BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

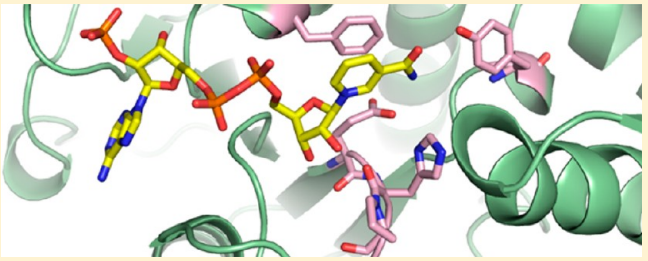
pubs.acs.org/biochemistry

# Two Structures of a Thiazolinyl Imine Reductase from *Yersinia enterocolitica* Provide Insight into Catalysis and Binding to the Nonribosomal Peptide Synthetase Module of HMWP1

Kathleen M. Meneely and Audrey L. Lamb*

Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66045, United States

**ABSTRACT:** The thiazolinyl imine reductase from *Yersinia enterocolitica* (Irp3) catalyzes the NADPH-dependent reduction of a thiazoline ring in an intermediate for the formation of the siderophore yersiniabactin. Two structures of Irp3 were determined in the apo (1.85 Å) and $NADP^+$-bound (2.31 Å) forms. Irp3 is structurally homologous to sugar oxidoreductases such as glucose-fructose oxidoreductase and 1,5-anhydro-D-fructose reductase, as well as to biliverdin reductase. A homology model of the thiazolinyl imine reductase from *Pseudomonas aeruginosa* (PchG) was generated. Extensive loop insertions are observed in the C-terminal domain that are unique to Irp3 and PchG and not found in the structural homologues that recognize small molecular substrates. These loops are hypothesized to be important for binding of the nonribosomal peptide synthetase modules (found in HMWP1 and PchF, respectively) to which the substrate of the reductase is covalently attached. A catalytic mechanism for the donation of a proton from a general acid (either histidine 101 or tyrosine 128) and the donation of a hydride from C4 of nicotinamide of the NADPH cofactor is proposed for reduction of the carbon−nitrogen double bond of the thiazoline.



Pathogenic bacteria must acquire iron from the environment, because iron is an essential element required as a cofactor by the enzymes of many pathways, including electron transport, DNA synthesis, and the tricarboxylic acid cycle. A typical bacterium requires ∼1 μM iron for optimal growth, but the concentration of available iron in the human host is $\sim 10^{-9}$ μM.[1] In addition, Fe(III) is very insoluble and frequently biologically inaccessible. For pathogenic bacteria, this requirement is a significant obstacle to overcome. Therefore, bacteria have developed elaborate systems for scavenging iron from the environment and/or host.[2] One such system is the use of siderophores, low-molecular weight iron chelators, which the bacteria synthesize, secrete, and then selectively take up in the iron-loaded form to survive and colonize human tissues.[2−4]

*Yersinia* and *Pseudomonas* spp. produce salicyl-capped siderophores that contain thiazoline and thiazolidine rings, both derived from cyclized cysteines (Figure 1). The latter ring is reduced from the thiazoline by an NADPH-dependent reductase, possibly to attain the appropriate geometry to coordinate ferric iron.[5] Two enzymes that catalyze this reduction are the focus of this study, in particular Irp3 from *Yersinia enterocolitica* and PchG from *Pseudomonas aeruginosa*, the sequences of which are 26% identical. The homologue from *Yersinia pestis*, YbtU, is different from Irp3 at 8 of 365 amino acids (98% identical). YbtU was originally identified as required for yersiniabactin formation, but with an unknown role.[6] A sequence analysis suggested that there was a possible transmembrane helix from residue 162 to 182.[6]

Functional characterization of PchG showed that the enzyme was a required reductase for pyochelin maturation that produced an intermediate, desmethyl-pyochelin, in which the second thiazoline ring is reduced (Figure 1).[7] The substrate for PchG, a hydroxyphenyl-bisthiazolinyl precursor of pyochelin, is covalently attached to the peptidyl carrier domain of PchF, a nonribosomal peptide synthetase (NRPS), through a phosphopantetheinyl (Ppant) post-translational modification of a serine (Figure 1B).[7] A complete in vitro reconstitution for the production of pyochelin required PchD (a salicylate adenylase), PchE and PchF (two nonribosomal peptide synthetases modified to have the Ppant group), PchG, NADPH, *S*-adenosylmethionine, ATP, cysteine, and salicylate.[5] The authors noted that PchG showed weak sequence similarity to biliverdin reductase (BVR) and proposed a mechanism of hydride transfer from the NADPH cofactor followed by protonation of the nitrogen of the developing thiazolidine.[5] The N-terminal portion of PchG was hypothesized to be structurally similar to the C-terminal portion of mouse major urinary protein, which had been determined with 2-(*sec*-butyl)thiazoline bound.[5] A similar role for the reduction of the second thiazoline ring of yersiniabactin was subsequently described for YbtU when the substrate is Ppant-tethered to the peptidyl carrier domain of the NRPS module of high-molecular weight protein 1 (HMWP1) (Figure 1A).[8] Kinetic

**Received:** August 14, 2012
**Revised:** September 11, 2012
**Published:** October 15, 2012





dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002−9013

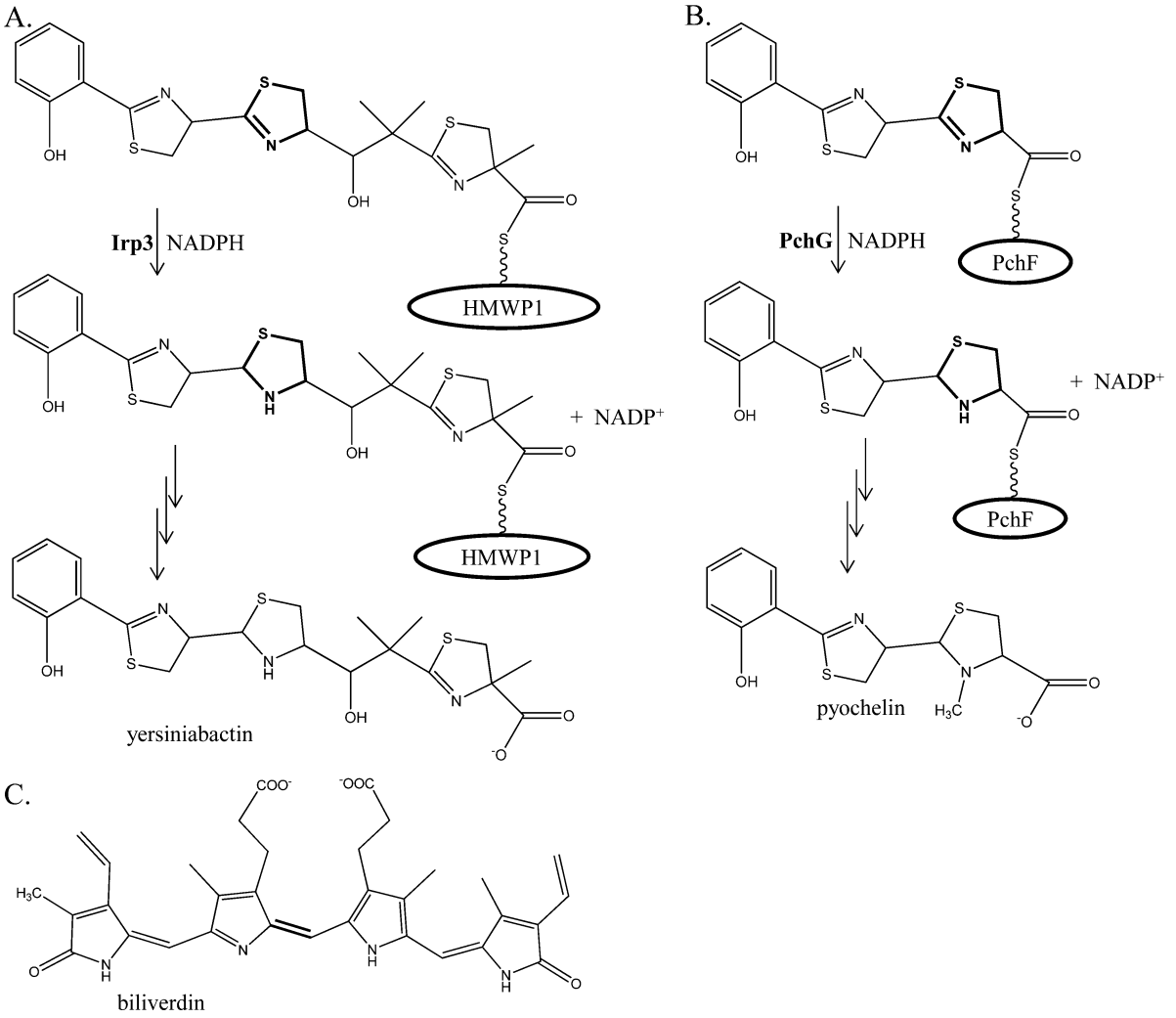



**Figure 1.** Yersiniabactin, pyochelin, and biliverdin. (A) Yersiniabactin, the siderophore produced by *Yersinia* spp., is the product of the biosynthetic pathway to which Irp3 belongs. Irp3 reduces a double bond in the center thiazoline ring to make the thiazolidine ring of the product. The substrate is reduced while covalently attached to a phosphopantetheinyl post-translationally modified serine of the peptidyl carrier domain of the high-molecular weight protein 1 (HMWP1). (B) Pyochelin is a siderophore produced by *P. aeruginosa*. PchG performs a reaction analogous to that of Irp3. The substrate is reduced while covalently attached to a phosphopantetheinyl post-translationally modified serine of the peptidyl carrier domain of the nonribosomal peptide synthetase PchF. This ring is subsequently methylated by the methylation domain of PchF. (C) Biliverdin is the substrate of biliverdin reductase, an enzyme that catalyzes the final step in heme degradation, reducing the center methylene bridge to form bilirubin. Biliverdin reductase is a structural homologue of Irp3, as found herein.

measurements for PchG, YbtU, and Irp3 are complicated by the nature of the substrate, an immature siderophore covalently tethered to an NRPS. Therefore, a next viable step for understanding these enzymes is acquiring structural information to provide context for catalysis and protein−protein interactions.

