Supplementary Information for

Discovery of a regioselectivity switch in nitrating P450s guided by MD simulations and Markov models

Sheel C. Dodani[a], Gert Kiss[b], Jackson K. B. Cahn[a], Ye Su[a], Vijay S. Pande[b]*, and Frances H. Arnold[a]*

[a]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, California, 91125, USA. [b]Department of Chemistry, SIMBIOS NIH Center for Biomedical Computation, and Center for Molecular Analysis and Design, Stanford University, 318 Campus Drive, Stanford, California, 94305, USA.

*Email: frances@cheme.caltech.edu, pande@stanford.edu

Table of Contents

Supplementary Computational Methods and Figures

Reconstruction of the F/G loop

We used the protein structure prediction tool 'Prime' from Schrödinger Inc.[1] to reconstruct the missing residues (176–183) of the TxtE F/G loop in the co-crystal with the PDB accession ID 4TPO (Supplementary Fig. S1). The resulting model was equilibrated with ten independent 5 ns molecular dynamics runs while all other residues were restrained to their crystallographic positions with a weak 5 kcal/mol harmonic potential. A total of one thousand snapshots were extracted at equidistant intervals from the trajectories and their geometries were optimized. Of these, the 100 lowest energy structures were chosen as starting points for the subsequent exploration of TxtE's conformational state space. Supplementary Fig. S1c gives an overview of the resulting structural loop diversity across the set.

MD simulations

All simulations, wildtype and variants, were initiated from the open-lid conformation with the substrate indole in its unflipped orientation. Force field parameters of the L-Trp ligand and of the peroxynitrite-heme were generated with the antechamber module of AMBER 12[2] using the general AMBER force field, with partial charges set to fit the electrostatic potential generated at HF/6-31G* by RESP.[3,4] The charges were calculated according to the Merz-Singh-Kollman scheme[5,6] using Gaussian 03.[7] The parameterized L-Trp and the ferric peroxynitrite-heme species were superimposed onto their respective coordinates in the 4TPO crystal structure. The heme-iron was parameterized in the presence of an axial methanethiol mimic of the coordinated Cys357. In the simulation setup, the methanethiol-carbon replaced CB of Cys357 and was covalently linked to the CA of Cys357, while distributing the remaining partial charge difference (-0.013182) across both carbons. This step was repeated for each of the 100 F/G loop conformational replicas. Each structure was immersed in an orthorhombic box of TIP3P[8] waters

with a 10 Å buffer, resulting in the addition of 13,800 solvent molecules on average. Each system was neutralized through the addition of counter ions. All subsequent simulations were performed using the Amber99sb-ildn force field.[9] The geometries were optimized in two stages, initially minimizing the positions of water molecules and ions, followed by an unrestrained minimization of all atoms. Each system was heated in six 50 K, 50 ps steps from 0 to 300 K at constant volume periodic boundary conditions. Harmonic restraints of 10 kcal/mol were applied to the solute, and the Langevin equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing any potential inhomogeneity to self-adjust. Each system was then equilibrated for 2 ns with a 2 fs time step at a constant pressure of 1 atm. Water molecules were triangulated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. An aggregate 110 μs of production MD simulations was run for each of the systems. Long-range electrostatic effects were modeled using the particle-mesh-Ewald method.

Markov state models

Studying the functional dynamics of TxtE presented us with the challenge of an unresolved F/G loop in the open-lid crystal structure and the inexistence of a catalytically competent closed state structure. As such, our study consisted of a significant exploratory component, which required that we first map out the conformational landscape of TxtE prior to investigating the details of its functional interactions. Since the open-to-closed transitions of the F/G loop occur on the microsecond timescale, attempting to solve this task with a single or even a handful of individual 0.5 microsecond MD trajectories proved insufficient and necessitated that we approach the conformational exploration on a larger scale - which is where the construction of Markov state models (MSMs) becomes advantageous.

Compared to conventional techniques, MSMs are particularly powerful when confronted with sparse structural data and in our work they were instrumental at organizing the vast raw data from these simulations into discrete conformational states. This allowed us to build a map of conformational states with insight into how these states are connected to one another. It enabled us to zero in on the conformational transition state, that separates open- from closed-lid states, and to identify a specific residue that is central to that transition, determine what type of mutation would be necessary to slow down this transition and also stabilize the closed-lid conformation of TxtE.

Attempting the same without MSMs would have been a far less tractable and non-trivial, but also non-obvious task. For instance, one would have to define a conformational coordinate that most accurately separates open- and closed-lid states, then ask which trajectories most closely follow this coordinate, and then try to isolate the event that constitutes the conformational transition state between open and closed. Constructing MSMs, on the other hand, is general and agnostic to user-defined 'reaction coordinates'.

Intermittently, as well as post-production, we generated L1-regularized reversible Gaussian hidden Markov models (HMMs)[10] to subdivide and cluster the protein dynamics data set. The feature space was trained on residues that outline the active site – including the B/C and F/G loops, which allowed us to monitor progress in the exploration of the most dynamic features of TxtE. Each of the 100 systems was simulated for an initial 500 ns, which was followed by three consecutive 200 ns rounds of equally weighted adaptive sampling in which new trajectories were seeded with equal probability from each of the explored conformational states. The final dataset was subdivided into 40 HMM initial states that were combined into 9 unique conformational states based on redundancies within the active site interaction network within the

set of 40. We populated a transition matrix with the counts of i-j transitions and applied transition path theory (TPT) to examine the connectivity of the conformational states. We used MSMExplorer[11] to visualize the populations of and flux through the interaction network, as well as to generate Fig. 2c. The thickness and shading of the arrows in Fig. 2c indicates the flux of MD trajectories between each pair of conformations. The color-depth of each state is coupled to the arrows and indicates the amount of flux that passes through a given state, ranging from white (low flux) to deep blue (maximum flux). The size of the each state label reflects its equilibrium population. The similarity between states (measured as the active site RMSD) is mapped onto the transition path diagram and is expressed as the Euclidian distance between state labels.

Since we began our study, the statistical models that were employed for the MD data analysis here (and many more) have become part of a vast python package (MSMBuilder) that is available free of charge at msmbuilder.org.

The hidden Markov state models (HMSMs) reported here were constructed with mixtape v.0.2.2, a predecessor code to the current HMSM implementation in MSMBuilder, which is available for download at http://zenodo.org/record/12638#.Vjtef7Rx1N0. The first step in this process is to featurize our MD trajectories into a vector space that consists of a subset of the raw cartesian coordinates that we care about. We focused on the active site residues, including the Fe-OONO heme and the substrate and specified this by creating a text file that contains the corresponding atom indices as a single vector (atomindices.dat). This step can be carried out on a UNIX command line with

grep " 59 \| 60 … \| 407 \| 408 " TxtE.pdb | awk '{print $2}' > "atomindices.dat"

where 59, 60, …, 407, and 408 correspond to the residue entries in TxtE.pdb that we care about.

The mixtape command for the subsequent featurization step is:

hmsm featurizer --top TxtE.pdb -a atomindices.dat -o features.pickl

In the next step, the features.pickl will be used to generate a hidden Markov state model (for a quick introduction to HMSMs, see http://msmbuilder.org/3.3.0/hmm.html). The mixtape command is

hmsm fit-ghmm --featurizer features.pickl -k $states -l $lagtime –dir TrajDir --ext lh5 --top TxtE.pdb

where $states requests the number of HMM states to be generated and $lagtime defines the lagtimes for which to generate Markov models. Both variables are specified in the form of a string that consist of space-separated values (e.g. states="2 4 6 8 10" to request 2, 4, 6, 8, and 10 state models to be built). The recommended approach to generating meaningful Markov state models is to build MSMs for an NxN matrix of states and lagtime combinations, plot the computed relaxation timescales for each MSM against the lagtimes and begin analyzing those state-to-lagtime combinations beyond which the relaxation timescale flattens off. A good introduction to this at the example of an alanine dipeptide simulation can be found here http://msmbuilder.org/3.3.0/examples/hmm-and-msm.html.

Any of the state/lagtime combinations can be inspected by pulling structures from the state ensemble (either from the mean or stochastically). Mixtape v.0.2.2 requires to first write out a .csv file for this:

hmsm means-ghmm -i hmsm.jsonlines --featurizer features.pickl --n-states 8 --n-per-state 20 --lag-time 100 -o csvfile.csv –dir ../Trajectories --ext lh5 --top TxtE.pdb

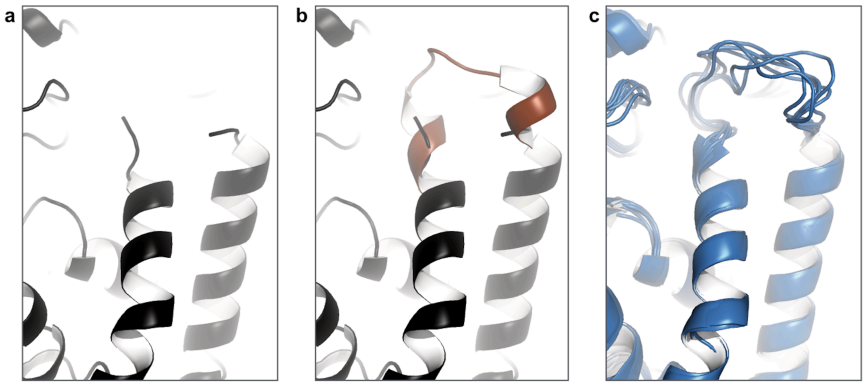
where a .csv file is written with which we can pull 20 structures from the mean of each of the states of an 8 state HMSM that was build with a lagtime of 100. Next up the .csv file can be used to extract the designated set of structures from the requested HMSM

hmsm structures --ext pdb --top TxtE.pdb csvfile.csv

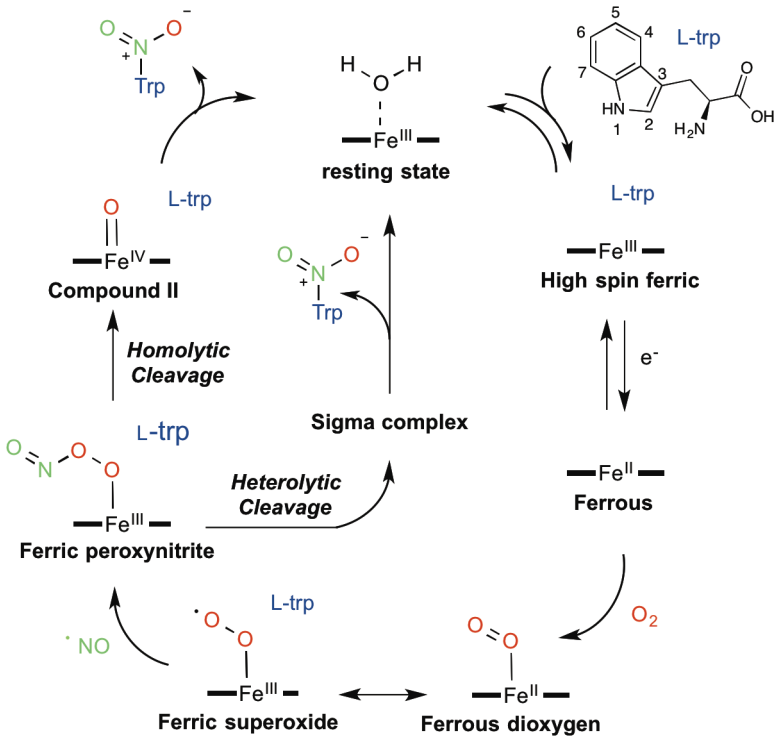
where 20 .pdb files will be generated from the mean of each of the 8 states specified, as specified by the .csv file. The resulting .pdb files can be visualized with a variety of tools, such as PyMOL or VMD.

PyMOL Session File

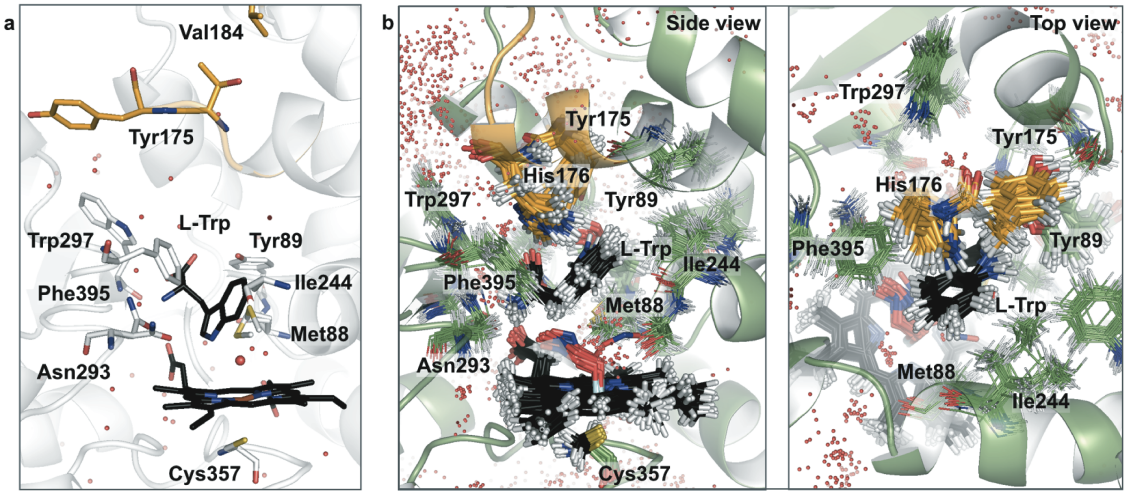
Upon opening of the PyMOL session file that is provided as part of the Supplementary Information, the view is that of the open-lid TxtE shown in Fig. 1b. The structural ensemble at hand consists of 20 frames that were extracted from the O1 open-lid conformation in Figure 2c. Note that the 'TxtE_open-lid_fromMD' object field is highlighted in the right-hand panel. The closed-lid ensemble can be displayed by pressing the 'F2' key (Fn+F2 on a Mac), or by first selecting the 'TxtE_closed-lid_fromMD' object, and then unselecting 'TxtE_open-lid_fromMD'. Pressing 'F1', will again display the open-lid ensemble. The PyMOL file also contains the crystal structures of WT, H176F, and H176Y TxtE. These can be selected by simply clicking on the corresponding objects in the right-hand menu.



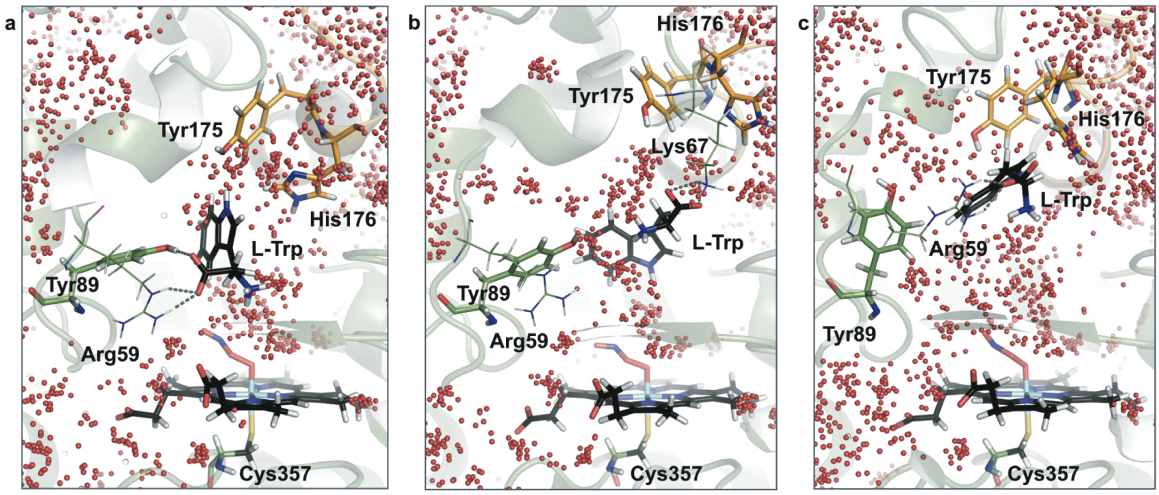
Supplementary Figure S1| F/G-loop reconstruction. **a,** Residues 176–183 of the F/G loop are missing from the co-crystal of TxtE with the PDB accession ID 4TPO. **b,** Reconstructed loop, using Prime.[1] **c,** Equilibration of the loop through short MD simulations gives a structural ensemble from which the production MD simulations on the full TxtE system were initiated.



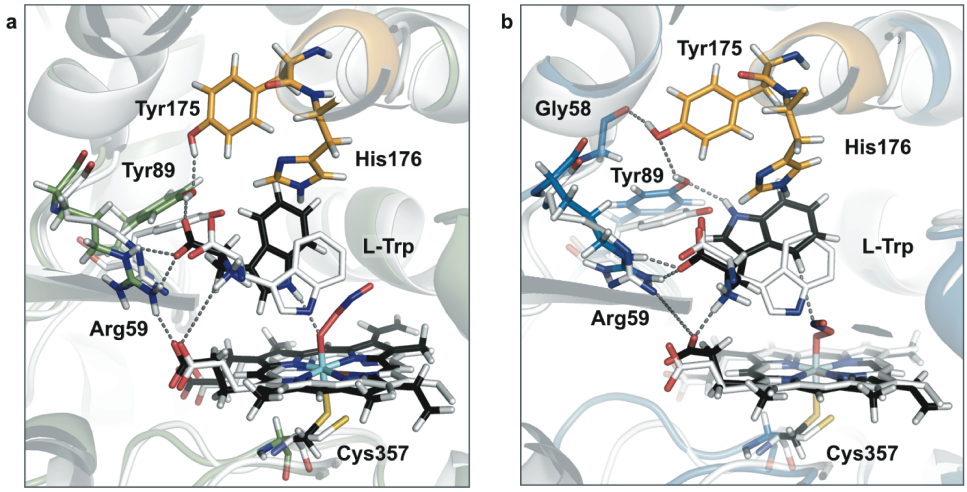
Supplementary Figure S2| Proposed mechanism of TxtE catalyzed nitration adapted from Barry et al.[12] The bold, horizontal line represents the cysteine ligated heme cofactor.



Supplementary Figure S3| Large-scale MD simulations of TxtE reveal direct interactions between the structurally unresolved F/G loop (orange) and the L-Trp substrate (black). a, Active site of the TxtE co-crystal structure (PDB ID: 4TPO) viewed from the side. **b**, MD structural ensemble of the closed-lid TxtE active site. His176 and the substrate engage in an edge-to-face interaction that pins the substrate-indole ring against Met88 and Ile244, aligning it proximal to the ferric peroxynitrite intermediate. The close fit supports only subtle vibrations about the optimally packed side chain orientations throughout the entire active site.



Supplementary Figure S4| The ensemble of open-lid TxtE structures features a solvent-accessible active site that supports various non-productive substrate binding modes. These range from poses in which substrate and ferric peroxynitrite-intermediate are separate by a solvent layer **(a)**, to inverted placements in which the amino acid moiety of the substrate binds to Lys67, instead of Arg59 **(b)**, and to partially bound modes that place the substrate at the opening of the solvent-flooded active site **(c)**. These non-productive binding modes enable transitions between the flipped and un-flipped substrate indole moiety. Substrate binding residues (e.g. Arg59, Lys67) are depicted as lines, the substrate and heme are in black stick rendition, the F/G loop is shown in orange, and the positions of water molecules are represented as clusters of red spheres.



Supplementary Figure S5| The ferric peroxynitrite intermediate active site supports two distinct orientations of the substrate indole moiety without jeopardizing the binding contacts with its amino acid moiety. **a,** The indole-NH is in contact with the Fe-bound peroxynitrite-oxygen, and resembles the orientation evident in the TxtE co-crystal structure (PDB ID: 4TPO, shown in white). **b,** The indole ring is rotated by 180º, which positions the C4 carbon proximal to the peroxynitrite-N. The indole-NH forms a hydrogen-bond with the Tyr89-OH. F/G loop residues are in orange; the substrate and heme are in black; polar contacts are indicated by dashed lines. Both substrate conformations support an edge-to-face interaction with His176.

Supplementary Experimental Methods and Figures

General

All reagents and chemicals were purchased from Sigma-Aldrich and were used as received unless otherwise stated. The 4-nitro-DL-tryptophan standard was prepared as previously described.[12]

Cloning, expression, and purification

The plasmid encoding the *Streptomyces scabies* 87.22 TxtE gene was constructed as previously described.[13] The TxtE His176Tyr, TxtE His176Phe, and Virginiae TxtE Trp176His mutants were made using the QuikChange™ Site-Directed Mutagenesis kit according to manufacturer's directions (Agilent Technologies). The PCR mixture was treated with *Dpn*I according to manufacturer's directions (New England Biolabs), followed by transformation of XL10-Gold ultracompetent *Escherichia coli* cells (Strategene). After DNA sequencing (Laragen, Inc.), the correct construct was used to transform *E. cloni*® EXPRESS BL21(DE3) Competent Cells (Lucigen). Site-saturation mutagenesis (SSM) was carried out as previously described.[14] The genes encoding putative TxtE genes from *Streptomyces virginiae*, *Streptomyces sp. Mg1*, *Streptomyces lavendulae*, and *Saccharomonspora marina* XMU15 were codon optimized for *E. coli* K12 (high) using the Integrated DNA Technologies Codon Optimization tool and then obtained as gBlocks from Integrated DNA Technologies. The gBlocks were amplified using Phusion High Fidelity DNA polymerase (Thermo Scientific) according to manufacturer's instructions and Gibson cloned into pET28b+.[15] The His_6-tagged constructs were used to transform XL10-Gold ultracompetent *E. coli cells* (Stratagene). After DNA sequencing, the correct construct was used to transform *E. cloni*® EXPRESS BL21(DE3) Competent Cells.

For large-scale expression, Hyper Broth™ (700 mL, 30 μg/mL kanamycin sulfate, 2-L flask) supplemented with glucose nutrient mixture (35 mL) (Athena Enzyme Systems) and 5-

aminolevulinic acid hydrochloride (1 mM final concentration, Acros Organics) was inoculated with an overnight culture (14 mL, Terrific Broth (TB, Research Products International Corp.) with 30 μg/mL kanamycin sulfate, ~20 mL in a 50 mL flask). After 3.5 h of incubation at 37 °C and with shaking at 180 rpm to 250 rpm, the culture was placed in an ice water bath for 15 min. Then expression was induced with the addition of isopropyl β-D-thiogalactopyranoside (IPTG, 0.5 mM final concentration, Research Products International Corp.). Protein expression was carried out at 20 °C for all proteins except Lavendulae TxtE, which was carried out at room temperature. After 18–24 h, the cells were harvested by centrifugation at 4 °C, 4,000*g* for 10 min and the pellets were stored at -20 °C.

For 96-well plate expression, each well was filled with a solution (700 μL) of Hyper Broth™ (300 mL, 30 μg/mL kanamycin sulfate) supplemented with glucose nutrient mixture (10 mL) and 5-aminolevulinic acid hydrochloride (1.5 mM final concentration) and was inoculated with an overnight culture (50 μL, TB with 30 μg/mL kanamycin sulfate, 500 μL in a 96-well plate). After 3.5 h of incubation at 37 °C with shaking at 250 rpm, the culture was placed in an ice water bath for 30 min. Then expression was induced with the addition of IPTG (50 μL/well, 0.5 mM final concentration). Protein expression was carried out 20 °C with shaking at 250 rpm. After 18–24 h, the cells were harvested by centrifugation at 4 °C, 5,000*g* for 5 min and the pellets were stored at -20 °C.

For the expression of the variants with intermediate activity from the site-saturation mutagenesis library, Hyper Broth™ (90 mL or 300 mL, 30 μg/mL kanamycin sulfate, 250-mL or 1-L flasks) supplemented with glucose nutrient mixture (4.5 mL or 15 mL) and 5-aminolevulinic acid hydrochloride (1 mM final concentration) was inoculated with an overnight culture (1.9 mL or 6 mL, 30 μg/mL kanamycin sulfate, 5–6 mL in 10 mL culture tubes). After 3.5–4 h of

incubation at 37 °C and with shaking at 170 rpm to 225 rpm, the culture was placed in an ice water bath for 15 min. Then expression was induced with the addition of IPTG (0.5 mM final concentration.). Protein expression was carried out at 20 °C with shaking 150 rpm to 225 rpm. After 20–24 hours, the cells were harvested by centrifugation at 4 °C at 5,000*g* for 10 min and the pellets were stored at -20 °C.

