









Structure-based engineering of substrate specificity for pinoresinol-lariciresinol reductases

Ying Xiao^{1,8}, Kai Shao^{2,3,8}, Jingwen Zhou ^{4,5,8}, Lian Wang⁴, Xueqi Ma¹, Di Wu⁴, Yingbo Yang¹, Junfeng Chen¹, Jingxian Feng¹, Shi Qiu¹, Zongyou Lv¹, Lei Zhang⁶  , Peng Zhang   & Wansheng Chen ^{1,7} 

Pinoresinol-lariciresinol reductases (PLRs) are enzymes involved in the lignan biosynthesis after the initial dimerization of two monolignols, and this represents the entry point for the synthesis of 8-8' lignans and contributes greatly to their structural diversity. Of particular interest has been the determination of how differing substrate specificities are achieved with these enzymes. Here, we present crystal structures of *li*PLR1 from *Isatis indigotica* and pinoresinol reductases (PrRs) AtPrR1 and AtPrR2 from *Arabidopsis thaliana*, in the apo, substrate-bound and product-bound states. Each structure contains a head-to-tail homodimer, and the catalytic pocket comprises structural elements from both monomers. β 4 loop covers the top of the pocket, and residue 98 from the loop governs catalytic specificity. The substrate specificities of *li*PLR1 and AtPrR2 can be switched via structure-guided mutagenesis. Our study provides insight into the molecular mechanism underlying the substrate specificity of PLRs/PrRs and suggests an efficient strategy for the large-scale commercial production of the pharmaceutically valuable compound lariciresinol.

¹Research and Development Center of Chinese Medicine Resources and Biotechnology, Institute of Chinese Materia Medica, Shanghai University of Traditional Chinese Medicine, Shanghai, China. ²National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China. ³University of Chinese Academy of Sciences, Beijing, China. ⁴National Engineering Laboratory for Cereal Fermentation Technology and School of Biotechnology, Jiangnan University, Wuxi, Jiangsu, China. ⁵Science Center for Future Foods, Jiangnan University, Wuxi, Jiangsu, China. ⁶Department of Pharmaceutical Botany, School of Pharmacy, Naval Medical University (Second Military Medical University), Shanghai, China. ⁷Department of Pharmacy, Changzheng Hospital, Naval Medical University (Second Military Medical University), Shanghai, China. ⁸These authors contributed equally: Ying Xiao, Kai Shao, Jingwen Zhou. email: zhanglei@smmu.edu.cn; pengzhang01@cemps.ac.cn; chenwansheng@shutcm.edu.cn

Lignans are a major group of secondary metabolites in plants¹. This family has numerous biological effects in humans (e.g., anticancer², antiviral³, antioxidant⁴, and immunosuppression⁵) owing to their structural diversity—nearly 2000 distinct lignans have been reported. For example, the furfuran lignans such as kandelisesquigan A/B and terminaloside K have antioxidant effects^{6,7}. The dibenzylbutyrolactone lignans including arctigenin, traxillagenin, arctiin, traxillaside, and their glycosides have neuroprotective activities⁸. Finally, the aryltetralin lignan podophyllotoxin⁹ is the precursor for the semi-synthesis of anticancer drugs such as etoposide¹⁰.

Lignans biosynthesis starts with the coupling of two coniferyl alcohols by an oxidase (laccase or peroxidase) with the aid of a dirigent protein to form pinoresinol¹¹. Pinoresinol/lariciresinol reductase (PLR), an NADPH-dependent reductase, converts pinoresinol to lariciresinol and subsequently to secoisolariciresinol¹². Because the reductive steps that give rise to lariciresinol and secoisolariciresinol represent entry points for the biosynthesis of the lignan subclasses furofurans, dibenzylbutane, dibenzylbutyrolactone, and aryltetrahydronaphthalene¹³, PLR is regarded as a pivotal enzyme that contributes to lignan structural diversity. Moreover, variation in both the composition and accumulation of lignans among different plant species, organs, and developmental stages can be ascribed, at least in part, to the characteristics of reactions catalyzed by PLRs as well as their expression patterns¹⁴. Therefore, characterization of the catalytic mechanisms of PLRs—especially their substrate selectivity—is particularly crucial for understanding the molecular basis of the remarkable diversity of both chemical structures and biological activities of lignans.

The substrate selectivity of PLRs has attracted considerable attention¹². Most PLRs that have been characterized reduce both pinoresinol and lariciresinol efficiently to produce lariciresinol and secoisolariciresinol, respectively¹². Known exceptions are *Arabidopsis thaliana* reductases that have substrate preference for pinoresinol, but only weak (*AtPrR1*) or no activity (*AtPrR2*) toward lariciresinol and are thus named pinoresinol reductases (PrRs)¹⁵. A recent study indicated that the L174I mutant of *Camellia sinensis* PLR1 (*CsPLR1*) loses the capacity to reduce pinoresinol and specifically catalyzes the conversion of lariciresinol to secoisolariciresinol, but the underlying mechanism is unclear¹⁶. The three-dimensional structure of *Thuja plicata* PLR1 (*TpPLR1*) has been elucidated and indicates that K138 is responsible for the basic catalysis, because the mutant K138A lacks the ability to convert pinoresinol¹⁷. However, the apo structure does not provide sufficient information to interpret the substrate-selective mechanism of PLRs/PrRs.

Phylogenetic analysis of PLRs/PrRs from different species has revealed that *Isatis indigotica* PLR1 (*IiPLR1*), a key enzyme involved in lariciresinol biosynthesis, has the closest relationship to *AtPrRs* with a high level of amino-acid sequence identity (>80%)¹⁸ and is grouped with *AtPrR2*, which cannot utilize lariciresinol as substrate (Fig. 1). Interestingly, in contrast to *AtPrRs* that has substrate preference for pinoresinol, *IiPLR1* from *I. indigotica* (family Cruciferae, same as *A. thaliana*) can reduce both pinoresinol and lariciresinol efficiently with comparable k_{cat}/K_m values¹⁸. The finding that *IiPLR1/AtPrRs*, which differ in substrate selectivity, are clustered together suggests that substrate specificity is independent of sequence conservation among PLRs/PrRs. Therefore, the amino-acid residues responsible for PLR/PrR substrate selectivity are difficult to determine merely through sequence analysis, and thus structural information on PLR/PrR enzymes is vital—as are data concerning how these two enzyme types can utilize two different substrates.

In the present work, we compare crystal structures of *IiPLR1*, *AtPrR1*, and *AtPrR2* and identify residues that may be responsible for the observed substrate selectivity of PLRs and PrRs.

Mutagenesis of these residues alters the substrate specificities for pinoresinol and lariciresinol. For example, mutagenesis of *IiPLR1* successfully eliminates the second reaction that converts lariciresinol to secoisolariciresinol, leading to a high accumulation of the pharmaceutically valuable compound lariciresinol. Our study will enable the synthesis of lignans with diverse chemical structures and bioactivities by biotechnological means or by enzyme-assisted chemistry.

Results

Characterization of *IiPLR1*, *AtPrR1*, and *AtPrR2* crystal structures. To understand both the catalytic mechanism of PLR and the mechanism underlying the substrate specificity of PLRs/PrRs, we chose *IiPLR1*, *AtPrR1*, and *AtPrR2* for a structure study. The crystal structures were captured in the apo, substrate-bound and/or product-bound forms (Table 1). We found that, for all 16 structures we solved, each enzyme adopts a similar head-to-tail dimer conformation (Fig. 2a and Supplementary Fig. 1), strongly suggesting that each PLR/PrR functions as a homodimer, consistent with the literature that *TpPLR1* exists as a dimeric entity in solution¹⁷.

Taking the structure of *IiPLR1*_NAP_+PIN for the purpose of a detailed description, each protomer contains two domains, namely the N-terminal NADPH binding domain (NBD) and the C-terminal substrate binding domain (SBD). The NBD comprises seven β -strands (β 1–6, β 8) surrounded by six α -helices (α 1–5, α 7), whereas the SBD comprises two β -strands (β 7, β 9) with five small α -helices (α 6, α 8–11). A large groove is formed between NBD and SBD (Fig. 2a). This groove can be roughly divided into two parts—the positively charged part that associates with the NBD and the hydrophobic part that associates with the SBD. The substrates or products can be clearly defined within the groove (Fig. 2a and Supplementary Fig. 2). Several regions (α 5 loop, α 9-helix, and α 9 loop) are partially disordered both in the *IiPLR1*_apo and *IiPLR1*_NAP structures, for which the differences can be characterized by a RMSD of 0.287 Å, whereas the β 4 loop is well defined in the apo structure (Fig. 2b). Intriguingly, the β 4 loop is well defined in the *AtPrR1*_NAP structure but disordered in the *AtPrR1/2*_apo structures (Supplementary Fig. 3). These structural differences suggest that the loops are somewhat flexible, and act as a switch to control the binding of NADPH and release of NADP⁺. Moreover, the β 2 loop moves slightly towards NADP⁺ in the *IiPLR1*_NAP structure, catering to the entry of the coenzyme (Fig. 2b). NADP⁺ forms strong hydrogen bonds and hydrophobic interactions with residues within the groove (Fig. 2c), among which the GXXGXXG motif (considered as the conserved NADPH-binding motif) binds the phosphate and deoxyribose groups, and residues Ala164, Cys165 together with Phe166 fix the position of the catalytically active nicotinamide group. Residue Lys144, which corresponds to the previously reported Lys138 in *TpPLR1* that serves as the general base for catalysis, forms direct hydrogen bonds with NADP⁺ in *IiPLR1*_apo and *IiPLR1*_NAP.

