Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine

Ella J. Watkins,[a] Patrick J. Almhjell,[b] and Frances H. Arnold*[a,b]

[a] Division of Biology and Biological Engineering

[b] Division of Chemistry and Chemical Engineering

California Institute of Technology,

1200 E. California Blvd, Mail Code 210-41

Pasadena, California 91125

Supporting Information

SI Table of Contents

1. Supplemental Figures

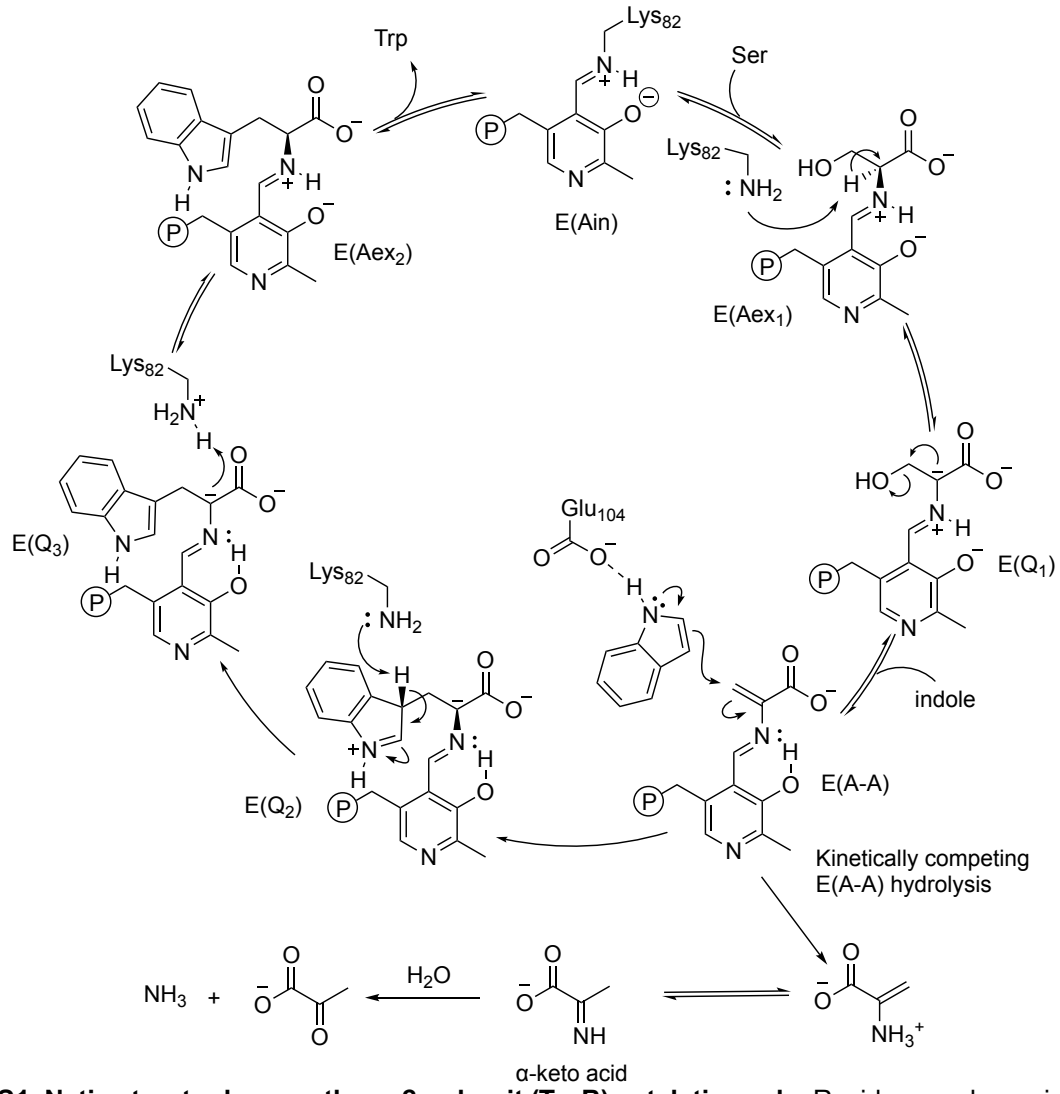


Figure S1. Native tryptophan synthase β-subunit (TrpB) catalytic cycle. Residue numbers given according to the *Pyrococcus furiosus* TrpB sequence.

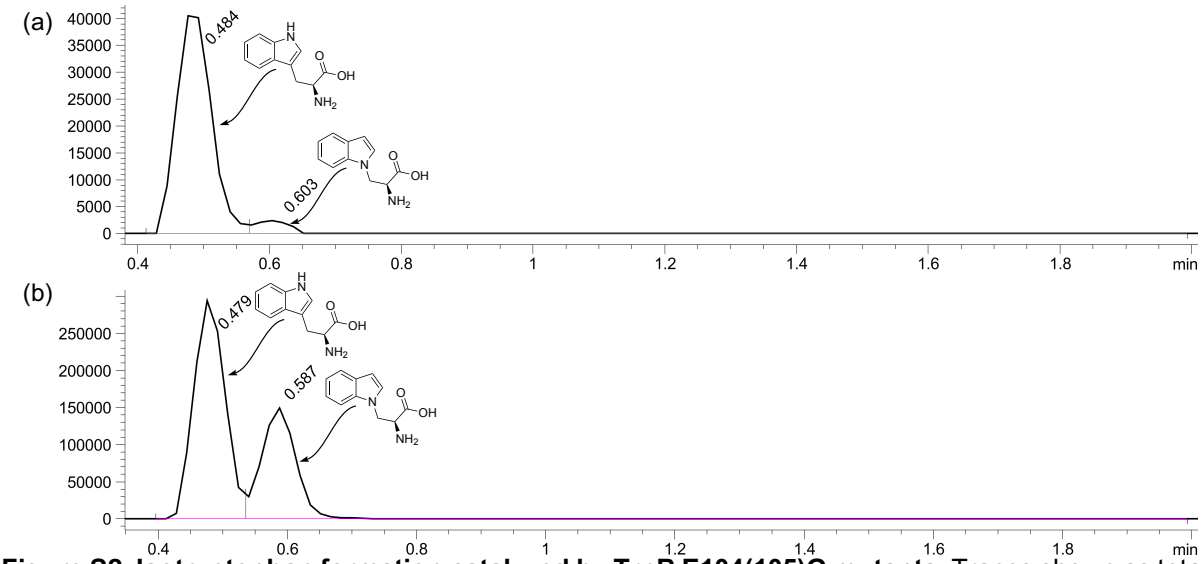


Figure S2. Isotryptophan formation catalyzed by TrpB E104(105)G mutants. Traces shown as total ion count filtered by $[M+H]^+$ for Trp (*m/z* = 205). **(a)** A trace amount of isoTrp is formed by *Pf*5G8 E104G alongside Trp. **(b)** *Tm*9D8* E105G also produces isoTrp as the minor product. Reactions prepared according to Section 3.7, with conditions given in Table S1.

