Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural insights into the role of *Bacillus subtilis* YwfH (BacG) in tetrahydrotyrosine synthesis

Malligarjunan Rajavel, Kumar Perinbam and B. Gopal*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka 560 012, India

Correspondence e-mail: bgopal@mbu.iisc.ernet.in

Received 17 September 2012
Accepted 12 November 2012

PDB References: *Bacillus subtilis* YwfH, 3u49; 3u4c; 3u4d

The synthesis of the dipeptide antibiotic bacilysin involves the sequential action of multiple enzymes in the *bac* operon. YwfH (also referred to as BacG) catalyzes the stereoselective reduction of dihydro-hydroxyphenylpyruvate (H_2HPP) to tetrahydro-hydroxyphenylpyruvate (H_4HPP) in this biosynthetic pathway. YwfH is an NADPH-dependent reductase that facilitates the conjugate addition of a hydride at the C^4 olefin terminus of H_2HPP. Here, the structure of YwfH is described at three conformational steps: the apo form, an apo-like conformation and the NADPH complex. YwfH is structurally similar to other characterized short-chain dehydrogenase/reductases despite having marginal sequence similarity. The structures of YwfH in different conformational states provide a rationale for the ping-pong reaction mechanism. The identification and role of the residues in the catalytic tetrad (Lys113–Tyr117–Ser155–Asn158) in proton transfer were examined by mutational analysis. Together, the structures and biochemical features revealed synchronized conformational changes that facilitate cofactor specificity and catalysis of H_4HPP formation en route to tetrahydrotyrosine synthesis.

1. Introduction

The nonribosomal peptide synthesis of bacilysin involves multiple enzymes encoded by genes in the *bac* operon as well as the adjacent *ywfGH* cluster (Inaoka *et al.*, 2003; Mahlstedt & Walsh, 2010). Bioinformatics analysis has suggested putative functional roles for individual proteins in this biosynthetic pathway; however, subsequent experimental characterization revealed several variations from the proposed biosynthetic mechanism (Mahlstedt & Walsh, 2010; Fig. 1). The first enzymatic step in the synthesis of bacilysin involves the prephenate dehydratase/dehydrogenase activity of BacA (Kleeb *et al.*, 2006; Hermes *et al.*, 1984). However, experimental studies revealed that BacA catalyzes the enantioselective decarboxylation and isomerization of prephenate (Rajavel *et al.*, 2009). BacB is involved in the second catalytic step. The characterization of the product of BacB suggested that this enzyme could function as an isomerase, resulting in the intermediate dihydro-hydroxyphenylpyruvate (H_2HPP), en route to L-anticapsin biosynthesis (Mahlstedt & Walsh, 2010). Indeed, both the BacA and the BacB products have been shown to be regioisomers of dihydro-4-hydroxyphenylpyruvate (Parker & Walsh, 2012). While the role of BacC is as yet unclear, BacD has been demonstrated to be a ligase that couples anticapsin with L-alanine to form bacilysin (Steinborn

et al., 2005). The observation that the catalytic activities of the Bac enzymes were insufficient to achieve bacilysin synthesis prompted an examination of the role of YwfH and YwfG in this process. In these studies, YwfH was noted to catalyze the conversion of H_2HPP (*exocyclic*-Δ^3,Δ^5-dihydro-4-hydroxyphenylpyruvate) to H_4HPP (tetrahydro-4-hydroxyphenylpyruvate), whereas YwfG catalyzes the transamination of the ketoacid to tetrahydrotyrosine (H_4Tyr) (Mahlstedt & Walsh, 2010). The nonproteogenic amino acid H_4Tyr is a well characterized component of cyanobacterial peptides in *Aeropyrum pernix* (Yamamura *et al.*, 2008); however, this amino acid has not been extensively characterized as part of the *B. subtilis* metabolome. The existence of these enzymes in *B. subtilis* suggests that the so-called *bac* gene cluster could also participate in the synthesis of these alternative nonribosomal peptides.

YwfH belongs to the short-chain dehydrogenase/reductase (SDR) family of enzymes. In general, the enzymes of the SDR family are NAD(P)H/NAD(P)$^+$-dependent oxidoreductases. Although SDR enzymes share low sequence similarity