During protein purification, all samples were kept at 4 °C. The cell pellet was resuspended in Ni-NTA buffer A (25 mM Tris-HCl, 100 mM NaCl, 30 mM imidazole, pH 7, 4 mL/g of cell wet weight) and lysed by sonication (Sonicator 3000, Misonix, Inc.). The lysate was clarified by centrifugation at 20,817*g* for 30 min, and the supernatant was loaded on a pre-equilibrated Ni-NTA column (GE Healthcare) using an AKTA purifier FPLC system (GE Healthcare). After washing with two column volumes (cv) of Ni-NTA buffer A, the protein was eluted with a linear gradient (12-cv gradient) of buffer B (25 mM Tris-HCl, 100 mM NaCl, 300 mM imidazole, pH 7). The fractions that showed an absorbance at 280 nm and 417 nm were combined and buffer exchanged into 20 mM Tris-HCl pH 7.5 on a 30-kDA MWCO centrifugal filter device (EMD Millipore). The protein was aliquoted into PCR tubes, flash frozen over powdered dry ice, and stored at -80 °C until further use.

For the purification of the variants with intermediate activity from the site-saturation library, a portion of the culture (~ 20–25 mL) was resuspended in Ni-NTA buffer A (700 µL, 25 mM Tris-HCl, 100 mM NaCl, 10 mM imidazole, pH 7) supplemented with lysozyme (2 mg/mL) and DNAseI (0.02 mg/mL). Lysis proceeded at 37 °C for 1 h followed by centrifugation at 12,000g for 15–30 min. The Ni-NTA spin column (Qiagen) was equilibrated with Buffer A (600 µL, 890g, 2 min). Then the supernatant was loaded onto the column (600 µL) and centrifuged (270g, 5 min) followed by washing with buffer B (600 µL x 2, 25 mM Tris-HCl, 100 mM NaCl,

20 mM imidazole, pH 7) and centrifugation between each wash step (890g, 2 min). The protein was eluted with elution buffer (300 μL x 2–3, 25 mM Tris-HCl, 100 mM NaCl, 300 mM imidazole, pH 7) followed by centrifugation (890g, 2 min). The eluted fractions were concentrated and stored as described above.

Protein concentration determination

The concentrations of purified holoenzymes were determined by quantifying the free hemin present using the pyridine/hemochrome assay.[16] The sample was prepared by vortexing a solution of protein in Tris-HCl (10 μL, 20 mM, pH 7.5) with potassium phosphate buffer (140 μL, 100 mM, pH 8.0) and sodium hydroxide/pyridine (750 μL of 1.75 mL of 1 M NaOH combined with 0.75 mL pyridine) followed by the addition of sodium dithionite (100 μL, 100 mM in water). This solution was promptly vortexed to ensure proper mixing, and the UV-visible absorption spectrum was acquired. A blank reference sample was prepared without protein. The previously reported molar extinction coefficient ε_{418} = 191.5 $mM^{-1}cm^{-1}$ was used.[16] For crystallography, total protein concentration was determined with the Quick Start™ Bradford Protein Assay (Biorad).

Chromatography

Low-resolution mass spectral analyses were carried out using an Agilent 1290 UHPLC system with a 6140 quadrupole detector and an Eclipse Plus 2.1 x 50 mm C18 column (1.8 μm particle size) with a water/0.1% acetic acid and acetonitrile mobile phase. A diode array detector signal at 360 nm or 400 nm with a reference signal at 500 nm was collected for the detection of nitrated products, and a signal at 280 nm with a reference signal at 500 nm was collected for the detection of tryptophan. Analytical high-performance liquid chromatography (HPLC) was carried out using an Agilent 1200 series and a Kromasil 4.6 mm x 50 mm C18 column (5 micron particle size) with a water/0.1% trifluoroacetic acid and acetonitrile mobile phase. A multi-

wavelength detector (MWD) signal at 330 nm with a reference signal at 550 nm was collected for the detection of 5-nitro-DL-tryptophan, at 400 nm with a reference signal at 550 nm was collected for the detection of 4-nitro-DL-tryptophan, and at 230 nm with a reference signal at 360 nm was collected for the detection of DL-tryptophan and the internal standard.

Preparation of standard curves

A solution of the *p*-toluenesulfonamide internal standard was prepared in DMSO at 1 M and then diluted to 4 mM in 0.15 M HCl for the standard curves. A stock solution of the 5-nitro-DL-tryptophan standard (Santa Cruz Biotechnology) was prepared at 50 mM in 1 M HCl and serially diluted from 500 μM to 0 μM in a solution of 10 M NaOH, 1 M Tris pH 8, and 25 mM Tris pH 8 (1:4:195 μL). Internal standard (25 μL) was added to each concentration (200 μL) of the standard curve and each solution was analyzed by HPLC in quadruplicate. The standard curve was prepared by plotting the concentration of the 5-nitro-DL-tryptophan standard versus the MWD signal of the 5-nitro-DL-tryptophan standard at 330 nm divided by the area of the internal standard at 230 nm. A stock solution of the 4-nitro-DL-tryptophan standard was prepared at 37 mM in 1 M HCl and serially diluted from 370 μM to 9.25 μM in a solution of 10 M NaOH, 1 M Tris pH 8, and 25 mM Tris pH 8 (1:4:195). The previously reported molar extinction coefficient ε_{400} = 3000 $M^{-1}cm^{-1}$ was used to determine the concentration of the stock solution.[17] Internal standard (25 μL) was added to each concentration (200 μL) of the standard curve and each solution was analyzed by HPLC in quadruplicate. Overlap from the 4-nitro-DL-tryptophan standard at 230 nm was corrected for by generating a standard curve of the 4-nitro-DL-tryptophan without the internal standard where the area at 400 nm was plotted versus the area at 230 nm. This standard curve was then used to subtract any overlap from the internal standard. The corrected standard curve was prepared by plotting the concentration of the 4-nitro-DL-tryptophan

standard versus the multi-wavelength detector (MWD) signal of the 4-nitro-DL-tryptophan standard at 400 nm divided by the area of the internal standard at 230 nm.

Purified enzyme and control reactions

For hemin control reactions, a stock solution of hemin (10 mg/mL in 0.1 M NaOH, Frontier Scientific, Inc.) was diluted and aliquoted into a 96-well plate (150 μL, 2.5 μM or 13.3 μM in 25 mM Tris pH 8) or stock solutions of hemin (10 mg/mL in 0.1 M NaOH) and BSA (250 mM in 25 mM Tris pH 8) were diluted and aliquoted into a 96-well plate (150 μL, 2.5 μM or 13.3 μM in 25 mM Tris pH 8). A solution (50 μL) containing the following was added to each well: L-tryptophan (5 μL, 20 mM in 25 mM Tris pH 8, 500 μM final concentration), sodium dithionite (1.79 μL, 100 mM in water, 890 μM final concentration), and DEANO (1 μL, 100 mM in 10 mM sodium hydroxide) in 42 μL of 25 mM Tris pH 8. The nitration reaction with 5-methyl-DL-tryptophan (2 μL, 100 mM in 250 mM NaOH, 500 μM final concentration of the L-enantiomer) was carried out as described in the *Methods* section. The nitration reactions with alpha-methyl-L-tryptophan and indole-3-acetamide were carried out as previously described.[13]

Reaction screening in 96-well plate format

The 96-well plate was thawed for 20 min at room temperature. The cell pellets were resuspended in lysis buffer (150 μL, 2 mg/mL lysozyme, 0.02 mg/mL DNase, 2 mM $MgCl_2$ in 25 mM Tris pH 8) by vortexing followed by incubation at 37 °C for 1 h. The lysate was clarified by centrifugation at 5,000 g for 30 min and 100 μL was transferred to a 96-well plate using an automated liquid handler (Hamilton). A solution (50 μL) containing the following was added to each well: L-tryptophan (6.25 μL, 50 mM in 25 mM Tris pH 8, 2.5 mM final concentration), spinach ferredoxin $NADP^+$ reductase (1.25 μL, 17 U/mL in 1 M Tris pH 8, 0.17 U/mL final concentration, spinach ferredoxin (0.625 μL, 2 mg/mL in Tris pH 8, 0.01 mg/mL final concentration), NADPH (1.25 μL, 89 mM in 25 mM Tris pH 8, 890 μM final concentration), and

DEANO (1.25 μL, 100 mM in 10 mM sodium hydroxide) in 15 μL of 25 mM Tris pH 8. The plate was covered, and wrapped in foil and the reactions proceeded with shaking. After 2 h, a solution of DEANO (0.625 μL) and NADPH (1.25 μL) in 8.125 μL of 25 mM Tris pH 8 was added to each well, and the reactions proceeded overnight with shaking. Each well from the 96-well plate was applied to an AcroPrep™ Advance 96-well 3K filter plate (Pall Corp.) and centrifuged at 2,500g for 30 min at room temperature. The filtrate was transferred to a 96-well assay plate and analyzed by HPLC. Variants with moderate to robust activity were further sequenced, stored as glycerol stocks, and analyzed for regioselectivity.

Crystallization and data collection

High-throughput screening of co-crystallization conditions for TxtE His176Phe and TxtE His176Tyr with L-tryptophan was conducted at the Beckman Molecular Observatory at the California Institute of Technology, and the optimal conditions were further refined using focused screens. Crystals for diffraction were grown in a sitting-drop format in 24-well plate (Hampton Research) by mixing 1 μL of 14 mg/mL TxtE His176Phe with 11.2 mM L-tryptophan or 12 mg/mL TxtE His176Tyr with 11.2 mM L-tryptophan in 20 mM Tris buffer (pH 7.5) with 1 μL of the well solution containing 200 mM $MgCl_2$, 100 mM Bis-Tris pH 5.5, 25% (w/v) PEG 3350. Moderately sized red-brown orthorhombic crystals appeared within a two weeks. All crystals were cryo-protected in mother liquor containing 25% (v/v) glycerol prior to flash-freezing in liquid nitrogen. Diffraction data were collected using a Dectris Pilatus 6M detector on beamline 12-2 at the Stanford Synchrotron Radiation Laboratory at 100 K. Diffraction datasets were indexed and integrated with XDS[18] and scaled using SCALA.[19]

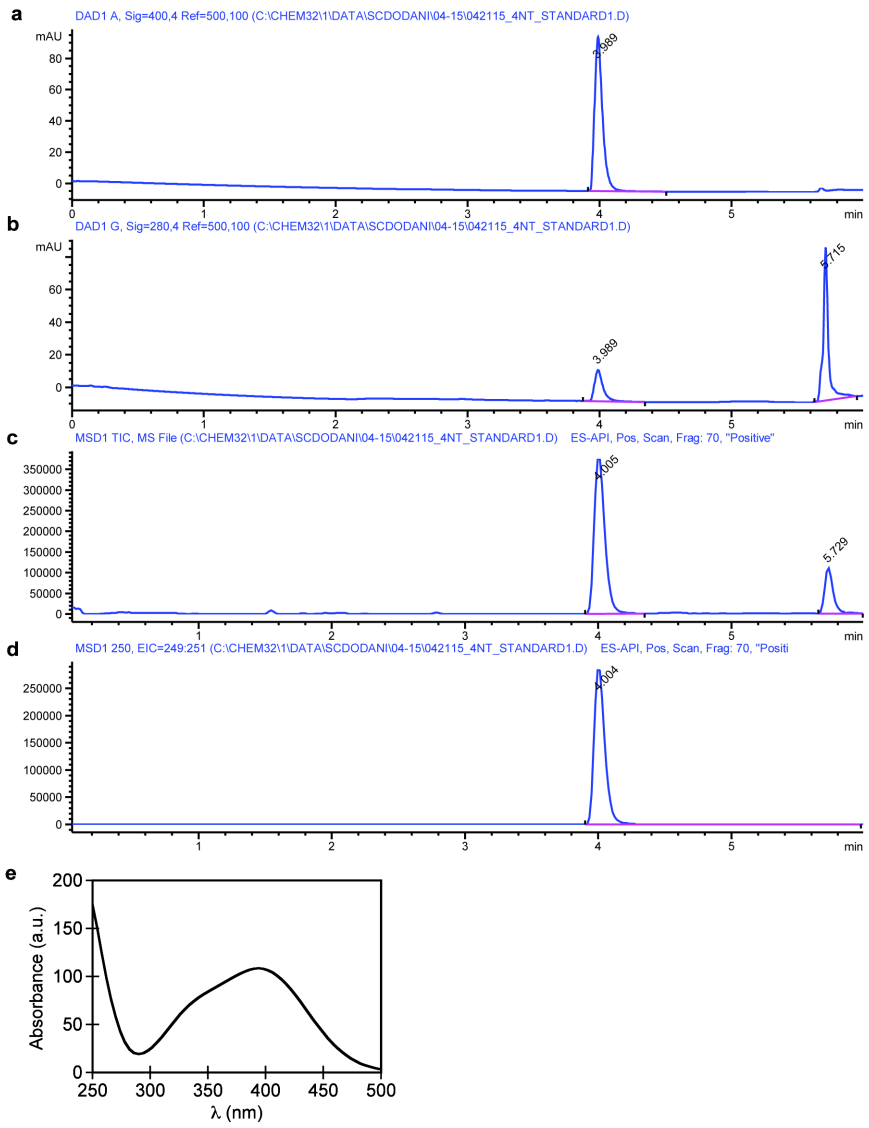
Structure solution

The substrate free TxtE structure (PDB ID: 4TPN) was used as a starting model to solve the 1.45 Å TxtE His176Phe and 1.51 Å TxtE His176Tyr L-tryptophan-bound data sets. After

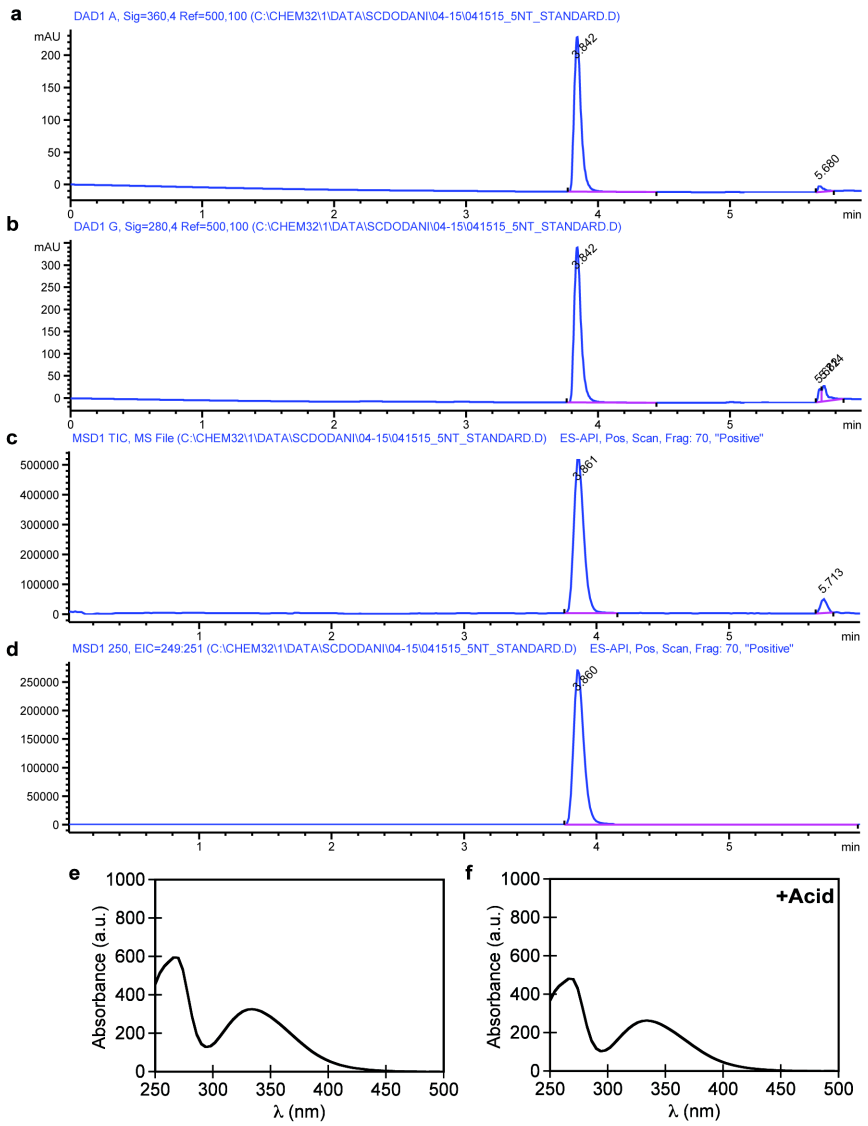
molecular replacement using Phaser,[20] refinement was performed by alternating cycles of automated refinement with Refmac5[21] and manual refinement in Coot.[22] To minimize phase bias, the same R-Free set was used for all structures solved. Anisotropic B-factors were calculated for both high-resolution structures. The structures were deposited to the RCSB Protein DataBank with accession codes 5D3U and 5D40 (Supplementary Table S2).

Sequence alignment, percent identity matrix, and phylogenetic tree

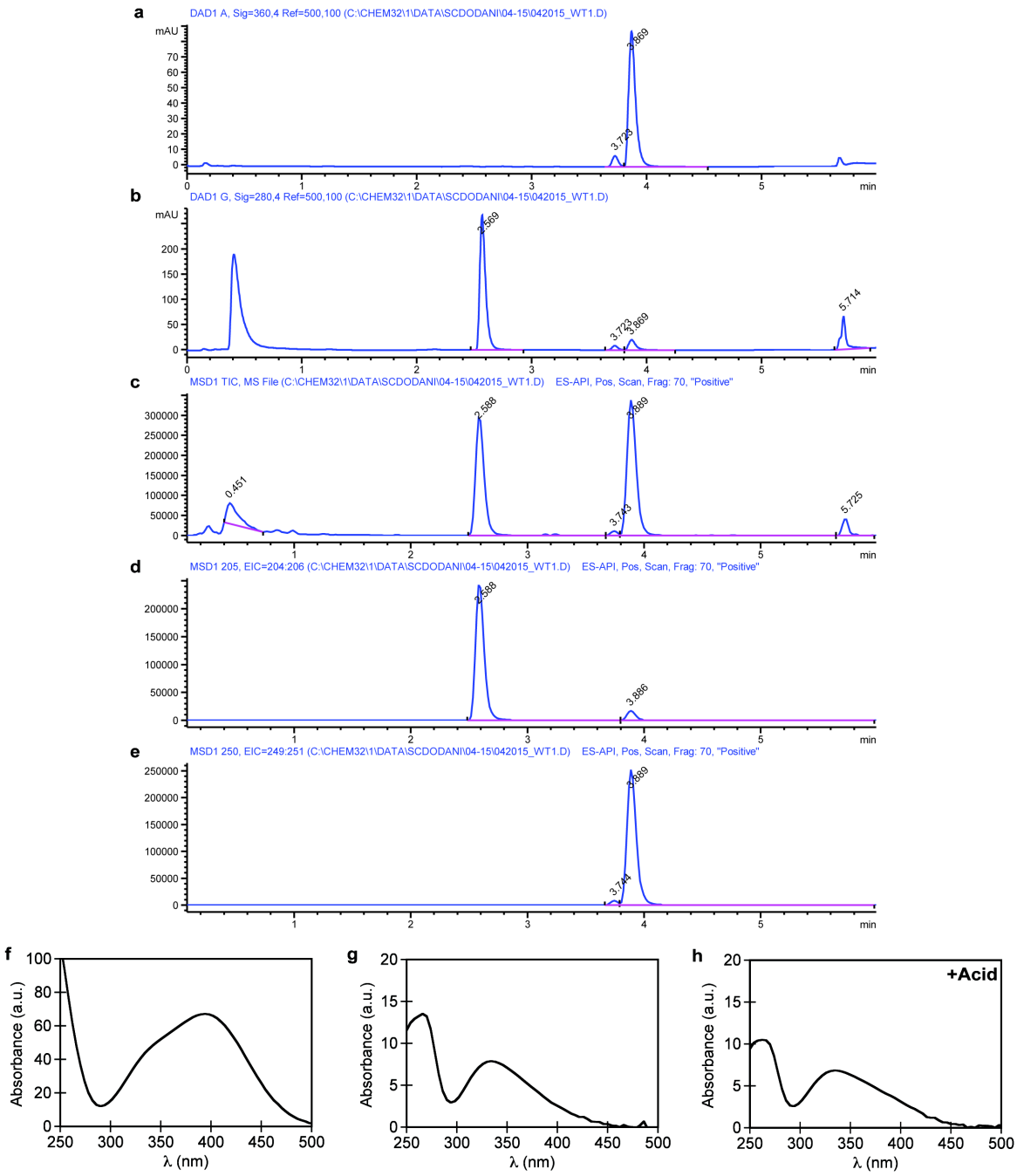
Uniprot was used to run a BLAST search and unique entries with a percent identity greater than 65% were aligned using ClustalW2.[23,24] The sequence logo plot was generated using WebLogo version 2.8.2.[25,26]



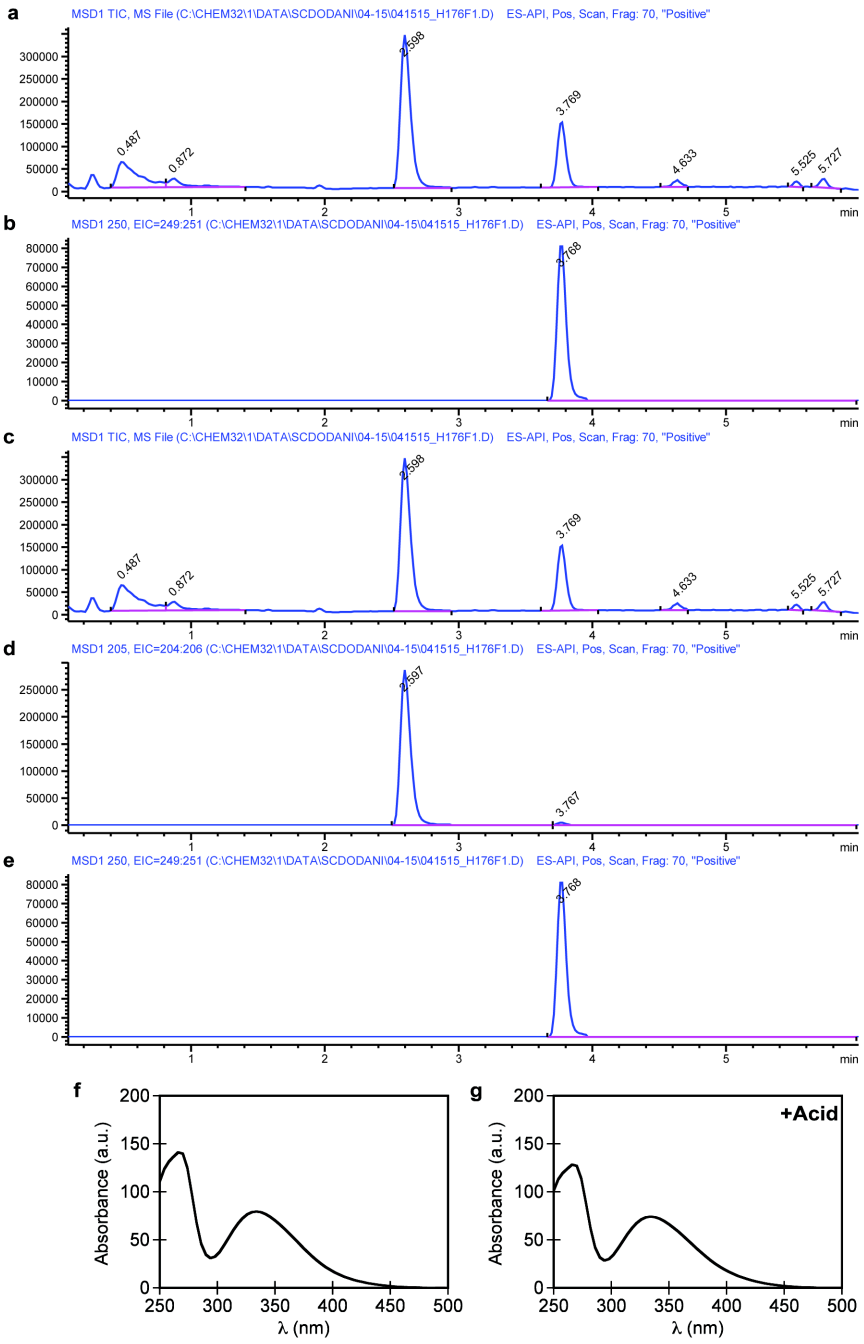
Supplementary Figure S6| LC-MS analysis of the 4-nitrotryptophan standard. Diode array detector (DAD) signal at **a,** 400 nm and **b,** 280 nm to detect the nitrated standard. **c,** LC-MS trace of the standard. Extracted ion chromatogram at **d,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated standard **f,** at *t* = 3.989 min corresponding to 4-nitrotryptophan.