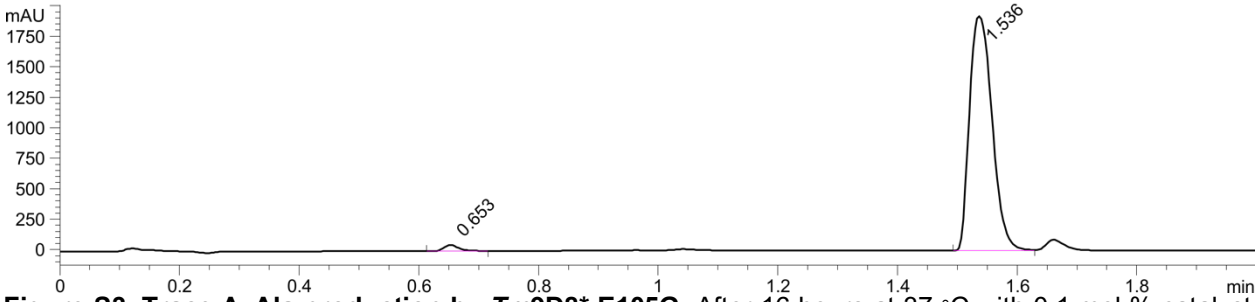


Figure S3. Trace AzAla production by *Tm*9D8* E105G. After 16 hours at 37 °C with 0.1 mol % catalyst loading, only a small amount of AzAla (peak at 0.653 min) is formed by *Tm*9D8* E105G. The reaction was prepared according to Section 3.7, but the reaction mixture was not diluted after centrifugation to increase the concentration of product in the sample. Trace AzAla yield is also seen for *Pf*5G8 E104G at 75 °C, but the sublimation of azulene interferes with analysis of the reaction at long timescales.

2. Supplemental Tables

Table S1. Rate comparisons among native and engineered TrpB enzymes[a]

	initial turnover frequency (min^{-1})			
	to Trp		to AzAla	
catalyst	E104(105)	G104(105)	E104(105)	G104(105)
*Tm*TrpS[b]	174 ± 1	—	12.0 ± 0.4	—
*Pf*5G8[b]	12.5 ± 0.3	6.8 ± 0.6	21.4 ± 1.4	n.d.
*Tm*9D8*[c]	19.0 ± 0.1	0.68 ± 0.1[d]	4.6 ± 0.6	n.d.
*Tm*AzuI[c]	—	—	14.0 ± 0.4	—

[a]Reactions performed according to Section 3.7 with 0.02 mol % catalyst loading and allowed to react for 15 minutes (Trp production) or 1 hour (AzAla production). [b]Reactions run at 75 °C. [c]Reactions run at 37 °C. [d]Reaction time and catalyst loading increased to 2 hours and 0.1 mol % catalyst loading to improve quantification. n.d. = not detected by LC-MS.

3. Experimental Procedures

3.1 General experimental methods

Chemicals and reagents were purchased from commercial sources and used without further purification. Proton and carbon NMR were recorded on a Bruker 400 MHz (100 MHz) spectrometer equipped with a cryogenic probe. Proton chemical shifts are reported in ppm (δ) relative to tetramethylsilane and calibrated using the residual solvent resonance (D_2O, δ 4.79 ppm). Data are reported as follows: chemical shift (multiplicity [singlet (s), doublet (d), doublet of doublets (dd), doublet of doublet of doublets (ddd), triplet (t), triplet of doublets (td), multiplet (m)], coupling constants [Hz], integration).

All cultures were grown in Terrific Broth supplemented with 100 μg/mL carbenicillin (TB_{carb}). Cultures were shaken in New Brunswick Innova 4000 (shaking diameter 19 mm), with the exception of 96-well plates, which were shaken in Multitron INFORS HT (shaking diameter 50 mm). Lysis buffer was composed of 50 mM potassium phosphate, pH 8.0 (KPi buffer), supplemented with 100 μM pyridoxal 5'-phosphate (PLP). Lysis was performed in 75 °C water bath (Fisherbrand™ Isotemp™ Digital-Control Water Baths: Model 220) for >1 h. Reactions were performed in 50 mM KPi. High-resolution mass spectrometry (HRMS) was conducted on an Agilent 6200 TOF using electrospray ionization (ESI) to ionize the sample. Liquid chromatography/mass spectrometry (LCMS) was performed on an Agilent 1290 UPLC-LCMS equipped with a C-18 silica column (1.8 μm, 2.1 × 50 mm) using CH_3CN/H_2O (0.1% acetic acid by volume): 5% to 95% CH_3CN over 2 min; 1 mL/min.

3.2 Cloning, expression, and purification of TrpB variants

The genes encoding *Tm*TrpA^{WT} (Uniprot P50908), *Tm*TrpB^{WT} (Uniprot G4FDT2), *Pf*5G8, *Pf*5G8 E104G, *Tm*9D8*, and *Tm*9D8* E105G were previously cloned into pET22b(+) with a C-terminal 6x His tag. Protein expression of the variants was carried out in *Escherichia coli* BL21(DE3) *E. cloni* Express® cells (Lucigen) by inoculating 5 mL TB_{carb} of with a single colony and incubating this pre-culture overnight at 37 °C and 230 rpm. For expression, 2.5 mL culture were used to inoculate 500 mL Tb_{carb} in a 2-L flask and incubated at 37 °C and 130 rpm for 2.5 hours to reach OD_{600} 0.6–0.8. Cultures were chilled on ice for 20–30 minutes and protein expression was induced with a final concentration of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Expression proceeded at 20 °C and 130 rpm for approximately 24 h. Cells

were harvested by centrifugation at 10,000*g* for 20 minutes at 4 °C and the supernatant was decanted. The pellet was stored at −20 °C until further use.

