Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Cofactors

Nathaniel Goldberg, Anders M. Knight, Ruijie Zhang, Frances H. Arnold

Transition-metal catalysis is a powerful tool for the construction of chemical bonds. Here we show that a non-heme iron enzyme can catalyze olefin aziridination and nitrene C–H insertion, and that these activities can be improved by directed evolution. The non-heme iron center allows for facile modification of the primary coordination sphere by addition of metal-coordinating molecules, enabling control over enzyme activity and selectivity using small molecules.

File list (2)

EFE_Manuscript_20191028.pdf (584.68 KiB)	view on ChemRxiv • download file
EFE_SI_20191028.pdf (1.25 MiB)	view on ChemRxiv • download file

Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Cofactors

Nathaniel W. Goldberg[†,§], Anders M. Knight[‡,§], Ruijie K. Zhang[†,∥], Frances H. Arnold[†,‡,*]

[†]Division of Chemistry and Chemical Engineering and [‡]Division of Biology and Bioengineering, California Institute of Technology, 1200 East California Boulevard, MC 210-41, Pasadena, California 91125, United States

Supporting Information Placeholder

ABSTRACT: Transition-metal catalysis is a powerful tool for the construction of chemical bonds. Here we show that a non-heme iron enzyme can catalyze olefin aziridination and nitrene C–H insertion, and that these activities can be improved by directed evolution. The non-heme iron center allows for facile modification of the primary coordination sphere by addition of metal-coordinating molecules, enabling control over enzyme activity and selectivity using small molecules.

Over the last century, chemists have developed myriad synthetic transition-metal catalysts to access new chemical transformations and modes of reactivity. Nature has been developing catalysts for far longer: over billions of years, she has evolved a rich repertoire of proteins that perform most of the chemical reactions of life. But nature's inventions do not include many of the best inventions of human chemists. Our efforts to merge abiological transition-metal chemistry with nature's vast toolbox of metalloproteins have focused on heme-binding proteins[1], as the heme cofactor and its analogues are well-studied in synthetic transition-metal chemistry. However, heme-binding proteins represent only a small fraction of the chemical diversity present in natural metalloproteins. Metalloproteins comprise greater than 30% of all proteins[2] and are responsible for some of the most fundamental chemical reactions in biology, including nitrogen fixation, photosynthesis, and DNA synthesis. Natural metalloproteins bind a variety of metals in a wide range of metal-binding sites, either coordinating the metal ion itself or a more complex metal-containing cofactor. Nearly any heteroatom-containing side chain can coordinate to a metal, in addition to the peptide backbone, allowing myriad possible coordination environments[3]. Many coordination environments in non-heme metalloenzymes have multiple open coordination sites at the metal center, a key feature of numerous synthetic transition-metal catalysts. Expanding new-to-nature catalysis to non-heme metalloenzymes would open a new world of transition-metal biocatalysis.

For decades, humans have used small-molecule ligands to understand and control Nature's catalysts. Nature employs allostery to regulate enzyme activity with small molecules (Figure 1A). Allosteric regulation of enzymes has been studied at length[4] and there have been successful examples of engineered[5] or designed[6] allosteric enzyme regulation, but allostery is still challenging to understand or adapt to engineered systems. Cofactor-dependent enzymes, in which a small molecule is directly involved in catalysis, are often more readily understood and more easily manipulated by small molecules. Biochemists have knocked out catalytic activity *via* mutagenesis and restored it with exogenous effectors[7] and sub-

stituted native cofactors with synthetic analogues with altered properties[8]. In the case of heme-binding proteins, scientists have substituted the metal ion[9] or modified the porphyrin prosthetic group[10] to modulate or expand enzymatic activity. A non-heme metalloprotein, in which a metal ion is coordinated directly by the protein and by exogenous small molecules, would allow for direct manipulation of the primary metal coordination environment.
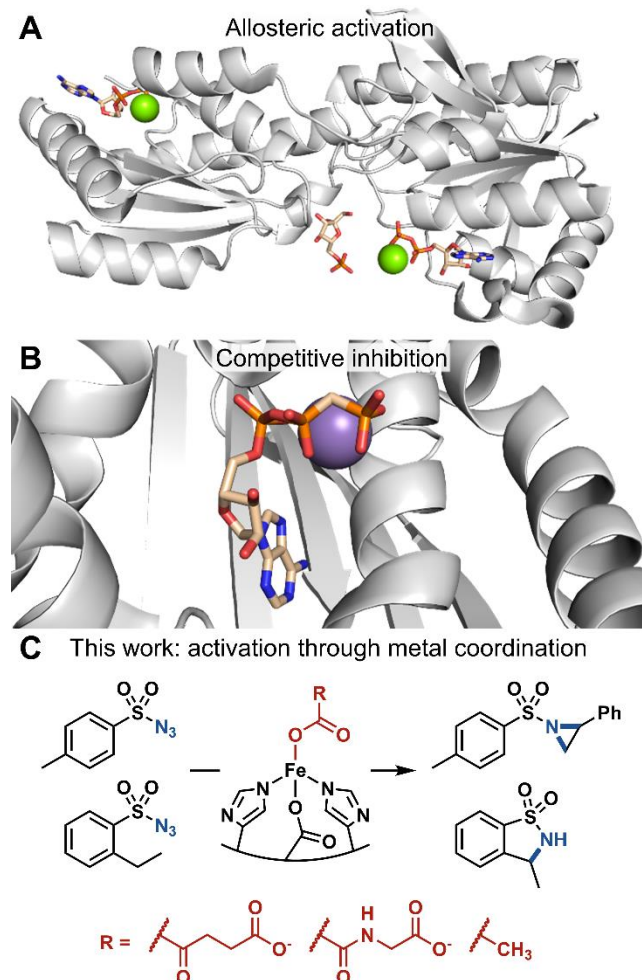

Figure 1. Examples of enzyme activity control using small molecules. (A) ADP acts as a natural allosteric regulator of phosphofructokinase (PDB ID: 4PFK). (B) Non-natural analogues of ATP act as competitive inhibitors of ATP-dependent enzymes (PDB ID: 1I5A). (C) This work: small-molecule activation of a non-heme iron center for nitrene transfer.

To search for abiological catalytic promiscuity among natural metalloproteins, we looked at α-ketoglutarate (αKG)-dependent iron enzymes, a family of enzymes which features a conserved metal-binding active site with iron coordinated by two histidines and one aspartate or glutamate[11]. In nature, these enzymes perform similar chemistry to the heme-binding cytochrome P450 family, in which a high-valent iron-oxo intermediate performs C–H hydroxylation, olefin epoxidation, or other oxidative transformations[12]. Though members of this enzyme family have been reported to catalyze reactions beyond their native functions, all the reactions reported proceed through the native iron-oxo mechanism[13]. We hypothesized that non-heme iron enzymes might also be able to catalyze abiological transformations similar to heme-binding proteins through a non-natural mechanistic pathway.
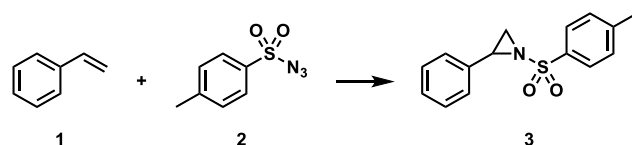
We screened a set of seven purified α-ketoglutarate (αKG)-dependent iron dioxygenases against the intermolecular aziridination reaction of styrene **1** and *p*-toluenesulfonyl azide **2**. Aziridination[14] and carbon–hydrogen (C–H) bond insertions of nitrenes[15] have been reported using engineered heme-binding proteins and were subsequently proposed in a natural product biosynthetic pathway[16]. Chang *et al*. speculated the existence of a transient iron-nitrene intermediate in their report of the transformation of alkyl azides to nitriles by an αKG-dependent iron dioxygenase, but this reaction still proceeds through the canonical iron-oxo catalytic cycle[13b]. To date, no reported non-heme iron enzyme, natural or engineered, has been reported to catalyze productive nitrene transfer.

From the set of enzymes we tested, only *Pseudomonas savastanoi* ethylene-forming enzyme (*Ps*EFE, UniProt ID P32021), formed aziridine **3** significantly above background (Supplementary Table S2). Compared to other members of the αKG-dependent iron dioxygenase family *Ps*EFE is mechanistically and structurally distinct. While most enzymes of this family catalyze the oxidation of a substrate, often C–H hydroxylation, *Ps*EFE natively catalyzes the fragmentation of the usual co-substrate α-ketoglutarate to ethylene, as well as the hydroxylation of L-arginine[17]. Structurally, *Ps*EFE possesses a hybrid fold, combining elements of both type I and type II αKG-dependent iron enzymes. It binds α-ketoglutarate in a strained conformation in an unusually hydrophobic pocket, which is likely responsible for the atypical catalytic activity[18].

As the iron-binding site in *Ps*EFE is quite unlike that of the heme-binding proteins that perform nitrene-transfer chemistry, we sought to characterize the necessary components of the reaction. Iron is required, with a single added equivalent of iron(II) sufficient to fully restore catalytic activity of the wild-type apoenzyme. *Ps*EFE has three coordination sites filled by amino-acid side chains (two histidines and one aspartate), leaving up to three additional sites open for binding. In the native catalytic mechanism of *Ps*EFE and other members of its family, α-ketoglutarate occupies two of these sites and is required for activity, as it is oxidatively decarboxylated to succinate to generate the reactive iron-oxo intermediate. *Ps*EFE has been shown to catalyze arginine hydroxylation with α-ketoadipate instead of α-ketoglutarate, but with 500-fold lower activity. Other α-ketoacids were reported to give no activity[19]. Nitrene transfer, however, does not proceed through the native catalytic cycle and therefore does not require αKG as a co-substrate; the αKG is now more a cofactor and as such could potentially be replaced by different small-molecule ligands. Intrigued by the possibility of modulating enzyme activity by changing the primary co-

ordination sphere of the catalytic iron, we tested *Ps*EFE for aziridination with a set of α-ketoglutarate mimics and related molecules as additives. We found that whereas addition of a carboxylate is beneficial for activity (though not required), the wild-type enzyme is significantly more active for aziridination with added acetate or *N*-oxalylglycine (NOG, a general αKG-dependent enzyme competitive inhibitor[11]) compared to α-ketoglutarate (Table 1).

Table 1. Aziridination catalyzed by wild-type *Ps*EFE

Deviation from standard conditions[1]	Relative activity
None	1.00
No iron	0.08
No αKG	0.52
Acetate[2] instead of αKG	6.72
N-oxalylglycine[3] instead of αKG	7.75
Acetate instead of αKG, no ascorbic acid	6.01

[1]Standard conditions: Reactions were performed in MOPS buffer (20 mM pH 7.0) with 5% ethanol co-solvent, with 20 μM apoenzyme, 1 mM Fe(NH$_4$)$_2$(SO$_4$)$_2$, 1 mM αKG (as disodium salt), 1 mM L-ascorbic acid, and 10 mM **1** and **2**. [2]Sodium salt. [3]Free acid.

We then sought to improve *Ps*EFE for aziridination *via* directed evolution, targeting active-site residues with site-saturation mutagenesis and screening for enhanced activity. During directed evolution, we screened with added acetate, as it enhanced the activity of the wild-type enzyme significantly more than the native α-ketoglutarate, it is biologically ubiquitous, and it is inexpensive. Although α-ketoglutarate is the native cofactor and is naturally present at near-millimolar intracellular concentration in *Escherichia coli*[20], we reasoned that by supplementing the reaction medium with acetate we could evolve *Ps*EFE to be dependent on acetate instead.

After two rounds of site-saturation mutagenesis and one round of recombination, we found a variant with five mutations from the wild type (T97M R171L R277H F314M C317M, *Ps*EFE MLHMM) which catalyzed the formation of **3** with 120 total turnover number (TTN) and 88% enantiomeric excess (ee) favoring the (*R*)-enantiomer (Figure 2A). Four of the five introduced mutations are in the binding pocket of the native substrate arginine and presumably are involved in substrate binding. The fifth beneficial mutation is at Arg-277, the residue whose guanidino group natively binds the distal carboxylate of α-ketoglutarate (Figure 2B). The R277H mutation presumably abolishes binding of the native cofactor α-ketoglutarate; as a result, *Ps*EFE MLHMM shows no significant increase in aziridination activity when α-ketoglutarate is added, but an 11-fold increase when acetate is added (Supplementary Table S4). Thus the evolved MLHMM variant is more activated by acetate than the wild type and is no longer activated by α-ketoglutarate at all, demonstrating the tunability of the cofactor dependence of *Ps*EFE.
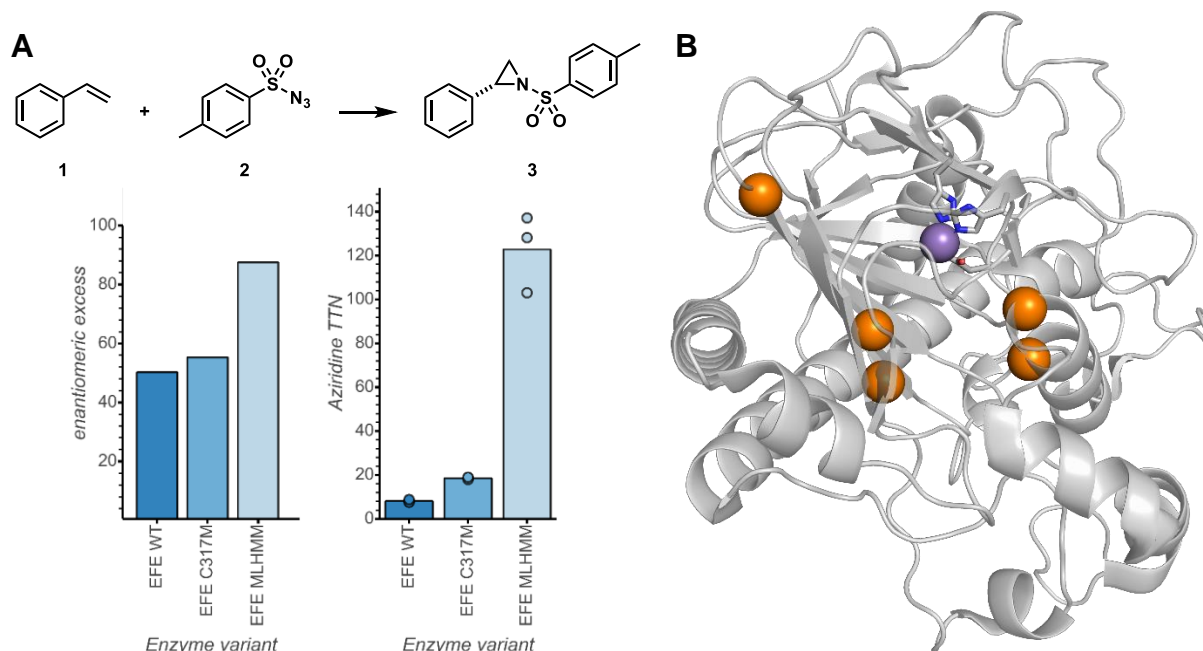
Figure 2. Directed evolution of *Ps*EFE for aziridination. (A) Evolutionary lineage. Reactions were performed in triplicate anaerobically with acetate and quantified by analytical HPLC-UV. Full experimental details are given in the Supporting Information. (B) Structural representation of *Ps*EFE with mutated sites highlighted in orange; metal-coordinating residues H189, D191, and H268 are represented in sticks and Mn (the metal with which the protein was crystallized) is represented as a purple sphere (PDB ID: 6CBA).