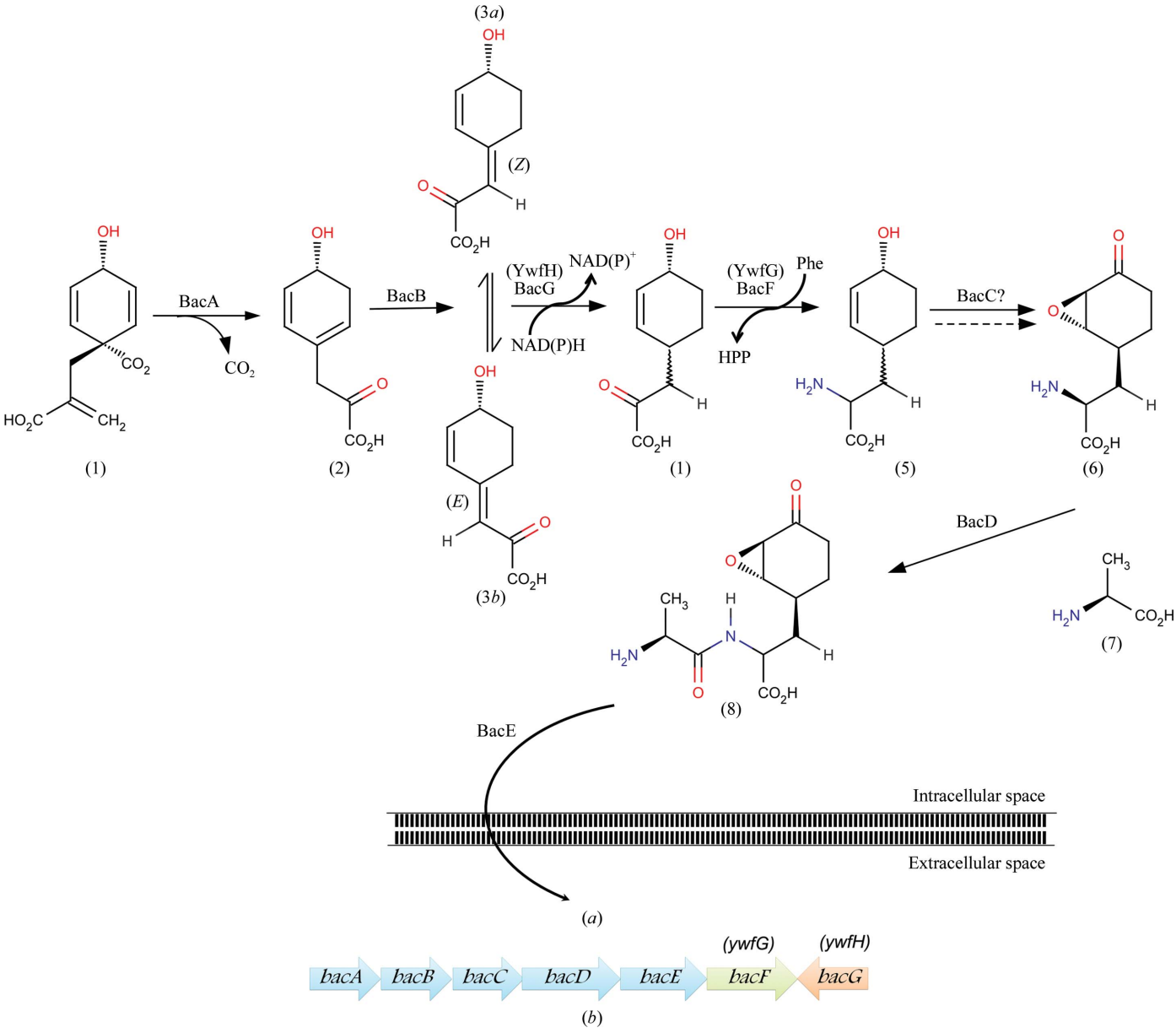

Figure 1
(*a*) Schematic representation of the chemical reactions performed by the enzymes encoded by the genes in the *bac* operon. Prephenate (1) is converted to *en*-H_2HPP (2) by BacA (prephenate decarboxylase; Rajavel *et al.*, 2009). *Z*-*ex*-H_2HPP (3*a*) and *E*-*ex*-H_2HPP (3*b*) isomers are synthesized by the action of BacB (YwfC; Parker & Walsh, 2012). The enzyme YwfH (BacG) is a short-chain reductase which converts *E*/*Z*-*ex*-H_2HPP to H_4HPP (4). This product serves as the substrate for BacF, which in turn produces H_4Tyr (5) (Mahlstedt & Walsh, 2010). The *in vitro* synthesis of L-anticapsin (6) remains unclear, as the role of the enzyme BacC remains to be experimentally determined. The L-alanine (7) synthesized from pyruvate is ligated with L-anticapsin by BacD. Bacilysin (8) is exported out of the cell by BacE (Steinborn *et al.*, 2005). (*b*) A representation of the polycistronic *bac* operon containing the genes *bacABCDE*, *bacF* and the monocistronic gene *bacG* (Inaoka *et al.*, 2003; Mahlstedt & Walsh, 2010).

(15–30%), their three-dimensional structures are well conserved and adopt the Rossmann (α/β) fold. Enzymes with the Rossmann fold bind an $NAD^+/NADP^+$ cofactor, primarily involving residues in the TG*xxx*G*x*G sequence motif (Jörnvall *et al.*, 1995; Filling *et al.*, 2002). However, substrate recognition involves considerably many more sequence and structural features that depend primarily on variations in the length of the substrate-binding domain. This in turn defines the specificity for a wide range of substrates such as lipids, amino acids, carbohydrates and hormones (Persson *et al.*, 2009). Although the keto reduction of hydroxyl/carbonyl groups constitutes the major class of enzymatic activity in SDR enzymes, they also catalyze the reduction of —C═C— and —C═N— double bonds (Persson *et al.*, 2009). Few SDR homologues have been shown to bind metal cofactors (Josefsson *et al.*, 1998). The composition as well as the spatial variation in the active-site residues is consistent with the diversity of substrates and reactions catalyzed by enzymes of the SDR family. The active-site residues referred to as a tetrad, most commonly R-S-Y-K or S-Y-K-N, mediate the transfer of a proton through a bound water between Lys and Asn to aid catalysis (Oppermann *et al.*, 2003; Miller *et al.*, 2006).

Here, we describe the structural features of YwfH, which reveal a novel NADPH-dependent —C═C— reductase. The structural and biochemical characteristics provide a basis for understanding both the composition and the spatial orientation of the catalytic residues as well as the features that affect cofactor specificity. The structure also reveals a hitherto unidentified aspect of conformational variation that regulates catalytic activity. We note that upon NADPH binding YwfH undergoes large conformational changes that result in the adoption of an open conformation by the substrate-binding domain. The synchronized conformational rearrangements observed in the crystal structures of YwfH provide a rationale for a ping-pong reaction mechanism.

2. Materials and methods

2.1. Cloning, expression and purification of YwfH and its mutants

The *ywfH* gene was PCR-amplified using the forward and reverse primers 5′-CGTGCTAGCATGTCAAAACGAACC-GCATTTGTT-3′ and 5′-ATTCTCGAGTATGCTTTTCATG-CTGCCGCCGTC-3′ incorporating *Nhe*I and *Xho*I restriction sites, respectively, to facilitate cloning. The PCR-amplified gene was cloned into a modified pET-15b expression vector. The plasmid containing the *ywfH* gene was transformed into *Escherichia coli* BL21 (DE3) cells for overexpression of recombinant YwfH. The cells were grown in Luria broth to an optical density (OD) of 0.6 at 600 nm and induced with 1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was reduced to 293 K after induction and the cells were grown for a further ∼12 h.

Recombinant YwfH with a polyhistidine tag at the N-terminus was partially purified using cobalt–NTA resin (Sigma–Aldrich Inc.) in a buffer consisting of 50 m*M* Tris–HCl, 250 m*M* NaCl pH 7.5 (Supplementary Fig. S1[1]). The protein was further purified by size-exclusion chromatography on a Sephacryl 16/60 S-200 column (GE Healthcare). The purity of the recombinant enzyme was analyzed using SDS–PAGE as well as mass spectrometry (MALDI-TOF; Bruker Daltonics Inc.).

The point mutations (K113A, Y117A, S155A, N158A and S250A) were introduced into *ywfH* using the PCR-based double-primer method. These utilized the pET-15b plasmid containing the wild-type *ywfH* gene as the template. The following primers were used for the point variants: K113A-forward, 5′-GCC ACG TTC ACT CAA GCG ACC GTT GCC TAT ATT-3′, and K113A-reverse, 5′-AAT ATA GGC AAC GGT CGC TTG AGT GAA CGT GGC-3′; Y117A-forward, 5′-CAA AAG ACC GTT GCC GCT ATT GAC GCA ATC-3′, and Y117A-reverse, 5′-GAT TGC GTC AAT AGC GGC AAC GGT CTT TTG-3′; S155A-forward, 5′-ATG TTC ACA AAC GCT ATG ATG AAT GCC GCT-3′, and S155A-reverse, 5′-AGC GGC ATT CAT CAT AGC GTT TGT GAA CAT-3′; N158A-forward, 5′-ACA AAC AGT ATG ATG GCT GCC GCT TTA ATT AAC-3′, and N158A-reverse, 5′-GTT AAT TAA AGC GGC AGC CAT CAT ACT GTT TGT-3′; S250A-forward, 5′-ACG GGA CAG CAA ATT GCC GCT GAC GGC GGC AGC-3′, and S250A-reverse, 5′-GCT GCC GCC GTC AGC GGC AAT TTG CTG TCC CGT-3′. The incorporation of these mutations was confirmed by DNA sequencing. The mutants were transformed in *E. coli* BL21 (DE3) cells and purified using an identical procedure to that adopted for wild-type YwfH.

