA–αZ) and 15 β-strands (β1–β18), among which 14 α-helixes and 15 β-strands are shown in Fig. 6a. The electron density clearly reveals the presence of FAD, which forms a covalent bond to H116 through the 8α-N1-histidyl-FAD linkage. The flavin moiety of FAD binds to MaDA through hydrogen-bonding interactions with Y192 and Y483, as shown in Fig. 6a. This covalently linked FAD together with one α-helix (αO), four β-strands (β10, β11, β14 and β15) and four loops (residues 175–182, 192–195, 359–367 and 371–378) form a large substrate cavity with a volume of 723 Å^3 (Supplementary Fig. 6).

Despite numerous attempts, we were not able to obtain a co-crystal structure of MaDA with any native D–A precursors or products, presumably because the remarkably fast kinetics of MaDA limit the possible substrate trapping inside enzyme. To explore possible interactions between MaDA and its substrates, D–A product **1** was docked into MaDA's cavity using UCSF DOCK 3.7, followed by 500 ns MD simulations to relax the protein–ligand

binding complex and obtain an energetically favourable binding conformation (Supplementary Fig. 6). It was found that the benzofuran group may interact with F356, N357, L358, L373, N374 and F375 (Fig. 6b). Possible interactions for the dienophile moiety include residues V177, I259, F292, Y192 and R443 (Fig. 6b). Probable hydrogen-bonding interactions include one between the C16″-OH and the C4 carbonyl of the flavin, and an intramolecular hydrogen bond between the C10″-OH and the C8″-carbonyl of **1** (Supplementary Fig. 6), consistent with the DFT calculations. Substrates **2** and **4** were also docked into the MaDA cavity simultaneously (Supplementary Fig. 7). When the two substrates are well positioned by MaDA through numerous interactions, the enone double bond in dienophile **2** and the conjugated diene in **4** are shown to be close enough (3.70 and 3.87 Å, Supplementary Fig. 7) to form the cyclohexene skeleton.

To further probe the involvement of key residues predicted to interact with the substrates by our docking results, we mutated a series of putative binding site residues and compared the D–A catalytic activities of the variants with the wild-type (WT) MaDA (Fig. 6c). The variants F356A, F375A, Y192A and R443A each caused >90% loss of enzymatic activity, which indicates the possible role of these residues in promoting the D–A reaction through π–π stacking and/or hydrogen-bonding interactions. Mutating the bulky hydrophobic residues V177 and I259 to Ala resulted in an ~60 and 90% losses in activity, respectively. The variants N357A and F292A showed little change in catalytic activity. Substitution of L358 with Ala resulted in a slightly higher activity than that the WT MaDA, whereas substitution of the same residue with Glu resulted in an ~80% loss of enzymatic activity. The kinetic characterizations of these MaDA variants (Supplementary Table 4) further support the key roles of these residues for substrate binding and enzyme activity.

To explore the function of the FAD cofactor in MaDA, we reduced MaDA with an excess of sodium dithionite. The reduced MaDA was stable for ~30 minutes (Extended Data Fig. 9a). Reduction of MaDA renders it essentially unable to catalyse the [4+2] cycloaddition (Fig. 6c). When His116 was mutated to Ala to break the covalent linkage between MaDA and FAD, partial loss of FAD was observed through monitoring the absorption change at 450 nm (Extended Data Fig. 9b). This mutation led to dramatically decreased [4+2] activity (Fig. 6c). We conclude that the FAD cofactor is essential for MaDA's catalytic activity in the intermolecular [4+2] cycloaddition, perhaps due to its role in maintaining substrate binding via hydrogen-bonding interaction.

In summary, we identified the enzyme functionally responsible for the intermolecular [4+2] cycloaddition reaction in the biosynthesis of chalcomoracin from *Morus* plants. Since the submission of this article, two new [4+2] cyclases, TiCorS[46] and EupF[47], were identified from plant and fungi, respectively. TiCorS involves an intramolecular D–A reaction, whereas EupF catalyses the formation of a highly reactive diene that then reacts with a dienophile substrate. These differ from MaDA, which is a stand-alone intermolecular D–A pericyclase that catalyses only the intermolecular D–A reaction; furthermore, MaDA involves a quite different mechanism and outstanding catalytic efficiency. MaMO, the enzyme responsible for the formation of the diene substrate for MaDA, was also functionally characterized. DFT calculations support a concerted, but asynchronous, pericyclic mechanism for the uncatalysed D–A reaction, and preliminary KIE measurements indicate that a similar asynchronous pathway is operational in the enzyme. We show that, due to MaDA's ability to accept a wide scope of diene substrates, a chemoenzymatic total syntheses of many natural products can be accomplished with MaDA. MaDA therefore provides an ideal starting point for future engineering efforts to develop novel intermolecular Diels–Alderases with a versatile catalytic repertoires. Our successful characterization of MaDA highlights the utility of the BIP-based target identification approach as a tool to speed up the discovery of unknown biosynthetic enzymes of plant origin.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-020-0467-7.

Received: 25 September 2019; Accepted: 14 April 2020;
Published online: 25 May 2020

References

1. Nicolaou, K. C., Snyder, S. A., Montagnon, T. & Vassilikogiannakis, G. The Diels–Alder reaction in total synthesis. *Angew. Chem. Int. Ed.* **41**, 1668–1698 (2002).
2. Oikawa, H. & Tokiwano, T. Enzymatic catalysis of the Diels–Alder reaction in the biosynthesis of natural products. *Nat. Prod. Rep.* **21**, 321–352 (2004).
3. Stocking, E. M. & Williams, R. M. Chemistry and biology of biosynthetic Diels–Alder reactions. *Angew. Chem. Int. Ed.* **42**, 3078–3115 (2003).
4. Minami, A. & Oikawa, H. Recent advances of Diels–Alderases involved in natural product biosynthesis. *J. Antibiot.* **69**, 500–506 (2016).
5. Klas, K., Tsukamoto, S., Sherman, D. H. & Williams, R. M. Natural Diels–Alderases: elusive and irresistable. *J. Org. Chem.* **80**, 11672–11685 (2015).
6. Jeon, B. S., Wang, S.-A., Ruszczycky, M. W. & Liu, H. W. Natural [4+2]-cyclases. *Chem. Rev.* **117**, 5367–5388 (2016).
7. Jamieson, C. S., Ohashi, M., Liu, F., Tang, Y. & Houk, K. N. The expanding world of biosynthetic pericyclase: cooperation of experiment and theory for discovery. *Nat. Prod. Rep.* **36**, 698–713 (2019).
8. Kasahara, K. et al. Solanapyrone synthase, a possible Diels–Alderase and iterative type I polyketide synthase encoded in a biosynthetic gene cluster from *Alternaria solani*. *ChemBioChem* **11**, 1245–1252 (2010).
9. Auclair, K. et al. Lovastatin nonaketide synthase catalyzes an intramolecular Diels–Alder reaction of a substrate analogue. *J. Am. Chem. Soc.* **122**, 11519–11520 (2000).
10. Ose, T. et al. Insight into a natural Diels–Alder reaction from the structure of macrophomate synthase. *Nature* **422**, 185–189 (2003).
11. Hudson, G. A., Zhang, Z., Tietz, J. I., Mitchell, D. A. & van der Donk, W. A. In vitro biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J. Am. Chem. Soc.* **137**, 16012–16015 (2015).
12. Kim, H. J., Ruszczycky, M. W., Choi, S. H., Liu, Y. N. & Liu, H. W. Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature* **473**, 109–112 (2011).
13. Hashimoto, T. et al. Biosynthesis of versipelostatin: identification of an enzyme-catalyzed [4+2]-cycloaddition required for macrocyclization of spirotetronate-containing polyketides. *J. Am. Chem. Soc.* **137**, 572–575 (2015).
14. Byrne, M. J. et al. The catalytic mechanism of a natural Diels–Alderase revealed in molecular detail. *J. Am. Chem. Soc.* **138**, 6095–6098 (2016).
15. Tian, Z. et al. An enzymatic [4+2] cyclization cascade creates the pentacyclic core of pyrroindomycins. *Nat. Chem. Biol.* **11**, 259–265 (2015).
16. Li, L. et al. Biochemical characterization of a eukaryotic decalin-forming Diels–Alderase. *J. Am. Chem. Soc.* **138**, 15837–15840 (2016).
17. Li, L. et al. Genome mining and assembly-line biosynthesis of the UCS1025A pyrrolizidinone family of fungal alkaloids. *J. Am. Chem. Soc.* **140**, 2067–2071 (2018).
18. Kato, N. et al. Control of the stereochemical course of [4+2] cycloaddition during *trans*-decalin formation by Fsa2-family enzymes. *Angew. Chem. Int. Ed.* **57**, 9754–9758 (2018).
19. Tan, D. et al. Genome-mined Diels–Alderase catalyzes formation of the *cis*-octahydrodecalins of varicidin A and B. *J. Am. Chem. Soc.* **141**, 769–773 (2019).
20. Zhang, Z. et al. Enzyme-catalyzed inverse-electron demand Diels–Alder reaction in the biosynthesis of antifungal ilicicolin H. *J. Am. Chem. Soc.* **141**, 5659–5663 (2019).
21. Ohashi, M. et al. SAM-dependent enzyme-catalysed pericyclic reactions in natural product biosynthesis. *Nature* **549**, 502–506 (2017).
22. Zhang, B. et al. Enzyme-catalysed [6+4] cycloadditions in the biosynthesis of natural products. *Nature* **568**, 122–126 (2019).
23. Siegel, J. B. et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels–Alder reaction. *Science* **329**, 309–313 (2010).
24. Preiswerk, N. et al. Impact of scaffold rigidity on the design and evolution of an artificial Diels–Alderase. *Proc. Natl Acad. Sci. USA* **111**, 8013–8018 (2014).
25. Hilvert, D., Hill, K. W., Nared, K. D. & Auditor, M.-T. M. Antibody catalysis of the Diels–Alder reaction. *J. Am. Chem. Soc.* **111**, 9261–9262 (1989).

