Structure of DnmZ, a nitrososynthase in the *Streptomyces peucetius* anthracycline biosynthetic pathway

Lauren Sartor,^a Charmaine Ibarra,^a Ahmad Al-Mestarihi,^{b,‡} Brian O. Bachmann^{b,c} and Jessica L. Vey^{a,*}

Received 13 June 2015

Accepted 28 July 2015

Edited by Z. S. Derewenda, University of Virginia, USA

‡ Current address: Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY 40506, USA.

Keywords: flavin monooxygenase; nitrososynthase; *N*-oxidation; biosynthetic pathway; protein crystal structure.

PDB references: DnmZ, ligand-free state, 4zxv; TDP-bound state, 4zyj

Supporting information: this article has supporting information at journals.iucr.org/f

^aDepartment of Chemistry and Biochemistry, California State University Northridge, Northridge, CA 91330, USA,

^bDepartment of Chemistry, Vanderbilt University, Nashville, TN 37235, USA, and ^cDepartment of Biochemistry, Vanderbilt University Medical Center, Nashville, TN 37232, USA. *Correspondence e-mail: jessica.vey@csun.edu

The anthracyclines are a class of highly effective natural product chemotherapeutics and are used to treat a range of cancers, including leukemia. The toxicity of the anthracyclines has stimulated efforts to further diversify the scaffold of the natural product, which has led to renewed interest in the biosynthetic pathway responsible for the formation and modification of this family of molecules. DnmZ is an *N*-hydroxylating flavin monooxygenase (a nitrososynthase) that catalyzes the oxidation of the exocyclic amine of the sugar nucleotide dTDP-L-epi-vancosamine to its nitroso form. Its specific role in the anthracycline biosynthetic pathway involves the synthesis of the seven-carbon acetal moiety attached to C4 of L-daunosamine observed in the anthracycline baumycin. Here, X-ray crystallography was used to elucidate the three-dimensional structure of DnmZ. Two crystal structures of DnmZ were yielded: that of the enzyme alone, solved to 3.00 Å resolution, and that of the enzyme in complex with thymidine diphosphate, the nucleotide carrier portion of the substrate, solved to 2.74 Å resolution. These models add insights into the structural features involved in substrate specificity and conformational changes involved in thymidine diphosphate binding by the nitrososynthases.

1. Introduction

The anthracyclines are glycosylated tetracyclic polyketide natural products that are currently used quite successfully as chemotherapeutics (Piekarski & Jelińska, 2013). The anthracyclines daunorubicin (DNR) and doxorubicin (DOX, which is a C14-hydroxylated semisynthetic derivative of DNR) are effective against a range of leukemias and lymphomas. Although DNR is the most prevalent isolatable form produced by the source organism *Streptomyces peucetius*, it is actually a degradation product of the even more complex baumycins (Takahashi *et al.*, 1977). These molecules are characterized by a unique acetal attached through a glycosidic linkage to the C4' hydroxyl of the daunosamine of DNR (Fig. 1). While the biosynthetic pathway of DNR has been well studied over the last several decades, the biosynthesis of the baumycin acetal has only recently been addressed. Recent work has shown that this acetal originates from thymidine diphosphate-linked L-epi-vancosamine, which undergoes a retro oxime-aldol-type cleavage at its C3'–C4' bond upon *N*-oxidation to the nitroso congener by the nitrososynthase DnmZ (Al-Mestarihi *et al.*, 2013). At this point, it is unclear whether the cleavage is directly catalyzed by DnmZ, is facilitated by unique features of the active site of DnmZ, or

occurs without enzymatic involvement after oxidation by DnmZ. After cleavage and a series of unknown subsequent steps, the acetal is transferred to the O4' atom of daunosamine by the glycosyltransferase DnrH (Scotti & Hutchinson, 1996).

According to its sequence homology and reaction requirements, DnmZ is a nitrososynthase from the class D subfamily of the flavin monooxygenases (FMOs; Al-Mestarihi *et al.*, 2013). The flavin monooxygenases are a diverse group of enzymes that catalyze oxygen insertions using molecular oxygen as the oxygen source. These enzymes use a conserved strategy to catalyze a variety of reactions, including hydroxylations, Baeyer–Villiger oxidations, epoxidations and even halogenations (Huijbers *et al.*, 2014). All FMOs activate molecular oxygen by catalyzing its reaction with reduced flavin to form a C4a-(hydro)peroxyflavin intermediate (Palfey & McDonald, 2010). The enzymes then use this intermediate to hydroxylate a nucleophilic or electrophilic substrate, depending on the enzyme active site.

Individually, the FMOs exhibit significant variability at many levels, from their three-dimensional folds, to the reactions that they catalyze, to the details of the active sites and catalytic steps (McDonald *et al.*, 2011; Huijbers *et al.*, 2014; Chaiyen *et al.*, 2012). This variability reflects the exquisite customization of the enzyme to its reaction, and presents researchers with a puzzle to decipher with each new FMO. As patterns in reactivity have emerged, researchers have classified the enzymes according to their structural and sequence homologies and selected aspects of catalysis (van Berkel *et al.*, 2006). At present, eight classes of FMOs (classes A–H) have been delineated (Huijbers *et al.*, 2014).