Catalytic mechanism of PLR based on its homodimerization. The dimers of *IiPLR1*_NAP and *IiPLR1*_NAP_+PIN have similar structures, as suggested by a RMSD of 0.362 Å, indeed, even the NADP⁺ moieties could be aligned almost in the same position (Fig. 2d). Both β 4 loops are disordered, further implying their flexibility, but β 2 loop and α 10-helix from neighboring molecules of *IiPLR1*_NAP_+PIN dimer make contacts with and stabilize the substrate (Fig. 2d, e). Similar conformational changes of β 2 loop and α 10-helix can be seen by comparing the structure of *AtPrR1*_NAP_+PIN with that of *AtPrR1*_NAP and the structure of *AtPrR2*_NAP_+PIN with that of *AtPrR2*_apo

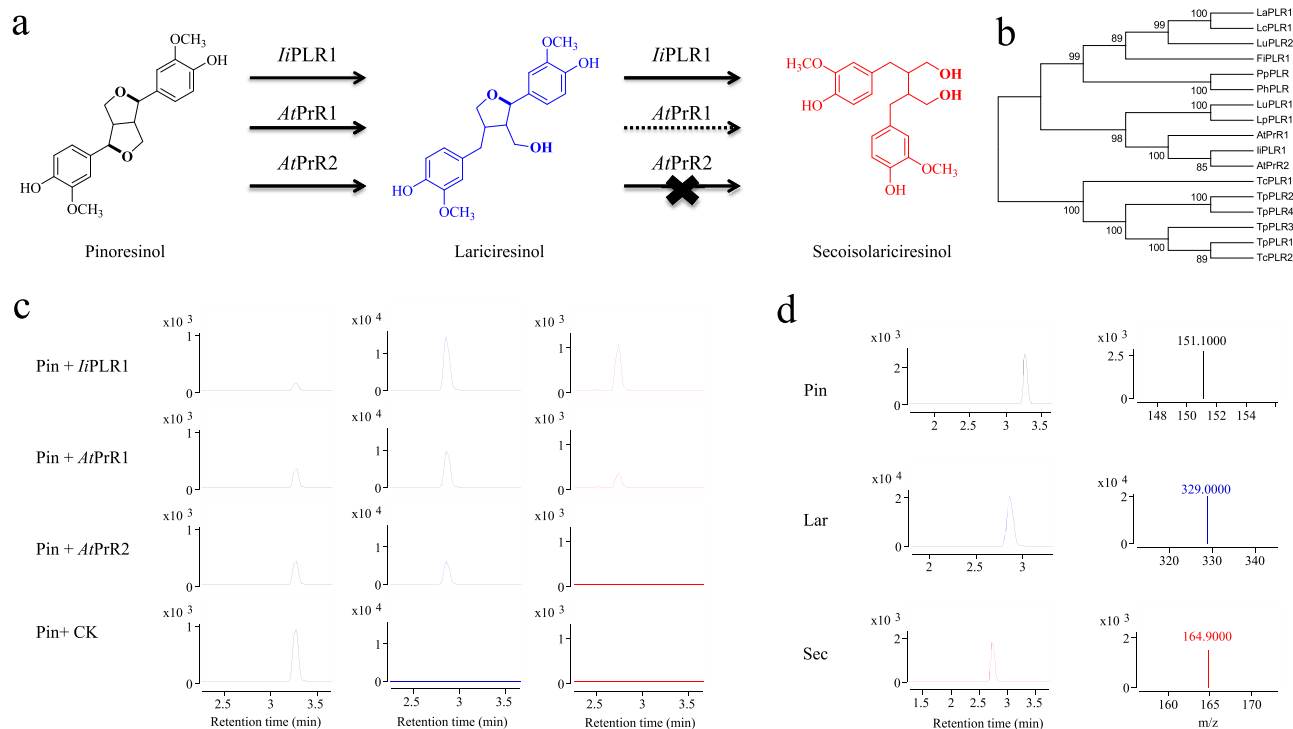


Fig. 1 Biochemical assays for *IiPLR1* and *AtPrRs* function. **a** *IiPLR1* efficiently catalyzes the conversion of pinoresinol into lariciresinol and also catalyzes the conversion of lariciresinol into secoisolariciresinol. In contrast, *AtPrR1/2* exhibit a substrate preference for pinoresinol, yet exhibit only weak activity (*AtPrR1*) or no activity (*AtPrR2*) for lariciresinol. **b** Phylogenetic tree of PLRs/PrRs from different species. **c** Conversion of pinoresinol into lariciresinol and then into secoisolariciresinol by recombinant *IiPLR1*, *AtPrR1*, and *AtPrR2*. The reaction products were analyzed by LC-MS. **d** Chromatograms for pinoresinol, lariciresinol, and secoisolariciresinol are denoted in black, blue, and red, respectively.

(Supplementary Fig. 4a, b). (+)-Pinoresinol is inserted as a straight chain deep into the hydrophobic groove, for which the hydrophilic ends are stabilized through the formation of hydrogen bonds with main-chain atoms of Met125 and Gly178, and the hydrophobic region is surrounded by a series of hydrophobic groups (Fig. 2e). The inner 2-methoxy-phenol group of (+)-pinoresinol forms a sandwich-like π - π stack comprising the nicotinamide head of NADP⁺ and Phe166. Two furan rings in the middle are surrounded by Tyr169 and Phe170 from α 6-helix and by His276 and Phe277 from α 10-helix. The outer 2-methoxy-phenol group is coordinated by Phe277 and Val46 of β 2 loop from a neighboring protomer, which is distant from the NADP⁺ and (+)-pinoresinol of the protomer (Fig. 2a, c, e). Further, Lys144 is far away from the furan rings, indicating that it may not participate in catalysis directly. Based on this structure analysis, we propose that both the entry and exit of NADPH are controlled by the β 4 loop of *IiPLR1*. Once one molecule of (+)-pinoresinol is captured by the narrow hydrophobic groove, each protomer forces the repositioning of the α 10-helix and β 2 loop in the other protomer, resulting in tight binding of the substrate. This allows H: transfer from the NADPH to the proximal furan ring of the substrate to produce one molecule of (+)-lariciresinol.

Regarding the mechanism of the second catalytic step, we further compared the structures of *IiPLR1*_{NAP}+PIN, *IiPLR1*_{NAP}-LAR and *IiPLR1*_{NAP}-SEC, which revealed a similar mode for substrate/product binding (Fig. 2f). Furthermore, Leu46 (corresponding to Val46 in *IiPLR1*), His276 and Phe277 of *AtPrR1* are positioned similar to the corresponding residues of *IiPLR1* to effect substrate binding or product release, except that the β 4 loops cover the substrate/product, which are disordered in the *IiPLR1* structures (Fig. 2d and Supplementary Fig. 4a, c, d). The importance of Val46 for catalysis in *IiPLR1* is underscored by data from a mutational analysis (Fig. 2g).

Mutation of Val46 to Ala improved the conversion of pinoresinol to lariciresinol by ~16%, and the subsequent conversion to secoisolariciresinol was greatly reduced. The *IiPLR1* mutant V46L had little ability to catalyze the conversion of lariciresinol to secoisolariciresinol. These data suggest that *IiPLR1* undergoes substrate-induced conformational changes upon homodimerization to achieve catalysis, and the principle of catalytic reactions using lariciresinol as substrate (second step) appears to be like that using pinoresinol as substrate (first step).

