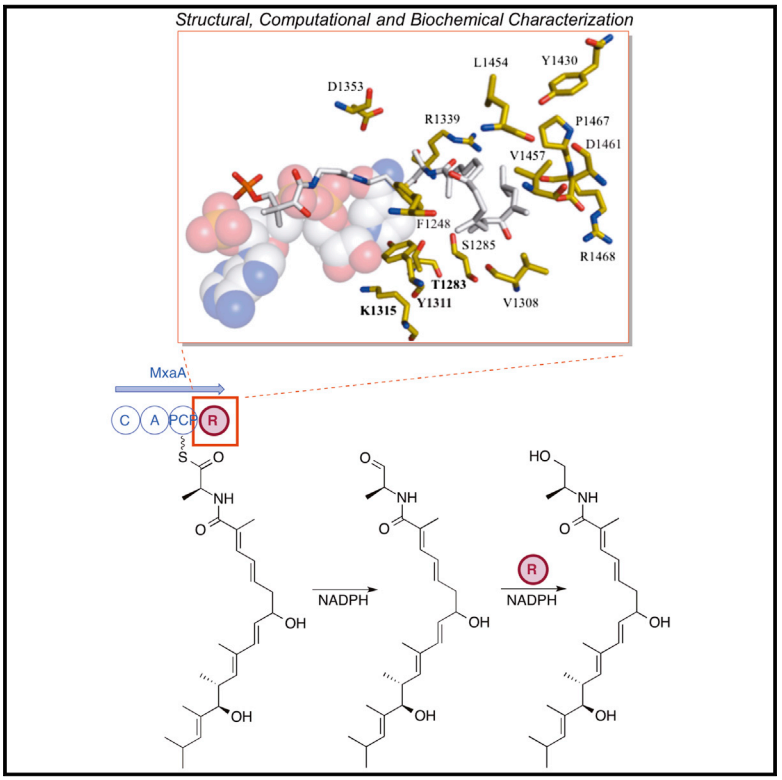
Article

Chemistry & Biology

Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols

Graphical Abstract



Authors

Jesus F. Barajas, Ryan M. Phelan, Andrew J. Schaub, ..., Ray Luo, Jay D. Keasling, Shiou-Chuan Tsai

Correspondence

keasling@berkeley.edu (J.D.K.), sctsai@uci.edu (S.-C.T.)

In Brief

Barajas et al. report the structure of a unique termination domain employed in the reductive release of NRPS-generated natural products. The crystal structure, combined with computational and biochemical investigations, provide a comprehensive understanding of key factors that govern catalysis in this class of termination domains.

Highlights

- Highest resolution and first cofactor-bound structure of a terminal reductase domain
- Computational modeling advances hypotheses made from the crystal structure
- Biochemical analysis defines residues critical for substrate specificity and catalysis
- Result-based engineering enabled improved reduction of highly reduced substrates

Accession Numbers

4U7W
4W4T

Barajas et al., 2015, Chemistry & Biology *22*, 1018–1029
August 20, 2015
http://dx.doi.org/10.1016/j.chembiol.2015.06.022

CellPress

CellPress

Chemistry & Biology
Article

Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols

Jesus F. Barajas,[1,6] Ryan M. Phelan,[2,3,6] Andrew J. Schaub,[1] Jaclyn T. Kliewer,[1] Peter J. Kelly,[1] David R. Jackson,[1] Ray Luo,[1] Jay D. Keasling,[2,3,4,5,*] and Shiou-Chuan Tsai[1,*]

[1]Department of Molecular Biology and Biochemistry, Chemistry, and Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697, USA

[2]Joint Bioenergy Institute, 5885 Hollis Street, Emeryville, CA 94608, USA

[3]QB3 Institute, University of California, Berkeley, Berkeley, CA 94270, USA

[4]Department of Chemical and Biomolecular Engineering and Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA

[5]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[6]Co-first author

*Correspondence: keasling@berkeley.edu (J.D.K.), sctsai@uci.edu (S.-C.T.)

http://dx.doi.org/10.1016/j.chembiol.2015.06.022

SUMMARY

The terminal reductase (R) domain from the non-ribosomal peptide synthetase (NRPS) module MxaA in *Stigmatella aurantiaca* Sga15 catalyzes a non-processive four-electron reduction to produce the myxalamide family of secondary metabolites. Despite widespread use in nature, a lack of structural and mechanistic information concerning reductive release from polyketide synthase (PKS) and NRPS assembly lines principally limits our ability to redesign R domains with altered or improved activity. Here we report crystal structures for MxaA R, both in the absence and, for the first time, in the presence of the NADPH cofactor. Molecular dynamics simulations were employed to provide a deeper understanding of this domain and further identify residues critical for structural integrity, substrate binding, and catalysis. Aggregate computational and structural findings provided a basis for mechanistic investigations and, in the process, delivered a rationally altered variant with improved activity toward highly reduced substrates.

INTRODUCTION

Polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) are large, multi-modular protein assemblies capable of generating chemically diverse and complex molecules. To produce these compounds, PKS and NRPS programmed assembly occurs through the use of either acyl-coenzyme A or amino acid building blocks, respectively, to provide natural products that have applications in numerous sectors in the world economy (e.g., pharmaceutical and agrochemical) (Hertweck, 2009). The myxobacterium *Stigmatella aurantiaca* Sga15 contains a modular PKS/NRPS hybrid responsible for the biosynthesis of the myxalamids (Figure 1), potent inhibitors of the respiratory electron transport chain (Gerth et al., 1983; Silakowski et al., 2001). Interestingly, this hybrid system uses a rare termination mechanism to release the final compound as a primary alcohol (Gaitatzis et al., 2001; Silakowski et al., 2001) as opposed to more common chain release mechanisms that produce, for instance, macrolactones or macrolactams (Du and Lou, 2010).

The biosynthesis of myxalamid is a multi-step process initiated by a type I modular PKS consisting of six modules to biosynthesize a polyene intermediate **3** (Figure 1A) that is translocated to the NRPS module for final processing (Silakowski et al., 2001). In the terminal NRPS module, MxaA, the adenylation (A) domain activates alanine as a building block while the condensation (C) domain catalyzes peptide bond formation between alanine and the PKS-generated intermediate **3** to yield **4** (Konz and Marahiel, 1999). The last step in biosynthesis requires the reductive release of myxalamid (mxa) from the phosphopantetheine (pPant) prosthetic group covalently attached to the peptidyl carrier protein (PCP). This action is catalyzed by a recently described class of NADPH-dependent terminal reductase (R) domains that execute chain termination by a $4e^-$ non-processive reduction to generate primary alcohols (Figure 1) (Chhabra et al., 2012; Du and Lou, 2010; Silakowski et al., 2001). To accomplish this, the PCP-bound thioester is first reduced to the aldehyde intermediate **5**, which, following reduction by a second NADPH equivalent, affords the final 2-aminopropanol-containing product **6** (Figure 1A).

Given the ability to rationally program PKS and NRPS biosynthesis (Hahn and Stachelhaus, 2006; Menzella et al., 2005; Poust et al., 2014; Sherman, 2005; Weissman and Leadlay, 2005), current research is focused on the study of R domains in order to expand the range of available characterized termination mechanisms. In light of the immense scope of possible products that could be generated by PKSs, terminating R domains can provide a unique route to numerous alcohol-containing compounds to replace existing petroleum-derived fuels or commodity chemicals (Atsumi et al., 2008; Chu and Majumdar, 2012; Peralta-Yahya et al., 2012). However, to reliably engineer new