Here we present the first structural characterization of an NADPH-dependent thiazolinyl imine reductase with the structures of apo and $NADP^+$-bound Irp3. We also present the homology model of PchG. We propose an interaction motif for Irp3 and PchG with the nonribosomal peptide synthetase modules to which their substrates are covalently attached. We further propose a catalytic mechanism based on the structures and by analogy to the closest structural homologues of known function, which are sugar oxidoreductases and biliverdin reductase.

## MATERIALS AND METHODS

**Preparation of Overexpression Plasmids To Generate Irp3 Protein with N- and C-Terminal Histidine Tags.** To generate a plasmid that produces Irp3 protein with a C-terminal histidine tag, the *irp3* gene was amplified from *Y. enterocolitica* 33114 genomic DNA (ATCC) by polymerase chain reaction by use of Master Mix (Eppendorf) supplemented with a 25 mM magnesium solution. The forward primer (5′-ATT CTT CAT ATG CCG TCC GCC TCC CCA AAA CA-3′) includes an *Nde*I site (underlined), whereas the reverse primer (5′-GGA TCC CTC GAG CGC CTC CTT ATC ATC ATC GTT G-3′) contains an *Xho*I site (underlined). The amplified 1119 bp fragment was digested with *Nde*I and *Xho*I and ligated into the pET29b plasmid (Novagen) digested with the same enzymes. The resulting plasmid encodes the *irp3* gene with a C-terminal histidine tag (Irp3-$his_6$-C).

The plasmid for producing Irp3 protein with an N-terminal histidine tag was generated by polymerase chain reaction by use


dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002−9013

**Table 1. Irp3 SeMet Data Collection Statistics[a]**

| | SeMet[b] | | |
|---|---|---|---|
| | inflection | peak | remote |
| wavelength (Å) | 0.97944 | 0.97915 | 0.91837 |
| space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| cell dimensions [*a*, *b*, *c* (Å)] | 82.97, 93.06, 181.37 | 82.93, 92.97, 181.26 | 83.00, 93.09, 181.34 |
| resolution (Å) | 51.14−2.60 (2.74−2.60) | 51.11−2.60 (2.74−2.60) | 51.15−2.60 (2.74−2.60) |
| $R_{sym}$[c] | 0.106 (0.373) | 0.105 (0.366) | 0.107 (0.361) |
| total no. of observations | 325730 (47180) | 324939 (47099) | 326255 (47276) |
| total no. of unique observations | 44022 (6307) | 43953 (6305) | 44045 (6306) |
| mean (*I*)/sd(*I*) | 12.5 (4.9) | 12.5 (5.1) | 12.6 (5.0) |
| completeness (%) | 100.0 (100.0) | 100.0 (100.0) | 100.0 (100.0) |
| redundancy | 7.4 (7.5) | 7.4 (7.5) | 7.4 (7.5) |
| anomalous completeness (%) | 100.0 (100.0) | 100.0 (100.0) | 100.0 (100.0) |
| anomalous redundancy | 3.8 (3.8) | 3.8 (3.7) | 3.8 (3.8) |
| no. of refined selenium sites | | 23 | |
| DelAnom correlation between half-sets[d] | 0.229 (−0.004) | 0.204 (0.011) | −0.009 (−0.009) |
| mean figure of merit[e] | | 0.41 | |
| midslope of anomalous probability[f] | 1.124 | 1.114 | 0.99 |

[a]All data were collected at beamline 9-2 at the Stanford Synchrotron Radiation Laboratory. Values in parentheses are for the highest-resolution shells. [b]Data indexed and scaled with XDS. [c]$R_{sym} = \sum_h|I_h - \langle I\rangle|/\sum_h I_h$, where $I_h$ is the intensity of reflection $h$ and $\langle I\rangle$ is the mean intensity of all symmetry-related reflections. [d]Anomalous correlation between half-sets = $\sum||FPH \pm FP| - FH_{calc}|/\sum|FPH|$ reported for all centric reflections. [e]Mean figure of merit = $\langle a\sum P_a e^{ia}/\sum P_a\rangle$, where $a$ is the phase and $P_a$ is the phase probability distribution. [f]Midslope of anomalous probability = $\langle|FH|/||FP + FH| - |FPH||\rangle$ reported for all reflections.

of Herculase II (Agilent) using the Irp3-his$_6$-C plasmid for the template, the Irp3-his$_6$-C forward primer, and a reverse primer (5′-GGA TCC CTC GAG TCA CAG CGC CTC CTT ATC ATC ATC GTT G-3′) containing an *Xho*I site (underlined) and a stop codon. The amplified 1122 bp fragment was digested with *Nde*I and *Xho*I and ligated into the pET28b plasmid (Novagen) digested with the same enzymes. The resulting plasmid encodes the *irp3* gene with an N-terminal histidine tag (N-his$_6$-Irp3).

**Irp3-his$_6$-C Protein Overexpression and Purification.** *Escherichia coli* BL21(DE3) cells containing the Irp3-his$_6$-C expression plasmid were grown in LB broth containing 50 μg/mL kanamycin at 22 °C while being shaken (225 rpm). The cells were harvested by centrifugation (6000*g* for 10 min at 4 °C) after 24 h. The cell pellet was resuspended in 15 mL of 25 mM Tris-HCl (pH 8), 500 mM NaCl, and 5 mM imidazole (buffer A) per liter of culture. Cells were disrupted by use of a French pressure cell (35000 psi), and cellular debris was removed by centrifugation (12000*g* for 30 min at 4 °C). The supernatant was applied to a chelating Sepharose fast-flow column (Amersham Biosciences) charged with nickel chloride and pre-equilibrated in buffer A. Irp3-his$_6$-C protein eluted at 300 mM imidazole in a linear gradient of 5 to 500 mM imidazole in buffer A. The pooled fractions were applied to a Superdex 200 size-exclusion column (Amersham Biosciences) equilibrated with 25 mM Tris-HCl (pH 8), 200 mM NaCl, and 2 mM dithiothreitol. The fractions containing Irp3-his$_6$-C were pooled and concentrated by use of an Amicon stirred cell with a YM-30 membrane to 19 mg/mL as determined by the Bradford assay and stored at −80 °C.

**N-his$_6$-Irp3 Protein Overexpression and Purification.** *E. coli* BL21(DE3) cells containing the N-his$_6$-Irp3 expression plasmid were grown in LB broth containing 50 μg/mL kanamycin at 37 °C while being shaken (225 rpm). When an optical density of ∼0.6 was reached, the temperature was reduced to 25 °C. Protein expression was induced with the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 200 μM. The cells were harvested by centrifugation (6000*g* for 10 min at 4 °C) after ∼18 h. The cell pellet was resuspended in 10 mL of buffer A per 4 L of culture and purified like Irp3-his$_6$-C with a nickel-charged chelating Sepharose fast-flow column followed by a Superdex 200 size-exclusion column (Amersham Biosciences) equilibrated with 50 mM potassium phosphate (pH 8) and 100 mM sodium citrate. The fractions containing N-his$_6$-Irp3 were pooled and concentrated by use of an Amicon stirred cell with a YM-10 membrane to 19 mg/mL as determined by the Bradford assay and stored at −80 °C.

**Selenomethionine-Substituted (SeMet) N-his$_6$-Irp3 Protein Overexpression and Purification.** The method for overexpression of SeMet N-his$_6$-Irp3 was adapted from the protocol of Van Duyne et al.[9] with several modifications. *E. coli* BL21(DE3) cells containing the N-his$_6$-Irp3 expression plasmid were grown in M9 minimal medium containing 50 μg/mL kanamycin and inoculated with 20 mL of overnight culture (grown in LB) per liter of medium at 37 °C while being shaken (225 rpm). When an optical density of ∼0.6 was reached, the temperature was reduced to 25 °C and the methionine biosynthesis pathway was inhibited by the addition of 100 mg each of L-lysine, L-phenylalanine, and L-threonine; 50 mg each of L-isoleucine, L-leucine, and L-valine; and 60 mg of DL-selenomethionine per liter of culture. Fifteen minutes after the addition of the amino acid mixture, protein expression was induced by the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 200 μM. Cultures were grown for ∼18 h. SeMet N-his$_6$-Irp3 was purified using the same protocol that was used for N-his$_6$-Irp3 with the addition of 1 mM β-mercaptoethanol to all buffers.

**Crystallization of SeMet N-his$_6$-Irp3.** SeMet N-his$_6$-Irp3 protein was crystallized using the hanging drop method at 24 °C with drops containing 1.5 μL of purified protein mixed with equal volumes of a reservoir solution composed of 0.1 M HEPES (pH 7.5), 29% (w/v) PEG 3350, and 0.225 M ammonium citrate. Large crystals (25 μm × 75 μm × 150 μm)


dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002−9013

formed after 1−2 weeks. For data collection, crystals were transferred to the reservoir solution supplemented with 20% (v/v) ethylene glycol as a cryoprotectant and flash-cooled by being plunged into liquid nitrogen.

