electronic reprint

STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

Crystal structure of pyrrolizidine alkaloid *N*-oxygenase from the grasshopper *Zonocerus variegatus*

Christian Kubitza, Annette Faust, Miriam Gutt, Luzia Gäth, Dietrich Ober and Axel J. Scheidig

Acta Cryst. (2018). D**74**, 422–432

STRUCTURAL BIOLOGY
ISSN 2059-7983

Crystal structure of pyrrolizidine alkaloid *N*-oxygenase from the grasshopper *Zonocerus variegatus*

Christian Kubitza,[a] Annette Faust,[a] Miriam Gutt,[b] Luzia Gäth,[a] Dietrich Ober[b] and Axel J. Scheidig[a]*

[a]Structural Biology, Zoological Institute, Kiel University, Am Botanischen Garten 1–9, 24118 Kiel, Germany, and [b]Biochemical Ecology and Molecular Evolution, Botanical Institute, Kiel University, Am Botanischen Garten 1–9, 24118 Kiel, Germany. *Correspondence e-mail: axel.scheidig@strubio.uni-kiel.de

Received 15 November 2017
Accepted 28 February 2018

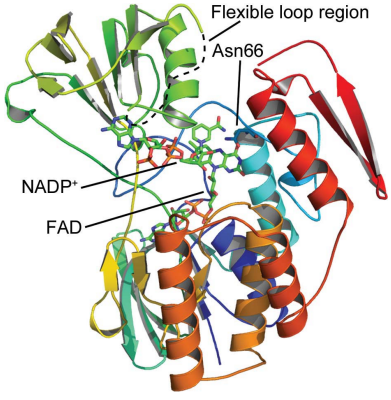
Edited by M. Rudolph, F. Hoffmann-La Roche Ltd, Switzerland

Keywords: flavin-dependent monooxygenase; FMO; pyrrolizidine alkaloids; PNO; *Zonocerus variegatus*.

PDB references: *Zv*PNO complex with FAD and NADP$^+$, 5nmx; *Zv*PNO complex with FAD, 5nmw

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

The high-resolution crystal structure of the flavin-dependent monooxygenase (FMO) from the African locust *Zonocerus variegatus* is presented and the kinetics of structure-based protein variants are discussed. *Z. variegatus* expresses three flavin-dependent monooxygenase (*Zv*FMO) isoforms which contribute to a counterstrategy against pyrrolizidine alkaloids (PAs). PAs are protoxic compounds produced by some angiosperm lineages as a chemical defence against herbivores. *N*-Oxygenation of PAs and the accumulation of PA *N*-oxides within their haemolymph result in two evolutionary advantages for these insects: (i) they circumvent the defence mechanism of their food plants and (ii) they can use PA *N*-oxides to protect themselves against predators, which cannot cope with the toxic PAs. Despite a high degree of sequence identity and a similar substrate spectrum, the three *Zv*FMO isoforms differ greatly in enzyme activity. Here, the crystal structure of the *Z. variegatus* PA *N*-oxygenase (*Zv*PNO), the most active *Zv*FMO isoform, is reported at 1.6 Å resolution together with kinetic studies of a second isoform, *Zv*FMOa. This is the first available crystal structure of an FMO from class B (of six different FMO subclasses, A–F) within the family of flavin-dependent monooxygenases that originates from a more highly developed organism than yeast. Despite the differences in sequence between family members, their overall structure is very similar. This indicates the need for high conservation of the three-dimensional structure for this type of reaction throughout all kingdoms of life. Nevertheless, this structure provides the closest relative to the human enzyme that is currently available for modelling studies. Of note, the crystal structure of *Zv*PNO reveals a unique dimeric arrangement as well as small conformational changes within the active site that have not been observed before. A newly observed kink within helix α8 close to the substrate-binding path might indicate a potential mechanism for product release. The data show that even single amino-acid exchanges in the substrate-entry path, rather than the binding site, have a significant impact on the specific enzyme activity of the isoforms.



1. Introduction

Flavin-dependent monooxygenases (FMOs), alongside cytochrome P450 enzymes, are one of two prominent families of monooxygenases in eukaryotes (Cashman, 2001). They are found throughout all phyla (Hao *et al.*, 2009), although they are expressed in different numbers of isoforms, ranging from only one FMO in yeast to five isoforms in vertebrates and to a large gene family in plants. FMOs directly recruit reducing equivalents from NAD(P)H without requiring any accessory proteins for enzyme activity (Phillips & Shephard, 2008).

FMOs, at least in vertebrates, mainly contribute to an efficient biotransformation and detoxification system for xenobiotics, converting nucleophilic heteroatom-containing

compounds (such as amines, amides, thiols and sulfides) into polar, readily excretable metabolites (Cashman, 2001; Ziegler, 2002).

During evolution, different insect lineages have developed specialized FMOs, pyrrolizidine alkaloid *N*-oxygenases (PNOs), as a counterstrategy to cope with pyrrolizidine alkaloids (PAs), which are toxic compounds that are produced by certain angiosperm species as part of their chemical defence against herbivores (Hartmann & Ober, 2008). PAs are produced by plants in their nontoxic polar *N*-oxide form. After ingestion by a vertebrate or insect herbivore, PAs are converted into the protoxic free base in the reducing gut milieu (Fig. 1). Owing to their lipophilic properties, free bases easily permeate membranes and are bioactivated to reactive pyrrolic compounds by cytochrome P450 monooxygenases that are part of the xenobiotic metabolism of the herbivore (Fu *et al.*, 2004).

PA *N*-oxygenases are the only functionally characterized FMOs found to date in insects. They are one of various strategies of adapted insects to avoid high concentrations of toxic PAs in their haemolymph by enzymatically stabilizing the nontoxic *N*-oxide (Fig. 1). This mechanism is realized in the larvae of arctiid moths (Hartmann *et al.*, 1990; Sehlmeyer *et al.*, 2010; Naumann *et al.*, 2002), *Longitarsus* flea beetle species (Narberhaus *et al.*, 2003) and the grasshopper genus *Zonocerus* (Bernays *et al.*, 1977; Wang *et al.*, 2012). Identical selection pressure upon different insect lineages has led to the independent yet convergent evolution of FMOs with almost identical substrate specificity in these organisms (Wang *et al.*, 2012).

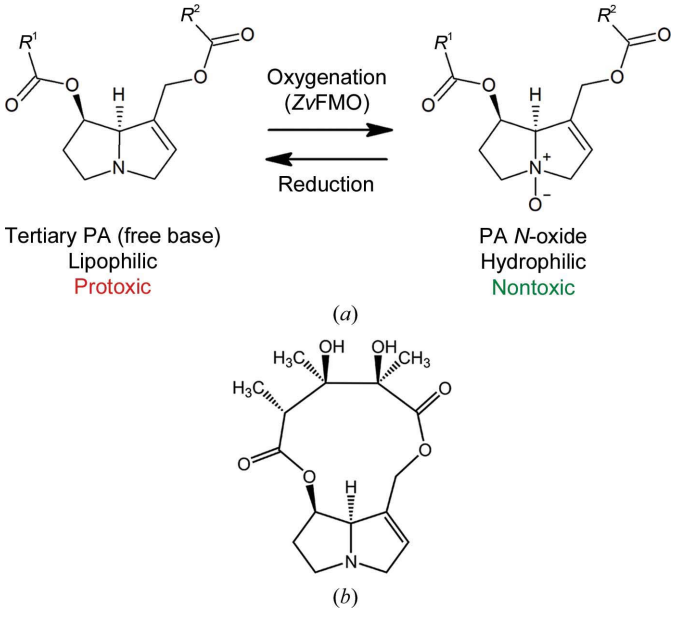


Figure 1
N-Oxygenation of pyrrolizidine alkaloids by *Zv*FMOs. Upon ingestion, plant-derived PAs are reduced to their protoxic tertiary amine form within the gut of the herbivore (Langel & Ober, 2011). (*a*) In adapted insects such as *Zonocerus*, protoxic PAs are efficiently converted to the respective PA *N*-oxides by *Zv*FMOs. PAs are esters of the bicyclic necine base moiety with one or more necic acids that are represented by R^1 and R^2 (Hartmann & Witte, 1995). (*b*) During this study, the PA monocrotaline was used as a substrate to characterize the different FMOs from *Zonocerus*.

Three isoforms of flavin-dependent monooxygenase from the locust *Z. variegatus* (*Zv*FMOs) have been identified, named *Zv*FMOa, *Zv*FMOc and *Zv*PNO. They all share a similar substrate spectrum and are able to oxygenate a variety of different PAs. However, they exhibit quite different enzyme activities. While *Zv*PNO is the most potent isoform, *Zv*FMOa and *Zv*FMOc show a specific activity that is eightfold to 300-fold lower, depending on the PA substrate. The high degree of amino-acid sequence identity (77–77.8%) between the various *Zv*FMOs combined with their similar substrate spectra indicates a duplication event of a common ancestor FMO-coding gene that already possessed PA *N*-oxygenating capabilities and was subsequently recruited and optimized for plant-derived PAs (Wang *et al.*, 2012).