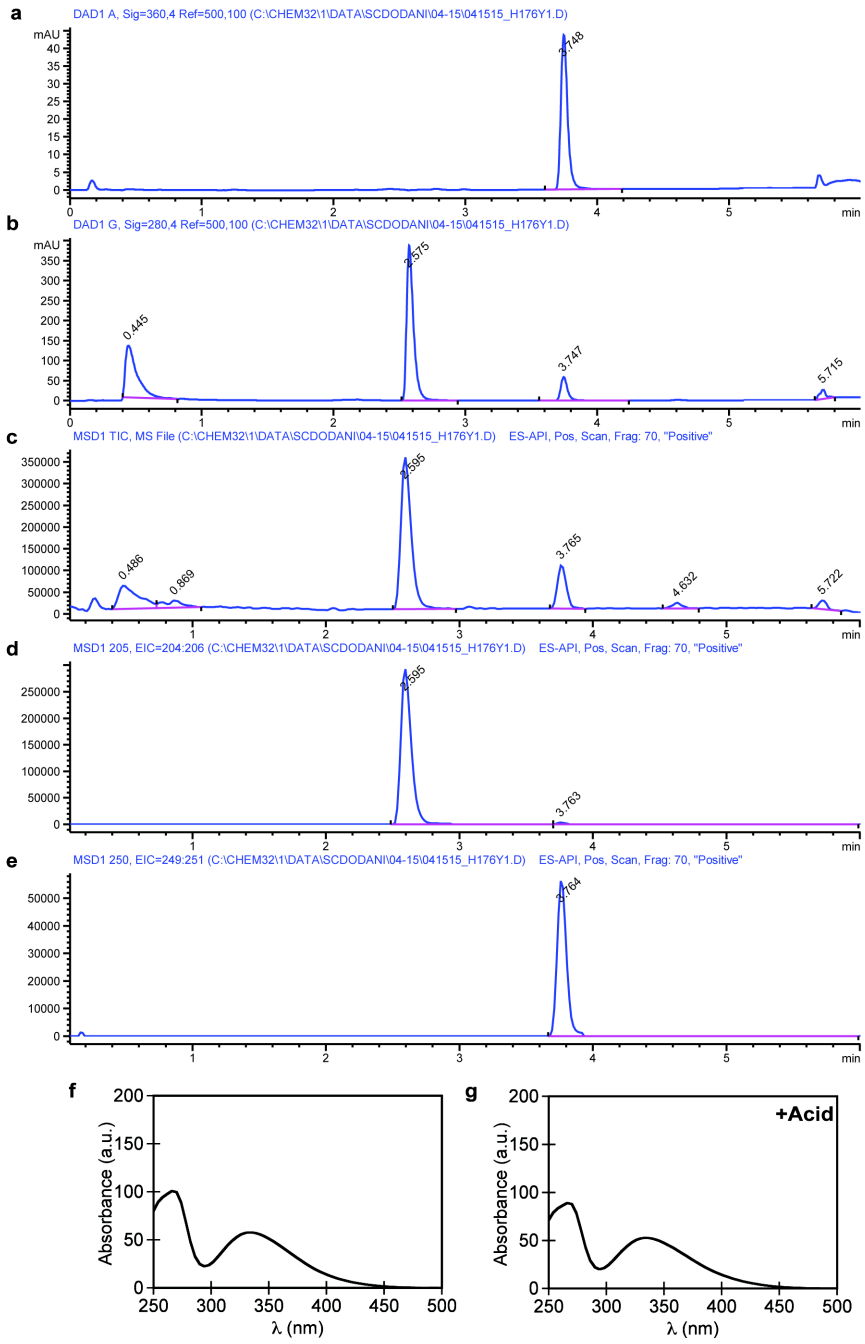
Supplementary Figure S7| LC-MS analysis of the 5-nitrotryptophan standard. DAD signal at **a,** 360 nm and **b,** 280 nm to detect the nitrated standard. **c,** LC-MS trace of the standard. Extracted ion chromatogram at **d,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated standard **e,** at *t* = 3.842 min corresponding to 5-nitrotryptophan and **f,** following treatment with acid, indicating no decomposition.



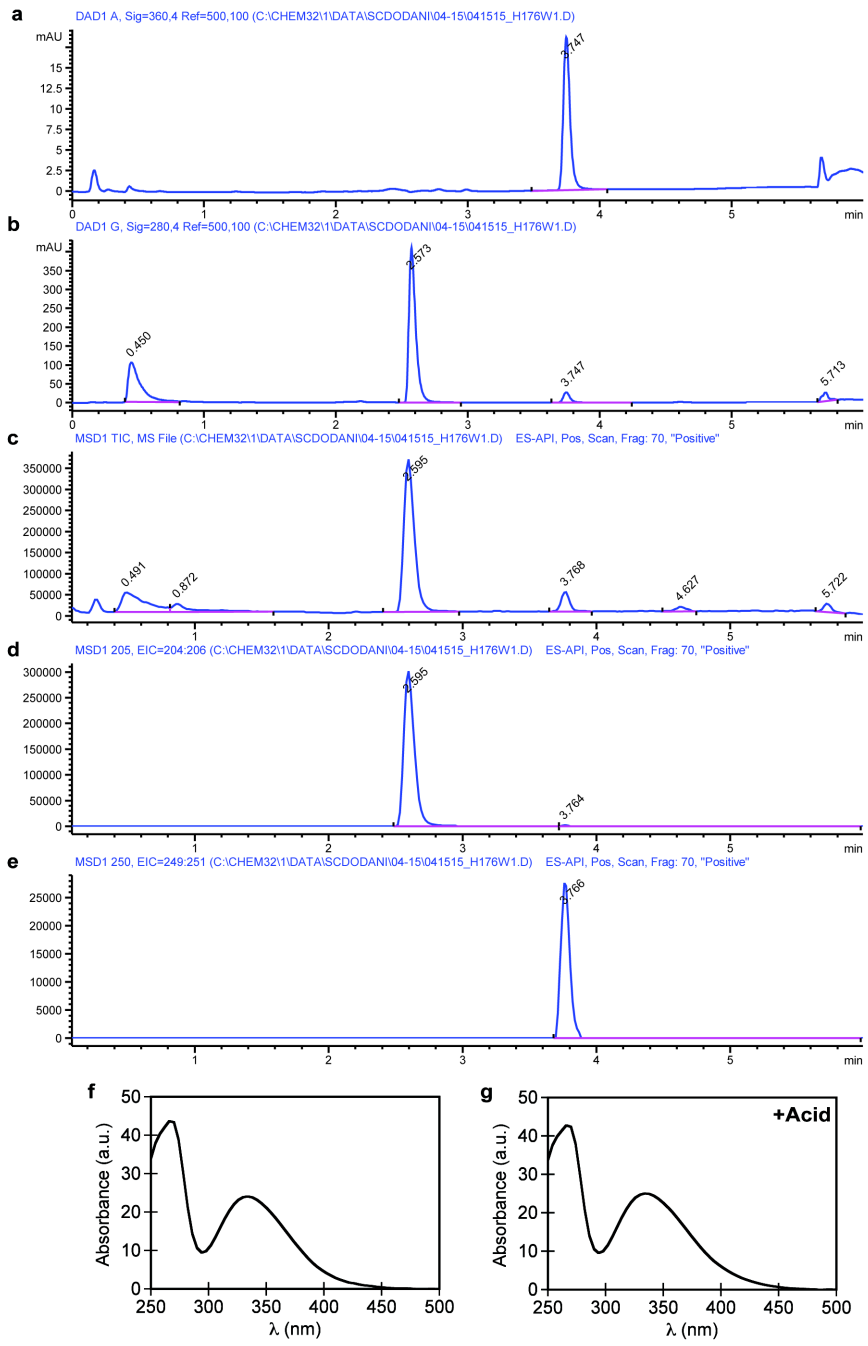
Supplementary Figure S8| LC-MS analysis of the nitration reaction with wild-type TxtE. DAD signal at **a,** 360 nm to detect all nitrated products and **b**, 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.869 min corresponding to 4-nitrotryptophan and **g,** at *t* = 3.723 min corresponding to 5-nitrotryptophan and **h,** following treatment with acid, indicating no decomposition.



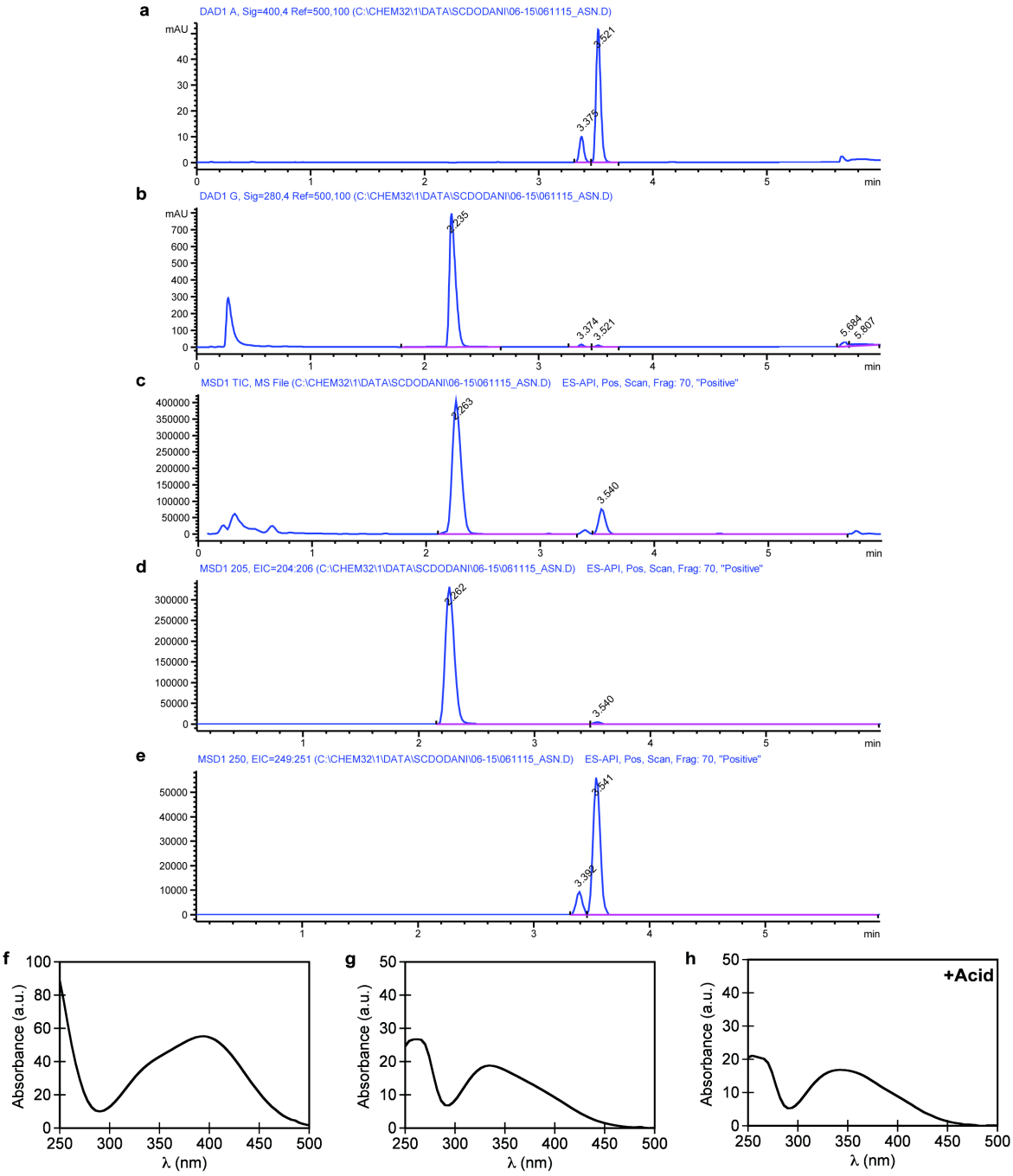
Supplementary Figure S9| LC-MS analysis of the nitration reaction with TxtE His176Phe. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.769 min corresponding to 5-nitrotryptophan and **g,** following treatment with acid, indicating no decomposition.



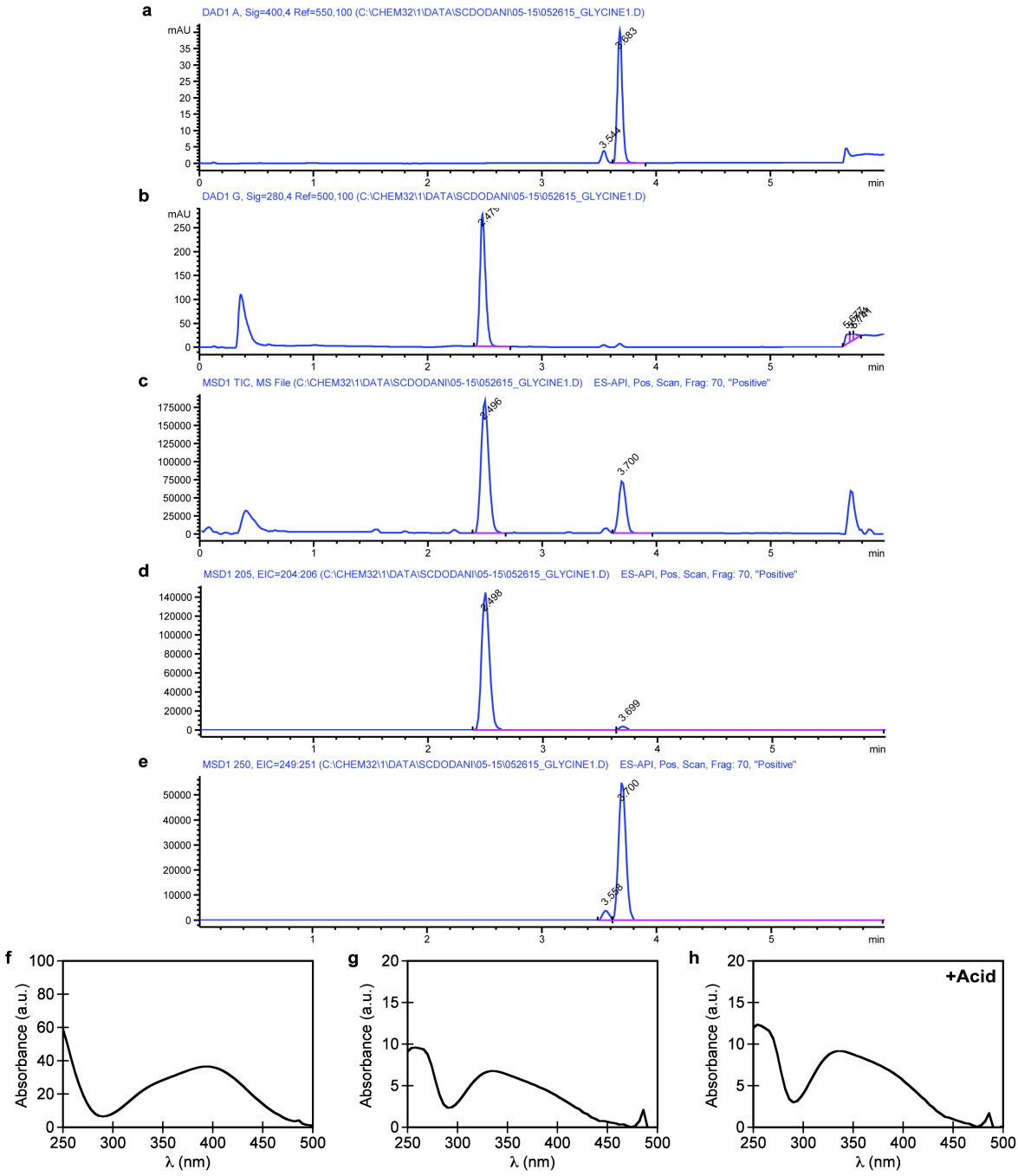
Supplementary Figure S10| LC-MS analysis of the nitration reaction with TxtE His176Tyr. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.748 min corresponding to 5-nitrotryptophan and **g,** following treatment with acid, indicating no decomposition.



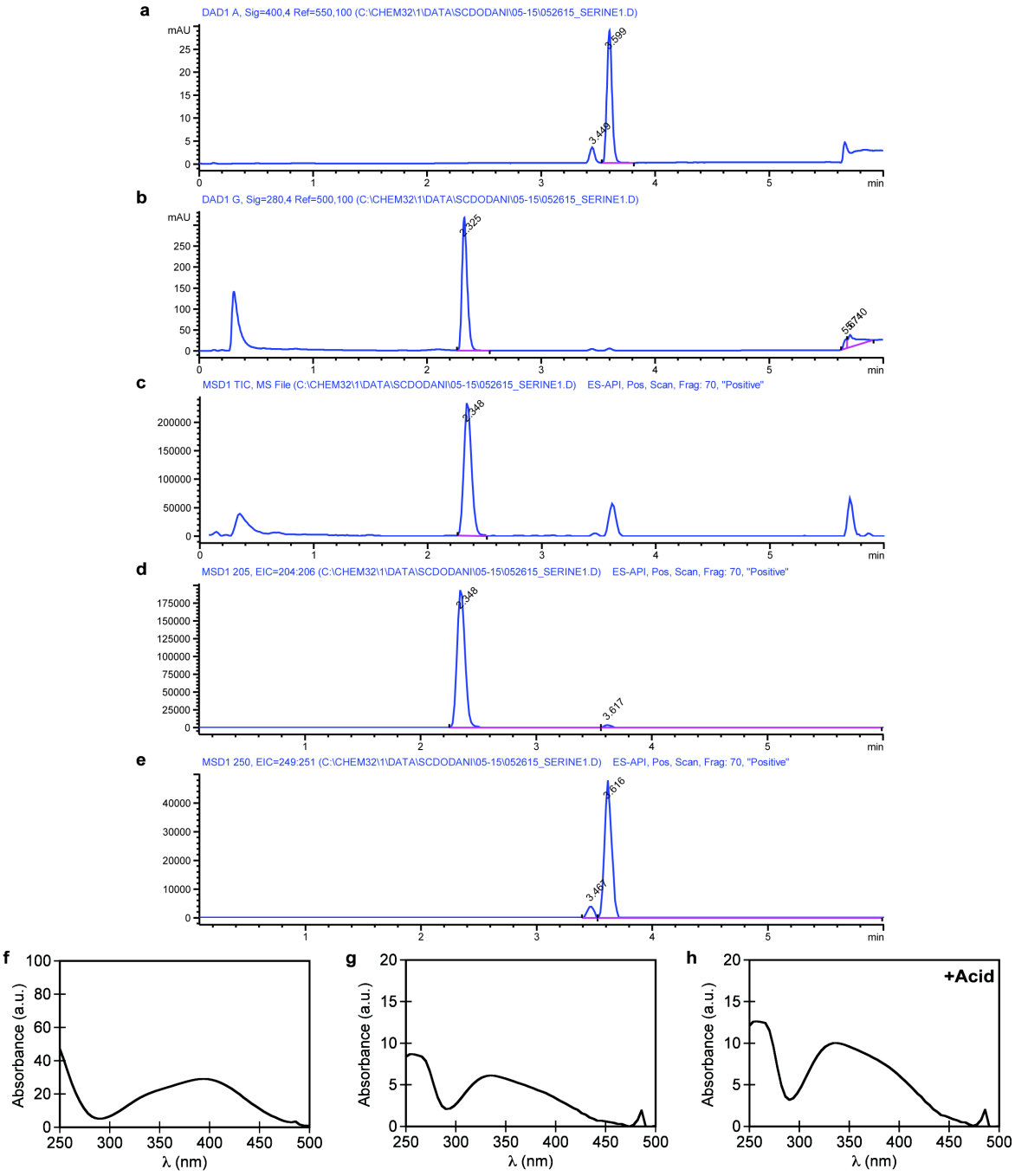
Supplementary Figure S11| LC-MS analysis of the nitration reaction with TxtE His176Trp. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.747 min corresponding to 5-nitrotryptophan and **g,** following treatment with acid, indicating no decomposition.



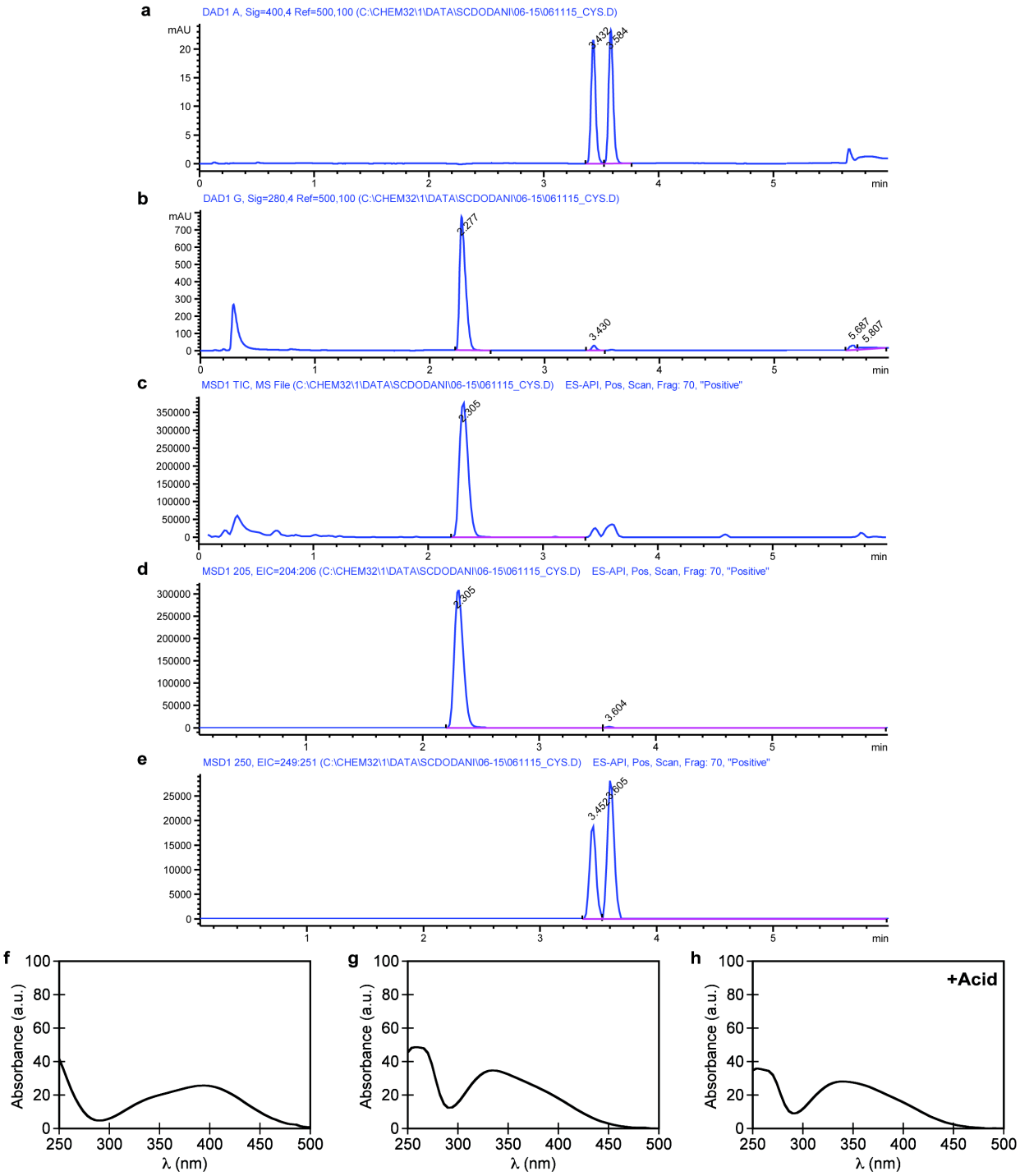
Supplementary Figure S12| LC-MS analysis of the nitration reaction with TxtE His176Asn. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated products **f,** at *t* = 3.521 min corresponding to 4-nitrotryptophan and **g,** at *t* = 3.375 min corresponding to 5-nitrotryptophan and **h**, following treatment with acid, indicating no decomposition.



