nature chemical biology

BRIEF COMMUNICATION

PUBLISHED ONLINE: 9 NOVEMBER 2015 | DOI: 10.1038/NCHEMBIO.1955

Structural determinants of reductive terpene cyclization in iridoid biosynthesis

Hajo Kries[1,3], Lorenzo Caputi[1,3], Clare E M Stevenson[1], Mohammed O Kamileen[1], Nathaniel H Sherden[1], Fernando Geu-Flores[2], David M Lawson[1] & Sarah E O'Connor[1]*

The carbon skeleton of ecologically and pharmacologically important iridoid monoterpenes is formed in a reductive cyclization reaction unrelated to canonical terpene cyclization. Here we report the crystal structure of the recently discovered iridoid cyclase (from *Catharanthus roseus*) bound to a mechanism-inspired inhibitor that illuminates substrate binding and catalytic function of the enzyme. Key features that distinguish iridoid synthase from its close homolog progesterone 5β-reductase are highlighted.

The iridoids are bicyclic cyclopenta[*c*]pyran monoterpenes synthesized by plants and insects to mediate interactions among these species[1,2]. Additionally, many iridoids have pharmacological value[3–5]. The pivotal step in the biosynthesis of plant-derived iridoids is catalyzed by iridoid synthase (ISY), which converts 8-oxogeranial into nepetalactol, the common intermediate for all plant-derived iridoids (**Fig. 1a**)[6,7]. We determined the structure of ISY from *C. roseus* (CrISY) with and without a bound inhibitor. The structure revealed the substrate-binding pocket, the probable mode of substrate binding, and structural features that are likely to control substrate selectivity for this unusual enzymatic reaction.

Cyclization in iridoid biosynthesis utilizes a different substrate and reaction mechanism than does canonical terpene cyclization[6]. Consequently, ISY does not resemble a terpene synthase but rather is a homolog of progesterone 5β-reductase (P5βR)[8], a short-chain NADPH-dependent dehydrogenase[9]. P5βR catalyzes the stereoselective reduction of the Δ^4-double bond in progesterone to yield 5β-pregnane-3,20-dione, a reaction in cardenolide biosynthesis (**Fig. 1b**)[10,11]. In contrast, ISY combines reduction with cyclization of its 8-oxogeranial substrate.

A full understanding of iridoid cyclization has not emerged. Rauhut-Currier cyclization, in which covalently assisted cyclization occurs before reduction[12] (**Fig. 1c**) is suggested by the presence of nonreduced iridoids isolated from beetles[13]. However, CrISY produces minor quantities of reduced, noncyclized products[6], suggesting that in plants reduction of 8-oxogeranial precedes cyclization. We proposed that, consistent with the P5βR mechanism[14], an enol or enolate intermediate initially forms via a 1,4 reduction of the substrate by NADPH[6]. ISY could then catalyze cyclization of the enol or enolate. One cyclization mechanism uses a hetero Diels-Alder reaction to form nepetalactol (**Fig. 1c**), though recent studies with substrate analogs argue against this concerted cycloaddition[15]. More likely, the enolate cyclizes via a Michael reaction to form iridodial (**Fig. 1c**), which is the open dial form of nepetalactol (**Fig. 1a**). Michael cyclization could be assisted by covalent catalysis, such as imine formation at C8 with an active-site lysine (**Fig. 1c**)[16].

To clarify the mechanism, CrISY's structure was solved in its apo form and with $NADP^+$ bound (PDB 5DCW and 5DCU, respectively) at 1.90-Å and 1.40-Å resolution, respectively (**Fig. 2a**; **Supplementary Results**, **Supplementary Table 1**). Crystallization required truncation of the N terminus, which had minimal impact on catalytic efficiency (**Supplementary Table 2**)[17]. The structure of P5βR from *Digitalis lanata* (DlP5βR, PDB 2V6G[14], 2.3-Å resolution, **Supplementary Figs. 1** and **2a** and **Supplementary Table 3**), which has an overall fold that is highly similar to that of CrISY, was used for molecular replacement. The tyrosine and lysine active site residues that are used in the model for P5βR catalysis, along with many of the residues that define the shape of the P5βR active site, are conserved in CrISY (**Supplementary Fig. 2b**)[14]. The apo and $NADP^+$-bound forms of CrISY show minor differences in conformation (**Supplementary Fig. 2**, **Supplementary Table 3**)[14]. An adventitious ligand, most likely an oxidized polyethylene glycol fragment (triethylene glycol carboxylic acid, TEG), was observed in the active site of the $NADP^+$ bound enzyme (**Fig. 2a** and **Supplementary Fig. 3a**,**b**). As efforts to co-crystallize CrISY with substrate, product and/or NADPH were unsuccessful, we searched for an inhibitor. We determined that geranic acid (GEA), which resembles the proposed enolate intermediate, acted as a noncompetitive inhibitor of CrISY activity (K_i ~4 μM) (**Supplementary Fig. 4**), presumably by binding to the inactive CrISY–$NADP^+$ complex. Geranic acid could be co-crystallized with $NADP^+$ and the enzyme, allowing the structure of the ternary complex to be solved at 1.75-Å resolution (PDB 5DF1) (**Fig. 2b**, **Supplementary Table 1**, **Supplementary Fig. 3c**,**d**).

These three structures of CrISY (apo, $NADP^+$, $NADP^+$–geranic acid) were used to construct a model of catalysis. The absence of suitably placed cysteine or histidine residues substantiated our hypothesis that covalently assisted cyclization before reduction is unlikely. No lysine residue was correctly oriented for imine-assisted catalysis; although Lys146 is in the active site, its ε-amino group points away from the carboxyl moiety of the inhibitor.