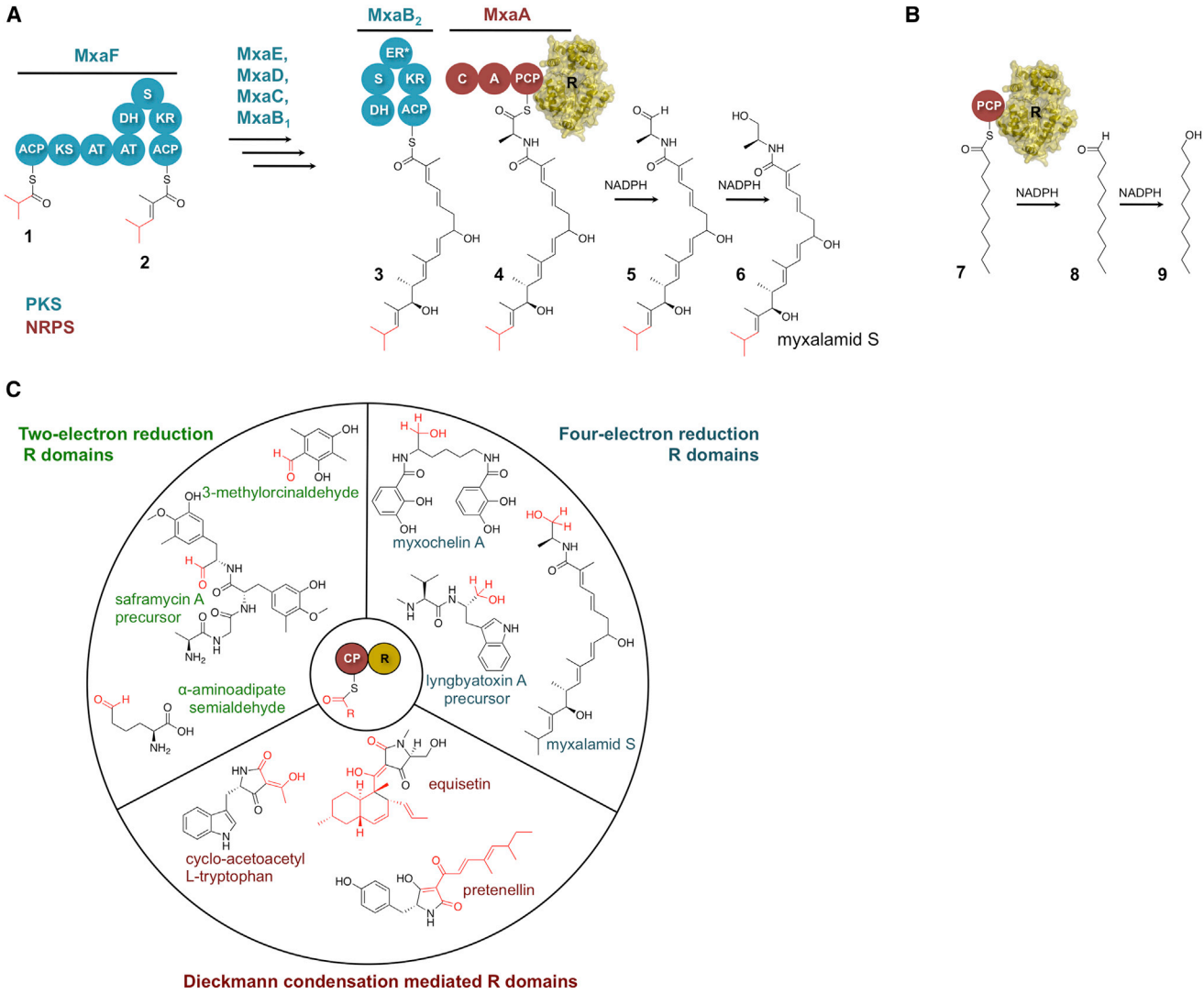


Figure 1. Myxalamid Biosynthetic Pathway and R Domain Substrate Specificity

(A) The myxalamid biosynthetic machinery is composed of six PKS modules (in cyan) and one terminal NRPS module (in red) containing a reductase terminal domain (in yellow). The starter unit of myxalamid is shown in red.

(B) The terminal R domain is capable of reducing non-native C_{10} derivatives.

(C) R domains can be classified into three distinct groups. The two- and four-electron reduction R domains and the Dieckmann condensation R domains.

A, adenylylation domain; ACP, acyl carrier protein; AT, acyltransferase; C, condensation domain; DH, dehydratase; KR, ketoreductase; KS, ketosynthase; PCP, peptidyl carrier protein; R, reductase.

megasynthases to produce biologically derived fuels or commodity chemicals, such as 1-decanol **9** (Figure 1B), a blueprint for the R domain is required; namely a high-resolution structure and biochemical evaluation.

Here we report the 1.90-Å and 1.84-Å structures of the MxaA R domain from *S. aurantiaca* Sga15 in the presence and absence of NADPH, respectively. This, in combination with molecular dynamics and structure-based mutagenesis, provided an unprecedented view of local and global interactions between the PCP and R domain, and those between the R domain and cofactor/substrate that are essential for catalysis. Furthermore, mutational analysis of the R domain enabled us to rationally mutate a key active site arginine that resulted in an MxaA variant with improved activity toward highly reduced substrates (e.g., dodecanoyl-PCP **7**, Figure 1B). The combined structural, computational, and biochemical results presented here provide a comprehensive understanding of these unique termination domains and, in the process, set a strong foundation for future efforts to generate new PKS- or NRPS-based routes to diverse terminal alcohol-containing compounds.

RESULTS AND DISCUSSION

Structure of the MxaA R Domain

To visualize MxaA R, we crystallized the R domain with and without the cofactor, NADPH. Using multiwavelength anomalous

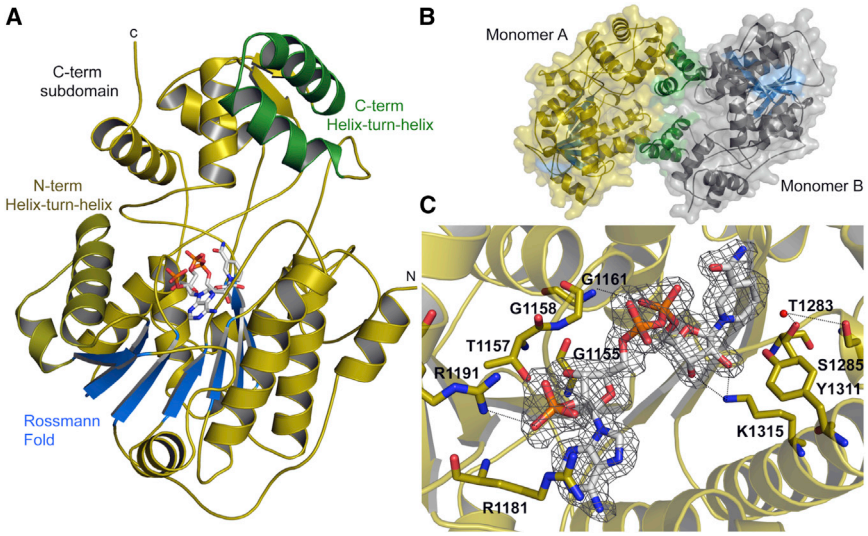


Figure 2. Structure of the MxaA R Domain

(A) The MxaA R domain monomer is composed of an N-terminal subdomain that contains an NADPH Rossmann fold (in blue) and a C-terminal subdomain that contains a helix-turn-helix motif (shown in green). The NADPH cofactor is displayed as gray sticks.

(B) The MxaA R domain crystallizes as a dimer; monomer A is shown in yellow and monomer B is shown in gray.

(C) The cofactor NADPH binds to the TGxxGxxG motif close to the T, Y, and K catalytic site. An SA-omit map of the NADPH cofactor contoured at 1.0σ is shown in the gray isomesh map.

See also Figure S3 and Table S1.

diffraction (MAD) with a selenomethionine-substituted protein, the structure of MxaA R was determined to a resolution of 1.95 Å. The apo MxaA R structure was further refined to 1.84 Å (Table S1). MxaA R was solved as a dimer with a root-mean-square deviation (RMSD) of 0.45 Å between monomers A and B. The overall structure contains strong architectural similarities to type E short-chain dehydrogenases/reductases (SDRs) that contain an N-terminal NADPH-binding region and a C-terminal substrate-binding subdomain (Figure 2) (Jornvall et al., 1995). Structural alignment with the type E SDR from *Agrobacterium tumefaciens* (PDB: 4ID9) displays an RMSD of 3.88 Å through 119 residues of the alpha carbon backbone. Hidden Markov models show that MxaA R has structural and sequence similarities with the extended type E SDRs based on the Kallberg et al. classification (Kallberg et al., 2010). The N-terminal subdomain contains an extended NADPH-binding α/β Rossmann fold with seven parallel beta sheets (β3-β2-β1-β4-β5-β6-β10) flanked by five alpha helices (α2-α3-α4-α6-α8-α11) (Figure 2A; Figures S2 and S3). These structural features correlate well with the previously solved Nrp R domain structure (PDB: 4DQV) with an RMSD of 2.19 Å through 279 residues of the alpha carbon backbone (Chhabra et al., 2012). Similar features include a canonical tyrosine-dependent catalytic triad (T1283, K1315, and Y1311) and a distinctive helix-turn-helix (HTH) motif (α16-α17) found in all structurally known terminating reductase domains.

Substrate recognition in the SDR family occurs in the C-terminal subdomain (Kavanagh et al., 2008). Consequently, while the N-terminal subdomains in SDRs are highly conserved, C-terminal domains often differ in sequence. The C-terminal subdomain of MxaA R consists of five helices (α12-α15-α16-α17-α20) and two parallel beta sheets (β9-β11), which are substantially larger (~130 residues) than those found in typical SDRs (Figures S3A–S3C) (Jornvall et al., 1995; Kallberg et al., 2010; Kavanagh et al., 2008). A notable inserted HTH motif (α16-α17) between residues Y1431 and Q1456 contains several conserved hydrophobic residues (W1433, L1437, L1450, L1451) frequently present in R domains that conduct PKS or NRPS chain termination with 2- or 4-electron reductions (Figure 1C) (Bergmann et al., 2007; Gaitatzis et al., 2001; Gomez-Escribano et al., 2012; Li et al., 2008; Masschelein et al., 2015; Silakowski et al., 2001).