The FMO reaction cycle had been investigated before the first structures became available, and is widely distributed in the literature. Class B FMOs such as those from *Z. variegatus* consist of two domains, each of which contains a dinucleotide-binding domain in the form of a Rossmann fold, which is responsible for cofactor binding. In its native state, oxidized FAD is tightly bound to the enzyme as a prosthetic group. NADPH is recruited as a co-substrate and transfers reducing equivalents to FAD. Upon the reaction of reduced FAD by molecular oxygen, a C4a-hydroperoxy-FAD intermediate is formed. This intermediate is capable of inserting one O atom into a substrate compound and ends up as C4a-hydroxy-FAD, which may then release the second O atom as part of a water molecule, restoring the oxidized FAD cofactor (Ziegler, 2002; Robinson *et al.*, 2013). A cocked-gun mechanism has been proposed for the reaction cycle of FMOs, meaning that the activated C4a-hydroperoxy-FAD intermediate can be stabilized within the enzyme until a substrate accesses the active site and is then immediately oxygenated (Cashman, 1995; Ziegler, 1993). The presence of $NADP^+$ seems to be crucial for the stabilization of this intermediate and therefore it has to remain bound to the enzyme throughout the whole catalytic cycle, making it the last compound to be released from the enzyme (Beaty & Ballou, 1981).

Here, we present two crystal structures: (i) the pyrrolizidine alkaloid *N*-oxygenase (*Zv*PNO) from *Z. variegatus* with bound FAD and (ii) *Zv*PNO in complex with FAD and $NADP^+$. In addition, we present data on *Zv*FMOa variants, which are discussed in terms of a better understanding of substrate conversion as well as the evolutionary background and specialization of FMOs in insects. The crystal structures presented in this study are the first available FMO structures to originate from more highly developed eukaryotes. Our data show a flexible helix close to the active site which is likely to be involved in substrate binding and/or product release and has not been described before in other FMO structures. Although *Zv*FMO isoforms are dimeric in solution, as is frequently observed for other members of the FMO family, their specific arrangement of the two subunits is unique. Of note, the contact area is rather small; nevertheless, the dimer is very stable in solution.

2. Experimental procedures

2.1. Generation of expression plasmids for *Zv*FMO isoforms and variants

The identification of three different *Zv*FMO isoforms (*Zv*FMOa, *Zv*FMOc and *Zv*PNO) and the generation of individual expression plasmids have been described previously (Wang *et al.*, 2012). Expression vectors for the *Zv*FMOa variants were generated according to the QuikChange site-directed mutagenesis protocol (Stratagene, La Jolla, California, USA) using Phusion high-fidelity DNA polymerase (Fermentas) with the pET-22b-ZvFMOa plasmid as a template and the following primers: *Zv*FMOa variant F307Y, forward, 5′-CCT CAC GAC GCC TGT TAT TCG ATC TTG TTT GAT C-3′; reverse, 5′-G ATC AAA CAA GAT CGA ATA ACA GGC GTC GTG AGG-3′; *Zv*FMOa variant Y356A, forward, 5′-CGC CCG CAC TTC ATG GCT AAC CGT CAG TGG AAG-3′; reverse, 5′-CTT CCA CTG ACG GTT AGC CAT GAA GTG CGG GCG-3′; *Zv*FMOa variant P388S, forward, 5′-G TTC GAT GAT CTG GCT TCT GGT TTG ACG AAG GAC-3′; reverse, 5′-GTC CTT CGT CAA ACC AGA AGC CAG ATC ATC GAA C-3′; *Zv*FMOa variant F354V, forward, 5′-GGC TTC CGC CCG CAC GTC ATG TAT AAC CGT CAG-3′; reverse, 5′-CTG ACG GTT ATA CAT GAC GTG CGG GCG GAA GCC-3′; ZvFMOa double variant F354V/Y356A, forward, 5′-CGC CCG CAC GTC ATG GCT AAC CGT CAG TGG AAG-3′; reverse, 5′-CTT CCA CTG ACG GTT AGC CAT GAC GTG CGG GCG-3′ (the underlined codons indicate the sites of mutagenesis). The template plasmid was digested with the restriction endonuclease DpnI prior to transformation of the reaction mixture into *Escherichia coli* XL1 Blue cells for subsequent screening and control sequencing.

2.2. Expression and purification of *Zv*FMOs

*Zv*FMO isoforms as well as their variants were heterologously expressed in *E. coli* BL21(DE3) cells harbouring the respective recombinant plasmids. Cells were grown at 210 K in LB medium containing 100 µg ml^{-1} ampicillin (for pET-22b-derived vectors) or 30 µg ml^{-1} kanamycin (for pET-28a-derived vectors). Protein expression was induced by adding 0.1 m*M* isopropyl *β*-D-1-thiogalactopyranoside (IPTG) to the bacterial cultures at an OD_{600} of 0.6 followed by incubation at 297 K for 16 h. The cells were lysed using an EmulsiFlex-C3 (Avestin, Mannheim, Germany) and subsequently centrifuged at 75 600*g* for 1 h. His_6-tagged recombinant proteins were purified from the resulting crude extract *via* immobilized metal ion-affinity chromatography (IMAC) using a HisTrap HP 5 ml column and an ÄKTApurifier FPLC system (both from GE Healthcare, Freiburg, Germany). The lysis and equilibration buffer was composed of 50 m*M* NaH_2PO_4 pH 8.0, 300 m*M* NaCl, 20 m*M* imidazole, 10 m*M* *β*-mercaptoethanol, 200 µ*M* phenylmethanesulfonyl fluoride (PMSF). Proteins were eluted *via* a linear gradient over five column volumes to a final concentration of 500 m*M* imidazole. Eluted target protein fractions were pooled and subsequently subjected to size-exclusion chromatography (HiLoad 16/60 Superdex 200 pg, GE Healthcare) with a running buffer composed of 20 m*M* glycine pH 9.0, 200 m*M* NaCl, 5 m*M* DTT. Purified protein was concentrated to approximately 6 mg ml^{-1} and stored in 20 m*M* glycine pH 9.0, 1 m*M* DTT at 193 K.

2.3. SEC-MALS analysis

In order to determine the oligomeric states of the purified *Zv*FMO proteins, HPLC-based size-exclusion chromatography (SEC) with refractive-index (RI) and multi-angle laser light-scattering (MALS) detectors was performed. Protein solutions were diluted to 4 mg ml^{-1} in phosphate-buffered saline (PBS) buffer pH 7.0 prior to SEC-MALS analysis. The HPLC–SEC system consisted of an online vacuum degasser (S 8515; SRI Instruments Europe), a high-pressure quaternary pump (G1311A; Agilent Technologies), a manual injector (7725i; Rheodyne), a pre-column (WTC-030N5G, 4.6 × 50 mm; Wyatt Technology), a SEC column (WTC-030N5, 4.6 × 300 mm; Wyatt Technology), a MALS detector (miniDAWN TREOS, λ = 658 nm; Wyatt Technology) and a RI detector (G1362; Agilent Technologies). Data acquisition and processing were carried out using the *ASTRA* software (Wyatt Technology). The mobile phase for the SEC-MALS analysis was PBS (adjusted to pH 7.0 with phosphoric acid) and the flow rate was 0.4 ml min^{-1}. Additionally, the SEC column was calibrated with a protein-standard mixture containing bovine thyroglobulin, bovine gamma globulin, chicken ovalbumin and bovine ribonuclease A. Retention times were correlated with molecular weight and used to determine the oligomeric states of the *Zv*FMOs.

2.4. Crystallization

Initial crystallization hits were identified by high-throughput screening performed at the SPC Facility at EMBL Hamburg. Crystallization experiments for the refinement of the initial conditions were carried out using the hanging-drop vapour-diffusion method at 291 K. *Zv*PNO crystals were obtained by mixing protein solution (5.6 mg ml^{-1} in 20 m*M* glycine–NaOH pH 9.0, 1 m*M* DTT) with an equal amount of precipitant solution [20 m*M* Tris–HCl pH 7.0, 200 m*M* $MgCl_2$, 15%(*w*/*v*) PEG 3350]. In order to generate crystals of *Zv*PNO in complex with the oxidized NADP cofactor, 1 m*M* $NADP^+$ was added to the precipitant solution for co-crystallization. Crystals of wild-type *Zv*FMOa were obtained by mixing the protein solution (6.5 mg ml^{-1} in 20 m*M* glycine–NaOH pH 9.0, 1 m*M* DTT) with an equal amount of precipitant solution [20 m*M* bis-tris propane–NaOH pH 7.5, 200 m*M* $NaNO_3$, 22%(*w*/*v*) PEG 3000, 1 m*M* TMA–HCl, 1 m*M* $NADP^+$]. Drops were equilibrated against reservoir solution. Crystals appeared after 2–4 d in the form of mostly rod-shaped clusters or displaying multiple lattices. A few single crystals could be isolated and were briefly equilibrated in cryoprotectant solution prior to flash-cooling in liquid nitrogen. *Zv*PNO–FAD crystals were equilibrated in 3.2 *M* trimethylamine *N*-oxide (TMAO) for cryoprotection. Crystals of *Zv*PNO in complex with oxidized NADP cofactor as well as *Zv*FMOa crystals

Table 1
Data collection and processing.

