Journal Name

COMMUNICATION

The amino-terminal segment in the β-domain of δ-cadinene synthase is essential for catalysis

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Verónica González, Daniel J. Grundy, Juan A. Faraldos and Rudolf K. Allemann

Despite its distance from the active site the flexible amino-terminal segment (NTS) in the β-domain of the plant sesquiterpene cyclase δ-cadinene synthase (DCS) is essential for active site closure and desolvation events during catalysis.

Sesquiterpene synthases catalyze the conversion of (E,E)-farnesyl diphosphate (FDP, **1**) to more than 300 distinct C_{15}-isoprenoid hydrocarbons, which serve as precursors to the >7,000 oxygenated sesquiterpenoids found in plants, fungi and bacteria.[1] Catalysis by class I sesquiterpene synthases is dependent on a shared α-helical fold (α domain),[2] in which two conserved aspartate-rich motifs (DDXXD and NSE/DTE[3]) bind and activate the diphosphate group of the linear substrate **1** through coordination to a tri-nuclear Mg^{2+} cluster.[4] After cleavage of the C_1-O bond in FDP, the resulting farnesyl cation undergoes a series of electrophilic cyclisations and rearrangements to generate a final carbocation. Neutralization of the positive charge either through loss of a proton or by nucleophilic water attack gives a unique sesquiterpene hydrocarbon or alcohol.[5,6]

Inspection of the available X-ray crystal structures of single domain microbial sesquiterpene synthases complexed with unreactive analogues of FDP (**1**) and carbocationic intermediates[5,7] as well as recent computational work[8] suggest that complexation of the essential metal ions[4] induces structural rearrangements of flexible loops leading to the formation of the closed enzyme conformation that binds diphosphate **1** in the reactive conformation. Plant mono- and sesquiterpene synthases contain an additional N-terminal α-helical fold known as the β-domain that was first observed in the X-ray crystal structures of 5-epi-aristolochene synthase from *Nicotiana tabacum* (TEAS)[9] and subsequently in the structures of bornyl diphosphate synthase from *Salvia officinalis*[10] (BPPS) and limonene synthase from *Mentha spicata* (LS).[11] The catalytic function of β-domains is unclear. The amino-terminal segment (NTS) of β-domains appears to be disordered in unliganded open structures, i.e. in the absence of Mg^{2+} ions, unreactive FDP analogues or mimics of reactive carbocationic intermediates. However upon ligand binding to the α-domain, the active site adopts a closed conformation and the NTS becomes well ordered. Complexation of ligands leads to the formation of stabilizing α–β interdomain hydrogen bond interactions that anchor the flexible NTS segment to the dynamic A-C and J-K loops in the closed conformation.[9] The conserved amino-terminal segments may therefore play a role in stabilising the fully closed Michaelis complex where the NTS caps the active site to shield the reactive cationic intermediates from water.

δ-Cadinene synthase (DCS) from *Gossypium arboreum* is a typical α,β plant sesquiterpene synthase that catalyses the metal-dependent conversion of FDP (**1**) to δ-cadinene (**3**), the biological precursor of several cotton phytoalexins such as gossypol.[12] The NTS of DCS consists of the first 30 amino acids of the β-domain and reaches across the α-domain to partially cover the active site of the apo-enzyme (Figure 1).[13] Solution of the co-crystal structure of DCS with three putative Mg^{2+}-ions and the substrate analogue 2-fluorofarnesyl diphosphate (2F-FDP) (Figure 1) yielded a structure that closely matched that of unliganded DCS (Figure 1) with an open and flexible active site conformation; interestingly whereas three stabilizing interdomain interactions (P267/A25, P267/D26 and A269/Q28) appear to fix the NTS to the ordered active site A-C loop segment in the native structure of DCS, only one (A269/Q28) stabilizes the fully liganded DCS complex (Figure 1).[13] Although weakly stabilised closed conformations with more plastic active sites might be expected for class I terpene enzymes that bind to structurally different isoprenyl diphosphates,[11] the unproductive conformation adopted by 2F-FDP in DCS suggests that this complexed structure is not a good representation of the Michaelis complex. Furthermore the apparent lack of electrostatic interactions between the diphosphate group of 2F-FDP and the putative trinuclear metal cluster is in stark contrast to the closed structures observed for the Mg^{2+}-bound 2F-FDP and 2-fluorolinalyl diphosphate (2F-LDP) complexes of TEAS and LS.[11,14] These observations may argue against an involvement of the NTS of DCS in the open to closed conformational transition commonly assumed to be required for the stabilization of the Michaelis complex.