The class D FMOs are set apart from the other FMOs by their acyl-CoA dehydrogenase (ACAD)-type fold and their dependence on a separate flavin reductase to supply reduced flavin (van Berkel *et al.*, 2006). The most common and best-characterized reaction type associated with the class D FMOs is electrophilic aromatic substitution, as exemplified by

4-hydroxyphenylacetate 3-monooxygenase, trichlorophenyl-4-monooxygenase and 3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione monooxygenase. Several heteroatom-oxidation reactions have been observed in this class and are now coming under more detailed study; for example, sulfoxidase, *N*-hydroxylase and nitrososynthase activity has been confirmed in specific class D FMO enzymes (in dibenzothiophene monooxygenase, isobutylamine *N*-hydroxylase and the nitrososynthases ORF36, RubN8, KijD3 and DnmZ, respectively).

N-Oxidation reactions such as that under study here are catalyzed by subsets of both the class B and class D FMOs. These particular oxygen-insertion reactions are involved in forming important functional groups in the products of their metabolic pathways. For example, the class B *N*-hydroxylases SidA (Olucha & Lamb, 2011), MbsG (Robinson *et al.*, 2014) and others form the hydroxylamines of hydroxamate-containing siderophores, completing the metal-binding functional group, while the class D nitrososynthases KijD3 (Bruender *et al.*, 2010; Thoden *et al.*, 2013), ORF36 (Hu *et al.*, 2008; Vey *et al.*, 2010) and DnmZ (Al-Mestarihi *et al.*, 2013) oxidize their respective amino-sugar substrates to generate moieties essential for the bioactivity of the final product. While there are some similarities between the *N*-hydroxylases from the two FMO classes (all currently known representatives oxidize primary amines, for example), the details of catalysis are likely to be quite distinct. For example, class B FMOs use NADPH to reduce an enzyme-bound FAD. After the reaction of the flavin with O₂, the NADP⁺ remains bound and helps to form and stabilize a long-lived C4a-(hydro)peroxyflavin intermediate. Meanwhile, class D FMOs do not bind NAD(P)H, and instead depend on an external flavin reductase to supply reduced flavin. Therefore, the mechanism of (hydro)peroxyflavin intermediate formation and stabilization must differ between the class B and class D FMOs.

Figure 1

Left, the proposed DnmZ reaction. Once formed, the product undergoes retro oxime–alcohol cleavage at the C3'–C4' bond before incorporation into baumycin (Al-Mestarihi *et al.*, 2013). Right, the baumycin A1 structure with the baumycin acetal highlighted in blue.

Here, we report the crystallization and X-ray crystal structure determination of DnmZ, the first of the class D FMO nitrososynthases known to trigger an on-pathway retro oxime–aldol cleavage reaction. Analysis of this structure provides insight into ligand binding and conformational changes during catalysis by the class D FMO nitrososynthases. Further investigations of DnmZ and other class D FMOs will lead to a deeper understanding of the various catalytic strategies used by the flavin monooxygenases.

2. Materials and methods

2.1. Macromolecule production

Purified DnmZ was obtained as described previously (Al-Mestarihi *et al.*, 2013). Briefly, *dnmZ* was amplified from *S. peucetius* genomic DNA and subcloned into pET-28a, yielding pET-28-ZN. This plasmid encodes N-terminally 6×His-tagged DnmZ (for complete details, see Supplementary Table S1). *Escherichia coli* BL21(DE3)/pET-28-ZN cells were induced with 0.1 mM IPTG at 28°C, harvested after 6 h and lysed using a French press in 20 mM Tris–HCl, 0.5 M NaCl, 20 mM imidazole pH 7.5. DnmZ was purified by Ni²⁺-affinity chromatography and exchanged into 20 mM Tris–HCl pH 7.5, 1 mM DTT, 5% glycerol using a desalting column. Purified DnmZ was stored at –80°C prior to crystallization.

2.2. Crystallization

Initial crystallization conditions for DnmZ were identified by sparse-matrix screening using the hanging-drop vapor-diffusion method at 298 K with 2 µl drops [1 µl DnmZ at 6 mg ml^{–1} and 1 µl 0.05 M glycine pH 9.5, 0.2 M ammonium sulfate, 10%(w/v) PEG 4000] equilibrated against 0.75 ml crystallization solution. Standard optimization methods, including adjusting the protein:crystallization solution ratio, were employed to increase the crystal size and diffraction quality. The rate of vapor diffusion was manipulated by varying the well volume, equilibrating against concentrated salt solutions and layering oil over the well. Co-crystallization experiments including thymidine diphosphate (dTDP), the nucleotide carrier portion of the substrate, were also carried out. For these experiments, DnmZ was pre-incubated with 5 mM dTDP for 30 min prior to crystallization. Co-crystallization and soaking experiments aimed at obtaining a structure of DnmZ complexed with dTDP-L-epi-vancosamine, FAD, FMN or combinations of the amino sugar and flavin were carried out using similar methods to those described here. These experiments have been unsuccessful thus far.

2.3. Data collection and processing

All crystals were flash-cooled in liquid nitrogen with or without cryoprotectant prior to data collection. Early crystals were screened on a Bruker MICROSTAR microfocussing rotating-anode source with Proteum CCD detectors. X-ray diffraction data were collected remotely on beamline 12-2 of the Stanford Synchrotron Radiation Laboratory (SSRL) at the Stanford Light Accelerator Center, Menlo Park, Cali-

fornia, USA using the *Blu-Ice* software (Stepanov *et al.*, 2011). The data sets were processed using *XDS* (Kabsch, 2010) or *HKL-2000* (Otwinowski & Minor, 1997).