The resemblance of negatively charged geranic acid to the proposed enolate intermediate supports the hypothesis that CrISY reduces 8-oxogeranial to an enolate. The backbone amide of Lys146 and/or the phenolic group of Tyr178 may bind to the C1 carbonyl of the substrate to stabilize the newly forming negative charge as reduction occurs (**Fig. 2b**). The backbone amide of Ile145 and the O2′ of the ribose of $NADP^+$ may also interact with the enolate to form an oxyanion hole (**Fig. 2b**). A proton transfer from Tyr178 to the hypothetical enolate intermediate is conceivable but seems catalytically undesirable, because the enolate is more reactive for Michael cyclization than the enol. Although mutation of Lys146

[1]The John Innes Centre, Department of Biological Chemistry, Norwich Research Park, Norwich, UK. [2]University of Copenhagen, Department of Plant and Environmental Sciences, Copenhagen Plant Science Centre & Section for Plant Biochemistry, Copenhagen, Denmark. [3]These authors contributed equally to this work. *e-mail: sarah.o'connor@jic.ac.uk

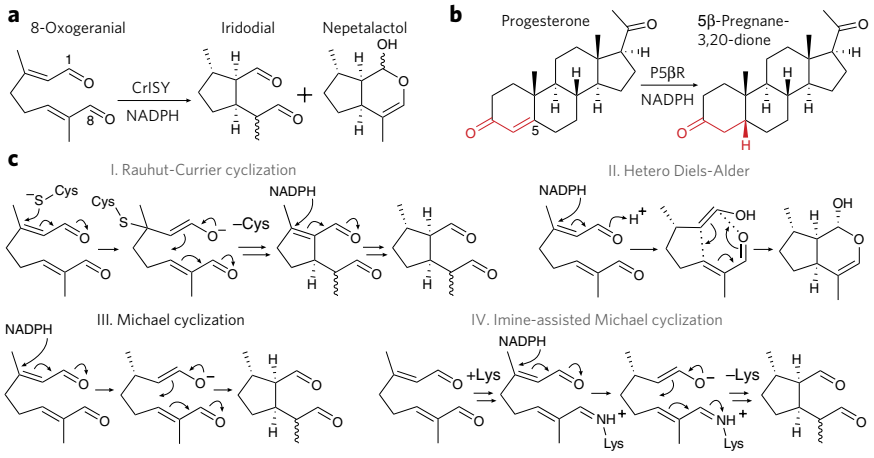

Figure 1 | Mechanism of iridoid synthase (ISY). (**a**) Conversion of 8-oxogeranial into iridodial and nepetalactol. Both products are observed in the ISY reaction. (**b**) Reduction of progesterone, catalyzed by the close ISY homolog progesterone 5β-reductase (P5βR). (**c**) Possible mechanistic scenarios for ISY. I. A Rauhut-Currier cyclization covalently assisted by cysteine (or histidine) is followed by reduction. II. Reduction followed by a hetero Diels-Alder reaction. III. Reduction followed by a Michael cyclization. IV. Reduction followed by an imine-assisted Michael cyclization.

to methionine or arginine increased the K_M, k_{cat} values were close to wild type (wild-type CrISY: k_{cat} 4.2 s^{-1}, K_M 1.9 μM; K146M: k_{cat} 2.2 s^{-1}, K_M 26 μM; K146R: k_{cat} 6.1 s^{-1}, K_M 17 μM; **Supplementary Table 2**) at a range of pH values (**Supplementary Fig. 13**), suggesting that this residue does not have a major role in catalysis. The activity of the Lys146 mutants argues against a tyrosine-lysine interaction that is proposed to modulate the pK_a of tyrosine in short-chain dehydrogenases[18]. Mutant CrISY^{Y178F} showed a 400-fold decrease in k_{cat}/K_M and a 30-fold decrease in k_{cat} as compared to wild type (**Supplementary Table 2**) but still yielded cyclized product (**Supplementary Fig. 7**).

The oxyanion hole anchors geranic acid directly above the NADP$^+$ cofactor, an orientation in which the pro-*S* hydride of NADPH would be poised for addition to C3 (**Fig. 2b**). Such a hydride transfer would generate the *S* stereocenter at this carbon, which is observed in the final product, further indicating a catalytically relevant orientation of the inhibitor. Overall, the ISY structures support the reduction of 8-oxogeranial to the enolate followed by noncovalently assisted Michael cyclization.

The shape of the geranic acid binding pocket implies that the enolate intermediate is generated in a tightly packed linear orientation, which is not compatible with cyclization (**Fig. 2b,c** and **Supplementary Fig. 5**). We assume that O1 of the enolate remains anchored in the oxyanion hole, which, after hydride transfer, is strengthened by the positive charge of the oxidized cofactor. Then, the enzyme must adopt a more open conformation to allow the intermediate to fold into the appropriate conformation to cyclize. Notably, the structure of NADP$^+$ bound CrISY reveals a mix of partially unresolved 'open' and 'closed' conformations involving the Gly150–Asp162 loop (**Fig. 3a**). In contrast, the inhibitor bound structure exists predominantly in the closed conformation, in which Phe152 locks the ligand in the linear conformation (**Supplementary Fig. 5**)[19].

We speculate that after formation of the enolate in the closed conformation, ISY rearranges to an open form to allow folding and cyclization.