26. Nomura, T., Hano, Y. & Fukai, T. Chemistry and biosynthesis of isoprenylated flavonoids from Japanese mulberry tree. *Proc. Jpn Acad. B* **85**, 391–408 (2009).
27. Nomura, T. & Hano, Y. Isoprenoid-substituted phenolic compounds of moraceous plants. *Nat. Prod. Rep.* **11**, 205–218 (1994).
28. Dat, N. T. et al. Hypoxia-inducible factor-1 inhibitory benzofurans and chalcone-derived Diels–Alder adducts from *Morus* species. *J. Nat. Prod.* **72**, 39–43 (2009).
29. Wang, M. et al. Diels–Alder adducts with PTP1B inhibition from *Morus notabilis*. *Phytochemistry* **109**, 140–146 (2015).
30. Fukai, T., Kaitou, K. & Terada, S. Antimicrobial activity of 2-arylbenofurans from *Morus* species against methicillin-resistant *Staphylococcus aureus*. *Fitoterapia* **76**, 708–711 (2005).
31. Esposito, F. et al. Kuwanon-L as a new allosteric HIV-1 integrase inhibitor: molecular modeling and biological evaluation. *ChemBioChem* **16**, 2507–2512 (2015).
32. Ueda, S., Nomura, T., Fukai, T. & Matsumoto, J. Kuwanon J, a new Diels–Alder adduct and chalcomoracin from callus culture of *Morus alba* L. *Chem. Pharm. Bull.* **30**, 3042–3045 (1982).
33. Hano, Y., Nomura, T. & Ueda, S. Biosynthesis of optically active Diels–Alder type adducts revealed by an aberrant metabolism of *O*-methylated precursors in *Morus alba* cell cultures. *J. Chem. Soc. Chem. Commun.* **8**, 610–613 (1990).
34. Hano, Y., Nomura, T. & Ueda, S. Biosynthesis of chalcomoracin and kuwanon J, the Diels–Alder type adducts, in *Morus alba* L. cell cultures. *Chem. Pharm. Bull.* **37**, 554–556 (1989).
35. Han, J. et al. Enantioselective biomimetic total syntheses of kuwanons I and J and brosimones A and B. *Angew. Chem. Int. Ed.* **53**, 9257–9261 (2014).
36. Qi, C. et al. Biomimetic dehydrogenative Diels–Alder cycloadditions: total syntheses of brosimones A and B. *Angew. Chem. Int. Ed.* **52**, 8345–8348 (2013).
37. Gunawan, C. & Rizzacasa, M. Mulberry Diels–Alder adducts: synthesis of chalcomoracin and mulberrofuran C methyl ethers. *Org. Lett.* **12**, 1388–1391 (2010).
38. Field, B. & Osbourn, A. E. Metabolic diversification-independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543–547 (2008).
39. Geu-Flores, F. et al. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138–142 (2012).
40. Chen, X. et al. A pathogenesis-related 10 protein catalyzes the final step in thebaine biosynthesis. *Nat. Chem. Biol.* **14**, 738–743 (2018).
41. Medema, M. H. & Osbourn, A. Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. *Nat. Prod. Rep.* **33**, 951–962 (2016).
42. Zhang, D.-W., Tao, X.-Y., Yu, L.-Y. & Dai, J.-G. New 2-arylbenzofuran metabolite from cell cultures of *Morus alba*. *J. Asian Nat. Prod. Res.* **17**, 683–688 (2015).
43. Zou, Y. et al. Tandem prenyltransferases catalyze isoprenoid elongation and complexity generation in biosynthesis of quinolone alkaloids. *J. Am. Chem. Soc.* **137**, 4980–4983 (2015).
44. Shoyama, Y. et al. Structure and function of Δ1-tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling the psychoactivity of *Cannabis sativa*. *J. Mol. Biol.* **423**, 96–105 (2012).
45. Yang, Z. et al. Relationships between product ratios in ambimodal pericyclic reactions and bond lengths in transition structures. *J. Am. Chem. Soc.* **140**, 3061–3067 (2018).
46. Farrow, S. C. et al. Biosynthesis of an anti-addiction agent from the iboga plant. *J. Am. Chem. Soc.* **141**, 12979–12983 (2019).
47. Chen, Q. et al. Enzymatic intermolecular hetero-Diels–Alder reaction in the biosynthesis of tropolonic sesquiterpenes. *J. Am. Chem. Soc.* **141**, 14052–14056 (2019).

Methods

General. PCRs were carried out using the standard thermocycling protocols with either TransStart FastPfu DNA Polymerase (Transgene Biotech) or KOD-Plus-Neo (Toyobo) and templates and buffers as specified for each reaction (detailed in the experimental sections below). All the protein quantification was done by the method of Bradford using Protein Assay Dye Reagent Concentrate (Bio-Rad) on protein sample dilutions in MilliQ water. Optical rotations were recorded on a Perkin-Elmer Model-343 digital polarimeter (PerkinElmer. Inc.). The ultraviolet–visible spectra were recorded on Shimadzu UV-3600 Plus UV-VIS-NIR spectrophotometer. Infrared spectra were recorded on a Thermo Fisher FTIR 200 spectrophotometer. NMR spectra were recorded on Varian 400 MHz, Bruker 400 MHz NMR (ARX400) and Bruker 600 MHz AVIIIHD spectrometers at ambient temperature with $CDCl_3$, CD_3OD, DMSO-d_6 and acetone-d_6 as the solvents. Chemical shifts are reported in parts per million and coupling constants are recorded in hertz. High-resolution electrospray ionization (ESI) MS data were measured using an Agilent Technologies 6520 Accurate-Mass Q-TOF LC/MS spectrometer (Agilent Technologies). The absorption at 450 nM was recorded in a NanoPhotometer (IMPLEN). Photocross-linking experiment was conducted by UVP CL-1000 ultraviolet crosslinker (Analytik Jena) at a wavelength of 365 nm with an intensity of 1,000 W for 1 h.

Chemicals. Chemical reagents were purchased from Sigma-Aldrich, J & K Scientific Ltd and Inno Chem Science & Technology Co., Ltd. Moracin C (**3**) and dienophile **15** were purchased from BioBioPha Co., Ltd (Kunming, China). KOD-Plus DNA polymerase was purchased from Toyobo Biotech Co., Ltd. Primer synthesis and DNA sequencing were conducted at the Tsingke Biotech Company. Restriction enzymes and DNA ligase were purchased from Takara Biotechnology Co. Ltd.

Chemical synthesis. Chemical synthesis of different dienophiles (**2**, **5**–**8** and **16**), the acetylated precursors of dienes **10**–**13** and deuterated diene **10** are depicted in detail in the Supplementary Information. Dienophile **14** was synthesized according to a literature method[35].

Chemical analyses. The analyses of enzymatic reaction mixtures and the determination of the conversion rates and kinetic parameters were performed on an Agilent 1200 series HPLC system (Agilent Technologies) coupled with an LCQ Fleet ion trap mass spectrometer (Thermo Electron Corp.) equipped with an ESI source. The LC chromatograph was equipped with a quaternary pump, a diode-array detector, an autosampler, a column compartment and a Shiseido Capcell Pak C_{18} MG III column (250 mm × 4.6 mm internal diameter (i.d.), 5 μm (Shiseido Co., Ltd)). The mobile phase consisted of a gradient elution of solvents A (that is, 0.1% (v/v) formic acid aqueous solution) and B (methanol) at a flow rate of 1.0 ml min^{-1} at 30 °C. The gradient programmes were 40–70% B, 0–15 min, 70–95% B, 15–40 min, 95–40% B, 40–41 min and 40% B, 41–46 min. For the HPLC–MS analyses, ultrahigh-purity helium was used as the collision gas, and high-purity nitrogen (N_2) was used as the nebulizing gas. The optimized ESI source parameters were: sheath gas flow rate, 20 arbitrary units (a.u.); auxiliary gas flow rate, 5 a.u.; spray voltage, 5.0 kV; capillary temperature, 350 °C; source collision-induced decomposition, 35 V; tube lens offset voltage, −75 V. The spectra were recorded in the 100–1,000 m/z range for a full-scan MS analysis. The split ratio of effluent from the LC to ion source was 2:1. The data were analysed using XCalibur software.

Activity-guided enzyme purification. The reaction mixture, which contained 20 mM Tris-HCl, pH 7.5, 100 μM morachalcone A (**2**) and 100 μM moracin C (**3**) as substrates and 9.8 μg of crude cell lysate in a final volume of 100 μl was incubated at 30 °C for 1 h. The reactions were terminated by the addition of 200 μl of ice-cold MeOH and were centrifuged at 15,000g for 30 min. The supernatants were analysed by HPLC–MS. Chromatography was performed using an ÄKTA prime plus (GE Healthcare). Three parallel assays for enzyme activity were routinely performed.

The fresh *M. alba* cell cultures[42] (200 g) was added into lysis buffer (400 ml) that consisted of 50 mM sodium phosphate, pH 7.4, 1 mM EDTA, 3 mM 2-mercaptoethanol and 100 mM phenylmethanesulfonyl fluoride at a ratio of 1:100 (v/v) and treated with a Waring blender at 4 °C. The mixture was centrifuged at 9,000g at 4 °C for 30 min, and solid AS was added to the protein sample to 80% saturation. After gentle agitation at 4 °C for 12 h, the mixture was centrifuged at 9,000g at 4 °C for 30 min. The resulting pellet was resuspended in a buffer that contained 20 mM Tris-HCl, pH 7.4, 2 mM EDTA and 3 mM 2-mercaptoethanol, and then centrifuged at 160,000g at 4 °C for 2 h. The supernatant was collected and concentrated using a centrifugal concentrator with Amicon Ultra-30K (Millipore), and the resulting solution (AS fraction) was used as a stock solution for the enzymatic assays.

Then, the AS fraction was loaded onto a Hitrap Butyl FF (GE Healthcare) column (5 ml) equilibrated with a 50 mM sodium phosphate buffer, pH 7.0, that contained 1.5 M AS. Protein elution was performed stepwise at a flow rate of 2 ml min^{-1} using a 50 mM sodium phosphate buffer, pH 7.0, with a gradient of 0–20 min at 0% (v/v), 20–70 min at 20% (v/v) and 70–120 min at 100% with respect to the equilibration buffer. the eluted fractions were tested for enzyme activity.

The active fractions (HIC fractions) were pooled, and the buffer was exchanged to 20 mM Tris-HCl, pH 8.0, and loaded onto a HiTrap Q FF (5 mL) column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 8.0. Protein elution was performed at a flow rate of 2 ml min^{-1} using a buffer that contained 20 mM Tris-HCl, pH 8.0, and 1 M NaCl with a gradient of 0–20 min at 0% (v/v), 20–40 min at 10% (v/v), 40–60 min at 20% (v/v) and 60–100 min at 100% with respect to the loading buffer. The eluted fractions were tested for activity.