Mechanism underlying the substrate selectivity of PLR/PrR. A previous study reported that the recombinant *AtPrR1* can only weakly reduce lariciresinol whereas *AtPrR2* lacks activity, which is in sharp contrast to all known PLRs¹⁵. To determine the mechanism underlying this difference in substrate specificity, we confirmed the relative lack of activity for *AtPrR1/2* (Fig. 1) and then carried out a structure analysis of *IiPLR1*, *AtPrR1*, and *AtPrR2*. Each of *IiPLR1*_{NAP}+PIN/*AtPrR1*_{NAP}+PIN/*AtPrR2*_{NAP}+PIN forms a homodimer, and superimposition of the protomers among the three complexes revealed RMSDs of 0.374, 0.365, and 0.308 Å, respectively (Fig. 3a, c and Supplementary Fig. 1). In contrast to *IiPLR1*_{NAP}+PIN, the β 4 loops of *AtPrR1*_{NAP}+PIN and *AtPrR2*_{NAP}+PIN can be clearly identified (Fig. 3 and Supplementary Fig. 4). These well-defined loops twist as an “8” shape and cover both the NADP⁺-binding and substrate-binding grooves. Within the twisted loop, His93 and His97 “grasp” helices α 5 and α 10, while Val92 and Phe94 interact directly with (+)-pinoresinol; Arg95 strongly interacts with NADP⁺ as well as each of the GXXGXXG motif and α -helix from neighboring protomer of the dimer. Similar β 4 loops are also observed in all other *AtPrR1* substrate/product-bound structures, whereas each β 4 loop is disordered in the

Table 1 The crystal structure information of *IiPLR1*, *AtPrR1*, and *AtPrR2* proteins in their apo, NADP⁺ and substrate/product binding forms.

apo	NADP ⁺		NADP ⁺		NADP ⁺		NADP ⁺	
	<i>IiPLR1</i>	<i>AtPrR1</i>	(+)-Pinoresinol	(-)-Pinoresinol	(+)-Lariciresinol	(-)-Lariciresinol	(+)-Secoisolariciresinol	(-)-Secoisolariciresinol
<i>IiPLR1</i>	<i>IiPLR1</i> _apo 2.7 Å	<i>IiPLR1</i> _NAP 2.4 Å	<i>IiPLR1</i> _NAP_+PIN 2.3 Å	<i>IiPLR1</i> _NAP_-PIN 2.1 Å	-	<i>IiPLR1</i> _NAP_-LAR 2.2 Å	<i>IiPLR1</i> _NAP_+SEC 2.3 Å	<i>IiPLR1</i> _NAP_-SEC 2.6 Å
<i>AtPrR1</i>	<i>AtPrR1</i> _apo 2.8 Å	<i>AtPrR1</i> _NAP 2.0 Å	<i>AtPrR1</i> _NAP_+PIN 2.0 Å	<i>AtPrR1</i> _NAP_-PIN 2.5 Å	<i>AtPrR1</i> _NAP_+LAR 1.8 Å	<i>AtPrR1</i> _NAP_-LAR 2.5 Å	-	<i>AtPrR1</i> _NAP_-SEC 2.0 Å
<i>AtPrR2</i>	<i>AtPrR2</i> _apo 2.0 Å	-	<i>AtPrR2</i> _NAP_+PIN 1.6 Å	-	-	-	-	-

corresponding *IiPLR1* structures, which indicates that the $\beta 4$ loop may participate in substrate selectivity and, hence, catalysis.

We further explored why the loops behaved differently between *IiPLR1* and the *AtPrRs*. The amino acid sequences of the $\beta 4$ loops in the three proteins are quite similar (Supplementary Fig. 5), but the residue corresponding to Ser98 at the C-terminal end of the loop in *IiPLR1* is replaced as Asn98 in *AtPrR1/2*. Combining the sequence and structural data, the difference can be explained reasonably as follows: the serine side chain is short enough to remain beneath the guanine group of NADPH, whereas the asparagine side chain cannot do so owing to steric hindrance. Consequently, the asparagine lies nearly vertical to the guanine group and points upward in the structure shown in Fig. 3b, d, and thus the swing of the $\beta 4$ loop is limited in the *AtPrR1*_NAP structure. As substrate enters the catalytic site, $\beta 4$ loop can fold and cover the substrate-binding groove (Fig. 3b, d and Supplementary Figs. 3b and 4a). Besides the $\beta 4$ loop, Val46 in *IiPLR1* is replaced with Leu46 in *AtPrR1*, which has the effect of compressing the substrate-binding pocket. Although Val46 is unchanged in *AtPrR2*, the $\alpha 2$ -helix and $\beta 2$ loop from the neighboring protomer move closer to the substrate upon its entry at the catalytic site, further condensing the pocket. The relative movement of dimers between *AtPrR2* and *IiPLR1* (as suggested by the $\sim 9^\circ$ shift shown in Fig. 3c) could be induced by different dimer orientations. Two molecules of the dimer exhibit relative torsion in *AtPrR2*, and consequently, Val46 is forced deeper into the substrate-binding pocket compared with what occurs in *IiPLR1*. Therefore, the entrance and orientation of the substrate in *AtPrR1/2* is more tightly controlled than in *IiPLR1*.

Mutagenesis-based alteration of the substrate selectivity. Based on the structural analysis of *IiPLR1*, *AtPrR1* and *AtPrR2*, the importance of the candidate amino acids controlling substrate specificity was verified through site-directed mutagenesis. Enzymatic assays using pinoresinol as substrate revealed that the *IiPLR1* mutations including V46A, V46L, S98A, S98H, and S98N somewhat enhanced the conversion rate of lariciresinol while significantly reduced that of secoisolariciresinol, and mutants V46A, S98A, and S98H had > 40% conversion rates for lariciresinol (Fig. 4a), suggesting that residues 46 and 98 are critical for the substrate preference. Taking V46A as an example for the kinetic analysis, its K_m value for pinoresinol ($29.4 \pm 1.62 \mu\text{M}$) is comparable with that for lariciresinol ($26.5 \pm 0.60 \mu\text{M}$), however its V_{max} for pinoresinol ($3.22 \pm 0.68 \mu\text{M min}^{-1}$) is 140-fold higher than that for lariciresinol ($0.023 \pm 0.0013 \mu\text{M min}^{-1}$), and its k_{cat}/K_m for pinoresinol ($3.88 \pm 0.65 \mu\text{M}^{-1} \text{min}^{-1}$) is 126-fold higher than that for lariciresinol ($0.031 \pm 0.002 \mu\text{M}^{-1} \text{min}^{-1}$) (Table 2). Compared with wild-type *IiPLR1*¹⁸, the activity of mutant V46A towards pinoresinol increases 4-fold, whereas that towards lariciresinol decreases 98% with regard to k_{cat}/K_m values. These results indicate mutant V46A enhances catalytic efficiency for the first reaction but dramatically eliminates the second reaction. Consistent with the data for *IiPLR1*, *AtPrR1* mutants L46A and L46V could enhance the conversion rate of lariciresinol and partially reduce that of secoisolariciresinol (Fig. 4a), which confirmed the importance of these two sites in substrate binding and product release thus in catalysis. As expected, mutants N98A and N98S in *AtPrR1* had increased activity for secoisolariciresinol production compared with wild type (Fig. 4a), strongly implying that residue 98 controls the swing of the $\beta 4$ loop, which affects substrate binding and catalysis. Interestingly, *AtPrR2* mutant N98S could utilize lariciresinol to produce secoisolariciresinol, with a conversion rate of 1.91%, in contrast to the wild type which lacks this activity (Fig. 4a). Other *AtPrR2* mutants, including V46A, V46L, and N98A, varied in their activities for pinoresinol,