To explore the catalytic relevance of the open and closed conformations, we biased the conformational equilibrium of the Gly150–Asp162 loop by mutating Gly150, the position at which the open and closed conformations bifurcate. Ramachandran plots show (**Supplementary Fig. 6**)[20] that a residue other than glycine would favor the open conformation. The structure of CrISY^{G150A} was solved with NADP$^+$ bound (PDB 5DCY), proving that an open form of the loop is trapped by this mutation (**Fig. 3b**, **Supplementary Fig. 14**). CrISY^{G150A} showed activity and product profiles comparable to those of wild-type enzyme (**Supplementary Table 2**, **Supplementary Fig. 7**), though the K_M did increase by five-fold (**Supplementary Table 2**), suggesting that the closed-loop conformation promotes substrate binding. Similarly, mutation of Phe152, which appears to lock the substrate in the linear conformation, caused a minor increase in K_M (**Supplementary Table 2**, **Supplementary Fig. 7**). The activity of these mutants suggests that this conformational flexibility near the active site (**Fig. 3**) is not essential for catalysis.

To validate the idea that the open conformation could accommodate the cyclized product, we modeled iridodial into the G150A structure (**Fig. 2c**). The model indicates that a number of hydrophobic residues define the 'open' binding pocket (Phe149, Ile151, Leu203, Met213, Phe342, Ile345). One side of the binding pocket is lined by a positive electrostatic potential (NADP$^+$-nicotinamide, Lys146, backbone amides), presumably stabilizing negative charge in the cyclizing enolate. The conformational flexibility at loop Gly150–Asp162 might allow both tight binding of the linear 8-oxogeranial and subsequent cyclization.

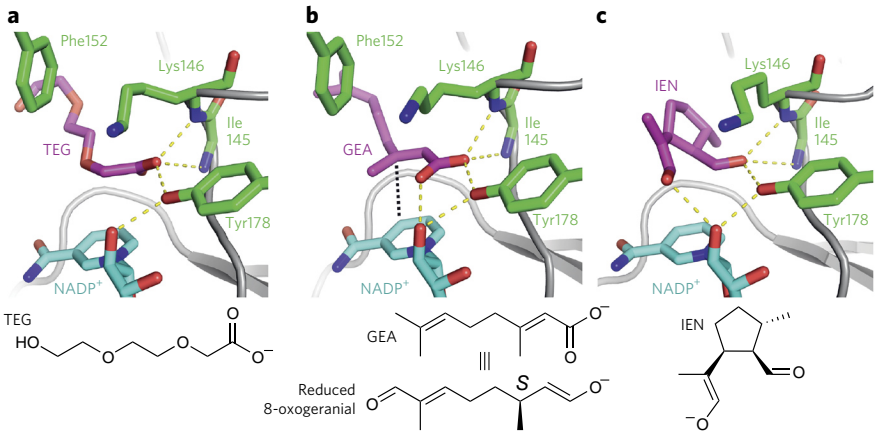

Figure 2 | Structure of ISY (*Catharanthus roseus*, CrISY). The protein backbone is shown in gray, binding-pocket residues and the backbone of residue Ile145 as sticks (carbon: green, oxygen: red, nitrogen: blue), NADP$^+$ as sticks with carbons in cyan, and additional ligands as sticks in magenta. The side chain of residue Ile145 is omitted for clarity. Relevant polar contacts with bonding distances between 2.5 and 3.5 Å are highlighted with yellow dashes. (**a**) CrISY co-crystallized with NADP$^+$ shows an adventitious ligand in the active site (PDB code 5DCU), most likely triethylene glycol carboxylic acid (TEG). (**b**) Active site of CrISY co-crystallized with geranic acid (GEA; PDB code 5DF1). The carbons equivalent to the start and end points of the hydride transfer from NADPH to 8-oxogeranial are connected with a black dotted line (distance: 3.4 Å). (**c**) The enolate form of iridodial (IEN) modeled into the active site of the 'open' conformation G150A mutant (see text and **Fig. 3**).

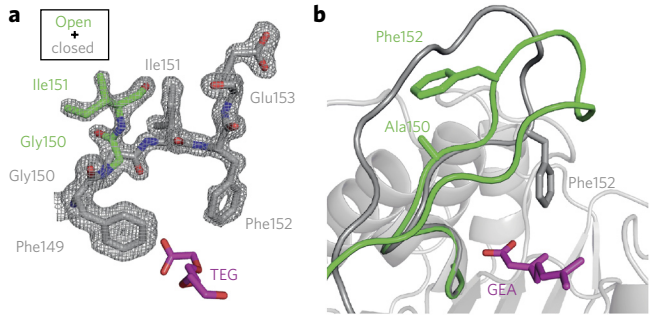

Figure 3 | Open and closed forms of loop Gly150–Val158 in CrISY. All electron density maps are of the form $2mF_{obs} - dF_{calc}$, contoured at 0.8 σ. (**a**) TEG-bound structure at 1.4-Å resolution revealed the Gly150–Val158 loop in 'open' and 'closed' conformations. In the closed conformation, Phe152 contacts the ligand. (**b**) Superposition of the GEA-bound structure (gray) and the loop of CrISY^{G150A} (PDB 5DCY, green). With GEA bound, this loop is exclusively in the closed conformation, and in mutant CrISY^{G150A}, the loop is exclusively in an open conformation (**Supplementary Fig. 14**).