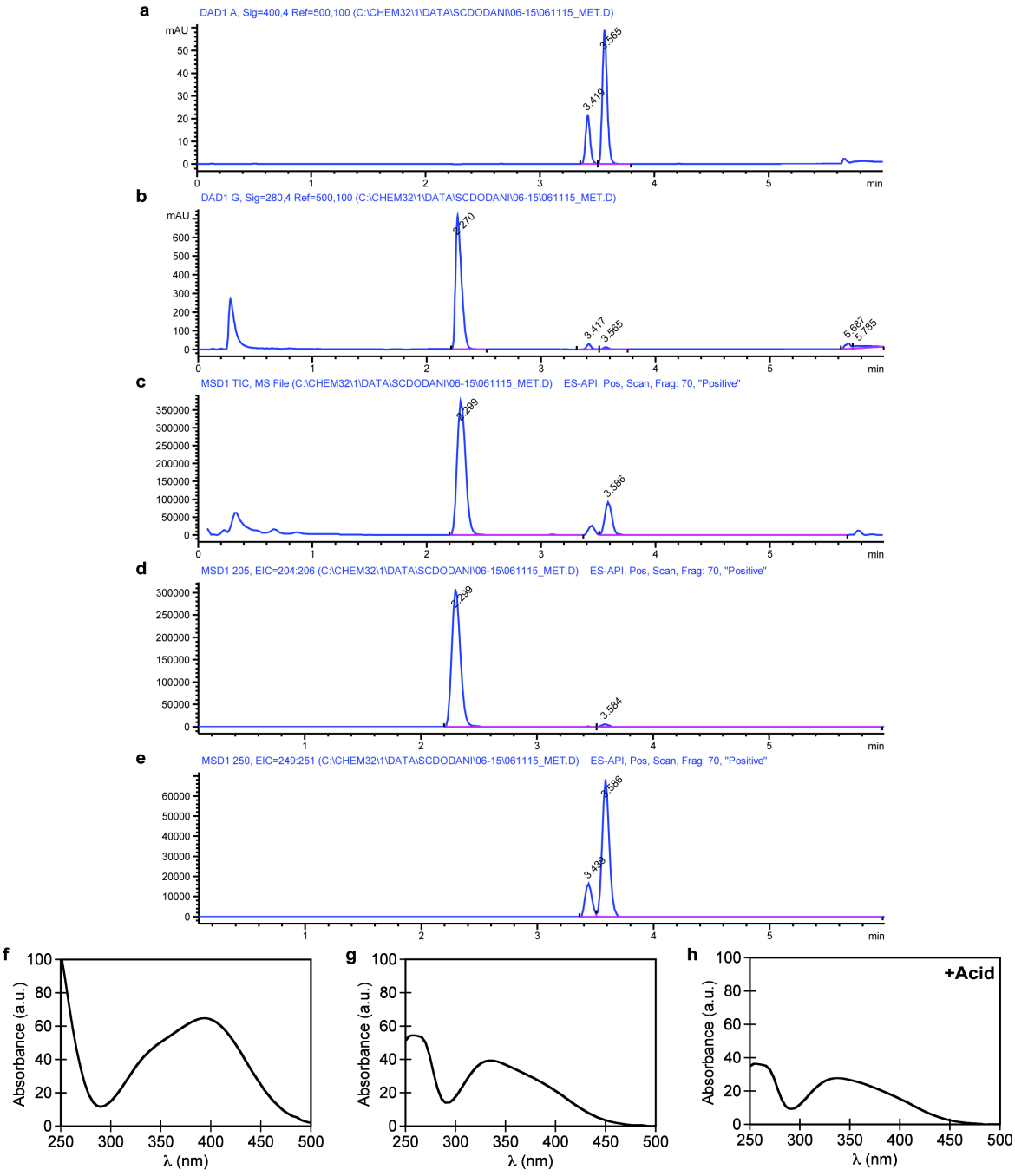
Supplementary Figure S13| LC-MS analysis of the nitration reaction with TxtE His176Gly. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated products **f,** at *t* = 3.683 min corresponding to 4-nitrotryptophan and **g,** at *t* = 3.544 min corresponding to 5-nitrotryptophan and **h**, following treatment with acid, indicating no decomposition.



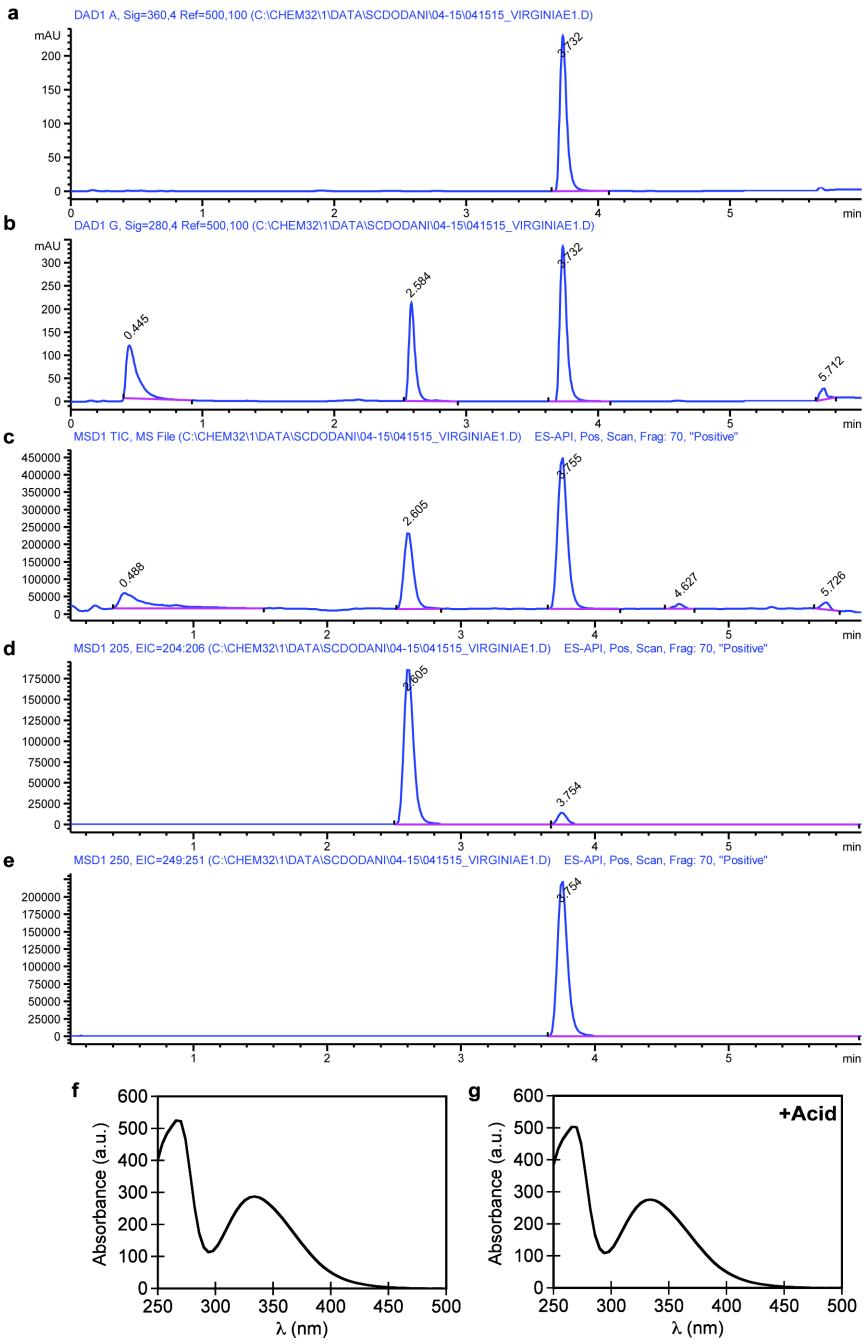
Supplementary Figure S14| LC-MS analysis of the nitration reaction with TxtE His176Ser. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated products **f,** at *t* = 3.599 min corresponding to 4-nitrotryptophan and **g,** at *t* = 3.449 min corresponding to 5-nitrotryptophan and **h**, following treatment with acid, indicating no decomposition.



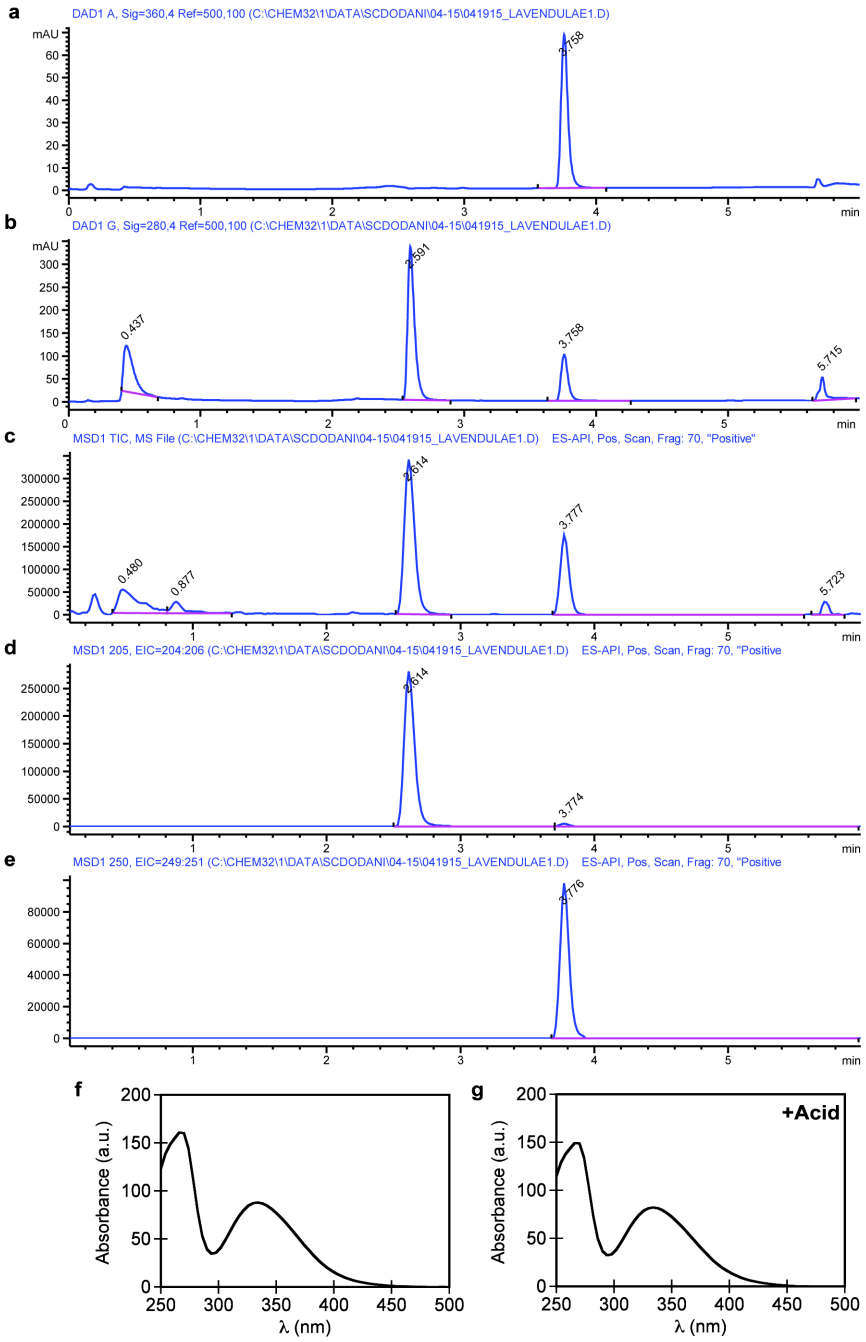
Supplementary Figure S15| LC-MS analysis of the nitration reaction with TxtE His176Cys. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated products **f,** at *t* = 3.584 min corresponding to 4-nitrotryptophan and **g,** at *t* = 3.432 min corresponding to 5-nitrotryptophan and **h**, following treatment with acid, indicating no decomposition.



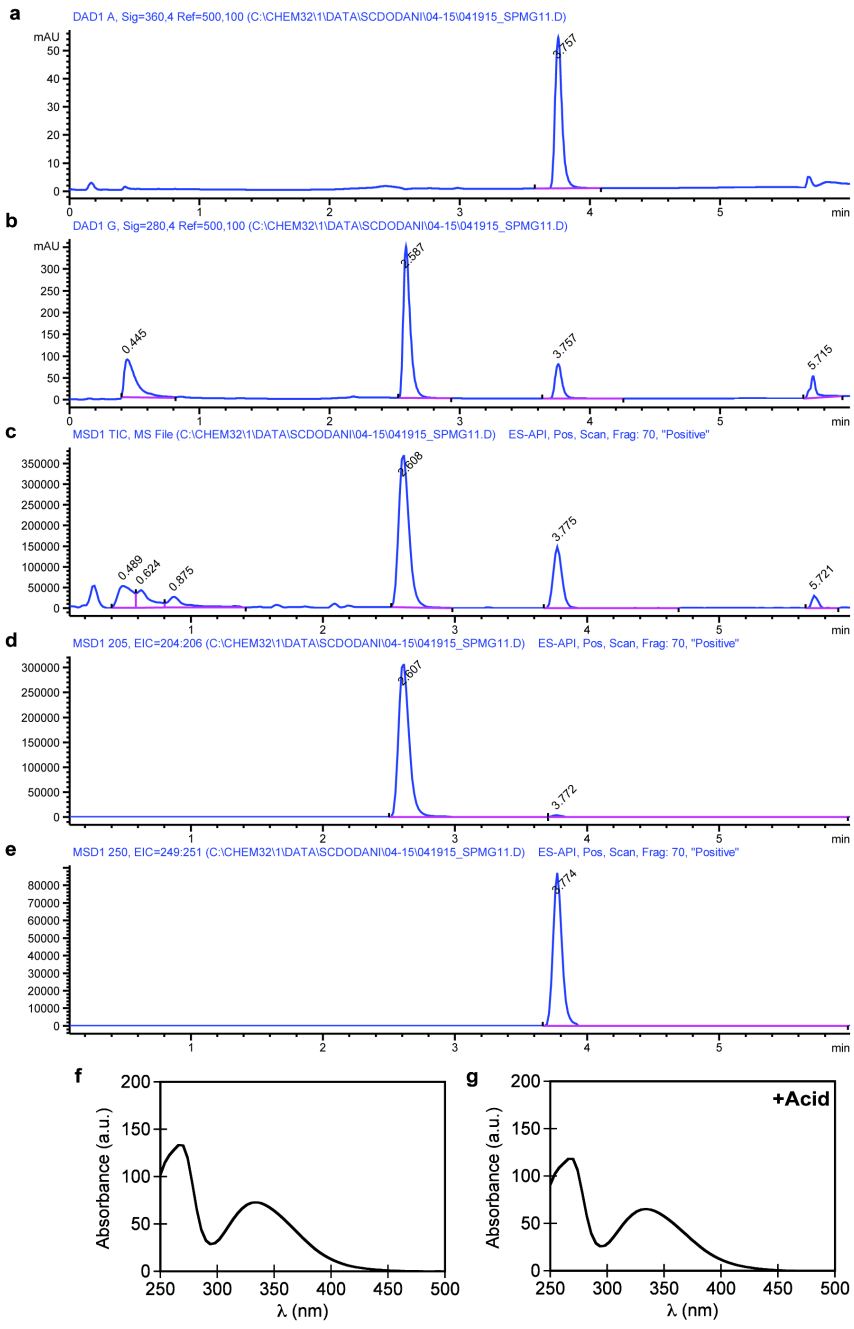
Supplementary Figure S16| LC-MS analysis of the nitration reaction with TxtE His176Met. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated products **f,** at *t* = 3.565 min corresponding to 4-nitrotryptophan and **g,** at *t* = 3.419 min corresponding to 5-nitrotryptophan and **h**, following treatment with acid, indicating no decomposition.



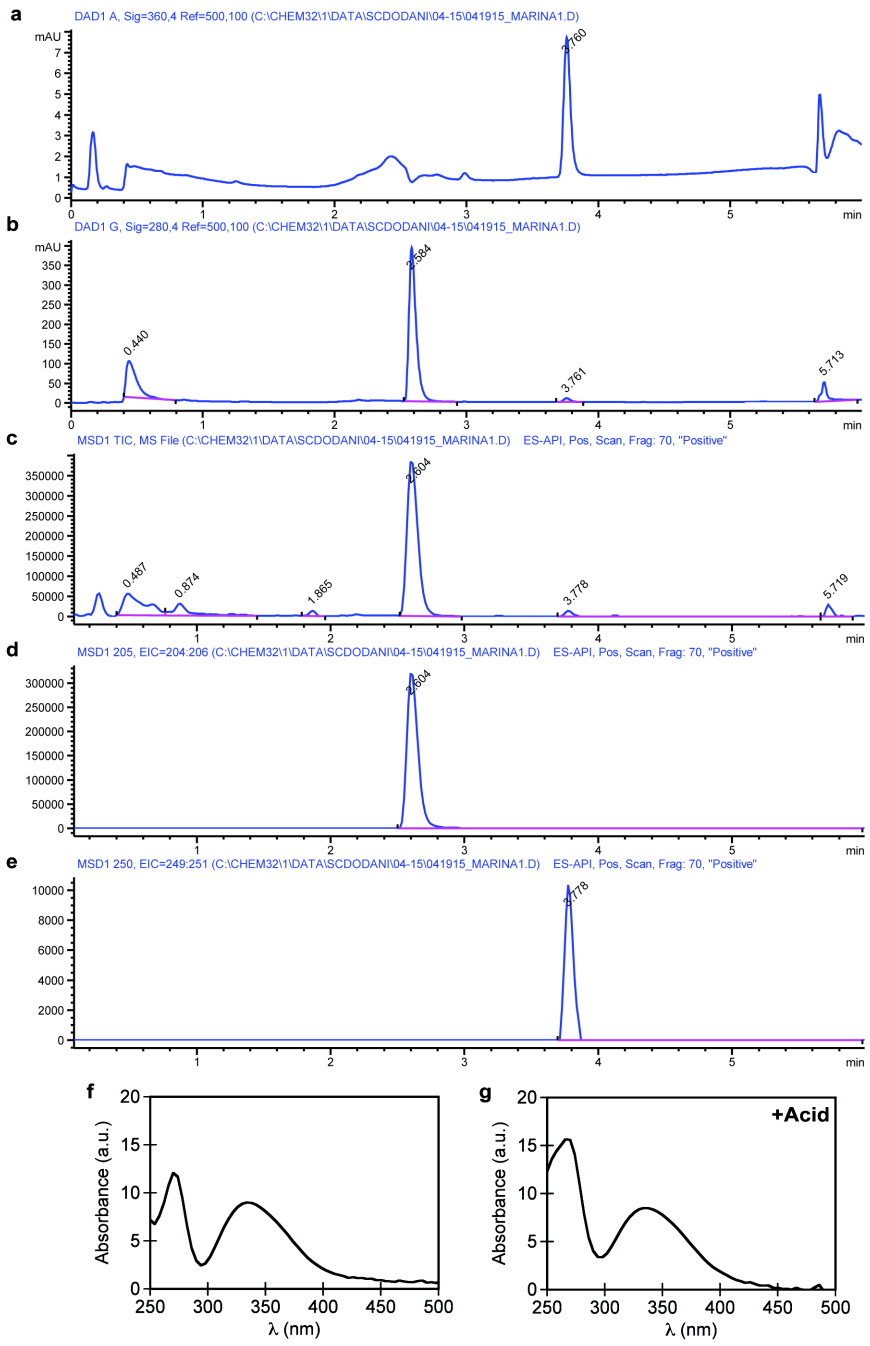
Supplementary Figure S17| LC-MS analysis of the nitration reaction with Virginiae TxtE. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.732 min corresponding to 5-nitrotryptophan and **g,** following treatment with acid, indicating no decomposition.



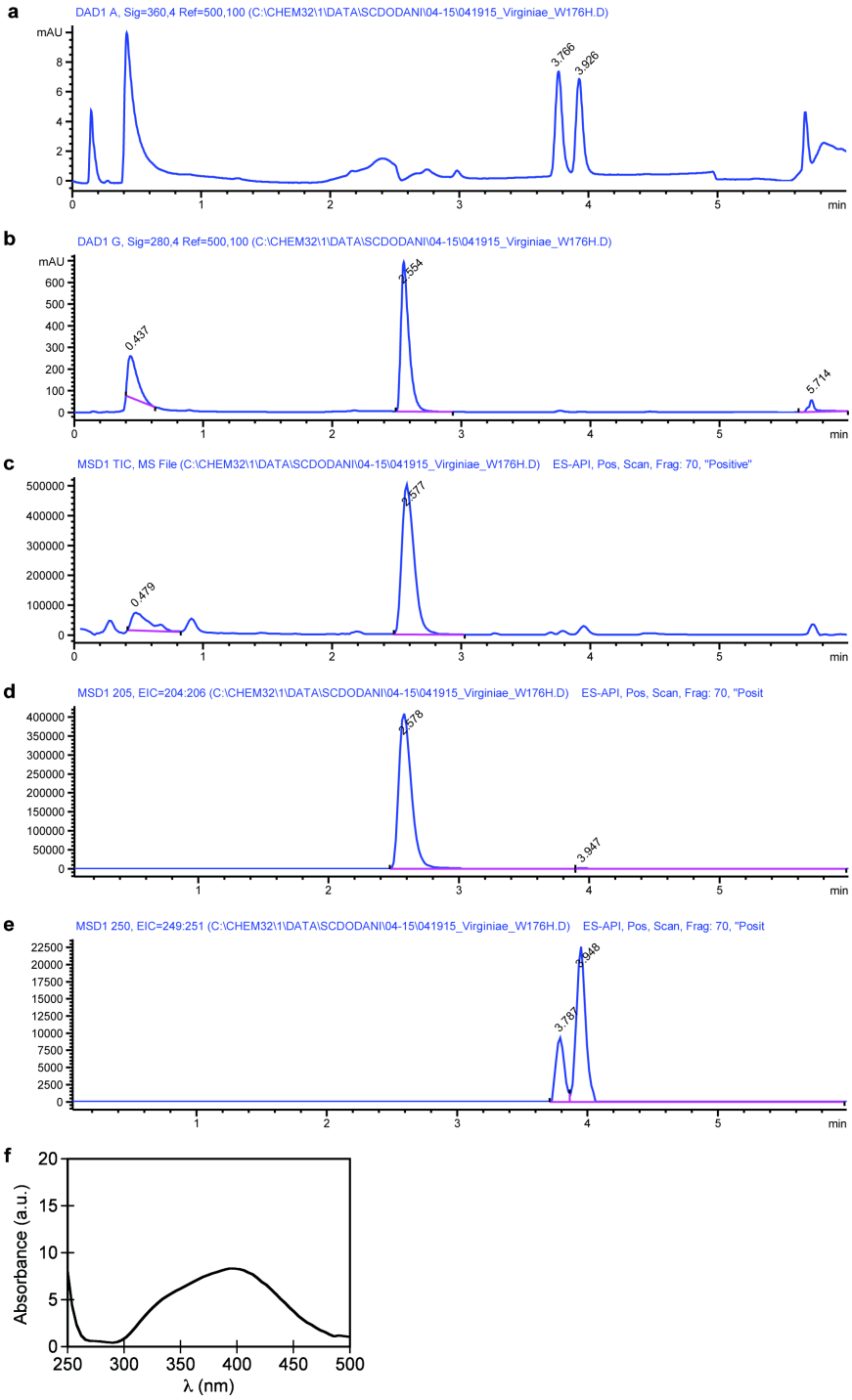
Supplementary Figure S18| LC-MS analysis of the nitration reaction with Lavendulae TxtE. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.758 min corresponding to 5-nitrotryptophan and **g,** following treatment with acid, indicating no decomposition.