To further distinguish true biological interfaces from lattice contacts in the crystal structure, we further analyzed the MxaA R domain utilizing the Evolutionary Protein-Protein Interface Classifier (EPPIC) server (Duarte et al., 2012), which relies on evolutionary data to detect biological interfaces and PDBePISA (Krissinel and Henrick, 2007). The EPPIC server was unable to reliably determine biologically relevant surface interfaces due to the lack of homolog sequences for comparison. PDBePISA generated a Complex Formation Significance Score of 0.00, suggesting that the surface interface displayed by the MxaA homodimer is a result of crystal packing. The average interface area between both monomers was calculated to be 656.9 $Å^2$, which is 3.85% of the total solvent accessible area. This constituted a total of 22 and 20 buried surface residues for monomers A and B, respectively. It is also well known that biological interfaces tend to exhibit large areas, with the majority of cases exceeding 1,000 $Å^2$ (Jones and Thornton, 1996). Furthermore, evidence for its biological monomeric state was gathered from analytical size exclusion chromatography experiments comparing the MxaA PCP-R didomain to known protein standards. Overall, these results suggest that MxaA R exists in a biologically monomeric form rather than the crystallographically observed homodimeric state.

Structure Analysis of NADPH-Bound R Domain

Currently the Nrp terminal R domain (PDB: 4DQV) from *Mycobacteria smegmatis* involved in glycopeptide biosynthesis and the AusA R domain (PDB: 4F6C, 4F6L) from *Staphyloccoccus aureus* involved in pyrazinone biosynthesis are the sole PKS- or NRPS-associated R domains to have a structure reported. While these monodomain structures have been solved with moderate resolution (2.30 Å for NRP and 2.81 Å for AusA), the lack of bound NADPH leaves key structural and mechanistic details rather unclear (Chhabra et al., 2012; Wyatt et al., 2012). In order to define residues required for cofactor binding in MxaA R, co-crystals of MxaA R complexed with NADPH were solved by molecular replacement of the apo structure to 1.90 Å (Table S1). NADPH binds to the well-known Rossmann fold, which has a conserved nucleotide-binding motif TGxxGxxG, with the

central diphosphate moiety hydrogen bonding to the peptide backbone of G1155, T1157, G1158, L1160, and G1161 (Figure 2C; Figure S3D). Furthermore, the G1155 carbonyl forms a hydrogen bond with the adenosine 3′-hydroxyl group while the adenosine 2′-phosphate oxygen interacts with highly conserved T1157, R1181, and R1191. Both the 2′- and 3′-hydroxyl groups of the nicotinamide-containing ribose ring hydrogen bond with K1315 and Y1311. The nicotinamide amine hydrogen bonds with the G1338 carbonyl. Together, these interactions serve to tightly bind NADPH (K_d = 45 ± 3.7 μM) and properly orient it in the active site for reduction of the pPant-bound intermediate to the terminal alcohol.

Several coordinated water molecules are present between the catalytic residues Y1311, T1283 and the non-catalytic S1285. One water molecule is positioned 2.7 Å from the hydroxyl of Y1311 and 2.8 Å from T1283, possibly occupying the oxyanion hole that these two residues create to assist in thioester and aldehyde reduction. T1283 and S1285 bind a second water molecule in the active site, although its positioning does not provide a clear role in catalysis. With respect to these observations, several SDR studies suggest that ordered water molecules in the active site might participate in a proton relay system involving the hydroxyl of Y1311, 2′-hydroxyl of the nicotinamide ribose and K1315 (Eklund et al., 1982; Oppermann et al., 2003). Structural comparison of the apo and NADPH-bound MxaA R domain show slight conformational changes with an overall RMSD of 0.63 Å through the entire backbone. From these small differences, the C-terminal subdomain experiences a slightly higher conformational change upon NADPH binding compared with the complete monomer, with an RMSD difference of 0.73 Å.

Molecular Dynamics Analysis

To elucidate the structural dynamics of NADPH and substrate binding in MxaA R, we conducted molecular dynamics (MD) simulations by analyzing conformational changes in 100-ns MD runs. Atomic coordinates of the MxaA R domain were obtained from the NADPH-bound MxaA R domain (chain B) crystal structure. The ff14SB forcefield in Amber14 was used for the protein and the general AMBER force field was used for the NADPH cofactor (Case et al., 2014a, 2014b; Gotz et al., 2012; Hornak et al., 2006; Wang et al., 2004, 2006; Wickstrom et al., 2009). NADPH was parameterized using Gaussian 09 to obtain the initial electrostatic potential using the HF/6-31G(d,p) basis set, followed by the use of antechamber to obtain the HF/6-31G(d,p) restricted electrostatic potential fit with final overall net charge of −4. The system was explicitly solvated with a buffer of 10 Å TIP3P waters in a truncated octahedron box after neutralizing with counter ions. A two-system minimization was performed using SANDER and PMEMD was used for production runs (Verdonk et al., 2003).

The NADPH-bound MxaA R domain was allowed to equilibrate after heating the system to 300 K and subsequently allowed to run over 100 ns. 2D RMSD maps were generated using Chimera and an in-house MATLAB script by comparing RMSD fluctuations of the protein backbone. The maps revealed an RMSD range of 0.61–2.41 Å for the MxaA R domain (Figure 3A). Further dissection of the N- versus C-terminal subdomains revealed RMSD ranges of 0.54–1.49 Å and 0.52–2.25 Å, respectively (Figures 3B and 3C). These results, combined with RMSD values found in our crystal structures, indicate higher flexibility and movement of the C-terminal subdomain.

The most noticeable region of flexibility was observed in the C-terminal HTH motif, specifically the conserved hydrophobic residues between Y1430 and Q1455 of α16-α17, which display an average RMSD of 0.82 Å in the NADPH-bound model (Figure 2C; Figures S4G–S4I). Numerous salt bridges are critical in stabilizing the α16-α17 HTH motif, such as R1426 and E1436 (Figures S4A–S4C). During the 100-ns NADPH-bound run, the ζC of R1426 maintains a distance of ≤6.0 Å with either εO of E1436. D1444, the turn residue between helix 16 and helix 17, also maintains a tight salt bridge interaction with a distance ≤3.5 Å between the ζC of R1364 through stochastic sampling of either helix 13 D1444 δO during 73.2% of the simulation (Figure S4A). Moderate electrostatic interactions were observed between helix 12 R1357 and helix 17 E1446, with the ζC of R1357 maintaining a distance ≤ 6.0 Å for 68.7% of the simulation. The catalytic triad (T1283, Y1311, and K1315) exhibits little movement with an average of 0.07 Å per residue throughout the entire 100-ns run. The phosphate attached to the nicotinamide ribose 5′ carbon remains stable within 1.99 and 2.09 Å of the Rossmann TGxxGxxG motif.

A representative cluster ensemble was generated from MD using RMSD scoring as implemented in Chimera (Pettersen et al., 2004). RMSD scoring reduced the initial set of 1,000 frames generated to the 46 most unique frames. In silico docking of the pPant-bound substrate using all of the 46 unique frames from the previous MD run by the program GOLD revealed a large binding cavity under the α16-α17 HTH motif (Figure 2A) (Verdonk et al., 2003). In order to identify substrate-binding residues for MD analysis, we docked the myxalamid substrate (see the next section on docking analysis) using the 46 unique clusters from the NADPH-bound MxaA R MD analysis. We ranked the docking solutions using the ChemPLP scoring function and identified the most consistent binding orientation of the myxalamid substrate by tallying residues involved in substrate binding (Figure 4B). The top ChemPLP docking solution that was prescreened by binding orientation was used for MD analysis of MxaA R with the myxalamid substrate. Using the same MD system parameters as before, we allowed the pPant-bound substrate and NADPH-bound R domain chain B simulation to run for 100 ns. RMSD 2D map analysis of the MxaA R domain in complex with the pPant substrate revealed an RMSD range of 0.60–1.99 Å (Figure 3D). This value is lower than the RMSD range for the R domain with NADPH bound but lacking substrate (Figure 3A) and indicates a decrease in protein motion upon substrate binding. The NADPH-binding N-terminal subdomain demonstrates a similar RMSD range of 0.55–1.49 Å, whether substrate is bound or not, while the C terminus reduces its flexibility upon substrate binding (Figures 3C and 3F).

In light of results that indicated that substrate binding stabilized the C terminus, we opted to focus additional attention on the C-terminal HTH motif, specifically those residues that are steadied through interactions with the substrate (Figures 4A–4C). The terminal HTH motif displays a slightly lower average RMSD of 0.76 Å, while the pPant moiety exhibits larger movements than the sequestered myxalamid segment. D1353 shows strong hydrogen bonding with the amide moiety of the pPant group, averaging a 3.0-Å distance for more than 70 ns of the MD run (Figure S4E). The amide carbonyl group of the terminal