The factors that control substrate specificity in the P5βR family are not well understood[21]. Although some P5βR enzymes from non-iridoid-producing plants cyclize 8-oxogeranial[22], the *D. lanata* P5βR that has been crystallized produces negligible iridodial or nepetalactol product (**Supplementary Fig. 7**). Moreover, progesterone consumption by CrISY is barely detectable (**Supplementary Fig. 8**). The inhibitor-bound structure of CrISY suggests that the Ala343–Ser354 loop plays a role in substrate recognition (**Supplementary Fig. 2h**). Ile347 of the DlP5βR sequence (Ala346 in CrISY) may clash with the linearly oriented 8-oxogeranial substrate, thus preventing binding (**Supplementary Fig. 9**). Indeed, when the Ala343–Ser354 loop was swapped with the corresponding DlP5βR sequence, ISY activity of the chimera was diminished by 2,000-fold (**Supplementary Table 2** and **Supplementary Fig. 10**), and this was accompanied by a substantial increase in progesterone turnover (**Supplementary Fig. 8**). Additionally, the Phe149–Pro160 loop exhibits a change in structure compared to DlP5βR (**Supplementary Fig. 2e**). Swapping this loop with the corresponding DlP5βR sequence improved progesterone reduction without greatly affecting 8-oxogeranial turnover (**Supplementary Figs. 7** and **8**). These data suggest that the features controlling substrate selectivity are surprisingly subtle, making it challenging to predict function directly from sequence. Continued analysis of the physiological functions of the many P5βR homologs found in the plant kingdom[21] will provide further insights into the basis of substrate specificity in the P5βR family.

Structural analyses have greatly facilitated the engineering and directed evolution of canonical terpene synthases[23,24]. These structures of CrISY will set the stage for efforts to produce variants that can be used in synthetic biology, metabolic engineering and biocatalysis applications. Repurposing iridoid synthase on the basis of mechanistic and structural information could unlock new monoterpene chemical space and also provide further insight into the balance between substrate reduction and cyclization.

Received 30 August 2015; accepted 8 October 2015;
published online 9 November 2015

Methods

Methods and any associated references are available in the online version of the paper.

Accession codes. CrISY in apo form, PDB code 5DCW; CrISY with NADP$^+$ bound, PDB code 5DCU; CrISY with geranic acid and NADP$^+$ bound, PDB code 5DF1; CrISY^{G150A} with NADP$^+$ bound, PDB code 5DCY.

References

1. Dobler, S., Petschenka, G. & Pankoke, H. *Phytochemistry* **72**, 1593–1604 (2011).
2. Birkett, M.A. & Pickett, J.A. *Phytochemistry* **62**, 651–656 (2003).
3. Tundis, R., Loizzo, M.R., Menichini, F., Statti, G.A. & Menichini, F. *Mini Rev. Med. Chem.* **8**, 399–420 (2008).
4. Viljoen, A., Mncwangi, N. & Vermaak, I. *Curr. Med. Chem.* **19**, 2104–2127 (2012).
5. Dinda, B., Debnath, S. & Banik, R. *Chem. Pharm. Bull. (Tokyo)* **59**, 803–833 (2011).
6. Geu-Flores, F. *et al. Nature* **492**, 138–142 (2012).
7. O'Connor, S.E. & Maresh, J.J. *Nat. Prod. Rep.* **23**, 532–547 (2006).
8. Lindemann, P. *Steroids* doi:10.1016/j.steroids.2015.08.007 (14 August 2015).
9. Kavanagh, K.L., Jörnvall, H., Persson, B. & Oppermann, U. *Cell. Mol. Life Sci.* **65**, 3895–3906 (2008).
10. Gärtner, D.E., Wendroth, S. & Seitz, H.U. *FEBS Lett.* **271**, 239–242 (1990).
11. Caspi, E. & Lewis, D.O. *Science* **156**, 519–520 (1967).
12. Bjelic, S. *et al. ACS Chem. Biol.* **8**, 749–757 (2013).
13. Kunert, M. *et al. ChemBioChem* **14**, 353–360 (2013).
14. Thorn, A. *et al. J. Biol. Chem.* **283**, 17260–17269 (2008).
15. Lindner, S., Geu-Flores, F., Bräse, S., Sherden, N.H. & O'Connor, S.E. *Chem. Biol.* **21**, 1452–1456 (2014).
16. Garrabou, X., Beck, T. & Hilvert, D. *Angew. Chem. Int. Edn Engl.* **54**, 5609–5612 (2015).
17. Rudolph, K., Bauer, P., Schmid, B., Mueller-Uri, F. & Kreis, W. *Biochimie* **101**, 31–38 (2014).
18. Winberg, J.O., Brendskag, M.K., Sylte, I., Lindstad, R.I. & McKinley-McKee, J.S. *J. Mol. Biol.* **294**, 601–616 (1999).
19. Petersen, J. *et al. J. Biomol. Struct. Dyn.* doi:10.1080/07391102.2015.1088797 (12 October 2015).
20. Lovell, S.C. *et al. Proteins* **50**, 437–450 (2003).
21. Munkert, J. *et al. Plant Biol.* doi:10.1111/plb.12361 (14 July 2015).
22. Munkert, J. *et al. Mol. Plant* **8**, 136–152 (2015).
23. Christianson, D.W. *Chem. Rev.* **106**, 3412–3442 (2006).
24. Cane, D.E. & Ikeda, H. *Acc. Chem. Res.* **45**, 463–472 (2012).

Acknowledgments

Funds were made available by the European Research Council (ERC R20359) and a BBSRC Institute Strategic Programme grant (MET; BB/J004561/1) to S.E.O'C., and an SNF Early Postdoc Mobility Fellowship to H.K. The Diamond Light Source provided access to beamlines I03, I04 and I04-1 (proposal MX9475).

Author contributions

L.C., H. K., N.H.S., F.G.-F., and S.E.O'C. designed the project; L.C., H.K., M.O.K., and F.G.-F. performed molecular cloning/enzyme assays; C.E.M.S. assisted with crystallization and X-ray data acquisition; H.K., F.G.-F., and N.H.S. performed chemical synthesis; H.K., L.C., and D.M.L. refined structures; S.E.O'C. supervised the work; H.K. and S.E.O'C. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to S.E.O'C.