[a] School of Chemistry, Cardiff University, Main Building Park Place, Cardiff CF10 3AT, United Kingdom
Electronic Supplementary Information (ESI) available: See DOI: 10.1039/x0xx00000x

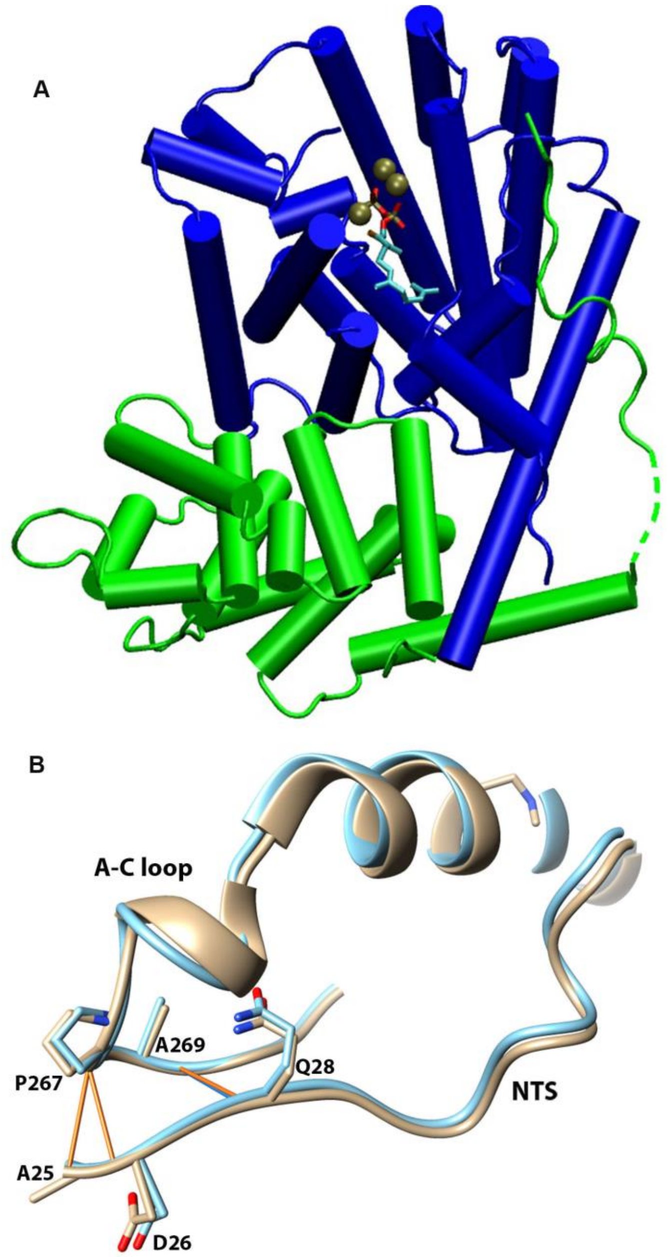
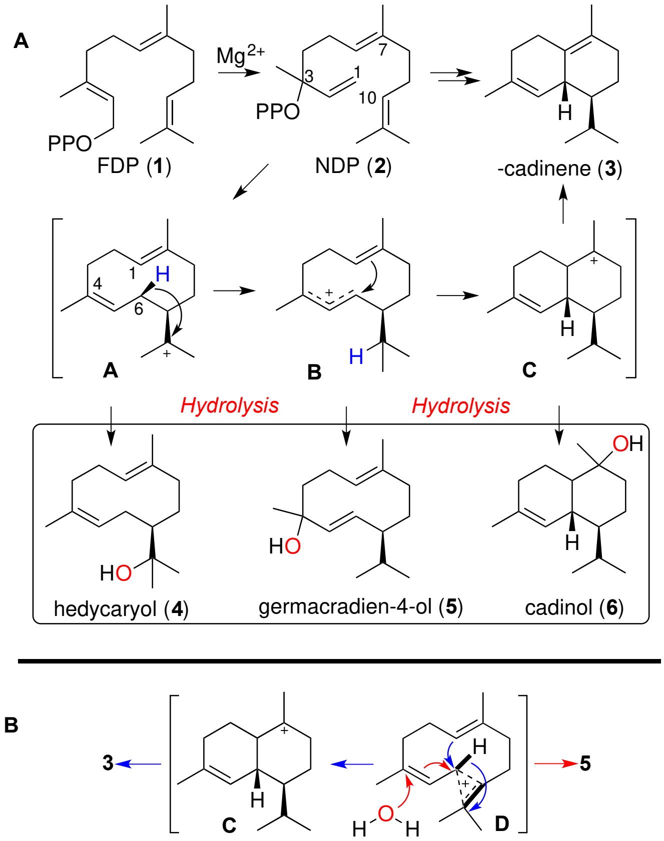