The active fractions from HiTrap Q chromatography (the IEC fractions) were concentrated and fractionated using Superdex 200 Increase 10/300 GL columns (GE Healthcare) connected in series. Isocratic protein elution was performed on a NGC^{T} chromatography system (Bio-Rad) at a flow rate of 0.25 ml min^{-1} using a buffer that contained 20 mM Tris-HCl, pH 7.2 and 0.15 M NaCl. the eluted fractions were tested for activity by HPLC to afford the active fraction (SEC fraction).

Photoaffinity labelling and pull down. To the PBS buffer (38 μl) were added 10 μl of the SEC fraction (1 mg ml^{-1}), 1 μl of **3** (2.5 mM) and 1 μl of probe **8** (2.5 mM). For the negative control, 1 μl of DMSO was used instead of probe **8**. The reaction mixture was irradiated at a 365 nm wavelength in ice for 1 h with an intensity of 1,000 W. The resulting mixture was conjugated with a biotin-enrichment tag by click chemistry (1 mM biotin–azide, 1 mM $CuSO_4$, 0.1 mM BTTAA and 10 mM L-ascorbic acid sodium salt for 2 h at room temperature). After conjugation, the proteins were precipitated by adding prechilled acetone (200 μl). The precipitated mixture was re-dissolved in 50 μl of PBS buffer that contained 1% SDS, and then added to 500 μl of PBS. The resulting solution was incubated with 100 μl of streptavidin agarose (Invitrogen) overnight at 4 °C. The beads were then washed with PBS (600 μl) six times and then heated to 95 °C for 10 min in the SDS loading buffer. The bead-bound proteins were separated by SDS–PAGE (8%) and visualized by silver staining (Beyotime, P0017S).

LC–MS/MS analysis of proteins. The enriched bands in the SEC fraction and the 'pull-down assay' were analysed following the protocol described here. Protein bands on the SDS–PAGE gel were destained and in-gel digested with sequencing-grade trypsin (10 ng μl^{-1} trypsin, 50 mM ammonium bicarbonate, pH 8.0) overnight at 37 °C. Peptides were extracted with 5% formic acid/50% acetonitrile and 0.1% formic acid/75% acetonitrile sequentially and then concentrated to ~20 μl. The extracted peptides were separated by an analytical capillary column (50 μm × 10 cm) packed with 5 μm spherical C18 reversed phase material (YMC). A Waters nanoAcquity UPLC system (Waters) was used to generate the HPLC gradient 0–30% B in 60 min and 30–70% B in 15 min (A = 0.1% formic acid in water and B = 0.1% formic acid in acetonitrile). The eluted peptides were sprayed into an LTQ ORBITRAP Velos mass spectrometer (ThermoFisher Scientific) equipped with a nano-ESI ion source. The mass spectrometer was operated in data-dependent mode with one MS scan followed by ten high-energy collisional dissociation MS/MS scans for each cycle. Database searches were performed on an in-house Mascot server (Matrix Science Ltd) against the proteome of *M. notabilis*[48] (download from the National Center for Biotechnology Information). The search parameters were 10 ppm mass tolerance for the precursor ions, 0.3 Da mass tolerance for the product ions and two missed cleavage sites were allowed for trypsin digestion. Methionine oxidation was set as a variable modification. The search results were filtered with both a peptide significance threshold and an expectation value of less than 0.05. The proteins identified by LC–MS/MS analysis are listed in Supplementary Tables 1 and 2, respectively.

RNA sequencing and cloning of *MaMO* and *MaDA*. Ten-day-old *M. alba* cell cultures were treated with 0.1 mM methyl jasmonate for 20 h. After that, the sample was sent to the Beijing Genomics Institute for transcriptome analysis and the total RNA was subsequently extracted from it with an EZNA Plant RNA kit (Omega Bio-tek Inc.) and reverse-transcribed using a SMARTer RACE cDNA Amplification kit (Clontech Inc.). The reverse-transcribed products were subjected to rapid amplification of the cDNA ends (RACE) according to the manufacturer's protocol. The 3′ ends of *MaMO* and *MaDA* were obtained using the primers MaMO-3′GSP and MaDA-3′GSP, respectively. The primer MaMO-3′GSP was specific for the sequence of Unigene ID CL1657.Contig4 in RNA-sequencing data, and MaDA-3′GSP was specific for the Unigene ID CL4729.Contig2 sequence. For amplification of the full-length coding sequence of *MaMO*, a forward primer MaMO-Fw and a reverse primer MaMO-Rv were prepared. The full-length *MaDA* coding sequence was acquired using the gene-specific primer pairs MaDA-Fw and MaDA-Rv.

Heterologous expression and purification of *MaMO*. The full-length *MaMO* with a C-terminal His tag was obtained by PCR using cDNA as a template with KOD-Plus-Neo (Toyobo) and the primer set MaMO-F/R. The pPIC3.5K (Invitrogen) was double digested using *Eco*RI and *Not*I (Takara Bio Inc.) and purified from agarose gel using Universal DNA Purification Kit (Tiangen Biotech. Co., Ltd). Insertion of *MaMO* into the linearized vector pPIC3.5K was performed with an In-Fusion HD Cloning Kit (Clontech Inc.) according to the manufacturer's protocol. The sequence was confirmed by Sanger sequencing. The expression vector was designed to express the recombinant protein with a C-terminal His tag in *K. phaffii* in the presence of methanol.

The pPIC3.5K expression vector, which contained the full-length *MaMO*, was introduced to *K. phaffii* SMD1168 (Invitrogen) by using a Frozen-EZ Yeast Tarnsformation II kit (Zymo Research). In vivo screening of multiple inserts was performed through the examination of the G418 (Invitrogen)-resistant level for each recombinant. His^+ recombinants for *MaMO* (96) were grown to approximately the same cell density by successive inoculations and spotted on yeast extract peptone dextrose plates that contained G418 at final concentrations of 0, 0.25, 0.5, 0.75, 1.0, 1.5, 1.75, 2.0, 3.0 and 4.0 mg ml^{-1}. The colonies resistant to G418 at a high concentration were chosen for protein expression. A single colony was inoculated into 50 ml Erlenmeyer flasks, added to 10 ml of BMGY medium that contained 100 mM potassium phosphate buffer, pH 6.0, 1% yeast extract, 2% peptone, 1.34% yeast nitrogen base, 4×10^{-5}% biotin and 1% glycerol, and cultivated at 28 °C, 220 r.p.m. for 12 h. The cultures were harvested after 12 h and centrifuged at 3,000*g* for 5 min to collect the *K. phaffii* strains that harboured the *MaMO* gene. The described strain was precultured in a large volume (1 l) of BMMY medium that contained 100 mM citric acid–sodium citrate buffer, pH 5.5, 1% yeast extract, 2% peptone, 1.34% yeast nitrogen base, 4×10^{-5}% biotin and 1% methanol and shaken at 18 °C, 100 r.p.m. for 6 days, and protein expression was induced by adding 10 ml of methanol every 24 h. The culture was harvested after 6 days and centrifuged at 6,000*g* at 4 °C for 15 min, and the supernatant was concentrated by using a centrifugal concentrator at 4,000*g* at 4 °C for 30 min. The protein sample was diluted 20-fold with binding buffer that contained 20 mM phosphate buffer, pH 7.4, 0.5 M NaCl and 20 mM imidazole. The soluble fraction was passed through a 0.45 μM syringe filter unit and the cleared supernatant was immediately applied to 1 ml of Ni-NTA resin (GE) loaded in a column, which was pre-equilibrated with binding buffer. The resin was subsequently eluted with 10 ml (10 ten column volumes) of binding buffer. Elution was carried out with 10 ml (10 column volumes) of a different gradient elution buffer (20 mM phosphate buffer, 0.5 M NaCl, 50–200 mM imidazole, pH 7.4) at a flow rate of 1 ml min^{-1} at 4 °C. The proteins were concentrated and buffer exchanged to buffer A, which contained 20 mM Tris-HCl buffer, pH 7.4. Protein purity was confirmed by SDS–PAGE to be >90% (Supplementary Fig. 8). The approximate protein yield for MaMO was 2.0 mg l^{-1} and stored at 4 °C.

Heterologous expression and purification of *MaDA*. The truncated N terminus of *MaDA* without a signal peptide (predicted by PROTTER, http://wlab.ethz.ch/protter/start/) was obtained by PCR using cDNA as the template with KOD-Plus-Neo and primer set MaDA-pInsect-F/R. PI-secSUMOstar was linearized using BamHI-HF (NEB) and XbaI (NEB), and purified from agarose gel using a Universal DNA Purification Kit. A ClonExpress II One Step Cloning Kit (Vazyme Biotech Co., Ltd) was used to generate the MaDA-pI-secSUMOstar construct. The sequence was confirmed by Sanger sequencing and then transformed to DH10Bac competent cells.

MaDA was expressed using the Bac-to-Bac system. Recombinant baculoviruses were generated and amplified using the Sf21 insect cells (maintained in the SIM SF medium (Sino Biological Inc.). Recombinant *MaDA* was expressed as a secreted protein in High Five (Hi5) cells with an N-terminal 6×His-SUMO fusion tag. Hi5 cells were infected with recombinant virus (multiplicity of infection of 2) at a density of 1.5–2.0 × 10^6 cells ml^{-1}. The medium was harvested 48 h after the infection and concentrated using a Hydrosart Ultrafilter (Sartorius), then exchanged into the binding buffer, which contained 25 mM Tris-HCl, pH 8.0 and 200 mM NaCl. The protein was purified using the Ni-NTA resin following same procedure as for *MaMO* and stored at 4 °C. The approximate protein yield for MaDA was 3.0 mg l^{-1}.

Activity assay, biochemical property and kinetics analyses of MaMO. A reaction mixture that contained 20 mM Tris-HCl, pH 8.0, and 100 μM **3** as substrates and 17 nM purified protein in a final volume of 100 μl was incubated at pH 8.0 and 50 °C for 5 min. The reactions were terminated by the addition of 200 μl of ice-cold methanol and were centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above. For quantification, three parallel assays were routinely carried out.

For the investigation on the effects of pH, temperature and divalent metal ions on MaMO activity, enzymatic reactions were performed in reaction buffers with various pH values in the ranges 4.0–6.0 (citric acid–sodium citrate buffer), 6.0–8.0 (Na_2HPO_4–NaH_2PO_4 buffer), 7.0–9.0 (Tris-HCl buffer) and 9.0–11.0 (Na_2CO_3–$NaHCO_3$ buffer). To assay the optimal reaction temperature, the reactions were incubated at different temperatures (20–70 °C). To test the necessity of divalent metal ions for MaMO activity, $BaCl_2$, $CaCl_2$, $CoCl_2$, $CuCl_2$, $FeCl_2$, $MnCl_2$, $ZnCl_2$, $NiSO_4$ and EDTA were used individually in a final concentration of 5 mM. All the assays were performed with 17 nM MaMO at 50 °C for 5 min. Aliquots were quenched with 200 μl of ice-cold methanol and centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above.