**Crystallization of Irp3-his$_6$-C.** Crystallization was conducted by the hanging drop method at 24 °C. Drops containing 1.5 μL of purified Irp3-his$_6$-C protein supplemented with 1 mM dichloroethylenediaminoplatinate were mixed with equal volumes of a reservoir solution composed of 0.1 M HEPES (pH 7.3), 32% (w/v) PEG 3350, and 0.25 M lithium sulfate. Large crystals (50 μm × 100 μm × 150 μm) formed after 1−2 weeks. For data collection, crystals were serially washed in a reservoir solution supplemented with 20% (v/v) ethylene glycol as a cryoprotectant and flash-cooled by being plunged into liquid nitrogen.

**Crystallization of N-his$_6$-Irp3 with NADP$^+$ Bound.** N-his$_6$-Irp3 was crystallized using the hanging drop method at 24 °C with drops containing 1.5 μL of purified N-his$_6$-Irp3 protein supplemented with 3 mM NADP$^+$ (Sigma) mixed with equal volumes of a reservoir solution composed of 0.1 M Tris-HCl (pH 8.0), 13% (w/v) PEG 3350, and 0.11 M magnesium formate. Rod crystals (25 μm × 25 μm × 150 μm) formed after 1−2 weeks. For data collection, crystals were transferred to a reservoir solution supplemented with 6 mM NADP$^+$ and 20% (v/v) ethylene glycol as a cryoprotectant and flash-cooled by being plunged into liquid nitrogen.

**Collection of Data for and Determination of the Structure of SeMet N-his$_6$-Irp3.** Multiwavelength anomalous dispersion data were collected at Stanford Synchrotron Radiation Laboratory (Stanford, CA) beamline 9-2. Peak (0.97915 Å), inflection (0.97944 Å), and remote (0.91837 Å) wavelengths were determined by a selenium fluorescence scan. Data (180°) were collected as 0.3° oscillation images (2 s exposure per frame) with a crystal−detector distance of 295 mm at 100 K. The crystals were assigned to space group $P2_12_12_1$ with the following unit cell dimensions: $a$ = 83 Å, $b$ = 93 Å, and $c$ = 181 Å. The data were indexed and scaled to 2.60 Å using the XDS program package.[10] AutoSolve and AutoBuild from the PHENIX software package[11] were used to determine the location of 23 selenium atoms (figure of merit = 0.41, overall score = 40.5), to calculate initial phase estimates, and to build the initial model with 1275 of the possible 1460 residues. Data processing and phasing statistics are listed in Table 1.

**Collection of Data for and Determination of the Structure of Irp3-his$_6$-C, Considered the "Native" Structure.** Irp3-his$_6$-C diffraction data (0.5° oscillation images for a total of 90.5°) were collected at Stanford Synchrotron Radiation Laboratory beamline 9-2 with a wavelength of 1.000034 Å at 100 K. The exposure time per frame was 5 s with a crystal−detector distance of 250 mm. The data were indexed with MOSFLM[12] and scaled with SCALA in the CCP4 program package[13] to 1.85 Å. The crystals were assigned to space group $P2_12_12_1$ with the following unit cell dimensions: $a$ = 83.87 Å, $b$ = 93.90 Å, and $c$ = 181.12 Å. Because the SeMet and native data were isomorphous, the initial SeMet model and phase estimates were used to build the higher-resolution native Irp3 model. Model building and refinement of the structure were accomplished using Coot[14] and Refmac5[15] from the CCP4 software suite. Waters were added manually using the water addition function in Coot, and suggested positions were manually verified in Coot following a refinement cycle in Refmac. Molecules from the crystallization solution were also modeled (sulfate, ethylene glycol, and HEPES). Alternate conformations were tested for several residues but in the end are not found in the final model because the density did not justify their inclusion. Data collection and refinement statistics are listed in Table 2.

**Collection of Data for and Determination of the Structure of NADP$^+$-Bound N-his$_6$-Irp3.** Oscillation images of N-his$_6$-Irp3 crystals grown in the presence of NADP$^+$ (90° with an oscillation angle of 0.6°) were collected at Stanford Synchrotron Radiation Laboratory beamline 9-2 with a wavelength of 0.97949 Å at 100 K. The exposure time per

**Table 2. Data Collection and Refinement Statistics for Native and NADP$^+$-Bound Irp3[a]**

| | native | NADP$^+$-bound |
|---|---|---|
| | Data Collection[b] | |
| wavelength (Å) | 1.000034 | 0.97944 |
| space group | $P2_12_12_1$ | $P2_12_12_1$ |
| cell dimensions [$a$, $b$, $c$ (Å)] | 83.87, 93.90, 181.12 | 83.42, 93.82, 181.99 |
| resolution (Å) | 76.11−1.85 (1.95−1.85) | 39.89−2.31 (2.43−2.31) |
| $R_{sym}$[c] | 0.054 (0.395) | 0.087 (0.389) |
| total no. of observations | 465318 (67058) | 234222 (33388) |
| total no. of unique observations | 122197 (17701) | 63447 (9044) |
| mean ($I$)/sd($I$) | 12.4 (3.1) | 11.2 (3.2) |
| completeness (%) | 99.8 (100.0) | 99.6 (98.5) |
| redundancy | 3.8 (3.8) | 3.7 (3.7) |
| | Refinement | |
| resolution (Å) | 40.97−1.85 (1.90−1.85) | 39.89−2.31 (2.37−2.31) |
| $R_{cryst}$[d] | 19.6 (26.8) | 19.7 (25.8) |
| $R_{free}$ | 24.2 (32.8) | 26.0 (33.1) |
| total no. of unique observations | 115840 | 60174 |
| no. of non-hydrogen atoms | 10938 | 11156 |
| protein | 10397 | 10854 |
| ligand | 20 | 192 |
| solvent | 23 | 0 |
| water | 498 | 110 |
| rmsd for bonds (Å) | 0.02 | 0.015 |
| rmsd for angles (deg) | 2.01 | 1.93 |
| overall mean $B$ factor (Å$^2$) | 28.3 | 28.7 |
| Ramachandran plot analysis[e] | | |
| favored region | 98.4 | 96.9 |
| allowed region | 1.4 | 2.9 |
| outlier reaion | 0.2 | 0.3 |

[a]All data were collected at beamline 9-2 at the Stanford Synchrotron Radiation Laboratory. Values in parentheses are for the highest-resolution shells. [b]Native data indexed with MOSFLM and scaled with Scala and NADP$^+$-bound data indexed and scaled with XDS. [c]$R_{sym} = \sum_h|I_h - \langle I\rangle|/\sum_h I_h$, where $I_h$ is the intensity of reflection $h$ and $\langle I\rangle$ is the mean intensity of all symmetry-related reflections. [d]$R_{cryst} = \sum||F_o| - |F_c||/\sum|F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure factor amplitudes, respectively. Five percent of the relections were reserved for the calculation of $R_{free}$. [e]Calculated with RAMPAGE.


dx.doi.org/10.1021/bi3011016 | Biochemistry 2012, 51, 9002−9013

**Table 3. Ordered Amino Acids in Structures by Monomeric Chain**

| | native | $NADP^+$-bound |
|---|---|---|
| monomer A | 7−75, 78−255, 270−357 | −7 to 254,[a] 262−357 |
| monomer B | 8−26, 31−74, 78−255, 270−357 | 7−255, 262−357 |
| monomer C | 7−27, 31−74, 78−254, 270−357 | 7−254, 270−357 |
| monomer D | 7−74, 78−254, 271−357 | 7−255, 264−357 |

[a]Includes seven N-terminal amino acids of the purification tag.