Values in parentheses are for the highest resolution shell.

Data set	*Zv*PNO–FAD	*Zv*PNO–FAD–NADP$^+$	*Zv*FMOa
PDB code	5nmw	5nmx	
Data collection			
Diffraction source	P14, PETRA III, EMBL Hamburg	P14, PETRA III, EMBL Hamburg	P14, PETRA III, EMBL Hamburg
Wavelength (Å)	0.976261	0.976200	0.976300
Temperature (K)	100	100	100
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M
Crystal-to-detector distance (mm)	395.2	320.1	463.0
Rotation range per image (°)	0.1	0.1	0.1
Exposure time per image (s)	0.1	0.1	0.1
Space group	*P*1	*P*1	$C222_1$
a, *b*, *c* (Å)	73.9, 76.3, 80.9	74.1, 76.1, 81.7	89.6, 212.2, 166.8
α, β, γ (°)	72.0, 81.5, 81.2	71.8, 81.6, 82.0	90, 90, 90
Resolution range† (Å)	76.46–1.89 (1.92–1.89)	77.13–1.60 (1.86–1.60)	89.53–3.00 (3.16–3.00)
Total No. of reflections†	268372	1034819	181082
No. of unique reflections†	119135	213835	32199
Completeness† (%)	90.2 (62.8)	96.1 (47.5)	99.8 (99.9)
Multiplicity†	2.3	4.9	5.6
$\langle I/\sigma(I)\rangle$†	8.8 (2.0)	39.3 (2.3)	5.9 (0.7)
Mean *I* half-set correlation $CC_{1/2}$†	99.5 (62.2)	99.7 (43.3)	98.9 (30.4)
R_{merge}†	0.07 (0.64)	0.08 (0.83)	0.28 (2.38)
R_{meas}†	0.09 (0.83)	0.09 (1.02)	0.35 (2.93)
$R_{p.i.m.}$†	0.06 (0.52)	0.04 (0.58)	0.20 (1.69)
Matthews coefficient (Å^3 Da^{-1})	2.16	2.19	2.71
Solvent content (%)	43.18	43.77	54.61
No. of molecules in asymmetric unit	4	4	3
Refinement			
Resolution (Å)	76.46–1.89 (1.94–1.89)	77.13–1.60 (1.64–1.60)	
R_{work}‡ (%)	19.1 (27.2)	17.4 (31.6)	
R_{free}‡ (%)	21.9 (33.1)	20.2 (35.4)	
MolProbity score§	1.70	2.05	
No. of non-H atoms			
Protein	13239	13310	
FAD cofactor	212	212	
NADP$^+$ cofactor	—	192	
Mg^{2+}	2	3	
Water	1207	1572	
Total	14660	15289	
Overall *B* factor from Wilson plot¶ (Å^2)	35.8	27.0	
Average *B* factor†† (Å^2)			
Protein (main chain)	29.7	18.3	
Protein (side chain)	32.3	22.9	
Protein (whole chain)	31.0	20.7	
FAD cofactor	25.9	12.8	
NADP$^+$ cofactor	—	29.4	
Mg^{2+}	34.9	23.4	
Water	35.0	28.6	
Ramachandran plot‡‡			
Favoured (%)	96.4	96.2	
Allowed (%)	3.6	3.7	
Outliers (%)	0.0	0.1	
R.m.s. deviations‡			
Bonds (Å)	0.006	0.023	
Angles (°)	1.000	2.305	

† Values as provided by *AIMLESS* after data processing, merging and scaling. ‡ Calculated by *REFMAC*5 (Murshudov *et al.*, 2011). § Calculated using the *MolProbity* server (Chen *et al.*, 2010). ¶ Calculated using *SFCHECK* (Vaguine *et al.*, 1999). †† Calculated using *BAVERAGE*. ‡‡ Calculated using *RAMPAGE*. All programs used here are implemented within the *CCP*4 program package (Winn *et al.*, 2011).

were equilibrated in 67%(*v*/*v*) reservoir solution in deionized water supplemented with a final concentration of 27.4%(*w*/*v*) PEG 3000.

2.5. Co-crystallization and soaking experiments

Co-crystallization experiments were carried out using the abovementioned crystallization conditions additionally supplied with 0.5–1 m*M* senecionine, 0.5–7.6 m*M* atropine, 3–33 m*M* homatropine hydrobromide or 14–154 m*M* atropine sulfate. The same PA substrate solutions and concentrations were used for soaking experiments. Soaking was performed by transferring crystals into the respective cryoprotectant solution (supplied with different amounts of substrates) for 5 min at 291 K before flash-cooling in liquid nitrogen.

2.6. *Zv*FMO activity assay

The activity of *Zv*FMO was determined photometrically using an Ultrospec 2100 pro (GE Healthcare, Freiburg, Germany). All substances were dissolved or diluted in glycine buffer (100 m*M* glycine–NaOH, pH 9.0). 340 µl of diluted enzyme (0.2 mg ml^{-1}) was incubated in a UV cuvette for 1 min at 303 K. 40 µl of 2 m*M* NADPH solution was added and incubated at 30°C for another minute. The enzymatic reaction was started by adding 20 µl of monocrotaline solution (2 mg ml^{-1}). NADPH turnover was monitored at 340 nm and directly correlated to *N*-oxygenation of monocrotaline.

2.7. Data collection, structure determination and representation

X-ray diffraction data were collected at 100 K on MX beamline P14, EMBL/DESY PETRA III (Hamburg, Germany) equipped with a PILATUS 6M detector. Diffraction data were indexed and integrated using the *XDS* software (Kabsch, 2010). Space-group determination, data scaling and merging were performed by the *AIMLESS* software as part of the *CCP*4 program suite (Winn *et al.*, 2011), while applying the free-*R* flag to 5% of reflections. The structure of *Zv*PNO without cofactors was solved by molecular replacement (MR) using the *MOLREP* software as implemented in the *CCP*4 program suite. The starting model for MR was built based on the bacterial FMO from *Methylophaga aminisulfidivorans* (PDB entry 2xve, subunit *A*,

33.3% sequence identity; Cho *et al.*, 2011) using *MODELLER* on the *HHpred* server (Söding *et al.*, 2005; Söding, 2005). For the rotation and translation search the model was manually trimmed to omit less well defined loop regions. The *MOLREP* scores for stepwise positioning of four molecules within the asymmetric unit were 0.136, 0.164, 0.200 and 0.224. The resulting *Zv*PNO model was iteratively completed by alternating refinement steps using *REFMAC*5 (Murshudov *et al.*, 2011) as well as manual inspection, modification and insertion of cofactors and water molecules using *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010). The refined model was used as the starting model for subsequent data sets from *Zv*FMO isoforms and variants. Refinement statistics are summarized in Table 1. The final model analysis, imaging and ray tracing were performed using *PyMOL* (v.1.8; Schrödinger).

3. Results

3.1. Expression, purification, crystallization and data collection

All *Zv*FMO isoforms were overproduced in *E. coli* and purified *via* FPLC. The enzymes eluted as stable, dimeric, 421–425-residue proteins (depending on the isoform) with a characteristic flavoprotein absorption maximum at 450 nm. The oligomeric state was separately confirmed by SEC-MALS analysis. All isoforms were subjected to crystallization trials, yet only *Zv*PNO yielded high-quality crystals that were suitable for X-ray diffraction experiments and subsequent structure determination. Crystal structures of the *Zv*PNO–FAD and *Zv*PNO–FAD–NADP$^+$ complexes were determined to resolutions of 1.9 and 1.6 Å, respectively. 12 C-terminal residues (including the hexahistidine affinity tag) as well as residues 216–223 could not be modelled owing to the absence of appropriate electron density. The electron-density maps for the rest of the model were of high quality. Co-crystallization and soaking experiments with low substrate concentrations yielded crystals that were suitable for structure determination. However, no additional electron density could be observed within the active site which could be interpreted as substrate. Using higher amounts of different substrates for co-crystallization resulted in the formation of amorphous precipitate rather than protein crystals. Crystals which were soaked with higher substrate concentrations immediately showed cracks when transferred to the soaking solution and began to dissolve after a few seconds. Protein crystals as well as data sets were also obtained for the isoform *Zv*FMOa, but owing to poor crystal quality and the resulting poor data statistics no reliable model could be determined for this isoform with regard to specific side-chain arrangements. However, the obtained diffraction data were sufficient for MR phasing, C^{α} tracing and analysis of the crystal packing as well as the dimeric arrangement. No crystallization condition was identified for the *Zv*FMOc isoform, even after screening using *Zv*PNO crystals for cross-seeding.