ONLINE METHODS

Chemicals. NADPH and NADP$^+$ were purchased from Sigma-Aldrich. Carbenicillin was purchased from Formedium. Substrate 8-oxogeranial was synthesized as previously described[6]. All salts and solvents used in this study were of analytical grade unless otherwise specified.

Construction of expression plasmids. For cloning purposes, *Escherichia coli* strain TOP10 was employed. KOD Hot Start polymerase (Merck Millipore) was used for all PCR amplifications. The native *CrISY* gene was amplified by PCR from pCR8-*ISY*[6] with primers CrISY22dN_f and CrISY_r (**Supplementary Table 4**) and cloned via In-Fusion (Clontech Laboratories) into expression vector pOPINF[25]. The resulting vector pOPINF-*CrISY22dNKAF* encodes protein CrISY$^{22\Delta N\text{-}KAF}$ identical to Uniprot entry K7WDL7 with an N-terminal truncation of 22 amino acids, an N-terminal purification tag (MAHHHHHHSSGLEVLFQGP), a mutation that corrects an error in the database (D87N), and a C-terminal appendage of three amino acids (KAF). After 3C cleavage of the tag, this protein was used for crystallization of the apo form and the NADP$^+$–TEG complex. A second version of the *CrISY* gene and a gene encoding DlP5βR were synthesized at Life Technologies with codons optimized for expression in *E. coli*.

CrISY. 5′-ATGAGTTGGTGGTGGAAACGTAGCATTGGTGCAGG
TAAAAATCTGCCGAATCAGAATAAAGAAAACGGCGTGTGC
AAAAGCTATAAAAGCGTTGCACTGGTTGTTGGTGTTACCGGTA
TTGTTGGTAGCAGCCTGGCAGAAGTTCTGAAACTGCCGGATAC
ACCGGGTGGTCCGTGGAAAGTTTATGGTGTTGCACGTCGT
CCGTGTCCGGTTTGGCTGGCAAAAAAACCGGTTGAATATAT
CCAGTGTGATGTGAGCAATAACCAAGAAACCATTAGCAAAC
TGAGTCCGCTGAAAGATATCACCCATATCTTTTATGTTAGCTG
GATTGGTAGCGAAGATTGTCAGACCAATGCAACCATGTT
TAAAAACATTCTGAACAGCGTTATTCCGAATGCAAGCAAT
CTGCAGCATGTTTGCCTGCAGACCGGTATCAAACATTATTTT
GGCATTTTTGAAGAGGGCAGCAAAGTTGTTCCGCATGAT
AGCCCGTTTACCGAAGATCTGCCTCGTCTGAATGTTCCGAACTTTTA
TCATGATCTGGAAGATATCCTGTATGAAGAAACCGGCAAAAA
TAACCTGACCTGGTCAGTTCATCGTCCGGCACTGGTGT
TTGGTTTTAGCCCGTGTAGCATGATGAATATTGTTA
GCACCCTGTGTGTTTATGCCACCATTTGCAAACATGA
AAATAAAGCCCTGGTTTACCCTGGCAGCAAAAATAGCTGGA
ATTGTTATGCAGATGCAGTTGATGCAGATCTGGTTGCCGAAC
ATGAAATTTGGGCAGCAGTTGATCCGAAAGCAAAAAATCAGGTTC
TGAATTGCAATAACGGCGACGTGTTTAAATGGAAACA
TATCTGGAAAAAACTGGCCGAAGAATTTGGCATTGAAAT
GGTTGGTTATGTGGAAGGTAAAGAACAGGTTAGCCTG
GCCGAACTGATGAAAGATAAAGATCAGGTTTGGGACGAAA
TCGTGAAAAAAAACAATCTGGTTCCGACCAAACTGAAAGA
AATTGCAGCATTTTGGTTTGCCGATATTGCCTTTTGTAGC
GAAAATCTGATTAGCAGCATGAACAAAAGCAAAGAACT
GGGTTTTCTGGGCTTTCGCAATAGCATGAAAAGCTTTGTGAGCT
GCATTGATAAAATGCGCGATTATCGTTTTATTCCG-3′.

DlP5βR. 5′-ATGAGTTGGTGGTGGGCAGGCGCAATTGGTGCA
GCAAAAAAACGTCTGGAAGAGGATGATGCACAGCCGAAACATA
GCAGCGTTGCACTGATTGTTGGTGTTACCGGTATTATTGGTAA
TAGCCTGGCCGAAATTCTGCCGCTGGCAGATACACCGGGT
GGTCCGTGGAAAGTTTATGGTGTTGCACGTCGTACCCGTCC
GGCATGGCATGAAGATAATCCGATTAACTATGTGCAG
TGCGATATTAGCGATCCGGATGATAGCCAGGCAAAAC
TGAGTCCGCTGACCGATGTTACCCATGTTTTTTATGTGACC
TGGGCAAATCGTAGCACCGAACAAGAAAATTGTGAAGCCA
ACAGCAAAATGTTTCGCAATGTTCTGGATGCCGTTATTCC
GAATTGTCCGAATCTGAAACACATTAGCCTGCAGAC
CGGTCGTAAACATTATATGGGTCCGTTTGAAAGCTATGGCA
AAATCGAAAGCCATGATCCGCCTTATACCGAAGATCTG
CCTCGTCTGAAATATATGAACTTCTATTACGATCTGGAAGA
TATCATGCTGGAAGAAGTCGAAAAAAAAGAAGGCCTGACCTGG
TCAGTTCATCGTCCGGGTAACATTTTTGGTTTTAGCCCGTATA
GCATGATGAATCTGGTTGGCACCCTGTGTGTTTATGCAG
CAATTTGTAAACATGAAGGTAAAGTGCTGCGTTTTACCGGTTG
TAAAGCAGCATGGGATGGTTATAGCGATTGTAGTGATGC
AGATCTGATTGCCGAACATCATATTTGGGCAGCAGTTG
ATCCGTATGCAAAAAATGAAGCCTTTAATGTGAGCAACGGC
GACGTGTTTAAATGGAAACACTTTTGGAAAGTTCTGGCCGA
ACAGTTTGGTGTTGAATGTGGTGAATATGAAGAAGGCGTTGATCTG
AAACTGCAGGATCTGATGAAAGGTAAAGAACCGGTTTGGGAAGA
AATTGTGCGTGAAAATGGTCTGACCCCGACCAAACTGAAAGAT
GTTGGTATTTGGTGGTTTGGTGATGTGATTCTGGGTAATGAA
TGTTTCCTGGACAGCATGAACAAAAGCAAAGAACATGGTT
TTCTGGGCTTCCGCAATAGCAAAAATGCCTTTATTAGCTGGAT
CGACAAAGCCAAAGCCTATAAAATCGTTCCG-3′.