We reasoned that variants of *Ps*EFE generated by directed evolution for aziridination could exhibit promiscuous activity for additional nitrene-transfer reactions beyond aziridination. Screening a panel of variants for the intramolecular C–H bond insertion reaction of 2-ethylbenzenesulfonyl azide **4** to form sultam **5**, we identified multiple variants that perform this reaction with good TTN, excellent chemoselectivity, and moderate enantioselectivity. Remarkably, *Ps*EFE R171V F314M C317M (*Ps*EFE VMM) is significantly more active and more chemo- and enantioselective with *N*-oxalylglycine added than with either acetate or α-ketoglutarate, forming **5** with up to 730 TTN and greater than 100:1 selectivity for insertion over reduction (Table 2). This chemoselectivity is much higher than that of the heme proteins previously reported to catalyze similar reactions[15a,b]. To probe the specific effect of cofactor binding on activity, we tested *Ps*EFE R171V R227H F314M C317M (*Ps*EFE VHMM), a variant in which αKG and NOG binding are disrupted by the R277H mutation. Whereas *Ps*EFE VHMM's activity is still enhanced nearly ten-fold with acetate, there is no significant difference between its activity with no additive, αKG, or NOG added to the reaction (Table 2). αKG or its analogues therefore appear to bind within the native αKG binding site, activating the protein through the primary metal coordination sphere.

Aziridination with *Ps*EFE is reasonably oxygen tolerant, with the MLHMM variant maintaining 20% activity for aziridination when performed in air. C–H insertion, however, does not proceed detectably in aerobic conditions with any variant tested. These observations suggest that formation of the putative iron-nitrene intermediate is not significantly inhibited by oxygen, but that a subsequent mechanistic step in the C–H insertion reaction is inhibited aerobically. Unlike nitrene transfer with heme proteins[15a], an additional reductant is not required when reactions are performed anaerobically (Supplementary Table S4).

Table 2. Nitrene C–H insertion catalyzed by *Ps*EFE

*Ps*EFE variant	Additive	TTN (**5**)	ee (%)	**5/6**
Wild type	Acetate	12	n.d.[1]	1.6
VHMM	None	27	n.d.[1]	3.4
VHMM	αKG	31	n.d.[1]	3.8
VHMM	Acetate	240	7.3	32
VHMM	NOG	33	n.d.[1]	4.1
VMM	None	25	n.d.[1]	0.9
VMM	αKG	130	61	9.0
VMM	Acetate	310	9.4	24
VMM	NOG	450	48	105
VMM[2]	NOG	730	47	100

Reactions were performed anaerobically in MOPS buffer (20 mM pH 7.0) with 2.5% ethanol co-solvent, 20 μM apoenzyme, 1 mM Fe(NH$_4$)$_2$(SO$_4$)$_2$, 1 mM additive, 1 mM L-ascorbic acid, and 10 mM **4** (maximum 500 TTN). Reactions were quantified by analytical HPLC-UV. TTNs are shown for **5** only. [1]Not determined due to low conversion. [2]10 μM enzyme concentration (max. 1000 TTN).

*Ps*EFE is highly expressed (>200 mg/L *E. coli* culture) and is catalytically active in whole *E. coli* cells and in cell lysate. We observe that in cell lysate the enzyme maintains high activity with no external additive; the enzyme presumably retains a cofactor from the intracellular environment during lysis. Addition of *N*-oxalylglycine nevertheless enhances C–H insertion yield and chemoselectivity when *Ps*EFE VMM is in cell lysate. In whole cells, however,

there is no significant change upon addition of *N*-oxalylglycine (Supplementary Table S6). This is unsurprising as *N*-oxalylglycine is not reported to be cell-permeable. Future biochemical studies and further mutagenesis will likely enhance the selectivity for cofactor analogues and impart *in vivo* activation to *Ps*EFE variants.

In conclusion, we have discovered a non-heme iron enzyme capable of performing nitrene-transfer chemistry and enhanced that activity *via* directed evolution. This is the first example of enzymatic nitrene transfer catalyzed by a non-heme metalloprotein. *Ps*EFE features a metal center whose primary coordination sphere can be altered by simple reaction additives, allowing for modulation of catalytic activity and selectivity. We anticipate that this biocatalytic system will lead to discovery of new metalloenzymatic transformations not possible with previously reported enzymes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Materials and experimental methods, compound characterization data (PDF)

Full nucleotide and amino-acid sequences for all reported enzyme variants (XLSX)

AUTHOR INFORMATION

Corresponding Author

*frances@cheme.caltech.edu

Present Address

|Amyris Biotechnologies, Emeryville, California 94608, United States

Author Contributions

§These authors contributed equally.

Notes

A provisional patent has been filed through the California Institute of Technology based on the results presented here.

REFERENCES

(1) Brandenberg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, 102–111.

(2) Degtyarenko, K. Metalloproteins. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*; Jorde, L.; Little, P.; Dunn, M.; Subramaniam, S., Eds.; John Wiley & Sons, Ltd., **2005**.

(3) Holm, R. H.; Kennepohl, P.; Solomon, E. I. Structural and Functional Aspects of Metal Sites in Biology. *Chem. Rev.* **1996**, *96*, 2239–2314.

(4) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The ensemble nature of allostery. *Nature* **2014**, *508*, 331–339.

(5) Guntas, G.; Mansell, T. J.; Kim, J. R.; Ostermeier, M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11224–11229.

(6) Langan, R. A.; Boyken, S. E.; Ng, A. H.; Samson, J. A.; Dods, G.; Westbrook, A. M.; Nguyen, T. H.; Lajoie, M. J.; Chen, Z.; Berger, S.; Mulligan, V. K.; Dueber, J. E.; Novak, W. R. P.; El-Samad, H.; Baker, D. De novo design of bioactive protein switches. *Nature* **2019**, *572*, 205–210.

(7) Qiao, Y.; Molina, H.; Pandey, A.; Zhang, J.; Cole, P. A. Chemical Rescue of a Mutant Enzyme in Living Cells. *Science* **2006**, *311*, 1293–1297.

(8) (a) Khanna, N.; Esmieu, C.; Mészáros, L. S.; Lindblad, P.; Berggren, G. *In vivo* activation of an [FeFe] hydrogenase using synthetic cofactors. *Energy Environ. Sci.* **2017**, *7*, 1563–1567. (b) Ji, X.; Mandalapu, D.; Cheng, J.; Ding, W.; Zhang, Q. Expanding the Chemistry of the Class C Radical SAM Methyltransferase NosN by Using an Allyl Analogue of SAM. *Angew. Chem. Int. Ed.* **2018**, *57*, 6601–6604. (c) Drenth, J.; Trajkovic, M.; Fraaije, M. W. Chemoenzymatic Synthesis of an Unnatural Deazaflavin Cofactor That Can Fuel F$_{420}$-Dependent Enzymes. *ACS Catal.* **2019**, *9*, 6435–6443.

(9) (a) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **2017**, *359*, 2076–2089. (b) Wolf, M. W.; Vargas, D. A.; Lehnert, N. Engineering of RuMb: Toward a Green Catalyst for Carbene Insertion Reactions. *Inorg. Chem.* **2017**, *56*, 5623–5635. (c) Natoli, S. N.; Hartwig, J. F. Noble-Metal Substitution in Hemoproteins: An Emerging Strategy for Abiological Catalysis. *Acc. Chem. Res.* **2019**, *52*, 326–335.

(10) (a) Reynolds, E. W.; McHenry, M. W.; Cannac, F.; Gober, J. G.; Snow, C. D.; Brustad, E. M. An Evolved Orthogonal Enzyme/Cofactor Pair. *J. Am. Chem. Soc.* **2016**, *138*, 12451–12458. (b) Sreenilayam, G.; Moore, E. J.; Steck, V., Fasan, R. Stereoselective Olefin Cyclopropanation under Aerobic Conditions with an Artificial Enzyme Incorporating an Iron-Chlorin e6 Cofactor. *ACS Catal.* **2017**, *7*, 7629–7633.

(11) Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21–68.

(12) Islam M. D.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem.* **2018**, *87*, 585–620.

(13) (a) Matthews, M. L.; Chang, W.-c.; Layne, A. P.; Miles, L. A.; Krebs, C.; Bollinger, J. M. Jr. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **2014**, *10*, 209–215. (b) Davidson, M.; McNamee, M.; Fan, R.; Guo, Y.; Chang, W.-c. Repurposing Nonheme Iron Hydroxylases To Enable Catalytic Nitrile Installation through an Azido Group Assistance. *J. Am. Chem. Soc.* **2019**, *141*, 3419–3423. (c) Neugebauer, M. E.; Sumida, K. H.; Pelton, J. G.; McMurry, J. L.; Marchand, J. A.; Chang, M. C. Y. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **2019**, *15*, 1009–1016.

(14) Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **2015**, *1*, 89–93.

(15) (a) McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold. F. H. Enantioselective Intramolecular C–H Amination Catalyzed by Engineered Cytochrome P450 Enzymes *In Vitro* and *In Vivo*. *Angew. Chem. Int. Ed.* **2013**, *52*, 9309– 9312. (b) Singh, R.; Bordeaux, M.; Fasan, R. P450-Catalyzed Intramolecular sp^3 C–H Amination with Arylsulfonyl Azide Substrates. *ACS Catal.* **2014**, *4*, 546–552. (c) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, intermolecular benzylic C–H amination catalysed by an engineered iron-haem enzyme. *Nat. Chem.* **2017**, *9*, 629–634.

(16) Tsutsumi, H.; Katsuyama, Y.; Izumikawa, M.; Takagi, M.; Fujie, M.; Satoh, N.; Shin-ya, K.; Ohnishi, Y. Unprecedented Cyclization Catalyzed by a Cytochrome P450 in Benzastatin Biosynthesis. *J. Am. Chem. Soc.* **2018**, *140*, 6631–6639.

(17) Fukuda, H.; Ogawa, T.; Tazaki, M.; Nagahama, K.; Fujii, T.; Tanase, S.; Morino, Y. Molecular cloning in *Escherichia coli*, expression, and nucleotide sequence of the gene for the ethylene-forming enzyme of *Pseudomonas syringae* pv. *phaseolicola* PK2. *Biochem. Biophys. Res. Commun.* **1992**, *188*, 483–489.

(18) (a) Zhang Z.; Smart T. J.; Choi, H.; Hardy, F.; Lohans, C. T.; Abboud, M. I.; Richardson, M. S. W.; Paton, R. S.; McDonough, M. A.; Schofield, C. J. Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 4467–4672. (b) Martinez, S.; Fellner, M.; Ferr, C. Q.; Ritchie, A.; Hu, J.; Hausinger, R. P. Structures and Mechanisms of the Non-Heme Fe(II)- and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist. *J. Am. Chem. Soc.* **2017**, *139*, 11980–11988.

(19) Martinez, S.; Hausinger, R. P. Biochemical and Spectroscopic Characterization of the Non-Heme Fe(II)- and 2-Oxoglutarate-dependent Ethylene-Forming Enyzme from *Pseudomonas syringae* pv. *phaseolicola* PK2. *Biochemistry* **2016**, *55*, 5989–5999.

(20) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **2009**, *5*, 593–599.

