Supporting Information

Molecular Basis for Spirocycle Formation in the Paraherquamide Biosynthetic Pathway

Amy E. Fraley, Kersti Caddell Haatveit, Ying Ye, Samantha P. Kelly, Sean A. Newmister, Fengan Yu, Robert M. Williams, Janet L. Smith, K. N. Houk, David H. Sherman

Experimental Methods:

Fungal strains and culture conditions

Penicillium fellutanum ATCC 20841 spores were generated on YPD agar plates over the course of 7 days. Spores were harvested into 5 mL sterile water per plate by gently scraping the surface of the culture with a sterile inoculating loop. Spores were stored at -80°C until ready to use. Genomic DNA was harvested using Wizard Genomic DNA Purification Kit from Promega.

cDNA preparation and cloning of *phqK*

Genomic DNA was extracted from *Penicillium fellutanum* statically cultivated in liquid medium (50% seawater with 2.0% malt extract and 0.5% peptone) at 28°C. To generate *phqK*, introns were predicted using Softberry Fgenesh-M, and the *phqK* exons were amplified from genomic DNA using overlapping PCR with primers in Table S1. Amplified genes were cloned into a pKLD116 vector using restriction enzyme digest and ligation.[1] *E. coli* DH5α were transformed with pKLD116-*phqK* for screening and plasmid maintenance.

Gene disruption of *phqK*

Gene disruption in *Penicillium simplicissimum* was performed using the CRISPR/Cas9 system for filamentous fungi.[2] For the preparation of the *in vitro* transcriptional gRNA, the gRNA cassettes containing the T7 promoter, the protospacer sequence, and the synthetic gRNA scaffold for targeting genes were PCR amplified from the plasmid pFC333 as template, using the primers listed in Supplementary Table 18, and inserted into pFC332 to generate the plasmid pFC332-*phqK*. For transformation of *Penicillium simplicissimum*, the strain was inoculated into 100 mL yeast extract peptone dextrose (YPD) medium and cultivated at 28 °C, 200 rpm, for two days. The mycelia were collected and digested using Vinoflow (64 mg/mL). The resulting protoplasts were then separated from mycelia by filtration and washed with STC solution (0.8 M sorbitol, 0.05 M Tris-HCl, 0.05 M $CaCl_2$, pH 8), and diluted to a concentration of 2×10^8 cells mL^{-1}. The circular plasmid was added to the 200 µL protoplast solution, and incubated on ice for 30 minutes, then blended with 2 mL 30% PEG solution (40% PEG8000, 50 mM $CaCl_2 \cdot 2H_2O$, 10 mM Tris-HCl, pH 8.0) and incubated at room temperature for 20 minutes. The resulting solution was then diluted with STC solution and distributed on selective PGA plates (PG broth, 1.2 M sorbitol, 100 µg/mL hygromycin B, 1.5% agar). The plates were incubated at 30 °C for 5-7 days. The colonies grown from these selective plates were cultured (stationary) in Czapek yeast agar (CYA) medium (1L containing Difco Czapek-Dox 35 g, yeast extract 5 g, $CuSO_4 \cdot 5H_2O$ 5 mg, $ZnSO_4 \cdot 7H_2O$; pH 6.3) for 7 days and analyzed by time of flight mass spectrometry (TOF-MS).

Extraction of Δ*phqK* strain of *Penicillium simplicissimum* and purification of paraherquamide K (14) and paraherquamide L (15)

The strain of *P. simplicissimum* produces 1.5 mg/mL of paraherquamide A.[3] For the experiments reported in this work, *P. simplicissimum* was grown stagnant for 7 days at 28°C on Czapek Yeast Autolysate media (CYA, 1L: 35g Difco Czapek-Dox broth, 5g yeast extract, 0.02 mM $CuSO_4$, and 0.035 mM $ZnSO_4$, pH 6.3)). The culture was filtered to obtain the liquid medium which was extracted with XAD16N resin. After treating the resin with methanol, the eluted extract was dried and further extracted with ethyl acetate. The ethyl acetate fractions were dried to generate the crude extract.

The crude extract was purified by gravity on a C18 column. The purification gradient started with two column volumes of 40% methanol, followed by three column volumes of 65% methanol and three column volumes of

90% methanol, at which point paraherquamide K (**14**) and L (**15**) eluted from the column. The samples were dried and further purified by chiral HPLC (Phenomenex Lux 5 µm Cellulose-3 250 x 10 mm column). The following HPLC time program was used for separation and purification of **14** and **15**: 40% acetonitrile for 0.5 minutes, gradient to 80.5% acetonitrile over 25 minutes, 80.5% acetonitrile for 0.5 minutes, 90% acetonitrile for 5 minutes, 40% acetonitrile for 10 minutes, at a flow rate of 3 mL/min. From 1L of culture, 2.1mg of paraherquamide K (**14**) and 1.8mg paraherquamide L (**15**) were obtained.