The synthetic genes were amplified with primer combinations indicated in **Supplementary Table 4** and cloned into expression vector pOPINF to give pOPINF-*CrISY,* pOPINF-*CrISY22dN*, and pOPINF-*DlP5bR*. pOPINF-*CrISY* and pOPINF-*CrISY22dN* lack the codons for the C-terminal KAF tag present in pOPINF-*CrISY22dNKAF* but also encode residue Asn87 that differs from the Uniprot entry. Protein CrISY$^{22\Delta N}$ expressed from plasmid pOPINF-*CrISY22dN* was used for crystallization of the NADP$^+$/GEA complex. pOPINF-*CrISY* encodes all N-terminal residues of wild-type CrISY. Point mutations and swaps were introduced into the codon optimized gene by PCR. Mutation G150A was introduced, for instance, by amplifying a first fragment with primer pair CrISY22dN_synth_f/CrISY_F149_r and a second fragment carrying the mutation with primer pair CrISY_G150A_f/CrISY_synth_r. Both fragments were assembled in a second PCR reaction containing the fragments and the outer primers CrISY22dN_synth_f and CrISY_synth_r and cloned into the pOPINF vector. The resulting plasmid pOPINF-*CrISY22dN-G150A* was used to express protein for crystallization of the CrISY$^{22\Delta N\text{-}G150A}$/NADP$^+$ complex. For the sake of simplicity, CrISY variants with the 22ΔN truncation and with or without C-terminal KAF tag are all referred to as 'CrISY' in the main text.

Protein expression and purification. Protein for the apo structure and the NADP$^+$ complex was expressed in strain Rosetta 2 (EMD Millipore) transformed with pOPINF-*CrISY22dNKAF*. All codon optimized genes were expressed in strain soluBL21 (DE3) (Genlantis). The pOPINF vector encodes an N-terminal His-tag that is cleavable with 3C protease. Cells were precultured overnight at 37 °C in LB medium containing 100 µg/mL carbenicillin. Aliquots of 1 mL each were used to inoculate two 2-L flasks containing 2YT medium and carbenicillin. When cultures reached an OD_{600} of 0.5–0.8 after ~6 h shaking at 37 °C, protein production was induced by adding IPTG at a final concentration of 500 µM. Protein was expressed at 18 °C for ~16 h. Cells were harvested by centrifugation, the supernatant was discarded and pellets were resuspended in 100 mL of 50 mM Tris-HCl buffer (pH 7.0) containing 300 mM NaCl, 1 mM DTT, one tablet of Complete EDTA free protease inhibitor (Roche), and 0.2 mg/mL lysozyme. The cells were disrupted by 7 min sonication on ice in cycles of 2 s sonication followed by a 3 s break. All subsequent steps were conducted at 4 °C. The lysate was centrifuged for 20 min at 35,000*g* and the supernatant containing the soluble protein was applied on a 5 mL HisTrap column connected to an Äkta Xpress purifier (GE healthcare). His-tagged protein was eluted with a linear gradient of 20 mM to 500 mM imidazole in 50 mM Tris/glycine buffer adjusted to pH 8.0 with hydrochloric acid and supplemented with 5% (v/v) glycerol, 0.5 M NaCl, and 1 mM DTT. Fractions containing the protein of interest were collected, incubated with 3C protease (120 µg) for 1 h at room temperature, and concentrated to 2–3 mL in an Amicon 10 kDa MWCO centrifugal filter (Millipore). The concentrated solution was further purified by size-exclusion chromatography on a Superdex 200 16/60 GF column (GE Healthcare). Fractions corresponding to the molecular weight of the CrISY dimer (83 kDa) were collected, combined, and washed with 20 mM Tris-HCl buffer (pH 8.0) containing 20 mM NaCl and 1 mM DTT (low salt buffer) in an Amicon 10 kDa MWCO centrifugal filter. After concentration to ~1 mL, protein was injected on a MonoQ 5/50 GL anion exchange column and eluted with a linear gradient of 20 mM to 1,000 mM NaCl in low salt buffer. Fractions containing the protein of interest were combined and stored at −20 °C after flash freezing in liquid nitrogen. For all mutants produced exclusively for kinetic characterization and DlP5βR, 3C cleavage and the final anion exchange chromatography step were omitted. Protein concentrations were determined spectrophotometrically at 280 nm using extinction coefficients calculated with the protparam tool (http://web.expasy.org/protparam/). A typical yield for CrISY produced with codon optimized genes on a 2-L scale was 4–5 mg after anion exchange chromatography. Proteins purified in this study were more than 95% pure as judged by SDS-PAGE (**Supplementary Fig. 12**).