To determine the kinetic values of **3**, reactions were performed with **3** from 1 to 50 μM at pH 8.0 and 50 °C in a total volume of 100 μl that contained 17 nM MaMO. Aliquots were quenched with 200 μl of ice-cold methanol and centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above. Data fitting was performed using GraphPad Prism 8, and K_M, k_{cat} and k_{cat}/K_M values represent the mean ± s.d. of three independent replicates.

Activity assay and biochemical property analysis of MaDA. Diene **4** (10 mM) was generated by hydrolysis *in situ* from the acetylated precursor (for details, see Chemical synthesis in the Supplementary Information). A reaction mixture that contained 20 mM Tris-HCl, pH 8.0, 100 μM, diene **4** and 100 μM **2** as substrates and 2.7 nM purified protein in a final volume of 100 μl was incubated at pH 8.0 and 50 °C for 5 min. The reactions were terminated by the addition of 200 μl of ice-cold methanol and centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above. For quantification, three parallel assays were routinely carried out.

For the investigation on the effects of pH, temperature and divalent metal ions on MaDA activity, enzymatic reactions were performed in reaction buffers with various pH values in the ranges 4.0–6.0 (citric acid–sodium citrate buffer), 6.0–8.0 (Na_2HPO_4–NaH_2PO_4 buffer), 7.0–9.0 (Tris-HCl buffer) and 9.0–11.0 (Na_2CO_3–$NaHCO_3$ buffer). To assay the optimal reaction temperature, the reactions were incubated at different temperatures (20–70 °C). To test the necessity of divalent metal ions for MaDA activity, $BaCl_2$, $CaCl_2$, $CoCl_2$, $CuCl_2$, $FeCl_2$, $MnCl_2$, $ZnCl_2$, $NiSO_4$ and EDTA were used individually in a final concentration of 5 mM. All the assays were performed with 2.7 nM MaDA at 50 °C for 5 min. Aliquots were quenched with 200 μl of ice-cold methanol and centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above.

Kinetic analysis of MaDA using 2 as a substrate. To determine the kinetic values of **2**, reactions were performed with **2** from 5 to 200 μM with the saturating **4** (1,000 μM) at pH 8.0 and 50 °C in a total volume of 100 μl that contained 2.7 nM MaDA. Aliquots were quenched with 200 μl of ice-cold MeOH and centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above. Data fitting was performed using GraphPad Prism 8, and K_M, k_{cat} and k_{cat}/K_M values represent the mean ± s.d. of three independent replicates.

Kinetic analysis of MaDA using diene 10 and dienophile 2 as substrates. To determine the kinetic values of **2**, reactions were performed with **2** (5, 15, 25, 35, 50, 75, 100, 150, 200 and 300 μM) at the saturating **10** (1.5 mM) at pH 8.0 and 50 °C in a total volume of 100 μl that contained 270 nM MaDA for 5 min. Meanwhile, to determine the kinetic values of **10**, reactions were performed with **10** (5, 10, 30, 50, 75, 100, 150, 250, 400, 600 and 800 μM) at the saturating **2** (4.0 mM) at pH 8.0 and 50 °C in a total volume of 100 μl that contained 270 nM MaDA for 5 min. Aliquots were quenched with 200 μl of ice-cold methanol and centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above. Data fitting was performed using GraphPad Prism 8, and K_M, k_{cat} and k_{cat}/K_M values represent the mean ± s.d. of three independent replicates.

Investigation for the substrate scope of MaDA. Reaction mixtures that contained 20 mM Tris-HCl, pH 8.0, 200 μM diene **4** and 100 μM dienophile **2**, **5**, **6**, **7**, **14**, **15** or **16** as substrates, 540 nM purified MaDA and 100 μM *p*-methoxyacetophone as the internal standard in a final volume of 100 μl were incubated at 50 °C for 5 min. Corresponding negative controls were carried out in the same conditions without adding MaDA. The reactions were terminated by the addition of 200 μl of ice-cold methanol and centrifuged at 15,000*g* for 30 min. Supernatants were analysed by analytical reverse-phase HPLC as described above. The conversion rates of the dienophiles were calculated from the peak areas of each dienophile in the enzymatic reaction mixture and its negative control.

To determine the conversion rates of different dienes, the corresponding precursors were hydrolysed to generate different dienes in situ (following same procedure as for synthesizing **4** (Supplementary Information) and then immediately used in the following assay: reaction mixtures that contained 20 mM Tris-HCl, pH 8.0, 200 μM **2** and 100 μM freshly prepared diene **4**, **10**, **11**, **12** or **13** as the substrates, 540 nM purified MaDA and 100 μM *p*-methoxyacetophone as the internal standard in a final volume of 100 μl was incubated at 50 °C for 5 min. The same procedure as that described above was used to calculate the conversion rate of the dienes.

The MaDA-mediated chemoenzymatic syntheses of natural products, that is, chalcomoracin (**1**), guangsangon E (**17**), kuwanol E (**18**), kuwanon J (**19**), deoxyartonin I (**20**) and 18″-*O*-methychalcomoracin (**21**), are described in Supplementary Information.

DFT calculations. DFT computations were performed in Gaussian 09[49]. Molecular geometries were optimized using the ωB97X-D[50] functional and the 6-31G(d) basis set. A pruned (99,590) grid (specified with the keyword int=ultrafine) was used for optimizations to minimize orientational variations in the evaluation of thermochemical corrections[51]. Frequency calculations were performed at the same level of theory as for geometry optimization to characterize the stationary points as either minima (no imaginary frequencies) or first-order saddle points (one imaginary frequency). Thermal contributions to free energies were calculated from vibrational frequencies using the quasi-rigid rotor-harmonic oscillator approach of Grimme[52]. Intrinsic reaction coordinate calculations were performed to ensure that the first-order saddle points found were true transition states that connected the reactants and the products. Single-point energies were calculated with the ωB97X-D functional and the 6-311++G(d,p) basis set. Solvation effects were

incorporated using the CPCM[53] model with diethyl ether or water as the solvent. Molecular structure visualizations were rendered using CYLview[54]. Monte Carlo conformational searches were performed with the Merck molecular force field implemented in Spartan'16 to ensure that the lowest energy conformations were located. The calculated energies (kcal mol^{-1}) for DFT-optimized structures are summarized in Supplementary Table 5.

KIE experimental method. The α-secondary deuterium KIE at each site of interest was measured using the method of internal competition in three separate experimental trials. In each trial, site-specifically labelled and unlabelled diene **10** were mixed together (~0.75 mM) in 2 ml of Tris buffer (20 mM, pH = 8.0) that contained 2% DMSO and 0.9 mM dienophile **2**. *p*-Methoxyacetophenone (0.35 mM) was also included in the reaction mixture as an internal standard to standardize the observed ultra-performance liquid chromatography (UPLC) peak integrations. A 0.5 ml aliquot of this pre-enzyme solution was added to 0.5 ml of cold MeCN and stored at −20 °C to await HPLC purification of diene **10**. A solution of 132 μM MaDA (7.5 μl) was added to the remaining 1.5 ml of pre-enzyme solution. The reaction mixture was kept at room temperature (~ 25 °C) and aliquots were quenched by 0.5 ml of cold MeCN on ice at three different reaction times (less than 7 min). The resulting mixture was immediately centrifuged (12,000 r.p.m., 10 min) at 4 °C to remove MaDA, and analysed at a flow rate of 0.3 ml min^{-1} at 40 °C by ACQUITY UPLC–MS (Waters, ACQUITY UPLC BEH C18 column, 50 mm × 2.1 mm i.d., 1.7 μm) using a gradient elution of water (A) and MeCN (B) to determine the fractions of the reaction[55]. The gradient programmes were 20% B, 0–1 min, 20–58% B, 1–4 min, 58–60% B, 4–8 min, 60–95% B, 8–10 min, 95–20% B, 10–11 min, and 20% B, 11–12 min. A non-enzymatic cycloaddition reaction between **10** and **2** was not observed even at 60 °C for 6 h (Supplementary Fig. 9).

The diene **10** that remained in the four different aliquots was isolated using C18 reversed-phase HPLC (Waters, XBridge pre-C18 OBD, 150 mm × 19 mm i.d., 5 μm) with ultraviolet detection at 280 nM. The column was eluted with water (A) and MeCN (B) using the gradient 30–40% B, 0-1 min, 40–64% B, 1–12 min, 64–66% B, 12–15 min, 66% B, 15–17 min, 66–95% B, 17–18 min, 95–30% B, 18–19 min and 30% B, 19–20 min. A solvent blank was run interposed between each HPLC separation to prevent memory effects. The fractions that contained diene **10** were collected and concentrated in vacuo at 30 °C to remove the MeCN. The remaining mixtures were lyophilized and re-dissolved in a 1:1 mixture of MeCN and water (300 μl) for MS analysis.

The MS data of KIE experiments were collected by directly injecting the sample into the mass spectrometer (Waters Vion IMS-QTOF) at a rate of 20 μl min^{-1}. Mass spectrometer parameters were: capillary voltage, 3 kV; source temperature, 100 °C; desolvation temperature, 350 °C; cone gas, 50 l h^{-1}; desolvation gas, 600 l h^{-1}; scan time, 1 s. The scan was operated in the ESI positive mode. The lockspray was injected every 2 min using 100 pg μl^{-1} leucine encephalin ($[M+H]^+ = 556.2766\ m/z$). The total runtime was 6 min and signal intensities obtained from 2.5 min to 5.5 min were averaged to obtain a final spectrum for analysis. A solvent blank was injected interposed between each sample to prevent memory effects.