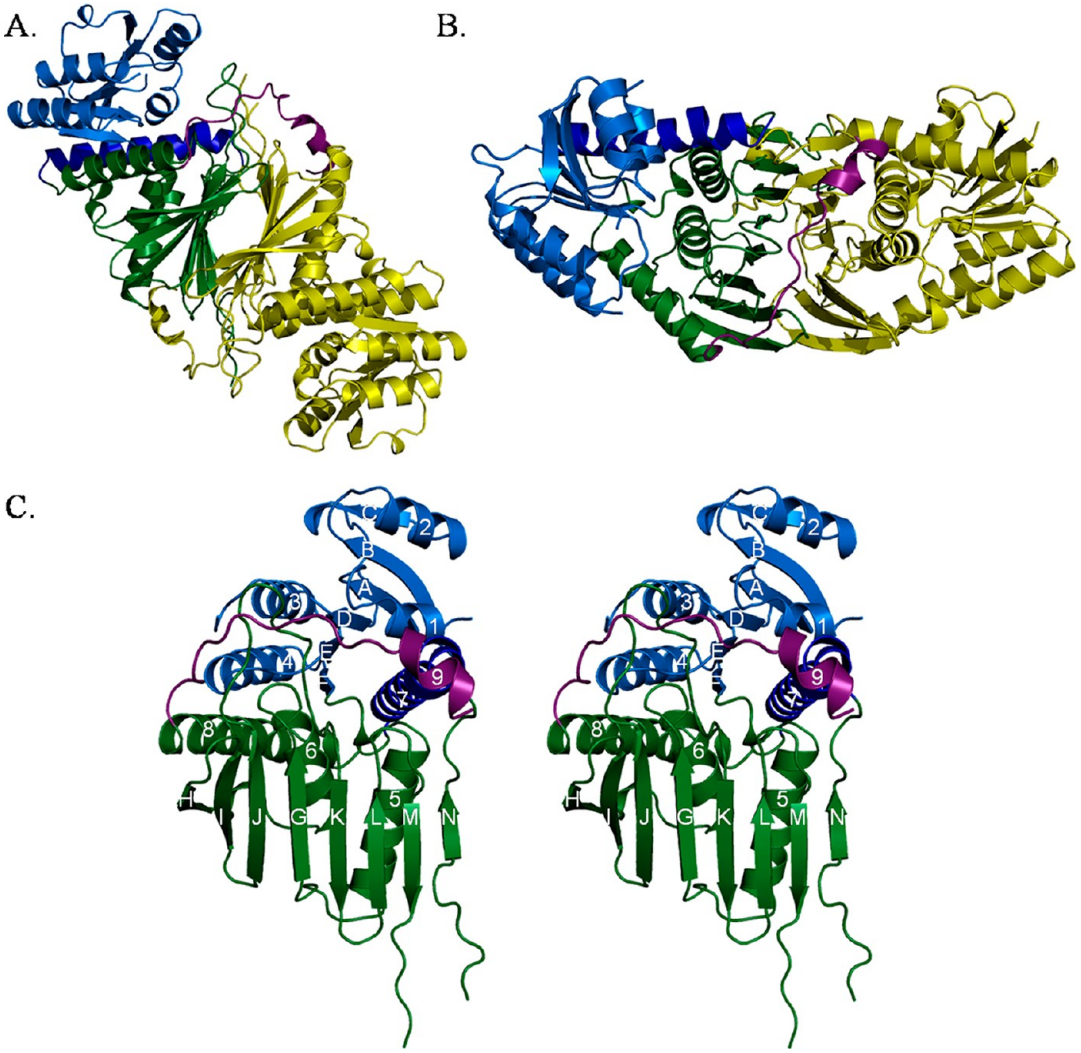



**Figure 2.** Overall architecture of native Irp3. (A and B) Dimeric structure of Irp3. Monomer A is colored yellow. Monomer B is colored blue (N-terminal domain) and green (C-terminal domain), with the C-terminal tail colored magenta. Panels A and B are related by a 90° rotation about the *x*-axis. The dimerization interface is evident where the open sides of the *β*-sheet in the C-terminal domains pack together. The magenta helix of the C-terminal tail packs against the opposing monomer, also contributing to the dimerization interface. (C) Stereoview of the monomer. The N-terminal, NADP(H) binding domain (residues 1−125) is colored blue, with helix 7 donated from the C-terminus colored dark blue (residues 285−310). The C-terminal substrate binding domain is colored green (residues 126−284 and 311−338). The C-terminal tail is colored magenta. Strands are denoted by letters A−N, starting from the N-terminus. Helices are counted from the N-terminus, 1−8. Monomer A of the native structure is depicted.

frame was 25 s with a crystal−detector distance of 330 mm. The data were indexed and scaled to 2.3 Å with XDS.[10] The crystals were assigned to space group $P2_12_12_1$ with the following unit cell dimensions: $a$ = 83.42 Å, $b$ = 93.82 Å, and $c$ = 181.99 Å. Molecular replacement calculations were performed using PHASER[16] and the data from 38.0 to 2.3 Å. Molecule A of the Irp3-$his_6$-C native structure was used as a search model yielding a clear solution with a log likelihood gain of 12806.61. The map generated with this solution showed unmodeled electron density in the active site corresponding to the $NADP^+$ and 15 additional residues at the N-terminus of molecule A compared with the native Irp3-$his_6$-C model. These differences in electron density relative to the molecular replacement model indicate a correct solution. Iterative rounds of model building and refinement were performed with Coot[14] and Refmac5.[15] Waters were added manually using the water addition function in Coot, and suggested positions were manually verified in Coot following a refinement cycle in


dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002−9013

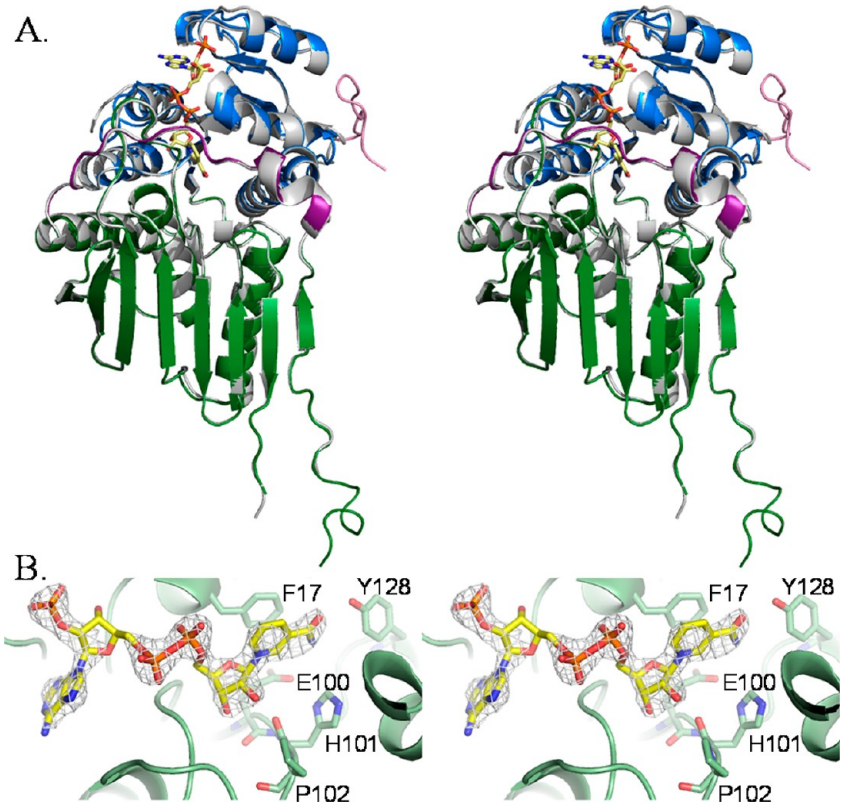



**Figure 3.** $NADP^+$ binding. (A) Superposition of the native Irp3 structure (gray) with the $NADP^+$-bound structure. The domains of the $NADP^+$-bound structure are colored as in Figure 2, with the addition of the pink N-terminal tail. $NADP^+$ is depicted as yellow sticks. The sulfate ion bound in the $NADP^+$ binding site of the native structure is also shown as sticks. The A monomers of both structures are overlaid. (B) $F_o - F_c$ difference density contoured at $3\sigma$ (gray cages) for $NADP^+$. Phenylalanine 17 is part of the glycine rich $^{14}$GAKFGE$^{19}$ sequence for nucleotide binding. The functionally important $^{100}$EHP$^{102}$ sequence is also shown, as is tyrosine 128. Oxygen atoms are colored red, nitrogen atoms dark blue, and phosphorus atoms orange. Carbon atoms are colored as per the molecule: light green for Irp3 and yellow for $NADP^+$.

Refmac. Data collection and refinement statistics are listed in Table 2.

**Native Irp3 Crystallographic Model.** The final native Irp3 model contains four monomers per asymmetric unit. The ordered residues in the structure can be found in Table 3. The structure also includes four sulfate ions, one HEPES molecule, and two ethylene glycol molecules from the crystallization conditions and 498 water molecules. Ramachandran analysis as calculated by RAMPAGE[17] showed good geometry (Table 2). The amino acid in the outlier region was glycine 276 in monomers A–C.

**$NADP^+$-Bound Irp3 Crystallographic Model.** The final native Irp3 model contains four monomers per asymmetric unit. The ordered residues in the structure can be found in Table 3. The structure also includes one $NADP^+$ in each active site and 110 water molecules. Ramachandran analysis as calculated by RAMPAGE[17] showed good geometry (Table 2). The amino acids in the outlier region were glycine 276 in all monomers and proline 2 in monomer A.

**Structural Analysis.** Root-mean-square deviations (rmsd) were calculated using the LSQMAN program suite.[18] Structural comparisons (including rmsd calculations therein) were conducted using PDBeFold,[19] and protein interaction interfaces were analyzed using PDBePISA,[20] both from the European Bioinformatics Institute server (http://www.ebi.ac.uk/pdbe/). Protein structure figures were generated using PyMOL.[21] The atomic coordinates and structure factors (entries 4GMF and 4GMG) have been deposited in the Protein Data Bank (PDB) (Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ).