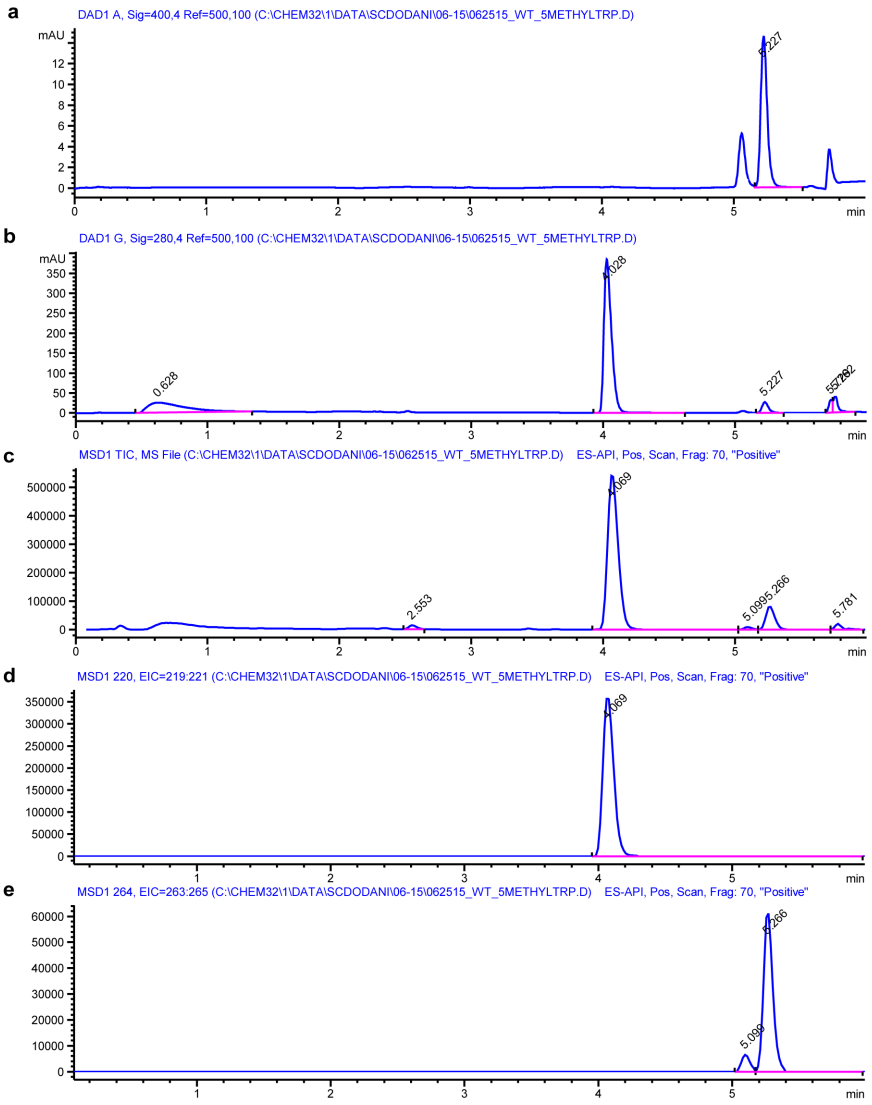
Supplementary Figure S19| LC-MS analysis of the nitration reaction with sp. Mg1 TxtE. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.757 min corresponding to 5-nitrotryptophan and **g,** following treatment with acid, indicating no decomposition.



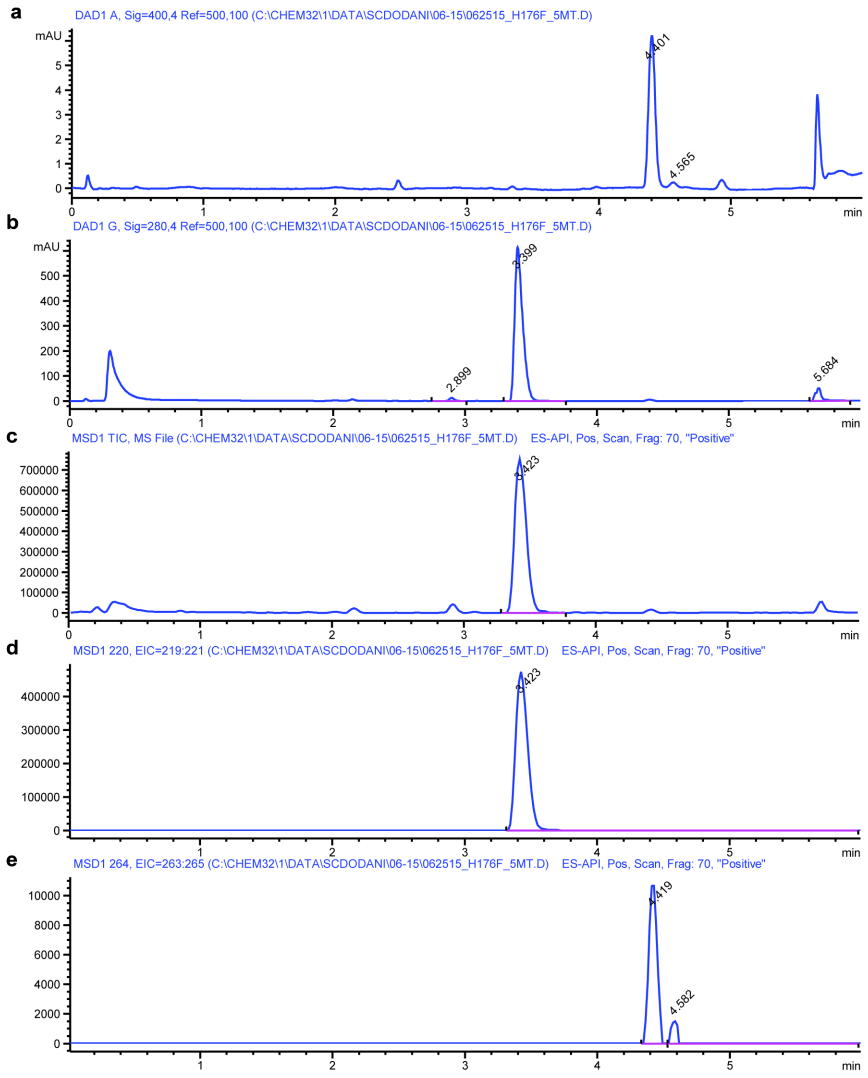
Supplementary Figure S20| LC-MS analysis of the nitration reaction with Marina TxtE. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.760 min corresponding to 5-nitrotryptophan and **g,** following treatment with acid, indicating no decomposition.



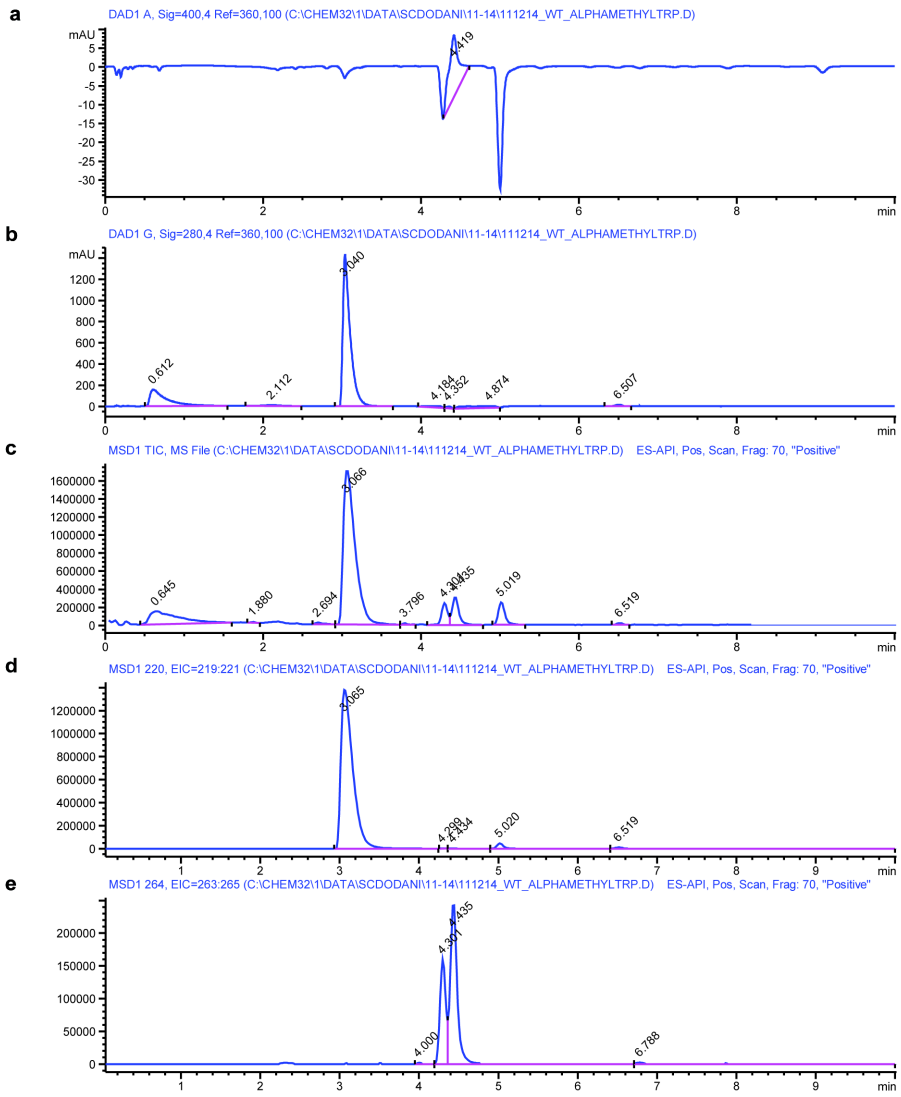
Supplementary Figure S21| LC-MS analysis of the nitration reaction with Virginiae TxtE Trp176His. DAD signal at **a,** 360 nm to detect all nitrated products and **b,** 280 nm to detect tryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 205 for tryptophan and **e,** *m/z* 250 for nitrotryptophan. DAD UV-visible absorption spectrum of the nitrated product **f,** at *t* = 3.926 min corresponding to 4-nitrotryptophan.



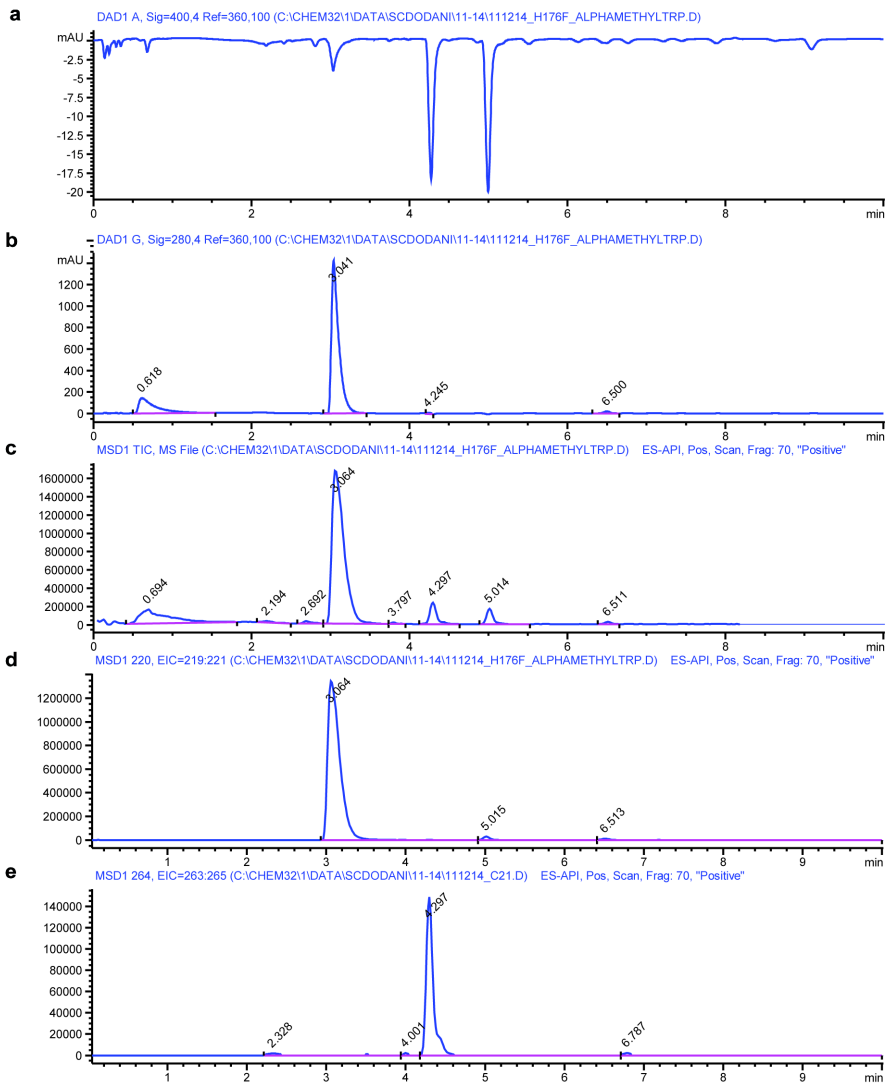
Supplementary Figure S22| LC-MS analysis of the nitration reaction with TxtE and 5-methyltryptophan. DAD signal at **a,** 400 nm to detect all nitrated products and **b,** 280 nm to detect 5-methyltryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 220 for 5-methyltryptophan and **e,** *m/z* 264 for 5-methyl-nitrotryptophan.



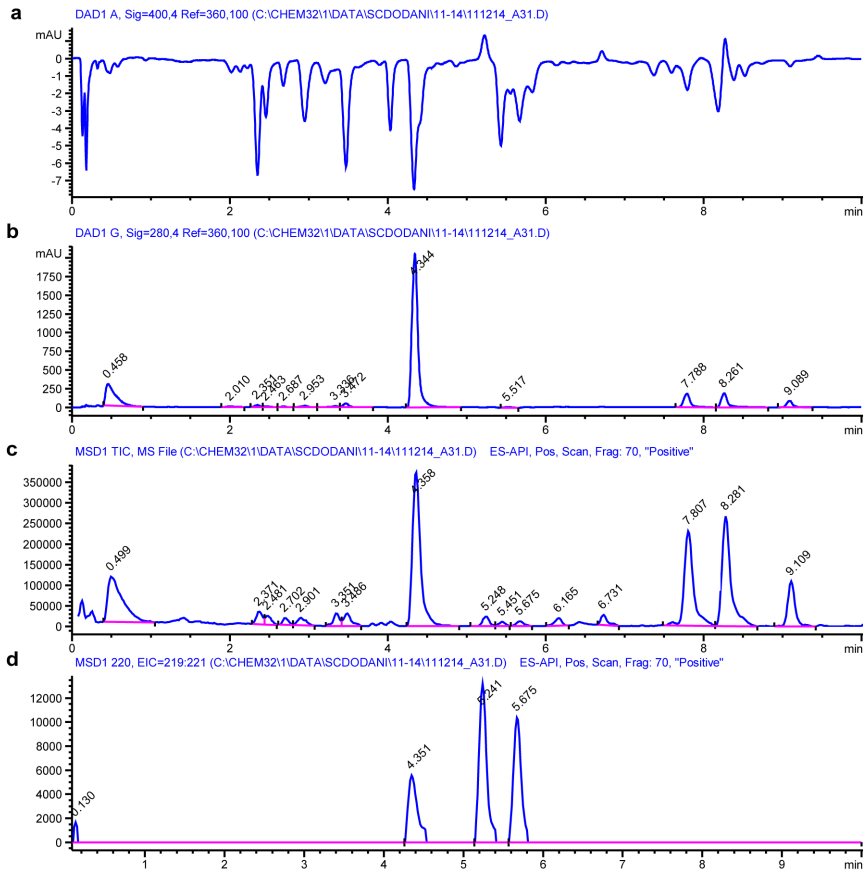
Supplementary Figure S23| LC-MS analysis of the nitration reaction with TxtE His176Phe and 5-methyltryptophan. DAD signal at **a,** 400 nm to detect all nitrated products and **b,** 280 nm to detect 5-methyltryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 220 for 5-methyltryptophan and **e,** *m/z* 264 for 5-methyl-nitrotryptophan.



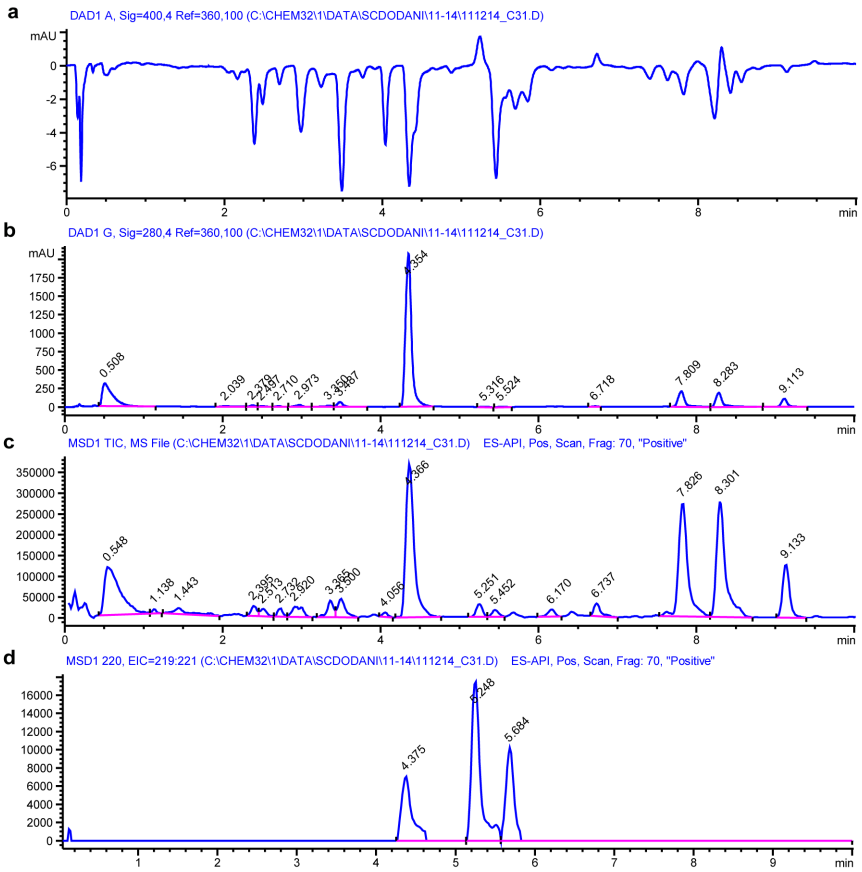
Supplementary Figure S24| LC-MS analysis of the nitration reaction with TxtE and alpha-methyltryptophan. DAD signal at **a,** 400 nm to detect all nitrated products and **b,** 280 nm to detect alpha-methyltryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 220 for alpha-methyltryptophan and **e,** *m/z* 264 for alpha-methyl-nitrotryptophan.



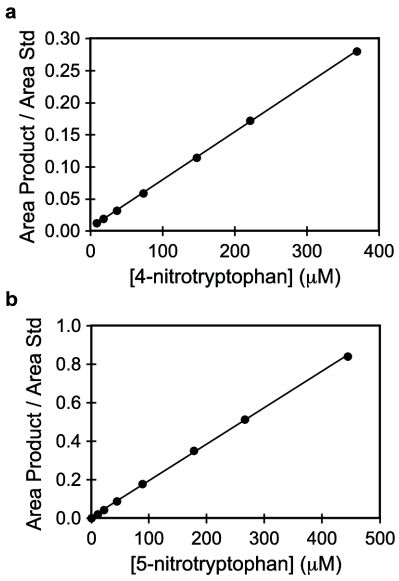
Supplementary Figure S25| LC-MS analysis of the nitration reaction with His176Phe and alpha-methyltryptophan. DAD signal at **a,** 400 nm to detect all nitrated products and **b,** 280 nm to detect 5-methyltryptophan. **c,** LC-MS trace of the reaction mixture. Extracted ion chromatograms at **d,** *m/z* 220 for alpha-methyltryptophan and **e,** *m/z* 264 for alpha-methyl-nitrotryptophan.



Supplementary Figure S26| LC-MS analysis of the nitration reaction with TxtE and indole-3-acetamide. DAD signal at **a,** 400 nm to detect all nitrated products and **b,** 280 nm to detect indole-3-acetamide. **c,** LC-MS trace of the reaction mixture. **d,** Extracted ion chromatogram at *m/z* 220 for nitro-indole-3-acetamide.

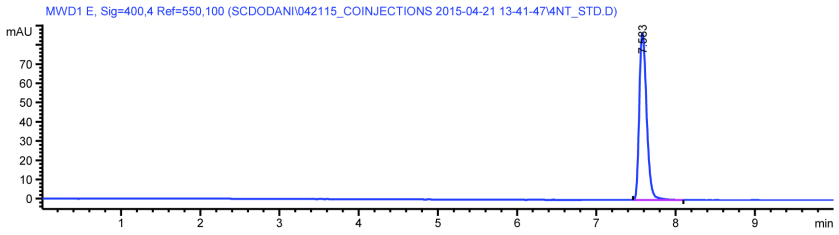


Supplementary Figure S27| LC-MS analysis of the nitration reaction with His176Phe and indole-3-acetamide. DAD signal at **a,** 400 nm to detect all nitrated products and **b,** 280 nm to detect indole-3-acetamide. **c,** LC-MS trace of the reaction mixture. **d,** Extracted ion chromatogram at *m/z* 220 for nitroindole-3-acetamide.



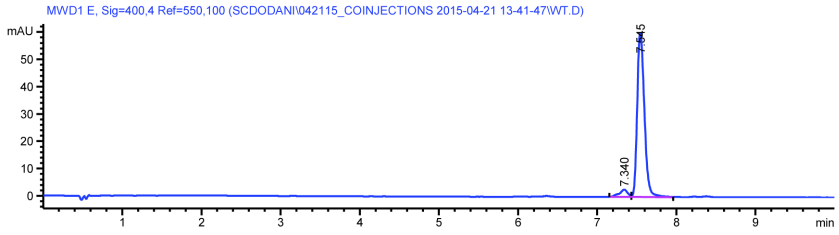
Supplementary Figure S28| Calibration curves for the **a,** 4-nitrotryptophan and **b,** 5-nitrotryptophan products with the concentration of product (x-axis) versus the ratio of the HPLC area of the product to the internal standard (y-axis).

a



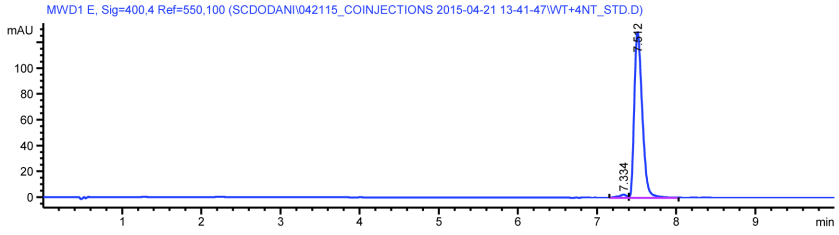
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.583	BB	0.1015	563.42383	87.07493	100.0000

b



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.340	BV	0.1069	20.07295	2.69434	4.8915
2	7.545	VB	0.0971	379.04486	60.46197	92.3681

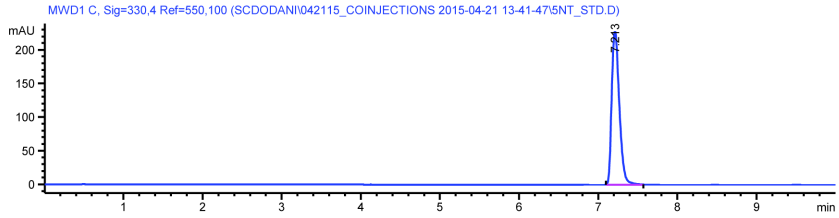
c



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.334	BV	0.1044	16.84738	2.38529	1.8207
2	7.512	VB	0.1091	897.49316	129.17084	96.9929

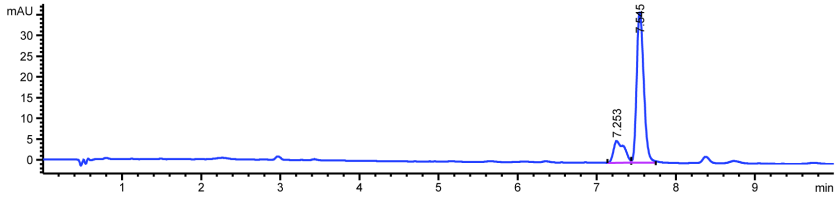
Supplementary Figure S29| Co-injection HPLC traces for the nitration reaction with wild-type TxtE. Multi-wavelength detector (MWD) signal at 400 nm of the **a,** synthetically prepared 4-nitrotryptophan standard, **b,** enzyme reaction mixture, and **c,** 4-nitrotryptophan standard and the enzyme reaction mixture.

a



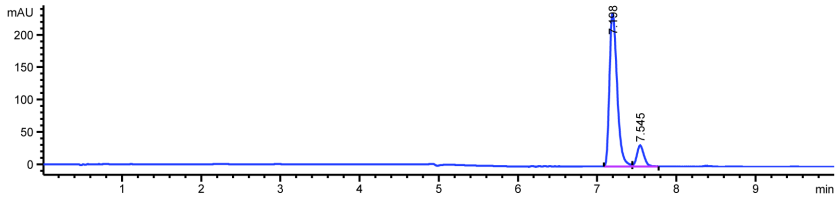
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.213	BB	0.1040	1532.26831	229.05194	100.0000

b



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.253	BV	0.1303	49.85480	5.27340	17.4610
2	7.545	VB	0.0963	226.10982	36.50229	79.1919

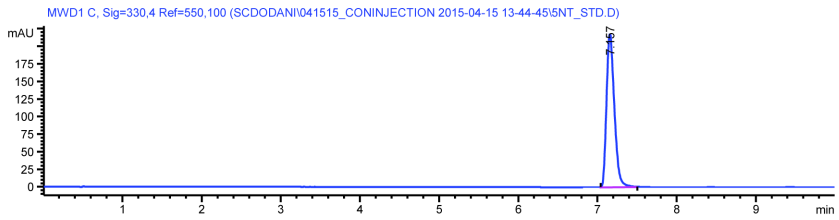
c



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.198	BV	0.1046	1597.72681	237.20305	87.6655
2	7.545	VB	0.0979	207.27501	32.72778	11.3730