The $M+1$ peak (309.1120 m/z) and $M+2$ peak (310.1183 m/z) in each sample from the resulting mass spectra were integrated by summing the observed intensities in 0.0023 m/z increments over the entire peaks using Origin 2018 software (Supplementary Fig. 10). These two peak integrations were used to determine the ratio of deuterated/non-deuterated diene **10** in different samples. The equation $\mathrm{KIE} = k_H/k_D = \log(1-F)/\log[(1-F)R_{SF}/R_0]$ uses these ratios to determine the deuterium KIE at two different sites of **10**, where F is the fraction of reaction completed, R_{SF} is the ratio of deuterated/non-deuterated diene **10** at fraction F and R_0 is the ratio of deuterated/non-deuterated diene **10** in the pre-enzyme fraction. R_{SF}/R_0 at different fractions was determined based on the peak integrations[56]. The KIEs of 10-*1″*D at three different reaction times were 0.9878, 0.9941 and 1.0533, to give an average value of 1.0117 ± 0.0511. The KIEs of 10-4″D at three different reaction times were 0.9550, 0.9310 and 0.9462, to give an average value of 0.9441 ± 0.0172.

Crystallization of MaDA, data collection and structure determination. To prepare the protein sample for crystallization, MaDA was digested with tobacco etch virus protease to remove the N-terminal 6 × His-SUMO fusion tag. Untagged MaDA was further purified by anion exchange chromatography using a HiTrap Q column, followed by the SEC using a Superdex 200 10/300 GL column. MaDA (59 kDa) existed as a monomer with a molecular weight of ~30 kDa as the SEC indicated (Supplementary Fig. 11). Protein was concentrated to 10 mg ml^{-1}. Crystals were grown using a sitting drop vapour diffusion method at 18 °C. MaDA was crystallized in 0.2 M zinc acetate dihydrate, 0.1 M sodium cacodylate trihydrate, pH 6.5 and 18% PEG8000.

Crystals were cryoprotected in 0.2 M zinc acetate dihydrate, 0.1 M sodium cacodylate trihydrate, pH 6.5, 20% PEG8000 and 20% ethylene glycol, and flash frozen in liquid nitrogen. X-ray diffraction data were collected at the Shanghai Synchrotron Radiation Facility (beamline BL17U) with an X-ray wavelength of 0.9795 Å at a temperature of 100 K. The data were processed using HKL2000 (HKL Research). The structure was solved by a molecular replacement method using Phaser[57], with the THCA synthase structure (PDB ID 3VTE) as the search model and refined using Coot[58] and Phenix[59]. Of the residues, 96.3% were in the Ramachandran-favoured regions, with 3.3% in the Ramachandran allowed regions and 0.4% as outliers.

The volume of the binding pocket was calculated using the program POVME 2.0[60]. The coordinates of the following atoms were used as centres for POVME inclusion spheres: N5 of the FAD cofactor, CB of the residue Phe359, CB of the residue Phe375 and CB of the residue Phe416. Points were generated in POVME with a grid spacing of 1 Å using inclusion spheres of 9 Å radius around the atom positions. The volume was calculated using the contiguous option (with contiguous seed spheres of radius 4 Å centred at the same coordinates as those of the inclusion spheres).

Computational methods for docking. The product chalcomoracin (**1**) or substrates (**2** and **4**) were docked into the binding pocket of MaDA using UCSF DOCK 3.7[61]. The top-ranked 1,000 poses were generated for structural filtering and conformational clustering as previously described[62]. Briefly, the generated docking poses were filtered based on the three types of structural descriptors calculated for each docking pose, namely the number of hydrogen bonds, the number of buried carbon atoms and the hydrophobic contact. For D–A product **1** or substrates **2** and **4**, the unreasonable docking poses were automatically removed if the calculated values of any type of descriptors were below the averaged values throughout the entire conformational ensemble. The docking poses survived from the filtering step were clustered based on Kelley–Gardner–Sutcliffe penalty function, which is cutoff-free, particularly efficient in clustering diverse poses in global space in an unbiased manner to obtain dissimilar docking poses. Then, a more physically rigorous approach (MM-GB/SA) was applied to refine and rescore the survived docking poses[63].

For the best-scored binding pose, three independent 500 ns MD simulations were further performed with the Desmond software package[64] and three copies of the MD simulations were also carried out for *apo*-MaDA as controls (Supplementary Fig. 12). We used the OPLS-AA 2005 force field[65] parameter set for the protein and ligand and the TIP3P model for water[66]. The system was solvated in a 12 Å cubic water box and neutralizing counterions were added. To mimic the experimental assay conditions, a 0.15 M NaCl salt bath was introduced. The prepared systems were first minimized using 5,000 steps of the steepest descent algorithm and then equilibrated as follows. The system was heated from 0 to 310 K in the NPT ensemble over 100 ps with harmonic restraints of 10.0 kcal mol^{-1} • Å^{-2} on heavy atoms of protein and ligand, and initial velocities sampled from the Boltzmann distribution. Further equilibration was performed at 310 K with harmonic restraints on the protein and ligand starting at 5.0 kcal mol^{-1} Å^{-2} and reduced by 1.0 kcal mol^{-1} Å^{-2} in a stepwise fashion every 1 ns for a total of 5 ns of additional restrained equilibration. Production runs were then made for 50 ns duration in the NPT ensemble. The M-SHAKE algorithm[67] was applied to constrain all the bonds that involved hydrogen atoms with a time step of 2 fs. The short-range electrostatic and Lennard–Jones interactions were cut off at 9 Å. Long-range electrostatic interactions were computed by the particle mesh Ewald method[68].

Single-site-directed mutagenesis and activity assay of MaDA variants. MaDA variants (H116A, V177A, Y192A, I259A, F292A, F356A, N357A, L358A, L358E, F375A and R443A) were engineered by site-directed mutagenesis (QuikChange, Stratagene) using the primers in Supplementary Table 6. Expression and purification of all these variants were done according to the same protocol described for the WT MaDA. The enzymatic reactions were performed under optimal conditions with 2.7 nM MaDA and 10 μM **2** and **4**.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available with this article and its Supplementary Information, or are available from the corresponding authors upon reasonable request. The gene sequences of MaDA and MaMO as amplified from cell cultures of *M. alba* are deposited in GenBank, accession no. MK573629 and no. MK573628, respectively. The structural factor and coordinate of MaDA are deposited in the Protein Data Bank under ID 6JQH.

References

48. He, N. et al. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nat. Commun.* **4**, 2445 (2013).
49. Frisch, M. J. et al. *Gaussian 09* (Gaussian Inc., 2009).
50. Chai, J. D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).
51. Bootsma, A. N. & Wheeler, S. Popular integration grids can result in large errors in DFT-computed free energies. Preprint at https://doi.org/10.26434/chemrxiv.8864204.v5 (2019).

52. Grimme, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem. Eur. J.* **18**, 9955–9964 (2012).
53. Takano, Y. & Houk, K. N. Benchmarking the conductor-like polarizable continuum model (CPCM) for aqueous solvation-free energies of neutral and ionic organic molecules. *J. Chem. Theory Comput.* **1**, 70–77 (2005).
54. Legault, C. Y. CYLview 1.0b (University of Sherbrooke, 2009); http://www.cylview.org
55. Jeona, B.-S. et al. Investigation of the mechanism of the SpnF-catalyzed [4+2]-cycloaddition reaction in the biosynthesis of spinosyn A. *Proc. Natl Acad. Sci. USA* **114**, 10408–10413 (2017).
56. Wilde, T. C., Blotny, G. & Pollack, R. M. Experimental evidence for enzyme-enhanced coupled motion/quantum mechanical hydrogen tunneling by ketosteroid isomerase. *J. Am. Chem. Soc.* **130**, 6577–6585 (2008).
57. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst. D* **66**, 486–501 (2010).
59. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* **66**, 213–221 (2010).
60. Durrant, J. D., Votapka, L., Sørensen, J. & Amaro, R. E. POVME 2.0: an enhanced tool for determining pocket shape and volume characteristics. *J. Chem. Theory Comput.* **10**, 5047–5056 (2014).
61. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS ONE* **8**, e75992 (2013).
62. Peng, S.-M., Zhou, Y. & Huang, N. Improving the accuracy of pose prediction in molecular docking via structural filtering and conformational clustering. *Chin. Chem. Lett.* **24**, 1001–1004 (2013).
63. Jacobson, M. P. et al. A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367 (2004).
64. Bowers, K. J et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proc. ACM/IEEE Conference on Supercomputing (SC06)* 84 (ACM, 2006).
65. Banks, J. L. et al. Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.* **26**, 1752–1780 (2005).
66. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
67. Kräutler, V., Van Gunsteren, W. F. & Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **22**, 501–508 (2001).
68. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An *N*log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

Acknowledgements

We thank X.-F. Fu (Peking University) for providing the crucial instrument and technical support to us; L. Li (NIBS) and W. Zhou (Peking University) for assistance in LC–MS/MS proteomics analysis; T. Cai (NIBS) for assistance in the transcriptome analysis; J. Y. Xiao (Peking University) for assistance in the expression of MaDA in insect cells using a baculovirus expression system. We also thank H. C. Lam and J. H. Snyder for proofreading the manuscript. We are grateful to the staff members of Shanghai Synchrotron Radiation Facility (beamline BL17U) for their support during X-ray data collection. This work was financially supported by National Natural Science Foundation of China grants (21625201, 21661140001, 21961142010, 91853202 and 21521003 to X.L.); the National Key Research and Development Program of China (2017YFA0505200 to X.L.); the Beijing Outstanding Young Scientist Program (BJJWZYJH01201910001001 to X.L.); CAMS Innovation Fund for Medical Sciences (CIFMS-2016-I2M-3-012 and 2019-I2M-1-005 to J.D.); the Drug Innovation Major Project (2018ZX09711001-001-006 to J.D.); Key Project at Central Government Level for the Ability Establishment of Sustainable Use for Valuable Chinese Medicine Resources (2060302 to J.D.); JSPS A3 Foresight Program to H.O. and the Fundamental Research Funds for the Central Universities (2017PT35001 to J.D.). Computational support was provided by the Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase) under grant no. U1501501. This work was also inspired by the international and interdisciplinary environment of the JSPS Core-to-Core Program 'Asian Chemical Biology Initiative'. K.N.H. is grateful to the National Science Foundation (grant no. CHE-1764328) for financial support. Calculations were performed on the Hoffman2 cluster at the University of California, Los Angeles, and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (grant OCI-1053575). L.G. is supported in part by a Postdoctoral Fellowship of Peking-Tsinghua Center for Life Sciences.