2.4. Structure determination, refinement and structural analysis

The DnmZ data were phased by molecular replacement in *Phaser* (McCoy *et al.*, 2007) from the *CCP4* program suite (Winn *et al.*, 2011) with ORF36 as the search model (PDB entry 3mxl; Vey *et al.*, 2010; 59% sequence identity to DnmZ). The *PRODRG* online server (Schüttelkopf & van Aalten, 2004) was used to create parameter and topology files for refinement of the dTDP molecule. The DnmZ models were built in *Coot* (Emsley & Cowtan, 2004) and refined in both *CNS* (Brünger *et al.*, 1998) and *REFMAC5* (Murshudov *et al.*, 2011). NCS restraints were used throughout the refinement in both *CNS* (using NCS restraint groups) and *REFMAC5* (using tight main-chain and medium side-chain restraints). NCS restraints were slowly relaxed for segments of the chains that clearly differed (exclusively in surface loops from the central β domain; see below), based on differences in the $2F_o - F_c$ and $F_o - F_c$ electron-density maps. DnmZ structural homologs were identified using the *DALI* server (Holm & Rosenström, 2010). Flavin and substrate were placed into our models by superimposing the ternary KijD3 model (PDB entry 4kcf; Thoden *et al.*, 2013), which contains bound flavin and the hydrated substrate analog dTDP-3-amino-2,3,6-trideoxy-4-keto-3-methyl-D-glucose), onto the dTDP-bound DnmZ model using *LSQMAN* (Kleywegt, 1996). This substrate analog differs from the actual DnmZ substrate in its stereochemistry at the C5' atom (L *versus* D configuration of the deoxy sugar) and in the substituents at the C4' atom. The substituent on C4' of the DnmZ substrate is a single hydroxyl group, whereas the substrate analog has a hydrated keto group at this position. The structure of the substrate analog was modified manually at C4' and C5' to generate the correct structure of the substrate, subjected to energy minimization using the *PRODRG* server (Schüttelkopf & van Aalten, 2004) and superimposed onto the coordinates of the original substrate analog for use in figures. The UniProtKB database was searched for nitrososynthases using the DnmZ sequence as an input for *BLAST* (Altschul *et al.*, 1997). Sequences with greater than 50% identity were considered to be likely nitrososynthases. A selection of those sequences (Supplementary Table S2) were aligned with *ClustalW* (Larkin *et al.*, 2007) and input into the *ESPrpt* server (Robert & Gouet, 2014) to calculate the sequence conservation of each DnmZ residue across the nitrososynthases. Figures were prepared using *PyMol* (v.1.3r1; Schrödinger).

3. Results

3.1. Purification and crystallization of DnmZ

DnmZ was purified to >95% homogeneity using previously described methods (Al-Mestarihi *et al.*, 2013) and crystallized in 0.1 M glycine, 0.12 M ammonium sulfate, 12% PEG 2000,

Table 1

Data collection and processing.

Values in parentheses are for the outer shell.

	Apo DnmZ	dTDP–DnmZ
Diffraction source	Beamline 12-2, SSRL	Beamline 12-2, SSRL
Wavelength (Å)	0.95370	0.97950
Temperature (K)	100	100
Detector	Dectris PILATUS 6M	Dectris PILATUS 6M
Crystal-to-detector distance (mm)	650	605
Rotation range per image (°)	0.2	0.15
Total rotation range (°)	90	180
Exposure time per image (s)	1	0.5
Space group	$P2_12_1$	$P2_12_1$
<i>a</i> , <i>b</i> , <i>c</i> (Å)	99.84, 134.49, 142.92	100.48, 134.67, 144.29
α , β , γ (°)	90, 90, 90	90, 90, 90
Mosaicity (°)	0.7	0.3
Resolution range (Å)	48.9–3.00 (3.11–3.00)	39.43–2.74 (2.88–2.74)
Total No. of reflections	123512 (11591)	325167 (33667)
No. of unique reflections	38251 (3691)	51414 (6704)
Completeness (%)	97.8 (95.3)	98.2 (88.9)
Multiplicity	3.2 (3.1)	6.3 (5.0)
$\langle I/\sigma(I) \rangle$	17.5 (2.3)	32.9 (4.1)
$R_{\text{r.i.m.}}$ † (%)	7.6 (56.5)	4.5 (39.1)
Overall <i>B</i> factor from Wilson plot (Å ²)	97.06	74.36

† The redundancy-independent merging *R* factor $R_{\text{r.i.m.}}$ was estimated by multiplying the conventional R_{merge} value by the factor $[N/(N-1)]^{1/2}$, where *N* is the data multiplicity.

pH 9.3–9.4. These needles grew to their largest size in 2–5 d. Increasing the rate of vapor diffusion by equilibrating drops against 2 *M* ammonium sulfate was required for the formation of three-dimensional crystals. Two distinct DnmZ crystal morphologies grew in these conditions: a large, diamond-shaped form that grew quickly to dimensions of 250 × 250 × 500 µm and a smaller, slower-growing cuboidal form that reached maximum dimensions of 20 × 50 × 75–100 µm (Supplementary Fig. S1). See Supplementary Table S3 for a summary of the DnmZ crystallization conditions.