***Malbranchea aurantiaca* culture conditions and extraction of malbrancheamides**

M. aurantiaca was grown and extracted as in Fraley et al.[4] Individual flasks of 75 mL potato dextrose broth were inoculated with 100 µL spore stock of *M. aurantiaca* and grown for three weeks, or until a white fungal mat was produced. Prior to the noticeably orange sporulation, the cultures were pulverized and extracted with dichloromethane. The crude extract was acid-base purified first with 1 M HCl, then neutralized with 2 M ammonium hydroxide to pH 9, and back extracted with dichloromethane. The extract was then purified by chiral HPLC with a Phenomenex Lux 5 µm Cellulose-3 250 x 10 mm column. The following HPLC time program was used for separation and purification of the malbrancheamide compounds: 50% acetonitrile for 18 minutes, gradient to 55% acetonitrile over 2 minutes, 55% acetonitrile for 2 minutes, gradient to 40% acetonitrile over 2 minutes, 40% acetonitrile for 5 minutes, at a flow rate of 4 mL/min. The mobile phase consisted of water and acetonitrile.

Overexpression and purification of PhqK for enzymatic reactions

The *Escherichia coli* BL21 *pRARE* transformant containing *pKLD116-MBP-phqK* was grown at 37°C overnight in LB media containing 50 µg/mL of ampicillin and 100 µg/mL of spectinomycin. 10 mL of culture were used to inoculate 1 L of TB media containing the aforementioned concentrations of antibiotic and 4% glycerol. Cells were grown at 37°C for roughly 4 hours until A_{600} reached 0.6-1.0, and isopropyl β-D-thiogalactoside (IPTG, 0.2 mM) and riboflavin (50 µM) were added to induce protein overexpression overnight at 18°C.

All purification steps were conducted at 4°C. Briefly, 1 L of expression culture was spun down at 5,500 rpm. The harvested cell pellet was resuspended in 35 mL of lysis buffer (10 mM imidazole pH 8, 50 mM NaH_2PO_4, 300 mM NaCl, 10% v/v glycerol, adjusted to pH 8) with the addition of 10 mg lysozyme, 4 mg DNase, 50 µM flavin adenine dinucleotide (FAD), 2 mM $MgSO_4$ and lysed by sonication. Insoluble material was removed by centrifugation at 20,000 rpm for 30 minutes, and the supernatant was filtered. The enzyme was purified through metal affinity chromatography with Ni^{2+}-NTA resin (Novagen) which was equilibrated with lysis buffer. The protein-bound resin was washed with 50 mL of lysis buffer, 50 mL of wash buffer (20 mM imidazole pH 8, 50 mM NaH_2PO_4, 300 mM NaCl, 10% v/v glycerol, adjusted to pH 8), and finally 10 mL of elution buffer (250 mM imidazole pH 8, 50 mM NaH_2PO_4, 300 mM NaCl, 10% v/v glycerol, adjusted to pH 8). Protein in the eluate was exchanged into storage buffer (10 mM HEPES pH 7.6, 50 mM NaCl, 0.1 mM EDTA, 0.2 mM TCEP, 10% v/v glycerol) using a PD-10 column. Samples were then flash frozen with liquid N_2 and stored at -80°C.

Purification of PhqK for crystallization and kinetic analysis.

A cell pellet from a 1L expression culture was re-suspended in 30 mL lysis buffer (10 mM HEPES pH 8, 10 mM imidazole pH 8, 300 mM NaCl, 10% glycerol) and supplemented with 50 µM FAD. Cell lysis was accomplished through addition of 10 mg lysozyme, 4 mg DNase, 2 mM $MgSO_4$, and sonication. Cell waste was cleared through centrifugation (18,000 rpm for 25 minutes). The supernatant was filtered and PhqK was purified through metal affinity chromatography on a 5 mL His-Trap column (GE Healthcare) with a 10-column volume gradient of elution buffer (10 mM HEPES pH 8, 280 mM imidazole pH 8, 300 mMM NaCl, 10% glycerol). The MBP and tag-free PhqK were separated from TEV protease and any remaining His_6-MBP-PhqK by metal affinity chromatography, and dialyzed into storage buffer (20mM HEPES pH 7.6, 300 mM NaCl, 10% glycerol).