**Homology Model of PchG.** Using monomer A of the Irp3 native structure, a model of the *P. aeruginosa* thiazolinyl imine reductase, PchG, was built with the SWISS-MODEL fully automated protein structure homology modeling server (http://swissmodel.expasy.org).[22,23]

## RESULTS

**Preparation of Irp3.** Irp3 protein was cloned into both the pET28b and pET29b vectors and overproduced in *E. coli* BL21(DE3) to produce protein with either an N- or C-terminal histidine tag (N-his$_6$-Irp3 or Irp3-his$_6$-C, respectively). Protein purification was completed in two steps using nickel chelating chromatography followed by gel filtration chromatography. The ~40 kDa protein eluted from the gel filtration column at a molecular mass consistent with an 80 kDa protein, indicating dimer formation in solution. This purification protocol yielded 100 mg of N-his$_6$-Irp3/L of culture and 24 mg of Irp3-his$_6$-C/L of culture at a purity estimated to be >90% by polyacrylamide gel electrophoresis. N-his$_6$-Irp3 was also produced in a selenomethionine-substituted form, with a yield of 24 mg/L of culture. The SeMet protein was purified using a similar protocol, with the addition of a reducing agent to all buffers. All three forms of the protein were crystallized using PEG 3350 as the precipitant, but with differing buffers and salt additives. Irp3-his$_6$-C and SeMet N-his$_6$-Irp3 were crystallized without


dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002–9013

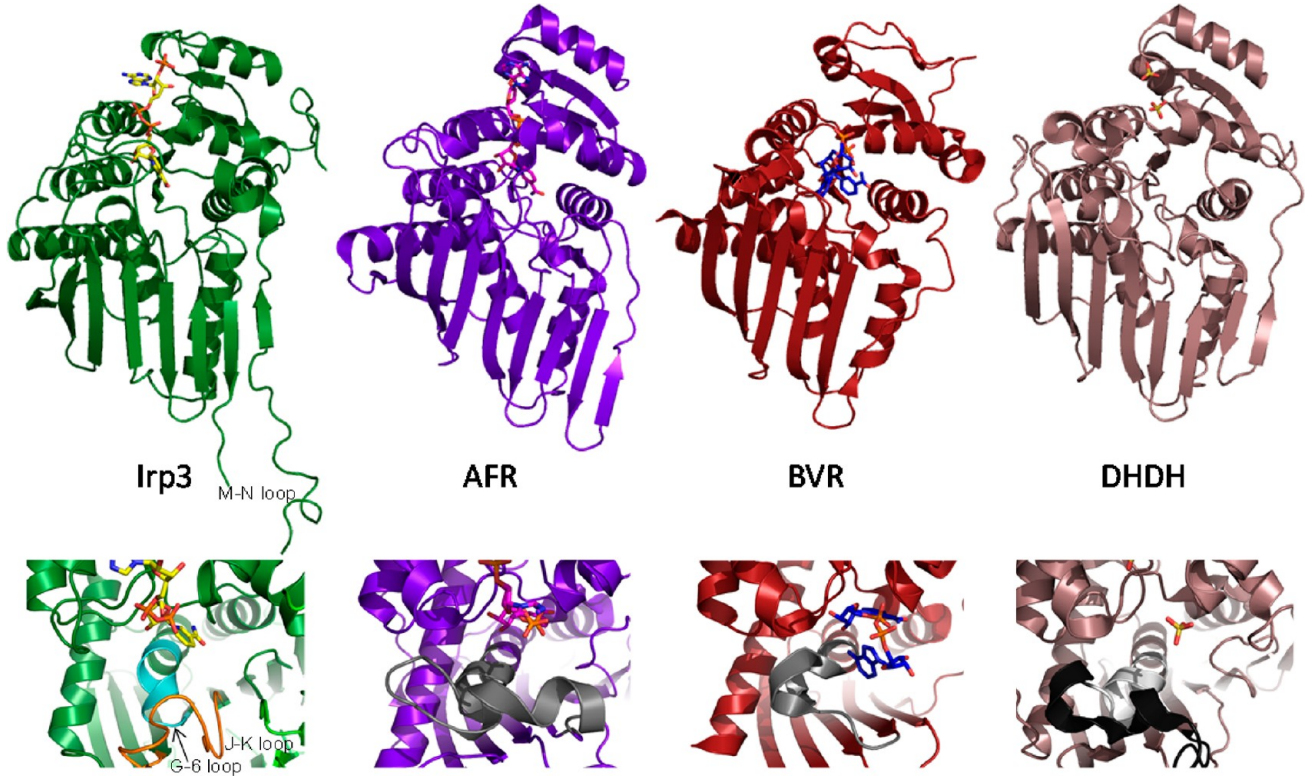



**Figure 4.** Cartoon representations of the NADP$^+$-bound Irp3, 1,5-anhydro-D-fructoase reductase (AFR, PDB entry 2GLX), biliverdin reductase (BVR, PDB entry 1LC3), and dimeric dihyrodiol dehydrogenase (DHDH, PDB entry 2O48). The top row shows the overall fold in cartoon and NADP$^+$ (or sulfates in DHDH) as sticks. It should be noted that NADP$^+$ binds in a similarly extended conformation in Irp3 and AFR. In BVR, NADP$^+$ is folded in a U-shaped conformation that is hypothesized to be the result of crystal packing.[35] The DHDH structure does not contain NADP$^+$ but has sulfates bound in locations similar to those at which the phosphates of NADP$^+$ are expected to bind.[30] In Irp3, the M−N loop, which has 6−15 residues disordered depending on the monomer, is labeled. The bottom row shows close-ups of the putative substrate binding region. In DHDH, the region of the $GGX_3DX_3(Y/H)$ consensus sequence is colored light gray and the remainder of the loop containing this sequence is colored black. The corresponding loops in AFR and BVR are also colored gray. In Irp3, the much shorter loop connecting strand G and helix 6 is colored cyan. The extended loop connecting strands J and K that occupies the corresponding space in three dimensions is colored orange.

ligands. N-his$_6$-Irp3 without selenomethionine substitution was crystallized in the presence of NADP$^+$.

**Overall Architecture of Irp3.** Initial phase estimates for N-his$_6$-Irp3 were determined by multiwavelength anomalous dispersion phasing using a selenomethionine-substituted form of the protein (Table 1). The SeMet model was used as a starting point for model building and structure refinement with the 1.85 Å isomorphous native (Irp3-his$_6$-C) data set. The high-resolution native model was used for molecular replacement to determine the NADP$^+$-bound (N-his$_6$-Irp3) structure to 2.31 Å (Table 2). In both structures, the asymmetric unit contains four monomers of Irp3, arranged as two dimers (Figure 2A,B). For each dimer, the two monomers are related by a noncrystallographic 2-fold axis, and the two dimers are related by a translation roughly along this axis.

**Monomer Architecture.** Each monomer is composed of two domains. The N-terminal domain is a Rossmann fold[24] comprising residues 7−125 with a helix donated from the C-terminal domain (residues 285−310) packing against the central six-stranded parallel $\beta$-sheet (Figure 2B). The faces of the $\beta$-sheet are packed with two or three $\alpha$-helices. The C-terminal domain (residues 126−284 and 311−337) is an eight-stranded mixed (nearly antiparallel) $\beta$-sheet with three $\alpha$-helices packed on one side. The N-his$_6$-Irp3 NADP$^+$-bound structure includes electron density for residues 1−6 and eight residues of the N-terminal purification tag (residues −7 to 0) of monomer A. Residues −7 to 1 form an additional crystal contact with monomer C. Residues 339−357 are evident in all four monomers of both the Irp3-his$_6$-C native and N-his$_6$-Irp3 NADP$^+$-bound structures. Residues 338−350 make up a random coil, with the last seven residues (351−357) forming helix 9, which contributes to the dimerization interface. In Irp3, eight additional amino acids are disordered at the C-terminus, as well as one (NADP$^+$-bound structure) or two or three (native structure) internal loops (Table 3). Finally, we note that the previously hypothesized transmembrane region (residues 162−182[6]) comprises helix 6 and the 6-H loop and does not play the role suggested.

**Comparison of Native and NADP$^+$-Bound Structures.** The overall architecture of the native and NADP$^+$-bound structures is quite similar (Figure 3A). The root-mean-square deviations for comparison between monomers of the same structure (native monomer A to native monomer B, for example) range from 0.22 to 0.55 Å for 328−343 C$\alpha$ residues. Comparison of monomers between structures (native monomer A to NADP$^+$-bound monomer A, etc.) yielded values ranging from 0.31 to 0.60 Å for 326−331 C$\alpha$ residues. Binding of NADP$^+$ causes the ordering of the loop connecting strand D and helix 3, which is disordered in all four monomers of the native structure (Table 3).