Native cofactor

Non-native cofactors
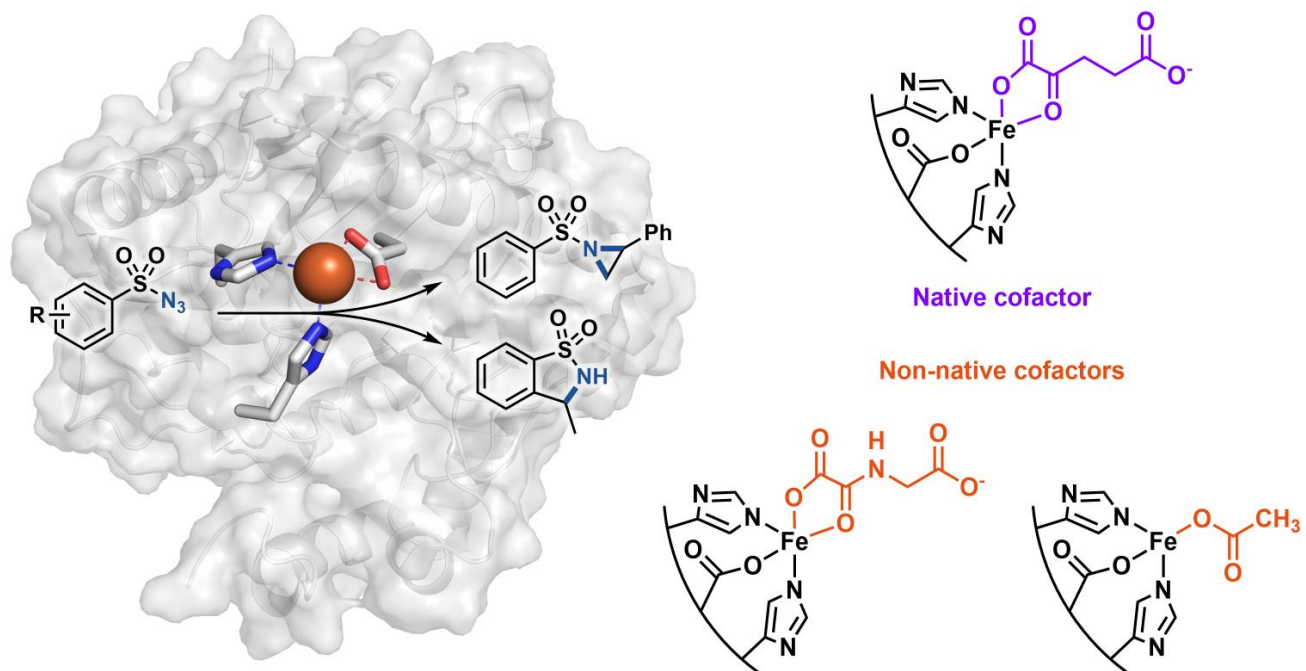
Non-natural nitrene transfer is enhanced by substitution of native iron-binding cofactor

Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Cofactors

Nathaniel W. Goldberg[†,§], Anders M. Knight[‡,§], Ruijie K. Zhang[†], Frances H. Arnold[†,‡,*]

†Division of Chemistry and Chemical Engineering and ‡Division of Biology and Bioengineering, California Institute of Technology, 1200 East California Boulevard, MC 210-41, Pasadena, California 91125, United States

*E-mail: frances@cheme.caltech.edu

TABLE OF CONTENTS

MATERIALS AND METHODS

Synthetic chemistry

All manipulations were performed using oven-dried glassware (130 °C for a minimum of 12 hours) and standard Schlenk techniques under an atmosphere of argon, unless otherwise stated.

Solvents

ACS- and HPLC-grade solvents were purchased from Fisher Chemical. Anhydrous tetrahydrofuran was obtained by filtration through a drying column and a deoyxgenation column on a Pure Process Technologies solvent system. High-purity water for PCR and HPLC was distilled after filtration through a deionizing column and organic removal column. Deuterated solvents were purchased from Cambridge Isotope Laboratories.

Chromatographic materials

Thin layer chromatography (TLC) was performed using EMD TLC plates pre-coated with 250 µm thickness silica gel 60 F_{254} and visualized by fluorescence quenching under UV light and staining with potassium permanganate or cerium ammonium molybdate. Preparative flash chromatography was performed using a Biotage Isolera automated chromatography instrument using columns hand-packed with silica gel (230–400 mesh, Silicyle Inc.).

Starting materials

All compounds were used as received from commercial suppliers, unless otherwise stated.

Analytical instrumentation

HPLC-MS analysis for initial activity determination was performed on an Agilent 1290 UPLC-MS equipped with a C18 silica column (1.8 µm packing, 2.1×50 mm). HPLC-MS analysis of site-saturation mutagenesis libraries was performed on an Agilent 1260 Infinity HPLC with an Agilent 6120 quadrupole mass spectrometer. Reverse-phase HPLC-UV analysis was performed with an Agilent 1200 series HPLC or an Agilent 1260 Series Infinity II HPLC using an Agilent Poroshell 120 EC-C18 column (4 µm packing, 2.1×50 mm) fitted with a Poroshell 120 guard column (1.7 µm packing, 2.1×5 mm). Normal-phase HPLC-UV analysis for chiral separations was performed with a Hewlett Packard Series 1100 HPLC instrument using a Daicel Chiralcel OJ-H column, (5 µm packing, 4.6×250 mm) or a Daicel Chiralpak IB column (5 µm packing, 4.6×250 mm).

NMR spectra were recorded on a Varian Unity/Inova 500 spectrometer operating at 500 MHz and 125 MHz for ^{1}H and ^{13}C respectively, or a Bruker Avance 400 spectrometer operating at 400 MHz and 100 MHz for ^{1}H and ^{13}C respectively. NMR data were analyzed in MestReNova (MestreLab Research). Chemical shifts are reported in ppm with the solvent resonance as the internal standard. For ^{1}H NMR: CDCl$_3$, δ 7.26. For ^{13}C NMR: CDCl$_3$, δ 77.16. Data are reported as follows: s = singlet, d = doublet, t = triplet, q = quartet, spt = septet, m = multiplet, br = broad; coupling constants in Hz; integration.

Biology and biocatalytic reactions

Materials

Oligonucleotides were purchased from IDT DNA. PCRs were run with Phusion® High-Fidelity PCR Kit (New England Biolabs). Gibson assembly mix[1] is prepared with isothermal master mix in-house and enzymes T5 exonuclease, Phusion® DNA polymerase, and Taq DNA ligase purchased from New England Biolabs.

Cloning

Plasmids encoding *Pseudomonas savastanoi* ethylene-forming enzyme (Uniprot ID P32021), *Streptomyces sp.* 2-aminobutyric acid chlorinase (UniProt ID D0VX22), and *Arabidopsis thaliana* anthocyanidin synthase (UniProt ID Q96323), with the coding sequences codon-optimized for *Escherichia coli* were purchased from Twist Biosciences. Plasmids encoding *Gluconobacter oxydans* leucine dioxygenase (UniProt ID Q5FQD2), *Streptomyces vinaceus* arginine hydroxylase (UniProt ID Q6WZB0), and *Streptomyces muensis* leucine hydroxylase (UniProt ID A0A0E3URV8) were obtained from the laboratory of Prof. Hans Renata (Scripps Research Institute). The plasmid encoding *Escherichia coli* taurine dioxygenase (UniProt ID P37610) was obtained from the laboratory of Prof. Harry Gray (Caltech). All genes were encoded with a C-terminal His_6-tag for purification and inserted between the NdeI and XhoI cut sites in the pET-22b(+) vector (Novagen).

Plasmids were used to transform *E. cloni* BL21(DE3) cells (Lucigen) by electroporation. SOC medium (0.75 mL) was added and the cells were incubated at 37 °C for 45 minutes before being plated on Luria-Bertani medium (Research Products International) supplemented with ampicillin (100 µg mL^{-1}, LB-amp) agar plates.

Protein expression and purification

Starter cultures of LB-amp were inoculated from a single *E. coli* colony on an agar plate harboring a plasmid encoding the protein of interest and grown overnight to stationary phase at 37 °C. Expression cultures of Terrific Broth (Research Products International) supplemented with ampicillin (100 mg L^{-1}, TB-amp) were inoculated from the starter cultures (1% v/v) and shaken at 37 °C and 160 rpm in a Multitron Infors incubator. When the expression cultures reached OD_{600} ~ 0.8 (typically 2–3 hours), they were cooled on ice for 20 minutes. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM). Cultures were incubated at 22 °C and 110 rpm overnight (16–24 hours). Cells were pelleted by centrifugation (5000×*g*, 10 minutes).

For reactions with whole cells, cell pellets were resuspended in MOPS buffer (20 mM pH 7.0) to OD_{600} 30. For reactions with cell lysate, the whole cell suspensions were lysed by sonication (QSonica Q500 sonicator, 25% amplitude, 33% duty cycle, 3 minutes). The lysate was clarified by centrifugation (20,817×*g*, 10 minutes).

For purification, cell pellets were frozen at −20 °C for at least 24 hours. Cells were resuspended in binding buffer (20 mM Tris·HCl, 100 mM sodium chloride, 20 mM imidazole, pH 7.0, ~5 mL/g wet cells) and lysed by

[1] Gibson, D.G.; Young, L.; Chuang, R.-Y.; Venter, J.C.; Hutchison, C.A. III; Smith, H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345.

sonication (QSonica Q500 sonicator, 25% amplitude, 33% duty cycle, 4 minutes). The lysate was clarified by centrifugation (20,817 *g*, 10 minutes) followed by filtration (0.45 μm syringe filter). The protein was purified using an Äkta Purifier with a HisTrap HP column (GE Healthcare), eluting with a gradient of 20–500 mM imidazole. Fractions containing the protein of interest were pooled and dialyzed at 4 °C against MOPS buffer (20 mM pH 7.0) containing 1 mM EDTA (>100:1 v/v) (Spectrum Laboratories Spectra/Por 12–14 kD membrane) for four hours, then against MOPS buffer (20 mM pH 7.0) overnight (12–16 hours). The dialyzed protein was concentrated by centrifugal filtration (Amicon Ultra-15 10 kD MWCO) to a final concentration of 40–100 mg mL^{-1}. The concentrated protein was divided into aliquots (50–100 μL), flash-frozen on powdered dry ice, and stored at −80 °C. Protein concentration was determined by Bradford assay (Bio-Rad Quick Start Bradford).

Site-saturation mutagenesis and library screening

Site-saturation mutagenesis was performed using the 22-codon method[2]. Oligonucleotides including the three 22-codon trick codons (NDT, VHG, TGG) and oligos within the ampicillin resistance cassette were used to amplify the plasmid in two pieces; oligo sequences are listed in Table S1. The two pieces were assembled *via* isothermal Gibson assembly (50 °C, 1 hour). The Gibson assembly product was used directly to transform *E. cloni* BL21(DE3) cells (Lucigen) by electroporation. SOC medium (0.75 mL) was added and the cells were incubated at 37 °C for 45 minutes before being plated on LB-amp agar plates. Single colonies from the agar plates were picked with sterile toothpicks and used to inoculate starter cultures (0.5 mL LB-amp) in 96 deep-well plates. The starter culture plates were grown at 37 °C, 250 rpm, and 80% humidity in a Multitron Infors shaker overnight (14–16 hours). The starter cultures (50 μL) were used to inoculate expression cultures (1 mL TB-amp) in 96 deep-well plates. In parallel, glycerol stock plates were prepared for long-term storage by mixing starter culture (50 μL) with sterile glycerol (50% v/v, 50 μL) and frozen at −80 °C. The expression cultures were grown at 37 °C, 250 rpm, and 80% humidity for three hours, then cooled on ice for 20 minutes. Protein expression was induced by addition of IPTG (0.5 mM). Cultures were incubated at 22 °C and 220 rpm overnight (18–20 hours). Cells were pelleted (5000×*g*, 5 minutes) and the cell pellets were frozen at −20 °C for at least 24 hours prior to use.

Cells were resuspended in MOPS buffer (20 mM pH 7.0) containing 1 mM sodium acetate. Under air, ferrous ammonium sulfate (40 mM in water, 10 μL, 1 mM final concentration, prepared immediately before use), L-ascorbic acid (40 mM in water, 10 μL, 1 mM final concentration, prepared immediately before use), styrene (400 mM in ethanol, 10 μL, 10 mM final concentration), and *p*-toluenesulfonyl azide (400 mM in ethanol, 10 μL, 10 mM final concentration) were added to each well. The plates were sealed with foil covers and shaken at room temperature for two hours. To quench the reactions, acetonitrile (400 μL) was added and the reaction plate was shaken for an additional 30 minutes. Insoluble material was pelleted by centrifugation (6000×*g*, 10 minutes) and 200 μL of the supernate was filtered through a 0.2 μm PTFE 96-well filter plate into a 96-well

[2] Kille, S.; Acevedo-Rocha, C.G.; Parra, L.P.; Zhang, Z.-G..; Opperman, D.J.; Reetz, M.T.; Acevedo, J.P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol*. **2013**, *2*, 83–92.

microplate (3000×g, 2 minutes). The microplate was sealed with a pierceable cover and analyzed via HPLC-MS (Analytical instrumentation).

Table S1. Oligonucleotides used for mutagenesis. Mutated codons are denoted here as NNN for simplicity; in practice they are a 12:9:1 ratio of NDT:VHG:TGG for site saturation or the appropriate single codon for site-directed mutagenesis.

| Mutations relative to wild type	Direction	Sequence
T97X	Forward	CCGACTTCCCCGAAATTTTC*NNN*GTCTGCAAAGATCTTTC
T97X	Reverse	GAAAATTTCGGGGAAGTCGGGCTTTCCAGCAGTCACCTC
R171X	Forward	GATGGATGGCACCACATG*NNN*GTGTTGCGTTTTCCGCC
R171X	Reverse	CATGTGGTGCCATCCATCGCGGGTCAAATCTG
R277X	Forward	GGTGAAACTTAATACACGTGAG*NNN*TTTGCTTGCGCGTACTTCCATGAGCCG
R277X	Reverse	CACGTGTATTAAGTTTCACCTTATGCGGAGTGCTAAGTAACTGTCCCCCG
F314X C317M	Forward	CACTATGGGGAACATTTCACGAACATG*NNN*ATGCGTATGTATCCTGACCG
F314X	Reverse	CATGTTCGTGAAATGTTCCCCATAGTGAATGCGCTCATTGGCC
C317X	Forward	TCACGAACATGTTCATGCGT*NNN*TATCCTGACCGCATTACCACACAGC
C317X	Reverse	CATGAACATGTTCGTGAAATGTTCCCCATAGTGAATGCGCTC

Analytical-scale biocatalytic aziridination reactions

Biocatalytic reactions were set up in 2-mL screw-cap vials (Agilent). Purified apoprotein (350 µL, 22.9 µM in 20 mM MOPS pH 7.0, final concentration 20 µM) was added to the vial. Solutions of ferrous ammonium sulfate and L-ascorbic acid were prepared immediately prior to use. Reactions to be set up anaerobically were brought into a Coy vinyl anaerobic chamber (nitrogen atmosphere, 0–10 ppm oxygen). To each reaction was added in order ferrous ammonium sulfate (40 mM in water, 10 µL, 1 mM final concentration), sodium acetate or other additive (40 mM in water, 10 µL, 1 mM final concentration), and L-ascorbic acid (40 mM in water, 10 µL, 1 mM final concentration). Each reaction was then charged with styrene (400 mM in ethanol, 10 µL, 10 mM final concentration) immediately followed by *p*-toluenesulfonyl azide (400 mM in ethanol, 10 µL, 10 mM final concentration, 500 max. TTN). The reactions were sealed and shaken at room temperature for three hours unless otherwise noted. To quench the reactions, acetonitrile (350 µL) was added to each vial, followed by internal standard propiophenone (0.1% v/v in acetonitrile, 50 µL). The sample was transferred to a 1.7-mL Eppendorf tube, vortexed, and then centrifuged (20817×g, 5 minutes). 250 µL of the supernate was transferred to HPLC vial inserts for reverse-phase HPLC analysis. The remaining supernate was partially concentrated *in vacuo* to remove acetonitrile and ethanol. Cyclohexane (500 µL) was added to the resulting aqueous suspension. The mixture was thoroughly shaken and then centrifuged (20817×g, 5 minutes). 250 µL of the organic layer was transferred to HPLC vial inserts for normal-phase chiral HPLC analysis.