2.2. Crystallization, diffraction data collection and crystal structure determination

The purified enzyme was concentrated to 20–25 mg ml^{-1} and initial crystallization trials were performed using commercially available crystallization kits (Hampton Research Inc). The conditions were examined using the microbatch method at 293 K, in which the drop consisted of 5 µl protein solution and 5 µl crystallization condition (1:1 ratio). A few conditions containing PEG 3350 resulted in poorly diffracting crystals within a week. A condition consisting of 0.1 *M* HEPES pH 7.5, 1.4 *M* trisodium citrate resulted in a few crystals within three weeks that diffracted to 1.9 Å resolution at the home source (a Rigaku MicroMax-007 HF X-ray generator equipped with a MAR345dtb detector) and to a Bragg spacing of 1.75 Å at the synchrotron (ESRF beamline BM-14). Cocrystallization studies were performed in the presence of NADH and NADPH. Solutions containing NADH and NADPH were directly added to the concentrated protein solution to a final concentration of ∼5 m*M* and were incubated on ice for 1 h prior to crystallization. Interestingly, only the NADPH-containing wells yielded crystals. These crystals diffracted to 2 Å resolution at the home source. The hexagonal-form crystals were incubated with the substrate in

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ5266). Services for accessing this material are described at the back of the journal.

Table 1
Crystallographic statistics for YwfH.

Values in parentheses are for the outer shell.

YwfH	Monoclinic form (apo YwfH)	Hexagonal form (NADPH-bound)	Hexagonal form (apo-like form)
Data collection			
Wavelength (Å)	1.5418	1.5418	1.5418
Resolution (Å)	68.50–1.75 (1.84–1.75)	22.68–2.03 (2.07–2.03)	28.70–2.70 (2.75–2.70)
Unit-cell parameters (Å, °)	a = 64.85, b = 136.99, c = 66.7, β = 117.99	$a = b$ = 69.29, c = 207.44	$a = b$ = 69.0, c = 207.04
Space group	$P2_1$	$P6_222$	$P6_222$
Total No. of reflections	773776 (114627)	334152 (10987)	95305 (1092)
No. of unique reflections	102319 (15067)	19411 (872)	8551 (331)
Wilson B factor (Å^2)	22.39	36.78	65.95
Completeness (%)	99.1 (100.0)	97.0 (90.6)	97.7 (79.4)
Multiplicity	7.6 (7.6)	17.2 (12.6)	11.1 (3.3)
$R_{\rm merge}$† (%)	8.9 (48.7)	6.2 (57.4)	5.6 (24.9)
$\langle I/\sigma(I)\rangle$	13.4 (4.1)	17.7 (3.0)	23.8 (2.9)
Refinement			
$R_{\rm cryst}$‡ (%)	18.8	20.1	21.7
$R_{\rm free}$§ (%)	22.1	24.3	27.4
No. of atoms			
Protein	7598	1922	1789
Water	410	48	16
Ligand	—	47	—
B factors (Å^2)			
Protein	28.15	47.93	81.99
Solvent	38.91	50.04	58.63
Ligand	—	49.64	—
R.m.s. deviations			
Bond lengths (Å)	0.007	0.008	0.008
Bond angles (°)	0.993	1.123	1.254
Ramachandran plot (%)			
Favoured regions	93.3	89.4	87.1
Allowed regions	6.5	10.6	12.4
Outliers	0.1	0	0.5

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith reflection and $\langle I(hkl)\rangle$ is the average intensity. ‡ $R_{\rm cryst} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$. § $R_{\rm free}$ was calculated as for $R_{\rm cryst}$ but using 5% of the data that were excluded from the refinement calculation.

an effort to obtain a ternary complex with H_2HPP. A reaction mixture consisting of BacA, BacB and ∼2–3 mM prephenate was prepared and incubated for 1 h for these experiments. The hexagonal-form crystals were directly transferred to a drop containing the mother liquor and the BacAB reaction product.

The diffraction data were processed using *iMOSFLM* and scaled using *SCALA* in the *CCP*4 suite (Battye *et al.*, 2011; Winn *et al.*, 2011). Phase information was obtained by molecular replacement (MR) using *Phaser* (McCoy *et al.*, 2007). Four molecules of YwfH were located in the asymmetric unit of the crystal in the apo form. Noncrystallographic symmetry (NCS) and translation–libration–screw (TLS) restraints were used during refinement. Iterative cycles of model building and refinement were performed using *Coot* and *REFMAC*5 (Emsley & Cowtan, 2004; Murshudov *et al.*, 2011). The crystal structure of apo YwfH was used as the intial model to obtain phase information for the cofactor-bound YwfH structure. This MR procedure, which was performed using *Phaser*, resulted in a solution in space group $P6_222$ with a log-likelihood gain (LLG) of 1000 and a Z-score of 20. The NADPH cofactor could be modelled in the electron-density map obtained after simulated annealing (Brünger *et al.*, 1998). Multiple programs from the *CCP*4 suite were used at various stages of model building, refinement and analysis (Winn *et al.*, 2011).

2.3. Catalytic activity of YwfH and active-site mutants

The substrate for the enzymatic reaction was prepared by adding 200 µM prephenate, 1 µM BacA and 0.5 µM BacB pre-incubated for 20 min. The time-course spectra were recorded at 340 nm for reactions containing varying concentrations of the cofactor (NADPH). The reduction of NADPH to $NADP^+$ was measured to calculate the kinetic parameters for the YwfH enzyme and its mutants. The UV absorbance spectra were recorded on a V-630Bio Jasco UV–Vis spectrophotometer.

2.4. HPLC and LC-ESI mass-spectrometric analyses of H_2HPP

HPLC and mass-spectrometric analyses were performed to determine the catalytic activity as well as the cofactor specificity of YwfH. HPLC experiments were carried out on a C-18 Zorbax analytical column at a flow rate of 0.5 ml min^{-1} with an acetonitrile gradient varying from 0 to 20% at room temperature. The analyte (H_2HPP) for the HPLC experiments was prepared by incubating 100 µM prephenate (Sigma–Aldrich) with 1 µM BacA and 0.2 µM BacB for 20 min. Following this, 200 µM NADPH and 1 µM YwfH were added to the reaction mixture and incubated for different time intervals varying from 1 to 20 min. The chromatograms were recorded at two different wavelengths (260 and 340 nm). LC-ESI mass-spectrometric analysis was performed using an HCT Ultra PTM Discovery System (ETD II, Bruker Daltonics) coupled to a 1100 series HPLC system (Agilent).

