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

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

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.)
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 Tsai1,*

1Department of Molecular Biology and Biochemistry, Chemistry, and Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697, USA
2Joint Bioenergy Institute, 5885 Hollis Street, Emeryville, CA 94608, USA
3QB3 Institute, University of California, Berkeley, Berkeley, CA 94720, USA
4Department of Chemical and Biomolecular Engineering and Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA
5Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
6Co-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 myxalamid 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 syntheses (PKSs) and non-ribosomal peptide syntheses (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 agro-chemical) (Hertweck, 2009). The myxobacterium Stigmatella aurantiaca Sga15 contains a modular PKS/NRPS hybrid respon-
megasynthases to produce biologically derived fuels or commodity chemicals, such as 1-decanol (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, 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
with 2- or 4-electron reductions (Figure 1C) (Bergmann et al., 2008). A notable inserted HTH motif (S3A–S3C) (Jornvall et al., 1995; Kallberg et al., 2010; Kavanagh et al., 2008). Structural alignment with the type E SDR from Agrobacterium tumefaciens (PDB: 4D9) displays an RMSD of 3.88 Å through 119 residues of the alpha carbon backbone. Similar features include a canonical tyrosine-dependent catalytic triad (T1283, K1315, and Y1311) and a distinctive helix-turn-helix (HTH) motif (x16–x17) 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 (x12–x15–x16–x17–x20) and two parallel beta sheets (b9–b11), 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 (x16–x17) 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-Escobar et al., 2012; Li et al., 2008; Marschelein 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 Å², 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 Å² (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 Mycobacterium smegmatis involved in glycopeptide biosynthesis and the AusA R domain (PDB: 4F6C, 4F6L) from Staphylococcus 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 (Kd = 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 o xoanion 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 0.61–2.41 Å for the MxaA R domain (Figure 3A). Further 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 Cα 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 εO of R1364 through stochastic sampling of either helix 13 D1444 εO during 73.2% of the simulation (Figure S4A). Moderate electrostatic interactions were observed between helix 12 R1357 and helix 17 E1446, with the εC 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 TGGxxGxxG 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 larger 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 S4B). The top ChemPLP docking solution that was pre-screened 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 steered 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...
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. 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 carboxyl groups in the substrate near the thioester linkage.

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. 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 carboxyl groups in the substrate near the thioester linkage.
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 (PDB 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 advanced biofuel production from NADPH, 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.

**Figure 4. Myxalamid Docking Analysis**
(A) Stereo image of the MxaA R domain bound to NADPH and docked with myxalamid-pPant. 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. (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.

**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...
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 thio-ester 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).
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 catalytically incompelete 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 bio-technologically 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 C₃₀ 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 (Ponder et al., 2004). Charges were calculated using the AMBER ff14SB force field. Selenium-methionine 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/molÅ², 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/molÅ² 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 the Chimera 1.9 (Ponder 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 His₆ tag were expressed in BL21 (DE3) E. coli cells (Novagen). Cells containing the MxaA R domain plasmid were grown to OD₆₀₀ = 0.6 at 37 °C in LB medium containing 50 μg/ml kanamycin. The cell cultures were cooled to 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. Resuspending cells were cooled on ice for 30 min and the cells were disrupted using sonication. The cell debris was cleared by centrifugation at 5,525 relative centrifugal force (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 His₆ 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 CaCl₂. 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 inoculated to 2 x 1 l of LB, which was allowed to grow at 37 °C in the presence of 50 μg/ml kanamycin until OD₆₀₀ = 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 transferred to 2 x 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 incubated 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 (8 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 Mosfim and Web-ice (Gonzalez 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 (reflection), 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 Mosfim for the P21 space group (Battey 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 Mosfim 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 (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 phosphopantetheinylated 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., 1999). 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 phosphopantetheinylated 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
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 Shaky 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


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 (λem = 340 nm, λex = 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 KAx 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), MgCl₂ (10 mM), Stp phosphopantetheiny/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. Stp-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 (Pouat 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 x 300 μl of hexanes containing an internal standard of 100 μM dodecane. 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 x 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 dodecane.

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).