3.2. Overall structure of *Zv*PNO

*Zv*PNO is composed of two structural domains, each of which exhibits a dinucleotide-binding Rossmann fold crucial for binding either the FAD or the NADP cofactor. While FAD is tightly bound to the large structural domain composed of amino-acid residues 1–155 and 260–425, NADP$^+$ is more loosely bound to the small domain (residues 156–259), where it can easily be exchanged for its reduced equivalent. The substrate-binding site is located within the cleft formed by the two domains, in direct proximity to the flavin moiety of the FAD, the ribose of NADP$^+$ and the highly conserved amino-acid residue Asn66 (Fig. 2).

3.3. FAD binding

The FAD cofactor is deeply anchored within the large structural domain *via* the formation of hydrogen bonds to

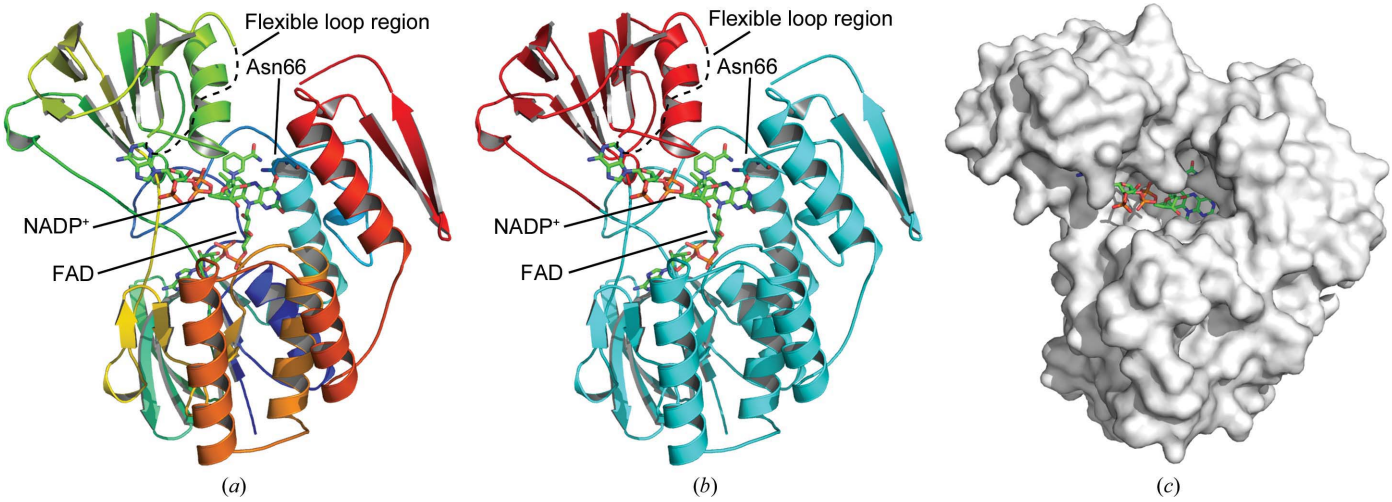


Figure 2
Representation of the overall structure of *Zv*PNO (PDB entry 5nmx) with the FAD and NADP$^+$ cofactors as well as Asn66 depicted as stick models. (*a*) Cartoon representation of one *Zv*PNO subunit. The protein backbone is rainbow-coloured from the N-terminus (blue) to the C-terminus (red). The flexible loop region between residues 216 and 223 could not be traced. (*b*) Colour-coded structural domains. Cyan, large FAD-binding domain (residues 1–155 and 260–425); red, small NADP$^+$-binding domain (residues 156–259). (*c*) Surface representation with visible substrate-entrance path leading towards the active site in front of the isoalloxazine-ring system of FAD.

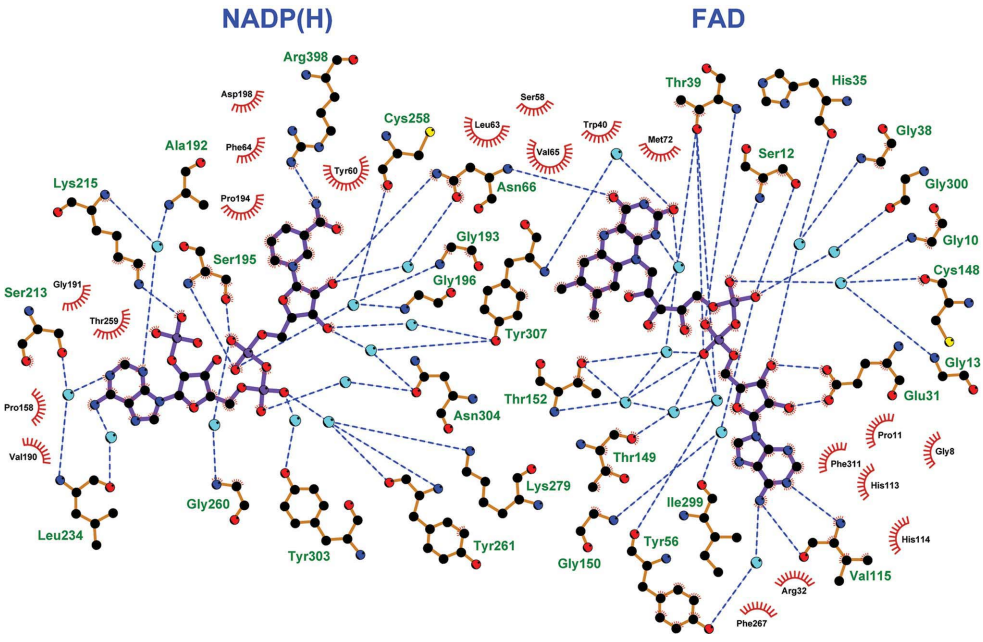


Figure 3
Cofactor-binding site of *Zv*PNO with bound FAD and NADP$^+$. The cofactors are shown in ball-and-stick representation; the bonds are indicated in purple. The protein residues are represented with side chains in ball-and-stick representation; the bonds are indicated in yellow. Hydrogen bonds are shown as blue dashed lines, and the spoked arcs represent protein residues that form hydrophobic interactions with the cofactors. The cyan spheres indicate water molecules which provide bridged hydrogen bonds between amino-acid residues and the cofactors. For clarity, the lengths of the hydrogen bonds are not given. The representation was derived from an analysis with *LigPlot*$^+$ (Laskowski & Swindells, 2011).

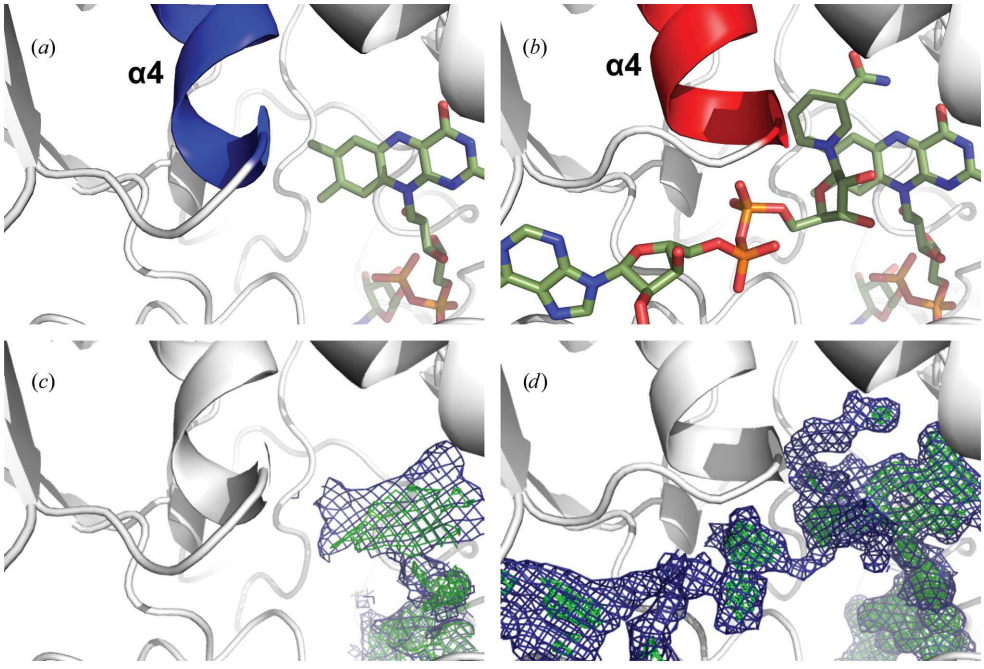


Figure 4
Conformational adjustment of helix α4 upon binding of the NADP cofactor. (*a*, *b*) Cofactors are represented as stick models. Helix α4 is colour-coded according to its conformational states: (*a*) blue, without NADP cofactor (PDB entry 5nmw); (*b*) red, with bound NADP cofactor (PDB entry 5nmx). (*c*, *d*) Composite OMIT maps of cofactors: (*c*) electron density indicating the FAD cofactor in PDB entry 5nmw; (*d*) electron density indicating the FAD and NADP$^+$ cofactors in PDB entry 5nmx. Blue, $2F_o - F_c$ map contoured at 1.0 σ; green, $F_o - F_c$ map contoured at 2.5σ. Composite OMIT maps were generated using *PHENIX* (Adams *et al.*, 2010). Simulated-annealing cycles were performed to remove model bias.