2.5. Docking *E*/*Z*-*ex*-H_2HPP into the active site of YwfH

A total of 256 docking trials for the *E*- and *Z*-*ex*-H_2HPP forms were conducted using the Lamarckian genetic algorithm (LGA) provided by the *AutoDock* program v.4.0 (http://www.autodock.scripps.edu/). Charges of the Gasteiger type were assigned for both *E*- and *Z*-*ex*-H_2HPP. The Ligand Torsion option was used to define the internal degrees of freedom and the torsional restraints for the ligand. The grid dimensions were kept constant for both *E*- and *Z*-*ex*-H_2HPP for ligand docking. Clustering of the ligand-docked structure was performed based on the lowest binding energy. The results were analyzed using *MGLtools* and were visualized using *LIGPLOT* (Wallace *et al.*, 1995; Morris *et al.*, 2009).

3. Results

3.1. Structural snapshots of YwfH at different steps of catalysis

The crystal structure of YwfH was determined in three conformational states. These represent apo YwfH, the YwfH–NADPH complex and an apo-like form. The phase information used to solve the structure of YwfH was obtained by molecular replacement (MR). The diffraction data and refinement statistics are compiled in Table 1. A sequence-based search for a protein with a similar structure in the PDB suggested entry 2z1n, a hypothetical protein from *A. pernix* K1 (Yamamura *et al.*, 2008), as a plausible template for MR (Supplementary Fig. S2). MR using the 2z1n structure (sequence similarity of ∼28%) resulted in a solution with an LLG of ∼352 and a *Z*-score of 6.3. The monomeric subunit of apo YwfH adopts the Rossmann fold with α/β topology. The substrate-binding domain consists of two α-helices (α6 and α8). A significant component of this domain is comprised of flexible loop segments. The monomeric subunit is comprised of residues 2–259. One loop (residues 207–216) could not be modelled in this structure. This polypeptide segment is also less conserved across SDR homologues. The secondary-structure alignment (based on the *DSSP* server) of apo YwfH and the NADPH complex is shown in Fig. 3(*a*) (Kabsch & Sander, 1983; Joosten *et al.*, 2011).

There are four monomers in the asymmetric unit of the apo YwfH crystal. The hexameric crystal form adopted by the YwfH–NADPH complex contained one monomer in the asymmetric unit. YwfH is a dimer in solution in both the apo and the NADPH-bound form. The monomeric subunit mass is ∼31 kDa. YwfH eluted at 13.7 ml from an analytical size-exclusion chromatography column (Superdex 200 10/300 Global, GE Healthcare), corresponding to a mass of ∼61.5 kDa (Supplementary Fig. S3). The buried surface area (BSA) corresponding to the tetrameric arrangement of YwfH was calculated using the *PISA* server. The BSA and $\Delta G^{\rm diss}$ values for a dimer (∼2680 Å^2 and ∼−14 kcal mol^{-1}, respectively) appeared to be more plausible compared with those corresponding to a tetrameric arrangement (∼8980 Å^2 and ∼−16 kcal mol^{-1}, respectively) (Krissinel & Henrick, 2007; Krissinel, 2010).

The crystal structure of the YwfH–NADPH complex was solved in space group $P6_222$ at 2.0 Å resolution using the apo YwfH structure as the starting model. The structure of the cofactor-binding domain in the YwfH–NADPH complex is similar to that of apo YwfH (with an r.m.s.d. of 0.9 Å between the C$^\alpha$ atoms in this domain). The electron-density map could be unambiguously interpreted for the NADPH cofactor (Fig. 2). A superposition of the structure of apo YwfH on that of the YwfH–NADPH complex revealed that cofactor binding induces substantial conformational changes in the substrate-binding domain. Similar to other NAD(P)H-binding SDR enzymes, YwfH also binds the cofactor in an extended conformation. Attempts to cocrystallize the ternary complex of YwfH with H_2HPP and NADPH were unsuccessful,