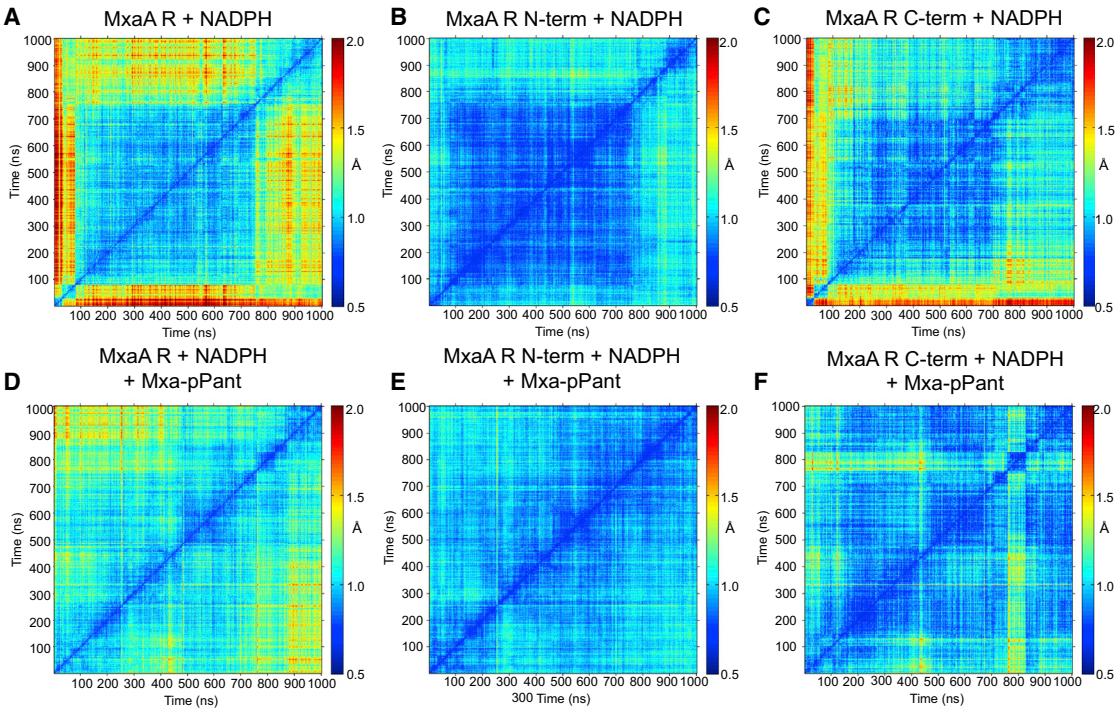


Figure 3. Molecular Dynamic Analysis

(A) 2D RMSD map analysis of the MxaA R domain backbone with NADPH over the entire 100-ns molecular dynamics simulation. Low RMSD is observed in blue and high RMSD is observed in red.

(B and C) Dissecting the N- versus C-terminal subdomain of MxaA R bound to NADPH reveals higher RMSD deviations in the C-terminal subdomain.

(D) 2D RMSD map was generated with the MxaA R domain bound to NADPH and docked with mxa-pPant.

(E and F) Dissection of the N- versus C-terminal subdomain of the bound NADPH mxa-pPant R domain demonstrates a decrease in movement of the C-terminal subdomain.

See also Figure S4.

alanine in the mxa intermediate generates a tight 2.0-Å interaction with R1339, highlighting the likely importance of electrostatic interactions between myxalamid and the R domain (Figure S4D). The methyl-branched diene moiety of the mxa substrate forms intramolecular hydrophobic interactions, kinking the aliphatic substrate back toward the pPant thioester to minimize its hydrophobic surface area. F1248 in MxaA R aids in stabilizing these hydrophobic interactions. The C_9 and C_{15} hydroxyl groups in myxalamid intermediate **4** hydrogen bond with S1285 and D1461, respectively (Figures 4A and 4C). S1285 forms an average 3.0-Å hydrogen bond with the terminal hydroxyl group in the myxalamid intermediate for more than 95% of the MD simulation. Similarly, D1461 forms an average 3.2-Å hydrogen bond with the first hydroxyl group for more than 80% of the MD run. Both S1285 and D1461 along with R1339 and F1248 appear to play a key role in myxalamid substrate recognition and orienting the substrate for reduction by NADPH. In summary, MD simulation results suggest that the C-terminal domain is highly mobile until myxalamid substrate binding quenches movement, more precisely at the HTH motif, for the first round of reduction.

Docking Analysis of pPant-Mxa Substrate and PCP Domain

Initial structural analysis in parallel with the NADPH-bound MxaA R MD analysis revealed a large substrate cavity with various potential substrate-binding residues. Using in silico docking, we further probed for residues important in substrate binding by docking the pPant-tethered mxa intermediate in the R domain active site. The 100-ns MD simulation of NADPH-bound MxaA R identified 46 unique clusters, indicative of 46 distinct MxaA R domain conformations. One frame from each cluster was obtained and was used as the receptor to dock against the pPant-mxa ligand using the program GOLD (Liebeschuetz et al., 2012). Each frame generated 100 solutions that were scored and ranked using the ChemPLP scoring function (Hildebrand et al., 2009). The program LIGPLOT, in parallel with visual inspection, was used to analyze and identify ligand-protein residue interactions (Wallace et al., 1995). The residue-ligand interactions between 2.5 and 4.0 Å for each frame were tallied and a heatmap was generated, indicative of ligand-residue proximity in different R domain conformations (Figure 4B). Not surprisingly, the catalytic triad was revealed to frequently associate with the pPant-bound substrate. T1283 showed interactions close to the thioester linkage in 28 out of the 46 frames. In addition, Y1311 interacted with the thioester in 18 of the 46 frames. The water-coordinating S1285 associated with the thioester in 35 of the 46 frames. The majority of residues that interacted with the mxa portion of the ligand were localized on the C-terminal subdomain. Of the 46 frames, 37 showed that R1339 engages in electrostatic interactions with one of the two carbonyl groups in the substrate near the thioester linkage.

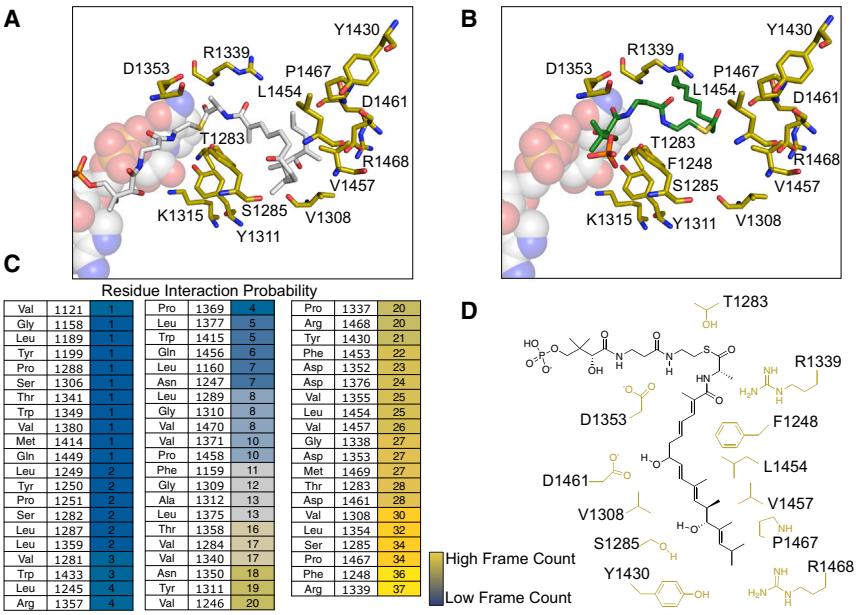


Figure 4. Myxalamid Docking Analysis

(A) Stereo image of the MxaA R domain bound to NAPDH and docked with myxalamid-pPant. Residues outlining the mxa-pPant substrate (gray sticks) are displayed as yellow sticks and the NADPH cofactor is represented as spheres.

(B) Residue interaction probability of all 46 distinct frames from MD docked with mxa-pPant.

(C) Figure represents mxa-pPant and close interacting MxaA R domain residues.

Residues that outline the back of the substrate-binding pocket revealed several hydrogen bonding and hydrophobic interactions with the substrate, specifically, V1308, Y1430, and R1468. Taken together, residues with high substrate contact probabilities, derived from resultant heatmaps, indicate the likely importance of specific residues in substrate recognition and orientation.

To situate the R domain in the perspective of the termination module MxaA, more specifically the protein interactions between PCP and R domain for the reductive release of the final product, we computationally docked the R domain with the PCP domain. The previously solved SrfA C-terminal module structure from *Bacillus subtilis* revealed that the PCP domain is positioned in close proximity to the catalytic C, A, and TE domains, thus providing evidence for the spatial relationships of the PCP in the termination module (Tanovic et al., 2008). Accordingly, we used the SrfA C-terminal domain structure (PBD ID: 2VSQ) as a template on which to base our protein-protein docking studies. Using the HHPRED server, we generated a tertiary homology model of the MxaA PCP domain and proceeded to dock the R domain using the ZDOCK server (Chen et al., 2003; Hildebrand et al., 2009; Pierce et al., 2011). Most of helix III in the PCP forms contacts with surface residues in both N- and C-terminal subdomains of the MxaA R domain (Figures 5A and 5B). Dissecting the PCP surface reveals electrostatic interactions between the surfaces of the R and PCP domains. The conserved serine (part of the signature D/HSL motif) contained in the PCP domain that covalently binds the pPant prosthetic group (S56) is located 12.0 Å away from the catalytic triad: T1283, Y1311, and K1315. The HTH motif of the MxaA R C-terminal subdomain has both electrostatic and pi-pi stacking interactions with helix III of the PCP domain. These include the Q1445 of the R domain with R77 from the PCP domain and the carbonyl backbone of S1442 of the R domain with the D73 side chain of the PCP. In addition, a face-to-face pi stacking interaction occurs between F1453 in the R domain and Y60 of the loop connecting helix II and helix III in the PCP domain (Figure 5B). These results highlight the importance of electrostatic and aromatic residues on the surfaces of both the R and PCP domains for protein-protein interactions and provide a basis for the engineering of chimeric ACP/PCP-R domain fusions.