polar or charged amino-acid side chains (residues Ser12, Glu31 and Thr39) as well as N and O atoms of the protein backbone (residues Ser12, Thr39, Asn66 and Val115). It is further held in position by numerous coordinated water molecules and hydrophobic residues, which define a major part of the binding site for this cofactor (Fig. 3). The typical dinucleotide-binding motif G*X*G*XX*G is represented by residues 8–13. Only the reactive flavin moiety extends towards the solvent-exposed cleft encircled by the two structural domains and is accessible to NADPH, molecular oxygen and substrates.

3.4. NADP$^+$ binding

The NADP$^+$ cofactor is bound to the G*X*G*XX*G motif (residues 191–196) of the smaller structural domain in an extended conformation *via* hydrogen bonds to its diphosphate moiety. The 2′ phosphate of NADP$^+$ is coordinated by Lys223 and His351. Interactions between the nicotinamide moiety and Phe64 and Arg398 as well as between the ribose moiety and Asn66 further support the binding and positioning of the cofactor (Fig. 3). In comparison to the *Zv*PNO–FAD complex, the presence of NADP$^+$ leads to a small conformational change of helix α4 such that it provides additional cofactor stabilization by interactions between a positive partial charge of the helix dipole and the diphosphate of NADP$^+$ (Fig. 4). The conformational change in the preceding loop region is caused by a flipping alanine, which reduces steric hindrance when binding the NADP$^+$ cofactor.

3.5. Active site and substrate-binding cleft

The active site of *Zv*PNO is located in direct proximity to the functional moieties of both dinucleotide cofactors in combination

with Asn66, which is highly conserved among FMOs. The ribose moiety of NADP$^+$ and Asn66 are supposed to coordinate molecular oxygen and stabilize the C4a-hydroperoxy-FAD intermediate (Eswaramoorthy *et al.*, 2006; Olucha *et al.*, 2011; Hille *et al.*, 2013). This is supported by an elongated electron-density feature found in this position in at least one chain of the *Zv*PNO–FAD–NADP$^+$ crystal structure, which was interpreted as two neighbouring water molecules but could also account for molecular oxygen. The rest of the substrate-binding pocket is formed by Asn66, Leu67, Pro194, Phe383, Val386, Tyr307 and Tyr303–Ala305, while the entrance is further restricted by a loop composed of Val354, Met356, Asn357 and Gln359 (Fig. 5).

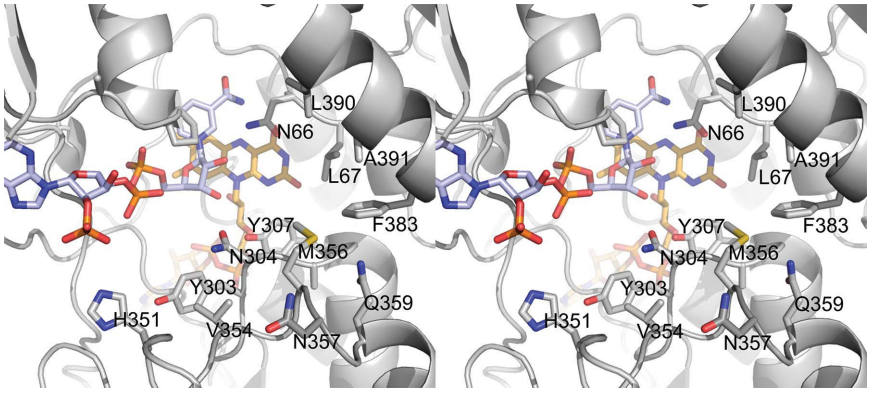


Figure 5
Stereo representation of the *Zv*PNO substrate-binding site. The protein backbone is shown in cartoon representation; cofactors and residues forming the active site and substrate-binding pocket are shown in stick representation. C atoms and backbone cartoon are in grey, N atoms are in blue, O atoms are in red and S atoms are in yellow; NAD$^+$ C atoms are in light blue and FAD C atoms are in light orange.

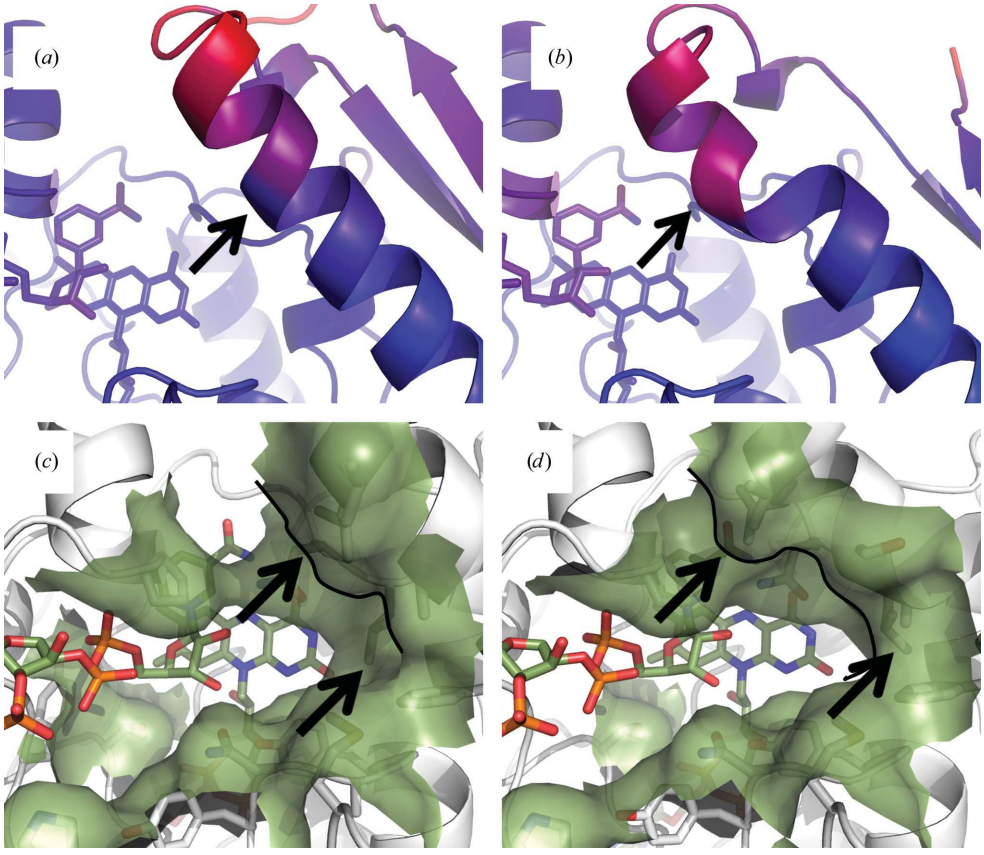


Figure 6
Binding-pocket limitations caused by conformational changes. (*a*, *b*) Cartoon representation of helix α8, colour-coded by *B*-factor distribution (spectrum from blue, 10 Å^2, to red, 65 Å^2). Left, chain *D*; right, chain *C* as found in the asymmetric unit of PDB entry 5nmx. Arrows highlight conformational changes. (*c*, *d*) Surface representation of residues forming the binding pocket; arrows highlight conformational changes that regulate the size of the substrate-binding pocket.

As in most enzymes, the binding pocket of *Zv*PNO is not static but exhibits some flexibility. Small conformational changes were observable on comparing different protein subunits within the unit cell, accounting for some degree of structural plasticity. The most prominent among them is helix α8 composed of Pro377–Lys392, which can be kinked to a certain degree. Thereby, the space available for substrate binding is partly occupied (Fig. 6). The flexibility of this part of helix α8 is further indicated by higher *B* factors. In addition, the electron density for the side-chain atoms of Leu67 reveals different rotamer conformations when comparing the four protein molecules comprising the asymmetric unit. Depending on the conformation of Leu67, the volume of the binding pocket at the reaction centre is slightly reduced or enhanced, respectively.