For protein purification, cells were thawed and were resuspended in 4 mL lysis buffer/g pellet. Cells were heat-treated at 75 °C for >1 h. The supernatant was collected from clarified lysate following centrifugation for 20 min at 14,000*g* and 4 °C. Purification was performed with a 1-mL Ni-NTA gravity flow column at room temperature. Buffer A: 20 mM imidazole, 25 mM KPi buffer, Buffer B: 500 mM imidazole, 25 mM KPi buffer. The column was equilibrated with 10 column volumes (CV) Buffer A. Subsequently, heat-treated lysate was loaded onto column and washed with 10 CV Buffer A. Next, 10 CV 1:1 Buffer A: Buffer B were added to elute non-target proteins from the column. Protein was eluted with 3 CV Buffer B. Proteins were dialyzed into 50 mM KPi buffer, flash frozen in liquid nitrogen, and stored at −80 °C.

To obtain the purified *Tm*TrpS complex, *Tm*TrpA^{WT} was co-purified with *Tm*TrpB^{WT}. Heat-treated lysate of both *Tm*TrpA^{WT} and *Tm*TrpB^{WT} were mixed together in a roughly 5:1 ratio (concentrations determined relative band intensities from SDS-PAGE analysis) and purification was followed as described above.

3.3 Construction of random mutagenesis libraries

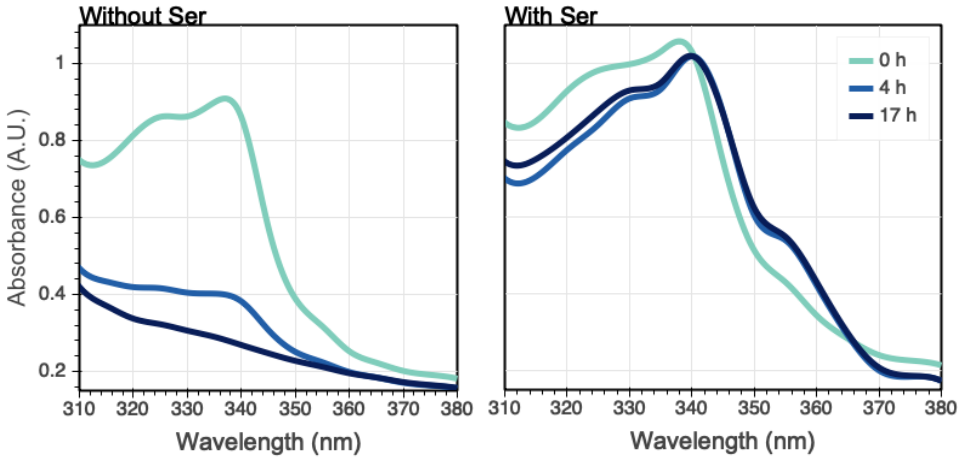
Random mutagenesis libraries were generated with the *Tm*9D8* gene as template by the addition of 200–400 μM MnCl*2* to a *Taq* (New England Biolabs) PCR reaction as previously reported.[1] PCR fragments were treated with DpnI for 1 h at 37 °C, purified by gel extraction, and then inserted into a pET22b(+) vector via Gibson assembly.[2] The Gibson assembly product was purified and concentrated using Zymo DNA Clean and Concentrate–5 kit (Catalog #: D4004). BL21(DE3) *E. cloni* Express® cells were transformed with the Gibson assembly product. Libraries generated with 200, 300, and 400 μM $MnCl_2$ were tested (one 96-well plate, each) to determine which library gave the optimal balance of high diversity and low rate of inactivation. The chosen library was then tested further (see below).

3.4 Library expression and screening

Individual colonies were grown in 300 μL TB_{carb} in deep-well 96-well polypropylene plates and grown overnight at 37 °C, 250 rpm, 80% humidity. The following day, 20 μL overnight culture were used to inoculate 630 μL TB_{carb} cultures in deep-well 96-well plates and grown at 37 °C, 250 rpm. After 2.5 h, cultures were chilled on ice for 20–30 minutes and protein expression was induced upon addition of 50 μL

IPTG (final conc. 1 mM) diluted in TB_{carb}. Cultures were shaken at 20 °C, 250 rpm for 20–24 h, after which they were subjected to centrifugation at 4,000*g* for 10 min. The cell pellets could be frozen at −20 °C until further use or used immediately.

Pellets were lysed in 300 µL lysis buffer and heat-treated lysate clarified by centrifugation at 4,000*g* for 10 min. To UV-transparent 96-well assay plates (Caplugs, catalog #:290-8120-0AF) charged with 10 µL azulene dissolved in DMSO (final conc. 0.625 mM), 30 µL heat-treated lysate was transferred using Microlab NIMBUS96 liquid handler (Hamilton), followed by addition of 70 µL serine (final conc. 10 mM), and 90 µL 50 mM KPi buffer with a 12-channel pipet. Reactions were sealed with Microseal 'B' PCR plate sealing film (BioRad, catalog #: MSB1001) and incubated in 37 °C water bath. Reaction progress was monitored by measuring absorption at 340 nm over the course of 24 h, in which more active variants retained AzAla in solution while the wells of inactive variants lost azulene due to absorption into the plastic assay plate and/or due to sublimation from solution. This can be seen below, in a case where *Tm*9D8* is not given Ser, resulting in no activity, as opposed to the case where the addition of Ser allows for AzAla formation and retention of signal at 340 nm:



3.5 Recombination of mutations

Site-directed mutagenesis was performed to combine mutations identified through screening to be beneficial for AzAla production. Gibson primers were designed to recombine the F184S and W286R mutations. Using *Tm*9D8* W286R as template, the F184S mutation was introduced. Two PCRs were performed using Phusion polymerase (New England Biolabs), each fragment originating from the F184S site and ending in the center of the ampicillin resistance gene. The PCR fragments were treated with DpnI

for 1 h at 37 °C, purified by gel extraction, and then combined via Gibson assembly.[2] The Gibson assembly product was purified using Zymo DNA Clean and Concentrate–5 kit (Catalog #: D4004). BL21(DE3) *E. cloni* Express® cells were transformed with the Gibson assembly product.