The DnmZ crystals are fragile and dissolve quickly in soaking solutions. Therefore, crystals were cryoprotected by extremely brief soaks in supplemented solutions (glycerol, ethylene glycol, 2-methyl-2,4-pentanediol and other common cryoprotectants were tested at a range of concentrations) or were not cryoprotected at all. Surprisingly, the highest-quality data (Table 1) were collected from crystals that were not cryoprotected.

The large, diamond-shaped crystals diffracted to 5–9 Å resolution and had a primitive triclinic lattice. We were not able to improve diffraction by this crystal form using standard optimization, additives or microseeding, and we therefore focused on the cuboidal crystal form. Crystals with this macroscopic shape diffracted to 3–5 Å resolution (Supplementary Figure S2) and had either a monoclinic or an orthorhombic lattice. Crystal size seemed to influence the lattice type: larger crystals typically adopted the lower symmetry lattice. Otherwise, crystals of the two lattice types were indistinguishable at the macroscopic level. Data collected from the orthorhombic lattice were used in our structure determinations and are summarized in Table 1.

Table 2

Structure solution and refinement.

Values in parentheses are for the outer shell.

	Apo DnmZ	dTDP–DnmZ
Resolution range (Å)	48.97–3.00 (3.081–3.00)	98.45–2.74 (2.81–2.74)
Completeness (%)	97.6 (92.52)	98.0 (77.02)
σ Cutoff	$F > 0.000\sigma(F)$	$F > 0.000\sigma(F)$
No. of reflections, working set	36262 (2357)	48689 (2699)
No. of reflections, test set	1937 (117)	2618 (123)
Final <i>R</i> factor (%)	21.0 (29.0)	17.4 (27.3)
Final R_{free} (%)	26.5 (34.8)	23.5 (35.6)
No. of non-H atoms		
Protein	11438	11558
Ligand	0	100
Water	27	113
Total	11465	11771
R.m.s.d.		
Bonds (Å)	0.011	0.013
Angles (°)	1.496	1.762
Average <i>B</i> factors (Å ²)		
Protein	100.116	61.685
Ligand	n/a	94.654
Water	76.0	50.705
Ramachandran plot		
Most favoured (%)	95.0	95.0
Allowed (%)	5.0	5.0

3.2. Quality and content of the DnmZ models

This work yielded two X-ray crystal structures (Table 2): DnmZ in the absence of ligands (apo DnmZ; solved to 3.00 Å resolution with an *R* factor of 21.0% and an R_{free} of 26.5%) and DnmZ with bound dTDP, the nucleotide carrier portion of the substrate (dTDP–DnmZ; solved to 2.74 Å resolution with an *R* factor of 17.4% and an R_{free} of 23.5%). The two models have relatively high average *B* factors that are consistent with the observed Wilson *B* factors. In both models the asymmetric unit consists of the DnmZ tetramer, which is the full biological assembly (*Z* = 4). As expected, the four chains of the apo DnmZ model are highly similar, with an average r.m.s.d. of 0.154 ± 0.041 Å over 384 atoms, compared with an r.m.s.d. of 0.470 ± 0.090 Å for the dTDP–DnmZ chains, of which chain *C* is the least similar. The lower average r.m.s.d. of the apo DnmZ model can be attributed to the tighter NCS restraints used during refinement owing to the lower resolution of the data set. Residues 10–405 and 11–405 were modeled in the apo DnmZ and dTDP–DnmZ models, respectively, each with several chain breaks. All of these chain breaks occur in the central β domain (see below) in loops at the protein surface. Unsurprisingly, this portion of each chain has consistently higher *B* factors than the helical domains. There are minor differences in each chain in terms of disordered residues and main-chain and side-chain conformations. These differences are most likely owing to conformational flexibility. See Supplementary Table S4 for a list of all of the missing residues in both apo DnmZ and dTDP–DnmZ.

3.3. The DnmZ fold and active-site cleft

DnmZ adopts the same tetrameric ACAD quaternary fold (Fig. 2*a*) as observed in the ORF36 (Vey *et al.*, 2010) and KijD3 structures, which include the dTDP-bound KijD3 model

(Bruender *et al.*, 2010) and the ternary KijD3 model with FMN and a substrate analog bound (Thoden *et al.*, 2013). The DnmZ monomer consists of three domains. The N-terminal domain (residues 11–139) is made up exclusively of helices (denoted as helices $\alpha 1$ – $\alpha 5$; Supplementary Fig. S3a), and the central domain (residues 140–250) is a nine-stranded β -sandwich (strands $\beta 1$ – $\beta 9$). The central β domain is connected by an eight-residue loop to the C-terminal α -helical domain (residues 251–405, helices $\alpha 6$ – $\alpha 10$), which provides the tetramerization interface. The β domain has the highest degree of conformational flexibility of the three domains, as illustrated by higher observed differences in conformation between each chain of the tetramer, chain breaks and significantly higher *B* factors.