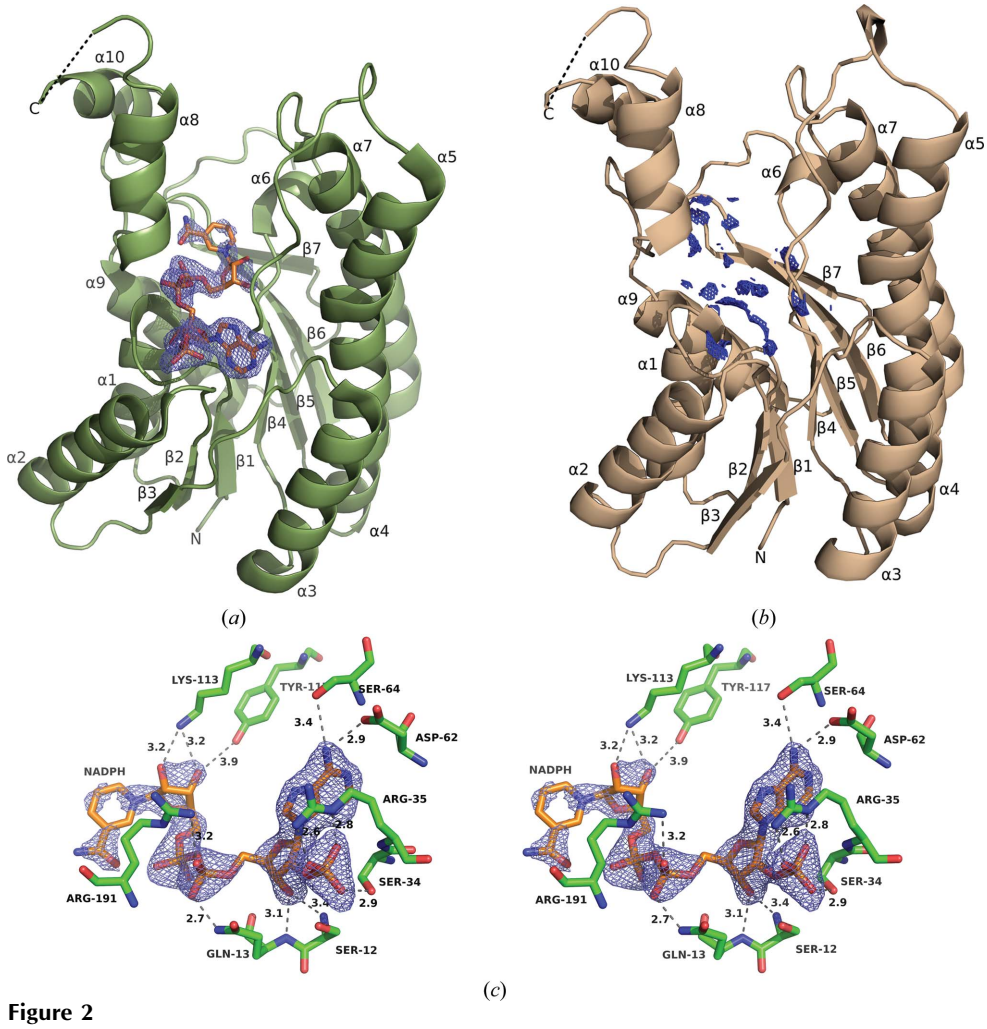

Figure 2
(*a*) The structure of YwfH with a bound cofactor (NADPH) at the active site. The σ_A-weighted ($mF_o - DF_c$) electron-density map (blue mesh) was obtained after simulated-annealing refinement (Brünger *et al.*, 1998). This map is contoured at the 3σ level around the NADPH cofactor. (*b*) The structure of the apo-like form. This structure was determined in the course of experiments designed to trap a ternary YwfH–substrate–cofactor complex. The NADPH cofactor could not be modelled owing to poor electron density. (*c*) Stereo representation of the simulated-annealed ($mF_o - DF_c$) electron-density map for the NADPH model (excluding NADPH) contoured at 3σ level.

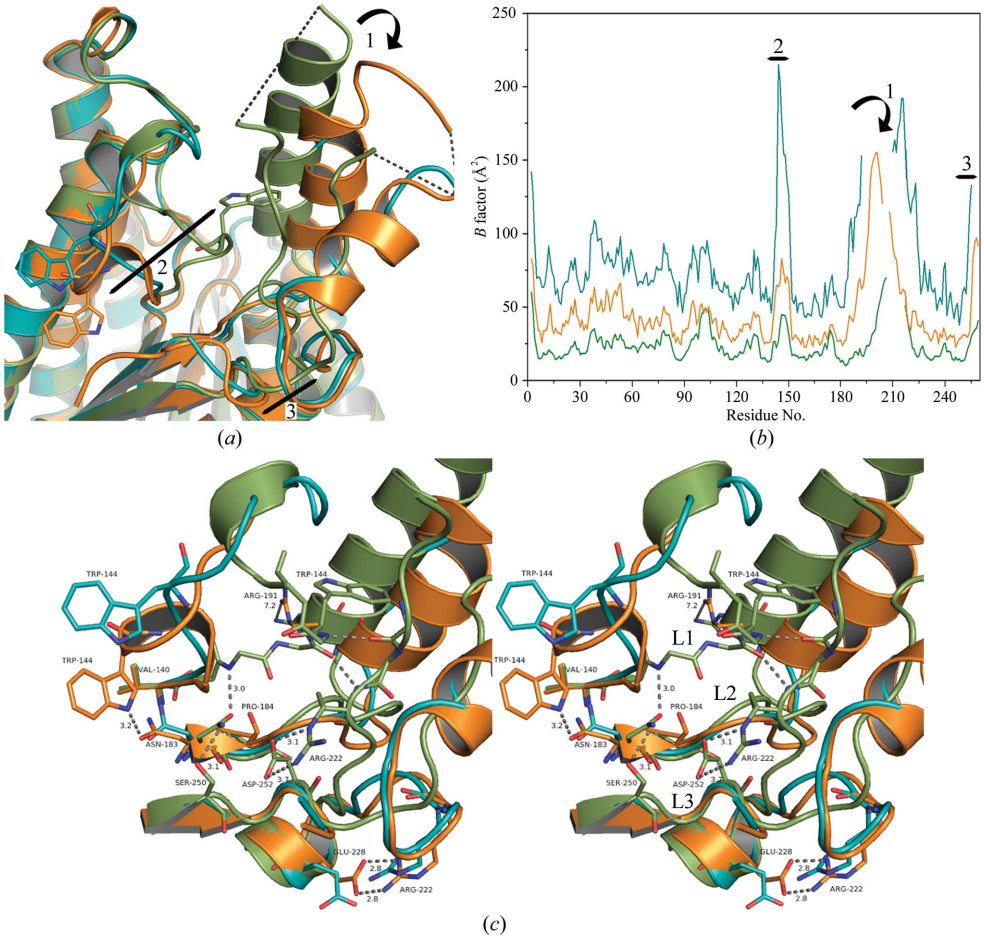

Figure 3
(*a*) The conformational changes upon NADPH binding in apo YwfH (green), holo YwfH (orange) and the apo-like form (teal) are shown. The flexible regions are labelled 1, 2 and 3 using arrows. Trp144 (stick representation) acts as a lid to the active site and exhibits a large movement of approximately 20 Å. The α8 helix is partially disordered to aid substrate entry into the active site. (*b*) The C^{α} *B*-factor plot compares the flexible regions in both the apo and substrate-bound forms of YwfH. (*c*) Stereo representation of the effect of NADPH binding and the concerted conformational changes in the loop segments L1, L2 and L3. These ribbon representations were prepared using *PyMOL* (Schrödinger LLC).

primarily owing to the rapid degradation of crystals upon soaking in the crystallization and cryoprotection buffer containing H_2HPP. After several attempts, a complete data set could be collected to 2.7 Å resolution. The structure corresponding to this data also revealed an extended conformation. However, neither NADPH nor H_2HPP could be modelled in this structure (Fig. 2*b*).

3.2. The NADPH-binding pocket

SDR enzymes bind nucleotide cofactors using a conserved glycine-rich sequence motif TG*xxx*G*x*G (Filling *et al.*, 2002; Oppermann *et al.*, 1997; Kavanagh *et al.*, 2008). These residues are ^{9}MGASQGIG^{16} (residue numbers in superscript) in YwfH (Supplementary Fig. S2). In addition to these residues, enzymes that prefer the NADPH cofactor possess a conserved basic residue prior to the second glycine in the glycine-rich motif (Gln13 in YwfH) and an invariant arginine (Arg35 in YwfH) located in the β2–α2 connecting loop (Zheng *et al.*, 2007; Kallberg *et al.*, 2002; Tanaka *et al.*, 1996). The NADPH is ensconced in a deep cleft of the cofactor-binding domain (Fig. 2*c*). In the structure of the YwfH–NADPH complex, Ser34 and Arg35 form a salt bridge with the 2′-phosphate group attached to the ribose. The hydroxyl (O3′A) group of the 2′-ribose interacts with the backbones of Gly11, Ser12 and Gln13. The side chains of Gln13 and Arg191 also interact with O1PA of the pyrophosphate moiety. The adenine ring of the NADPH is accommodated in a largely hydrophobic pocket comprising Met9, Gly10, Asp62, Met63 and Ser64. The amino group of the adenine ring forms hydrogen bonds with Asp62. Lys113 produces a kink in the α4 helix to hold the NADPH at the active site. Tyr117 is located structurally close to O3′N. Both Lys113 and Tyr117 are part of the catalytic tetrad in YwfH.