Fragment	Size (Kb)	Forward primer (5' to 3')	Reverse primer (5' to 3')
1	4.8	cctgcagaccacctattacgtgtccggctctgtggttggtccgcatccatatccg	ctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcg
2	1.7	ccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgc	gcgtgactggattaccaacctgcagaccacctattacgtg

3.6 Variant sequences

Variants identified through screening were DNA sequenced using Sanger Sequencing (Laragen) to determine their identities. The DNA sequences of all TrpB genes tested in this paper are included here. All variants were cloned into a pET22b(+) vector as described above. T7 and T7-terminator primers were used to sequence all variants.

Primer	Direction	Sequence
T7	Forward (5' to 3')	taatacgactcactataggg
T7-term	Reverse (3' to 5')	ctagttattgctcagcggtg

***Tm*TrpA^{WT}**

ATGAAAGGTTTTATCGCGTACATCCCGGCTGGTTTTCCGGATCTGGAAACCACCCGTAAAATTCTGAT
CGCACTGAACGAGCTGGGTATTACCGGTGTTGAAATTGGTGTCCCGTTCTCCGACCCGGTTGCGGA
TGGTCCGGTGATCCAACTGGCGCATAGCGTTGCTCTGCGTAACGGTGTGACTATTAAAAAAATTCTG
GAAATGCTGTCCGAGATTTCCGTAGATTACGACCTGTACCTGATGTCTTACCTGAACCCGATCGTTAA
TTACCCTGAAGGCAAAGAGAAACTGCTGGACGAACTGAAGAAGCTGGGCGTTAAAGGCCTGATTATC
CCAGACCTGCCGCTGCGTGAAGTAAAAAACGTTGACATCGCTTACCCGATCGTTCCATTCGTTGCAC
CGAATACCAAAGACGAAGAGATCGACCTGATCAACTCCGTGCAGGCTCCGTTCGTGTACTATATCTC
TCGTTACGGTGTAACTGGTGAACGCGAAGACCTGCCGTTTGCAGATCACATCAAACGCGTGAAAGAA
CGTATCAAACTGCCACTGTTCGTCGGTTTCGGTATCTCCCGTCACGAACAAGTTAAAAAAGTTTGGG
AAATCGCTGATGGTGTTATTGTTGGCAGCGCACTGGTCCGCATCATGGAAGAAAACCCGAAAGATGA
GATCCCACGTAAAGTTGTTGAAAAAGTTAAAGAGCTGCTGGGCAAAtga

***Tm*TrpB^{WT}**

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGGAAATCCTGATGCCAGCTCTGGAA
GAACTGGAAGCTGCGTACGAAGAAATCATGAAAGATGAGTCTTTCTGGAAAGAATTCAATGACCTGC
TGCGCGATTATGCGGGTCGTCCGACTCCGCTGTACTTCGCACGTCGTCTGTCCGAAAAATACGGTG
CTCGCATCTATCTGAAACGTGAAGACCTGCTGCATACTGGTGCGCATAAAATCAATAACGCTATCGG
CCAGGTTCTGCTGGCAAAAAAAATGGGCAAAACCCGTATCATTGCTGAAACGGGTGCTGGTCAGCA
CGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGA
AGACACGATCCGCCAGAAACCGAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTA
AAATCCGGTAGCCGTACCCTGAAAGACGCAATTAACGAAGCTCTGCGTGACTGGATTACCAACCTGC
AGACCACCTATTACGTGATCGGCTCTGTGGTTGGTCCGCATCCATATCCGATTATCGTACGTAACTTC
CAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCTGGAAAAAGAAGGCCGTCTGCCGGACTAC
ATCGTTGCGTGCGTGGGTGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTATCGATTCTGGTG
TGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAAACCGGTAAACATGCGGCTTCTCTGC
TGAAAGGTAAAATCGGCTACCTGCACGGTTCTAAGACGTTCGTTCTGCAGGATGACTGGGGTCAAGT

TCAGGTGACGCACTCCGTCTCCGCTGGCCTGGACTACTCCGGTGTCGGTCCGGAACACGCCTATTG
GCGTGAGACCGGTAAAGTGCTGTACGATGCTGTGACCGATGAAGAAGCTCTGGACGCATTCATCGA
ACTGTCTCGCCTGGAAGGCATCATCCCAGCCCTGGAGTCTTCTCACGCACTGGCTTATCTGAAGAAG
ATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCGTGGTGACAAGGATCTGGAATCTG
TACTGAACCACCCGTATGTTCGCGAACGCATCCGCCtcgagcaccaccatcaccatcactga

***Pf*5G8:**

ATGTGGTTCGGTGAATTTGGTGGTCAGTACGTGCCAGAAACGCTGGTTGGACCCCTGAAAGAGCTG
GAAAAAGCTTACAAACGTTTCAAAGATGACGAAGAATTCAATCGTCAGCTGAATTACTACCTGAAAAC
CTGGGCAGGTCGTCCAACCCCACTGTACTACGCAAAACGCCTGACTGAAAAAATCGGTGGTGCTAA
AGTCTACCTGAAACGTGAAGACCTGGTTCACGGTGGTGCACACAAGACCAACAACGCCATCGGTCA
GGCACTGCTGGCAAAGCTCATGGGTAAAACTCGTCTGATCGCTGAGACCGGTGCTGGTCAGCACGG
CGTAGCGACTGCAATGGCTGGTGCACTGCTGGGCATGAAAGTGGACATTTACATGGGTGCTGAGGA
CGTAGAACGTCAGAAATTGAACGTATTCCGTATGAAGCTGCTGGGTGCAAACGTAATTCCAGTTAAC
TCCGGTTCTCGCACCCTGAAAGACGCAATCGACGAGGCTCTGCGTGATTGGGTGGCTACTTTTGAAT
ACACCCACTACCTAATCGGTTCCGTGGTCGGTCCACATCCGTATCCGACCATCGTTCGTGATTTTCA
GTCTGTTATCGGTCGTGAGGCTAAAGCGCAGATCCCGGAGGCTGAAGGTCAGCTGCCAGATGTAAT
CGTTGCTTGTGTTGGTGGTGGCTCTAACGCGATGGGTATCTTTTACCCGTTCGTGAACGACAAAAAA
GTTAAGCTGGTTGGCGTTGAGGCTGGTGGTAAAGGCCTGGAATCTGGTAAGCATTCCGCTAGCCTG
AACGCAGGTCAGGTTGGTGTGTCCCATGGCATGCTGTCCTACTTTCTGCAGGACGAAGAAGGTCAG
ATCAAACCAAGCCACTCCATCGCACCAGGTCTGGATTATCCAGGTGTTGGTCCAGAACACGCTTACC
TGAAAAAAATTCAGCGTGCTGAATACGTGGCTGTAACCGATGAAGAAGCACTGAAAGCGTTCCATGA
ACTGAGCCGTACCGAAGGTATCATCCCAGCTCTGGAATCTGCGCATGCTGTGGCTTACGCTATGAAA
CTGGCTAAGGAAATGTCTCGTGATGAGATCATCATCGTAAACCTGTCTGGTCGTGGTGACAAAGACC
TGGATATTGTCCTGAAAGCGTCTGGCAACGTGCtcgagcaccaccaccaccaccactga