A wide cleft, approximately 35 Å wide and 15 Å deep, is formed between the β domain and the C-terminal domain, with $\alpha 4$ of the N-terminal domain forming the 'base' of the opening (Fig. 2*b* and Supplementary Figs. S3*b* and S3*c*). The active site is presumably located in this cleft, by analogy to KijD3 and other class D FMOs and ACADs that have been structurally characterized. The flavin-binding site, predicted based on sequence homology and structural superposition with the ternary KijD3 model (Thoden *et al.*, 2013), and the dTDP-binding site observed in this structure are found on opposite sides of the cleft. The site of catalysis, where the substrate deoxy amino sugar and flavin isoalloxazine come into contact, is centrally located and was again identified by superposition with the ternary KijD3 model. This site is located directly adjacent to a loop containing two tandem *cis*-peptide bonds (residues His387–Tyr389) that connects helices $\alpha 9$ and $\alpha 10$ (Supplementary Fig. S4). This 'cis-peptide loop' is a unique feature of the nitrososynthases and has been observed in all three structurally characterized enzymes thus far. A large portion of the binding and catalytic sites are pre-ordered at the base of the cleft, as judged by comparisons between the apo and dTDP-bound DnmZ models and the two KijD3 models.

3.4. The dTDP-binding site and predicted flavin-binding site

DnmZ binds dTDP in much the same manner as observed in the KijD3 structures. dTDP binds at one end of the cleft of the protein, where the nucleotide-binding site is formed primarily by residues from helices $\alpha 4$ and $\alpha 6$ (Fig. 3*a*). As in KijD3, the thymidine nucleoside is well positioned by several hydrogen bonds to Thr113, Glu117, Ala322 and Arg332 and π -stacking interactions with Phe116. Thymine is recognized by interactions between the C4 carbonyl O atom and Arg332, and between both N3 and the C4 carbonyl O atom and a water molecule that hydrogen-bonds to the backbone carbonyl of Ala322. The ribose 3' hydroxyl group is hydrogen-bonded by Glu117, which is suitably positioned to form a salt bridge with Arg243. In stark contrast to the nucleoside moiety, the dTDP diphosphate makes no apparent interactions with the protein. Indeed, this group is observed in a different conformation in each chain of the dTDP tetramer. This was a surprising find, although not unexpected. In the ternary KijD3 model the

phosphates form hydrogen bonds to Arg106 (Arg110 in DnmZ) and Ser257 (Ser260 in DnmZ), both of which are

Figure 2

The DnmZ structure. (*a*) The dTDP–DnmZ tetramer is shown viewed down one of its twofold axes. Chain *A* is shown in color to visualize the three domains, with the N-terminal helical domain in purple, the central β domain in green and the C-terminal domain in pink. The other three chains are shown in different shades of gray. The bound dTDP molecules are shown in thick stick representation with carbon shown in teal, oxygen in red, nitrogen in blue and phosphate in orange. (*b*) A closer view of the DnmZ monomer with selected secondary-structural elements labeled and ligands modeled into the active site by superposition with KijD3 (PDB entry 4kcf). Modeled FMN and substrate are displayed as in (*a*), with FMN C atoms in yellow and substrate C atoms in teal.

conserved in the nitrososynthases (Supplementary Fig. S5). However, in the dTDP-bound structure of KijD3 (PDB entry 3m9v) the dTDP diphosphate group makes only one of these binding interactions (to Ser257), and indeed the terminal

phosphate group points away from the flavin-binding site. A possible explanation of the differences in diphosphate binding mode observed between the dTDP-bound structures and the ternary KijD3 model is that the full dTDP-deoxy sugar and/or

Figure 3

Stereoviews of the dTDP-binding site and putative active site. (a) dTDP is shown colored as in Fig. 2(a) with $F_o - F_c$ electron density contoured at 3.0σ . All dTDP coordinates were omitted from the model, followed by refinement in *REFMAC* and recalculation of the electron density to generate the difference map shown here. Nearby residues that interact with dTDP are shown in stick representation with C atoms colored gray. Possible hydrogen-bonding interactions are shown as black dashed lines. (b) Residues found in the proposed flavin-binding and catalytic sites are shown in stick representation along with local secondary structure. Modeled FMN and substrate are shown as thin sticks colored as in Fig. 2(b).

flavin cofactor must be present in order to achieve the correct binding interactions in these enzyme active sites. This explanation is consistent with the currently available structures.

The predicted flavin-binding site is formed by residues emanating from all three DnmZ domains and the loop connecting $\alpha 8$ and $\alpha 9$ from a neighboring chain (Fig. 3*b*). Numerous nonpolar amino-acid side chains form a hydrophobic pocket that can accommodate the dimethylbenzene portion of the flavin cofactor. A hydrogen-bonding network including Ser174, Ser223, Ser225 and Trp215 holds the protein surface in the correct conformation to allow hydrogen-bonding interactions with the flavin isoalloxazine, including a hydrogen bond between the flavin N5 atom and Ser174. This interaction is conserved in the flavin monooxygenases and is thought to play a role in stabilizing the hydroperoxyflavin intermediate (Chaiyen *et al.*, 2012).