Author contributions

X.L., J.D. and L.G. initiated the project. X.L., J.D., L.G., C.S., X.D. and R.W. conceived and designed the experiments, analysed the data and prepared the manuscript, with input from all the authors. L.G., X.L. and R.T. conducted the chemical synthesis. L.G. and She Chen performed the pull-down assay and LC–MS/MS analysis. C.S. and K.X. conducted the activity-based protein purification and the enzymology experiments. X.D. performed the X-ray crystallization and data analysis of MaDA. R.W. cloned the genes of MaDA and MaMO. Y.Z. and N.H. performed docking and MD studies. L.G., C.S., X.D. and Q.G. performed the protein mutagenesis and expression. C.L., A.M. and H.O. helped the transcriptome analysis. C.S. cultivated the cell cultures. Y.H. and M.S. provided important information from the enzymatic assay. Shuming Chen performed DFT calculations under the guidance of K.N.H. L.G. and L.Z. performed the KIE experiments. L.G., C.S., X.D. and R.W. contributed equally to this work. X.L., J.D. and L.H. managed the whole project.

Competing interests

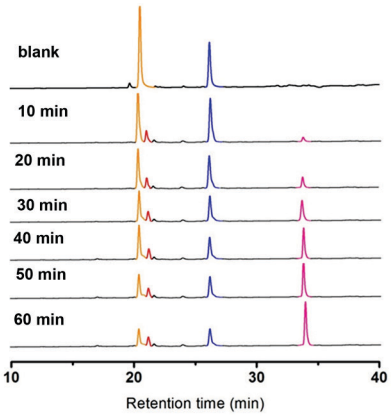
The authors declare no competing interests.

Additional information

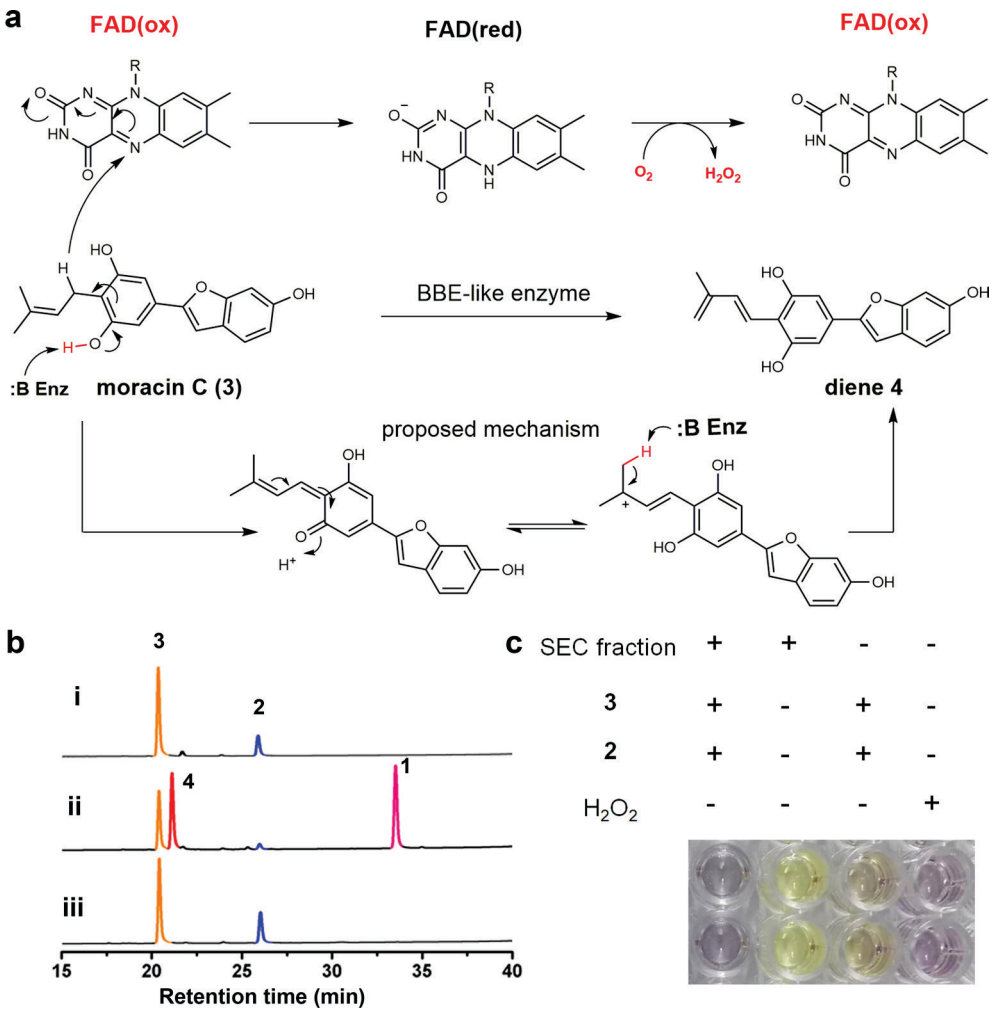
Extended data is available for this paper at https://doi.org/10.1038/s41557-020-0467-7.

Supplementary information is available for this paper at https://doi.org/10.1038/s41557-020-0467-7.

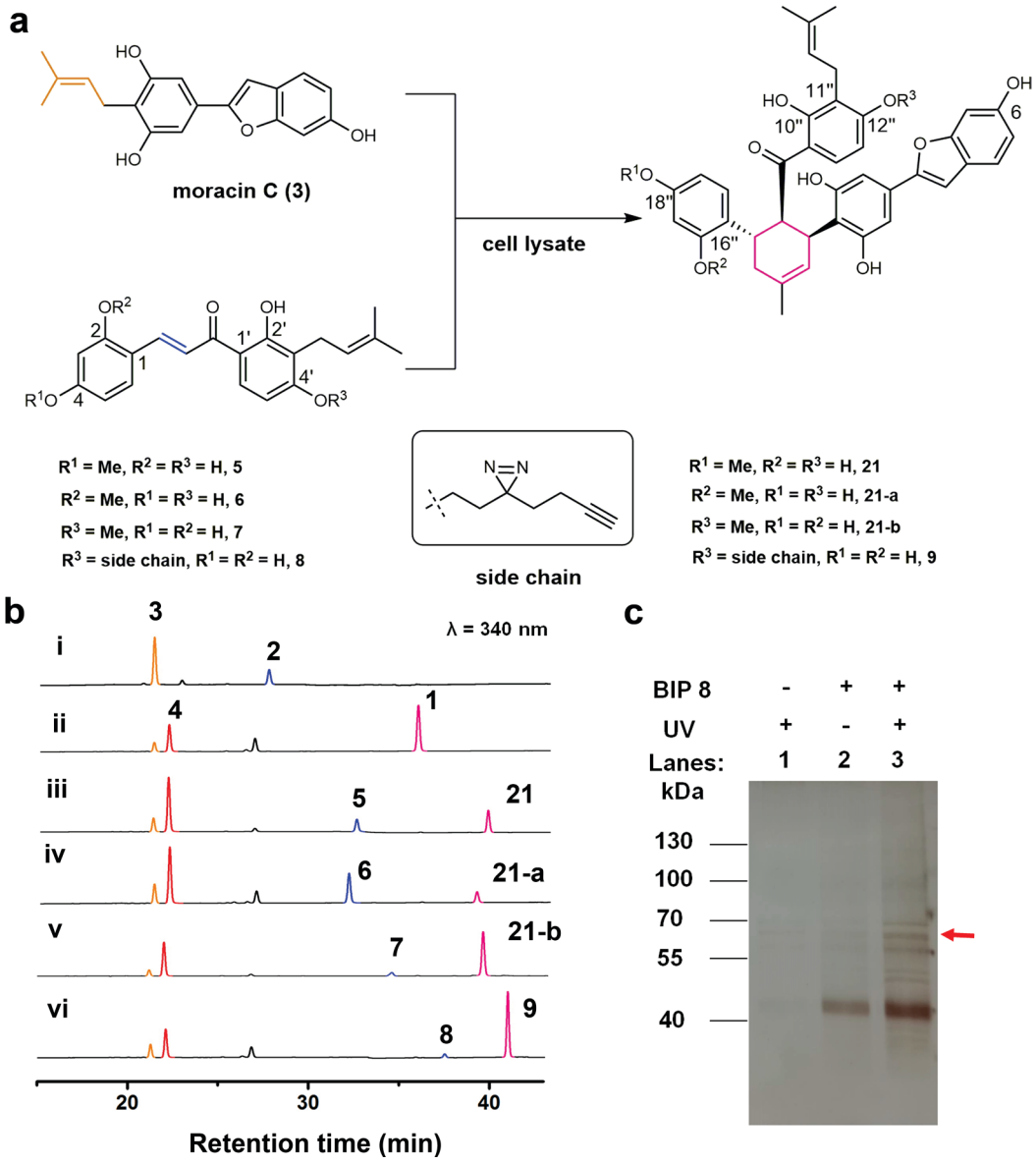
Correspondence and requests for materials should be addressed to L.H., J.D. or X.L.



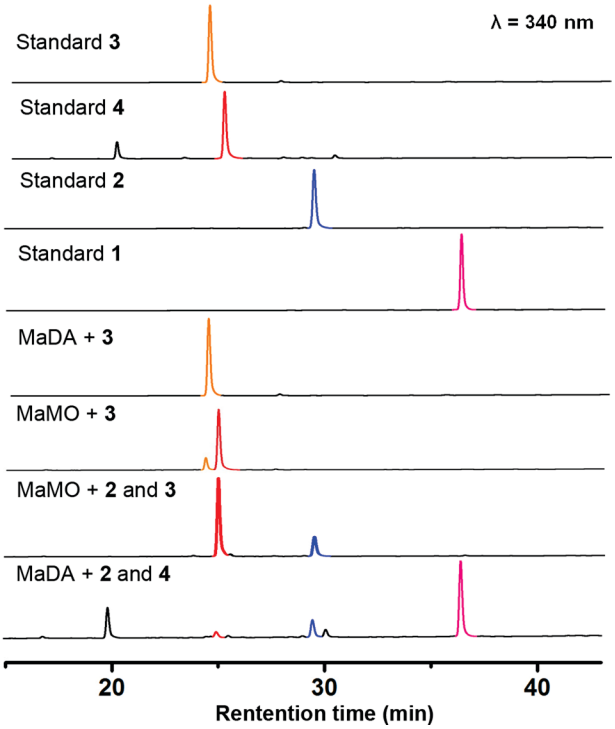
Extended Data Fig. 1 | The time dependent assay using SEC fraction of crude enzyme. The reaction mixture containing 20 mM Tris-HCl, pH 7.5, 100 mM morachalcone A (**2**), 100 μM moracin C (**3**) as substrates and 9.8 μg of crude cell lysate in a final volume of 100 μL was incubated at 30 °C for 1 h. The reactions were terminated by the addition of 200 μL of ice-cold MeOH and were centrifuged at 15,000 g for 30 min. The supernatants were analysed by the LC/MS. The experiments were repeated three times independently with similar results.