Biochemical Analysis of the NADPH and Substrate-Binding Pockets

Moving beyond the structural and computational results pertaining to the reduction of **4** to **6**, and set in the context of advanced biofuel production **7** to **9** (Figure 1B), we aimed to provide biochemical support for residues involved in NADPH binding and the requirement of the Lys, Tyr, Thr catalytic triad. As the structural data revealed that NADPH contacts the amide backbone of G1155, G1158, and G1161, we generated single Gly to Ala mutants for each residue to examine the effect on NADPH-binding capacity. Likewise, residues comprising the putative catalytic triad were mutated one at a time (T1283A, K1315A, and Y1311F) to similarly determine effects on both NADPH binding and catalytic activity. NADPH-binding studies, as determined by intrinsic fluorescence measurements (Chhabra et al., 2012; Wilson et al., 2010) with the wild-type (WT) and mutant PCP-R didomains, clearly demonstrated disruption of NADPH binding by alanine mutation in the TGxxGxxG motif, as well as with mutants Y1311F and K1315A (Figure 6; Figure S2). Surprisingly, although not without precedent (Chhabra et al., 2012), the T1283A mutant possessed an approximate 50% increase in NADPH-binding affinity ($K_{d(WT)}$ = 45 ± 3.7 μM versus $K_{d(T1283A)}$ = 31.7 ± 1.6 μM). This could be rationalized by the fact that T1283 has no direct contact with the cofactor and removal of its steric bulk likely improves access of NADPH to the binding site. These mutants were additionally investigated for activity toward the reduction of decanal **8**. All mutants were catalytically inactive, including T1283A, which further verified the requirement of the intact NADPH-binding pocket and the complete catalytic triad for activity (Figures S5C and S6).

Moving beyond characteristic analysis of the NADPH-binding motif and catalytic residues, we sought to further investigate residues that may interact with the mxa substrate as determined through MD simulations and docking studies. We focused on five residues. The first, and perhaps most interesting, was R1339, which was determined by MD simulations to possess the highest probability of contact with the substrate, in particular the thioester-bound alanine moiety. Four additional residues with high probability of substrate interaction in the predominantly hydrophobic mxa-binding pocket were mutated to reverse their

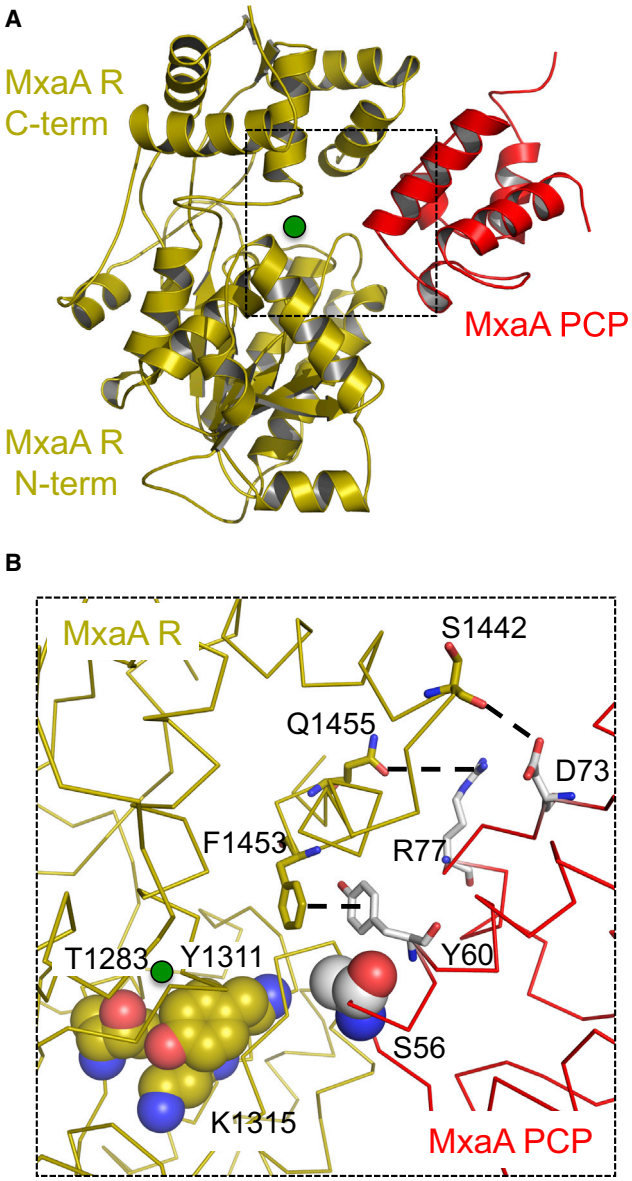


Figure 5. Model of MxaA PCP-R Domain Interactions

(A) Cartoon model of MxaA PCP-R domain interactions. The green dot represents the active site cavity of the MxaA R domain.

(B) Model of the MxaA PCP-R interface reveals possible electrostatic and pi stacking interactions between these two domains.

The square in (A) outlines the boundary of the area observed in (B). Spheres represent the catalytic triad of the R domain (yellow) and the pPant attachment site serine of the PCP (gray).

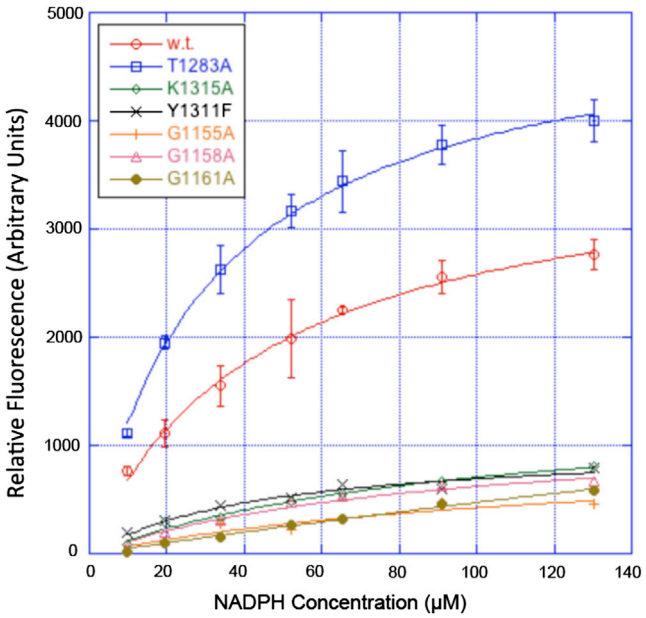


Figure 6. Determination of NADPH-Binding Constants for Wild-Type MxaA R and Select Mutants of NADPH-Interacting Residues

Error bars represent the coefficient of variance between triplicate reads. See also Figures S5 and S6 and Table S2.

polarity or knockout key functional groups (F1248N, V1308T, Y1430F, R1468A). Because of the complexity of the substrates (pPant-mxa **4** and mxa aldehyde **5**) and concerns of their aqueous solubility, we opted to conduct the enzyme assay using simplified substrates: decanoyl-PCP **7** and decanal **8**. Moreover, as our ultimate goal is to use the information gained in these studies for the production of biologically derived replacement fuels and commodity chemicals, examination of the active site in the context of a target compound provides valuable knowledge to enable our desired goal. Studies additionally offer critical information pertaining to the mechanism of the R domain. Each mutant was assayed for the full reductive reaction (**7** to **8**) and the second half reaction (**8** to **9**). Owing to the fact that aldehyde reduction is several orders of magnitude faster than PCP thioester reduction (Table 1), assay of the full reaction provides rates that are specific to the first half reaction. Therefore, we were able to obtain rates for both reductions: *k*1 (**7** to **8**) and *k*2 (**8** to **9**). Due to the fact that both decanoyl-CoA and decanoyl-loaded MxaA PCP monodomain were not turned over by MxaA R, we developed a single turnover assay by loading decanoyl-CoA to the PCP-R didomain with the promiscuous phosphopantetheinyl transferase Sfp (Quadri et al., 1998) in order to obtain kinetic parameters for the first reduction. Given the slow rate of the first-half reaction, multiple time points were taken within the first 3 hr without depleting the enzyme-substrate complex below 5% of the total concentration. This allowed the assays to be kept under pseudo-saturating (k_{cat}) conditions. While data concerning the first reduction provided turnover numbers that are significantly slower than those found with the second-half reaction, a clear dependence on residue identity was observed in the course of our studies. Moreover, as this system is a truncated portion of the complete MxaA module, changes in protein structure or substrate positioning, a parameter that may be altered by both protein truncation and the use of a substrate lacking a PCP-bound amide bond, may have consequences that affect the upper limit of k_{cat} but still clearly represent changes brought on by amino acid substitution. In contrast, observing NADPH consumption under saturating, multiple turnover conditions with the intermediate aldehyde **8**, yielded values similar to those obtained with the Nrp R domain (Table 1).