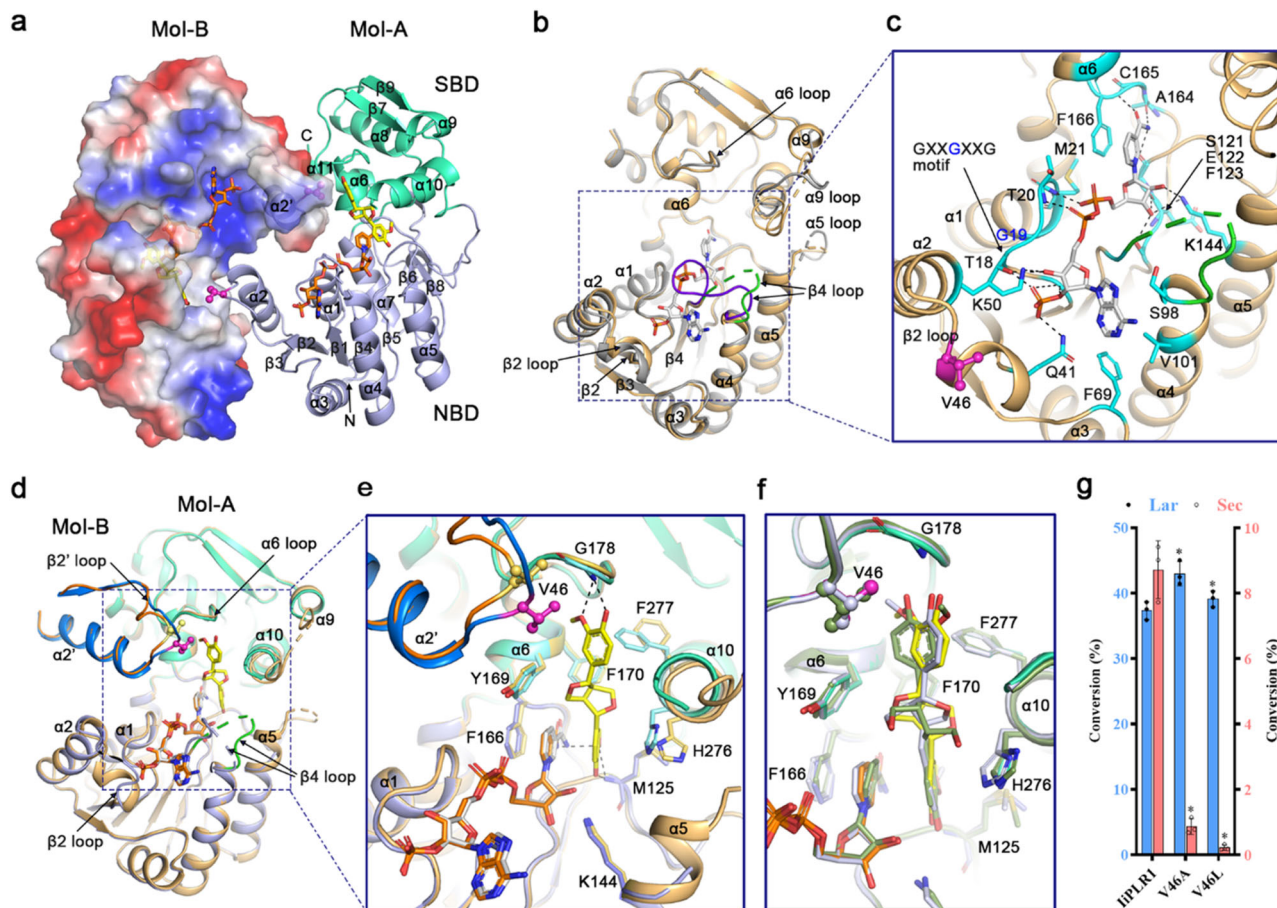


Fig. 2 Structural mechanism of continuous catalytic reactions by *IiPLR1* based on homodimerization. **a** Dimer formation of *IiPLR1_NAP* + PIN. Mol-A is shown as cartoon model, with its NBD and SBD colored in light blue and green cyan, respectively. Mol-B is represented as an electrostatic-surface model, on which blue and red colors represent positive and negative charges, respectively. NADP⁺ and (+)-pinosresinol are shown as sticks and colored in orange and yellow, respectively. **b** Conformational changes were assessed by comparing monomer structures of *IiPLR1_apo* (gray) and *IiPLR1_NAP* (light orange). The NADP⁺ bound to *IiPLR1_NAP* is colored gray. The $\beta 4$ loops of *IiPLR1_apo* and *IiPLR1_NAP* are highlighted as purple and green, respectively. **c** Zoom-in view of the NADPH-binding groove of *IiPLR1_NAP*. Residues interacting with NADP⁺ are colored cyan. The conserved GXXGXXG motif is indicated. Residue Val46 involved in dimer formation and substrate binding is shown as a ball-and-stick model and colored magenta. Dotted lines denote possible hydrogen bonds. **d** Structure comparison of *IiPLR1_NAP* + PIN and *IiPLR1_NAP*. Mol-Bs of *IiPLR1_NAP* + PIN and *IiPLR1_NAP*, Val46s in *IiPLR1_NAP* + PIN, and *IiPLR1_NAP* are colored in marine, orange, magenta, and light yellow, respectively. **e** Zoom-in view of the substrate-binding groove. Residues of Mol-As in *IiPLR1_NAP* + PIN and *IiPLR1_NAP* are colored green cyan (or light blue) and light yellow, respectively. **f** Structural comparison of the substrate/product-binding grooves in *IiPLR1_NAP* + PIN, *IiPLR1_NAP-LAR* (blue white), and *IiPLR1_NAP-SEC* (smudge). The cartoons are generated by PyMOL. **g** Enzyme assays for wild-type *IiPLR1* and its mutants V46A and V46L. Data are mean \pm s.d. ($n = 3$ independent experiments). Asterisk * indicates significant difference from the wild-type enzyme ($P < 0.05$) analyzed by one-way ANOVA with Tukey's multiple comparisons test. Source data underlying Fig. 2g are provided as a Source Data file.

as indicated by the relative rates of conversion to lariciresinol (Fig. 4a). Similar results were obtained for conversion of lariciresinol to secoisolariciresinol (Fig. 4b). Altogether, the structure-guided mutagenesis indeed could switch the substrate specificity of PLR/PrR, e.g., the *IiPLR1* mutant V46A had increased preference for pinosresinol but little catalytic activity for lariciresinol, whereas the *AtPrR2* mutant N98S gained the activity to catalyze the conversion of lariciresinol to secoisolariciresinol (Fig. 4c).

Taking structure and enzymology data together, we proposed a three-step catalytic mechanism for PLR based on its homodimerization. First, the protomers of dimeric PLRs recruit free NADPH through the very flexible $\beta 4$ loop. Second, pinosresinol binds into one protomer via the substrate-binding groove, and the other protomer of the homodimer helps stabilize the substrate. Subsequently, pinosresinol receives H⁺ from NADPH and be reduced to lariciresinol released later. Third, free lariciresinol is bound by another reductive PLR molecule and fixed by another

homodimer, and then the lariciresinol is reduced to secoisolariciresinol and finally released (Fig. 5 and Supplementary Movie 1).

Importantly, the PrRs have more strict requirements for the binding and orientation of lariciresinol compared with PLRs, so PrRs cannot efficiently carry out the third step (Fig. 5). Hence, the substrate-specificity mechanism of PLRs/PrRs could be that residues located around the substrate-binding pocket and within the loop, together with residues that promote homodimerization, form the appropriate hydrophobic environment for binding a specific substrate.

Mutation increases lariciresinol and reduces secoisolariciresinol production in vivo. Enzymatic assays indicated that certain *IiPLR1* mutants had increased activity for producing lariciresinol from pinosresinol in vitro (Fig. 4). Therefore, these *IiPLR1* mutant genes were selected for lariciresinol production using pinosresinol-producing *E. coli*¹⁹. Because matairesinol,