**Structural Homologues.** A search with the PDBeFold server[19] revealed that the closest structurally related enzymes


dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002−9013

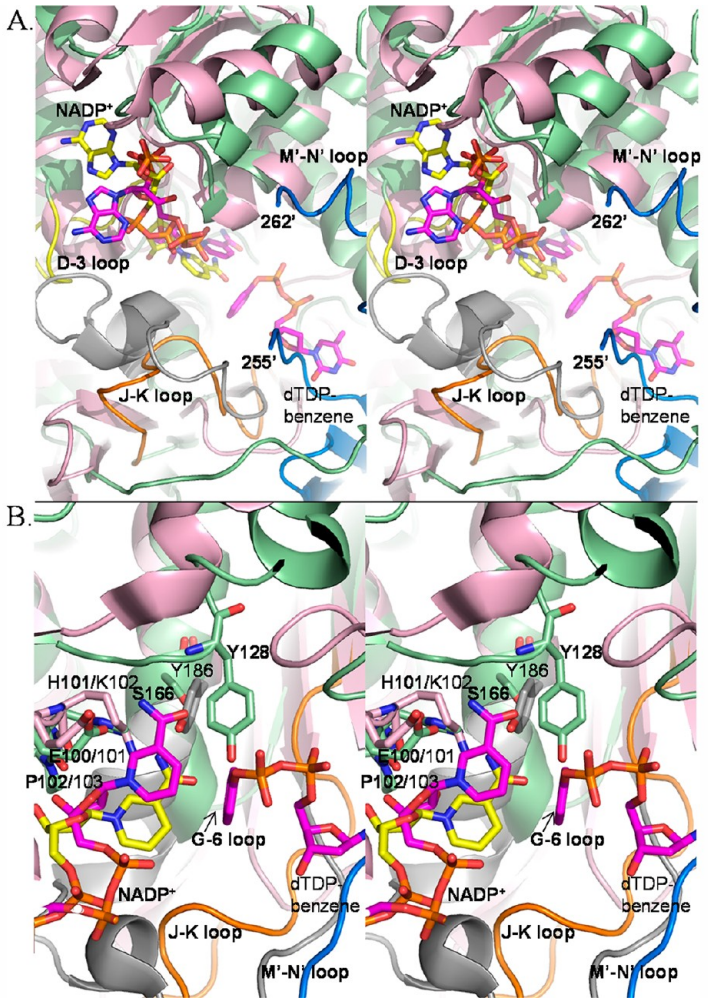



**Figure 5.** Overlay of the Irp3 and KijD10 active sites, highlighting differences in loops (A) and residues (B). The colors are maintained in panels A and B. $NADP^+$-bound Irp3 is shown as a green cartoon. KijD10 (PDB entry 3RC1) is shown as a pink cartoon. The $NADP^+$ molecules bound in the active sites are shown as yellow (Irp3) and magenta (KijD10) sticks. The substrate analogue, dTDP-benzene, for KijD10 is also shown as magenta sticks. KijD10 is a member of the glucose-fructose oxidoreductase family and contains the $GGX_3DX_3(Y/H)$ consensus and following loop shown as the gray cartoon, similar to Figure 4. The D−3 loop of Irp3 is colored yellow. The J−K loop of Irp3 is colored orange. The blue cartoon is for the M′−N′ loop of the opposing dimer, which would be connected by six disordered residues (256−261). (A) The gray loop in KijD10 undergoes a disorder to order transition when the protein goes from the open to the closed conformation, and with this transition, the nicotinamide of the $NADP^+$ flips from *syn* to *anti*.[33] The nicotinamide ring in Irp3 is *anti* as seen in the closed conformation of KijD10 shown here. The loop in Irp3 that undergoes a disorder to order transition upon $NADP^+$ binding is the D−3 loop (yellow). The orange (J−K) loop of Irp3 is in a position comparable to that of the gray loop of KijD10 but does not contain the consensus sequence or connect comparable secondary structure (this is the G−6 loop in Irp3) as discussed in Figure 4. We hypothesize that the very long M′−N′ loop (blue) resides in a different location in the presence of substrate and aids in binding the protein HMWP1 to which the substrate is covalently attached. (B) The EKP sequence of KijD10 and the corresponding EHP of Irp3 are shown as sticks and labeled. The K/H at the center of this loop is hypothesized to be the general acid, which is supported by mutagenic analyses in KijD10.[33] Y186 of the $GGX_3DX_3(Y/H)$ consensus sequence in KijD10 is displayed as sticks to show the location relative to S166, the structurally equivalent amino acid, and Y128, found in the linker between the N-terminal and C-terminal domains (the F−5 loop), both in Irp3. Labels in bold are for Irp3, and those in plain text are for KijD10.

for which a function has been determined belong predominantly to the GFO_IDH_MocA oxidoreductase family and also the dimeric dihydrodiol dehydrogenase. There are many structures that are similar to Irp3 that have been deposited by structural genomics consortia with little annotation, but labeled "oxidoreductase" or "dehydrogenase". The similarity in fold is evident, even by eye (Figure 4). For the top 100 hits, the *Z* scores range from 12.2 to 9.7, with a range in rmsd for these same comparisons of 2.2−2.9 Å aligning 235−265 $\alpha$-carbons with 8−17% identical sequences. More simply, this means that the structure of Irp3 falls into the general structural class of oxidoreductase. This comparison led us to compare the structure Irp3 to the sugar oxidoreductases of known function, with hopes of deducing functional significance. These enzymes include the glucose-fructose oxidoreductase from *Zymomonas mobilis* (GFOR[25,26]), the 1,5-anhydro-D-fructose reductase from *Sinorhizobium morelense* (AFR[27]), monkey dimeric dihydrodiol dehydrogenase (DHDH[28−30]), WlbA from *Bordetella pertussis* and *Chromobacterium violaceum*, which are required to oxidize a C-3′ hydroxyl group of a UDP-linked sugar for formation of 2,3-diacetamido-2,3-dideoxy-D-manuronic acid,[31,32] and KijD10, a ketoreductase in the biosynthetic pathway of L-digitoxose derived from the bacterium *Actinomadura kijaniata*.[33] Irp3 was also compared to biliverdin reductase[34,35] and


dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002−9013

mouse major urinary protein,[36] because Irp3 had previously been hypothesized to be structurally similar to these proteins.[5] Biliverdin reductase is roughly as structurally similar to Irp3 as the oxidoreductases identified by PDBeFold, with an rmsd of 1.88 Å for 199 $\alpha$-carbons. Mouse major urinary protein, which is a $\beta$-barrel, shares no structural similarity with Irp3.

**Dimerization Interface.** The open side of the $\beta$-sheet of the C-terminal domain comprises the most extensive part of the dimerization interface, a common oligomerization motif in this class of enzymes,[25,29,31,33,37] with an additional interaction from the C-terminal tail (Figure 2A,B). Interestingly, GFOR has an N-terminal "arm" that seems to perform a similar role, controlling oligomerization.[26] Analysis via PDBePISA[20] shows that the interface between monomers A and B (dimer 1) and monomers C and D monomers (dimer 2) ranges from 3260 to 3300 Å$^2$, with an average of 3280 Å$^2$. No higher-order oligomers are evident in the structure, consistent with the dimer observed in solution.

**NADP$^+$ Binding.** The N-terminal, NADP$^+$-binding domain of Irp3 is a Rossmann fold with a topology like that found in lactate and malate dehydrogenase.[38] None of the canonical consensus sequences for cofactor binding (for example, GXGXXG) are found.[38−40] Nevertheless, there is a glycine rich sequence $^{14}$GAKFGE$^{19}$ in the first loop of the first $\beta\alpha\beta$ motif to which the NADP$^+$ binds. As is seen in GFOR, an aromatic amino acid stacks with the nicotinamide ring on the side away from the substrate binding site. A tyrosine is observed in GFOR,[25,37] and a phenylalanine (residue 17) is found at this site in Irp3 (Figure 3B). Unlike the GFOR and AFR, Irp3 does not have a kink in the first helix from the inclusion of a proline or glycine in the third turn, which is proposed to influence the helix dipole moment.[27]

The NADP$^+$-bound structure has clear density for the cofactor (Figure 3B). We assume that the cofactor remained oxidized, but the electron density of the nicotinamide ring is insufficient to determine whether the cofactor is oxidized (planar) or became reduced (not planar). The density is sufficient for a supposition of stereochemistry. The nicotinamide ring is bound in the *anti* conformation, which is consistent with the A-side specific oxidoreductases[27,33] and in contrast to the B-side specific biliverdin reductase,[35] suggesting that Irp3 may also be A-side specific. In Irp3, the D−3 loop becomes ordered upon binding of NADP$^+$ (yellow loop in Figure 5A). KijD10 with NADP$^+$ and a substrate analogue bound crystallized in "open" and "closed" conformations in which the nicotinamide ring was *syn* and *anti*, respectively. An extended loop in KijD10 (gray in Figure 5) undergoes a disorder to order transition from an open to a closed conformation, which is hypothesized to be responsible for the change in the conformation of the nicotinamide ring (from *syn* to *anti*). Note that the loops that become ordered upon binding of NADP$^+$ in Irp3 and KijD10 are not comparable. The loop that became ordered in KijD10 is comparable to the G−6 loop of Irp3. The adenine ring is bound in the *anti* conformation in Irp3. In the determined structures of the oxidoreductases,[27,31−33] the conformation of the adenine is varied, with both *anti* and *syn* observed.

In GFOR, AFR, and DHDH, there is a second consensus sequence for cofactor binding. In GFOR and AFR, this sequence is $GX_2VXCEKP$,[25,27] and in DHDH, it is $HX_2HX_{16}EKP$,[28−30] with the last residues being most important (EKP). The glutamic acid hydrogen bonds to the carboxamide of the nicotinamide ring. The proline residue is in the *cis* configuration, which allows for positioning of the lysine residue in the active site to form a cation−$\pi$ interaction with the nicotinamide ring.[27] In WlbA and KijD10, the EKP sequence is conserved, although the remainder of the longer consensus sequence is not found,[31,33] whereas biliverdin reductase has an EYP sequence at this site.[35] Irp3 has $^{100}$EHP$^{102}$ in the comparable loop (Figures 3B and 5B), and the glutamic acid is H-bonded to the carboxamide of the nicotinamide ring; the proline is *cis*, and the histidine is 3.5−4.0 Å from the carboxamide of the nicotinamide ring.

**Substrate Binding.** A consensus sequence for substrate binding has been hypothesized for GFOR, AFR, and DHDH, $GGX_3DX_3(Y/H)$,[28,29] originally proposed from sequence alignment studies of DHDH.[41] This sequence is found at the C-terminus of a 28-residue loop and includes the two N-terminal turns of an active site $\alpha$-helix (Figure 4, bottom row). In GFOR,[37] KijD10,[33] and WlbA from *B. pertussis*,[31] which also have the $GGX_3DX_3(Y/H)$ substrate binding consensus sequence (gray in Figure 5 for KijD10), the aspartic acid is hypothesized to be important for substrate binding. The final amino acid of the sequence has been hypothesized to serve as the general acid/base for catalysis. Biliverdin reductase does not have the consensus sequence, but the comparable loop or helix is extensive (19 residues in all) and similarly located. Irp3 does not have the consensus sequence or the loop. Instead, the comparable loop connecting strand G to helix 6 is five residues long (residues 157−162), and the helix is a turn shorter, effectively removing this potential substrate specificity determinant (cyan in Figure 4). BVR and Irp3 have a serine (residue 166 in Irp3) at the site structurally equivalent to the (Y/H) of the consensus sequence. Irp3 has a tyrosine (residue 128) derived from the linker between the N- and C-terminal domains (F−5 loop) that is in a similar three-dimensional location and hydrogen bonded to the carboxamide of the nicotinamide of NADP$^+$ (Figure 5B). The orange loop of Irp3 in Figures 4 and 5 connects strands J and K, occupies the space filled in the other proteins by the comparable G−6 loop, and may provide the specificity determinants for substrate binding. Interestingly, the putative substrate binding site in Irp3 is filled in part by the very long loop (residues 250−277) that connects strands M and N from the opposing monomer (blue M′−N′ loop in Figure 5). A portion of the M−N loop is disordered in the structure, the number of disordered residues depending on the monomer (Table 3). This very long loop is unique to Irp3 compared to the others of this structural class, as seen best in Figure 4 (top row). The sugar oxidoreductases have a short turn connecting the strands comparable to strands M and N. Indeed, BVR does not have strands comparable to strands M and N.