3.6. Dimer formation

Despite the differences in sequence, the overall structures of the flavin-dependent monooxygenases are very similar. This indicates the need for high conservation of the three-dimensional structure for this type of reaction. For the biological assembly, a broader variation can be observed. Many functional assemblies of FMOs are known to be homo-oligomers (dimers, tetramers or hexamers), but the interfaces that contribute to oligomer formation seem to be

unpredictable. The published crystal structures of dimeric FMOs reveal quite different dimeric arrangements (Malito *et al.*, 2004; Leisch *et al.*, 2012; Eswaramoorthy *et al.*, 2006; Cho *et al.*, 2011), none of which are shared by the structures presented in this study. For all *Zv*FMO isoforms the major fraction was identified as a dimer by SEC-MALS analysis and calibrated SEC (Fig. 7; results are only shown for *Zv*PNO). Symmetrical homodimers with twofold symmetry can be observed within the *Zv*PNO crystal structure, however, with a new type of orientation (Fig. 8). Contacts are provided by hydrogen bonds and salt bridges between the large domain of one subunit (residues Trp47, Val55, Glu116, Trp153 and Arg264) and the small domain of the symmetry mate (residues Arg167, Glu239 and Trp240). With the exception of Val55, which is only involved *via* its backbone atoms, these residues are shared by all three *Zv*FMO isoforms. Of note, Arg167 is also part of the signature sequence (residues 166–176), which is highly conserved among FMOs (Fraaije *et al.*, 2002). Despite the few contact sites and the rather small interface area of approximately 900 Å^2 (calculated by *PDBePISA*; Krissinel & Henrick, 2007) a crystallographic artefact can be ruled out by the nonrefined crystal structure of *Zv*FMOa. It exhibits the same dimeric arrangement of subunits, but a different crystal packing of these dimers. The substrate-binding sites of the two enzyme molecules forming the biologically active *Zv*FMO dimer are not occupied by the interface area and are fully solvent-accessible. The two active sites point in opposite directions, which enhances the possibility of cofactor and substrate recruitment in a rather orientation-independent manner. Based on this new dimeric arrangement, it seems that the relative orientation of the protein subunits is not crucial for the enzymatic reaction.

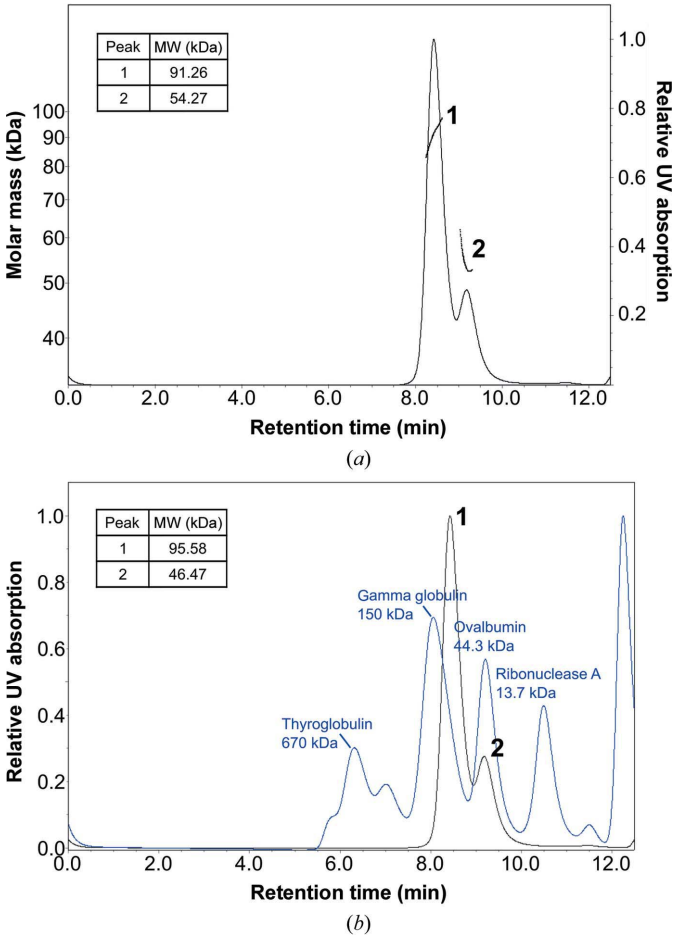


Figure 7
SEC-MALS analysis of *Zv*PNO. (*a*) SEC-MALS analysis of *Zv*PNO with mean molecular weights calculated by the *ASTRA* software for observable peaks. (*b*) Calibrated SEC analysis with molecular weights calculated from the retention times of observable *Zv*PNO peaks.

3.7. Sequence alignment of *Zv*FMO isoforms

All *Zv*FMO isoforms share high sequence identity (77.7% between *Zv*PNO and *Zv*FMOa; 77.1% between *Zv*PNO and *Zv*FMOc; 82.8% between *Zv*FMOa and *Zv*FMOc). Still, their specific enzyme activities significantly diverge from each other, probably owing to differences in the substrate-binding pocket or substrate-entry path (Fig. 9). In contrast, the sequence identities of *Zv*PNO compared with its modelling templates from *M. aminisulfidivorans* (34% sequence identity; r.m.s.d. of 2.5 Å) and *Schizosaccharomyces pombe* (27% sequence identity; r.m.s.d. of 2.3 Å) are rather low.

3.8. *Zv*FMOa variants

Variants of the less active isoform *Zv*FMOa were generated to identify amino-acid residues which have an impact on the specific activity of this enzyme and are likely to be involved in substrate binding or turnover. We investigated residues which are part of the substrate-binding pocket and differed between the *Zv*PNO and *Zv*FMOa isoforms. Based on initial docking experiments with the crystal structure of *Zv*PNO and different PA substrates, Tyr307 was identified as the most promising part of the binding pocket, potentially forming hydrogen bonds to a variety of PA substrates. Additionally, when comparing the crystal structures of *Zv*PNO with a homology model of *Zv*FMOa, a tyrosine residue (Tyr356) blocking the substrate entrance in the latter model was observed. Thus, the *Zv*FMOa variant F307Y and a double variant F307Y/Y356A

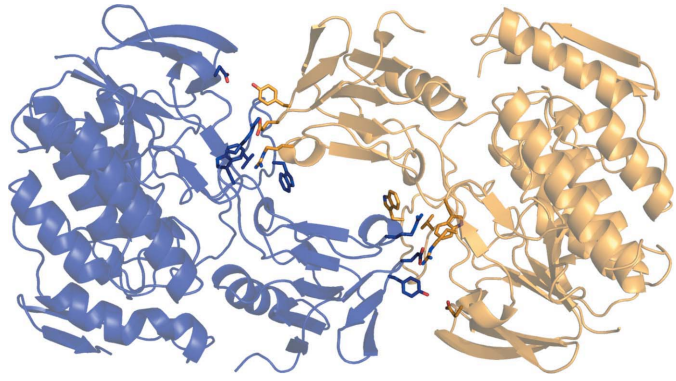


Figure 8
Dimeric arrangement of *Zv*PNO. Cartoon representation of two *Zv*PNO subunits (blue and yellow). Interface residues are depicted as stick models. The view is along the noncrystallographic twofold dyad. The surface-complementarity score was calculated to be 0.8 (Lawrence & Colman, 1993).

were investigated. In contrast to preliminary substrate-docking predictions, the *Zv*FMOa F307Y variant resulted in a decline in specific activity by a factor of 0.81 in comparison with wild-type *Zv*FMOa (Fig. 10), while the double variant exhibited a slight increase by a factor of 1.51. Consequently, a variant containing a single Y356A exchange was generated and analysed. The specific activity of this variant was increased by a factor of 2.45 compared with the wild-type enzyme. Therefore, another variant (F354V) was created which likewise contained an amino-acid exchange within the loop limiting the entrance to the binding site. This exchange also resulted in an increased specific activity (by a factor of 2.75). However, combining both favourable amino-acid exchanges does not result in a higher level of specific activity, but rather diminishes their individual effects. Still, the observed specific activity for this double variant is increased by a factor of 1.5 compared with the control experiment. The beneficial effect of a more accessible substrate-entry path might possibly be partly antagonized by a loss of substrate affinity when both bulky aromatic amino acids are exchanged for smaller substituents. Based on our identification of the flexible helix α8, which might have an influence on substrate binding and turnover, we also generated a P388S variant. However, the specific activity of this variant did not differ significantly from that of wild-type *Zv*FMOa, indicating that the presence of proline within the *Zv*FMOa helix does not interfere with its potential role in substrate turnover.