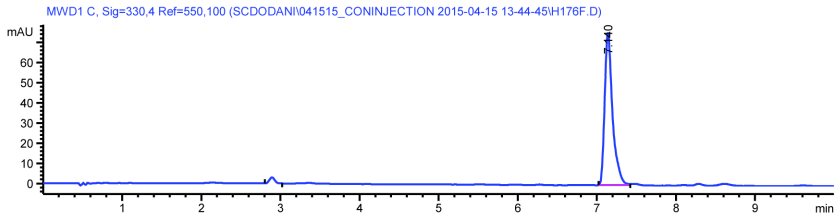
Supplementary Figure S30| Co-injection HPLC traces for the nitration reaction with wild-type TxtE. MWD signal at 330 nm of the **a,** commercial 5-nitrotryptophan standard, **b,** enzyme reaction mixture, and **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture.

a



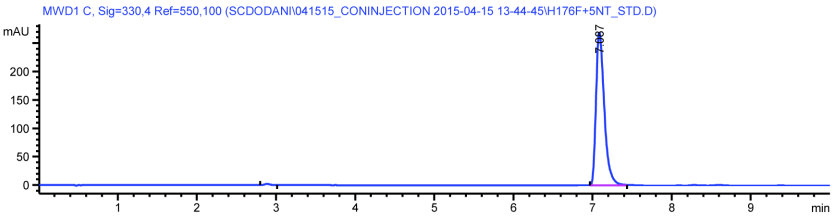
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.157	BB	0.1060	1500.49316	218.86983	100.0000

b



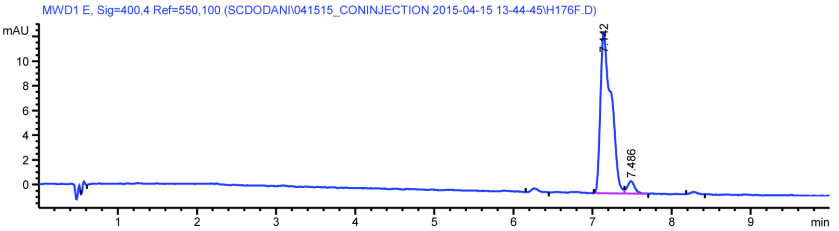
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.140	BB	0.1043	521.45386	75.76968	97.4167

c



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.087	BB	0.1105	1960.95752	270.65097	99.5612

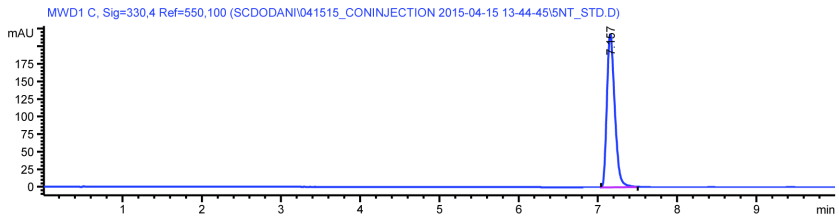
d



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.142	BV	0.1314	122.80553	13.09828	89.1892
2	7.486	VB	0.0993	6.61063	9.98269e-1	4.8011

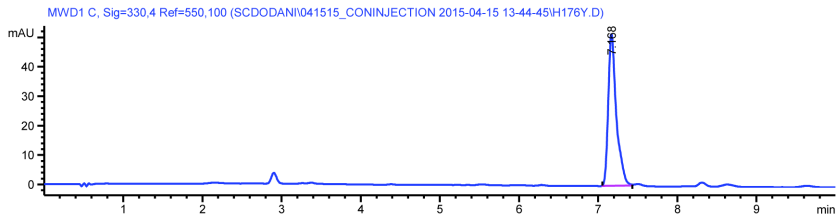
Supplementary Figure S31| Co-injection HPLC traces for the nitration reaction with TxtE His176Phe. MWD signal of the **a,** commercial 5-nitrotryptophan standard at 330 nm, **b**, enzyme reaction mixture at 330 nm, **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture at 330 nm, and **d,** enzyme reaction mixture at 400 nm.

a



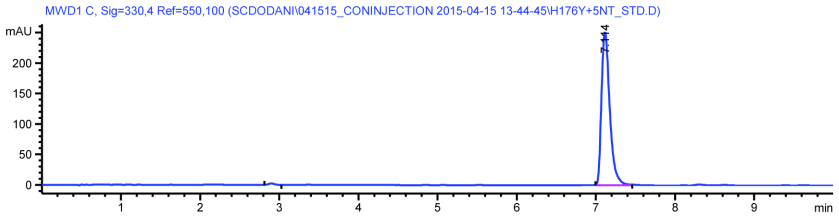
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.157	BB	0.1060	1500.49316	218.86983	100.0000

b



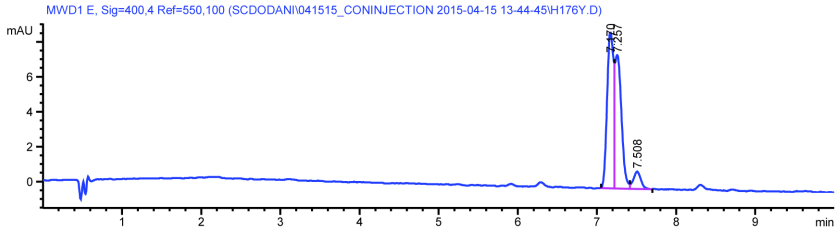
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.168	BB	0.1047	366.81516	51.74812	93.7126

c



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.114	BB	0.1104	1825.82715	252.29512	98.8082

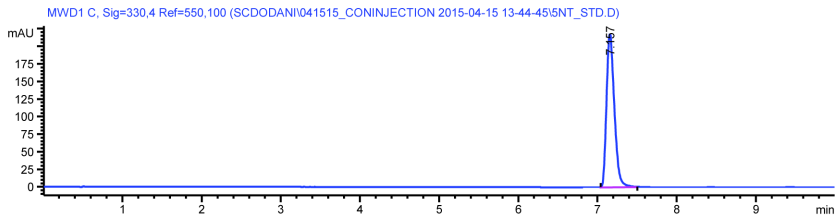
d



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.170	BV	0.0873	51.84539	8.99170	39.8097
2	7.257	VV	0.0844	43.52043	7.64252	33.4174
3	7.508	VB	0.0988	6.50141	9.88002e-1	4.9921

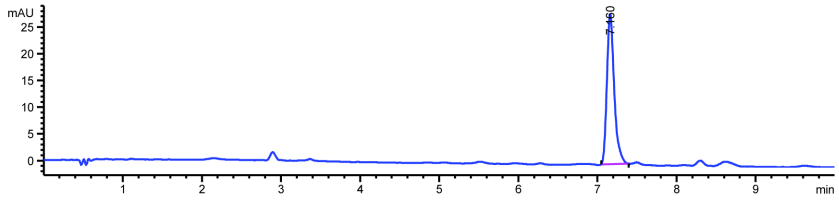
Supplementary Figure S32| Co-injection HPLC traces for the nitration reaction with TxtE His176Tyr. MWD signal of the **a,** commercial 5-nitrotryptophan standard at 330 nm, **b,** enzyme reaction mixture at 330 nm, **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture at 330 nm, and **d,** enzyme reaction mixture at 400 nm.

a



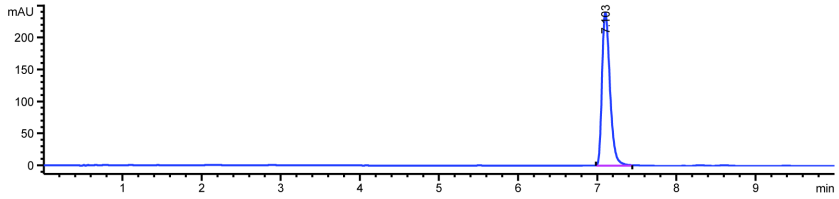
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.157	BB	0.1060	1500.49316	218.86983	100.0000

b



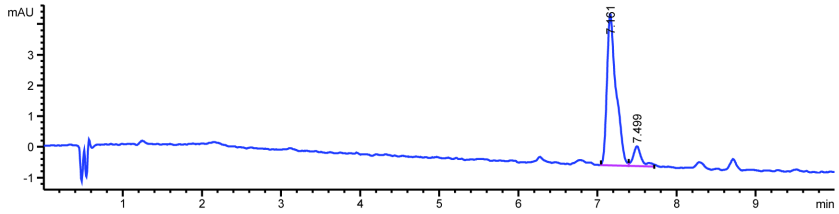
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.160	BB	0.0969	181.24802	28.23374	93.3094

c



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.103	BB	0.1071	1670.66504	240.34787	100.0000

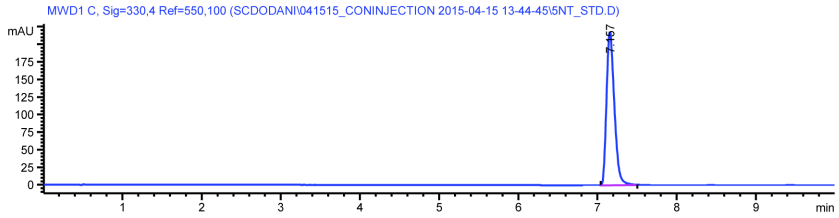
d



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.161	BV	0.1152	39.52575	4.94924	72.6399
2	7.499	VB	0.1075	4.70099	6.41190e-1	8.6394

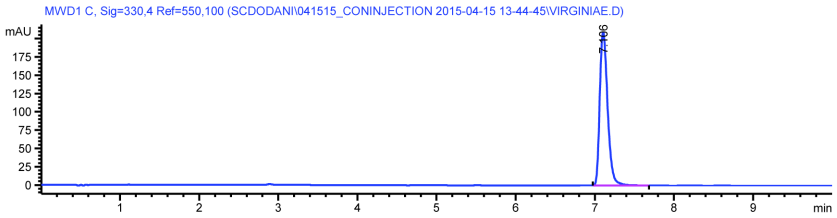
Supplementary Figure S33| Co-injection HPLC traces for the nitration reaction with TxtE His176Trp. MWD signal of the **a,** commercial 5-nitrotryptophan standard at 330 nm, **b,** enzyme reaction mixture at 330 nm, **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture at 330 nm, and **d,** enzyme reaction mixture at 400 nm.

a



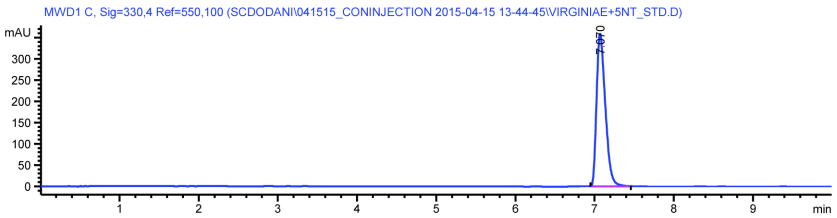
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.157	BB	0.1060	1500.49316	218.86983	100.0000

b



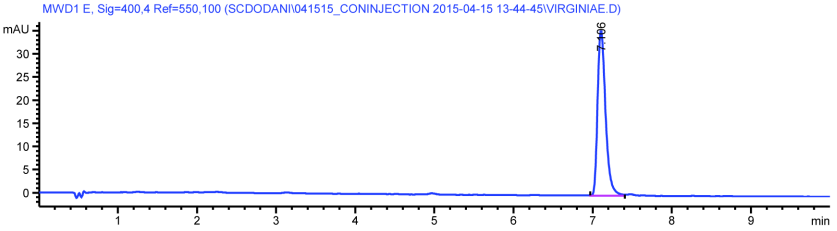
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.106	BB	0.1067	1453.48523	210.25711	100.0000

c



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.070	BB	0.1173	2770.19189	361.86639	100.0000

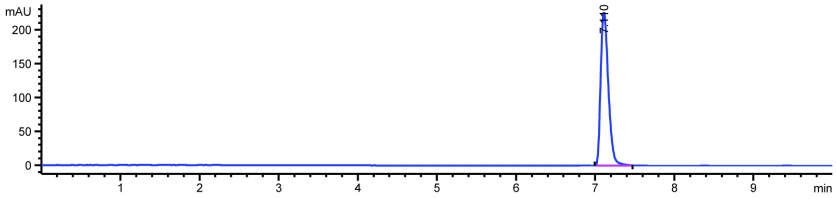
d



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.106	VV	0.1073	249.36900	35.79520	100.0000

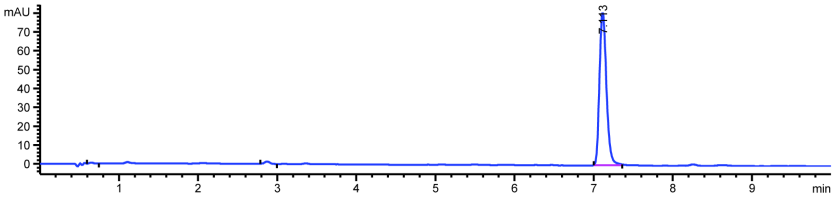
Supplementary Figure S34| Co-injection HPLC traces for the nitration reaction with Virginiae TxtE. MWD signal of the **a,** commercial 5-nitrotryptophan standard at 330 nm, **b,** enzyme reaction mixture at 330 nm, **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture at 330 nm, and **d,** enzyme reaction mixture at 400 nm.

a



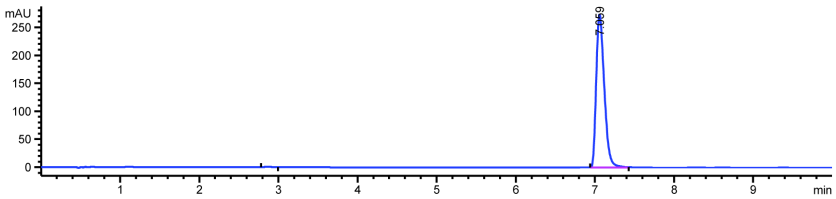
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.110	BB	0.1023	1527.08618	227.60956	100.0000

b



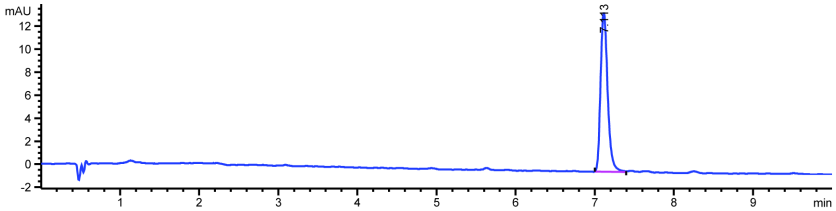
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.113	BB	0.0949	493.87756	81.29138	97.6660

c



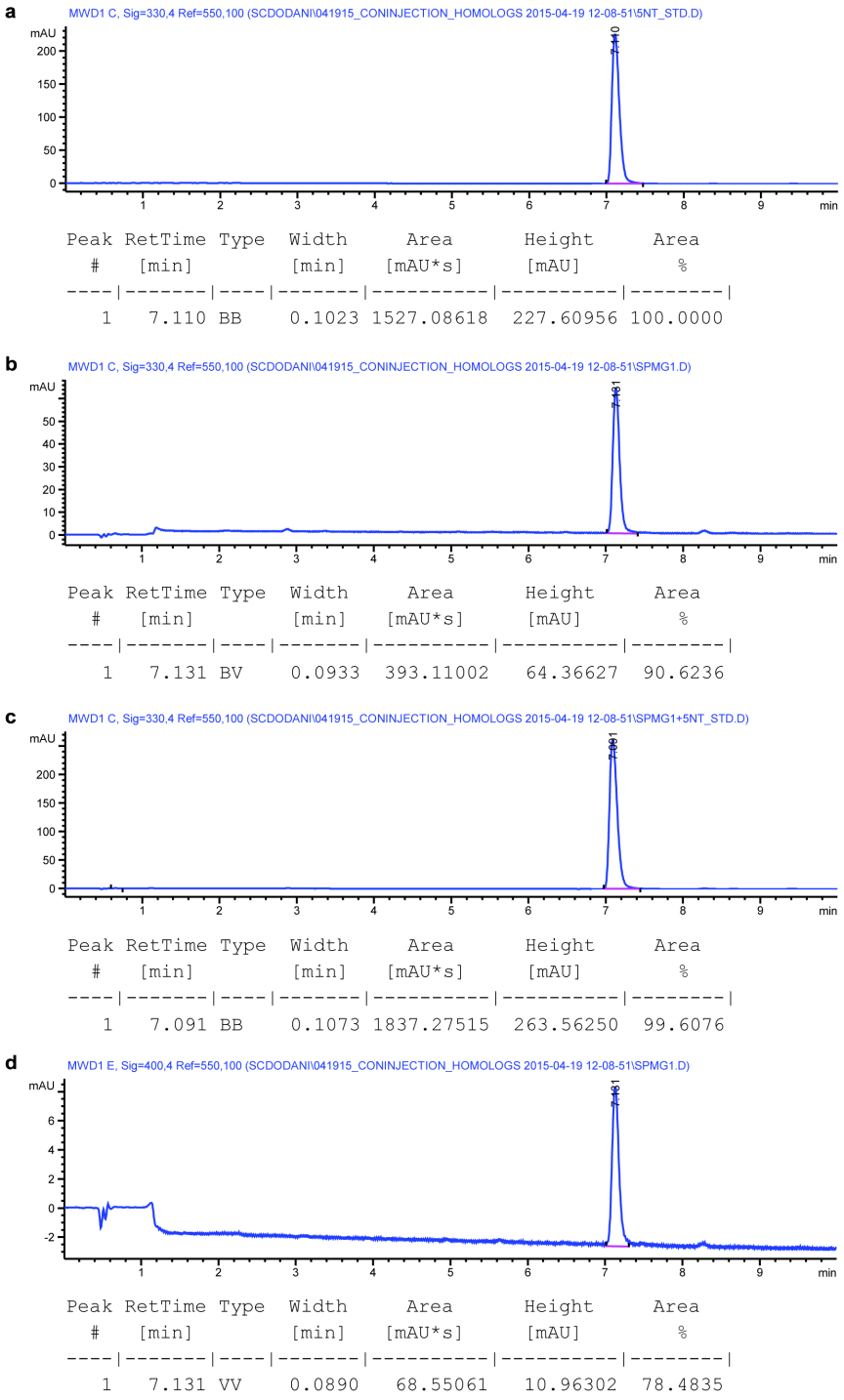
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.059	BB	0.1103	1937.45166	274.72665	99.7489

d

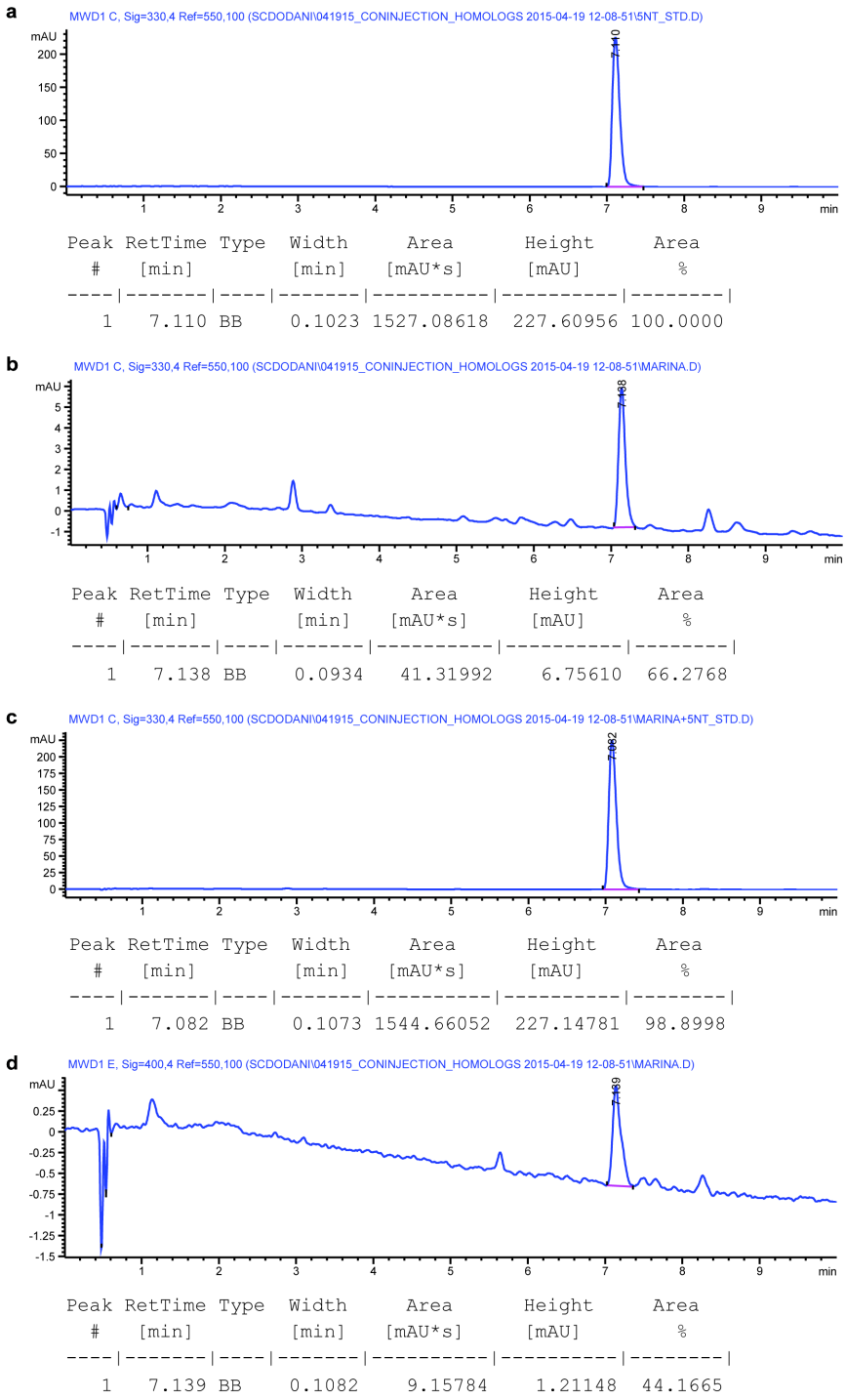


Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	7.113	BB	0.0962	85.89352	13.88463	89.3180

Supplementary Figure S35| Co-injection HPLC traces for the nitration reaction with Lavendulae TxtE. MWD signal of the **a,** commercial 5-nitrotryptophan standard at 330 nm, **b,** enzyme reaction mixture at 330 nm, **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture at 330 nm, and **d,** enzyme reaction mixture at 400 nm.

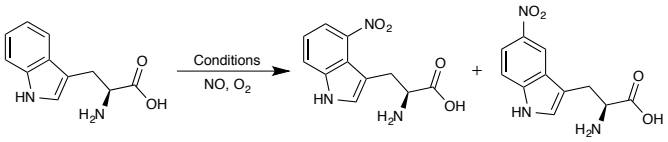


Supplementary Figure S36| Co-injection HPLC traces for the nitration reaction with sp. Mg1 TxtE. MWD signal of the **a,** commercial 5-nitrotryptophan standard at 330 nm, **b,** enzyme reaction mixture at 330 nm, **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture at 330 nm, and **d,** enzyme reaction mixture at 400 nm.