Crystallization of PhqK protein

The purified PhqK was dialyzed overnight into 20 mM HEPES pH 7.6 buffer with 300 mM NaCl to remove glycerol and then supplemented with 2x FAD. For active site complexes with malbrancheamide B (**21**),

paraherquamide K (**14**), and paraherquamide L (**15**), the protein was incubated with 5x concentration of substrate prior to crystallization. For active site complexes with malbrancheamide C (**22**), the protein was incubated with 2x concentration of substrate prior to crystallization. Crystals were grown by vapor diffusion from 1:1 mixture of 8-10 mg/mL PhqK preincubated with substrate and a well solution containing 25% PEG 3350, 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 5.5, and 2% 2,2,2-trifluoroethanol. Crystals with no substrate bound were produced in a similar manner, except without the addition of substrate. Crystals were cyoprotected in well solution at 30% PEG 3350 and flash-cooled in liquid nitrogen.

Data collection for x-ray crystallography

Data were collected at GM/CA beamline 23ID-B at the Advanced Photon Source (APS) at Argonne National Laboratory. For the SeMet-PhqK crystal, 360º of diffraction data were collected in inverse-beam geometry using 30º wedges. All data were processed using XDS.[5] The SeMet PhqK monooxygenase structure was solved by single-wavelength anomalous diffraction (SAD) using AutoSol in the Phenix suite to locate the Se sites, determine initial phases and perform density modification (figure of merit = 0.320).[6] AutoBuild in the Phenix suite was used to build an 82% complete starting model. The SeMet PhqK model was used as a template in molecular replacement to solve the native PhqK structures using Phaser in the Phenix suite. A progression of model building and refinement were carried out to complete the models using Coot and Phenix Refine with seven translation/libation/screw groups.[7]

Enzymatic reactions and HPLC analysis of PhqK

The standard enzyme assay containing 200 µM FAD, 200 µM substrate, 5 mM NADH, and 40 µM enzyme in 50 µL reaction buffer (10 mM HEPES pH 7.6, 50 mM NaCl, 0.1 mM EDTA, 0.2 mM TCEP 10% v/v glycerol, pH 7.6) was performed at 28°C for 2 hours. The reactions were quenched with 100µL LC/MS grade methanol and centrifuged to remove solid material. The samples were analyzed on a Shimadzu HPLC using a Phenomenex Lux 5 µm Cellulose-3 LC column (250x4.6mm) with the following time program: 30% acetonitrile for 1 minute, 30-95% acetonitrile over 15 minutes, 95% acetonitrile for 1 minute, 95-30% acetonitrile over 1 minute, and 30% acetonitrile for 7 minutes. The flow rate was 1.5 mL/min and the reactions were monitored at 240nm.

Enzymatic reactions and HPLC analysis of MalA bromination reactions

The standard enzyme assay containing 100 µM FAD, 50 µM NaBr, 250 µM substrate, 5 mM NADH, 40 µM MalA, and 160 µM HpaC in 50 µL reaction buffer (50 mM NaH_2PO4, 250 mM imidazole pH 7, 100 mM NaBr, 10% glycerol, 0.2 mM TCEP) was performed at 28°C for 2 hours. The reactions were quenched with 100µL LC/MS grade methanol and centrifuged to remove solid material. The samples were analyzed on a Shimadzu HPLC using a Phenomenex Lux 5 µm Cellulose-3 LC column (250x4.6mm) with the following time program: 20% acetonitrile for 1 minute, 20-95% acetonitrile over 15 minutes, 95% acetonitrile for 1 minute, 95-20% acetonitrile over 1 minute, and 30% acetonitrile for 7 minutes. The flow rate was 1.5 mL/min and the reactions were monitored at 240nm.

Enzymatic reactions and HPLC analysis for determination of PhqK kinetics

The standard enzyme assay containing 200 µM FAD, varying concentrations of substrate between 20 and 700 µM, 5 mM NADH, and 1 µM enzyme in 250 µL reaction buffer (10 mM HEPES pH 7.6, 50 mM NaCl, 0.1 mM EDTA, 0.2 mM TCEP 10% v/v glycerol, pH 7.6) was performed at 28°C for varying time points (4, 10, 20, 30 minutes). At each time point, 50 µL of the reaction mix were quenched with 100µL ethyl acetate, vortexed, and extracted. The samples were resuspended in 50 µL methanol and analyzed on a Shimadzu HPLC using a Phenomenex Lux 5 µm Cellulose-3 LC column (250x4.6mm) with the following time program: 15-75% acetonitrile over 7 minutes, 95% acetonitrile for 2.5 minutes, 75-15% acetonitrile over half a minute, and 15% acetonitrile for 5 minutes. The flow rate was 1.4 mL/min and the reactions were monitored at 240nm.