doi:10.1038/nchembio.1955

Protein crystallization. For crystallization, protein was buffer exchanged into crystallization buffer (5 mM MOPS buffer pH 7.0, 40 mM NaCl and 5 mM tris(2-carboxyethyl)phosphine [TCEP]) in Amicon 10 kDa MWCO centrifugal filters and concentrated to 7–10 mg/mL. Crystallization buffer and, if applicable, $NADP^+$ (Sigma N0505) were added to reach a final concentration of 7 mg/mL protein and 2 mM $NADP^+$. The resulting solution was passed through a 0.1 µm PVDF filter (Durapore). Crystal screens were conducted by sitting-drop vapor diffusion in MRC2 96-well crystallization plates (Swissci) with a mixture of 0.3 µL well solution from the PEGs and PACT suite (Qiagen) and 0.3 µL protein solution. Solutions were dispensed either by an OryxNano or an Oryx8 robot (Douglas Instruments). The initial screens provided several hits after only 1–2 days of incubation at 20 °C. The best crystals for the complex $CrISY^{22\Delta N\text{-}KAF}/NADP^+$ were obtained with well solution containing 0.1 M MIB buffer (pH 6.0) and 25% PEG 1500. The apo enzyme crystallized with 0.2 M Na/K tartrate and 20% PEG 3500, and $CrISY^{22\Delta N\text{-}G150A}/NADP^+$ crystallized with 0.1 M SPG buffer pH 6.0 and 25% PEG 1500. MIB buffer is a mixture of malonate, imidazole, and borate with a molar ratio of 2:3:3 and SPG buffer is a mixture of succinate, phosphate, and glycine with a molar ratio of 2:7:7. Co-crystallization of $CrISY^{22\Delta N}$ with $NADP^+$ and geranic acid (Aldrich 427764) was performed by sitting-drop vapor diffusion combining 0.1 µL of 3 mM geranic acid diluted in crystallization buffer from a 100 mM stock in THF, 0.2 µL filtered protein solution containing 10.5 mg/mL $CrISY^{22\Delta N}$ and 1.5 mM $NADP^+$ in crystallization buffer, and 0.3 µL well solution containing 100 mM MIB buffer pH 6.0 and 20% PEG 1500. Crystals were harvested from the mother liquor using LithoLoops (Molecular Dimensions), cryoprotected with well solution supplemented with 25% (v/v) ethylene glycol and cryo-cooled in liquid nitrogen.

Crystallographic methods. X-ray data sets were recorded on one of three beamlines at the Diamond Light Source (Oxfordshire, UK) (beamline 5DCU, I04; 5DCW, I04; 5DF1, I03; 5DCY, I04-1) at wavelengths of 0.9000–0.9795 Å (5DCU, 0.9000; 5DCW, 0.9000; 5DF1, 0.9795; 5DCY, 0.9173) using a Pilatus 6M detector (Dectris) with the crystals maintained at 100 K by a Cryojet cryocooler (Oxford Instruments). Diffraction data were integrated using XDS[26] and scaled and merged using AIMLESS[27]; data collection statistics are summarized in **Supplementary Table 1**. An initial model of the $CrISY^{22\Delta N\text{-}KAF}/NADP^+$ complex was obtained by molecular replacement using PHASER[28] with the DlP5βR structure (PDB code 2V6G)[14] as the search model. Manual corrections were made in COOT[29] and the model was refined with REFMAC5[30]. As the crystals of the other two $NADP^+$ complexes belonged to the same space group ($C222_1$), their structures could be solved by refining against the $CrISY^{22\Delta N\text{-}KAF}/NADP^+$ model. The structure of the apo complex, belonging to space group H32, was solved by molecular replacement with the $CrISY^{22\Delta N\text{-}KAF}/NADP^+$ model. Ligand restraints were calculated on the Grade Web Server (http://grade.globalphasing.org/). In general, parameters for REFMAC5 refinement were optimized with PDB-REDO[31]. All models were subsequently refined with TLS group definitions obtained from the TLS-MD server[32]. Model geometries were validated with the MOLPROBITY[33] tool before submission to the PDB. The statistics of the final models are summarized in **Supplementary Table 1**. Additional statistics for R_{meas}: 5DCU, 0.074 (1.548); 5DCW, 0.131 (1.030); 5DCY, 0.113 (1.702); 5DF1, 0.114 (1.822) and $CC_{½}$: 5DCU, 0.999 (0.569); 5DCW, 0.999 (0.799); 5DCY, 0.998 (0.522); 5DF1, 0.995 (0.543) (where values in parentheses are for highest-resolution shell) were also noted. Ramachandran statistics (Favored/allowed/outlier (%)) are 5DCU, 98.4/1.6/0.0; 5DCW, 98.1/1.9/0.0; 5DCY, 98.1/1.9/0.0; 5DF1, 98.3/1.7/0.0.

Structure 5DCW crystallized as a monomer. Where there are two subunits per asymmetric unit (5DCU, 5DCY and 5DF1), these do not constitute the biological dimer. In both crystal forms of CrISY, biological dimers consistent with that observed for DlP5βR were detected using the PISA server (http://www.ebi.ac.uk/pdbe/pisa/) (**Supplementary Fig. 11**)[34]. Although the $C222_1$ crystal form observed for all the holo CrISY structures contains two monomers in the asymmetric unit (ASU), these do not correspond to the biological dimer. Instead, the ASU comprises two halves of two discrete biological dimers. Each of these is generated through the application of a different two-fold crystallographic symmetry operator. Similarly, an equivalent dimer is generated by two-fold crystallographic symmetry from the single monomer in the ASU of the H32 crystal form observed for the apo CrISY structure. All possible structural superpositions were made within the resultant set of CrISY structures at both the monomer and dimer level, and the corresponding r.m.s. deviation values are summarized in **Supplementary Table 3**. These showed that there was no significant variation in the relationship between the individual subunits within the biological dimers since the r.m.s. deviation values calculated for the dimer superpositions were relatively low and not significantly larger than those calculated for the monomer superpositions. Similar values were observed for comparisons between the CrISY structures and those of DlP5βR. Moreover, the protein surface areas buried at the dimer interfaces were reasonably consistent, lying in the range 918–979 $Å^2$ for CrISY and 888–919 $Å^2$ for DlP5βR. All structural figures were prepared using PyMOL (PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.).