Figure 1: Panel A. Cartoon representation of the X-ray crystal structure of DCS complexed with 2F-FPP and three putative Mg^{2+}-ions (3g4f.pdb) illustrating the capping of the active site in the α-domain (blue) by the amino-terminal segment (NTS) of the β-domain (green).[13] The first amino acid of the β-domain is Ala-25; segments M1 to K24 and K42 to I44 (dotted line) were disordered. Panel B. Cartoon representation of the NTS in apo-DCS (3g4d.pdb, sand) and DCS complexed with 2-fluoro-FDP (3g4f.pdb, light blue). Interactions P267-A25, P267-D26 and A269-Q28 fixing the NTS to the A-C loop of the α-domain are highlighted in orange (apo DCS) and blue (liganded DCS).

The NTS in class I plant terpene synthases is on average more than 11 Å away from their respective active sites. Nevertheless, due to the proposed capping function of the NTS, inefficient anchoring of the NTS to the α-domain will most likely lead to poor desolvation, reduced catalytic activity[15] and alteration of the product profile. To explore the effects of partial or complete removal of the NTS, and those of single amino acid substitutions in the NTS on the formation of the Michaelis complex of DCS, a series of truncated (M8, M20 and M30), hybrid (CH-DCS) and single substitution CH-DCS and DCS variants (CH-DCS-S24W and DCS-S30W) were produced and characterised (Table 1 and ESI).



Scheme 1. Panel A. Conversion of FDP (**1**) to δ-cadinene (**3**) by DCS via NDP (**2**) and carbocations **A**, **B** and **C**. Framed are the expected hydrolysis products arising from these cations. For more mechanistic details see reference 19. Compound **5** was assigned as (-)-germacradien-4-ol based on chiral GC measurements and GC/MS comparisons with an authentic sample (see ESI). Panel B. Postulated enzymatic formation of **3** and **5** from a bridged carbocation **D**.

GC-MS analysis of the pentane extractable products generated from incubations of FDP (**1**) with the N-terminally truncated proteins M8, M20 and M30, in which the first 8, 20 and 30 amino acid residues of DCS had been removed respectively, revealed that while the catalytic efficiencies (k_{cat}/K_M) of M8 and M20 were similar to that of the wild type enzyme, deletion of 30 residues (M30) severely impaired the catalytic function of DCS (Table 1). Interestingly, the progressive deletion of NTS segments of approximately 10-residues in length led to the production of increasing amounts of a sesquiterpene alcohol, reaching 57% in M20 and 94% in M30 (Table 1). Comparison of the MS fragmentation patterns and the gas chromatograms of this alcohol and the product (**5**) generated by the microbial (-)-germacradiene-4-ol synthase[16] established the chemical identity of the M8-M30-produced sesquiterpenol as (-)-germacradiene-4-ol (**5**) (ESI). The production of **5** by M8, M20 and M30 is most likely the result of premature quenching of germacradiene-6-yl cation (**B**) by a water molecule (Scheme 1) due to less efficient active site closure and desolvation in the truncated enzymes.[7,17] The absence of any detectable hedycaryol (**4**)[18] from incubations with M8, M20 and M30 demonstrates that the formation of carbocation **B** occurs more rapidly than the neutralisation of its precursor carbocation **A** by a water molecule (Scheme 1). In principle, the competing hydrolytic and alkylation pathways that convert **B** to **5** or