3.3. Conformational changes induced by NADPH binding

Superposition of the NADPH-binding domains of apo YwfH on the YwfH–NADPH complex (r.m.s.d. of ∼0.9 Å) suggests that the cofactor-binding domain does not undergo conformational changes upon NADPH binding. However, the inherently flexible substrate-binding domain undergoes substantial conformational changes in these two forms (Fig. 3*a*). This flexibility of the substrate-binding domain is reflected in the *B*-factor plots (shown as regions 1, 2 and 3) in a comparison between the apo form (green), YwfH–NADPH complex (orange) and apo-like form (teal) structures (Fig. 3*b*). Arg191 at the base of the α8 helix acts as a gating residue to the NADPH-binding pocket and undergoes a shift of ∼7.2 Å upon binding NADPH. The conformational changes in Arg191 and the hinge residue Pro184 in the L2 loop (^{183}NPGFIATD^{190}) disrupt the hydrogen-bonding interactions with residues in the L1 loop (^{140}VGNLWKEPGANM^{151} containing Trp144) and salt-bridge interactions (Arg222–Asp252) in the L3 loop (^{217}GIPMKRVGSA^{226}). However, these loop movements lead to another salt bridge between Arg222 and Glu228, thereby inducing the disordered L3 loop in apo YwfH to form a helix in the YwfH–NADPH complex. This motion in the α8 helix is depicted as region 1 in the *B*-factor plot (Figs. 3*a* and 3*b*).

The structural rearrangements upon NADP binding are most noticeable in the flip of the Trp144 residue that acts as

Table 2
Kinetic parameters for YwfH and active-site mutants.

Protein	K_m (μ*M*)	V_{max} (nmol s^{-1} μg^{-1}) $\times 10^{-5}$	k_{cat} (s^{-1}) $\times 10^{-3}$	k_{cat}/K_m (M^{-1} s^{-1})
Native	178.0 ± 29.7	92.82 ± 5.7	34.38 ± 2.0	193.15
K113A	365.3 ± 123.2	39.52 ± 6.4	14.63 ± 2.3	40.05
Y117A	528.2 ± 187.2	41.18 ± 9.3	17.48 ± 3.5	33.09
S155A	613.6 ± 221.1	58.32 ± 11.8	21.60 ± 4.3	35.20
N158A	344.7 ± 124.8	38.67 ± 6.6	14.32 ± 2.3	41.54
S250A	323.3 ± 62.1	77.32 ± 6.9	28.63 ± 2.5	88.56

the lid of the active-site cavity. Trp144 moves to a diagonally opposite position approximately 20 Å away from its original site (Fig. 3*a*). Asn183, which is adjacent to Pro184 in the L2 loop, is involved in stabilizing the Trp144 conformation in the YwfH–NADPH structure. The L1 loop movement is shown as region 2 in the *B*-factor plot (Fig. 3*b*). Indeed, in the apo-like form (obtained by soaking the NADPH complex crystal with H_2HPP), Trp144 adopts a distinctly different conformation than in either the apo or the NADPH-bound forms. The Ramachandran map also depicts these conformational changes: Val140 in apo YwfH moves from the favoured region ($\varphi = -133.8^\circ$, $\psi = 99.7^\circ$) to a less favoured conformation ($\varphi = -113.6^\circ$, $\psi = -96.6^\circ$) upon NADPH binding. The loss of interactions between the L2 loop and the L1 loop is also reflected in the higher temperature-factor values for this region in the apo-like form (Fig. 3*b*). The salt-bridge interactions between Arg222 and Glu228 in the YwfH–NADPH complex also make the C-terminal loop re-orient during these conformational changes. This conformational change is denoted as region 3 in the *B*-factor plot (Fig. 3*b*). All of these concerted conformational changes suggest an induced-fit mechanism in YwfH upon binding to the cofactor involving an energetically less favoured 'primed' state to aid catalysis.

3.4. Identification of the catalytic residues

The keto-reductases that predominate in the SDR enzyme family have a conserved catalytic site tetrad Asn-Ser-Tyr-Lys (Oppermann *et al.*, 2003). However, the reaction catalyzed by YwfH revealed a novel C=C reductase. The NADPH-binding region is similar to the classical SDR family of enzymes. The spatial location of the catalytic tetrad differs from the classical SDR in the case of YwfH. It was therefore necessary to experimentally verify the veracity of the catalytic tetrad using mutational analysis. The active site of the YwfH–NADPH complex provided the most significant structural confirmation of the residues in the catalytic tetrad (Fig. 2*c*). Based on this structure, Lys113 and Tyr117 are the two core members of the catalytic tetrad as they directly interact with the nicotinamide ribose of NADPH. The search for the Asn and Ser residues was less conclusive from a structural perspective, as two asparagine residues, Asn154 and Asn158, proximal to Lys113 were involved in coordinating water molecules. This Asn-coordinated water chain possibly plays a role in the water-mediated proton transfer. The crystal structure of the N158A mutant confirmed the identity of Asn158 as a bona fide member of the catalytic tetrad. In the case of the serine residue, Ser155 and Ser250 are located proximal to the active site. Apart from Lys113 and Tyr117, the other residues Ser155, Asn158 and Ser250 were assigned by their proximity and were subsequently mutated to alanine. Enzyme assays were performed with native YwfH and mutants (K113A, Y117A, S155A, N158A and S250A). The kinetic parameters are compiled in Table 2. The activity measurements revealed the role of Ser155, thus completing the identification of the catalytic tetrad as Ser155–Asn158–Lys113–Tyr117 in YwfH.

3.5. Docking of H_2HPP isomers in the active site of the YwfH–NADPH complex

The conformations of both the *E*- and *Z*-*ex*-H_2HPP forms were clustered based on the binding energy upon docking in the active site of the YwfH–NADPH complex (Supplementary Fig. S4). This analysis suggested that the *Z*-*ex*-H_2HPP conformation is preferred (176 conformations out of 256 trials) over the *E*-*ex*-H_2HPP conformer (117 conformations out of 256 trials). The binding energy corresponding to both conformers is similar (−6.05 kcal mol^{-1} for the *Z*-*ex*-H_2HPP conformer and −6.0 kcal mol^{-1} in the case of *E*-*ex*-H_2HPP). Indeed, both *E*- and *Z*-*ex*-H_2HPP interact with Gly92, Arg192 and the NADPH cofactor. However, *Z*-*ex*-H_2HPP packs more closely to NADPH compared with *E*-*ex*-H_2HPP (Supplementary Fig. S4).