Analytical-scale biocatalytic C–H insertion reactions

Biocatalytic reactions were set up in 2-mL screw-cap vials (Agilent). Purified apoprotein (360 µL, 22.2 µM in 20 mM MOPS pH 7.0, final concentration 20 µM) was added to the vial. Solutions of ferrous ammonium sulfate and L-ascorbic acid were prepared immediately prior to use. Reactions to be set up anaerobically were brought into a Coy vinyl anaerobic chamber (nitrogen atmosphere, 0–10 ppm oxygen). To each reaction was added in order ferrous ammonium sulfate (40 mM in water, 10 µL, 1 mM final concentration), sodium acetate or other additive (40 mM in water, 10 µL, 1 mM final concentration), L-ascorbic acid (40 mM in water, 10 µL, 1 mM final concentration). Each reaction was then charged with 2-ethylbenzenesulfonyl azide (400 mM in ethanol, 10 µL, 10 mM final concentration, 500 max. TTN). The reactions were sealed and shaken at room temperature for six hours unless otherwise noted. To quench the reactions, acetonitrile (350 µL) was added to each vial, followed by internal standard propiophenone (0.5% v/v in acetonitrile, 50 µL). The sample was transferred to a 1.7-mL Eppendorf tube, vortexed, and then centrifuged (20817×g, 5 minutes). 250 µL of the supernate was transferred to HPLC vial inserts for reverse-phase HPLC analysis. The remaining supernate was partially concentrated *in vacuo* to remove acetonitrile and ethanol. Hexanes (250 µL, HPLC grade) and ethyl acetate (250 µL, HPLC grade) were added. The resulting mixture was thoroughly shaken and then centrifuged (20817×g, 5 minutes). 250 µL of the organic layer was transferred to HPLC vial inserts for normal-phase chiral HPLC analysis.

EXPERIMENTAL DATA

## Initial evaluation of α-ketoglutarate-dependent iron dioxygenases
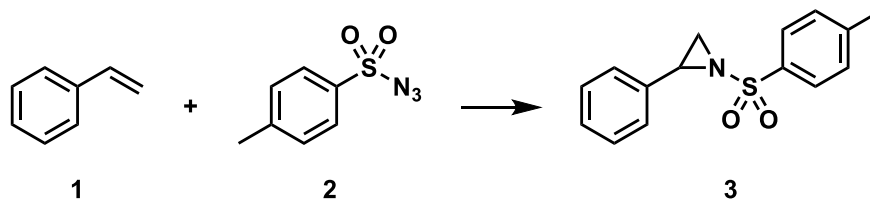
Reactions were performed as described above (Analytical-scale biocatalytic aziridination reactions) except enzyme concentrations were 50 µM, disodium α-ketoglutarate was used as the additive, and acetonitrile was used as the co-solvent. Activity was assayed by HPLC-MS (Analytical instrumentation). Activities are normalized to the negative control, bovine serum albumin.

Table S2. Activities of α-KG-dependent iron enzyme towards aziridination to form **3**.

Enzyme	Relative activity
P. savastanoi **ethylene-forming enzyme**	**12.0**
Streptomyces sp. 2-aminobutyric acid chlorinase	0.93
A. thaliana anthocyanidin synthase	0.54
G. oxydans leucine dioxygenase	1.11
E. coli taurine dioxygenase	0.61
S. vinaceus arginine hydroxylase	0.57
S. muensis leucine hydroxylase	0.61
Bovine serum albumin (negative control)	1.00

Reaction condition controls

Aziridination reaction
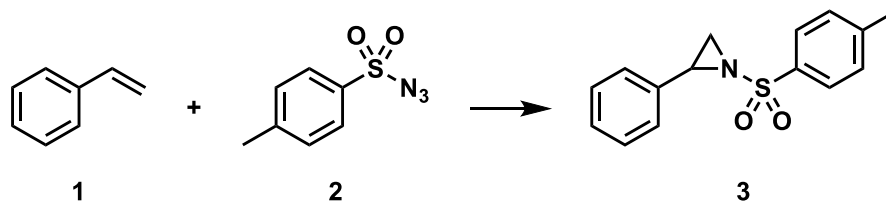
Reactions were performed in triplicate as described above with acetate as additive (Analytical-scale biocatalytic aziridination reactions) except where noted. Yields given are the average of the triplicates.

Table S3. Aziridination reaction controls with wild-type *Ps*EFE.

Deviation from standard conditions	Aziridine yield (%)
None	0.56%
No iron	0.01%
No ascorbate	0.50%
No acetate	0.04%
αKG instead of acetate	0.08%
Succinate instead of acetate	0.11%
N-oxalylglycine instead of acetate	0.64%

Table S4. Aziridination reaction controls with *Ps*EFE T97M R171L R277H F314M C317M

Deviation from standard conditions	Aziridine yield (%)
None	23.8
No iron	0.1
No ascorbate	15.2
No acetate	2.1
α-Ketoglutarate instead of acetate	2.6
N-oxalylglycine instead of acetate	2.4
Aerobic	4.7
Aerobic, no ascorbate	2.9

C–H insertion reaction
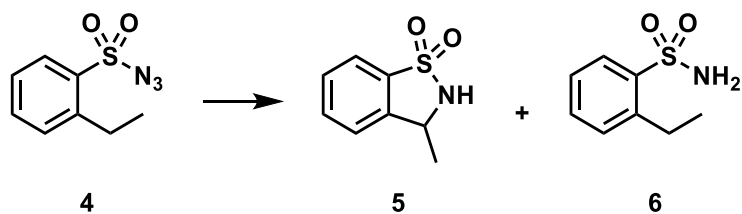
4   **5**   **6**

Reactions were performed in triplicate as described above (Analytical-scale biocatalytic C–H insertion reactions) with acetate as the additive, except where otherwise noted. Data shown are the average of the triplicates.

Table S5. C–H insertion reaction controls (max. 500 TTN)

Variant	Additive	TTN (**5**)	**5/6**
Wild type	Acetate	12.3	1.6
R171V F314M C317M	Acetate	313	24
R171V F314M C317M	None	24.6	0.9
R171V F314M C317M	α-Ketoglutarate	130	9.0
R171V F314M C317M	*N*-Oxalylglycine	447	105
R171V R277H F314M C317M	Acetate	243	32
R171V R277H F314M C317M	None	27.0	3.4
R171V R277H F314M C317M	α-Ketoglutarate	30.8	3.8
R171V R277H F314M C317M	*N*-Oxalylglycine	33.3	4.1

C–H insertion reaction with whole cells and cell lysate

Reactions were performed in triplicate as described above (Analytical-scale biocatalytic C–H insertion reactions), except whole cell suspensions and cell lysates were used instead of purified protein solutions. Data shown are the average of the triplicates. Note that the chemoselectivities are generally much lower than with purified protein; this is presumably due to increased reduction by the cellular background.

Table S6. C–H insertion with whole cells and lysate

Variant	Formulation	Additive	% Yield (5)	5/6
VMM	Whole cells	None	36%	1.6
VMM	Whole cells	Acetate	37%	1.7
VMM	Whole cells	αKG	36%	1.6
VMM	Whole cells	NOG	30%	1.6
VMM	Lysate	None	76%	7.7
VMM	Lysate	Acetate	72%	7.0
VMM	Lysate	αKG	69%	6.3
VMM	Lysate	NOG	90%	12.7
VHMM	Whole cells	None	33%	3.5
VHMM	Whole cells	Acetate	34%	3.7
VHMM	Whole cells	αKG	33%	3.4
VHMM	Whole cells	NOG	27%	3.1
VHMM	Lysate	None	73%	15.8
VHMM	Lysate	Acetate	75%	16.9
VHMM	Lysate	αKG	73%	14.3
VHMM	Lysate	NOG	77%	17.0

SDS-PAGE of *Ps*EFE variants

Large-scale protein expression, lysis, and purification was carried out as described in the methods section (Protein expression and purification). Whole-cell samples were taken after resuspension, clarified lysate samples were taken following sonication and centrifugation, and purified protein samples were taken after HisTrap purification. These samples were mixed 1:1 with 2X Laemmli loading buffer (Bio-Rad Laboratories, Inc.) with added 2-mercaptoethanol. Samples were heated to 95 °C in a thermomixer block, briefly centrifuged, and loaded on an Any kD™ Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad Laboratories, Inc.). Gels were run at 150 V for 30–45 minutes. Gels were washed with water, then stained by microwaving the gels with Coomassie solution. Gels were destained with successive rounds of microwaving with water, followed by gentle shaking overnight in water.

Aziridination variant protein purification
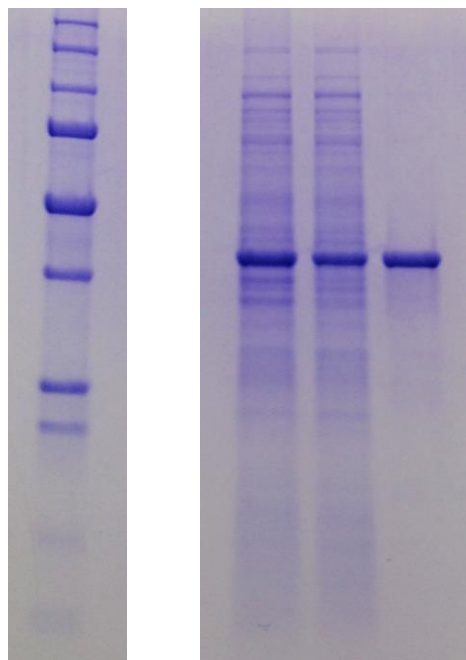

Figure S1. SDS-PAGE of *Ps*EFE aziridination variant T97M R171L R277H F314M C317M (PsEFE MLHMM). The protein is shown from left to right as whole-cell sample, clarified lysate sample, and purified protein sample. Whole cells and lysates were diluted 25-fold; purified protein was diluted 50-fold (each dilution is prior to addition of 2X Laemmli loading dye). The ladder and sample were run on the same gel; unrelated protein samples were cropped out for image clarity. The SDS-PAGE image brightness was increased in Microsoft Word for image clarity and is not being used for quantitation.

C–H insertion variant protein purification
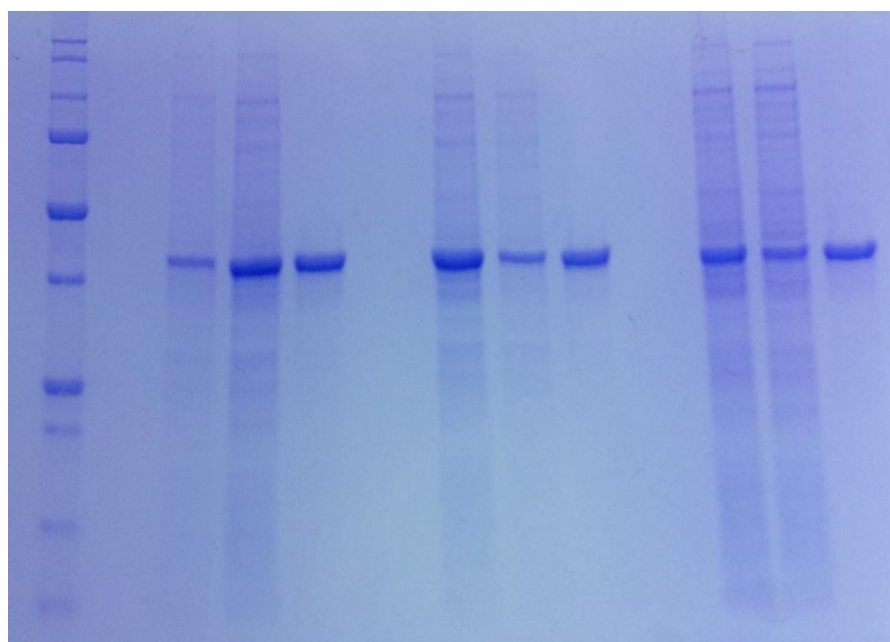
Figure S2. SDS-PAGE of *Ps*EFE C–H insertion variants. Each protein is shown from left to right as whole-cell sample, clarified lysate sample, and purified protein sample. Whole cells and lysates were diluted 25-fold; purified protein was diluted 50-fold (each dilution is prior to addition of 2X Laemmli loading dye). Protein 1:

R171V F314M C317M, Protein 2: R171V R277H F314M C317M. Protein 3: T97M R171V R277H F314L C317M. The SDS-PAGE image brightness was increased in Microsoft Word for image clarity and is not being used for quantitation.