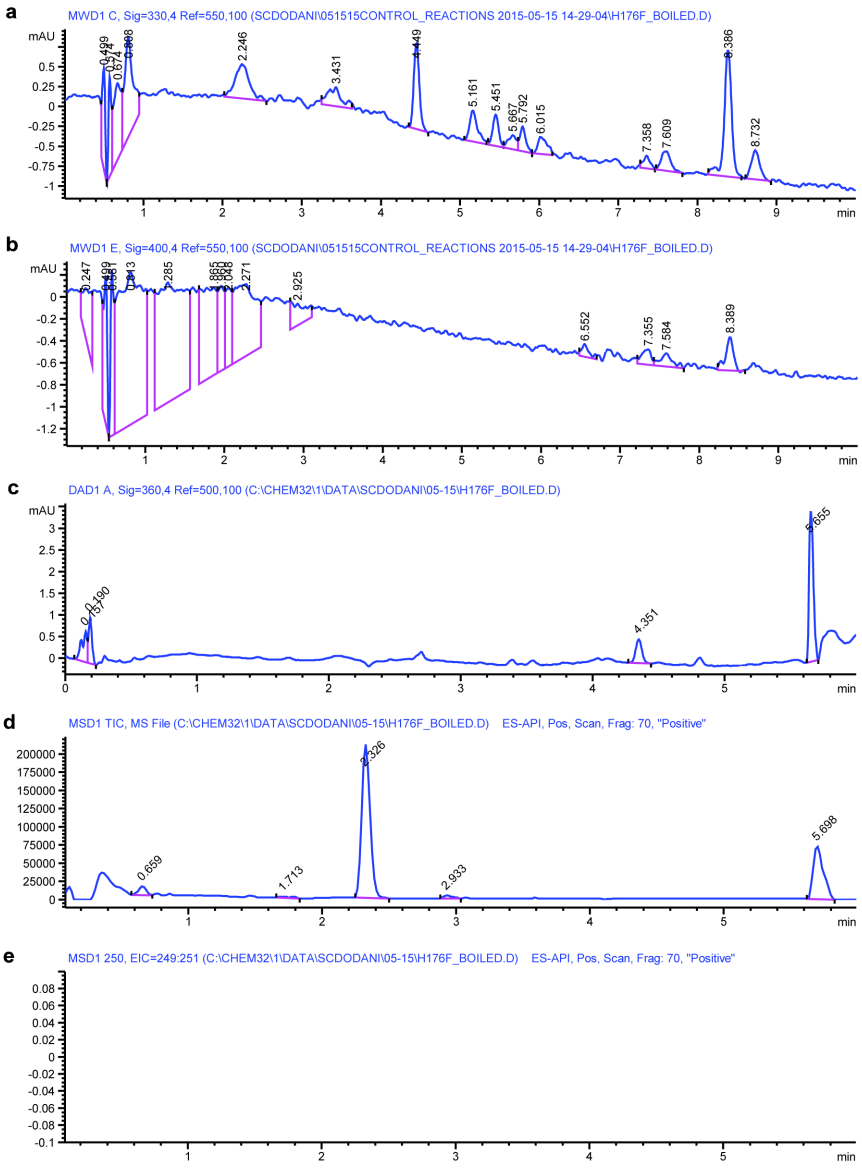
Supplementary Figure S37| Co-injection HPLC traces for the nitration reaction with Marina TxtE. MWD signal of the **a,** commercial 5-nitrotryptophan standard at 330 nm, **b,** enzyme reaction mixture at 330 nm, **c,** commercial 5-nitrotryptophan standard and the enzyme reaction mixture at 330 nm, and **d,** enzyme reaction mixture at 400 nm.

Supplementary Table S1| Control experiments.

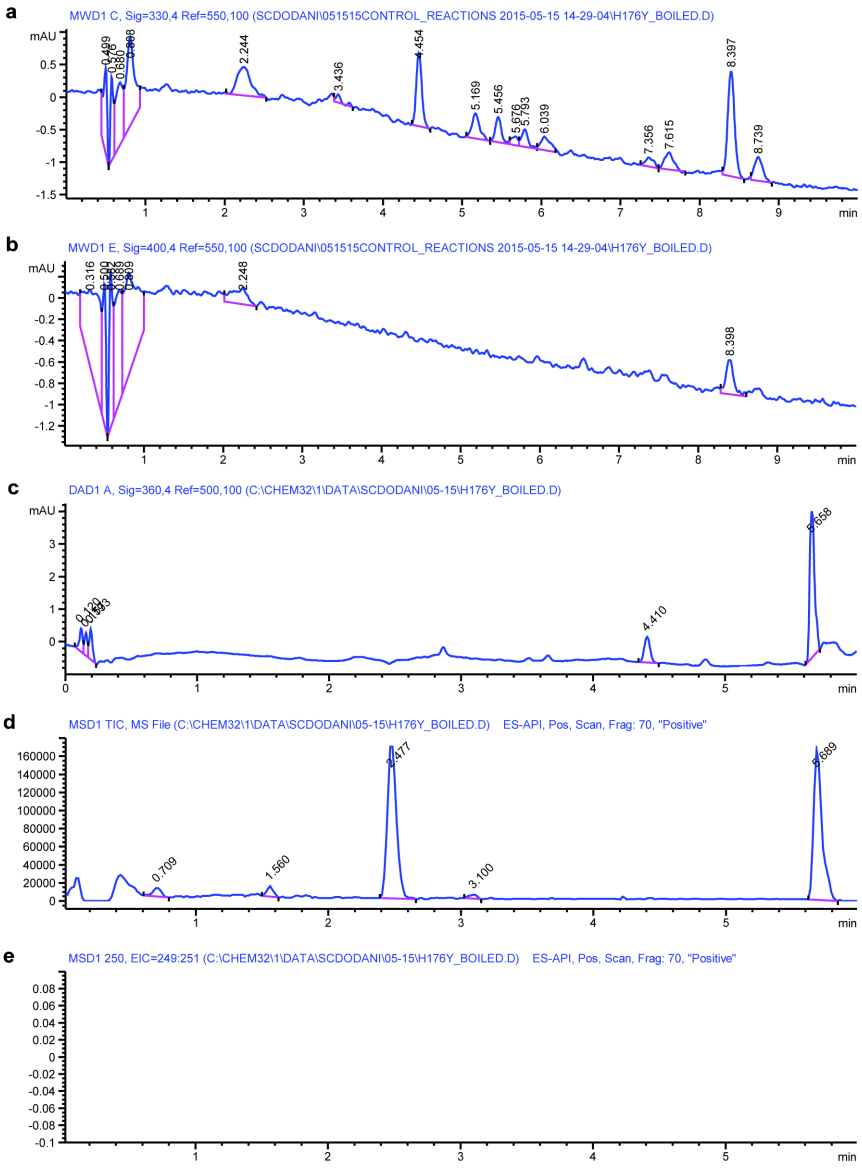


Entry	Condition	TTN[a]
1	TxtE + CS[b]	84
2	TxtE His176Phe + CS	38
3	TxtE His176Tyr + CS	17
4	TxtE His176Trp + CS	6
5	Virginiae TxtE + CS	57
6	Boiled TxtE +CS	0
7	Boiled TxtE His176Phe +CS	0
8	Boiled TxtE His176Tyr + CS	0
9	Boiled TxtE His176Trp + CS	0
10	Boiled Virginiae TxtE + CS	0
11	Hemin + $Na_2S_2O_4$	0
12	Hemin + $Na_2S_2O_4$ + BSA	0

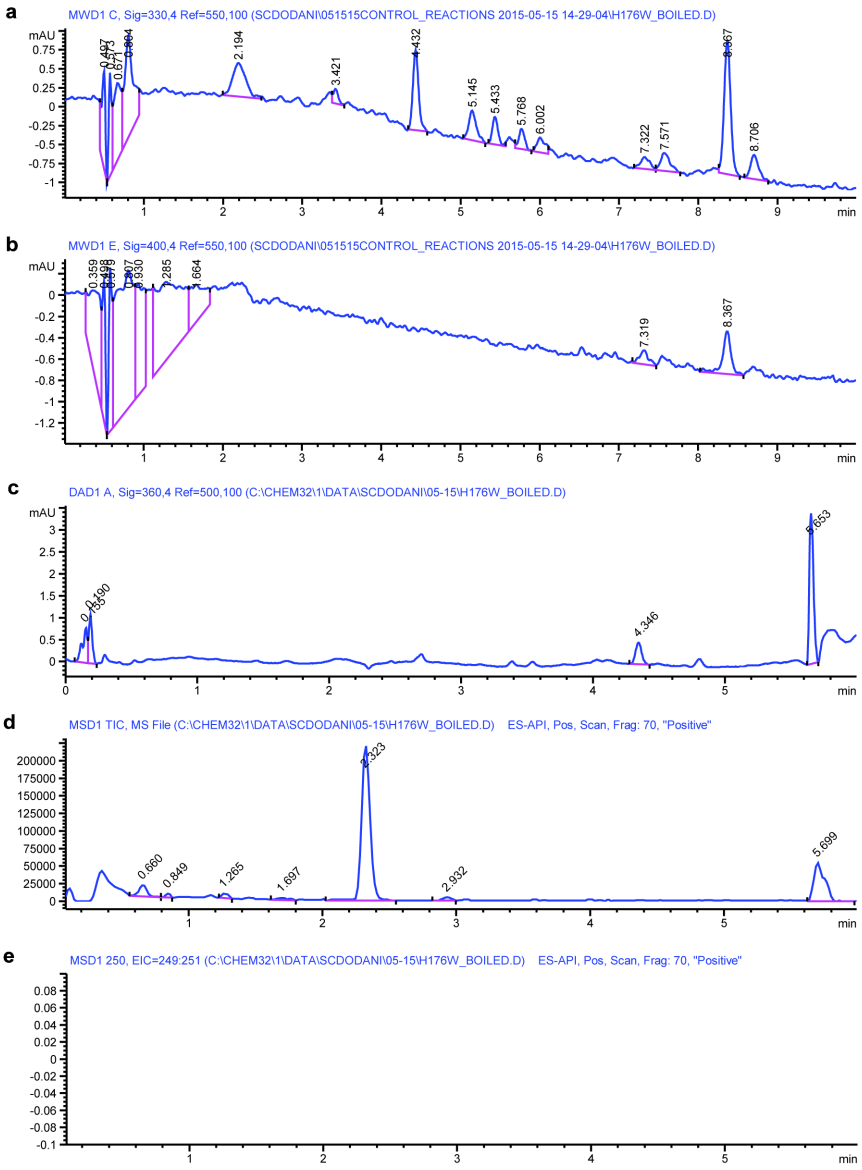
[a]Total turnover numbers (TTN) reported for the major product. [b] CS is a mixture which includes the Fd/Fr reductase system and NADPH. Reaction conditions and protein sequences are described in the *Methods* section. TTNs and regioselectivities were determined by HPLC.



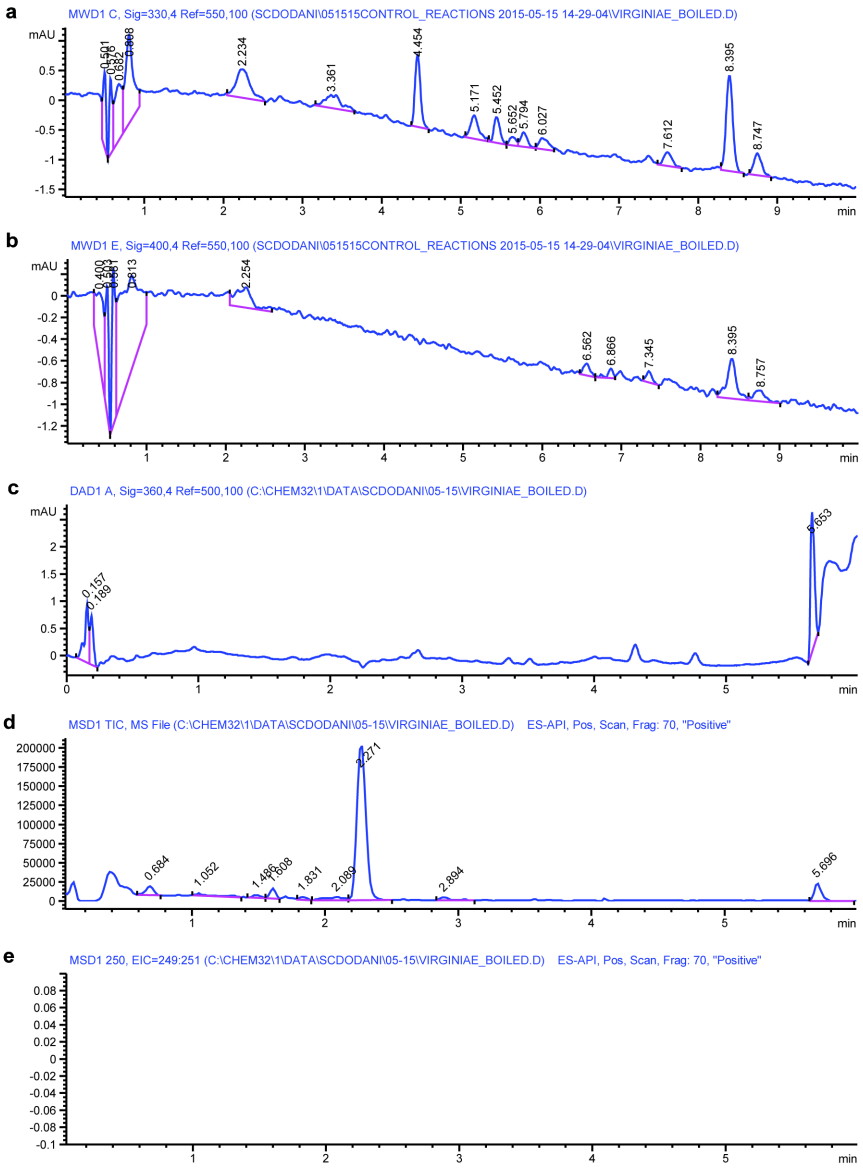
Supplementary Figure S38| Analysis of the nitration reaction with boiled TxtE His176Phe. HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from the LC-MS analysis at *m/z* 250 for nitrotryptophan.



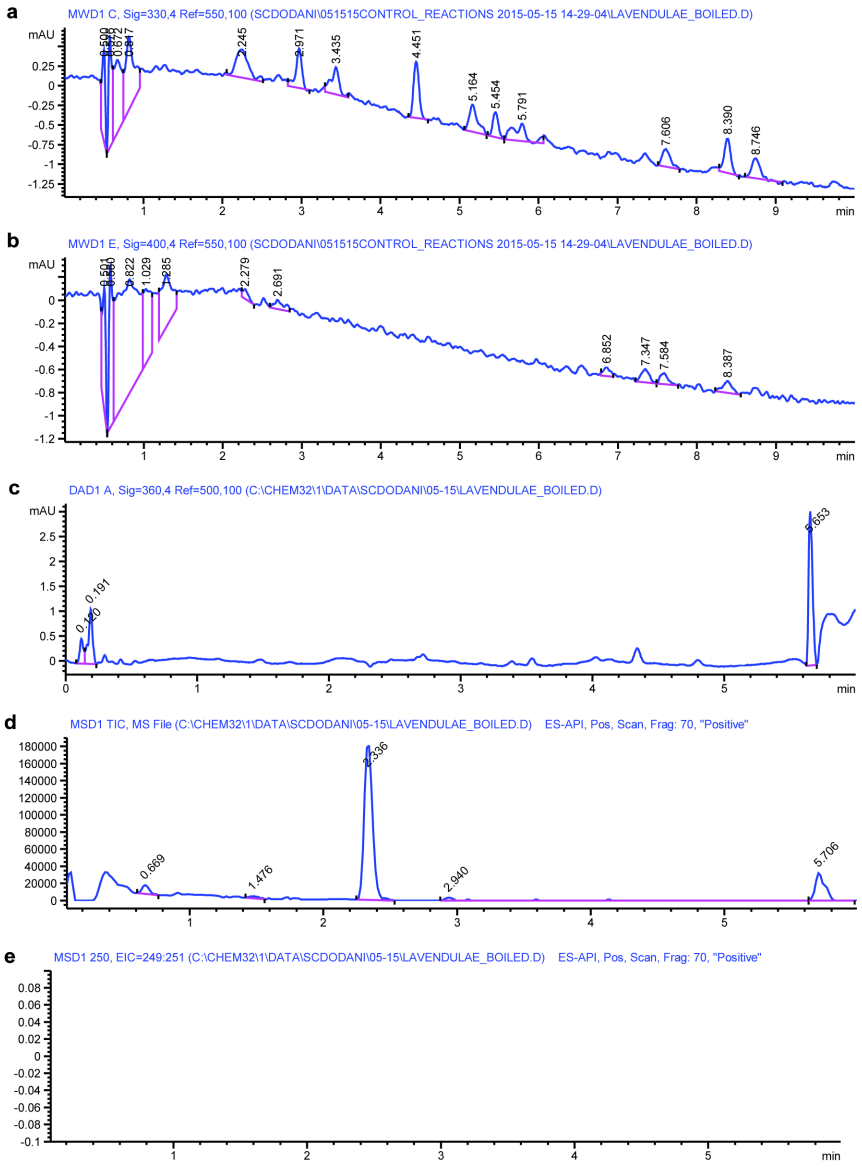
Supplementary Figure S39| Analysis of the nitration reaction with boiled TxtE His176Tyr. HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



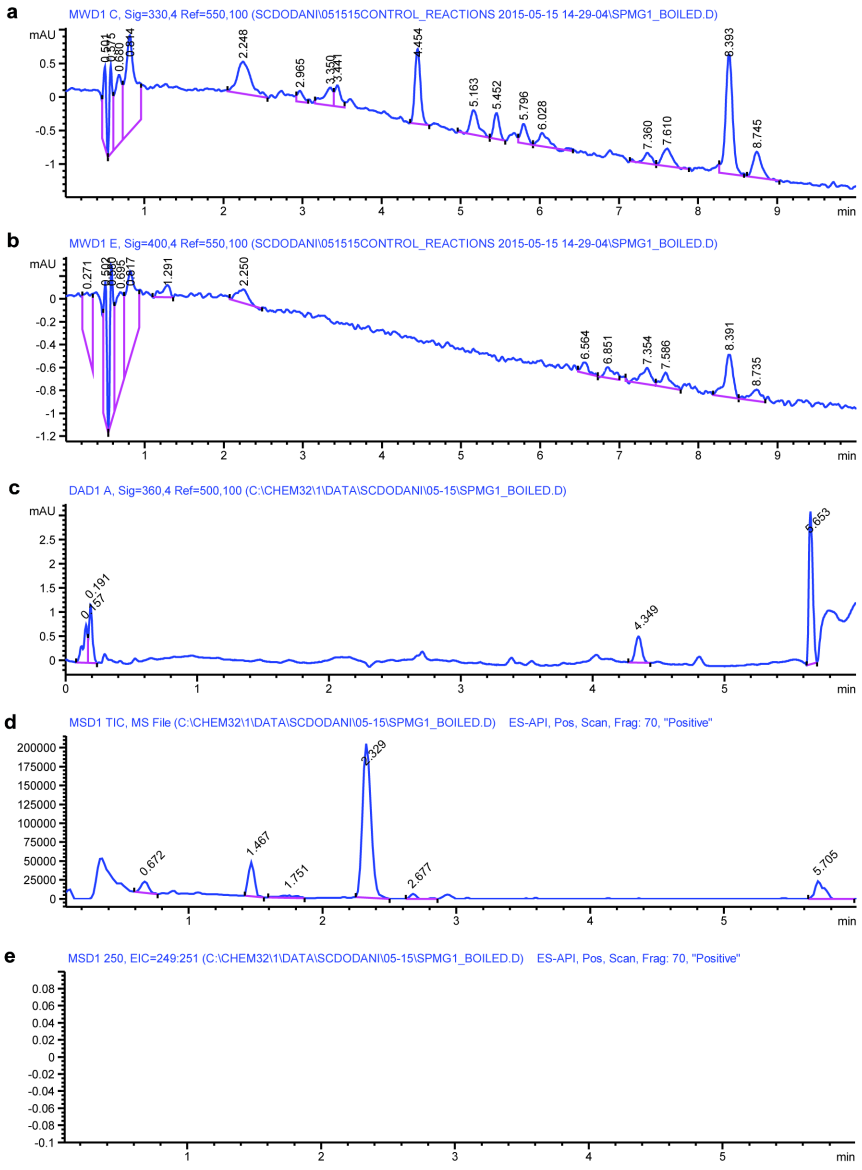
Supplementary Figure S40| Analysis of the nitration reaction with boiled TxtE His176Trp. HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



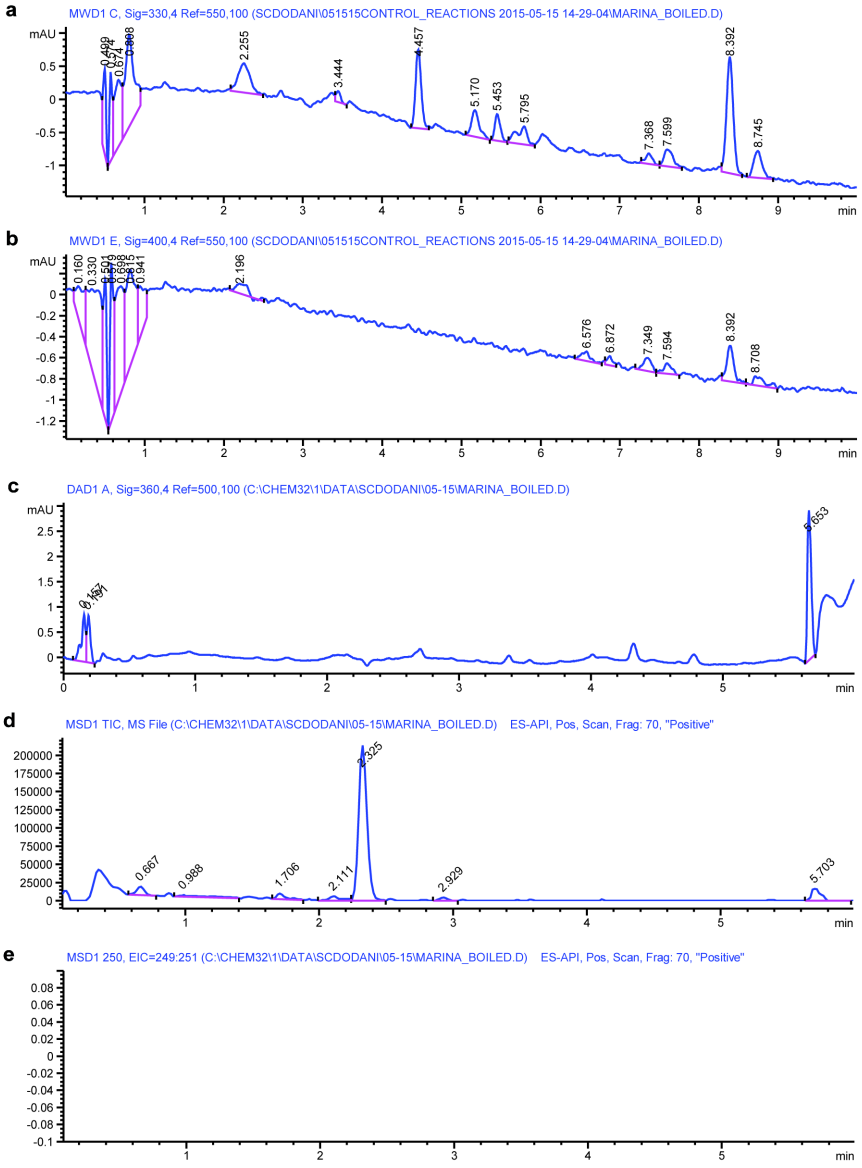
Supplementary Figure S41| Analysis of the nitration reaction with boiled Virginiae TxtE. HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



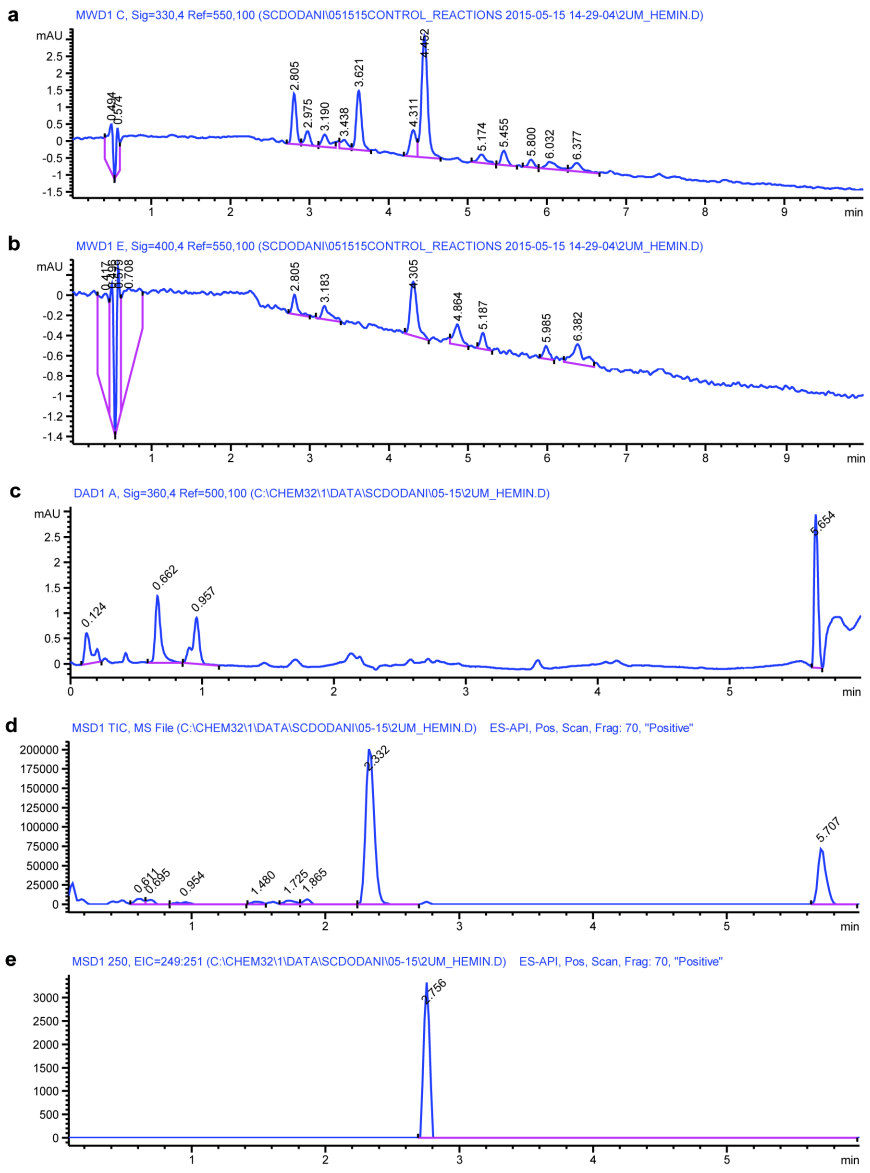
Supplementary Figure S42| Analysis of the nitration reaction with boiled Lavendulae TxtE. HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