3.6. HPLC and LC-ESI mass-spectrometric analyses of YwfH products

YwfH has been demonstrated to be involved in the formation of 3-(4-hydroxycyclohex-2-enyl)-2-oxopropanoic acid, also known as H_4HPP, which is an intermediate in bacilysin biosynthesis, in the presence of the cofactor NADH (Mahlstedt & Walsh, 2010). However, we could only obtain crystals of the YwfH–NADPH complex. HPLC and mass-spectrometric experiments were performed to evaluate the catalytic activity in the presence of NADPH. In this experiment, the formation of $NADP^+$ was measured rather than the production of H_4HPP because of the absence of a chromophore in H_4HPP (Fig. 4*a*). In the absence of YwfH, the NADPH peak (340 nm) appears at a retention time of approximately 11 min and the H_2HPP (260 nm) peak appears at 14 min. Addition of YwfH to the reaction mixture (BacA, BacB, prephenate and NADPH) led to a change in the HPLC profiles, with the disappearance of both the 11 and 14 min peaks. A new peak appeared at 7 min, confirming the formation of $NADP^+$ (Fig. 4*a*). We note here that even freshly prepared NADPH dissolved in the assay buffer showed minor auto-oxidation of NADPH to $NADP^+$ upon incubation for 20 min (Fig. 4*a*). An identical set of reactions were used for mass-spectrometric studies. When the peak at 7 min was subjected to LC-ESI/MS analysis, it revealed a product with an *m*/*z* of 183.1 Da [$(M–H)^-$; Fig. 4*b*]. A similar experiment performed by replacing the H_2O with D_2O resulted in deuteration of the 7 min product (Supplementary Fig. S5).

4. Discussion

The H_2HPP reductase YwfH adopts the Rossmann fold despite poor sequence similarity to other members of the SDR family of enzymes. While the MR strategy to obtain phase information for apo YwfH was straightforward, a substantial proportion of the NADPH-binding domain and the entire substrate-binding domain of YwfH had to be remodelled. After the determination of the structure of YwfH in the apo form, we turned our efforts towards determining the structure of the YwfH–NADH complex based on the observations reported previously (Mahlstedt & Walsh, 2010). However, the crystallization trials were not successful at the reported concentration (used in the activity assays) as well as at higher concentrations (∼5 m*M*) of NADH. Subsequently, activity assays and mass-spectrometric experiments suggested a clear preference for NADPH over NADH. The cocrystallization experiments resulted in a binary complex of YwfH with NADPH. A closer look at the NADPH-binding domain also suggested that the 2′-phosphate group attached to the ribose is held in place by salt-bridge interactions with Arg35. This arginine is also a conserved feature of NADPH-binding domains in proteins with the Rossmann fold (Kallberg *et al.*, 2002; Tanaka *et al.*, 1996). The crystal structure of YwfH complexed with NADPH provided significant clues to identify the residues of the catalytic tetrad, as the Lys and Tyr residues were observed to interact directly with the bound cofactor. Furthermore, in the structure of apo YwfH the substrate-binding domain was compact and the substrate-entry tunnel was blocked by Trp144 (located in the L1 loop). The assignment of the other catalytic residues was less straightforward than anticipated, primarily owing to several potential Ser and Asn candidates for the catalytic tetrad (Supplementary Fig. S2). The role of Asn158 as an active-site residue was further substantiated by the structure of the N158A mutant of YwfH, which lacks the bound water that is crucial for proton transfer. A combination of structural inputs and mutational analysis revealed the identity of the catalytic tetrad: Ser155–Asn158–Lys113–Tyr117.

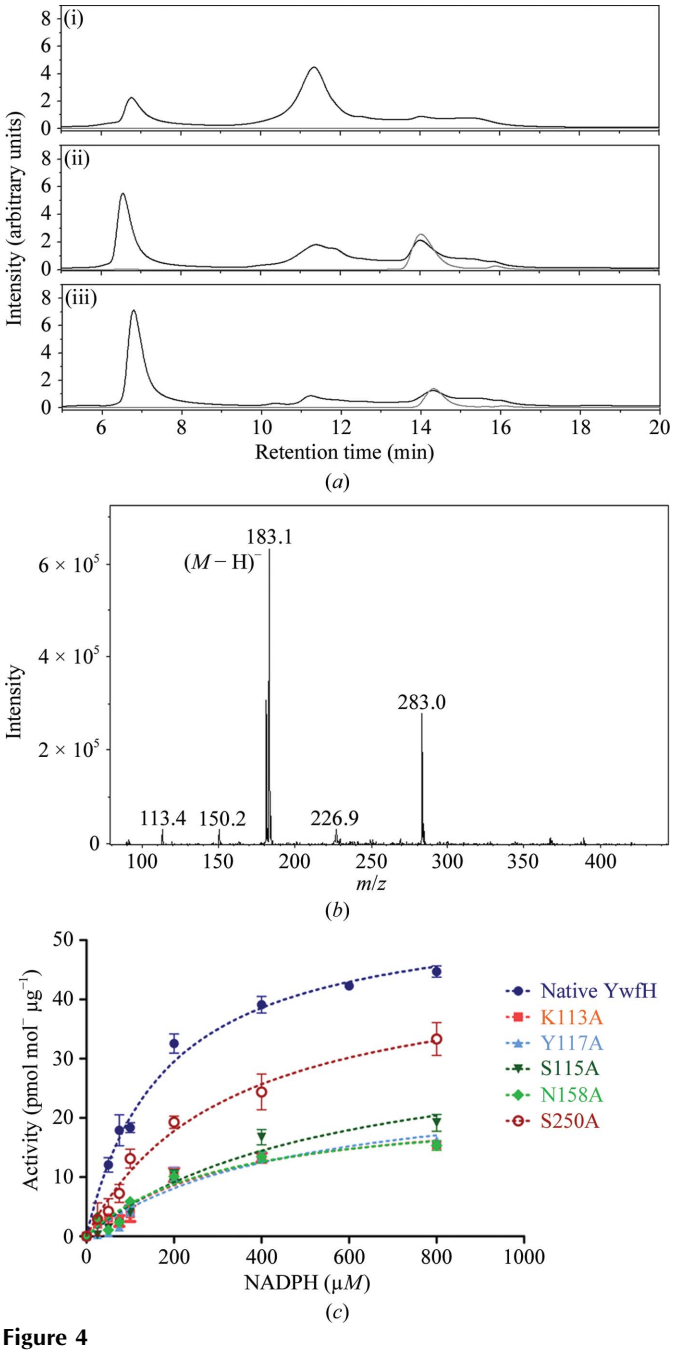

Figure 4
(*a*) The HPLC profile corresponding to $NADP^+$ (7 min) and NADPH (11 min) (i). The experimental details are described in §2. (ii) The reductase activity is indicated by the increase in $NADP^+$ (at 7 min). (iii) After 20 min, most of the H_2HPP is converted to H_4HPP. The peak corresponding to NADPH is reduced. (*b*) LC-ESI mass spectrum of H_4HPP. The predicted mass is 184.19 Da, while the observed mass is 183.1 Da $[(M–H)^-]$ in the negative ionization mode. (*c*) Michaelis–Menten plots of YwfH and the active-site mutants. All of the mutants except S250A show a substantial reduction in catalytic activity. This evidence was crucial to identify the catalytic tetrad in YwfH.