Reaction time courses

Aziridination time course

A time course was run with *Ps*EFE WT and *Ps*EFE MLLHMM, a variant with one additional mutation (I186L) relative to *Ps*EFE MLHMM with decreased activity and slightly increased enantioselectivity. Purified protein reactions were set up in triplicate both anaerobically and aerobically, as described in the methods section above (Analytical-scale biocatalytic aziridination reactions). Time points were taken at 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours, 4 hours, and 8 hours, at which point those reactions were quenched by addition of acetonitrile (350 µL) and the internal standard propiophenone (1 µL mL^{-1} in acetonitrile, 50 µL).

As we see in Figure S3, the reaction with wild-type *Ps*EFE is essentially done by the 15 minute mark, while the yield with MLLHMM continues to increase until approximately 2 hours. The reaction appears to proceed for a longer time anaerobically than aerobically for *Ps*EFE MLLHMM; the low activity and fast reaction completion for wild-type *Ps*EFE makes such comparisons challenging.
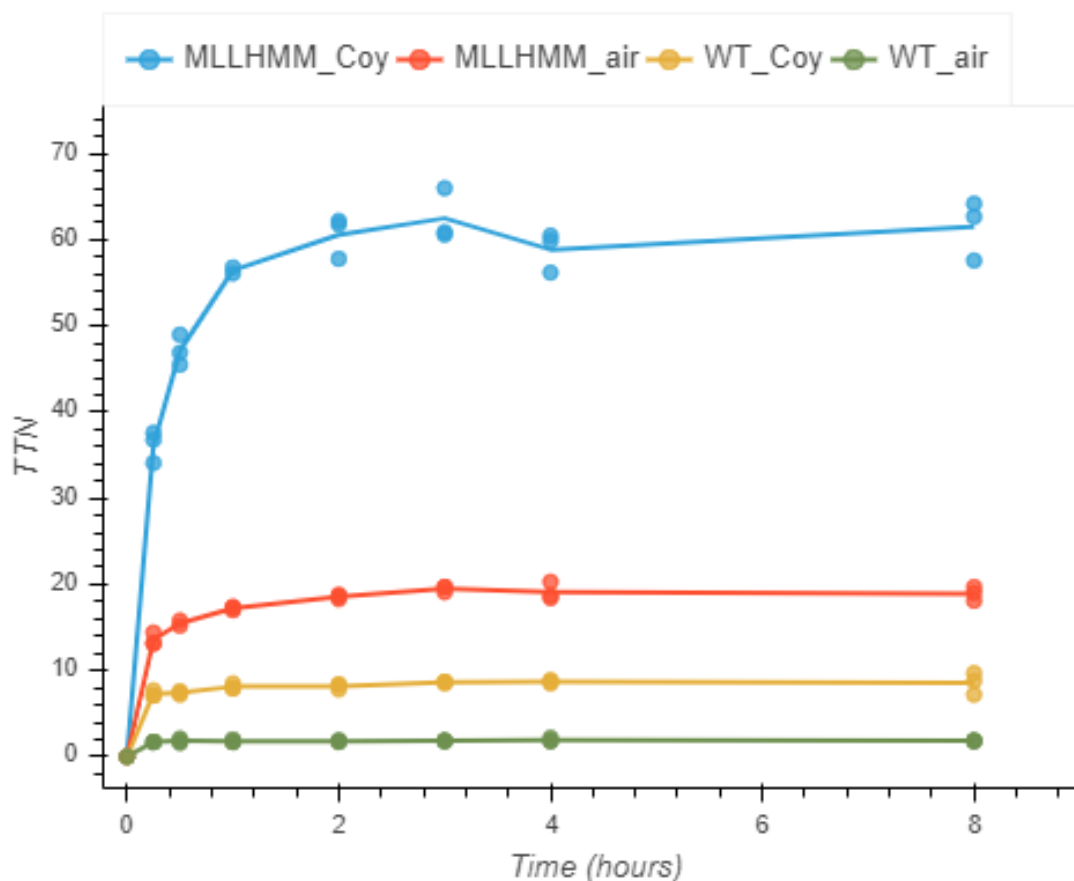
Figure S3. Aziridination time course. Time courses denoted "air" were set up aerobically; time courses denoted "Coy" were set up anaerobically.

Evolved variant thermostability

During our purified protein reactions, we do observe what is ostensibly protein precipitation over time. This visual observation, together with our time-course data, indicate that one possible reason for limited activity might be the protein's stability to the reaction conditions over time. The thermostability of wild-type *Ps*EFE has been measured with ITC[3]; not surprisingly, the protein is reported to have increased stability in the presence of iron and α-ketoglutarate. We used the thermal shift assay[4] using SYPRO orange (Thermo Fisher Scientific).

Thermal shift assay samples were prepared in triplicate anaerobically under similar conditions as described above (Reaction condition controls). To a PCR tube with purified *Ps*EFE wild type or a *Ps*EFE variant (stripped and dialyzed, 10–15 μM final concentration) was added (to a final concentration of 1.25 mM each) either:

- ferrous ammonium sulfate
- ferrous ammonium sulfate, L-ascorbic acid, and α-ketoglutarate
- ferrous ammonium sulfate, L-ascorbic acid, and sodium acetate

Following these additions, to each tube was added 5 μL SYPRO orange (25-fold diluted in water). The PCR tubes were sealed, brought out of the anaerobic chamber, and analyzed on an Stratagene Mx3005P qPCR machine (Agilent Technologies, Inc.). The temperature program ran from 25 °C to 99 °C, holding for 30 seconds per degree before measuring fluorescence on the SYPRO channel and increasing temperature. The melting temperature for a given temperature was taken as the maximum of the numerical first derivative, representing the inflection point of the protein's melt curve.

The data are presented in Figure S4. We can see that, even though beneficial mutations were only chosen based on activity and stereoselectivity, the protein's stability improved from wild type to the final variants. We also see a significant enhancement in thermostability for early variants upon addition of α-ketoglutarate (noted as 2OG for 2-oxoglutarate), which is not observed in the later evolved variants.

---

[3] Li, M.; Martinez, S.; Hausinger, R. P.; Emerson, J. P. Thermodynamics of Iron(II) and Substrate Binding to the Ethylene-Forming Enzyme. *Biochemistry* **2018**, *57*, 5696-5705.

[4] Ericsson, U. B.; Hallberg, B. M.; DeTitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **2006**, *357*, 289–298.
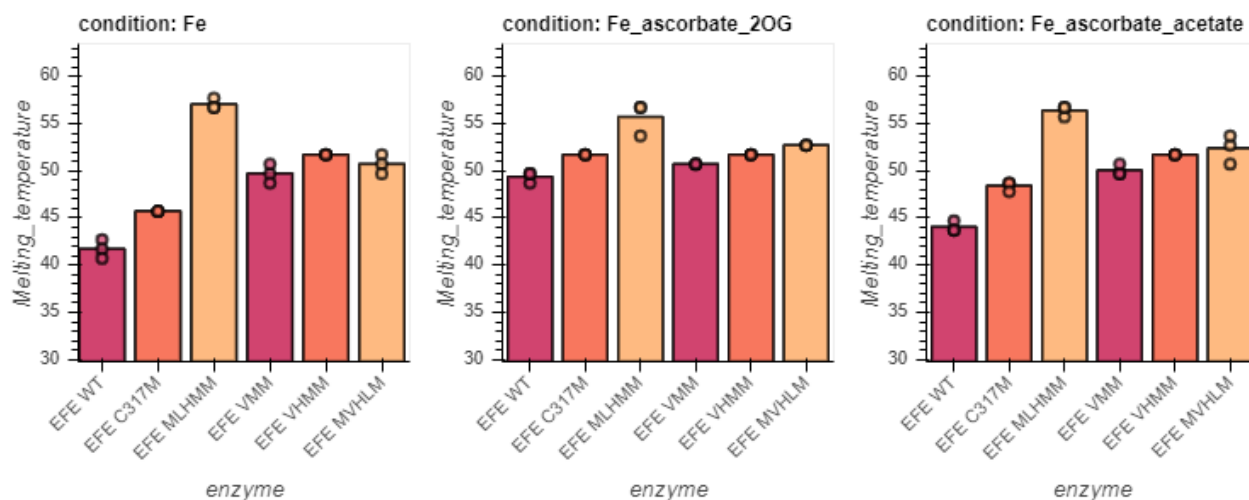
Figure S4. Thermostability of wild-type and evolved *Ps*EFE variants for aziridination and intramolecular C-H insertion.

Synthesis of sulfonyl azide substrates

Safety statement

Organic azides are potentially explosive compounds. We have not observed any problems in our handling of the compounds described, but care should be taken, especially on large scales.

p-Toluenesulfonyl azide (2)
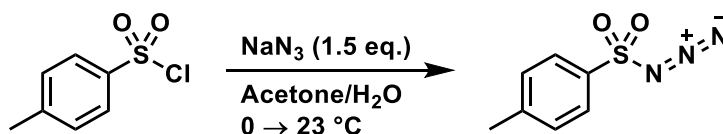
Under air, *p*-toluenesulfonyl chloride (19.1 g, 100 mmol, 1.00 equiv.) was dissolved in acetone (200 mL) in a 500 mL round-bottomed flask with a magnetic stir bar and cooled to 0 °C. A solution of sodium azide (9.75 g, 150 mmol, 1.50 equiv.) in water (60 mL) was added dropwise over one hour with stirring. Once addition was complete, the reaction was allowed to warm to room temperature and stirred for 16 hours. The acetone was removed *in vacuo*, and the resulting mixture was extracted with diethyl ether (2×100 mL). The combined organic layers were washed with water (2×100 mL), saturated aqueous sodium bicarbonate (100 mL), and brine (100 mL), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a colorless oil that solidified upon storage at −20 °C (19.1 g, 97%).

NMR Spectroscopy:

^{1}H NMR (500 MHz, CDCl$_3$, 23 ºC, δ): 7.80 (d, *J* = 8.5 Hz, 2 H), 7.38 (d, *J* = 8.5 Hz, 2 H), 2.45 (s, 3 H)

^{13}C NMR (125 MHz, CDCl$_3$, 23 ºC, δ): 146.3, 135.3, 130.3, 127.4, 21.7

2-Ethylbenzenesulfonyl chloride (S1)[5]
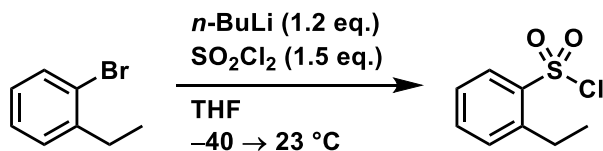
Under argon, 1-bromo-2-ethylbenzene (2.07 mL, 2.78 g, 15.0 mmol, 1.00 equiv.) was dissolved in anhydrous tetrahydrofuran (30 mL) in a 100 mL round-bottomed flask with magnetic stirring and cooled to −40 °C. *n*-Butyllithium (2.5 M solution in hexanes, 7.20 mL, 18.0 mmol, 1.20 equiv.) was added dropwise by syringe over two minutes. The reaction was stirred at −40 °C for thirty minutes, then sulfuryl chloride (1.82 mL, 3.04 g, 22.5 mmol, 1.5 equiv.) was added dropwise by syringe over two minutes. The reaction was allowed to warm to room temperature and stirred for 16 hours. The reaction was cooled to 0 °C, then carefully quenched by the addition of ice-cold water (50 mL). The resulting mixture was extracted with diethyl ether (2×50 mL). The combined organic layers were dried over sodium sulfate and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica (50 g), eluting with a gradient of 0 to 20% diethyl ether/hexanes, to afford the title compound as a slightly yellow oil (1.10 g, 36%).

NMR Spectroscopy:

¹H NMR (400 MHz, CDCl₃, 23 ºC, δ): 8.07 (d, *J* = 8.1 Hz, 1 H), 7.66 (t, *J* = 7.5 Hz, 1 H), 7.49 (d, *J* = 7.6 Hz, 1 H), 7.41 (t, *J* = 7.9 Hz, 1 H), 3.20 (q, *J* = 7.5 Hz, 2 H), 1.37 (t, *J* = 7.5 Hz, 3 H)

2-Ethylbenzenesulfonyl azide (4)[5]
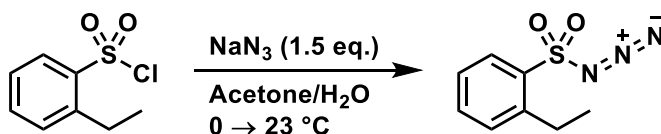
Under air, 2-ethylbenzenesulfonyl chloride **S1** (1.00 g, 4.89 mmol, 1.00 equiv.) was dissolved in acetone (8 mL) in a 20 mL scintillation vial with magnetic stirring and cooled to 0 °C. A solution of sodium azide (476 mg, 7.33 mmol, 1.50 equiv.) in water (2.5 mL) was added dropwise over two minutes with stirring. Once addition was complete, the reaction was allowed to warm to room temperature and stirred for six hours. The acetone was removed *in vacuo*, and the resulting mixture was extracted with diethyl ether (2×15 mL). The combined organic layers were washed with water, saturated aqueous sodium bicarbonate, and brine (15 mL each), then dried over magnesium sulfate and concentrated *in vacuo* to afford the title compound as a yellow oil (1.01 g, 98%).

NMR Spectroscopy:

¹H NMR (500 MHz, CDCl₃, 23 ºC, δ): 8.01 (dd, *J* = 8.0, 1.3 Hz, 1 H), 7.62 (td, *J* = 7.6, 1.3 Hz, 1 H), 7.46

[5] Ichinose, M.; Suematsu, H.; Yasutomi, Y.; Nishioka, Y.; Uchida, T.; Katsuki, T. Enantioselective Intramolecular Benzylic C–H Bond Amination: Efficient Synthesis of Optically Active Benzosultams. *Angew. Chem. Int. Ed.* **2011**, *50*, 9884–9887.