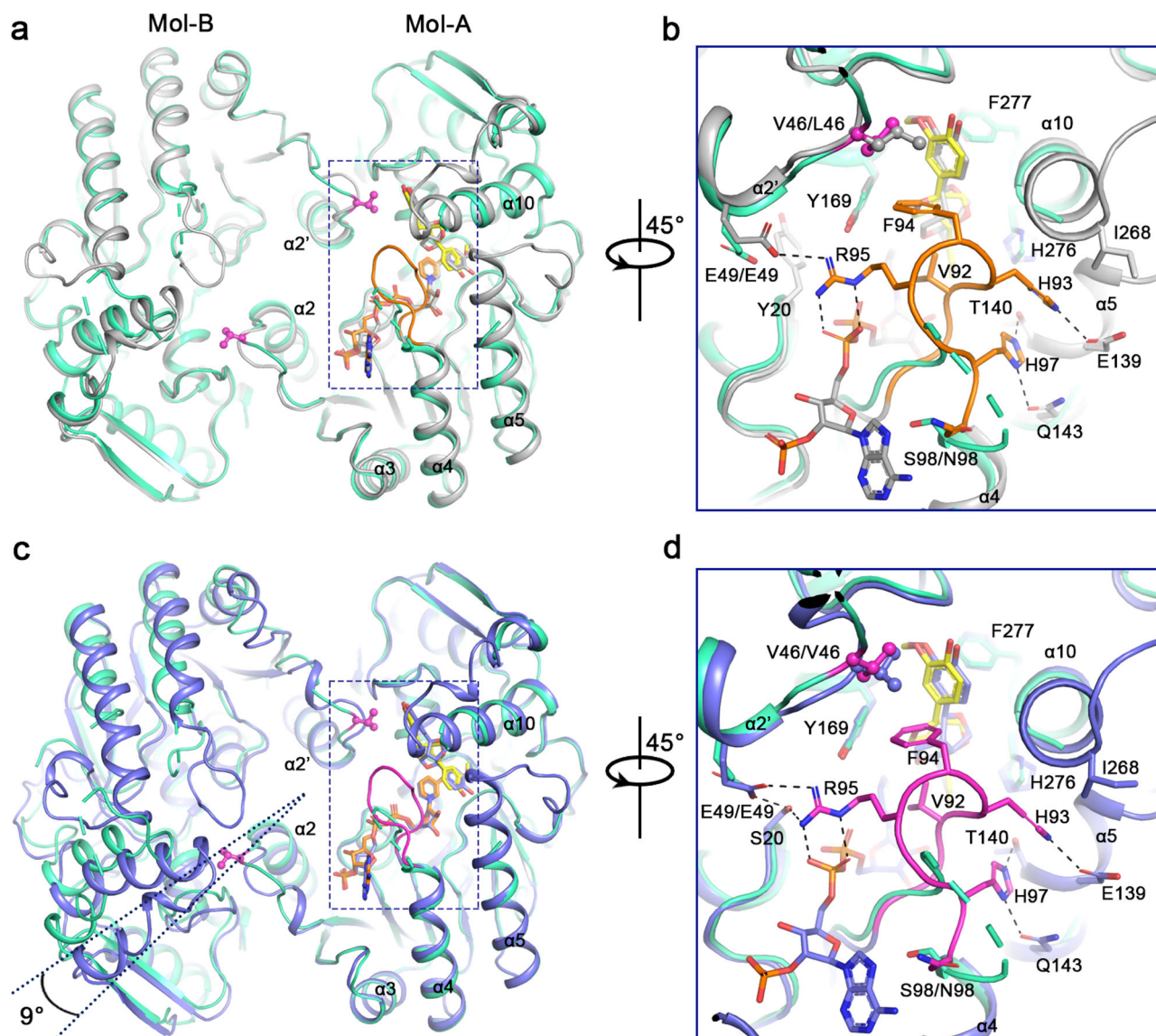


Fig. 3 Structural differences among *IiPLR1*, *AtPrR1*, and *AtPrR2* indicating different catalytic capacities. **a** Structural alignment of dimerized *IiPLR1*_{NAP_} + PIN (green cyan) and *AtPrR1*_{NAP_} + PIN (light gray). **b** Zoom-in view of dashed box in **a**. The $\beta 4$ loop of *AtPrR1*_{NAP_} + PIN is highlighted in orange. **c** Structural alignment of dimerized *IiPLR1*_{NAP_} + PIN (green cyan) and *AtPrR2*_{NAP_} + PIN (slate blue). **d** Zoom-in view of the dashed box in **c**. The $\beta 4$ loop of *AtPrR2*_{NAP_} + PIN is highlighted in magenta. For brevity, NADP⁺ and some residues in *IiPLR1* are hidden. The cartoons are generated by PyMOL.

which is derived from secoisolariciresinol, is detectable only when CueO (multicopper oxidase), PLR and SDH (secoisolariciresinol dehydrogenase) are individually expressed in cells²⁰, each of wild-type *IiPLR1* and its mutants were co-cultured in pinoresinol-producing *E. coli*¹⁹. Consistent with the enzyme assay results, *IiPLR1*_{V46A} produced the greatest amount of lariciresinol (997.79 mg L⁻¹ compared with 936.14 mg L⁻¹ for wild-type). However, mutants V46L, S98A, S98H, and S98N were not as efficient as wild-type cells at producing lariciresinol, which was opposite to the results from in vitro enzyme assays. This may reflect the potential effects of complex metabolic networks and feedback mechanisms in vivo, which are not relevant to in vitro enzyme assays. Moreover, the provision of NADPH is tightly regulated in prokaryotic systems, which also may influence the activity of PLRs.

Notably, all the *IiPLR* mutants produced significantly less secoisolariciresinol than wild-type cells, i.e., by 22.7–52.5%; in particular, *IiPLR*_{V46A} produced 46.4% less secoisolariciresinol

than wild type (Fig. 6). These results paralleled those obtained in vitro with the *IiPLR1* mutants in which there was elimination of the second catalytic step, i.e., the conversion of lariciresinol to secoisolariciresinol (Fig. 4). Taken together, our results establish a promising route for the production of lariciresinol by synthetic biology strategies, and mutant *IiPLR*_{V46A} mutant would be a good candidate for use in the large-scale production of the pharmaceutically valuable compound lariciresinol.

Discussion

The molecular mechanism of substrate selectivity of PLR/PrR has attracted particular interest owing to the key role of these enzymes in lignan biosynthesis. However, the lack of structural results—especially for PLRs/PrRs in complex with different substrates—has limited our understanding of the mechanism underlying enzyme specificity. In the present study, we characterized crystal structures of *IiPLR1*, *AtPrR1*, and *AtPrR2* in

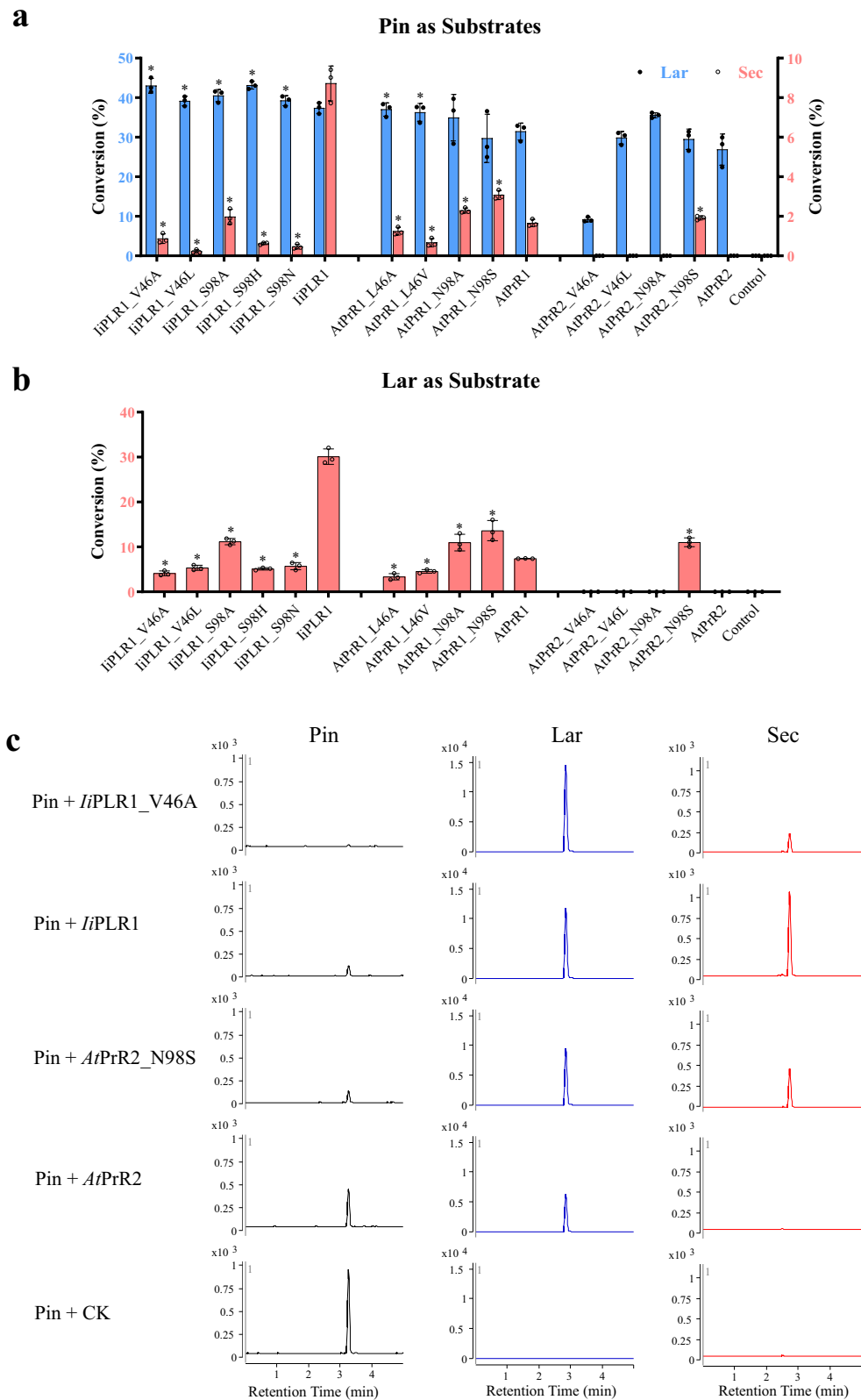


Fig. 4 Percent conversion of pinoresinol to lariciresinol and subsequently to secoisolariciresinol by mutants of *IiPLR1*, *AtPrR1* and *AtPrR2*. a Conversion of pinoresinol into lariciresinol and subsequently to secoisolariciresinol. **b** Conversion of lariciresinol into secoisolariciresinol. Data are mean \pm s.d. ($n = 3$ independent experiments). Asterisk (*) indicates significant difference from the wild-type enzyme ($P < 0.05$) analyzed by one-way ANOVA with Tukey's multiple comparisons test. Source data underlying Fig. 4a, b are provided as a Source Data file. **c** LC-MS determination of the products as catalyzed by *IiPLR1*, *IiPLR1_V46A*, *AtPrR2* and *AtPrR2_N98S*.

Table 2 Kinetic properties of *li*PLR1_V46A.

Substrate	K_m (μM)	V_{max} ($\mu\text{M}/\text{min}$)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu\text{M}^{-1} \text{min}^{-1}$)
(±)-pinoresinol	29.4 ± 1.62	3.22 ± 0.68	114.7 ± 24.1	3.88 ± 0.65
(±)-lariciresinol	26.5 ± 0.60	0.023 ± 0.0013	0.81 ± 0.045	0.031 ± 0.002

Data are expressed as mean ± s.d. with three independent experiments. Source data are provided as a Source Data file.