***Pf*5G8 E104G**

ATGTGGTTCGGTGAATTTGGTGGTCAGTACGTGCCAGAAACGCTGGTTGGACCCCTGAAAGAGCTG
GAAAAAGCTTACAAACGTTTCAAAGATGACGAAGAATTCAATCGTCAGCTGAATTACTACCTGAAAAC
CTGGGCAGGTCGTCCAACCCCACTGTACTACGCAAAACGCCTGACTGAAAAAATCGGTGGTGCTAA
AGTCTACCTGAAACGTGAAGACCTGGTTCACGGTGGTGCACACAAGACCAACAACGCCATCGGTCA
GGCACTGCTGGCAAAGCTCATGGGTAAAACTCGTCTGATCGCTGGGACCGGTGCTGGTCAGCACGG
CGTAGCGACTGCAATGGCTGGTGCACTGCTGGGCATGAAAGTGGACATTTACATGGGTGCTGAGGA
CGTAGAACGTCAGAAATTGAACGTATTCCGTATGAAGCTGCTGGGTGCAAACGTAATTCCAGTTAAC
TCCGGTTCTCGCACCCTGAAAGACGCAATCGACGAGGCTCTGCGTGATTGGGTGGCTACTTTTGAAT
ACACCCACTACCTAATCGGTTCCGTGGTCGGTCCACATCCGTATCCGACCATCGTTCGTGATTTTCA
GTCTGTTATCGGTCGTGAGGCTAAAGCGCAGATCCCGGAGGCTGAAGGTCAGCTGCCAGATGTAAT
CGTTGCTTGTGTTGGTGGTGGCTCTAACGCGATGGGTATCTTTTACCCGTTCGTGAACGACAAAAAA
GTTAAGCTGGTTGGCGTTGAGGCTGGTGGTAAAGGCCTGGAATCTGGTAAGCATTCCGCTAGCCTG
AACGCAGGTCAGGTTGGTGTGTCCCATGGCATGCTGTCCTACTTTCTGCAGGACGAAGAAGGTCAG
ATCAAACCAAGCCACTCCATCGCACCAGGTCTGGATTATCCAGGTGTTGGTCCAGAACACGCTTACC
TGAAAAAAATTCAGCGTGCTGAATACGTGGCTGTAACCGATGAAGAAGCACTGAAAGCGTTCCATGA
ACTGAGCCGTACCGAAGGTATCATCCCAGCTCTGGAATCTGCGCATGCTGTGGCTTACGCTATGAAA
CTGGCTAAGGAAATGTCTCGTGATGAGATCATCATCGTAAACCTGTCTGGTCGTGGTGACAAAGACC
TGGATATTGTCCTGAAAGCGTCTGGCAACGTGCtcgagcaccaccaccaccaccactga

Tm*9D8

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGGAAATCCTGATGGGAGCTCTGGAA
GAACTGGAAGCTGCGTACGAAGGAATCATGAAAGATGAGTCTTTCTGGAAAGAATTCAATGACCTGC
TGCGCGATTATGCGGGTCGTCCGACTCCGCTGTACTTCGCACGTCGTCTGTCCGAAAAATACGGTG
CTCGCGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTGCGCATAAAATCAATAACGCTATCGG
CCAGGTTCTGCTGGCAAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGTGCTGGTCAGCA

CGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGA
AGACACGATCCGCCAGAAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTA
AAATCCGGTAGCCGTACCCTGAAAGACGCAATTGACGAAGCTCTGCGTGACTGGATTACCAACCTGC
AGACCACCTATTACGTGTTCGGCTCTGTGGTTGGTCCGCATCCATATCCGATTATCGTACGTAACTTC
CAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGAAAAAGAAGGCCGTCTGCCGGACTAC
ATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTATCGATTCTGGTG
TGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAAACCGGTAAACATGCGGCTTCTCTGC
TGAAAGGTAAAATCGGCTACCTGCACGGTTCTAAGACGTTCGTTCTGCAGGATGACTGGGGTCAAGT
TCAGGTGAGCCACTCCGTCTCCGCTGGCCTGGACTACTCCGGTGTCGGTCCGGAACACGCCTATTG
GCGTGAGACCGGTAAAGTGCTGTACGATGCTGTGACCGATGAAGAAGCTCTGGACGCATTCATCGA
ACTGTCTCGCCTGGAAGGCATCATCCCAGCCCTGGAGTCTTCTCACGCACTGGCTTATCTGAAGAAG
ATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCGTGGTGACAAGGATCTGGAATCTG
TACTGAACCACCCGTATGTTCGCGAACGCATCCGCCtcgagcaccaccaccaccaccactga