(dd, J = 7.7, 0.7 Hz, 1 H), 7.37 (td, J = 8.0, 1.2 Hz, 1 H), 3.03 (q, J = 7.5 Hz, 2 H), 1.31 (t, J = 7.5 Hz, 3 H)

13C NMR (125 MHz, CDCl$_3$, 23 ºC, δ): 144.5, 136.3, 134.9, 131.4, 129.4, 126.4, 26.1, 15.2

Synthesis of authentic product standards

2-Phenyl-1-(*p*-toluenesulfonyl)aziridine (*rac*-3)
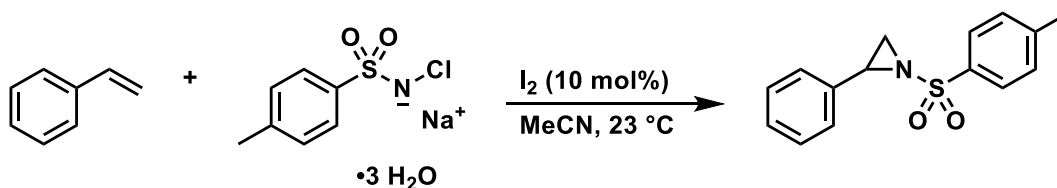
Under argon, chloramine-T trihydrate (4.23 g, 15.0 mmol, 1.00 equiv.) and iodine (381 mg, 1.50 mmol, 0.100 equiv.) were dissolved in acetonitrile (100 mL). Styrene (3.45 mL, 3.13 g, 30.0 mmol, 2.00 equiv.) was added dropwise, and the reaction was stirred at room temperature for 18 hours. The reaction mixture was partitioned between water (50 mL) and dichloromethane (100 mL), and the layers were separated. The aqueous layer was extracted with dichloromethane (2×100 mL). The combined organic layers were concentrated *in vacuo* and the residue was purified by flash column chromatography on silica (100 g) eluting with hexanes/ethyl acetate (6:1 v/v) containing 1% triethylamine to afford the title compound as a colorless solid (3.55 g, 87%).

NMR Spectroscopy:

1H NMR (500 MHz, CDCl$_3$, 23 ºC, δ): 7.87 (d, J = 8.3 Hz, 2 H), 7.33 (d, J = 8.3 Hz, 2 H), 7.30–7.25 (m, 3 H), 7.24–7.20 (m, 2 H), 3.78 (dd, J = 7.2, 4.5 Hz, 1 H), 2.99 (d, J = 7.2 Hz, 1 H), 2.44 (s, 3 H), 2.40 (d, J = 4.4 Hz, 1 H)

13C NMR (125 MHz, CDCl$_3$, 23 ºC, δ): 144.8, 135.1, 135.0, 129.9, 128.7, 128.4, 128.1, 126.7, 41.2, 36.1, 21.8

3-Methylbenzo[*d*]isothiazole 1,1-dioxide (S2)[6]
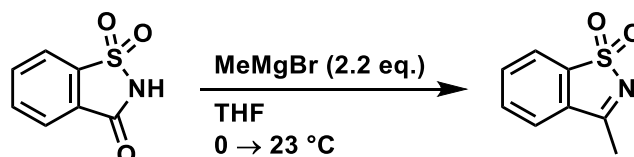
Under argon, saccharin (1.00 g, 5.46 mmol, 1.00 equiv.) was dissolved in anhydrous tetrahydrofuran (25 mL) in a 100 mL round-bottomed flask with magnetic stirring and cooled to 0 °C. Methylmagnesium bromide (3 M solution in diethyl ether, 4.00 mL, 12.0 mmol, 2.20 equiv.) was added dropwise by syringe over 10 minutes.

[6] Li, B.; Chen, J.; Zhang, Z.; Gridnev, I.Y., Zhang. W. Nickel-Catalyzed Asymmetric Hydrogenation of *N*-Sulfonyl Imines. *Angew. Chem. Int. Ed.* **2019**, *58*, 7329–7334.

After addition, the reaction was stirred at 0 °C for five minutes, then slowly warmed to room temperature and stirred at room temperature for 16 hours. The reaction was then cooled to 0 °C and carefully poured into ice-cold hydrochloric acid (1 M, 30 mL). The resulting mixture was extracted with dichloromethane (2×50 mL). The combined organic layers were dried over sodium sulfate and concentrated *in vacuo* to afford the title compound as a colorless solid (crude yield 1.04 g, 105%) which was used in the next step without further purification.

NMR Spectroscopy:

¹H NMR (400 MHz, CDCl$_3$, 23 ºC, δ): 7.94–7.90 (m, 1 H), 7.77–7.73 (m, 2 H), 7.71–7.67 (m, 1 H), 2.67 (s, 3 H)

3-Methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide (*rac*-5)
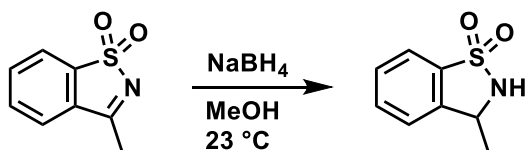
Under air, 3-methyl-[*d*]isothiazole 1,1-dioxide **S2** (500 mg, 2.76 mmol, 1.00 equiv.) was dissolved in methanol (20 mL) in a 50 mL round-bottomed flask with magnetic stirring. Sodium borohydride (522 mg, 13.8 mmol, 5.00 equiv.) was slowly added over two minutes. The reaction mixture bubbled vigorously and became warm to the touch. The reaction was stirred at room temperature for thirty minutes to ensure complete reaction. The reaction was cooled to 0 °C, then poured carefully into cold hydrochloric acid (2.5 M, 40 mL). The methanol was removed *in vacuo* and the resulting mixture was extracted with dichloromethane (3×25 mL). The combined organic layers were dried over sodium sulfate and concentrated *in vacuo*. The residue was purified by flash column chromatography on silica (25 g), eluting with a gradient of 10 to 60% ethyl acetate/hexanes to afford the title compound as a colorless solid (378 mg, 75%).

NMR Spectroscopy:

¹H NMR (500 MHz, CDCl$_3$, 23 ºC, δ): 7.74 (d, *J* = 8.0 Hz, 1 H), 7.61 (td, *J* = 7.6, 1.0 Hz, 1 H), 7.50 (t, *J* = 7.6 Hz, 1 H), 7.38 (d, *J* = 7.7 Hz, 1 H), 5.15 (br d, *J* = 4.8 Hz, 1 H), 4.78 (qd, *J* = 6.7, 4.8 Hz, 1 H), 1.59 (d, *J* = 6.7 Hz, 3 H)

¹³C NMR (125 MHz, CDCl$_3$, 23 ºC, δ): 141.8, 135.4, 133.3, 129.2, 124.0, 121.2, 53.5, 21.5

2-Ethylbenzenesulfonamide (6)
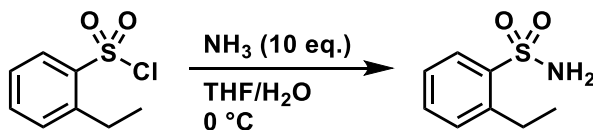
Under air, 2-ethylbenzenesulfonyl chloride **S1** (50.0 mg, 244 µmol, 1.00 equiv.) was dissolved in tetrahydrofuran (1 mL) in a 4 mL vial with magnetic stirring and cooled to 0 °C. Ammonia (28% w/v in water, 149 µL, 2.44 mmol, 10.0 equiv.) was added dropwise over one minute. After stirring for five minutes, the

reaction mixture was partitioned between water and ethyl acetate (10 mL each). The layers were separated and the aqueous layer was extracted with ethyl acetate (2×10 mL). The combined organic layers were washed with brine (10 mL), dried over sodium sulfate, and concentrated *in vacuo* to afford the title compound as a colorless solid (41.8 mg, 92%).
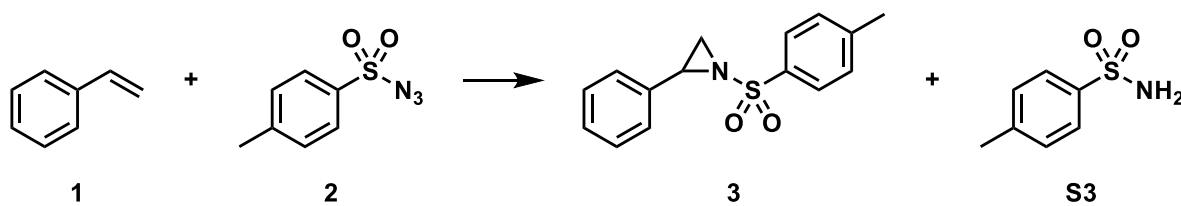
NMR Spectroscopy:

^{1}H NMR (500 MHz, CDCl$_3$, 23 ºC, δ): 7.99 (d, *J* = 8.0 Hz, 1 H), 7.51 (t, *J* = 7.5 Hz, 1 H), 7.39 (d, *J* = 7.6 Hz, 1 H), 7.29 (t, *J* = 7.7 Hz, 1 H), 5.02 (br s, 2 H), 3.07 (q, *J* = 7.5 Hz, 2 H), 1.33 (t, *J* = 7.5 Hz, 3 H)

^{13}C NMR (125 MHz, CDCl$_3$, 23 ºC, δ):143.0, 139.6, 133.1, 130.7, 128.3, 126.2, 26.1, 15.3

HPLC analytical methods and calibration curves

Aziridination reaction

Samples for HPLC calibration curves were prepared as simulated reaction samples. To MOPS buffer (20 mM pH 7.0, 380 µL) was added a solution of the appropriate reaction product in acetonitrile (0–100 µM, 20 µL, final concentration 0–5 mM). To this sample was added the internal standard propiophenone (0.1% v/v in acetonitrile, 50 µL) and acetonitrile (350 µL). The product concentration in the curves below corresponds to the concentration in the reaction mixture; the final analytical sample is two-fold diluted.

Analysis was performed on an Agilent 1200 series HPLC with water/acetonitrile mobile phase (1 mL min^{-1} flow), with an Agilent Poroshell 120 EC-C18 column (4 µm packing, 2.1×50 mm) fitted with a Poroshell 120 guard column (1.7 µm packing, 2.1×5 mm), injecting 5 µL. Detection was at 230 nm (16 nm bandwidth). The gradient program and retention times are given in Table S7 and Table S8, respectively.

Table S7. HLPC gradient program for aziridination analysis

Time (minutes)	% Acetonitrile
0.00	20
0.50	20
1.00	40
5.00	65
5.50	95
6.00	95
6.01	20
7.00	20

Table S8. HPLC retention times for aziridination analysis

Compound	Retention time (minutes)
p-Toluenesulfonamide	0.58
Propiophenone	2.24
p-Toluenesulfonyl azide	2.81
Styrene	3.01
2-Phenyl-1-(p-toluenesulfonyl)aziridine	3.45

2-Phenyl-1-(p-toluenesulfonyl)aziridine (3) calibration curve

$y = 1.5815x$
$R^2 = 0.9927$
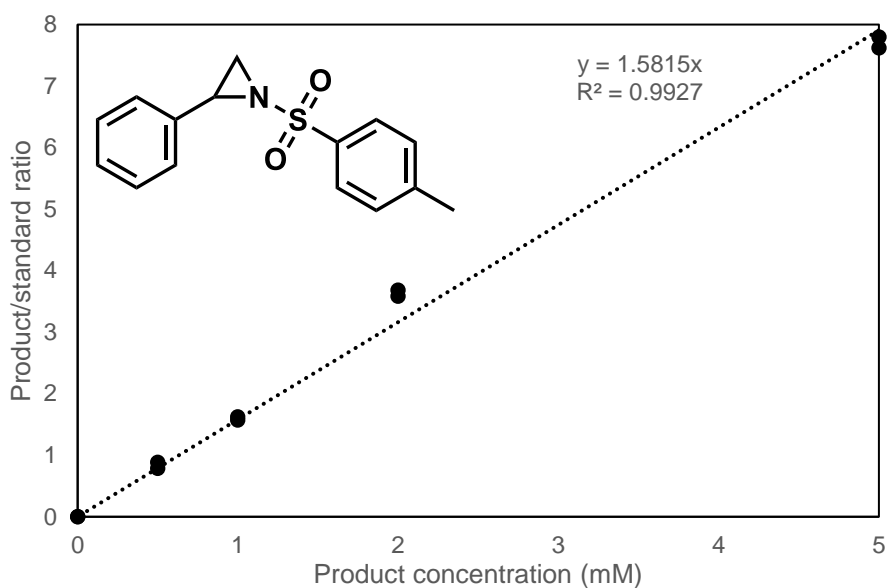
p-Toluenesulfonamide (S3) calibration curve

$y = 1.2648x$
$R^2 = 0.9992$
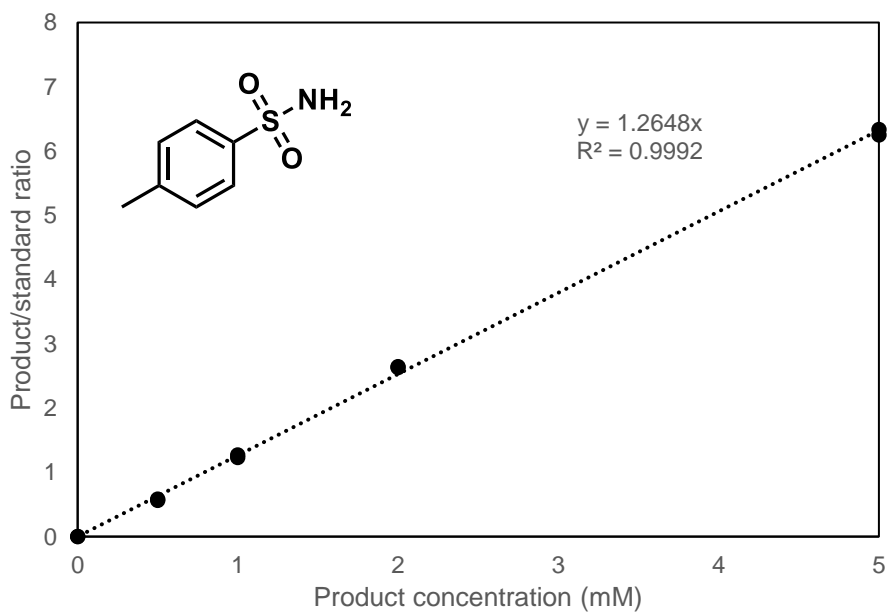
C–H insertion reaction
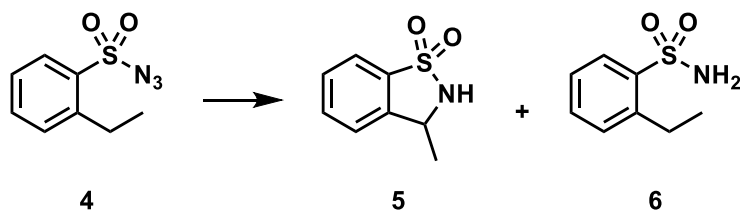
Calibration curve samples were prepared as described above for the aziridination reaction, except the internal standard used was propiophenone (0.5% v/v in acetonitrile, 50 μL) and the final product concentrations ranged from 1–10 mM. The product concentration shown corresponds to the concentration in the reaction mixture; the final analytical sample is twofold diluted.