**Homology Model of PchG.** Using the fully automated protein structure homology modeling server SWISS-MODEL (http://swissmodel.expasy.org),[22,23] and the native Irp3 structure as a template, a model of the thiazolinyl imine reductase involved in pyochelin production from *P. aeruginosa* (PchG) was generated (Figure 6). The level of sequence identity between the proteins is 26%, with Irp3 comprising 365 residues and PchG containing 349 residues. The PchG model includes residues 4−336, with the majority of the unmodeled residues at the C-terminus. The QMEAN score for the PchG model of −0.757 indicates that the model is of good quality,[42] and as expected, the overlay of the PchG model with the determined structure of Irp3 shows a marked degree of similarity [rmsd of 0.464 Å for 315 $\alpha$-carbons (Figure 6)],

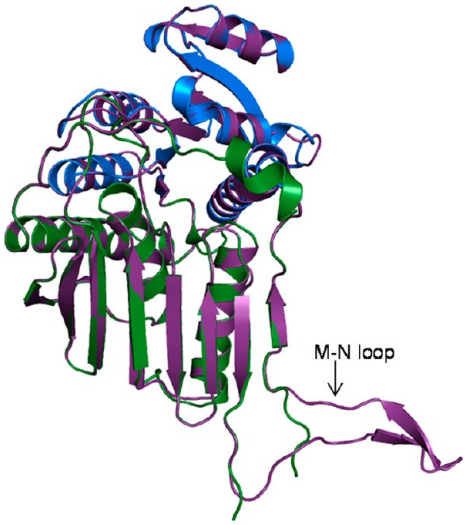



**Figure 6.** Homology model of PchG from *P. aeruginosa* generated with SWISS-MODEL.[22,23] The N-terminal, NADP(H) binding domain of the native Irp3 structure (residues 1−125 and 285−310) is colored blue. The C-terminal substrate binding domain is colored green (residues 126−284 and 311−338). The PchG homology model is shown as a purple cartoon. The loop connecting strands M and N, which is partially disordered in the Irp3 structure, contains two additional $\beta$-strands in the PchG model.

including a high degree of conservation of active site residues. The notable difference is in the loop connecting strands M and N, which is partially disordered in the Irp3 structure but built in the PchG model and includes two additional short $\beta$-strands.

## DISCUSSION

**Dimerization May Promote Binding of the Substrate.** This work provides the first structures of a thiazolinyl imine reductase important in siderophore biosynthesis, including an apo Irp3 crystal structure, Irp3 with $NADP^+$ bound, and a homology model for PchG generated from the Irp3 structures. An N-terminal Rossmann fold nucleotide binding domain is followed by a C-terminal substrate binding domain, with the active site cleft lying between the two domains. The closest structural homologues are sugar oxidoreductases, along with dihydrodiol dehydrogenase and biliverdin reductase. The structural homologues range in oligomerization state from monomeric to octameric. In solution and in the crystal, Irp3 is dimeric, using the open face of the $\beta$-sheet in the C-terminal domain as the most extensive dimeric interface (Figure 2). The question of dimerization in this structural class has been tested in DHDH, where dimer interface mutations were generated and shown to result in active but temperature sensitive monomers, suggesting dimerization promotes stability and thus activity.[28] This may also be true in Irp3.

The substrate for Irp3 (depicted in Figure 1) is covalently bound to the peptidyl carrier domain of the very large HMWP1 (350 kDa), consisting of both polyketide synthase and nonribosomal peptide synthetase modules.[8,43] The substrate is bound to the peptidyl carrier domain of the NRPS module.[8] Neither the structure nor the oligomerization state of HMWP1 is documented; however, there is evidence of oligomerization of polyketide synthases,[44−46] and the predominance of evidence supports monomers for nonribosomal peptide synthetases.[47−49] Binding of dimeric Irp3 to HMWP1 (monomeric or dimeric) allows for insertion of the substrate covalently attached to the peptidyl carrier domain of HMWP1 into the active site of Irp3. This binding, which is a catalytic interaction and is most likely transient, may be facilitated by two loops on Irp3. The M−N loop of Irp3 is very long compared to the comparable loop in homologous proteins (~25 residues longer). This loop from one monomer projects into the active site of the opposing monomer (Figure 5), with some residues disordered. While the precise placement of the loop may be crystallographic, dimerization orients the M−N loop for participation in the binding of HMWP1. If this is true, the C-terminal tail may also be involved in binding to HMWP1, because of the proximity to the M−N loop, instead of serving an additional role in Irp3 oligomerization. These extended loops (the M−N loop and the C-terminal tail) are conserved in PchG (homology model depicted in Figure 6). The potential functional role may be conserved as well, with the comparable loops in PchG promoting interaction with PchF, the NRPS to which the PchG substrate is covalently attached. Finally, the G−6 loop is significantly shorter in the thiazolinyl imine reductases relative to the sugar oxidoreductases, preventing this loop from involvement in binding of the substrate. Instead, the J−K loop is extended and occupies a similar location in three dimensions, possibly providing binding residues for the siderophore biosynthetic intermediate (Figures 4 and 5).

**Implications for Catalysis.** A mechanism has evolved in the literature for the sugar oxidoreductases that includes a general acid/base residue for donation or abstraction of a proton from the substrate and donation of a hydride to or from the nicotinamide cofactor to the substrate, depending on whether oxidation or reduction of a carbon−oxygen bond is being undertaken.[25,27,31,33] A similar mechanism is proposed for reducing the carbon−carbon bond in biliverdin, with a proton donated from a tyrosine general acid and a hydride donated from C4 of the nicotinamide.[35] The residue hypothesized to serve as the general acid/base has been hypothesized to be (1) the middle residue of the EKP consensus sequence[31,33,35] (tyrosine in BVR) or (2) the final residue of the $GGX_3DX_3(Y/H)$ consensus sequence.[25,27−29] These hypotheses have been tested by mutagenesis. In AFR, the K94G variant (EKP) and the H180A variant [$GGX_3DX_3(Y/H)$] produced similar changes in kinetic parameters, leading the authors to suggest that these residues were important in substrate binding and catalysis.[27] In KijD10, a battery of K102 variants (EKP) had either no activity or, in the case of K102E, a reduction of 4 orders of magnitude in $k_{cat}/K_m$, whereas the Y186F variant [$GGX_3DX_3(Y/H)$] showed mostly unchanged kinetic parameters. The authors propose the lysine of EKP to be the general acid/base with the glutamic acid of the K102E variant serving the role ineffectively.[33] In BVR, the tyrosine that corresponds to the lysine of EKP has been mutated to phenylalanine, producing the Y97F variant, which showed 50% of the activity of the wild type. The authors conclude that Y97 is the general acid, but that the transfer of the hydride from NADPH was the dominant process in catalysis.[35] BVR does not contain the $GGX_3DX_3(Y/H)$ consensus sequence, and neither do Irp3 and PchG. However, there is a structurally equivalent residue at the (Y/H) site, which is serine (a possible general acid) in BVR and Irp3 and alanine (incapable of serving as a general acid) in PchG. Taken together, the middle residue at the EKP sequence is more likely the general acid/base, especially considering the following conserved *cis*-proline for the orientation of the acid/base in the proper conformation. Therefore, by analogy, histidine 101 of Irp3 (EHP) is the general acid required for


dx.doi.org/10.1021/bi3011016 | *Biochemistry* 2012, 51, 9002−9013

catalysis, which is conserved in PchG (histidine 97). However, we cannot rule out the possibility that tyrosine 128 of Irp3 that is derived from the loop connecting strand F on the N-terminal domain to helix 5 of the C-terminal domain that is in a location similar to that of the final residue of the $GGX_3DX_3(Y/H)$ sequence in the sugar oxidoreductases may also serve as a general acid (Figure 5B). This residue is conserved in PchG (tyrosine 124), but not in the other proteins discussed here.

**Conclusions.** Thiazolinyl imine reductases (Irp3 and PchG) are dimeric enzymes that are structurally similar to sugar oxidoreductase such as GFOR, AFR, KijD10, and WlbA and to biliverdin reductase. Irp3 and PchG may interact with the nonribosomal peptide synthetase modules to which their substrates are attached using loops that are unique from their structural homologues. The reduction of the C−N double bond of the substrate thiazoline ring is hypothesized to be conducted by donation of a proton from a general acid, most likely histidine 101 or alternatively tyrosine 128, and transfer of a hydride from C4 of the nicotinamide of the NADPH cofactor.

## AUTHOR INFORMATION


### Corresponding Author

*Phone: (785) 864-5075. Fax: (785) 864-5294. E-mail: lamb@ku.edu.



### Funding

This publication was made possible by National Institutes of Health (NIH) Grants R01 AI77725 and K02 AI093675 from the National Institute for Allergy and Infectious Disease (A.L.L.) and by the Graduate Training Program in Dynamic Aspects of Chemical Biology via NIH Grant T32 GM08545 (K.M.M.).