Table 1. Specific and Relative Activities for Wild-Type and Select Mutants with Respect to the First- and Second-Half Reactions

	Full Reaction (7 to 9)		Second-Half Reaction (8 to 9)	
	Enzyme Activity (pmol/min/mg MxaA)	Activity Relative to Wild-Type	Enzyme Activity (μmol/min/mg MxaA)	Activity Relative to Wild-Type
WT	3.69 ± 0.19	1.00	21.5 ± 1.7	1.00
F1248N	1.24 ± 0.03	0.34	27.2 ± 4.9	1.26
V1380T	1.86 ± 0.10	0.50	34.1 ± 0.7	1.58
R1339A	15.19 ± 0.46	4.11	134.41 ± 12.5	6.22
Y1430F	2.45 ± 1.56	0.66	37.4 ± 4.1	1.73
R1468A	1.10 ± 0.10	0.30	26.8 ± 3.1	1.24

With respect to the first reduction, we found mutations of the four residues that define the mxa-binding pocket to cause significant reductions in activity (Table 1). Mutation of residues closer to the NADPH-binding site (F1248N, approximately 65% reduction in activity) caused a greater reduction in activity than those buried deeper in the pocket Y1430F and V1380T, (approximately 45% reduction in activity). This is likely due to reduced substrate-residue interactions, as indicated by docking simulations with the non-native substrates **7** and **8**. R1468, while buried deep in the binding pocket, still appeared to have an important role in the first-half reaction as demonstrated by the sharp reduction in activity with the R1468A mutant. Interestingly, for the second reduction, the same mutations moderately increased activity compared with the WT. Aggregate results reveal a high probability that the first-half reaction is the rate-limiting step of this overall process and acutely sensitive to binding pocket mutations, while aldehyde reduction appears to be more robust. In our investigation, the second-half reaction turnover rate was actually improved by disruption of the binding pocket and active site entrance, suggesting that, for the second-half reduction, product release might be rate limiting.

In addition to exploring the mutational tolerance of the binding pocket, we were interested to determine if R1339, as indicated by MD and docking simulations, interacted with the substrate. Computational data hinted at an electrostatic interaction between the R1339 guanidino group and the terminal alanine moiety contained within mxa. Therefore, we aimed to determine if R1339 in fact has an impact on catalysis. While kinetic analysis of substrates lacking a terminal alanyl thioester or alanal moiety, as found in mxa substrates **4** and **5**, cannot definitively demonstrate the role that R1339 plays during catalysis with the native substrate, comparison of the C10 substrate used in our studies with both WT and R1339A MxaA R provides a general understanding as to the nature of the residue-substrate interaction. Of significant importance, particularly in light of our goals to apply this enzyme in the production of fully reduced alcohols, we found that R1339A dramatically improved the ability of MxaA R to reduce C10 substrates with a 4.1- and 6.2-fold increase in activity for the first and second reduction, respectively. The large increases in activity can be rationalized by the fact that reduction of the thioester or aldehyde is guided by interactions between the PCP, R domain, and pPant arm and R1339 appears to be poised to interact with incoming substrates (Figure 4). Both the first and second reductions with alternate substrates are improved by removal of the mismatched residue-substrate polarity (i.e., hydrocarbon-guanidino interaction) and, accordingly, are facilitated by an increase in the hydrophobicity of the active site tunnel (Figure S6). These biochemical findings support the combined crystal structure and computational data and set the stage for future endeavors to further tune the active site to increase the turnover of aliphatic substrates.

Conclusions

Products generated by PKSs and NRPSs require release from pPant-tethered carrier proteins contained in megasynthases. Both thioesterase and R domains mediate chain release to provide distinct terminal functional groups to enrich the chemical diversity of polyketide and non-ribosomal peptide natural products (Du and Lou, 2010). R domains are an NADPH-dependent class of SDR-like enzymes capable of reductively releasing acyl and peptide intermediates from the pPant-tethered carrier protein. Prior to this study, no cofactor bound structure was available for modular enzyme-associated terminal R domains. Here, we report the crystal structure, with significantly increased resolution, of the myxalamid PKS-NRPS terminal R domain that catalyzes the non-processive four-electron reduction of **4** to **6** and decanoyl-PCP (**7**) to 1-decanol (**9**). Computational MD and biochemical analysis support assertions that the C-terminal subdomain of the R domain is the most flexible region, responsible for substrate binding and selectivity. With respect to kinetic parameters, the first reduction of decanoyl-PCP (**7**) to yield decanal (**8**) is significantly slower than the second reduction of decanal (**8**) to 1-decanol (**9**), thus providing insight into the rate-limiting step during R domain-mediated product release. Structure-based mutations helped to determine residues important for substrate binding and reduction. Furthermore, mutational analysis of the putative gatekeeping residue (R1339) improved reduction of both **7** and **8**. Combined, the mechanistic insights gained by our comprehensive investigation of MxaA R provide not only a deeper understanding of the structural and catalytic features required for activity but set a foundation for future engineering efforts using modular catalyst-associated R domains. Efforts in combining R domains with novel PKS- or NRPS-based assembly lines could produce alternate substrates that, for example, could be screened for new bioactivity or used in the production of biologically derived commodity chemicals.

SIGNIFICANCE

Termination domains found in modular catalysts (i.e. polyketides synthases and non-ribosomal peptide synthetases) are responsible for the release of covalently attached intermediates and, in the process, generate functional group diversity contingent on the mechanism employed. We focus on a member of the 4-electron reducing domain (R) class, MxaA R from myxalamid biosynthesis, and report the highest resolution structure to date of the apo- and, for the first time, cofactor-bound enzyme. Molecular dynamics simulations delivered an improved picture, beyond traditional structural studies, of key protein-protein and protein-substrate interactions, which, combined with structural data, provided the basis for biochemical investigations. Mutational analysis

focused both on the putative catalytic residues and substrate-binding pocket to define the necessity for the catalytic triad and reveal select residues that are highly influential in catalysis. The combined data provide an unparalleled view of this unique termination mechanism that spans from macromolecular movements essential for catalysis to the identification of key substrate-residue interactions. In aggregate, the studies presented here will aid efforts to improve these domains for the production of diverse primary alcohols. This possibility was highlighted by the enhancement of activity toward fully saturated compounds, specifically C_{10} derivatives, through mutation guided by our structural and biochemical results.

EXPERIMENTAL PROCEDURES

Molecular Dynamics

MD was carried out using AMBER 14 (Case et al., 2014a, 2014b). Both protein and ligand were prepared for docking using the program Chimera (Pettersen et al., 2004). Charges were calculated using the AMBER ff14SB force field. Selenomethionine residues were converted to methionine residues, solvent was deleted, and hydrogens were added. LEaP was used to neutralize the system by adding eight Na^+ ions and solvating the apoenzyme in a 10-Å water buffer TIP3P truncated octahedron box. The fully solvated system contained 42,865 atoms. Minimization using SANDER was performed in two stages to remove any steric clashes present in the initial crystal structure. The initial stage was carried out over 2,500 steps for the solvent and ions with the protein and cofactor restrained by a force constant of 500 kcal/mol/Å^2, followed by a second stage carried out over 5,000 steps of the entire system. A short 20-ps simulation with weak restraints (force constant of 10 kcal/mol/Å^2 on the protein and cofactor) was used to heat up the system to a temperature of 300 K using a Langevin temperature equilibration scheme. Periodic boundary conditions were used, along with a non-bonded interaction cutoff of 10 Å. For the simulation, hydrogen atoms were constrained using the SHAKE algorithm, allowing for a 2-fs time step. The simulation was run over 100 ns (50,000,000 time steps). Simulation speeds of 4.0 ns/day were observed. A representative cluster ensemble was generated from MD using RMSD scoring as implemented in Chimera 1.9 (Pettersen et al., 2004). RMSD scoring reduced the initial set of 1,000 frames generated to the 46 most unique frames. Molecular graphics and analysis were performed with the UCSF Chimera package. RMSD scoring was also used to calculate changes in the C-terminal and N-terminal domains. Highly mobile residues were identified in a similar approach.

Protein Expression and Purification

The recombinant WT and mutant MxaA R monodomains with an N-terminal His_{6x} tag were expressed in BL21 (DE3) *E. coli* cells (Novagen). Cells containing the MxaA R domain plasmid were grown to OD_{600} = 0.6 at 37°C in LB medium containing 50 μg/ml kanamycin. The cell cultures were cooled to 18°C and expression was induced using 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The cell cultures were incubated for an additional 16 hr at 18°C and harvested by centrifugation at 5,525 relative centrifugal force (RCF) for 15 min. The cell pellets were resuspended in 50 mM Tris-HCl (pH 7.5), 10% glycerol, 10 mM imidazole, 300 mM NaCl, and 1 mg/ml lysozyme. Resuspended cells were cooled on ice for 30 min and the cells were disrupted using sonication. The cell debris was cleared by centrifugation at 21,036 RCF for 1 hr. The supernatant was collected and batch bound to HisPur Cobalt Resin (Thermo Scientific) for 1 hr at 4°C. MxaA R was purified according to the manufacturer's instructions using an imidazole step gradient. Fractions containing pure protein were determined by SDS-PAGE and fractions containing MxaA R were combined and dialyzed against 50 mM Tris-HCl (pH 7.5), 10% glycerol, 300 mM NaCl at 4°C for 12 hr. Removal of the N-terminal His_6 tag was conducted by incubating the dialyzed MxaA R at 18°C for 24 hr with thrombin from bovine plasma (Sigma-Aldrich) at a concentration of 2 U/mg of MxaA R protein and 3.5 mM $CaCl_2$. Removal of thrombin and further purification of MxaA R was conducted by anion exchange chromatography using HiTrap Q FF (GE Healthcare) according to the manufacturer's instructions. Purified MxaA R was dialyzed against crystallization buffer, which consisted of 25 mM Tris-HCl (pH 7.5), 5% glycerol, and 1 mM DTT.