Analysis was performed on an Agilent 1260 Infinity II HPLC instrument with water/acetonitrile mobile phase (1 mL min^{-1} flow), with an Agilent Poroshell 120 EC-C18 column (4 μm packing, 2.1×50 mm) fitted with a Poroshell 120 guard column (1.7 μm packing, 2.1×5 mm), injecting 5 μL. Detection was at 220 nm (4 nm bandwidth). The gradient program and retention times are given in Table S9 and Table S10, respectively.

Table S9. HPLC gradient program for C–H insertion analysis

Time (minutes)	% Acetonitrile
0.00	12
1.00	12
3.50	95
4.00	95
4.01	12
5.00	12

Table S10. HPLC retention times for C–H insertion analysis

Compound	Retention time (minutes)
3-Methyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide	1.05
2-Ethylbenzenesulfonamide	2.38
Propiophenone	3.00
2-Ethylbenzenesulfonyl azide	3.48

3-Methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide (5) calibration curve
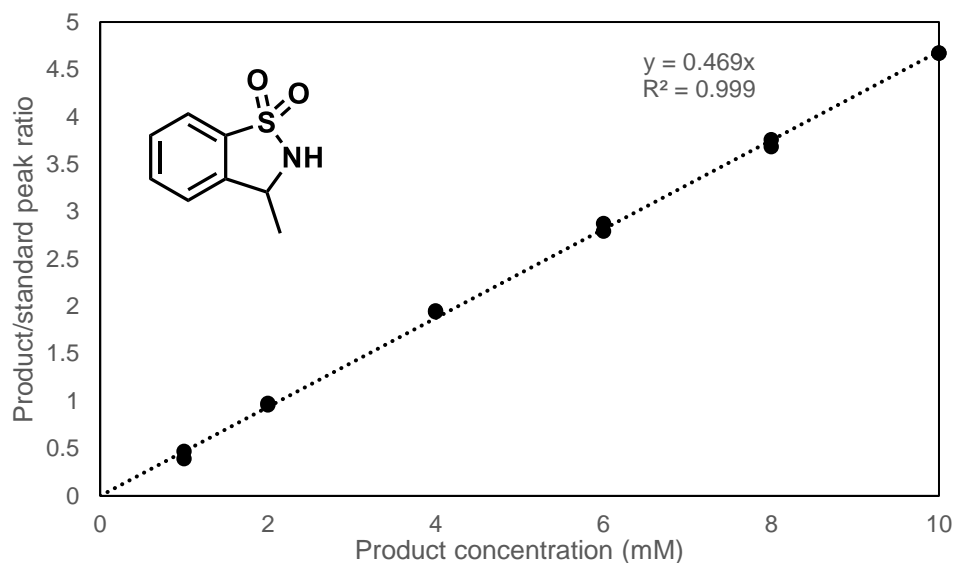
2-Ethylbenzenesulfonamide (6) calibration curve
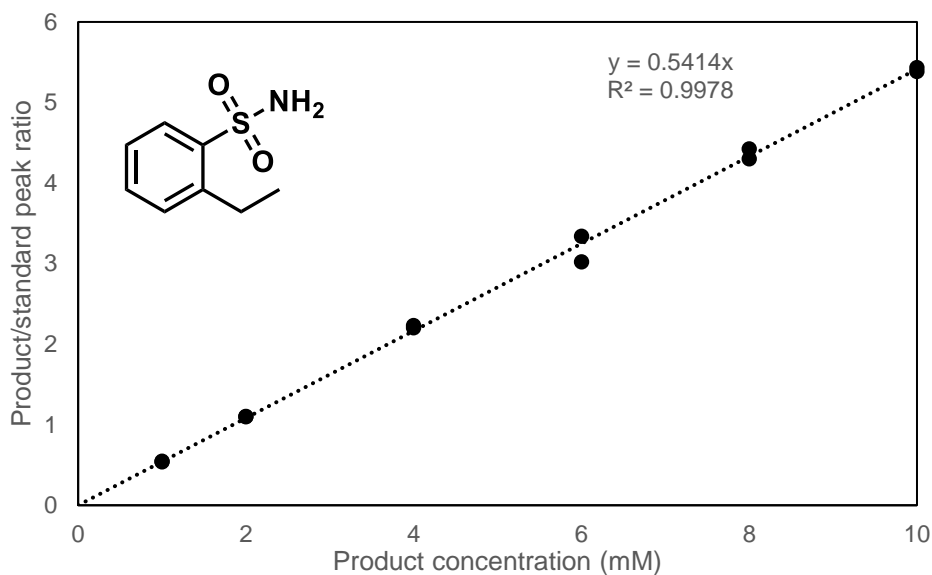
Chiral analysis

Chiral analysis was performed by HPLC with a chiral stationary phase, using a Hewlett Packard Series 1100 HPLC instrument with hexanes/2-propanol mobile phase (1 mL min^{-1} flow).
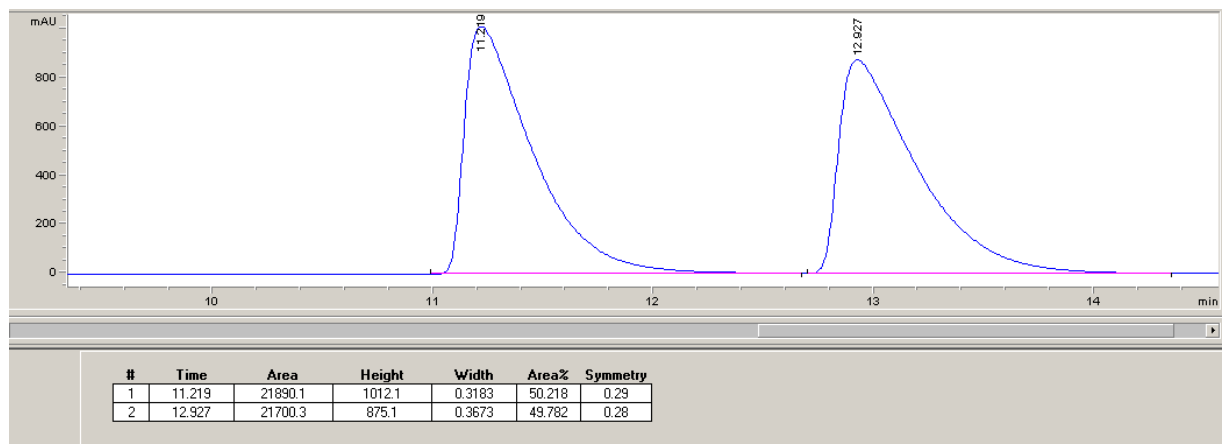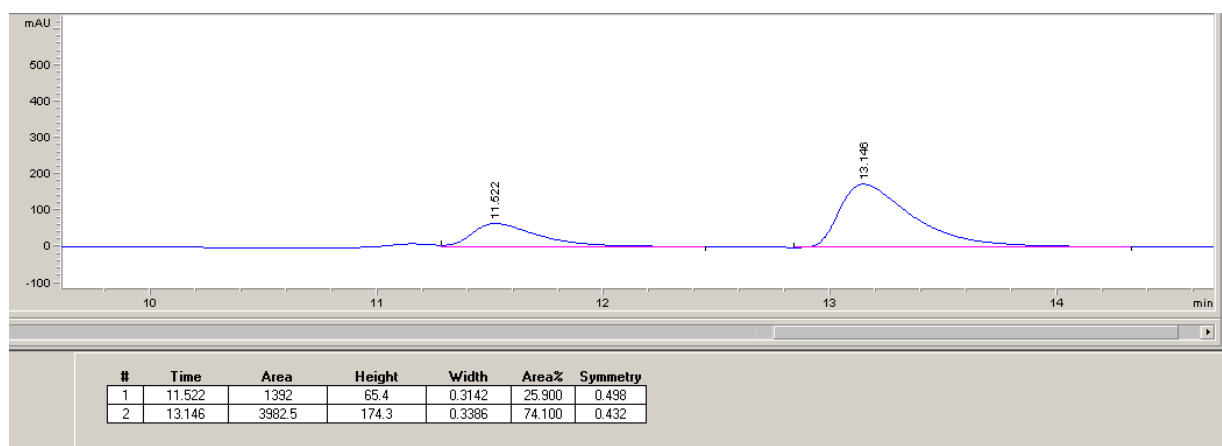
2-Phenyl-1-(*p*-toluenesulfonyl)aziridine (3)

Analysis was performed with a Daicel Chiralcel OJ-H column, (5 μm packing, 4.6×250 mm), with an isocratic 30% 2-propanol/70% hexanes mobile phase. The peak areas were analyzed at 235 nm (16 nm bandwidth).

DAD1 D, Sig=235,16 Ref=360,100 (C:\CHEM32\1\DATA\AMK\20191001_AZIR_LINEAGE 2019-10-01 16-04-14\001-0901.D)

#	Time	Area	Height	Width	Area%	Symmetry
1	18.327	6817.5	190.4	0.5357	49.722	0.447
2	22.626	6893.7	162.6	0.6346	50.278	0.51
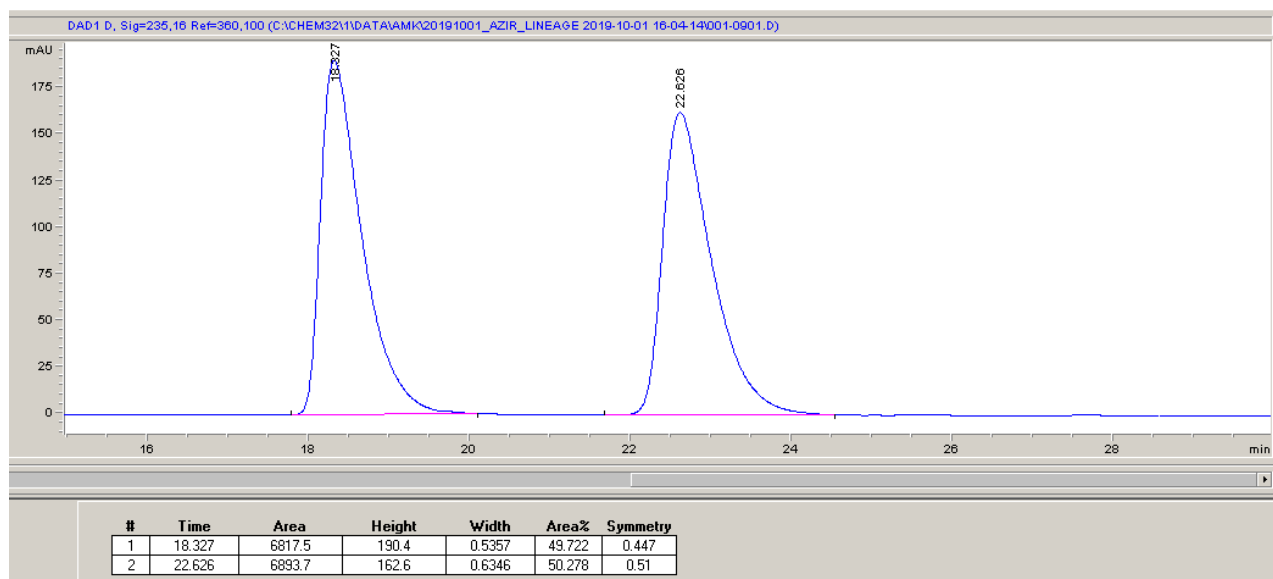
Figure S5. Chiral HPLC trace of *rac*-**3**.

#	Time	Area	Height	Width	Area%	Symmetry
1	18.334	446.1	13.2	0.4913	6.075	0.682
2	22.324	6897	164.8	0.6303	93.925	0.517
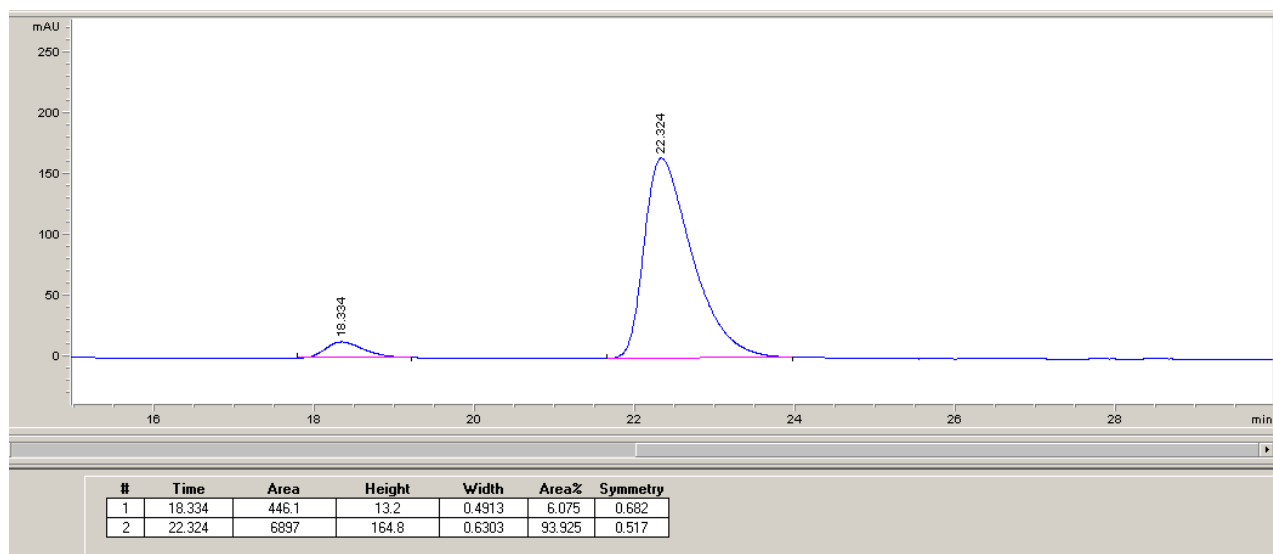
Figure S6. Chiral HPLC trace of *Ps*EFE MLHMM-catalyzed product **3**.