Based on superposition with KijD3, the site of catalysis is located in a region of the active-site cleft formed by $\alpha 4$, $\beta 1$, $\alpha 6$ and the *cis*-peptide loop (Fig. 3*b*). The residues that form the probable interaction surfaces for the deoxy-sugar moiety are His105, Met106, Ser109, Ser260, Ser261 and Met264 on one side of the cleft (the histidine and serine residues are all conserved in the nitrososynthases) and Gly145 and Val147 on the other side of the cleft. Lastly, the *cis*-peptide loop forms the upper entrance to the catalytic site. In DnmZ, His387 is located on this loop, with its side chain facing away from the active site. As in the KijD3 and ORF36 structures, no obvious general acids or bases are located within reach of the predicted catalytic site unless a significant conformational change were to occur, such as in the *cis*-peptide loop to move His387 into the active site. Such a conformational change seems unlikely given the ternary KijD3 model, which has a very similar backbone conformation in this region as in our DnmZ structures. If a *cis*- to *trans*-peptide isomerization were to occur in this loop to bring His387 (or its equivalent) into the active site, the corresponding isomerization should have been observed in the ternary KijD3 model, assuming that the enzymes have a conserved mechanism of action.

4. Discussion

This paper presents the first structures of DnmZ, the third nitrososynthase to be structurally characterized and the first nitrososynthase known to trigger an on-pathway retro oxime–aldol cleavage reaction. The availability of both apo and dTDP-bound DnmZ models has allowed comparison of the three nitrososynthase structures to identify potential functional differences and conformational changes associated with dTDP binding.

There are currently four characterized dTDP-deoxy amino sugar nitrososynthases: ORF36, KijD3, RubN8 and DnmZ. These enzymes share high sequence homology and structural similarity: the enzymes share 55–65% sequence identity and the r.m.s.d.s between ORF36, KijD3 and DnmZ range from 0.84 to 0.95 Å (see Supplementary Table S5 for the individual r.m.s.d.s). A BLAST search suggests that there are numerous uncharacterized homologs (see Supplementary Table S2 for

representative examples). The current state of knowledge on nitrososynthases tells us that KijD3, ORF36, RubN8 and DnmZ all oxidize the exocyclic amine of 2,3,6-trideoxy-3-amino sugars, and that each enzyme prefers a unique pathway-specific deoxy sugar (Hu *et al.*, 2008; Bruender *et al.*, 2010; Vey *et al.*, 2010; Al-Mestarihi *et al.*, 2013). For example, the true substrate of DnmZ is dTDP-L-epi-vancosamine (Al-Mestarihi *et al.*, 2013), while that of ORF36 and RubN8 is dTDP-L-evernosamine, the 4-*O*-methylated derivative of dTDP-L-epi-vancosamine (Vey *et al.*, 2010), although all three enzymes do have varying levels of activity towards both sugars. KijD3 is active against dTDP-3-amino-2,3,6-trideoxy-4-keto-3-methyl-D-glucose (Bruender *et al.*, 2010), which differs from the aforementioned sugars in the configuration at C5' and the substituent at C4' (a keto group *versus* a hydroxyl or hydroxymethyl group). The molecular details that determine substrate specificity in the nitrososynthases are currently an open question, and with several structures in hand we can begin to consider the differences between the enzymes that may impact specificity and other catalytic details.

The majority of the residues in the active-site cleft are highly conserved in the nitrososynthases. This includes the residues involved in the proposed flavin-binding, dTDP-binding and catalytic sites. There are some notable exceptions to this observation, including His387 of the DnmZ *cis*-peptide loop and Met106 of the proposed catalytic site of DnmZ (Supplementary Figure S5). His387 is a glutamine in the majority of the nitrososynthases and a histidine in other class D FMOs such as *p*-hydroxyphenylacetate 3-hydroxylase, where it was shown to be involved in formation of the C4a-(hydro)peroxyflavin intermediate (Thotsaporn *et al.*, 2011). The conformation of the DnmZ protein backbone keeps the His387 side chain sequestered above and outside the DnmZ active site, as occurs for the corresponding glutamine of KijD3 and ORF36. This difference between the nitrososynthases and the other class D FMOs could have implications for the catalytic mechanism in terms of how the peroxyflavin intermediate is formed and whether it is stabilized in the absence of substrate by the nitrososynthases, among other catalytic steps. Further study of the nitrososynthases using, for example, transient-state kinetic methods will help to address this question.

Meanwhile, the location of Met106 in the active site and its variability in the nitrososynthase sequences suggests that it may play a role in substrate selectivity. Because our structures have only the nucleotide carrier portion of the actual substrate bound, discussions of substrate selectivity must focus on the ternary KijD3 model, comparisons between that structure and those presented here, and sequence conservation of substrate-binding residues. Residue numbering corresponding to the KijD3 model will be denoted by a 'k' in this discussion. As noted by the authors of the ternary KijD3 model, the residues located closest to the substrate deoxy sugar are kHis101 (equivalent to His105 in DnmZ), kMet102 (Met106 in DnmZ), kSer141 (Gly145 in DnmZ) and kIle143 (Val147 in DnmZ) (Thoden *et al.*, 2013). kIle143, which is always a small hydrophobic residue in the nitrososynthases (Val, Ile or Leu;

Supplementary Fig. S5) packs against the deoxy sugar C2' atom, while kSer141 (which is always a Ser or Gly) is located on the surface of the cleft underneath the O5' and O4A atoms of the sugar. kHis101, which is completely conserved in the nitrososynthases, is nearest to O4B, the second oxygen substituent on C4' of the sugar, although in the KijD3 structure the distance between the histidine N atom and the substrate O atom, 3.6 Å, is longer than expected for a hydrogen bond. Finally, kMet102 (which is always a small hydrophobic residue, either Met, Val, Leu or Ala) packs against C4' and C6' of the substrate deoxy sugar. In addition to these residues, the KijD3 active site is hydrated, with several ordered water molecules mediating interactions between the substrate and protein. For example, the kTyr409 C-terminal tail, which has no counterpart in DnmZ, appears to form hydrogen bonds to the substrate amine and phosphoryl O atom through water molecules 684 and 647, respectively.