### Notes

The authors declare no competing financial interest.


## ACKNOWLEDGMENTS


We are grateful to A. S. Chilton for technical assistance. Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. We thank the staff of the SSRL for their support and assistance.


## ABBREVIATIONS

AFR, 1,5-anhydro-D-fructose reductase from *S. morelense*; BVR, rat biliverdin reductase; DHDH, monkey dimeric dihydrodiol dehydrogenase; GFOR, glucose-fructose oxidoreductase from *Z. mobilis*; HMWP1, high-molecular weight protein 1, a nonribosomal peptide synthetase from *Y. enterocolitica*; Irp3, thiazolinyl imine reductase from *Y. enterocolitica*; Irp3-$his_6$-C, Irp3 with a C-terminal histidine purification tag; KijD10, ketoreductase from *A. kijaniata*; N-$his_6$-Irp3, Irp3 with an N-terminal histidine purification tag; NRPS, nonribosomal peptide synthetase; Ppant, phosphopantetheine; PchD, salicylate adenylase from *P. aeruginosa*; PchE, nonribosomal peptide synthetase from *P. aeruginosa*; PchF, nonribosomal peptide synthetase from *P. aeruginosa*; PchG, thiazolinyl imine reductase from *P. aeruginosa*; rmsd, root-mean-square deviation; SeMet, selenomethionine; WlbA, from *B. pertussis*, *C. violaceum*, and *P. aeruginosa*, enzymes involved in 2,3-diacetamido-2,3-dideoxy-D-manuronic acid production; YbtU, thiazolinyl imine reductase from *Y. pestis*.

## REFERENCES


(1) Vasil, M. L., and Ochsner, U. A. (1999) The response of *Pseudomonas aeruginosa* to iron: Genetics, biochemistry and virulence. *Mol. Microbiol. 34*, 399−413.

(2) Cornelis, P., and Matthijs, S. (2002) Diversity of siderophore-mediated iron uptake systems in fluorescent pseudomonads: Not only pyoverdines. *Environ. Microbiol. 4*, 787−798.

(3) Crosa, J. H., and Walsh, C. T. (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol. Mol. Biol. Rev. 66*, 223−249.

(4) Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev. 97*, 2651−2674.

(5) Patel, H. M., and Walsh, C. T. (2001) In vitro reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin: Characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. *Biochemistry 40*, 9023−9031.

(6) Geoffroy, V. A., Fetherston, J. D., and Perry, R. D. (2000) *Yersinia pestis* YbtU and YbtT are involved in synthesis of the siderophore yersiniabactin but have different effects on regulation. *Infect. Immun. 68*, 4452−4461.

(7) Reimmann, C., Patel, H. M., Serino, L., Barone, M., Walsh, C. T., and Haas, D. (2001) Essential PchG-dependent reduction in pyochelin biosynthesis of *Pseudomonas aeruginosa*. *J. Bacteriol. 183*, 813−820.

(8) Miller, D. A., Luo, L., Hillson, N., Keating, T. A., and Walsh, C. T. (2002) Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem. Biol. 9*, 333−344.

(9) 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.

(10) Kabsch, W. (2010) XDS. *Acta Crystallogr. D66*, 125−132.

(11) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D66*, 213−221.

(12) Leslie, A. G. W., and Powell, H. R. (2007) Processing diffraction data with Mosflm. *Evolving Methods for Macromolecular Crystallography 245*, 41−51.

(13) Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D50*, 760−763.

(14) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126−2132.

(15) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240−255.

(16) McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D61*, 458−464.

(17) Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by C$\alpha$ geometry: $\phi$, $\psi$ and C$\beta$ deviation. *Proteins 50*, 437−450.

(18) Kleywegt, G. J., and Jones, T. A. (1997) Detecting folding motifs and similarities in protein structures. *Methods Enzymol. 277*, 525−545.

(19) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D60*, 2256−2268.

(20) Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774−797.
(21) DeLano, W. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.
(22) Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics 22*, 195−201.
(23) Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res. 37*, D387−D392.
(24) Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G. (1974) Structure determination of crystalline lobster D-glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol. 82*, 563−585.
(25) Kingston, R. L., Scopes, R. K., and Baker, E. N. (1996) The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: An osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure 4*, 1413−1428.
(26) Lott, J. S., Halbig, D., Baker, H. M., Hardman, M. J., Sprenger, G. A., and Baker, E. N. (2000) Crystal structure of a truncated mutant of glucose-fructose oxidoreductase shows that an N-terminal arm controls tetramer formation. *J. Mol. Biol. 304*, 575−584.
(27) Dambe, T. R., Kuhn, A. M., Brossette, T., Giffhorn, F., and Scheidig, A. J. (2006) Crystal structure of NADP(H)-dependent 1,5-anhydro-D-fructose reductase from *Sinorhizobium morelense* at 2.2 Å resolution: Construction of a NADH-accepting mutant and its application in rare sugar synthesis. *Biochemistry 45*, 10030−10042.
(28) Carbone, V., Hara, A., and El-Kabbani, O. (2008) Structural and functional features of dimeric dihydrodiol dehydrogenase. *Cell. Mol. Life Sci. 65*, 1464−1474.
(29) Carbone, V., Endo, S., Sumii, R., Chung, R. P., Matsunaga, T., Hara, A., and El-Kabbani, O. (2008) Structures of dimeric dihydrodiol dehydrogenase apoenzyme and inhibitor complex: Probing the subunit interface with site-directed mutagenesis. *Proteins 70*, 176−187.
(30) Carbone, V., Sumii, R., Ishikura, S., Asada, Y., Hara, A., and El-Kabbani, O. (2008) Structure of monkey dimeric dihydrodiol dehydrogenase in complex with isoascorbic acid. *Acta Crystallogr. D64*, 532−542.
(31) Thoden, J. B., and Holden, H. M. (2011) Biochemical and structural characterization of WlbA from *Bordetella pertussis* and *Chromobacterium violaceum*: Enzymes required for the biosynthesis of 2,3-diacetamido-2,3-dideoxy-D-mannuronic acid. *Biochemistry 50*, 1483−1491.
(32) Thoden, J. B., and Holden, H. M. (2010) Structural and functional studies of WlbA: A dehydrogenase involved in the biosynthesis of 2,3-diacetamido-2,3-dideoxy-D-mannuronic acid. *Biochemistry 49*, 7939−7948.
(33) Kubiak, R. L., and Holden, H. M. (2011) Combined structural and functional investigation of a C-3″-ketoreductase involved in the biosynthesis of dTDP-L-digitoxose. *Biochemistry 50*, 5905−5917.
(34) Kikuchi, A., Park, S. Y., Miyatake, H., Sun, D., Sato, M., Yoshida, T., and Shiro, Y. (2001) Crystal structure of rat biliverdin reductase. *Nat. Struct. Biol. 8*, 221−225.
(35) Whitby, F. G., Phillips, J. D., Hill, C. P., McCoubrey, W., and Maines, M. D. (2002) Crystal structure of a biliverdin IX$\alpha$ reductase enzyme-cofactor complex. *J. Mol. Biol. 319*, 1199−1210.
(36) Perez-Miller, S., Zou, Q., Novotny, M. V., and Hurley, T. D. (2010) High resolution X-ray structures of mouse major urinary protein nasal isoform in complex with pheromones. *Protein Sci. 19*, 1469−1479.
(37) Nurizzo, D., Halbig, D., Sprenger, G. A., and Baker, E. N. (2001) Crystal structures of the precursor form of glucose-fructose oxidoreductase from *Zymomonas mobilis* and its complexes with bound ligands. *Biochemistry 40*, 13857−13867.
(38) Bellamacina, C. R. (1996) The nicotinamide dinucleotide binding motif: A comparison of nucleotide binding proteins. *FASEB J. 10*, 1257−1269.
(39) Hanukoglu, I., and Gutfinger, T. (1989) cDNA sequence of adrenodoxin reductase. Identification of NADP-binding sites in oxidoreductases. *Eur. J. Biochem. 180*, 479−484.
(40) 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.
(41) Arimitsu, E., Aoki, S., Ishikura, S., Nakanishi, K., Matsuura, K., and Hara, A. (1999) Cloning and sequencing of the cDNA species for mammalian dimeric dihydrodiol dehydrogenases. *Biochem. J. 342* (Part 3), 721−728.
(42) Benkert, P., Biasini, M., and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics 27*, 343−350.
(43) Suo, Z., Tseng, C. C., and Walsh, C. T. (2001) Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase. *Proc. Natl. Acad. Sci. U.S.A. 98*, 99−104.
(44) Leibundgut, M., Maier, T., Jenni, S., and Ban, N. (2008) The multienzyme architecture of eukaryotic fatty acid synthases. *Curr. Opin. Struct. Biol. 18*, 714−725.
(45) Maier, T., Leibundgut, M., and Ban, N. (2008) The crystal structure of a mammalian fatty acid synthase. *Science 321*, 1315−1322.
(46) Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E., and Khosla, C. (2006) The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U.S.A. 103*, 11124−11129.
(47) Mitchell, C. A., Shi, C., Aldrich, C. C., and Gulick, A. M. (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry 51*, 3252−3263.
(48) Sieber, S. A., Linne, U., Hillson, N. J., Roche, E., Walsh, C. T., and Marahiel, M. A. (2002) Evidence for a monomeric structure of nonribosomal peptide synthetases. *Chem. Biol. 9*, 997−1008.
(49) Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C., and Gulick, A. M. (2012) Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem. Biol. 19*, 188−198.