Enzyme assay by NADPH consumption. Kinetics of NADPH consumption were determined spectrophotometrically on a PerkinElmer Lambda 35 instrument at a wavelength of 340 nm in cuvettes with 1 cm path length. Reactions contained 50 µM NADPH, 200 mM MOPS buffer pH 7.0, 100 mM sodium chloride and 2 nM to 4 µM enzyme, depending on the activity, in a total volume of 800 µL. 8-oxogeranial was added from a stock solution in inhibitor free THF. A co-solvent concentration of 1% was maintained in the assay to ensure solubility. Cuvettes were equilibrated to 25 °C for 3 min before the reaction was initiated by addition of enzyme. Absorbance values were recorded at a rate of 1 Hz and initial velocities were linearly fit to at least 10 data points. Plots of initial velocities versus substrate concentration were nonlinearly fit to the Michaelis-Menten equation in Kaleidagraph 4.0 in order to obtain values of k_{cat} and K_M.

GC-MS analysis. For analyzing reaction products by GC-MS, reactions contained 0.5–4 µM of enzyme, 300–600 µM of NADPH (Sigma N7505), 100 mM sodium chloride, and 300–500 µM of 8-oxogeranial or 200 µM of progesterone in 200 mM MOPS buffer pH 7.0 in a volume of 50 µL. 8-oxogeranial was solubilized with 0.5% (v/v) inhibitor free THF as co-solvent and progesterone with 2% MeOH (v/v). Reactions were incubated for 40–120 min at 25–30 °C. The reaction mixture was transferred to a cylindrical 400 µL flat bottom glass insert (Agilent 5181-3377) in a GC-MS vial closed with a PTFE septum. Hydrophobic compounds were extracted with 100 µL ethyl acetate. The supernatant (1–5 µL) was injected in splitless mode at an inlet temperature of 220 °C on a Hewlett Packard 6890 GC-MS equipped with a 5973 mass selective detector (MSD), and an Agilent 7683B series injector and autosampler. Separation was performed on an Agilent HP-5MS column (5% phenyl methyl siloxane; length: 30 m; diameter: 320 µm) for 8-oxogeranial reactions and on a Phenomenex ZB5-HT column (5% phenyl methyl siloxane; length: 35 m; diameter: 250 µm) with guard column for progesterone reactions. Helium was used as mobile phase at a constant flow rate of 1.0 mL/min and average velocity 36 cm/s. After 5 min at 60 °C, the column temperature was increased to 150 °C at a rate of 20 K/min, then to 280 °C at 45 K/min, and kept at 280 °C for another 4 min. A solvent delay of 5 min was allowed before collecting MS spectra at a fragmentation energy of 70 eV.

Modeling. For modeling the iridodial 7,8-enolate (IEN) into the structure of $CrISY^{22\Delta N\text{-}G150A}/NADP^+$, the lowest energy conformer of the relevant stereoisomer of iridodial 7,8-enolate (IEN) was predicted in Chem3D Pro 13.0 (PerkinElmer). Initial coordinates for the IEN ligand were obtained by pair fitting atoms O1, C1, and C2 of GEA in the $NADP^+$/GEA bound structure to the corresponding atoms in IEN. The resulting coordinates were then copied into the $CrISY^{22\Delta N\text{-}G150A}/NADP^+$ model. The IEN bound structure was prepared for energy minimization by deleting water molecules and alternate side chain conformations from the model, and rotating the side chains of Lys146 and Ile145 to avoid steric clashes with the docked ligand using PyMOL. Hydrogens and charges were added to the ligands $NADP^+$ and IEN in UCSF chimera (IEN: -1). The resulting mol2 files were converted to params files readable by Rosetta with the molfile2params script. Then, the structure was energy minimized in Rosetta energy terms with a Rosetta relax script[35].

Circular dichroism. Far ultraviolet (UV) CD spectra were recorded on a Chirascan Plus spectropolarimeter (Applied Photophysics) at 20 °C in 10 mM potassium phosphate buffer, pH 7.0. Protein concentrations were in the range 0.20–0.25 mg/mL. Spectra are presented as the CD absorption coefficient calculated on a mean residue ellipticity (MRE) basis.

25. Berrow, N.S. *et al. Nucleic Acids Res.* **35**, e45 (2007).
26. Kabsch, W. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
27. Evans, P.R. & Murshudov, G.N. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
28. McCoy, A.J. *et al. J. Appl. Crystallogr.* **40**, 658–674 (2007).
29. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
30. Winn, M.D., Murshudov, G.N. & Papiz, M.Z. *Methods Enzymol.* **374**, 300–321 (2003).
31. Joosten, R.P. *et al. J. Appl. Crystallogr.* **42**, 376–384 (2009).
32. Painter, J. & Merritt, E.A. *J. Appl. Crystallogr.* **39**, 109–111 (2006).
33. Chen, V.B. *et al. Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
34. Krissinel, E. & Henrick, K. *Computational Life Sciences* (Springer-Verlag, Berlin, Heidelberg, 2005).
35. Leaver-Fay, A. *et al. Methods Enzymol.* **487**, 545–574 (2011).

doi:10.1038/nchembio.1955