There appears to be a relationship between the residues corresponding to kSer141 and kIle143 in that any nitrososynthase that has a Ser at the first of these two positions always has an Ile at the second position, while in a second group of nitrososynthases a Gly in the first position is paired with either a Val or Leu in the second position. This pattern may be involved in substrate specificity, possibly to select a specific sugar conformation or configuration, but this is not clear from the currently available data. On the other hand, Met106 (kMet102) may play a role in substrate specificity at the deoxy-sugar C4' atom. This residue is a Met in DnmZ and KijD3, both of which are thought to prefer substrates with hydroxyl substituents on C4', while ORF36 and RubN8 (both of which prefer the 4-*O*-methylated deoxy sugar dTDP-L-evernosamine) have the smaller residues Val and Ala at this position, respectively. The size of the side chain in this position may in part dictate the size of the substituent allowed at the C4' atom of the substrate. Further study of the nitrososynthases, including structural studies and experimentation aimed at clarifying the substrate preferences of each enzyme, is necessary to clarify the specific roles of each residue in selectivity.

The average r.m.s.d. between our apo and dTDP-bound DnmZ monomers is 0.55 ± 0.042 Å over 386 main-chain atoms. The major conformational differences between the two models are highlighted in Fig. 4. These differences are concentrated within the β -sheet domain, while the N-terminal and C-terminal domains remain very similar, with the base of the active-site cleft generally unchanged. The surface of the β domain exhibits significant differences between the models, but as differences in these areas are observed between different monomers of the same tetramer, these are attributed to protein flexibility. Two conformational changes that do appear to be associated with dTDP binding occur in two loops located adjacent to the site of catalysis (loop A in Fig. 4) and the dTDP-binding site (loop B in Fig. 4). These two changes are observed consistently in all (or a majority) of the four chains of the tetramers, and neither region directly acts as a crystal lattice contact in either model. Firstly, the loop following β 1, which carries residues that form the expected

flavin- and deoxy sugar-binding sites, forms a five-residue turn that is held in place by a hydrogen bond between Asp149 and the backbone and side chain of Thr154. This turn bends 'towards' the active-site cleft in all four chains of the apo DnmZ model and shifts 'away' from the cleft in the dTDP-DnmZ chains. A more significant change occurs in the last β -strand and loop of the β domain. This region, which carries Arg243 of the dTDP-binding site, shifts towards the dTDP and the N-terminal helical domain in the dTDP-DnmZ chains. The extent of this shift in the dTDP-DnmZ model varies between the four chains, but a consistent shift is indeed observed. This movement allows the formation of several additional interactions between the β domain and N-terminal

Figure 4

Illustration of the conformational changes that occur on dTDP binding in two orientations. (a) Apo DnmZ and dTDP-DnmZ are both colored gray and shown in cartoon representation. The two regions that undergo conformational changes are designated loop A (named L2 in the ORF36 structure) and loop B (named L6 in the ORF36 structure). These two regions are colored pink in the apo DnmZ model and blue in the dTDP-DnmZ model. The bound dTDP molecule is shown as described in Fig. 2(a) along with modeled FMN and substrate, which are represented as described in Fig. 2(b). Hydrogen-bonding interactions between DnmZ and bound dTDP are shown as dashed black lines. (b) A closer view of the differences between the two models in the loop B region.

domain. For example, the Arg243 side chain (which is disordered in most of the apo DnmZ chains) moves to interact with the dTDP 3'-hydroxyl and form a salt bridge with Glu117. In addition, in the dTDP-DnmZ chains Leu241 makes more extensive hydrophobic interactions with N-terminal domain residues, and Arg239 moves ~ 2.5 Å to form a hydrogen bond to Asp131. Although these conformational changes do appear to be associated with dTDP binding, no side-chain or backbone rearrangements are observed within the flavin-binding or catalytic site of the active-site cleft (see Supplementary Movie S1 for an animated depiction of the conformational changes).