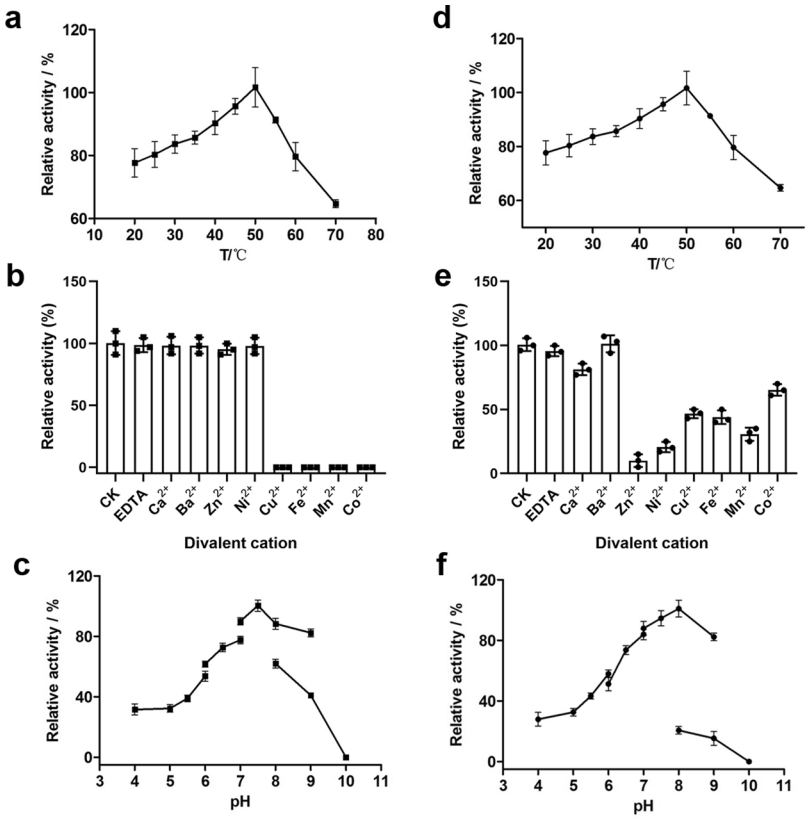
Extended Data Fig. 2 | The evidence suggesting BBE-like enzyme as the putative oxidase. a, The proposed mechanism for the formation of diene **4** catalysed by BBE-like enzyme. **b**, *In vitro* reaction analysis of **2** and **3** with or without dioxygen using SEC fraction: i) without SEC fraction (negative control), ii) with SEC fraction and dioxygen (positive control), iii) with SEC fraction but without dioxygen. The reaction buffer containing **2** and **3** was degassed in −78 °C and charged with argon, and then SEC fraction was added under argon atmosphere at room temperature. The experiments were repeated two times independently with similar results. **c**, Detection of hydrogen peroxide using Hydrogen Peroxide Assay Kit (Beyotime, S0038). The purple color indicated the existence of hydrogen peroxide in the reaction buffer. The experiments were repeated three times independently with similar results.



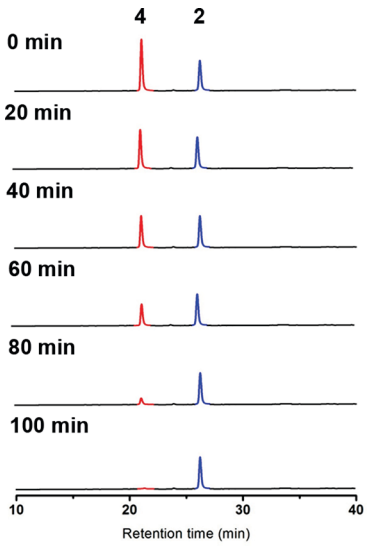
Extended Data Fig. 3 | The activities of the probe 8 and its analogues and the silver staining of pull-down assay using 8 as the photoaffinity probe. **a**, The structures of morachalcone A derivatives and the corresponding chalcomoracin derivatives. **b**, *In vitro* analysis using AS fraction (0.25 mg/mL)with different morachalcone A derivatives (100 μM): i) **2** and **3** without the AS fraction, ii) **2** and **3** with the AS fraction, iii) **5** and **3** with the AS fraction, iv) **6** and **3** with the AS fraction, v) **7** and **3** with the AS fraction, vi) **8** and **3** with the AS fraction. The experiments were repeated three times independently with similar results. **c**, The SEC fraction was incubated on ice with or without BIP **8** under irradiation of 365 nm UV light or without UV irradiation for 1 h. The following lysates were used for streptavidin-agarose pull-down assays, and the precipitates were resolved by 8% SDS-PAGE, followed by silver staining. The indicated bands were excised and subjected to LC-MS/MS proteomics analysis. The experiments were repeated two times independently with similar results.



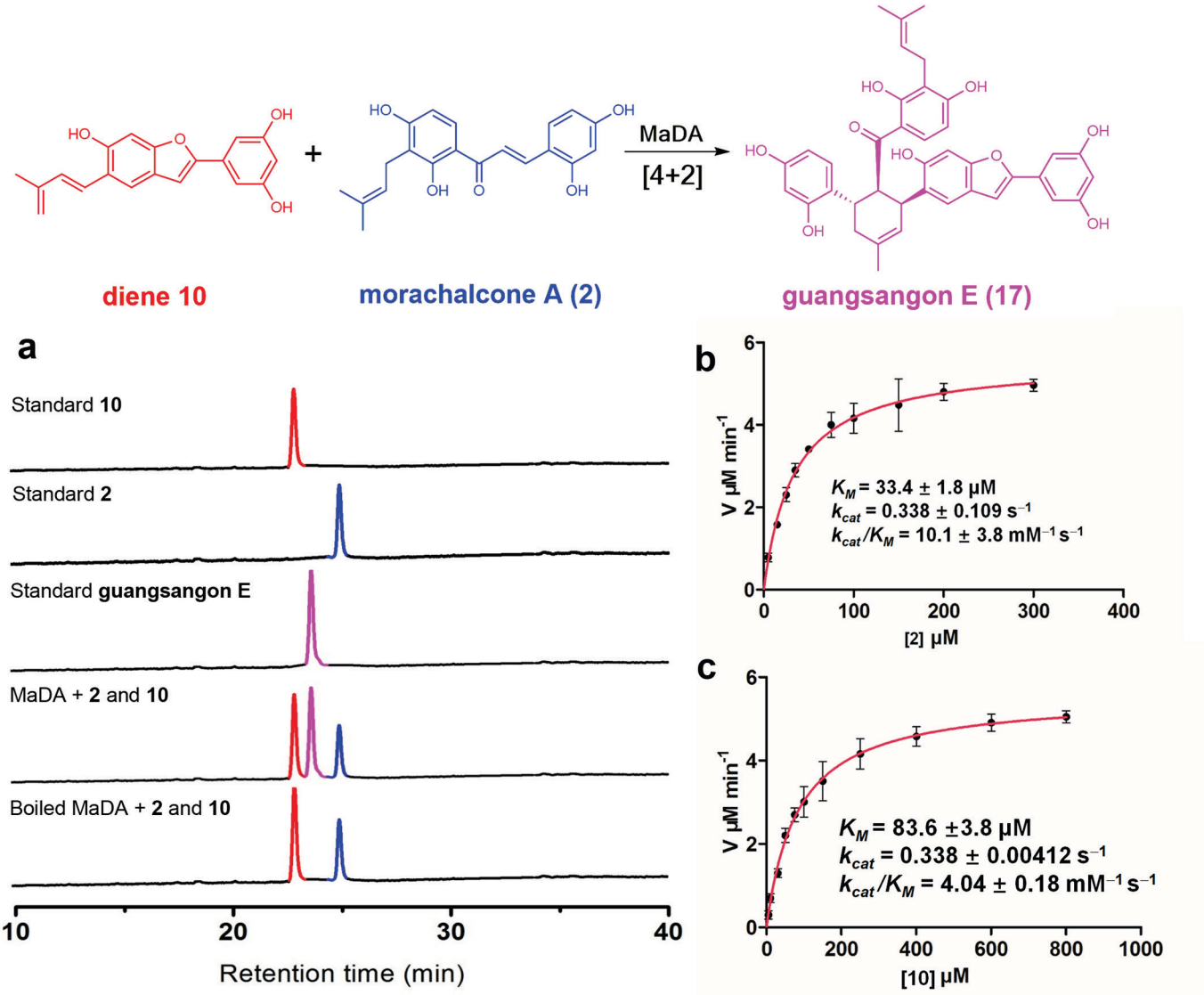
Extended Data Fig. 4 | *In vitro* reaction analysis of MaMO and MaDA. When MaDA was incubated with **3** alone, no new peak appeared, which revealed the MaDA did not have the oxidative function. In contract, when MaMO was used, diene **4** was formed. On the other hand, when MaMO was incubated with **2** and **3**, **3** was completely transformed to diene **4**, but no chalcomoracin or oxidative product of **2** was observed which indicated MaMO neither oxidized **2** nor catalysed the [4 + 2] cycloaddition between **2** and **4**. The experiments were repeated three times independently with similar results.