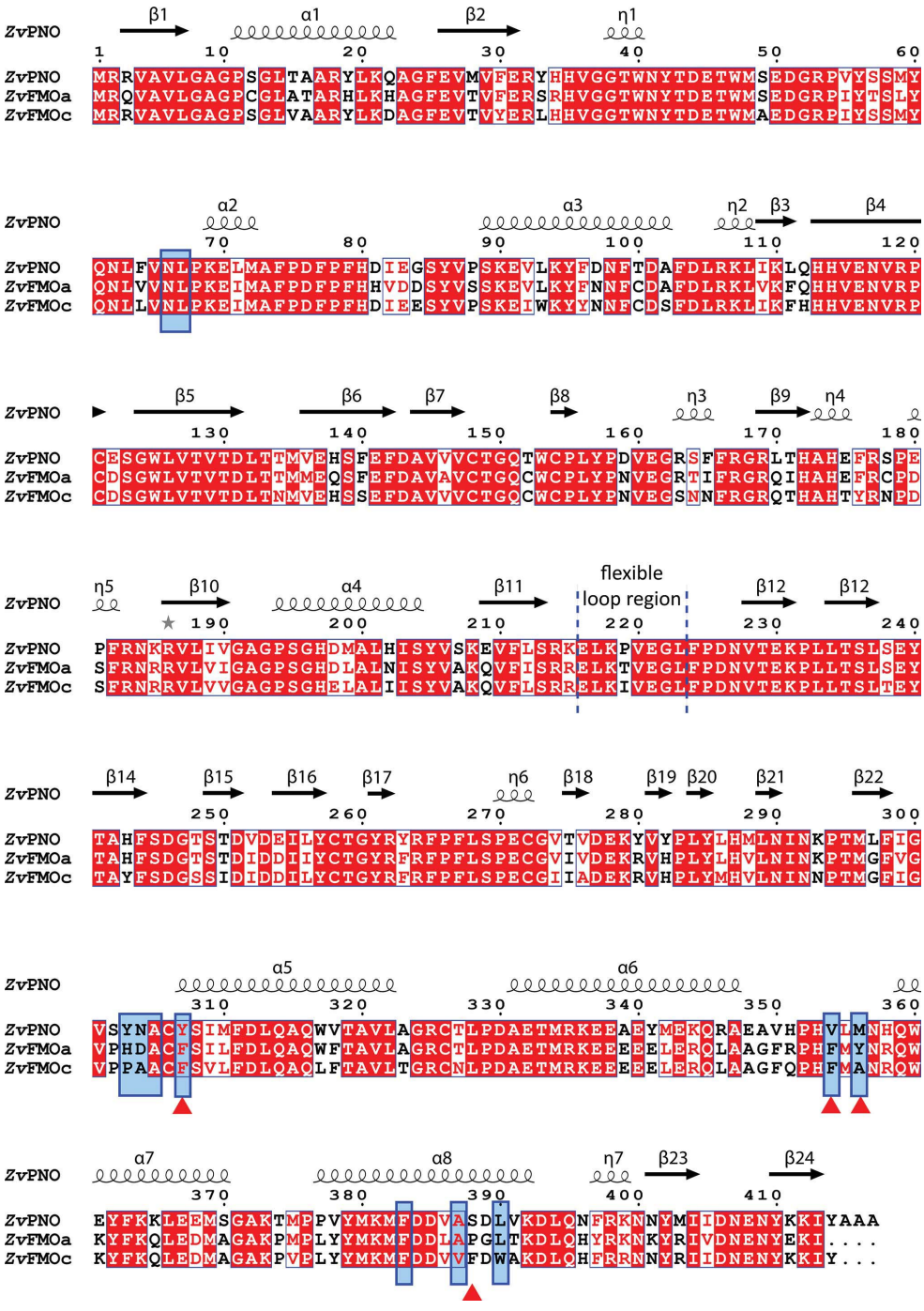


Figure 9
Structure-based sequence alignment of *Zv*FMO isoforms. Red background, identical amino acids; bold letters, similar amino acids; blue boxes, residues forming the predicted substrate-binding pocket. The multiple sequence alignment was performed with *Clustal Omega* (Sievers *et al.*, 2011); the final figure was prepared using *ESPript* (Robert & Gouet, 2014). The secondary structure was derived by *DSSP* (Kabsch & Sander, 1983) and is depicted above the sequences. Residues which were exchanged by site-directed mutagenesis are highlighted by a triangle below the sequence.

4. Discussion

The structure of *Zv*PNO represents that of an FMO from the highest eukaryotic organism currently available. FMOs oxidize a large number of structurally very diverse substrates. The determinants for the selectivity or even specificity for the individual substrates are defined by residues surrounding the substrate-binding path. The specific activity of *Zv*FMO seems to depend on the accessibility of the active site, rather than on specific interactions between the binding pocket and different PA substrates. In order to be able to accommodate a rather diverse set of substrates, a need for rather large structural

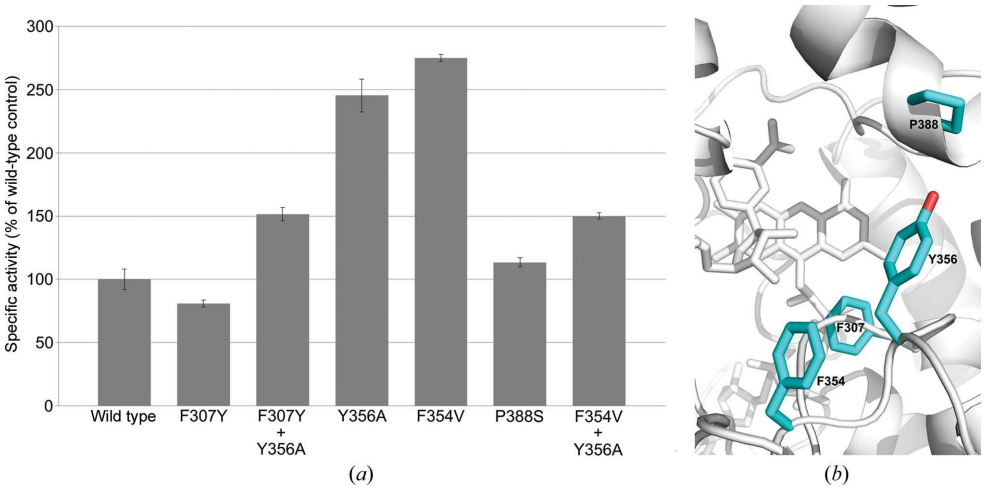


Figure 10
Specific activity of *Zv*FMOa variants towards monocrotaline. (*a*) The enzyme activities of the variants were determined photometrically in triplicate. (*b*) Cyan stick models show the residues in the binding pocket of wild-type *Zv*FMOa which were exchanged to create variants.

plasticity within this pocket can be envisaged. Substrates are presumably not firmly bound to the enzyme but are instantly oxygenated upon reaching the active site. These findings coincide with the broad PA substrate spectrum of *Zv*FMOs as well as the proposed cocked-gun mechanism for these enzymes. It might also be possible that an active C4a-hydroperoxy-FAD intermediate structure may preferably bind the PA ligand compared with the oxidized FAD structure as presented. These circumstances might be the reason for our unsuccessful co-crystallization and soaking experiments and the lack of structural evidence for an enzyme–substrate complex.

Still, the active site and substrate-binding cleft of *Zv*PNO show some flexibility, especially in helix α8, which either reduces or enhances the volume of the binding pocket depending on its bending angle. This flexibility might account for either one of two features: (i) it might contribute to an induced fit to best adapt to different PA substrates or (ii) it might serve as some kind of product-removal mechanism by mechanically pushing the oxygenated PA out of the active site.

Of note, only the space in direct proximity to the reactive flavin moiety is quite limited, while the rest of the substrate-binding pocket is largely solvent-exposed and exhibits few steric boundaries. Therefore, the small active site only accepts small compounds such as molecular oxygen, which is needed for the oxygenation reaction, and is probably selective for structural features that resemble a necine base, which is shared by all PA substrates. Otherwise, there seems to be very limited selectivity regarding accepted substrates apart from the steric hindrance present in the entrance site of the binding cleft. This is backed up by our findings that single amino-acid exchanges within the *Zv*FMOa sequence could significantly increase its specific activity towards one of the sterically most demanding PA substrates, monocrotaline.

Therefore, *Zv*FMOs are not specialists with high efficiency towards a single substrate, but rather generalists which accept a variety of compounds with similar features. This harbours a great advantage for the host organism *Zonocerus*, which can easily adapt to a variety of PAs produced by food plants and thereby circumvent their chemical defence against herbivory as part of the so-called arms race between plants and their attackers (Pieterse & Dicke, 2007). Evolutionary pressure caused by ever-developing PAs in the host plants might be a reason for this organism to retain three isoforms of *Zv*FMOs with differences in their substrate-binding cleft in its genome, although two of them are far less active towards currently known PAs. A few mutations during the course of evolution might turn them into highly active enzymes capable of oxygenating PAs, as our studies of *Zv*FMOa variants have already indicated. However, it cannot be ruled out that the real substrates for isoform A and isoform C have not yet been identified.