The conformational changes described above are not directly observed in the KijD3 and ORF36 structures, as the corresponding pair of structures is not currently available for either of those two enzymes. However, the KijD3 and ORF36 models are generally consistent with our observations. The ORF36 structure (no ligands bound) mimics the apo DnmZ conformation in the loop A and loop B regions, while the two KijD3 structures (both of which have dTDP bound) recall the dTDP-DnmZ conformation, particularly in the loop B region (Supplementary Fig. S6). The most significant structural difference between the four models is the extreme C-terminus, corresponding to residues 416–437 in KijD3. This portion of the C-terminus is disordered in the dTDP-bound form of KijD3 (Bruender *et al.*, 2010) and the ORF36 structure (Vey *et al.*, 2010). In the ternary KijD3 model the C-terminus becomes ordered and extends into the active-site cleft, contributing residue side chains (Tyr409 in particular) to interact with the ligands and bury the active site (Thoden *et al.*, 2013). Surprisingly, the C-terminus of DnmZ is truncated compared with the other two enzymes: the protein chain ends just after the final helix (Supplementary Fig. S5) and therefore cannot extend into the cleft of the protein to bury the active site. A similarly shortened C-terminus occurs in a putative oxidoreductase from *S. griseus* (UniProtKB entry Q54197), which also carries a histidine at the position equivalent to His387 in DnmZ instead of the glutamine found in the other nitrososynthases (Supplementary Fig. S5). Because this reaction is expected to proceed *via* a peroxyflavin intermediate that could undergo nonproductive decomposition, we expect these enzymes to employ an alternative strategy to close off the active site during catalysis. Analysis of our structures did not yield an obvious answer to this question, although the structures do hint at possible strategies, such as a conformational change involving the *cis*-peptide loop, a rigid-body-type motion of the central β domain to close the opening of the cleft, or a combination of the two. As discussed above, a change in the *cis*-peptide loop does not appear to be likely based on the KijD3 structures, although the combined histidine substitution and truncated C-terminus observed in both DnmZ and UniProtKB entry Q54197 could indicate a difference in this respect from KijD3. On the other hand, the high apparent mobility of the central β domain of DnmZ would be expected if a rigid-body motion of this domain were to occur. Further structural and biochemical characterization of DnmZ and other nitrososynthases will shed more light on the

substrate selectivity, conformational changes and reaction mechanisms of this interesting group of enzymes.

Acknowledgements

The authors acknowledge Joel Harp, Jens Kaiser, Douglas Rees and Irimpan Mathews for access to data-collection facilities and experimental support. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. JLV acknowledges NIH Grant 5SC2AI109500, Research Corporation CCSA 22672, the CSUPERB New Investigator Grant Program and the CSUN Office of Research and Sponsored Projects. BOB acknowledges the Office of Naval Research Grant N00014-09-1-012.

References

- Al-Mestarihi, A., Romo, A., Liu, H. & Bachmann, B. O. (2013). *J. Am. Chem. Soc.* **135**, 11457–11460.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
- Berkel, W. J. H. van, Kamerbeek, N. M. & Fraaije, M. W. (2006). *J. Biotechnol.* **124**, 670–689.
- Bruender, N. A., Thoden, J. B. & Holden, H. M. (2010). *Biochemistry*, **49**, 3517–3524.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst. D* **54**, 905–921.
- Chaiyen, P., Fraaije, M. W. & Mattevi, A. (2012). *Trends Biochem. Sci.* **37**, 373–380.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst. D* **60**, 2126–2132.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Hu, Y., Al-Mestarihi, A., Grimes, C. L., Kahne, D. & Bachmann, B. O. (2008). *J. Am. Chem. Soc.* **130**, 15756–15757.
- Huijbers, M. M., Montersino, S., Westphal, A. H., Tischler, D. & van Berkel, W. J. (2014). *Arch. Biochem. Biophys.* **544**, 2–17.
- Kabsch, W. (2010). *Acta Cryst. D* **66**, 125–132.
- Kleywegt, G. J. (1996). *Acta Cryst. D* **52**, 842–857.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- McDonald, C. A., Fagan, R. L., Collard, F., Monnier, V. M. & Palfey, B. A. (2011). *J. Am. Chem. Soc.* **133**, 16809–16811.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst. D* **67**, 355–367.
- Olucha, J. & Lamb, A. L. (2011). *Bioorg. Chem.* **39**, 171–177.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Palfey, B. A. & McDonald, C. A. (2010). *Arch. Biochem. Biophys.* **493**, 26–36.
- Piekarski, M. & Jelińska, A. (2013). *Mini Rev. Med. Chem.* **13**, 627–634.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
- Robinson, R. M., Rodriguez, P. J. & Sobrado, P. (2014). *Arch. Biochem. Biophys.* **550–551**, 58–66.
- Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst. D* **60**, 1355–1363.

- Scotti, C. & Hutchinson, C. R. (1996). *J. Bacteriol.* **178**, 7316–7321.
- Stepanov, S., Makarov, O., Hilgart, M., Pothineni, S. B., Urakhchin, A., Devarapalli, S., Yoder, D., Becker, M., Ogata, C., Sanishvili, R., Venugopalan, N., Smith, J. L. & Fischetti, R. F. (2011). *Acta Cryst. D* **67**, 176–188.
- Takahashi, Y., Naganawa, H., Takeuchi, T., Umezawa, H., Komiyama, T., Oki, T. & Inui, T. (1977). *J. Antibiot.* **30**, 622–624.
- Thoden, J., Branch, M., Zimmer, A., Bruender, N. & Holden, H. (2013). *Biochemistry*, **52**, 3191–3193.
- Thotsaporn, K., Chenprakhon, P., Sucharitakul, J., Mattevi, A. & Chaiyen, P. (2011). *J. Biol. Chem.* **286**, 28170–28180.
- Vey, J. L., Al-Mestarihi, A., Hu, Y., Funk, M. A., Bachmann, B. O. & Iverson, T. M. (2010). *Biochemistry*, **49**, 9306–9317.
- Winn, M. D. *et al.* (2011). *Acta Cryst. D* **67**, 235–242.