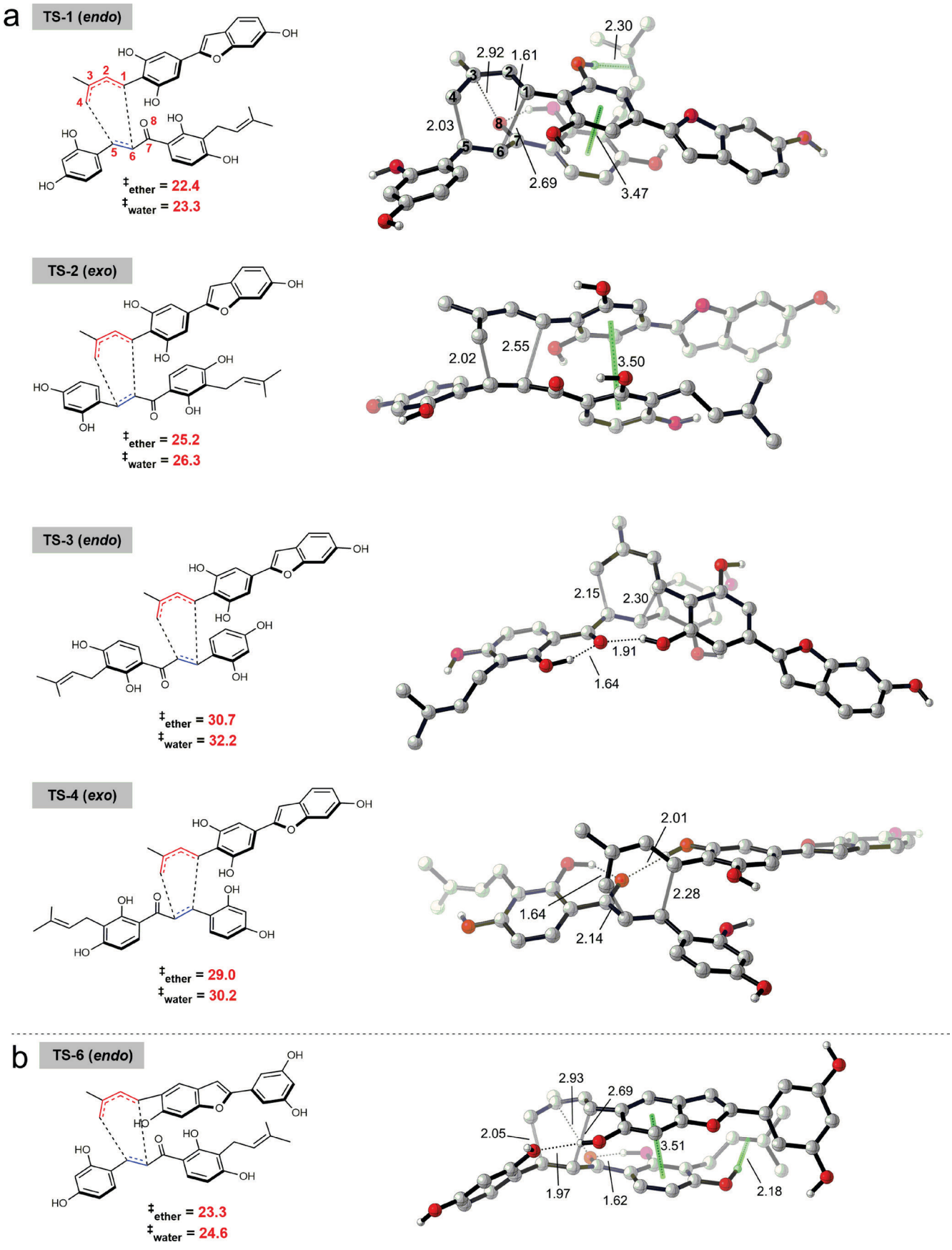
Extended Data Fig. 5 | Effects of pH, temperature and divalent metal ions on the activity of MaMO and MaDA. a, Effect of temperature on MaMO's activity. **b**, Effect of divalent metal ions on MaMO's activity. **c**, Effect of pH value on MaMO's activity. **d**, Effect of temperature on MaDA's activity. **e**, Effect of divalent metal ions on MaDA's activity. **f**, Effect of pH value on MaDA's activity. CK means 'control check'. Enzyme activity values represent mean ± standard deviation (s.d.) of three independent replicates.



Extended Data Fig. 6 | The time course for the incubation of morachalcone A (2) and 4 without MaDA. The reaction mixture containing 20 mM Tris-HCl, pH 8.0, 100 μM morachalcone A (**2**), 100 μM diene (**4**) as substrates without MaDA in a final volume of 100 μL was incubated at 50 °C for 100 min. The reactions were terminated by the addition of 200 μL of ice-cold MeOH and were centrifuged at 15,000 g for 30 min. The supernatants were analysed by the LC/MS. The experiments were repeated three times independently with similar results.

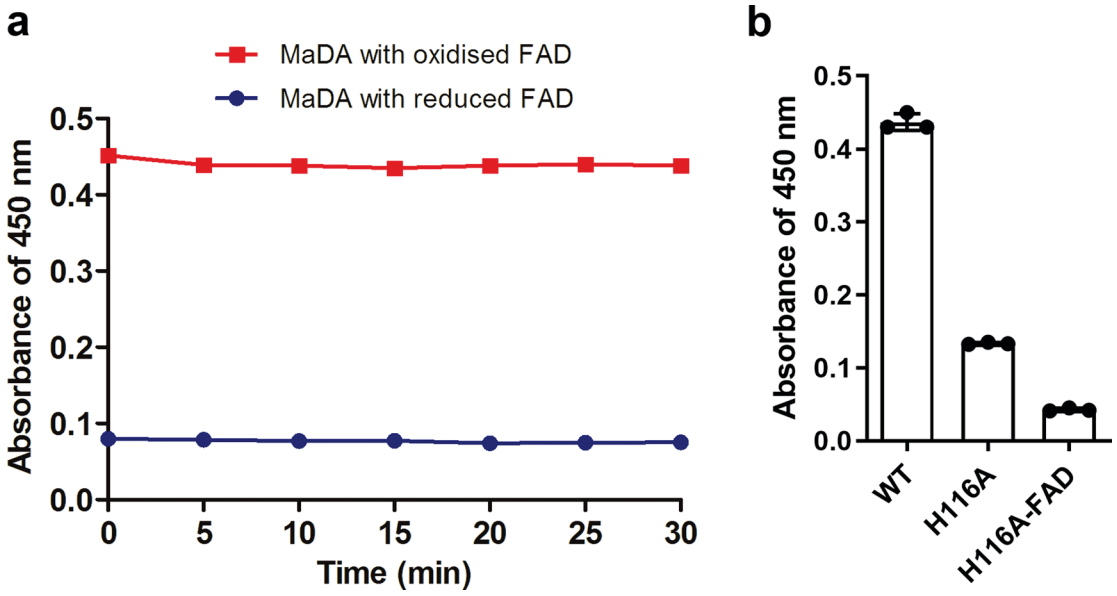


Extended Data Fig. 7 | Determination of the kinetic parameters of MaDA using stable diene 10 and morachalcone A (2). a, Enzymatic assay of **2** and **10** with MaDA. Compounds **2** and **10** were incubated with 270 nM MaDA or boiled MaDA for 5 min. The experiments were repeated three times independently with similar results. **b**, Kinetic parameters of MaDA for **2** (5–300 μM) using **10** (1.5 mM) as diene. **c**, Kinetic parameters of MaDA for **10** (5–800 μM) using **2** (4 mM) as dienophile. K_M, k_{cat} and k_{cat}/K_M values represent mean ± standard deviation (s.d.) of three independent replicates.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Calculated transition states of the Diels-Alder reaction between dienophile 2 and dienes 4 or 10. a, This data shows the calculated Diels-Alder transition states of Diels-Alder reaction between **2** and **4** leading to the four possible product regio- and stereoisomers. The computed barriers (all with respect to isolated reactants) show that TS-1 is the favored TS, suggesting that the reaction displays intrinsic regio- and stereoselectivity for the experimentally observed product isomer under enzyme catalysis. **b**, The DFT calculation supports an *endo* transition state and a concerted but asynchronous mechanism of the Diels-Alder reaction between **2** and **10**. C–H hydrogen atoms are omitted for clarity. Interatomic distances are in Å. Energies are in kcal/mol.



Extended Data Fig. 9 | Different absorbance data of MaDA at 450 nm at different conditions. a, The absorbance of MaDA (34 μM) at 450 nm was detected every 5 min with or without adding 6.6 mM sodium dithionite (about 200 eq), indicating the reduced form of MaDA is stable under excess of sodium dithionite for at least 30 minutes. The experiments were repeated three times independently with similar results. **b**, The absorbance of 450 nm of $MaDA^{wt}$, $MaDA^{H116A}$ and $MaDA^{H116A}$-FAD were detected at the same protein concentration (34 μM). The $MaDA^{H116A}$ was dialysed in 1 M KBr solution to remove the non-covalent FAD. Absorbance values of 450 nm represent mean ± standard deviation (s.d.) of three independent replicates.

Extended Data Table 1. Data collection and refinement statistics of MaDA-FAD

	MaDA
Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	74.66, 115.38, 95.87
α, β, γ (°)	90, 96.23, 90
Resolution (Å)	44.0-2.30(2.39-2.30) *
R_{merge}	0.12(0.59)
$I / \sigma I$	12.6(4.5)
Completeness (%)	92.1(92.2)
Redundancy	7.0(7.0)
Refinement	
Resolution (Å)	2.30
No. reflections	65799(6541)
R_{work} / R_{free}	0.186/0.227
No. atoms	
Protein	8122
Ligand/ion	106
Water	473
B-factors	28.8
Protein	28.5
Ligand/ion	19.3
Water	35.8
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.24

Extended Data Fig. 10 | Data collection and refinement statistics of MaDA-FAD.

nature research

Corresponding author(s): Luqi Huang, Jungui Dai, Xiaoguang Lei

Last updated by author(s): Feb 24, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	UCSF DOCK 3.6; XDS; CCP4;Gaussian 09 Revision D.01; Spartan 16; Desmond v4.6
Data analysis	UCSF Chimera 1.13.1; GraphPad Prism 8.0.2 (GraphPad Software); Chembiodraw 14.0, PyMol 2.07; Origin 2018; ExPASy ProtParam tool (https://web.expasy.org/protparam/); PROTTER server(http://wlab.ethz.ch/protter/start/);Phenix; CCP4; COOT;CYLView; XCalibur 2.1 W/ Foundation 1.0.2; Jalview 2.10.0 software

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The gene sequences of the MaDA and MaMO as amplified from cell cultures of M. alba are deposited in GenBank, accession no. MK573629 and no. MK573628, respectively. The structural factor and coordinate of MaDA have been deposited in the Protein Data Bank under the ID 6JQH.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Sample size of n=3 was selected in this study to conduct three biologically independent experiments to determine the activities, Km, kcat, Kcat/Km of MaMO, MaDA and its variants. These sample sizes are sufficient for the biochemical studies in this work.
Data exclusions	No data were excluded from this study.
Replication	All experiments were repeated independently and showed similar results. Number of repeats is given in the figure legends and table notes.
Randomization	No randomization was required because genetically and biochemically identical samples were used.
Blinding	No blinding was required because biochemical experiments were performed in this work by individual researchers.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The unique Morus alba cells in this study are available from the one of the corresponding authors, Jungui Dai, upon reasonable request; Komagataella phaffii SMD1168 (Invitrogen, USA); Sf21 and Hi5 insect cells (Invitrogen, USA).
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	All cell lines were negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	Commonly misidentified lines were not used in this study.