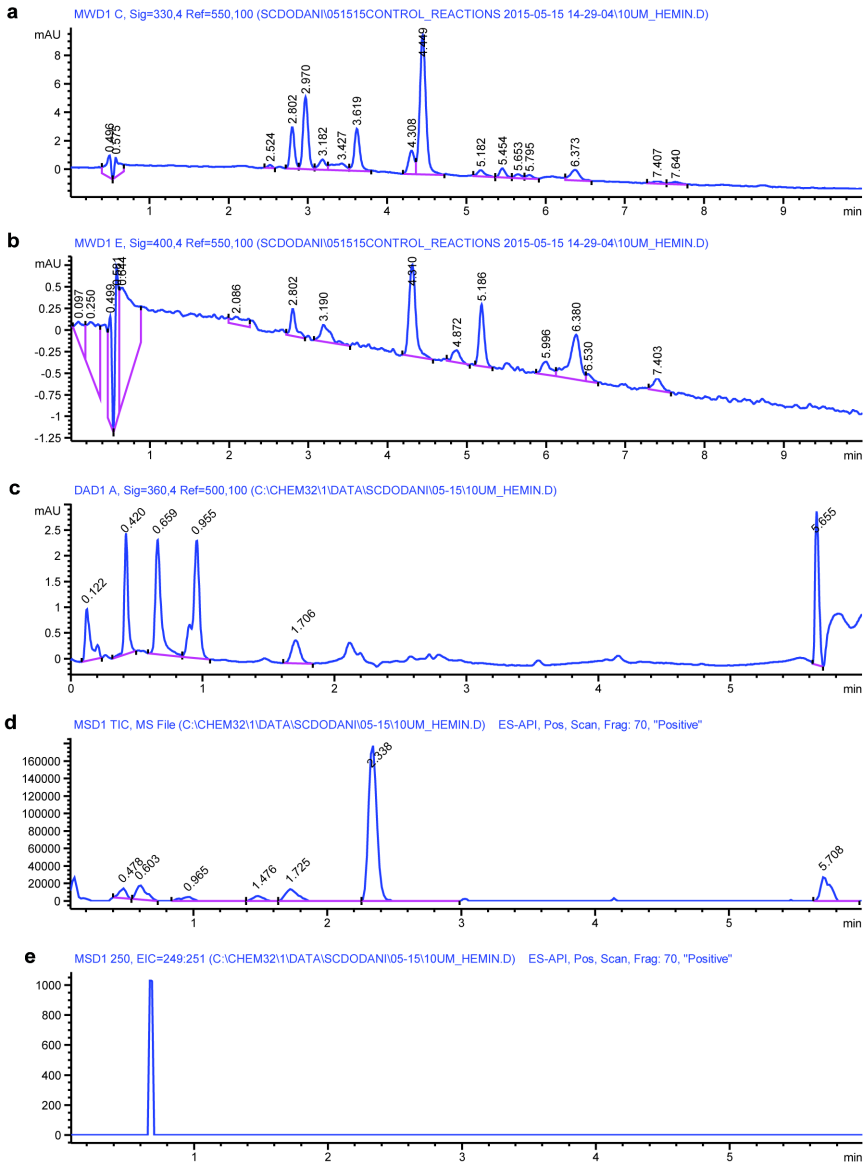
Supplementary Figure S43| Analysis of the nitration reaction with boiled sp. Mg1 TxtE. HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



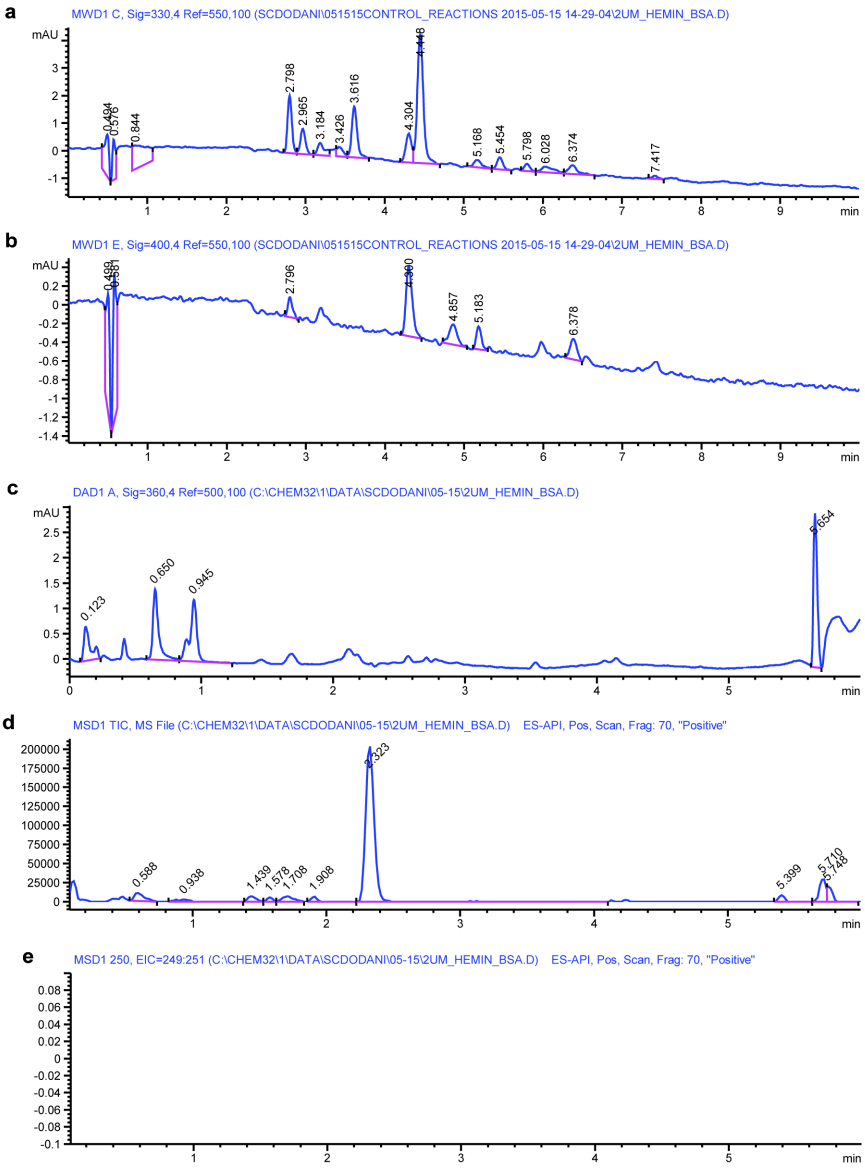
Supplementary Figure S44| Analysis of the nitration reaction with boiled Marina TxtE. HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



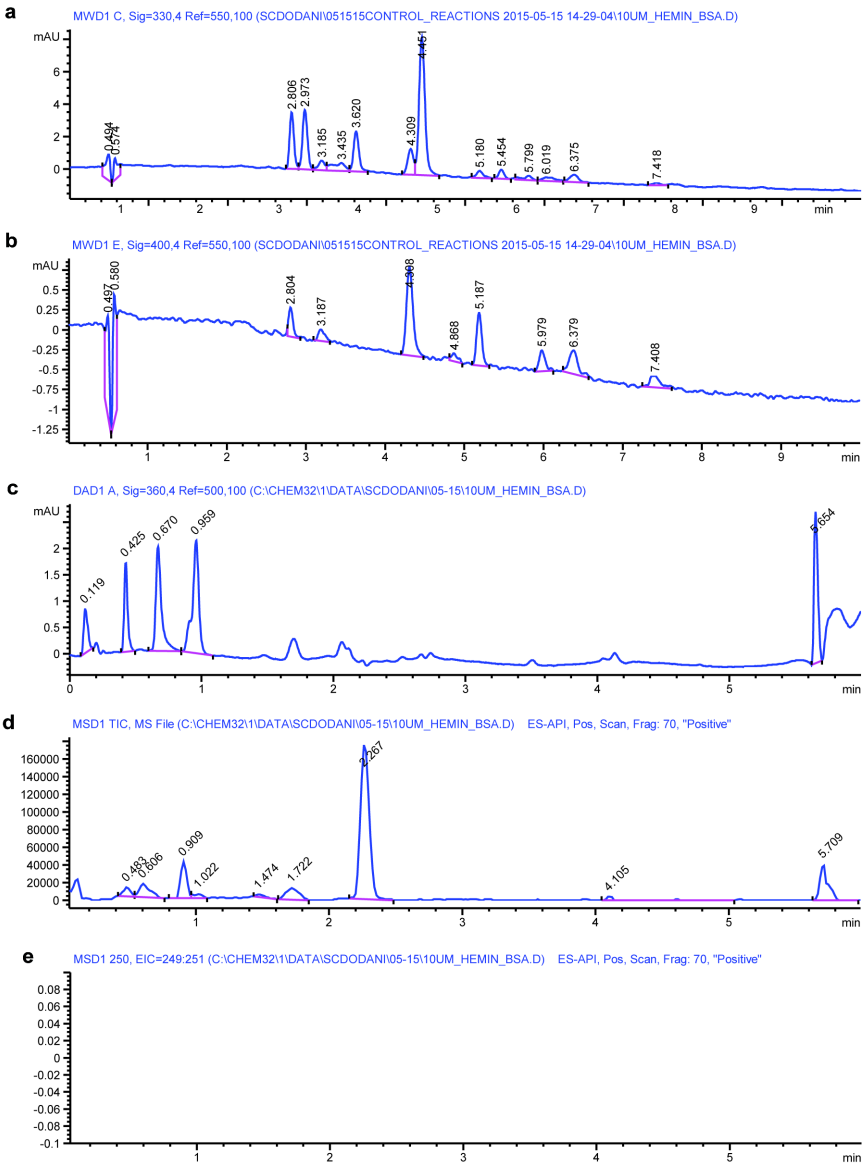
Supplementary Figure S45| Analysis of the nitration reaction with hemin (2 μM). HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



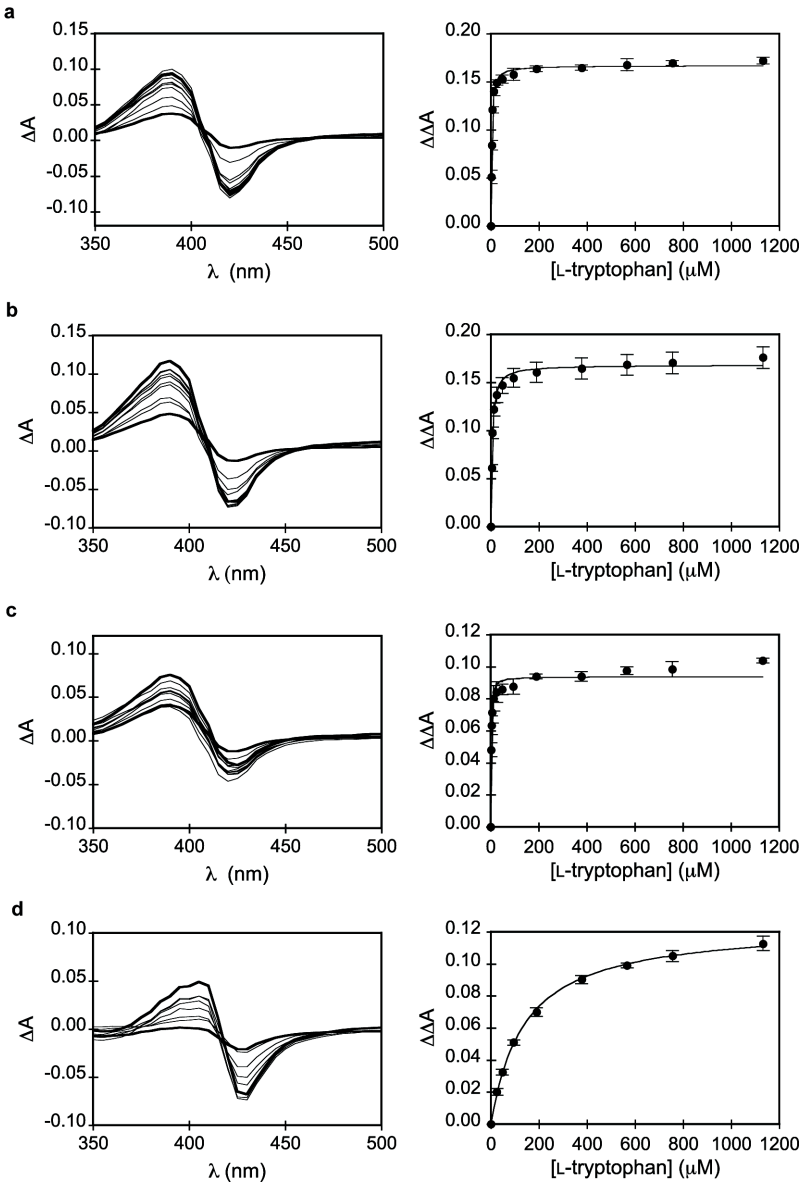
Supplementary Figure S46| Analysis of the nitration reaction with hemin (10 μM). HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



Supplementary Figure S47| Analysis of the nitration reaction with hemin/BSA (2 μM). HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.



Supplementary Figure S48| Analysis of the nitration reaction with hemin/BSA (10 μM). HPLC MWD signal at **a,** 330 nm to detect 5-nitrotryptophan and **b,** 400 nm to detect 4-nitrotryptophan. **c,** LC-MS DAD signal at 360 nm to detect all nitrated products. **d,** LC-MS trace of the reaction mixture. **e,** Extracted ion chromatogram from LC-MS analysis at *m/z* 250 for nitrotryptophan.

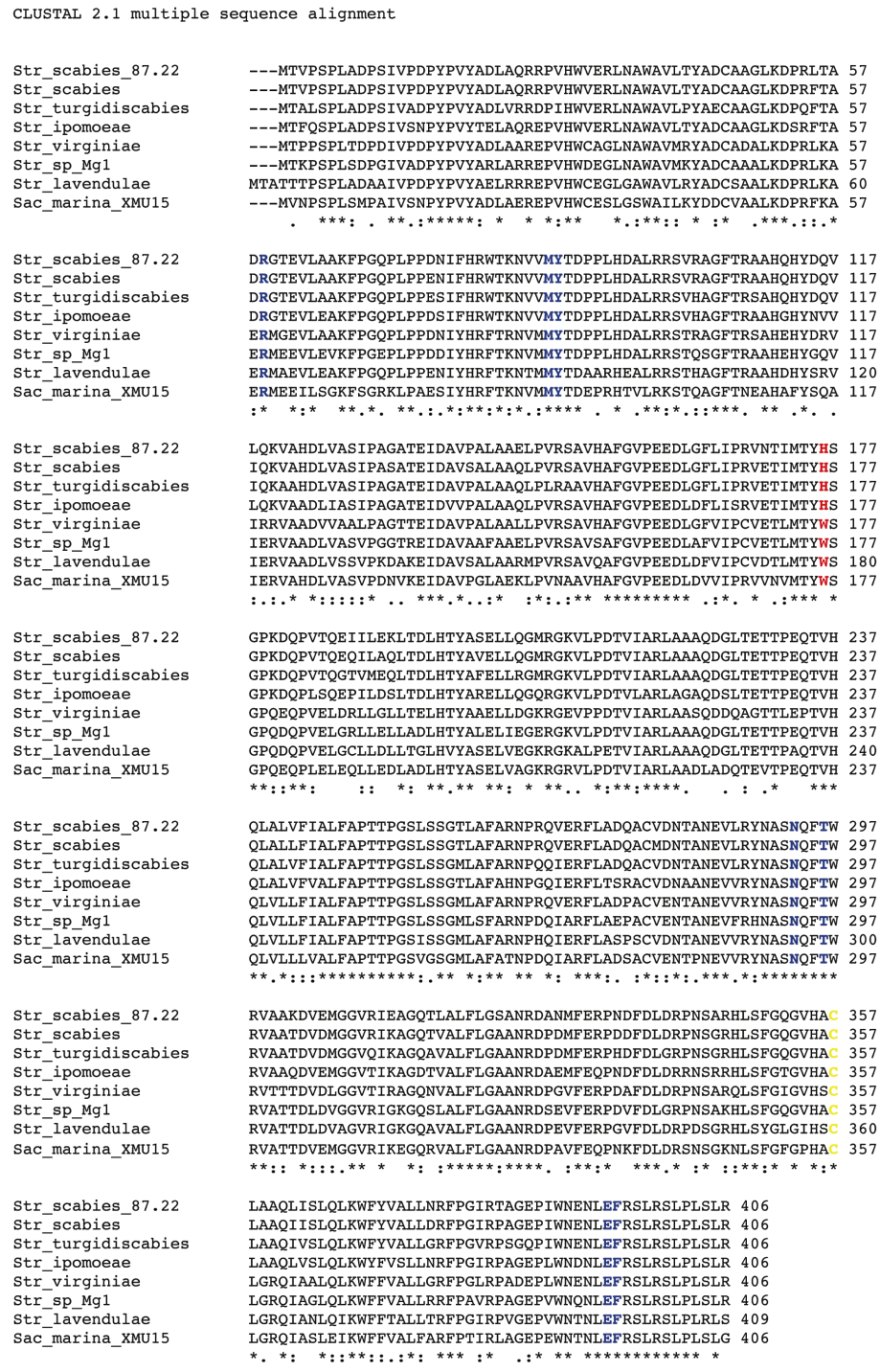


Supplementary Figure S49| Analysis of L-tryptophan binding to a, TxtE His176Phe, b, TxtE His176Tyr, c, TxtE His176Trp and d, Virginiae TxtE. For each enzyme, differential UV-visible spectra are plotted from the titration of the enzyme with L-tryptophan. The bold black lines mark the starting and end points of the substrate titration (left). The plots show the difference in the maximum and minimum absorption (ΔΔA) versus the substrate concentration for each enzyme (right). Each data set is fitted to a binding isotherm model using KaleidaGraph. All points are the mean of at least three measurements and error bars show ± one standard deviation.

Supplementary Table S2| Crystallographic parameters of L-tryptophan bound TxtE His176Phe and TxtE His176Tyr.

Parameter	**TxtE His176Phe + L-tryptophan** PDB ID: 5D3U	**TxtE His176Tyr + L-tryptophan** PDB ID: 5D40
Data Collection		
Space Group	$P2_12_12_1$	$P2_12_12_1$
a (Å)	74.80	74.72
b (Å)	99.35	99.21
c (Å)	105.50	105.29
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	72.33-1.45 (1.53-1.45)	72.21-1.51 (1.59-1.51)
R_{merge} (%)	7.2 (63.7)	2.2 (33.8)
<I>/<σI>	7.2 (1.5)	8.0 (2.2)
Completeness (%)	98.1 (97.5)	97.5 (85.4)
Redundancy	4.8 (4.8)	4.7 (4.0)
Refinement		
No. reflections	129589	114245
R_{work}/R_{free} (%)	18.8/25.1 (31.5/34.8)	14.9/21.5 (26.2/29.2)
No. atoms		
Protein	6370	6346
Ligand	129	171
Water	980	1155
RMSD		
Bond lengths (Å)	0.020	0.019
Bond angles (°)	1.875	1.859
Ramachandran plot (%)		
Favored	808	809
Allowed	9	6
Outliers	0	0

[a]Values in parentheses represent statistics from the highest resolution shell.

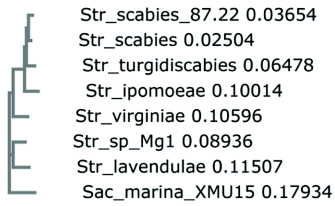


Supplementary Figure S50| Alignment of the sequences of TxtE and related homologs. Abbreviations: *Streptomyces* (Str) and *Saccharomonospora* (Sac). The axial cysteine conserved in all P450s is in yellow. Residues that are predicted to bind tryptophan are in blue. The residue corresponding to position 176 in the F/G loop is in red.

a

1: Str_scabies_87.22	100.00	93.84	87.68	85.47	76.60	76.11	71.67	67.00
2: Str_scabies	93.84	100.00	90.39	84.48	77.83	76.35	73.89	68.47
3: Str_turgidiscabies	87.68	90.39	100.00	81.53	74.63	75.12	71.43	66.26
4: Str_ipomoeae	85.47	84.48	81.53	100.00	71.18	71.43	69.21	65.76
5: Str_virginiae	76.60	77.83	74.63	71.18	100.00	80.30	76.35	69.70
6: Str_sp_Mg1	76.11	76.35	75.12	71.43	80.30	100.00	79.56	72.17
7: Str_lavendulae	71.67	73.89	71.43	69.21	76.35	79.56	100.00	70.44
8: Sac_marina_XMU15	67.00	68.47	66.26	65.76	69.70	72.17	70.44	100.00

b



Supplementary Figure S51| Percent identity matrix (a) and phylogenetic tree (b) of TxtE homologs generated with Clustal 2.1. Abbreviations: *Streptomyces* (Str) and *Saccharomonospora* (Sac).

References

1. Prime in Maestro Release 1-2015, Schrödinger, Inc., New York (2015).
2. Case, D. A. *et al*. *AMBER 12*, University of California, San Francisco (2012).
3. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
4. Wang, J., Cieplak, P. & Kollmann, K. P. How well does a restrained electrostatic potential(RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).
5. Besler, B. H., Merz, K. M., Jr & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11**, 431–439 (1990).
6. Singh, U. C., Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1990).
7. Frisch, M. J. *et al*. Gaussian 03, Revision C.02. (2004).
8. Jorgensen W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L.. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
9. Lindorff-Larsen, K. *et al*. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
10. McGibbon, R. T., Ramsundar, B., Sultan, M. M., Kiss, G. & Pande, V. S. Understanding protein dynamics with L1-regularzied reversible hiddne markov models. *arXiv* (2014).
11. Cronkite-Ratcliff, B. & Pande, V. MSMExplorer: visualizing Markov state models for biomolecule folding simulations. *Bioinformatics* **29**, 950–952 (2013).
12. Barry, S. M. *et al*. Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat. Chem. Biol.* **8**, 814–816 (2012).
13. Dodani, S. C. *et al*. Structural, functional, and spectroscopic characterization of the substrate scope of the novel nitrating cytochrome P450 TxtE. *Chembiochem* **15**, 2259–2267 (2014).
14. Engqvist, M. K. *et al*. Directed evolution of Gloeobacter violaceus rhodopsin spectral properties. *J. Mol. Biol.* **427**, 205–220 (2015).
15. Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **498**, 349–361 (2011).
16. Fuhrhop, J. H. & Smith, K. M. in *Porphyrins and Metalloporphyrins* (ed. K. M. Smith) Ch. 19, 804–807 (Elsevier, Amsterdam, 1975).
17. Buddha, M. R., Tao, T., Parry, R. J. & Crane, B. R. Regioselective nitration of tryptophan by a complex between bacterial nitric-oxide synthase and tryptophanyl-tRNA synthetase. *J. Biol. Chem.* **279**, 49567–49570 (2004).
18. Kabsch, W. XDS. *Acta Crystallogr., Section D: Biol. Crystallogr.* **66**, 125–132, (2010).
19. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Section D: Biological Crystallogr.* **62**, 72–82, (2006).
20. McCoy A. J., *et al*. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
21. Murshudov, G. N., *et al*. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Section D: Biol. Crystallogr.* **67**, 355c367 (2011).

22. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Section D: Biol. Crystallogr.* **60**, 2126–2132 (2004).
23. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, W695–699 (2010).
24. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
25. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
26. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).