***Tm*9D8* E105G**

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGGAAATCCTGATGGGAGCTCTGGAA
GAACTGGAAGCTGCGTACGAAGGAATCATGAAAGATGAGTCTTTCTGGAAAGAATTCAATGACCTGC
TGCGCGATTATGCGGGTCGTCCGACTCCGCTGTACTTCGCACGTCGTCTGTCCGAAAAATACGGTG
CTCGCGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTGCGCATAAAATCAATAACGCTATCGG
CCAGGTTCTGCTGGCAAAACTAATGGGCAAAACCCGTATCATTGCTGGTACGGGTGCTGGTCAGCA
CGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGA
AGACACGATCCGCCAGAAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTA
AAATCCGGTAGCCGTACCCTGAAAGACGCAATTGACGAAGCTCTGCGTGACTGGATTACCAACCTGC
AGACCACCTATTACGTGTTCGGCTCTGTGGTTGGTCCGCATCCATATCCGATTATCGTACGTAACTTC
CAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGAAAAAGAAGGCCGTCTGCCGGACTAC
ATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTATCGATTCTGGTG
TGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAAACCGGTAAACATGCGGCTTCTCTGC
TGAAAGGTAAAATCGGCTACCTGCACGGTTCTAAGACGTTCGTTCTGCAGGATGACTGGGGTCAAGT
TCAGGTGAGCCACTCCGTCTCCGCTGGCCTGGACTACTCCGGTGTCGGTCCGGAACACGCCTATTG
GCGTGAGACCGGTAAAGTGCTGTACGATGCTGTGACCGATGAAGAAGCTCTGGACGCATTCATCGA
ACTGTCTCGCCTGGAAGGCATCATCCCAGCCCTGGAGTCTTCTCACGCACTGGCTTATCTGAAGAAG
ATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCGTGGTGACAAGGATCTGGAATCTG
TACTGAACCACCCGTATGTTCGCGAACGCATCCGCCtcgagcaccaccaccaccaccactga

***Tm*Azul:**

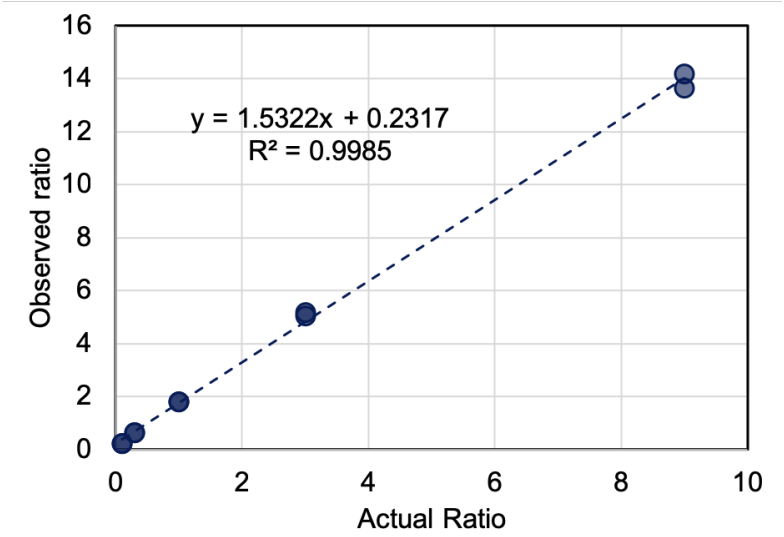
ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGGAAATCCTGATGGGAGCTCTGGAA
GAACTGGAAGCTGCGTACGAAGGAATCATGAAAGATGAGTCTTTCTGGAAAGAATTCAATGACCTGC
TGCGCGATTATGCGGGTCGTCCGACTCCGCTGTACTTCGCACGTCGTCTGTCCGAAAAATACGGTG
CTCGCGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTGCGCATAAAATCAATAACGCTATCGG
CCAGGTTCTGCTGGCAAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGTGCTGGTCAGCA
CGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGA
AGACACGATCCGCCAGAAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTA
AAATCCGGTAGCCGTACCCTGAAAGACGCAATTGACGAAGCTCTGCGTGACTGGATTACCAACCTGC
AGACCACCTATTACGTGTCCGGCTCTGTGGTTGGTCCGCATCCATATCCGATTATCGTACGTAACTT
CCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGAAAAAGAAGGCCGTCTGCCGGACTA
CATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTATCGATTCTGGT
GTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAAACCGGTAAACATGCGGCTTCTCTG
CTGAAAGGTAAAATCGGCTACCTGCACGGTTCTAAGACGTTCGTTCTGCAGGATGACAGGGGTCAA
GTTCAGGTGAGCCACTCCGTCTCCGCTGGCCTGGACTACTCCGGTGTCGGTCCGGAACACGCCTAT
TGGCGTGAGACCGGTAAAGTGCTGTACGATGCTGTGACCGATGAAGAAGCTCTGGACGCATTCATC
GAACTGTCTCGCCTGGAAGGCATCATCCCAGCCCTGGAGTCTTCTCACGCACTGGCTTATCTGAAGA
AGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCGTGGTGACAAGGATCTGGAATC
TGTACTGAACCACCCGTATGTTCGCGAACGCATCCGCCtcgagcaccaccaccaccaccactga

3.7 Small-scale analytical reactions

All analytical reactions were performed in 2-mL glass HPLC vials (Agilent) charged with 10 µL azulene or indole (final conc. 10 mM) dissolved in DMSO (5% v/v), followed by addition of serine (final conc. 10 mM) and purified enzyme diluted in 50 mM KPi buffer to a final volume of 200 µL. Reactions were incubated at 75 °C or 37 °C for 24 h. The reaction was then diluted with 800 µL of 1:1 CH_3CN/1 M aq. HCl and subjected to centrifugation at 20,000*g*. For indole, this reaction mixture was analyzed directly by UHPLC-MS at 277 nm, representing the isosbestic point between indole and tryptophan and allowing quantification of yield by comparing the substrate and product peak areas.[3] For azulene, the reaction mixture was further diluted by adding 10 µL of this mixture into 190 µL of 1:1 CH_3CN/1 M aq. HCl and analyzed by UHPLC-MS. The yield was estimated comparing the integration of the product peak at 254 nm to a calibration curve (Section 3.8).

3.8 Calibration for measuring HPLC yield of AzAla

Solutions of azulene and an authentic AzAla standard were made in 1:1 CH_3CN/1 M aq. HCl (total conc. 1 mM) and mixed in different ratios (9:1, 3:1, 1:1, 1:3, 1:9) in duplicate and analyzed by HPLC. The ratios of substrate to product peaks at 254 nm were correlated to the actual ratios by a linear relationship.