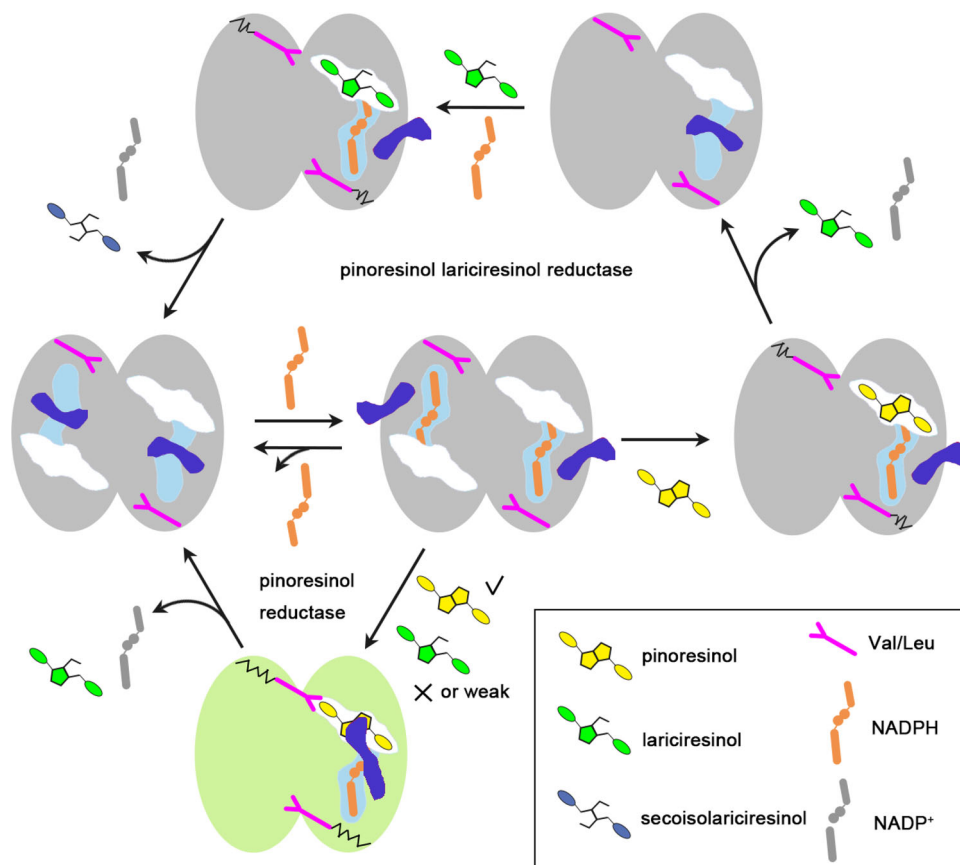


Fig. 5 Model depicting the catalytic processes of PLRs and PrRs. The blue dumbbell-shaped objects represent switches composed of $\beta 4$ loops. A movie showing how the enzymes change conformation throughout a single round of catalysis can be found in Supplementary Movie 1.

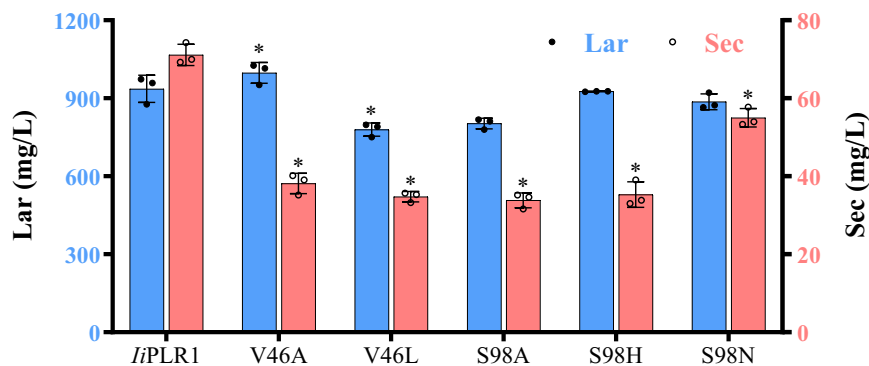


Fig. 6 Lariciresinol production through co-culture of different strains harboring a plasmid encoding *li*PLR1 or its single-site mutants with pinoresinol-producing cells. Data are mean ± s.d. ($n = 3$ independent experiments). Asterisk (*) indicates significant difference from the wild-type enzyme ($P < 0.05$) analyzed by one-way ANOVA with Tukey's multiple comparisons test. Source data are provided as a Source Data file.

complex with their various substrates. Several residues participating in substrate binding and catalysis were identified either directly or indirectly based on structural analysis, and these residues were validated by enzyme assays. All these data provide solid evidence to explore the mechanistic basis of substrate selectivity for PLRs/PrRs. Besides residues 46 and 98 in *IiPLR1* and *AtPrR1* that we identified as being critical for binding and catalysis, residues Phe166, Tyr169, Phe170, His276, and Phe277 within the substrate-binding groove were also strongly correlated with enhanced substrate binding and catalysis. We further deduced that any residue in PLRs/PrRs around the hydrophobic groove or affecting homodimerization may impact the conformation of the active site, thereby dictating substrate selectivity (Supplementary Figs. 5 and 6). Consequently, it is not difficult to understand why mutant L174I of *C. sinensis* PLR1 can hardly reduce pinoselinol and specifically catalyze the conversion of lariciresinol to secoisolariciresinol¹⁶, i.e., because Leu174 points directly toward Tyr163 and thus may indirectly promote substrate recognition.

In addition, PLRs also display substrate stereochemical selectivity, which contributes to the enantiomeric diversity of lignans^{21–24}. We found this to be true for *IiPLR1*, which gave comparable k_{cat}/K_m values for both (±)-pinoselinol and (±)-lariciresinol in the range of 0.9–1.6 $\mu\text{M}^{-1}\text{min}^{-1}$, although no experiments with respect to the enantio-specificity of this enzyme have been performed^{12,18}. Despite past research on this topic, however, the mechanism underlying the substrate stereochemical selectivity of PLRs remains unclear. Comparison of the enantiospecifically opposite PLRs *TpPLR1* and *TpPLR2* suggests that F164, V268, and L272 in *TpPLR1* contribute to the catalysis of (–)-pinoselinol, whereas L164, G268, and F272 in *TpPLR2* prefer to bind (+)-pinoselinol¹⁷. Nevertheless, site-directed mutagenesis carried out in flax indicates that these positions are insufficient to determine enantiospecificity²¹. Based on amino-acid sequence and structural analyses, it seems that residues Phe94 and Phe277 in *IiPLR1/AtPrR1* may act in concert to determine the enantiospecificity of PLRs (Supplementary Fig. 6). These two residues are highly conserved in PLRs that have no enantio-specificity, whereas Ile94 and Tyr277 are present in PLRs that have a strict substrate preference for (+)-pinoselinol. Moreover, in *Linum usitatissimum* PLR1 (*LuPLR1*), which has strict enantiospecificity for (–)-pinoselinol, a leucine residue is deleted as are two other residues on β 4 loop (corresponding to the β 4 loop of *IiPLR1*, where Phe94 is located). Unfortunately, we could not obtain sufficient amounts of the enantiomerically pure substrates to carry out the experiments necessary to establish the enantioselectivity.

Nature uses a dazzling array of enzymes to produce diverse natural products. However, some modifications are challenging to control because the relative lack of substrate specificity often generates undesired byproducts. *IiPLR1* plays an important role in the biotechnological production of lariciresinol¹⁸, which represents the most important component for the antibacterial, antiviral, and the immune-regulatory effects of the traditional Chinese medicine *Radix Isatidis*^{3,25,26}. The fact that *IiPLR1* can efficiently utilize both pinoselinol and lariciresinol as substrates¹⁸ suggests that the biosynthetic efficiency towards the pharmaceutically valuable compound lariciresinol in *Radix Isatidis* has been hampered by the relatively low substrate specificity of *IiPLR1*. In our present work, structure-guided mutagenesis successfully switched the substrate specificity of *IiPLR1*, leading to overproduction of lariciresinol and reduced production of secoisolariciresinol by *E. coli*. Our study provides insight into the molecular mechanism underlying the substrate specificity of PLRs/PrRs, and paves the way for the manufacture of lariciresinol through microbial fermentation. Moreover, this work suggests the

possibility of using targeted mutagenesis of *IiPLR1* to improve the efficiency of synthesizing bioactive compounds in *I. indigotica* using gene-editing technologies^{27,28}.