Please do not adjust margins

3 (Scheme 1) could involve a bridged carbocation corresponding to the common branching intermediate **D**, from which germacradiene-4-ol and δ-cadinene can be formed (Scheme 1). It has been suggested previously that the reaction pathways to germacrene and humulenes might be connected through a similar bridged carbocation intermediate that resembles bicyclogermacrene.[19] Mechanistic studies with DCS[20] have shown that FDP (**1**) is first converted to nerolidyl diphosphate (NDP, **2**) *via* an enzyme bound farnesyl cation / $^{-}$OPP ion pair that most likely prevents the intermolecular nucleophilic attack of a water molecule, thus explaining the absence of farnesol or nerolidol when M8, M20 and M30 are incubated with **1**.

Table 1. Steady-state kinetic parameters and product profile for all proteins.[a]

Enzyme	**3**[b]	**5**[b]	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)
DCS	98	2	0.010	3.2	3.1
M8	85	15	0.008	2.9	2.8
M20	43	57	0.003	2.5	1.2
M30	6	94		n.m.[c]	
CH-DCS	25	75	0.006	4.7	1.3
CH-DCS-S24W				Inactive	
DCS-S30W				Inactive	

[(a)] Kinetic data are from 3 independent measurements with an error smaller than 10%. [(b)] Percentage (%). [(c)] Not measurable in the kinetic assay used here.

The low catalytic efficiency of M30 together with the near wild type efficiency of M8 and M20 suggest that the sequence M[21]RPKADF[27]QPS[30] in DCS, which is conserved as XRPXXXFXPS in plant sesquiterpene cyclases, is vital for efficient catalysis and product fidelity. For DCS, this deca-peptide is the minimal amino-terminal segment required to stabilise the Michaelis complex. Despite the unusual, open X-ray crystal structure of DCS in its ligand bound form,[13] the production of increasing amounts of germacradien-4-ol (**5**) with increasing extent of N-terminal truncation clearly implicates the NTS in the shielding of the active site from bulk solvent.

Sequence alignments of DCS with other class I plant sesquiterpene synthases (Figure 2 and ESI) indicate that there is essentially no N-terminal sequence similarity outside the consensus motif XRPXXXFXPS, i.e. M[21]RPKADF[27]QPS[30] in DCS. To assess the function of the XRPXXXFXPS sequence as a determinant of product specificity, a domain-swapping strategy[21] was adopted for DCS, in which the hybrid CH-DCS was constructed by extending the catalytically impaired M30 at its N-terminus with the first 24 amino acids (M1-S24) of (E)-β-farnesene synthase (EBFS) from *Mentha* x *piperita* (Fig. 2).[22]

```
                  Amino N-terminal segment (NTS)
                        M8          M20       M30
DCS      1  MASQVSQMPSSSPLSSNKDEMRPKADFQPSIWGDFFLNCPDKNI 44
EBFS     1          MATNGVVISCLREVRPPMTKHAPSMWTDTFSNFSLDDK 38
CH-DCS   1          MATNGVVISCLREVRPPMTKHAPSIWGDLFLNCPDKNI 38
GAS      1  MAAVEANGTLQANTKTTTEPVRPLANFPPSVWGDRFLSFSLDNK 44
ADS      1                    MSLTEEKPIRPIANFPPSIWGDQFLIYEKQVE 32
TEAS     1            MASAAVANYEEEIVRPVADFSPSLWGDQFLSFSIDNQ 37
```

Figure 2. Amino-terminal sequence alignment of 6 plant sesquiterpene synthases, including the hybrid CH-DCS, and showing the truncation points to generate M8, M20 and M30. (GenBank accession numbers are X96429 for DCS, AJ786642 for EBFS, AJ304452 for GAS, JQ319661 for ADS, and JQ812050 for TEAS.

The catalytic efficiency (k_{cat}/K_M) of CH-DCS was similar to that of M20 (Table 1) and the product profile resembled that of M30. The increased production of germacradiene-4-ol (**5**) by CH-DCS when compared to DCS and M20 further underlines the catalytic relevance of the NTS motif and the existence of catalytic and product specificity determinants within the consensus sequence XRPXXXFXS. In agreement, CH-DCS-S24W, which was produced as a cloning artifact, was inactive, thus supporting the essential role of this serine residue in stabilising the closed conformation of CH-DCS. This rationale was further corroborated by the observation that no activity could be detected for DCS-S30W. Remarkably, this simple swapping experiment illustrates how both the length and the composition of the NTS modulate the desolvation process that is required for effective catalysis by DCS. Specific changes to the NTS can also alter the product specificity and lead to novel enzymes (CH-DCS, Table 1) that produce distinct product(s) with high efficiency.