Experiments to obtain a ternary complex with H_2HPP in the active site were not successful. However, a crystal which was soaked in the reaction mixture for an extended time period (2 h) resulted in an apo-like form which revealed an intermediate state of YwfH between the unprimed (apo) and primed (NADPH-bound) states. This finding is similar to that noted by Zheng and coworkers in the case of the human protein HSCARG that adopts the Rossmann fold (Zheng *et al.*, 2007). HSCARG, like YwfH, is a dimer in solution. The crystal structure of HSCARG revealed an asymmetric dimer with only one subunit containing a bound NADP cofactor. The monomer representing the apo form without the NADP cofactor revealed a deformation of the core α-helix of the Rossmann fold. This comparison lends support to the hypothesis that the apo-like form represents an intermediate conformational step during the catalytic cycle of YwfH.

While this manuscript was in preparation, the kinetic parameters corresponding to H_2HPP activity were reported (Parker & Walsh, 2012). In this paper, Parker and Walsh suggested different preferences for YwfH towards the two products of BacB (BacB produces two isomers, 3*E*-*ex*-H_2HPP and 3*Z*-*ex*-H_2HPP, in a 3:1 ratio). This report also revealed that these two isomers could be individually purified. Activity

assays were carried out at varying concentrations of these individual isomers but at saturating concentrations of NADH (~2 m*M*). In the study reported here, the enzyme assays were conducted at varying concentrations of NADPH but in the presence of both isomers, eventually looking at the reduction of NADPH. It is thus likely that the variations in the kinetic parameters reported in this manuscript differ because of the enantiomeric selection of YwfH and variations in the assay conditions. The high-resolution NMR studies reported by Parker & Walsh (2012) suggest that YwfH produces two diastereomeric products: 4*S*-H_4HPP and 4*R*-H_4HPP. The LC-ESI/MS data, in addition to D_2O-exchange studies, support the formation of H_4HPP. The crystal structure of the YwfH–NADPH complex enabled a conformational study to evaluate the enantiomeric preferences. The *in silico* docking analysis supports the conclusions on stereoselectivity: the 3*Z*-*ex*-H_2HPP enantiomer is preferred over the other enantiomer. Enantiomeric selection thus appears to perform a regulatory role in the bacilysin-biosynthesis pathway. The crystal structure and biochemical data presented here put forth several chemical hypotheses that need to be examined. Indeed, it still remains to be demonstrated that H_4HPP formation takes place as part of bacilysin synthesis. This aspect remains tentative, as the role of BacC remains to be determined.

In conclusion, the structural characterization of YwfH reveals facets of substrate recognition and nucleotide cofactor specificity as well as a rationale for the reaction mechanism. While the structures provide an insight into a novel C=C reductase, the biochemical insights gained from an analysis of the substrate specificity provide a detailed understanding of the steps that facilitate a ping-pong reaction mechanism. Put together, the structures and biochemical data reveal distinct features of the Rossmann fold that dictate cofactor specificity in the rigid nucleotide-binding domain while embedding substrate-specificity and enantiomeric recognition features in the flexible substrate-recognition domain.

This work was supported by a grant from the Council of Scientific and Industrial Research, Government of India.

References

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.

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.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Filling, C., Berndt, K. D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jörnvall, H. & Oppermann, U. (2002). *J. Biol. Chem.* **277**, 25677–25684.

Hermes, J. D., Tipton, P. A., Fisher, M. A., O'Leary, M. H., Morrison, J. F. & Cleland, W. W. (1984). *Biochemistry*, **23**, 6263–6275.

Inaoka, T., Takahashi, K., Ohnishi-Kameyama, M., Yoshida, M. & Ochi, K. (2003). *J. Biol. Chem.* **278**, 2169–2176.

Joosten, R. P., te Beek, T. A., Krieger, E., Hekkelman, M. L., Hooft, R. W., Schneider, R., Sander, C. & Vriend, G. (2011). *Nucleic Acids Res.* **39**, 411–419.

Jörnvall, H., Persson, B., Krook, M., Atrian, S., Gonzàlez-Duarte, R., Jeffery, J. & Ghosh, D. (1995). *Biochemistry*, **34**, 6003–6013.

Josefsson, E., O'Connell, D., Foster, T. J., Durussel, I. & Cox, J. A. (1998). *J. Biol. Chem.* **273**, 31145–31152.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kallberg, Y., Oppermann, U., Jörnvall, H. & Persson, B. (2002). *Protein Sci.* **11**, 636–641.

Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. (2008). *Cell. Mol. Life Sci.* **65**, 3895–3906.

Kleeb, A. C., Kast, P. & Hilvert, D. (2006). *Biochemistry*, **45**, 14101–14110.

Krissinel, E. (2010). *J. Comput. Chem.* **31**, 133–143.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Mahlstedt, S. A. & Walsh, C. T. (2010). *Biochemistry*, **49**, 912–923.

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.

Miller, D. J., Zhang, Y.-M., Rock, C. O. & White, S. W. (2006). *J. Biol. Chem.* **281**, 18025–18032.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. (2009). *J. Comput. Chem.* **30**, 2785–2791.

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.

Oppermann, U. C., Filling, C., Berndt, K. D., Persson, B., Benach, J., Ladenstein, R. & Jörnvall, H. (1997). *Biochemistry*, **36**, 34–40.

Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B. & Jörnvall, H. (2003). *Chem. Biol. Interact.* **143–144**, 247–253.

Parker, J. B. & Walsh, C. T. (2012). *Biochemistry*, **51**, 5622–5632.

Persson, B. *et al.* (2009). *Chem. Biol. Interact.* **178**, 94–98.

Rajavel, M., Mitra, A. & Gopal, B. (2009). *J. Biol. Chem.* **284**, 31882–31892.

Steinborn, G., Hajirezaei, M. R. & Hofemeister, J. (2005). *Arch. Microbiol.* **183**, 71–79.

Tanaka, N., Nonaka, T., Nakanishi, M., Deyashiki, Y., Hara, A. & Mitsui, Y. (1996). *Structure*, **4**, 33–45.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Yamamura, A., Ichimura, T., Mimoto, F., Ohtsuka, J., Miyazono, K., Okai, M., Kamo, M., Lee, W.-C., Nagata, K. & Tanokura, M. (2008). *Proteins*, **70**, 1640–1645.

Zheng, X., Dai, X., Zhao, Y., Chen, Q., Lu, F., Yao, D., Yu, Q., Liu, X., Zhang, C., Gu, X. & Luo, M. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 8809–8814.