Methods

Phylogenetic analysis of plant PLRs. Phylogenetic relationships were analyzed using the maximum likelihood method with the pairwise deletion option in MEGA 6.06. Tree reliability was estimated using a bootstrap analysis of 1000 replicates²⁹. Plant PLR amino-acid sequences used in the phylogenetic analysis were retrieved from GenBank, including *TpPLR1* (AAF63507.1), *TpPLR2* (AAF63508.1), *TpPLR3* (AAF63509.1), *TpPLR4* (AAF63510.1), *PpPLR* (AHL21381.1), *LuPLR1* (CAH60857.1), *LuPLR2* (CAH60858.1), *LuPLR3* (ABW24501.1), *LpPLR1* (ABM68630.1), *LcPLR1* (ABW86959.1), *PhPLR* (ACF71492.1), *TcPLR1* (AZL88516.1), *TcPLR2* (AZL88517.1), *FiPLR1* (AAC49608.1), *IiPLR1* (AEA42007.1), *AtPrR1* (NP_174490.1), and *AtPrR2* (NP_193102.1).

Heterologous expression of *IiPLR1*, *AtPrR1*, and *AtPrR2* in *E. coli*. Total RNA was extracted from leaves of wild-type *I. indigotica* or *A. thaliana* using TRIzol Reagent (GIBCO BRL). The mRNA was reverse transcribed with oligo dT to generate cDNA as a template for PCR. Full-length cDNA sequences of *IiPLR1* (GenBank accession no. JF264893), *AtPrR1* (AY065214) and *AtPrR2* (BT002882) were cloned into pET-duet-1 (Novagen, USA) to generate *IiPLR1*-pET, *AtPrR1*-pET, and *AtPrR2*-pET, respectively. The primers used are listed in Supplementary Table 1. *E. coli* Rosetta (DE3) cells were transformed with purified plasmid DNA and then grown at 37 °C to an OD₆₀₀ of 0.8. Then, protein expression was induced by adding 0.5 mM isopropyl β -D-thiogalactoside (IPTG, final concentration) with incubation overnight at 16 °C. Cells were collected, resuspended in buffer A (20 mM Tris-HCl pH 8.0, 100 mM NaCl), and lysed with a French press. The lysate was centrifuged at 20,000 \times g for 45 min, and the supernatant was applied to a Ni-NTA column equilibrated with buffer A supplemented with 25 mM imidazole. Bound protein was eluted using buffer A containing 250 mM imidazole and was concentrated for further purification on a Superdex-200 column equilibrated with buffer A. Protein purity was assessed by SDS-PAGE (12% polyacrylamide), and protein concentration was determined by the Bradford method³⁰.

Crystallization, data collection, and structure determination. The full-length *IiPLR1/AtPrR1/AtPrR2* were purified as described above and concentrated to 5–10 mg mL⁻¹ for crystallization. Aliquots of each concentrated protein sample were mixed 1:1 with reservoir solution, and crystals were grown at 20 or 4 °C in one week using the sitting-drop vapor-diffusion method. For co-crystals, protein was combined with NADP⁺ at a 1:5 molar ratio, and protein with NADP⁺ and substrate/product at a 1:5:10 molar ratio. For reservoir solutions, *IiPLR1* apo and co-crystals were grown with 0.2 M sodium citrate tribasic, 0.1 M sodium citrate/citric acid, pH 4.0 and 20% polyethylene glycol (PEG) 3350; *AtPrR1* apo crystals were grown in 0.2 M lithium chloride, 20% w/v PEG 3350; *AtPrR1*_NAP, *AtPrR1*_NAP_+ PIN, *AtPrR1*_NAP_+ LAR and *AtPrR1*_NAP_-SEC were grown in 0.2 M sodium fluoride, 20% w/v PEG 3350; *AtPrR1*_NAP_-PIN and *AtPrR1*_NAP_-LAR crystals were grown in 0.2 M sodium malonate, pH 6.0, 20% w/v PEG 3350; *AtPrR2* apo crystals were grown in 0.2 M magnesium chloride, 0.1 M sodium HEPES, pH 7.5 and 25% PEG 3350; *AtPrR2*_NAP_+ PIN crystals were grown in 2.1 M DL malic acid, pH 7.0. The crystals were cryoprotected by serial transfers into reservoir solutions supplemented with 30% (v/v) glycerol and then flash-cooled in liquid nitrogen. Data collections were performed at the BL17U1 and BL19U1 beamline of the Shanghai Synchrotron Radiation Facility. The data were processed with HKL3000³¹, and the initial phase was determined by molecular replacement with Phenix³² using the crystal structure of *TpPLR1* (PDB ID: 1QYD [<https://doi.org/10.2210/pdb1qyd/pdb>]) as a template. The structure models were firstly auto-built in Coot³³ and then refined by iterative rounds of manual adjustment with Coot and refinement with Phenix. The statistics of data collection and structure refinement are shown in Supplementary Tables 2–5.

Site-directed mutagenesis of *IiPLR1*, *AtPrR1*, and *AtPrR2* and enzymatic assays

Single-site mutagenesis was achieved through one-step PCR, and mutants were verified with Sanger sequencing. All primers are listed in Supplementary Table 1. After expression and purification of recombinant enzymes under the aforementioned conditions, the results for the enzyme assays for mutants were compared with those for wild-type recombinant *IiPLR1*, *AtPrR1*, and *AtPrR2* as follows.

Enzyme activity assays were conducted strictly according to our previous work¹⁸. Assay mixtures (1 mL) consisted of TG buffer (50 mM Tris-HCl, 10% [w/v] glycerol, pH 7.0), 150 μM NADPH, 100 μM pinoselinol, or 100 μM lariciresinol and 5 μg of purified protein. Assays without a fusion protein were used as controls. Protein, buffer, and substrate were pre-incubated for 5 min at 30 °C, and each reaction was initiated by addition of NADPH and terminated after 30 min by addition of 300 μL ethyl acetate. Each assay mixture was extracted with ethyl acetate (3 \times 300 μL total). The combined ethyl acetate phases were dried under vacuum, and the residue was dissolved in 1 mL methanol. Conversion rate was then determined. The content of pinoselinol, lariciresinol and secoisolariciresinol was determined by LC-MS using a triple-quadrupole mass spectrometer (Model 6410, Agilent, Santa Clara, CA)

following our published methods¹⁸. MassHunter Qualitative Analysis B.06.00 was used for the data analysis. The selected transitions of *m/z* were 357 → 151 for pinoresinol, 359 → 329 for lariciresinol, and 361 → 164 for secoisolariciresinol. All standards were purchased from Sigma-Aldrich (St. Louis, MO).

For determination of V_{max} and K_m values for *IiPLR1_V46A*, 10 different concentrations of substrate (pinoresinol or lariciresinol; 5–200 μ M) and 1 μ g purified protein were used. Samples were incubated at 30 °C for 5 min (during which substrate consumption was \leq 10%). Samples without protein were used as controls. The rate of substrate consumption was calculated for kinetic analysis. V_{max} and K_m values were determined from Lineweaver-Burk plots, and k_{cat} was determined by dividing V_{max} by the enzyme concentration.

Bioconversion. For the production of lariciresinol, biotransformation was divided into two modules, namely the accumulation and conversion of the precursor, pinoresinol. *E. coli* strain strOpr2 carrying plasmid pET28a-Prx02-PsVAO was used to produce pinoresinol¹⁹, whereas *E. coli* BL21(DE3) carrying a plasmid encoding *IiPLR1* or its mutants was used for conversion of pinoresinol to lariciresinol. These *E. coli* strains were cultured in LB medium at 37 °C with shaking (220 rpm) for 12 h as seed cultures, and then a 2% seed culture was transferred to a 250-mL shaker flask containing 25 mL TB medium. After culturing for 2–2.5 h at 37 °C and 220 rpm, 500 μ M IPTG (final concentration) was added to the medium with continued cultivation for 12 h at 25 °C and 220 rpm. These cells were used for pinoresinol accumulation and conversion, respectively. Cells from *E. coli* strain strOpr2 were harvested by centrifugation at 4 °C and 3724 \times g for 30 min and then resuspended in phosphate-buffered saline (pH 7.0) to adjust the OD₆₀₀ value to 20. Then 0.15% (v/v) eugenol was added into 15 mL of the resuspension at 0, 1, 3, 5, and 7 h for pinoresinol accumulation (20 °C, 220 rpm). At 9 h, 15 mL of a culture of *E. coli* expressing *IiPLR1* and each mutant (OD₆₀₀ = 20) was added to determine the conversion of pinoresinol to lariciresinol (25 °C, 220 rpm), and samples were taken after 20 h. The concentration of each of lariciresinol and secoisolariciresinol was determined by HPLC.

Statistical analysis. All the experiments in this paper were repeated at least three times and results from representative data sets are presented. GraphPad Prism (version 9.1.0) was used for the statistical analysis. The statistical evaluations used one-way analysis of variance (ANOVA) with multiple comparisons, followed by Tukey tests. The results were considered statistically significant at $*P < 0.05$.