In conclusion, despite minimal sequence identity and varying lengths, the N-terminal polypeptide segment of plant sesquiterpene synthases is essential for their high catalytic activity and the product specificity. Indeed, NTSs contain specific residues such as Ser30 and/or segments like M[21]RPKADF[27]QPS[30] in DCS implicated in active site closure and desolvation events upon formation of the Michaelis complex. Despite its distance from the active site, the NTS seems to help shape the active site contour that dictates substrate folding and ultimately product specificity of plant sesquiterpene synthases. As a consequence of its late involvement in catalysis, changes to the NTS by single mutation, truncation and domain swap only compromise the turnover number k_{cat}, while the Michaelis constants K_M remain largely unchanged (Table 1). Furthermore, our work reveals the indirect contribution of NTS in defining the function of dual sesquiterpene / sesquiterpenol synthases. Comparisons between CH-DCS, M20 and M30 reveal that the sesquiterpene / sesquiterpenol ratio is modulated by precisely balancing the k_{cat} values with the degree of solvation / desolvation. Key modulators are the conserved Ser30 in DCS at the C-terminal end of the consensus motif XRPXXXFXS and the length of the NTS. Although progressive shortening of the NTS leads to enzymes with a lower hydrocarbon / alcohol ratio (M8, M20 and M30), the increase in the alcohol fraction is paralleled by a severe reduction of the catalytic activity. NTSs can be moved between different plant sesquiterpene synthases as shown here by importing the NTS from EBFS into DCS in CH-DCS. The increasing amount of (-)-germacradien-4-ol (**5**) produced by progressive N-terminal truncation of DCS strongly suggests that the NTS contributes to the shielding of the active site from water. The absence of hedycaryol (**4**) and cadinol (**6**) (Scheme 1) shows that carbocations **A** and **C** are not quenched by water, either due to the formation of bridged carbocation (**D**) or because they are protected as diphosphate ion pairs. The sustainable production of high value terpenes and terpene alcohols using metabolically engineered microbial hosts[23] with catalytically optimized and altered versions of terpene synthases[24] profits from a more detailed understanding of the molecular mechanisms that control product specificities and reaction rates in terpene synthase chemistry.

We thank Dr Xiao-Ya Chen (Institute of Plant Physiology and Ecology, and Institute for Biological Sciences, Shanghai, China) for plasmid pXC1 harboring the cDNA of (+)-δ-cadinene, Dr. Linda Field (Rothamsted Research, Harpenden, UK) for a cDNA coding for (E)-β-farnesene synthase, Dr. Y. Ohnishi (University of Tokyo) for the germacren-4-ol synthase expression vector (pET16b-SC1) and Dr David Miller (Cardiff University) for helpful comments on the manuscript. This work was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC) through grant BB/H01683X/1, the Engineering and Physical Sciences Research Council (EPSRC) through grants EP/D069580/1 and EP/K301635/1 and Cardiff University.

Notes and references

1 D. E. Cane, *Acc. Chem. Res.*, 1985, **18**, 220.

2 L. C. Tarshis, M. Yan, C. D. Poulter and J. C. Sacchettini, *Biochemistry*, 1994, **33**, 10871.

3 D. W. Christianson, *Chem. Rev.*, 2006, **106**, 3412.

4 J. A. Aaron, and D. W. Christianson *Pure Appl. Chem.*, 2011, **82**, 1585.

5 E. Y. Shishova, F. Yu, D. J. Miller, J. A. Faraldos, Y. Zhao, R. M. Coates, R. K. Allemann, D. E. Cane and D. W. Christianson, *J. Biol. Chem.*, 2008, **283**, 15431.