3.9 Large-scale preparation of AzAla

Two 500 mL cultures of BL21(DE3) cells expressing *Tm*Azul were grown according to expression conditions described in Section 3.2. The expression cultures were centrifuged for 20 min, 14,000*g*. The

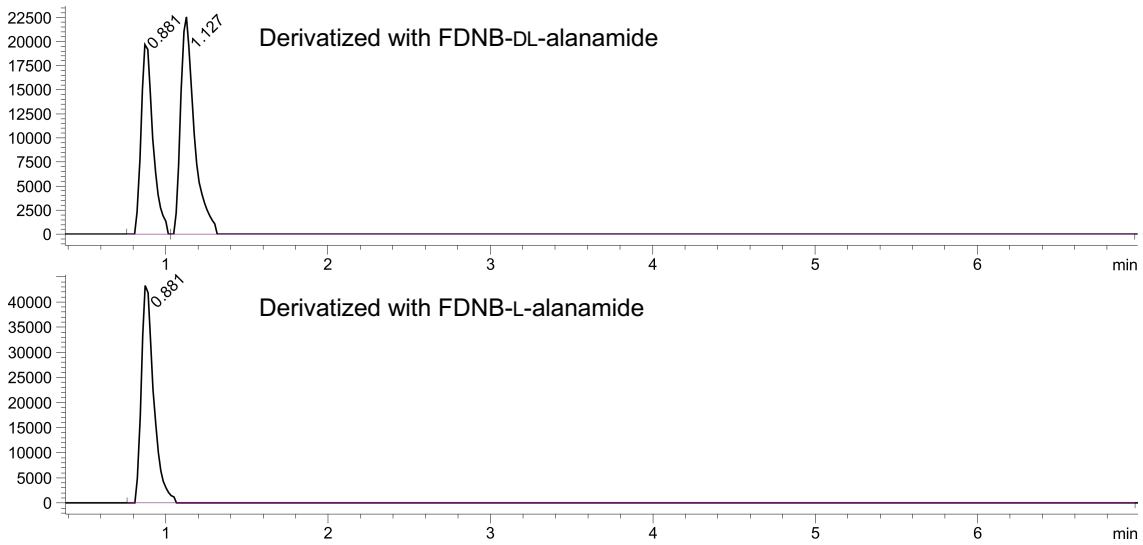
pellets (14 g) were resuspended in 50 mL lysis buffer (100 μM PLP, 50 mM KPi buffer) and lysed at 75 °C for >1 h. Heat-treated lysate was centrifuged 20 min, 14,000*g* and the lysate decanted into a fresh container.

To a 500-mL Erlenmeyer flask azulene (5 mmol, 0.6415 g) and serine (5.5 mmol, 0.5792 g), 5 mL DMSO (5% v/v), 40 mL 50 mM KPi buffer, and 55-mL heat-treated lysate were added. The reaction was covered in aluminum foil, placed in 37 °C incubator, and shaken at 130 rpm. After 72 h, the reaction was put on ice and filtered. The reaction was extracted with chloroform to remove any remaining azulene, allowing crude starting material to be recovered from chloroform by gently evaporating off the solvent under a constant stream of nitrogen. The aqueous layer was concentrated *in vacuo* and the product purified on a Biotage Isolera One purification system, using C-18 silica as the stationary phase, CH_3OH as the strong solvent, and H_2O as the weak solvent. The purified product fractions were combined and concentrated *in vacuo* to afford pure AzAla (blue solid, 0.25 g, 23% yield).

3.10 Characterization of AzAla

The enzymatic AzAla product was characterized by NMR, HRMS, and chiral derivatization. The proton NMR spectrum was taken in D_2O to reduce deuterium exchange on the five-membered ring. The ^{13}C NMR spectrum was taken in a mixture of D_2O/DCl to increase solubility. ^1H NMR (400 MHz, D_2O) δ 8.52 (d, *J* = 9.6 Hz, 2H), 7.94 (d, *J* = 3.9 Hz, 1H), 7.80 (t, *J* = 9.9 Hz, 1H), 7.52 (d, *J* = 3.9 Hz, 1H), 7.36 (dt, *J* = 13.2, 9.8 Hz, 2H), 4.15 (dd, *J* = 7.8, 5.2 Hz, 1H), 3.82 (dd, *J* = 15.3, 5.2 Hz, 1H), 3.67 (dd, *J* = 15.2, 7.8 Hz, 1H). ^{13}C NMR (100 MHz, D_2O) δ 172.35, 140.79, 138.75, 137.59, 137.54, 137.48, 136.66, 133.71, 123.74, 123.09, 120.83, 54.63, 27.64. HRMS (*m/z*) for $[M+H]^+$ $C_{13}H_{14}NO_2$ requires 216.1019, observed 216.1019.