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

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. The atomic coordinates and structure factors for the structures have been deposited in the Protein Data Bank with accession codes 7CS2 (Apo structure of dimeric *IiPLR1*), 7CS3 (*IiPLR1* with NADP⁺), 7CS4 (*IiPLR1* with NADP⁺ and (+)pinoresinol), 7CS5 (*IiPLR1* with NADP⁺ and (–)pinoresinol), 7CS6 (*IiPLR1* with NADP⁺ and (–)lariciresinol), 7CS7 (*IiPLR1* with NADP⁺ and (+)secoisolariciresinol), 7CS8 (*IiPLR1* with NADP⁺ and (–)secoisolariciresinol), 7CS9 (*AtPrR1* in apo form), 7CSA (*AtPrR1* with NADP⁺), 7CSB (*AtPrR1* with NADP⁺ and (+)pinoresinol), 7CSC (*AtPrR1* with NADP⁺ and (–)pinoresinol), 7CSD (*AtPrR1* with NADP⁺ and (+)lariciresinol), 7CSE (*AtPrR1* with NADP⁺ and (–)lariciresinol), 7CSF (*AtPrR1* with NADP⁺ and (–)secoisolariciresinol), 7CSG (*AtPrR2* in apo form), 7CSH (*AtPrR2* with NADP⁺ and (+)pinoresinol). The initial phase was determined by molecular replacement using the crystal structure of *TpPLR1* (PDB ID: 1QYD [<https://doi.org/10.2210/pdb1qyd/pdb>]) as a template. The source data underlying Figs. 2g, 4a, b, and 6, as well as Table 2 are provided as a Source Data file. All data generated and analyzed during the current study are available from the corresponding authors upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file. Source data are provided with this paper.

Received: 10 July 2020; Accepted: 14 April 2021;

Published online: 14 May 2021

References

1. Teponno, R. B., Kusari, S. & Spittler, M. Recent advances in research on lignans and neolignans. *Nat. Prod. Rep.* **33**, 1044–1092 (2016).
2. Teodor, E. D., Moroceanu, V. & Radu, G. L. Lignans from medicinal plants and their anticancer effect. *Mini Rev. Med. Chem.* **20**, 1083–1090 (2020).
3. J, L. et al. Lariciresinol-4-O- β -D-glucopyranoside from the root of *Isatis indigotica* inhibits influenza A virus-induced pro-inflammatory response. *J. Ethnopharmacol.* **174**, 379–386 (2015).
4. Xiao, P., Huang, H., Chen, J. & Li, X. *In vitro* antioxidant and anti-inflammatory activities of Radix Isatidis extract and bioaccessibility of six bioactive compounds after simulated gastro-intestinal digestion. *J. Ethnopharmacol.* **157**, 55–61 (2014).
5. Hirano, T., Wakasugi, A., Oohara, M., Oka, K. & Sashida, Y. Suppression of mitogen-induced proliferation of human peripheral blood lymphocytes by plant lignans. *Planta Med.* **57**, 331–334 (1991).
6. Haihan et al. Two new furofuran lignans from *Kandelia obovata*. *Heterocycles* **87**, 1093–1098 (2013).
7. Chin, Y., Chai, H., Keller, W. J. & Kinghorn, A. D. Lignans and other constituents of the fruits of *Euterpe oleracea* (Acai) with antioxidant and cytoprotective activities. *J. Agric. Food Chem.* **56**, 7759–7764 (2008).
8. Jang, Y. P., Kim, S. R. & Kim, Y. C. Neuroprotective dibenzylbutyrolactone lignans of *Torreya nucifera*. *Planta Med.* **67**, 470–472 (2001).
9. Canel, C., Moraes, R. M., Dayan, F. E. & Ferreira, D. Podophyllotoxin. *Phytochemistry* **54**, 115–120 (2000).
10. Lau, W. & Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **349**, 1224–1228 (2015).
11. Laurence et al. Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science* **275**, 362–366 (1997).
12. Markulin, L. et al. Pinoresinol–lariciresinol reductases, key to the lignan synthesis in plants. *Planta* **249**, 1695–1714 (2019).
13. Dinkova-Kostova, A. T. et al. (+)-Pinoresinol/(+)-lariciresinol reductase from *Forsythia intermedia*. Protein purification, cDNA cloning, heterologous expression and comparison to isoflavone reductase. *J. Biol. Chem.* **271**, 29473–29482 (1996).
14. Umezawa, T. Diversity in lignan biosynthesis. *Phytochem. Rev.* **2**, 371–390 (2003).
15. Tomoyuki, N., Masaharu, M., Shiro, S., Takefumi, H. & Toshiaki, U. Characterization of *Arabidopsis thaliana* pinoresinol reductase, a new type of enzyme involved in lignan biosynthesis. *J. Biol. Chem.* **283**, 15550–15557 (2008).
16. Wu, Y. et al. A variable loop involved in the substrate selectivity of pinoresinol/lariciresinol reductase from *Camellia sinensis*. *Phytochemistry* **162**, 1–9 (2019).
17. Tongpil, M. et al. Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. *J. Biol. Chem.* **278**, 50714–50723 (2003).
18. Xiao, Y. et al. Combined transcriptome and metabolite profiling reveals that *IiPLR1* plays an important role in lariciresinol accumulation in *Isatis indigotica*. *J. Exp. Bot.* **66**, 6259–6271 (2015).
19. Lv, Y., Cheng, X., Du, G., Zhou, J. & Chen, J. Engineering of an H₂O₂ auto-scavenging *in vivo* cascade for pinoresinol production. *Biotechnol. Bioeng.* **114**, 2066–2074 (2017).
20. Decembrino, D., Girhard, M. & Urlacher, V. B. Use of copper as a trigger for *in vivo* activity of *E. coli* laccase CueO: a simple tool for biosynthetic purposes. *ChemBioChem* **22**, 1470–1479 (2021).
21. Heimendahl, C. B. I. V. et al. Pinoresinol-lariciresinol reductases with different stereospecificity from *Linum album* and *Linum usitatissimum*. *Phytochemistry* **66**, 1254–1263 (2005).
22. Chu, A., Dinkova, A., Davin, L. B., Bedgar, D. L. & Lewis, N. G. Stereospecificity of (+)-pinoresinol and (+)-lariciresinol reductases from *Forsythia intermedia*. *J. Biol. Chem.* **268**, 27026–27033 (1993).
23. Fujita, M., Gang, D. R., Davin, L. B. & Lewis, N. G. Recombinant pinoresinol-lariciresinol reductases from western red cedar (*Thuja plicata*) catalyze opposite enantiospecific conversions. *J. Biol. Chem.* **274**, 618–627 (1999).
24. Sicilia, T., Niemeier, H. B., Honig, D. M. & Metzler, M. Identification and stereochemical characterization of lignans in flaxseed and pumpkin seeds. *J. Agric. Food Chem.* **51**, 1181–1188 (2003).
25. Yang, Z. et al. Antiviral activity of *Isatis indigotica* root-derived clemastanin B against human and avian influenza A and B viruses *in vitro*. *Int. J. Mol. Med.* **31**, 867–873 (2013).
26. Council, N. P. *Pharmacopoeia of China 2015 Edition*. (Chemical Industry, 2015).
27. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
28. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
29. Tamura, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
30. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
31. Cymborowski, M., Otwinowski, Z., Chruszcz, M. & Minor, W. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
32. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

Acknowledgements

This work was sponsored by National Key R&D Program of China (2019YFC1711100, 2018YFA0900600, 2019YFC1711000, SQ2018YFA090071), National Natural Science Foundation of China (81874335, 32025020, 31872665, 31670292 and 31970316), the National Science Fund for Excellent Young Scholars (21822806), Shanghai Rising-Star Program (18QB1402700), the Chinese Academy of Sciences Strategic Priority Research Program (XDB27020103), Shanghai local Science and Technology Development Fund Program guided by the Central Government (YDZX20203100002948), and Shanghai Science and technology Commission (19XD1424500). We thank the staff members at SSRF BL17U1 and BL19U1 for their technical assistance in X-ray diffraction data collection, and the core facility center of Institute of Plant Physiology and Ecology (CAS) for X-ray diffraction test.

Author contributions

Y.X., W.C., K.S., P.Z., J.Z., and L.Z. designed experiments. K.S. carried out protein expression, purification, crystallization, structure determination, and data analysis. Y.X., X.M., Y.Y., J.C., J.F., S.Q., and Z.L. carried out enzymatic assays. J.Z., L.W., and D. W. performed bioconversion and optimization. Y.X., K.S., J.Z., W.C., P.Z., and L.Z. analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23095-y>.

Correspondence and requests for materials should be addressed to L.Z., P.Z. or W.C.

Peer review information *Nature Communications* thanks Clint Chapple, Jiahai Zhou, and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021