6 J. A. Faraldos, S. Wu, J. Chappell and R. M. Coates, *J. Am. Chem. Soc.*, 2010, **132**, 2998.

7 a) P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll, and J. S. Dickschat, Angew. Chem. Int. Ed. 2014, **53**, 7652. b) M. Chen, W. K. W. Chou, N. Al-Lami, J. A. Faraldos, R. K. Allemann, D. E. Cane, and D. W. Christianson, *Biochemistry*, 2016, ASAP. c) M. Chen, N. Al-Lami, M. Janvier, E. L. D'Antonio, J. A. Faraldos, D. E. Cane, R. K. Allemann, and D. W. Christianson, *Biochemistry*, 2013, **52**, 5441.

8 M. W. van der Kamp, J. Sirirak, J Zurek, R. K. Allemann and A. J. Mulholland, *Biochemistry*, 2013, **52**, 8094.

9 C. M. Starks, K. Back, J. Chappell and J. P. Noel, *Science*, 1997, **277**, 1815.

10 D. A. Whittington, M. L. Wise, M. Urbansky, R. M. Coates, R. B. Croteau and D. W. Christianson, *Proc. Natl. Acd. Sci. U.S.A.*, 2002, **99**, 15375.

11 D. C. Hyatt, B. Youn, Y Zhao, B Santhamma, R. M. Coates, R. B. Croteau and C. Kan, *Proc. Natl. Acad. Sci. U.S.A.,* 2007, **104**, 5360.

12 K. J. Abraham, M. L. Pierce and M. Essenberg, *Phytochem.*, 1999, **52**, 829.

13 H. A. Gennadios, V. Gonzalez, L. Di Costanzo, A. Li, F. Yu, D. J. Miller, R. K. Allemann and D. W. Christianson, *Biochemistry*, 2009, **48**, 6175.

14 J. P. Noel, N. Dellas, J. A. Faraldos, M. Zhao, B. A. Hess, L. Smentek, R. M. Coates and P. E. O'Maille, *ACS Chem. Biol.* 2010, **5**, 377.

15 M. Y. Kim, Y. J. Chang, M. H. Bang, N. I. Baek, J. Jin, C. H. Lee and S. U. Kim, *J. Plant Biol*., 2005, **48**, 178.

16 C. Nakano, F. Kudo, T. Eguchi and Y. Ohnishi, *ChemBioChem*, 2011, **12**, 2271.

17 D. J. Grundy, M. Chen, V. Gonzalez, S. Leoni, D. J. Miller, D. W. Christianson and R. K. Allemann, *Biochemistry*, 2016, **55**, 2112.

18 P. Baer, P. Rabe, C. A. Citron, C. C. O. Mann, N. Kaufmann, M. Groll and J. S. Dickschat, *ChemBioChem*, 2014, **15**, 213.

19 V. Gonzalez, S. Touchet, D. J. Grundy, J. A. Faraldos and R. K. Allemann, *J. Am. Chem. Soc.,* 2014, **136**, 14505.

20 J. A. Faraldos, D. J. Miller, V. González, Z. Yoosuf-Aly, O. Cascón, A. Li and R. K. Allemann, *J. Am. Chem. Soc*., 2012, **134**, 5900.

21 Y. Back and J. Chappell, *Proc. Nat. Acad. Sci. U.S.A*., 1996, **93**, 6841.

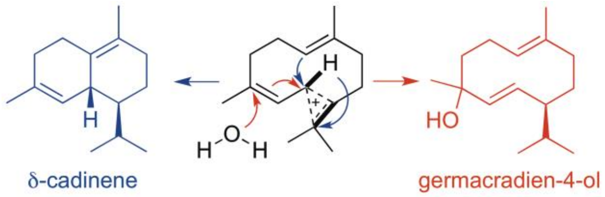
22 J. A. Faraldos, V. Gonzalez, A. Li, F. Yu, M. Koksal, D. W. Christianson and R. K. Allemann, *J. Am. Chem. Soc.,* 2012, **134**, 20844.

23 J. C. Paddon and J. D. Keasling, *Nature Rev. Microbiol*., 2014, **12**, 355.

24 J. X. Li, X. Fang, Q. Zhao, J. X. Ruan, C. Q. Yang, L. J. Wang, D. J. Miller, J. A. Faraldos, R. K. Allemann, X. Y. Chen and P. Zhang, *Biochem. J.,* 2013, **451**, 417.

Journal Name

COMMUNICATION



The β-domain of δ-cadinene synthase (DCS) directs desolvation of the active site.