PhqK enzymatic reactions for substrate conversion studies

The enzyme activity was analyzed in the presence of catalase and a cofactor regeneration system (Glucose-6-phosphate dehydrogenase, G-6-P DH) to mimic an environment similar to the producing organism. The reactions involving G-6-P DH contained the following components: 40 μM FAD, 200 μM substrate, 100 μM $NADP^+$, 1 unit G-6-P-DH, 10 mM G-6-P, and 40 μM PhqK in 50 μL reactions in the reaction buffer from above. Comparison reactions contained no G-6-P DH/G-6-P/$NADP^+$ and instead contained 5 mM NADH or NADPH. The negative controls contained either no PhqK or no NADPH/NADH source. The reactions were run for two hours at 28°C and quenched with 100 μL methanol prior to analysis by HPLC using the same method as above. The presence of the regeneration system increased the conversion of malbrancheamide (**19**) by PhqK, but the conversions of the native substrates paraherquamide K (**14**) and L (**15**) were unaffected.

The reactions involving catalase contained the following components: 200 μM FAD, 200 μM substrate, 5 mM NADH, 40 μM PhqK, and 5 mM catalase (0.3 mg/mL) in 50 μL reactions in the reaction buffer from above. A comparison reaction contained no catalase and the negative control contained no PhqK. The reactions were run for two hours at 28°C and quenched with 100 μL methanol prior to HPLC analysis using the same method as above. The addition of catalase increased the reactivity of the enzyme toward the non-native substrate malbrancheamide (**19**), but the conversions of the native substrates paraherquamide K (**14**) and L (**15**) were unaffected.

These experiments were performed to optimize a system for high conversion that may be similar to the enzyme's natural environment in the producing organism. While, the scope of this work involved the characterization of the stand-alone enzyme, it is possible that with the aid of other enzymes commonly found in fungi, such as catalase,[8] PhqK may be more efficient at large scale-conversion of non-native substrates.

Enzymatic reactions to analyze reactivity of hydrogen peroxide generated as a byproduct

The spirocyclization of malbrancheamide by hydrogen peroxide was analyzed to ascertain the role of this byproduct in epoxidation. The reactions contained 200 μM malbrancheamide (**19**) and varying concentrations of hydrogen peroxide (0.2, 0.25, 0.5, and. 1.0 mM) in 50 μL reactions in the reaction buffer from above. The negative control contained no hydrogen peroxide. The reactions were run for two hours and quenched with 100 μL methanol prior to analysis by HPLC using the same method as above. Under all conditions, no conversion was observed.

Computational Details

DFT calculations: All quantum mechanical calculations were performed with Gaussian 09.[9] Geometry optimizations were calculated with the M06-2X[10] density functional with the IEFPCM model (water)[11], and the 6-31G(d) basis set. Single point energies were calculated using M06-2X[10] with the IEFPCM model (water)[11], and the 6-311++G(d,p) basis set. Previous computational work on epoxide-opening reactions with similar methods provided results that aligned with experiements.[12, 13] Monte Carlo conformational searches were performed to identify lowest energy conformations with the OPLS3 force field[14] in Maestro/Macromodel. Structure graphics were generated using CYLview.[15]

Molecular dynamics simulations: Molecular dynamics (MD) simulations were performed using the GPU code (pmemd)[16] of the AMBER 16 package.[17] Parameters for FAD cofactor and paraherquamides K-L (**14**, **15**) were generated within the antechamber module using the general AMBER force field (gaff),[18] with partial charges set to fit the electrostatic potential generated at the HF/6-31G(d) level by the RESP model.[19] The charges were calculated according the Merz-Singh-Kollman scheme[20, 21] using the Gaussoan 09 package.[9] Each protein was immersed in a pre-equilibrated truncated cuboid box with a 10 Å buffer of TIP3P[19] water molecules using the leap module. The systems were neutralized by addition of explicit counter ions (Na^+ and Cl^-). All subsequent calculations were performed using the Amber14 force field (ff14sb).[22] The substrate and enzyme were optimized for a total of 1,000,000 steps, with 750,000 steepest descent steps and 250,000 conjugate gradient steps. The systems were gently heated using six 50 ps steps, incrementing the temperature by 50 K for each step (0-300 K)

under constant-volume and periodic-boundary conditions. Water molecules were treated with the SHAKE algorithm such that the angle between hydrogen atoms were kept fixed. Long-range electrostatic effects were modelled using the particle-mesh-Ewald method.[23] An 8Å cutoff was applied to Lennard-Jones and electrostatic interactions. Harmonic restraints of 30kcal/(mol $Å^2$) were applied to the solute and the Andersen equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2 fs time step at a constant volume. Production trajectories were then run for an addition 1500 ns under the same simulation conditions.