Selenomethionine-substituted (SeMet) MxaA R protein was produced in Bl21 (DE3) *E. coli* strain in M9 minimal medium using metabolic inhibition of the methionine biosynthetic pathway (Van Duyne et al., 1993). Five milliliters of an LB culture grown overnight was used to inoculate 2 × 1 l of LB, which was allowed to grow at 37°C in the presence of 50 μg/mL kanamycin until OD_{600} = 0.6 was reached. The resulting cells were pelleted at 5,525 RCF for 15 min and washed three times by suspension in 40 ml of M9 medium and then transferred to 2 × 1 l of M9 medium containing 50 μg/ml kanamycin and the following amino acids: lysine, phenylalanine, and threonine (100 mg/L); isoleucine, leucine, and valine (50 mg/L); and L-selenomethionine (40 mg/L) (Sigma). The temperature was reduced to 18°C and the mixture was induced with 0.5 mM IPTG and allowed to grow overnight for 16 hr. The cells were harvested and purified following the WT procedure. The incorporation of selenomethionine (10 residues in total) was confirmed by MALDI-TOF mass spectrometry.

Crystallization, Data Processing, Refinement and Analysis

Both native and SeMet crystals of the WT MxaA R domain (9 mg/ml) grew in 0.22 M ammonium acetate, 28% PEG 3350, and 0.1 M HEPES (pH 7.7) overnight at 25°C using the hanging drop vapor diffusion method. NADPH-bound MxaA R crystals formed similarly with the exception of incubating NADPH and MxaA R at a 5:1 M ratio for 1 hr at 4°C prior to crystal tray setup. Crystals were cryoprotected in well solution and flash frozen in liquid nitrogen prior to data collection. Data was collected at beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) for SeMet crystals. Prior to data collection, initial frames were assessed for quality and redundancy using Mosflm and Web-ice (González et al., 2008; Leslie and Powell, 2007). MAD data was collected to 1.70 Å for SeMet MxaA R at λ = 0.9792 Å (selenium peak), λ = 0.9611 Å (inflection), λ = 0.9794 Å (remote). For MAD data collection, the exposure time was set to 0.2 s; 0.15° oscillation width for 1,920 frames. All data were processed using Mosflm to the P21 space group (Battye et al., 2011). Native NADPH-bound MxaA R data were collected at the Advance Light Source beamline 822 at the Lawrence Berkeley National Laboratory. Single monochromatic X-ray diffraction data (λ = 0.9775 Å, 700 frames at 0.5° oscillation width for 1-s exposure) were collected to 1.84 Å and processed with Mosflm using the P21 space group. Resolution cutoff was based on a combination of data completeness, R values, and CC values. Initial phases for the MAD dataset were obtained using PHENIX Autosol and 9 of the 10 heavy-atom derivatives were located (Terwilliger et al., 2009). The initial model was constructed using PHENIX Autobuild. Refinement was done using PHENIX.REFINE and COOT (Emsley and Debreczeni, 2012; Adams et al., 2010). Improved phases were used in COOT to model missing side residues manually and waters were added during the last refinement cycles. For the NADPH co-crystal structure, PHENIX LigandFit was used to model the NADPH upon obtaining the initial model and phases from PHENIX Autosol (Adams et al., 2010). Both apo and NADPH-bound structures were validated using PROCHECK and PDB_REDO (Joosten et al., 2012; Laskowski et al., 1993). Structural analyses such as structural superimposition, electrostatic potentials, and figure generation used in the manuscript were done using PyMOL (Schrödinger LLC, 2013).

In Silico Docking

The docking program GOLD was used for docking between the MxaA R domain and the phosphopantetheine-tethered myxalamid intermediate (Verdonk et al., 2003). Both protein and ligand were prepared for docking by removing waters, adding hydrogens, and converting the PDB files to Mol2 files using the program Chimera (Huang et al., 1996). The MxaA R ligand-binding pocket was defined as residues within 20 Å of the hydrogen atom on the hydroxyl group of T1283. Docking was performed using the default settings with 100 docking trials performed. The docking solutions were ranked using the ChemPLP scoring functions. MD simulations generated 46 clusters with significant RMSD differences. A frame from each cluster was used to dock the phosphopantetheine-tethered myxalamid using the same docking parameters. Prior to MxaA PCP-R domain docking, a PCP homology model was generated using the structure prediction HHpred (Hildebrand et al., 2009). The R domain monomer was docked with the PCP homology model using

the protein-protein docking server ZDOCK (Pierce et al., 2011). The ZDOCK 3.0/3.02 scoring function was used to identify the correct binding motif (Chen et al., 2003).

Circular Dichroism

All samples, both mutant and WT, were prepared by diluting protein to 0.2 mg/ml in 20 mM Tris-HCl (pH 7.5). The circular dichroism (CD) data was collected using a Jasco J810 CD spectropolarimeter. Spectral scans were collected at 20°C from 190 to 260 nm using 0.5-nm steps with five repeats.

NADPH Consumption Time Course

Consumption of NADPH by MxaA, or variants thereof, was measured by the decrease in absorption at 340 nm. 5.0 μM MxaA R was incubated in 100 mM potassium phosphate (pH 7.0) buffer containing 200 μM NADPH and 1.0 mM decanal in order to keep the substrate near saturating conditions. Measurements were recorded in triplicate and averaged; spontaneous NADPH degradation was accounted for in a control reaction lacking enzyme.

Fluorescence Titration of MxaA R and Mutants

Assays were prepared by adding NADPH (1 mM stock solution, 1–130 μM final concentration) to 10 μM MxaA or mutant R in buffer containing 100 mM phosphate and 300 mM NaCl at pH 7.25. Fluorescence was measured on a Teacan Safire fluorometer (λ_{ex} = 340 nm, λ_{em} = 460 nm with excitation and emission slits set to 7.5 nm) and the relative increase in fluorescence was measured by subtracting the autofluorescence of NADPH samples in the absence of enzyme from those interacting with the reductase domain. Plotting these data and fitting to the Michaelis-Menten equation determined the K_d and relative maximum fluorescence.

Determination of Enzyme-Specific Activities with Decanal

MxaA (WT or mutant) (20 μM final concentration) was added with NADPH (250 μM) and decanal (2 mM, saturating) to the reaction buffer (150 mM sodium phosphate, 200 mM NaCl) at a total volume of 200 μl. These reactions were monitored at 340 nM for the depletion of NADPH over 6 min, corrected for background NADPH consumption; the resultant slope was used to calculate the specific activity. Conversion to decanol was verified by gas chromatography-mass spectrometry.

Single Turnover Assay for R Domain Reduction

MxaA (WT or mutant) (50 μM final concentration) was combined with decanoyl-CoA (200 μM), $MgCl_2$ (10 mM), Sfp phosphopantetheinyl transferase (10 μM) (Quadri et al., 1998) in the reaction buffer (150 mM sodium phosphate, 200 mM NaCl) at a total volume of 300 μl. Sfp-mediated pPant loading proceeded for 2 hr at which point the extent of loading was determined by liquid chromatography-tandem mass spectrometry to provide 19 μM decanoyl-loaded MxaA PCP (Poust et al., 2015) (see Supplemental Information). The reaction was initiated with NADPH (250 μM). Control reactions showed no reduction of decanoyl-CoA in the absence of being loaded to the MxaA PCP. Reactions, done in duplicate, were stopped at 1, 2, and 3 hr with the addition of 30 μl of 10% (v/v) acetic acid and extracted with 2 × 300 μl of hexanes containing an internal standard of 100 μM dodecanol. Combined extracts were concentrated ~10-fold and mixed with an equal volume of *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and analyzed on a Hewlett Packard 6890 series gas chromatograph fitted with an Agilent 5973Network mass detector with a 30 m × 0.25 mm DB-5MS column (Agilent). Samples were injected at 80°C, held at that temperature for 2.0 min, and then ramped to 300°C at 25°C/min, held at 300°C for 1.0 min, and returned to the initial temperature. Samples were compared with an authentic decanol standard curve and normalized to internal dodecanol.

Details of the cloning of the Mxa PCP and R domain, site-directed mutagenesis, the enzyme assay, the liquid chromatography-tandem mass spectrometry PCP loading assay, size exclusion chromatography analysis, structural comparison of MxaA R, and Figures S1–S6 and Tables S1 and S2 can be found in the Supplemental Information.

ACCESSION NUMBERS

The MxaA R structures are accessible in the PDB (PDB: 4U7W, 4W4T).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Materials and Methods, six figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2015.06.022.