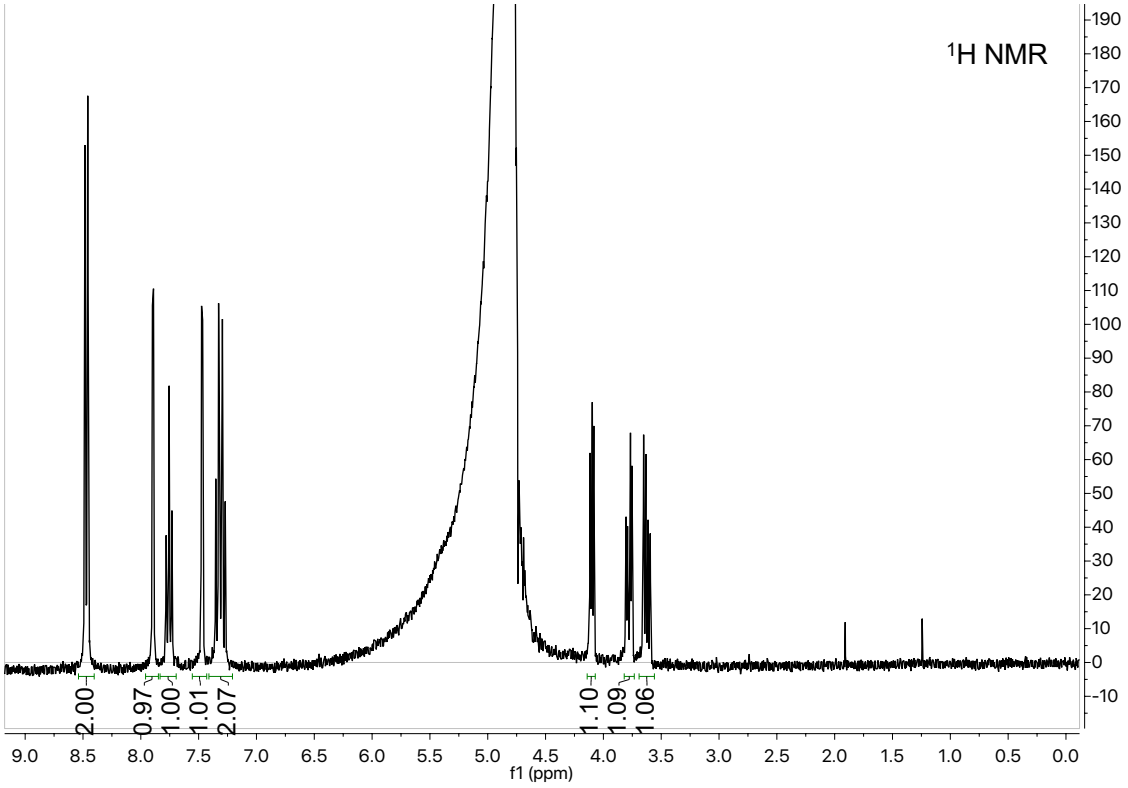
Enantiopurity was determined by derivatization with enantiopure and racemic FDNP-alanamide. In a 2 mL vial, AzAla (0.5 μmol) was dissolved in 1 M aq. $NaHCO_3$ (100 μL), to which 10 μL of a 33-mM FDNP-alanamide solution in acetone (0.33 μmol) was added. The vial was shaken for 2 h at 230 rpm, 37 °C. The reaction was allowed to cool to room temperature, then diluted with 1:1 CH_3CN/1 M aq. HCl (600 μL). The solution was analyzed via LC-MS (40% to 95% CH_3CN, monitored by using total ion count filtered for the expected mass of 468). Absolute stereochemistry for AzAla was inferred by analogy to L-tryptophan and determined to have >99% enantiomeric excess.

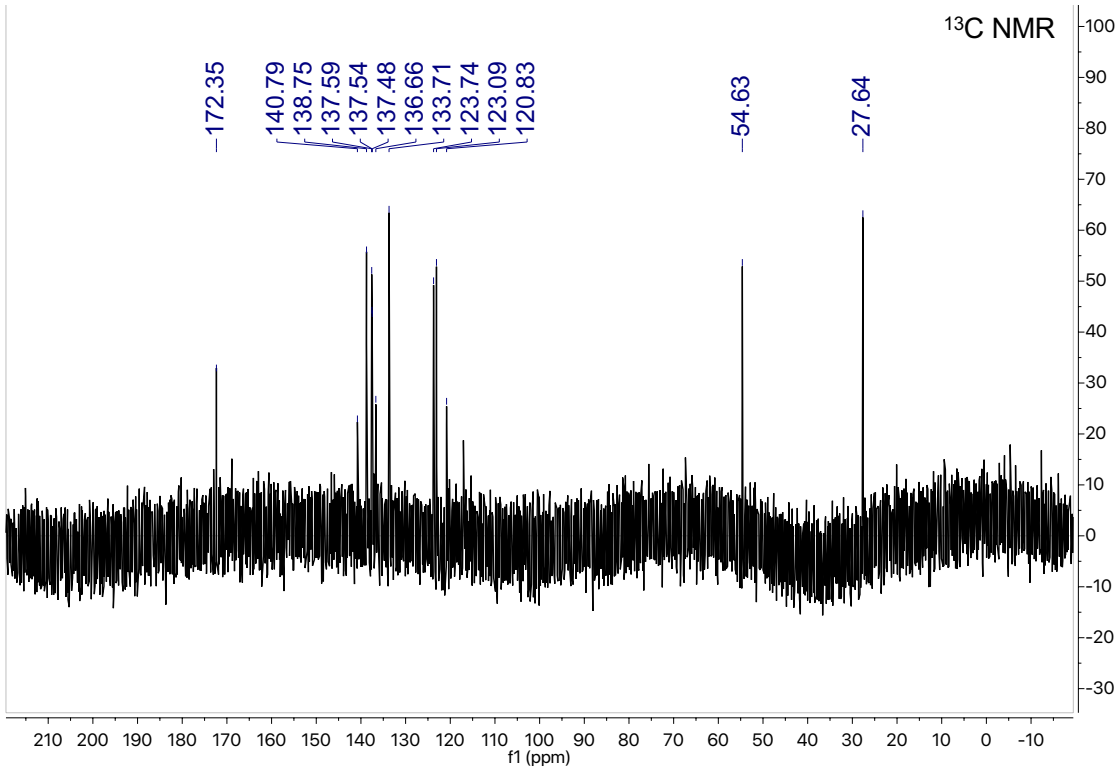


3.11 Structural modeling

A crystal structure of a homologous TrpS complex (*Salmonella typhimurium*, *St*TrpS) has been reported with the amino-acrylate in the β-subunit stabilized in the active site by benzimidazole, a competitive inhibitor (PDB: 4HPX).[4] This TrpB structure served as the template for construction of a homology model of *Tm*9D8* (57% sequence identity) using the SWISS-MODEL program.[5] The alignment of *Tm*9D8* homology model and the *St*TrpB structure allowed benzimidazole and the amino-acrylate to be placed within the *Tm*9D8* structure in PyMOL, and the five-membered ring of azulene was aligned directly to the five-membered ring of benzimidazole to simulate a productive binding pose.

4. NMR spectra





5. Citations

[1] C. E. Boville, R. A. Scheele, P. Koch, S. Brinkmann-Chen, A. R. Buller, F. H. Arnold, *Angew. Chemie - Int. Ed.* **2018**, *57*, 14764–14768.

[2] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343–345.

[3] D. K. Romney, J. Murciano-Calles, J. Wehrmüller, F. H. Arnold, *J. Am. Chem. Soc* **2017**, *139*, 10769–10776.

[4] D. Niks, E. Hilario, A. Dierkers, H. Ngo, D. Borchardt, T. J. Neubauer, L. Fan, L. J. Mueller, M. F. Dunn, *Biochemistry* **2013**, *52*, 6396–6411.

[5] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, et al., *Nucleic Acids Res.* **2018**, *46*, W296–W303.