Owing to their structural and catalytical properties, the *Zv*FMO isoforms belong to the class B flavoprotein monooxygenases (van Berkel *et al.*, 2006). To date, our structural knowledge about this specific FMO subclass is solely based on FMO crystal structures from bacteria, yeast and fungi (Malito *et al.*, 2004; Eswaramoorthy *et al.*, 2006; Cho *et al.*, 2011; Mirza *et al.*, 2009; Olucha *et al.*, 2011; Yachnin *et al.*, 2012; Leisch *et al.*, 2012; Jensen *et al.*, 2012, 2014; Franceschini, Fedkenheuer *et al.*, 2012; Franceschini, van Beek *et al.*, 2012; Binda *et al.*, 2015; Setser *et al.*, 2014; Ferroni *et al.*, 2016; Romero *et al.*, 2016; Fürst *et al.*, 2017). Therefore, the structures presented here are the first to derive from a more highly developed organism and give insight into the evolutionary conservation of FMOs. Although displaying a unique dimer interface, the overall structure of the FMO subunits has remained largely unchanged throughout evolution. However, the detailed mechanisms of cofactor and substrate binding as well as product release from the active site, which are conducted by minor and major conformational changes within the active site, are still not fully understood. The identification of a kink within the flexible helix α8 presented in this work might contribute to a better future understanding of the complete machinery.

Because of its closer evolutionary relationship, the crystal structure of *Zv*PNO will serve as a more suitable template for the related human enzyme involved in the biotransformation of currently available xenobiotics.

Acknowledgements

We gratefully acknowledge access to the core facilities of the BiMo/LMB of Kiel University. We thank Felix Helfrich, Sebastian Krossa and Christina Hopf for helpful discussions

and support during data collection. Diffraction data were collected on beamlines P13 and P14 operated by EMBL at the PETRA III storage ring. We are grateful to the beamline staff for providing assistance in using the beamlines.

Funding information

The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under BioStructX (grant agreement No. 283570). Additionally, we are grateful for access to the HTX crystallization facility by means of a grant from P-Cube and BioStructX at the EMBL Outstation Hamburg. Beamtime at P14 at the EMBL Outstation Hamburg was also funded by a BioStructX grant. This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant OB162/9-1.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Beaty, N. B. & Ballou, D. P. (1981). *J. Biol. Chem.* **256**, 4619–4625.
Berkel, W. J. H. van, Kamerbeek, N. M. & Fraaije, M. W. (2006). *J. Biotechnol.* **124**, 670–689.
Bernays, E., Edgar, J. & Rothschild, M. (1977). *J. Zool.* **182**, 85–87.
Binda, C., Robinson, R. M., Martin Del Campo, J. S., Keul, N. D., Rodriguez, P. J., Robinson, H. H., Mattevi, A. & Sobrado, P. (2015). *J. Biol. Chem.* **290**, 12676–12688.
Cashman, J. R. (1995). *Chem. Res. Toxicol.* **8**, 166–181.
Cashman, J. R. (2001). *Enzyme Systems that Metabolise Drugs and Other Xenobiotics*, edited by C. Ioannides, pp. 67–93. Chichester: John Wiley & Sons.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cho, H. J., Cho, H. Y., Kim, K. J., Kim, M. H., Kim, S. W. & Kang, B. S. (2011). *J. Struct. Biol.* **175**, 39–48.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Eswaramoorthy, S., Bonanno, J. B., Burley, S. K. & Swaminathan, S. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 9832–9837.
Ferroni, F. M., Tolmie, C., Smit, M. S. & Opperman, D. J. (2016). *PLoS One*, **11**, e0160186.
Fraaije, M. W., Kamerbeek, N. M., van Berkel, W. J. & Janssen, D. B. (2002). *FEBS Lett.* **518**, 43–47.
Franceschini, S., Fedkenheuer, M., Vogelaar, N. J., Robinson, H. H., Sobrado, P. & Mattevi, A. (2012). *Biochemistry*, **51**, 7043–7045.
Franceschini, S., van Beek, H. L., Pennetta, A., Martinoli, C., Fraaije, M. W. & Mattevi, A. (2012). *J. Biol. Chem.* **287**, 22626–22634.
Fu, P. P., Xia, Q., Lin, G. & Chou, M. W. (2004). *Drug Metab. Rev.* **36**, 1–55.
Fürst, M. J., Savino, S., Dudek, H. M., Gómez Castellanos, J. R., Gutiérrez de Souza, C., Rovida, S., Fraaije, M. W. & Mattevi, A. (2017). *J. Am. Chem. Soc.* **139**, 627–630.
Hao, D. C., Chen, S. L., Mu, J. & Xiao, P. G. (2009). *Genetica*, **137**, 173–187.
Hartmann, T., Biller, A., Witte, L., Ernst, L. & Boppré, M. (1990). *Biochem. Syst. Ecol.* **18**, 549–554.
Hartmann, T. & Ober, D. (2008). *Induced Plant Resistance to Herbivory*, edited by A. Schaller, pp. 213–231. Dordrecht: Springer.
Hartmann, T. & Witte, L. (1995). *Alkaloids: Chemical and Biological Perspectives*, edited by S. W. Pelletier, Vol. 9, pp. 155–233. New York: Elsevier.
Hille, R., Miller, S. & Palfey, B. (2013). Editors. *Handbook of Flavoproteins.* Berlin: De Gruyter.
Jensen, C. N., Ali, S. T., Allen, M. J. & Grogan, G. (2014). *J. Mol. Catal. B Enzym.* **109**, 191–198.
Jensen, C. N., Cartwright, J., Ward, J., Hart, S., Turkenburg, J. P., Ali, S. T., Allen, M. J. & Grogan, G. (2012). *Chembiochem*, **13**, 872–878.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langel, D. & Ober, D. (2011). *Phytochemistry*, **72**, 1576–1584.
Laskowski, R. A. & Swindells, M. B. (2011). *J. Chem. Inf. Model.* **51**, 2778–2786.
Lawrence, M. C. & Colman, P. M. (1993). *J. Mol. Biol.* **234**, 946–950.
Leisch, H., Shi, R., Grosse, S., Morley, K., Bergeron, H., Cygler, M., Iwaki, H., Hasegawa, Y. & Lau, P. C. (2012). *Appl. Environ. Microbiol.* **78**, 2200–2212.
Malito, E., Alfieri, A., Fraaije, M. W. & Mattevi, A. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 13157–13162.
Mirza, I. A., Yachnin, B. J., Wang, S., Grosse, S., Bergeron, H., Imura, A., Iwaki, H., Hasegawa, Y., Lau, P. C. K. & Berghuis, A. M. (2009). *J. Am. Chem. Soc.* **131**, 8848–8854.
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.
Narberhaus, I., Theuring, C., Hartmann, T. & Dobler, S. (2003). *J. Comp. Physiol. B*, **173**, 483–491.
Naumann, C., Hartmann, T. & Ober, D. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 6085–6090.
Olucha, J., Meneely, K. M., Chilton, A. S. & Lamb, A. L. (2011). *J. Biol. Chem.* **286**, 31789–31798.
Phillips, I. R. & Shephard, E. A. (2008). *Trends Pharmacol. Sci.* **29**, 294–301.
Pieterse, C. M. & Dicke, M. (2007). *Trends Plant Sci.* **12**, 564–569.
Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
Robinson, R., Badieyan, S. & Sobrado, P. (2013). *Biochemistry*, **52**, 9089–9091.
Romero, E., Castellanos, J. R., Mattevi, A. & Fraaije, M. W. (2016). *Angew. Chem. Int. Ed.* **55**, 15852–15855.
Sehlmeyer, S., Wang, L., Langel, D., Heckel, D. G., Mohagheghi, H., Petschenka, G. & Ober, D. (2010). *PLoS One*, **5**, e10435.
Setser, J. W., Heemstra, J. R. Jr, Walsh, C. T. & Drennan, C. L. (2014). *Biochemistry*, **53**, 6063–6077.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.
Söding, J. (2005). *Bioinformatics*, **21**, 951–960.
Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Wang, L., Beuerle, T., Timbilla, J. & Ober, D. (2012). *PLoS One*, **7**, e31796.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yachnin, B. J., Sprules, T., McEvoy, M. B., Lau, P. C. K. & Berghuis, A. M. (2012). *J. Am. Chem. Soc.* **134**, 7788–7795.
Ziegler, D. M. (1993). *Annu. Rev. Pharmacol. Toxicol.* **33**, 179–199.
Ziegler, D. M. (2002). *Drug Metab. Rev.* **34**, 503–511.