AUTHOR CONTRIBUTIONS

J.F.B. performed the crystallization experiments and bioinformatic analysis of the MxaA R domain. R.M.P and P.J.K. completed the in vitro protein experiments and biochemical analysis. A.J.S. performed all in silico experiments. J.F.B., R.M.P., and A.J.S. wrote the manuscript, which was proofread and approved by all authors.

ACKNOWLEDGMENTS

We would like to acknowledge Benjamin Adler for assistance with the biochemical assay of the MxaA didomain. We would also like to thank Dr. Gaurav Shakya for his insights in chemistry and Dr. P. Adams, C. Petzold, and L. Chan for assistance with the LC-MS/MS assay. This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. In addition, R.M.P. and J.D.K. would like to acknowledge the National Science Foundation for support through the Catalysis and Biocatalysis Program (CBET-1437775). A portion of this work was supported by grant ES001670 by the Pew Foundation. Crystallographic studies were performed at the Stanford Synchrotron Radiation Laboratory, a national user facility by Stanford University on behalf of the US Department of Energy Office of Basic Energy Sciences and the Advance Light Source at the Lawrence Berkeley National Laboratory. J.D.K. has financial interests in Amyris, LS9, and Lygos.

Received: April 23, 2015
Revised: June 5, 2015
Accepted: June 22, 2015
Published: July 30, 2015

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Atsumi, S., Hanai, T., and Liao, J.C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature *451*, 86–89.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. D Biol. Crystallogr. *67*, 271–281.

Bergmann, S., Schumann, J., Scherlach, K., Lange, C., Brakhage, A.A., and Hertweck, C. (2007). Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. Nat. Chem. Biol. *3*, 213–217.

Case, D.A., Babin, V., Berryman, J.T., Betz, R.M., Cai, Q., Cerutti, D.S., Cheatham, T.E., III, Darden, T.A., Duke, R.E., Gohlke, H., et al. (2014a). AMBER 14 (University of California).

Case, D.A., Darden, T., Cheatham, T.E., III, Adrian Roitberg, C.S., Wang, J., Duke, R.E., Luo, R., Roe, D.R., Walker, R.C., Legrand, S., et al. (2014b). Amber 14 Reference Manual (University of California).

Chen, R., Li, L., and Weng, Z. (2003). ZDOCK: an initial-stage protein-docking algorithm. Proteins *52*, 80–87.

Chhabra, A., Haque, A.S., Pal, R.K., Goyal, A., Rai, R., Joshi, S., Panjikar, S., Pasha, S., Sankaranarayanan, R., and Gokhale, R.S. (2012). Nonprocessive [2 + 2]e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. Proc. Natl. Acad. Sci. USA *109*, 5681–5686.

Chu, S., and Majumdar, A. (2012). Opportunities and challenges for a sustainable energy future. Nature *488*, 294–303.

Du, L., and Lou, L. (2010). PKS and NRPS release mechanisms. Nat. Prod. Rep. *27*, 255–278.

Duarte, J.M., Srebniak, A., Scharer, M.A., and Capitani, G. (2012). Protein interface classification by evolutionary analysis. BMC Bioinformatics *13*, 334.

Eklund, H., Plapp, B.V., Samama, J.P., and Branden, C.I. (1982). Binding of substrate in a ternary complex of horse liver alcohol dehydrogenase. J. Biol. Chem. *257*, 14349–14358.

Emsley, P., and Debreczeni, J.E. (2012). The use of molecular graphics in structure-based drug design. Methods Mol. Biol. *841*, 143–159.

Gaitatzis, N., Kunze, B., and Muller, R. (2001). In vitro reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. Proc. Natl. Acad. Sci. USA *98*, 11136–11141.

Gerth, K., Jansen, R., Reifenstahl, G., Hofle, G., Irschik, H., Kunze, B., Reichenbach, H., and Thierbach, G. (1983). The myxalamids, new antibiotics from *Myxococcus xanthus* (Myxobacterales). I. Production, physico-chemical and biological properties, and mechanism of action. J. Antibiot. *36*, 1150–1156.

Gomez-Escribano, J.P., Song, L.J., Fox, D.J., Yeo, V., Bibb, M.J., and Challis, G.L. (2012). Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of *Streptomyces coelicolor* M145. Chem. Sci. *3*, 2716–2720.

González, A., Moorhead, P., McPhillips, S.E., Song, J., Sharp, K., Taylor, J.R., Adams, P.D., Sauter, N.K., and Soltis, S.M. (2008). Web-Ice: integrated data collection and analysis for macromolecular crystallography. J. Appl. Crystallogr. *41*, 176–184.

Gotz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S., and Walker, R.C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. J. Chem. Theory Comput. *8*, 1542–1555.

Hahn, M., and Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. Proc. Natl. Acad. Sci. USA *103*, 275–280.

Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. Angew. Chem. Int. Ed. Engl. *48*, 4688–4716.

Hildebrand, A., Remmert, M., Biegert, A., and Soding, J. (2009). Fast and accurate automatic structure prediction with HHpred. Proteins *77* (*Suppl 9*), 128–132.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins *65*, 712–725.

Huang, C.C., Couch, G.S., Pettersen, E.F., and Ferrin, T.E. (1996). Chimera: an extensible molecular modeling application constructed using standard components. Pac. Symp. Biocomput. *1*, 724.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA *93*, 13–20.

Joosten, R.P., Joosten, K., Murshudov, G.N., and Perrakis, A. (2012). PDB_REDO: constructive validation, more than just looking for errors. Acta Crystallogr. D Biol. Crystallogr. *68*, 484–496.

Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). Biochemistry *34*, 6003–6013.

Kallberg, Y., Oppermann, U., and Persson, B. (2010). Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. FEBS J. *277*, 2375–2386.

Kavanagh, K.L., Jornvall, H., Persson, B., and Oppermann, U. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. Cell. Mol. Life Sci. *65*, 3895–3906.

Konz, D., and Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? Chem. Biol. *6*, R39–R48.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Leslie, A.G.W., and Powell, H.R. (2007). Processing diffraction data with MOSFLM. In Evolving Methods for Macromolecular Crystallography, J.L.S. Randy and J. Read, eds. (Springer), pp. 41–51.

Li, Y., Weissman, K.J., and Muller, R. (2008). Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. J. Am. Chem. Soc. *130*, 7554–7555.

Liebeschuetz, J.W., Cole, J.C., and Korb, O. (2012). Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. J. Comput. Aided Mol. Des. *26*, 737–748.

Masschelein, J., Clauwers, C., Awodi, U.R., Stalmans, K., Vermaelen, W., Lescrinier, E., Aertsen, A., Michiels, C., Challis, G.L., and Lavigne, R. (2015). A combination of polyunsaturated fatty acid, nonribosomal peptide and polyketide biosynthetic machinery is used to assemble the zeamine antibiotics. Chem. Sci. *6*, 923–929.

Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A., and Santi, D.V. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat. Biotechnol. *23*, 1171–1176.

Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B., et al. (2003). Short-chain dehydrogenases/reductases (SDR): the 2002 update. Chem. Biol. Interact. *143–144*, 247–253.

Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B., and Keasling, J.D. (2012). Microbial engineering for the production of advanced biofuels. Nature *488*, 320–328.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pierce, B.G., Hourai, Y., and Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. PLoS One *6*, e24657.

Poust, S., Hagen, A., Katz, L., and Keasling, J.D. (2014). Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. Curr. Opin. Biotechnol. *30*, 32–39.

Poust, S., Phelan, R.M., Deng, K., Katz, L., Petzold, C.J., and Keasling, J.D. (2015). Divergent mechanistic routes for the formation of gem-dimethyl groups in the biosynthesis of complex polyketides. Angew. Chem. Int. Ed. Engl. *54*, 2370–2373.

Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry *37*, 1585–1595.

Schrödinger LLC. (2013). The PyMOL Molecular Graphics System (Schrödinger LLC).

Sherman, D.H. (2005). The Lego-ization of polyketide biosynthesis. Nat. Biotechnol. *23*, 1083–1084.

Silakowski, B., Nordsiek, G., Kunze, B., Blocker, H., and Muller, R. (2001). Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15. Chem. Biol. *8*, 59–69.

Tanovic, A., Samel, S.A., Essen, L.O., and Marahiel, M.A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. Science *321*, 659–663.

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr. D Biol. Crystallogr. *65*, 582–601.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 105–124.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., and Taylor, R.D. (2003). Improved protein-ligand docking using GOLD. Proteins *52*, 609–623.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. *8*, 127–134.

Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. J. Comput. Chem. *25*, 1157–1174.

Wang, J., Wang, W., Kollman, P.A., and Case, D.A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. J. Mol. Graph. Model. *25*, 247–260.

Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol. *3*, 925–936.

Wickstrom, L., Okur, A., and Simmerling, C. (2009). Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. Biophys. J. *97*, 853–856.

Wilson, R.A., Gibson, R.P., Quispe, C.F., Littlechild, J.A., and Talbot, N.J. (2010). An NADPH-dependent genetic switch regulates plant infection by the rice blast fungus. Proc. Natl. Acad. Sci. USA *107*, 21902–21907.

Wyatt, M.A., Mok, M.C., Junop, M., and Magarvey, N.A. (2012). Heterologous expression and structural characterisation of a pyrazinone natural product assembly line. Chembiochem *13*, 2408–2415.