3-Methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide (5)

Analysis was performed with a Daicel Chiralpak IB column (5 µm packing, 4.6×250 mm), with an isocratic 25% 2-propanol/75% hexanes mobile phase. The peak areas were analyzed at 220 nm (16 nm bandwidth).

#	Time	Area	Height	Width	Area%	Symmetry
1	11.219	21890.1	1012.1	0.3183	50.218	0.29
2	12.927	21700.3	875.1	0.3673	49.782	0.28

Figure S7. Chiral HPLC trace of *rac*-**5**.

#	Time	Area	Height	Width	Area%	Symmetry
1	11.522	1392	65.4	0.3142	25.900	0.498
2	13.146	3982.5	174.3	0.3386	74.100	0.432

Figure S8. Chiral HPLC trace of *Ps*EFE VMM-catalyzed product **5**.

SPECTROSCOPIC DATA

¹H NMR spectrum of *p*-Toluenesulfonyl azide (2)

500 MHz, CDCl₃, 23 °C
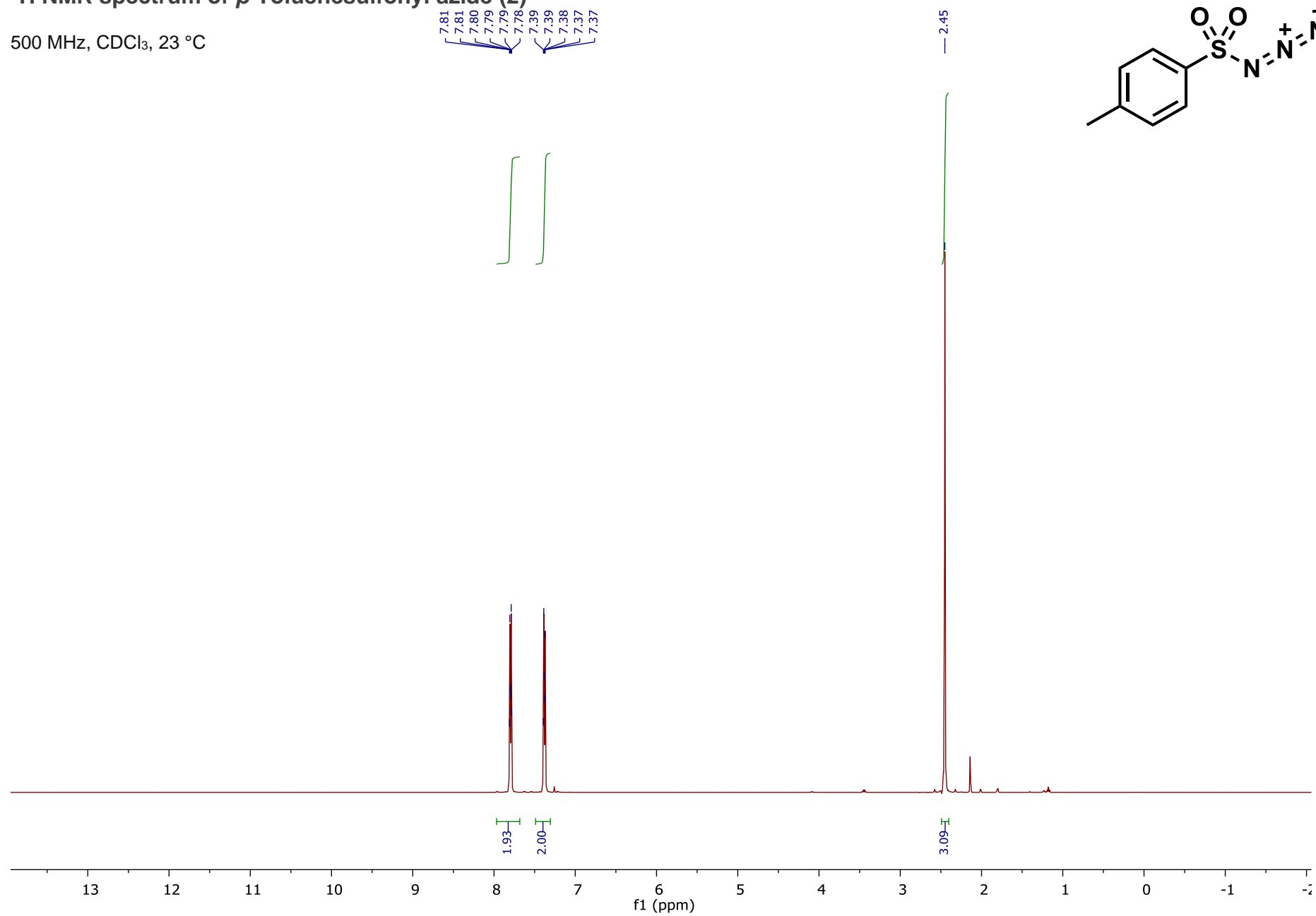
^{13}C NMR spectrum of *p*-Toluenesulfonyl azide (2)

125 MHz, CDCl$_3$, 23 °C
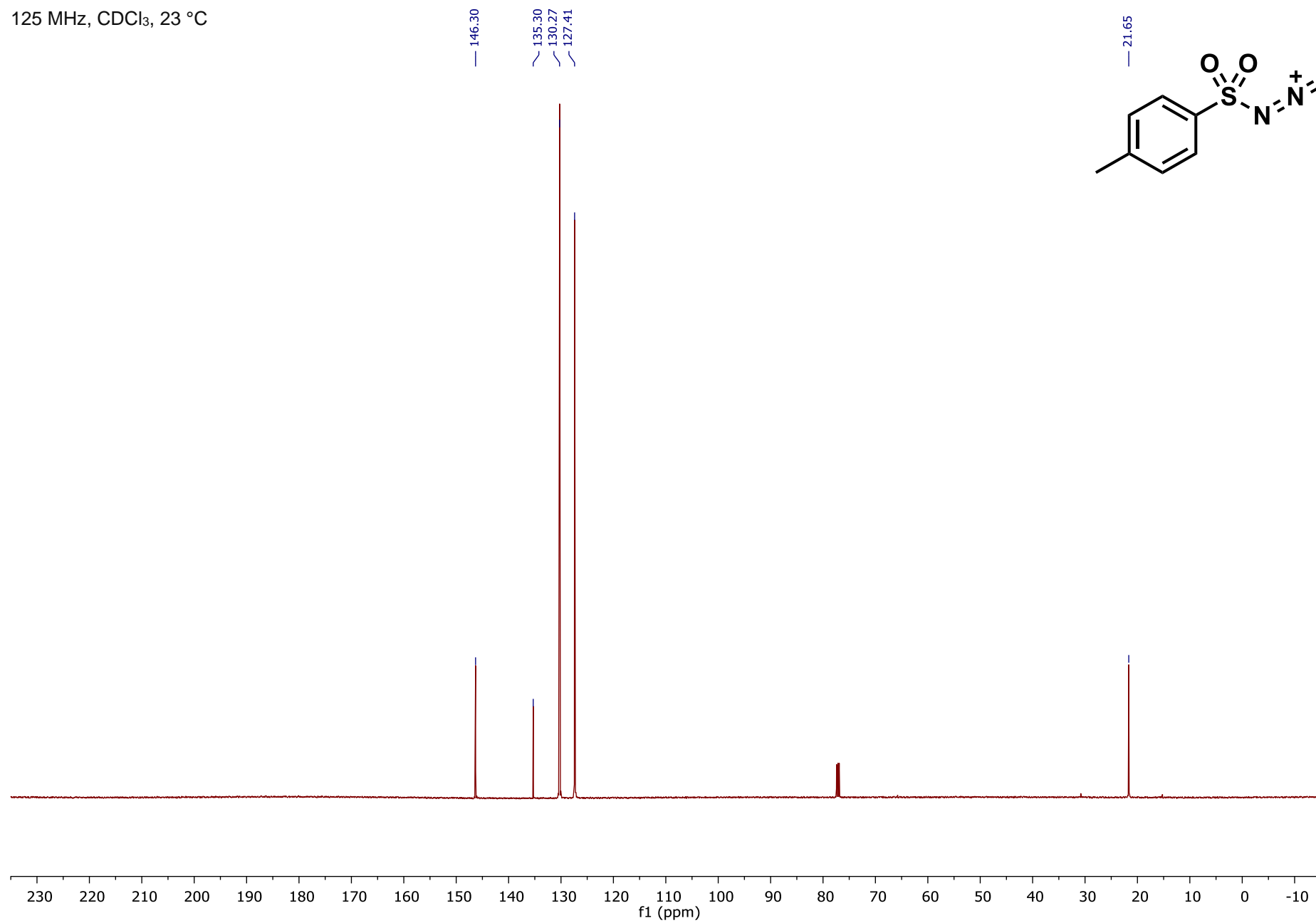
^{1}H NMR spectrum of 2-Ethylbenzenesulfonyl chloride (S1)

400 MHz, CDCl$_3$, 23 °C
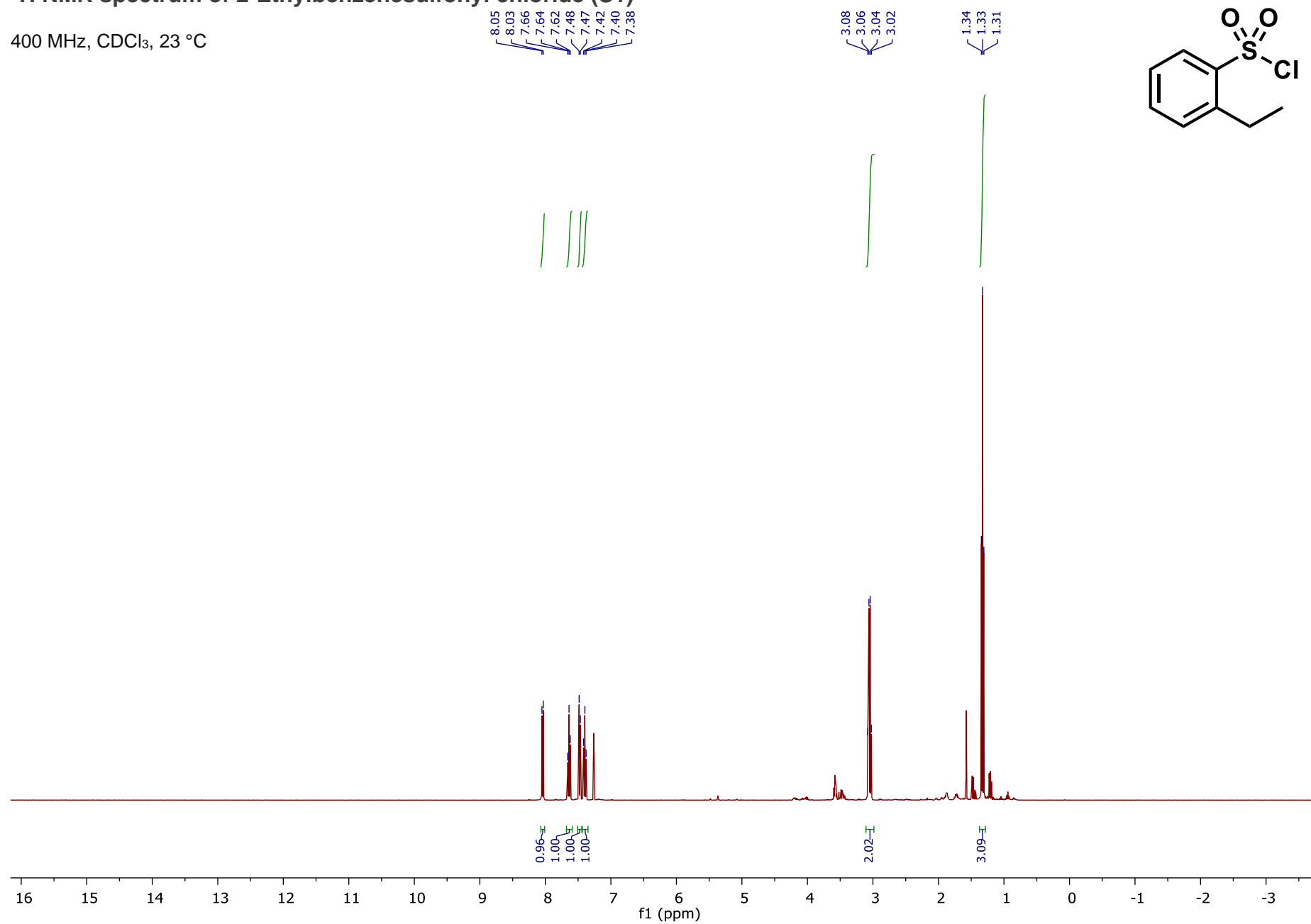
^{1}H NMR spectrum of 2-Ethylbenzenesulfonyl azide (4)

500 MHz, CDCl$_3$, 23 °C
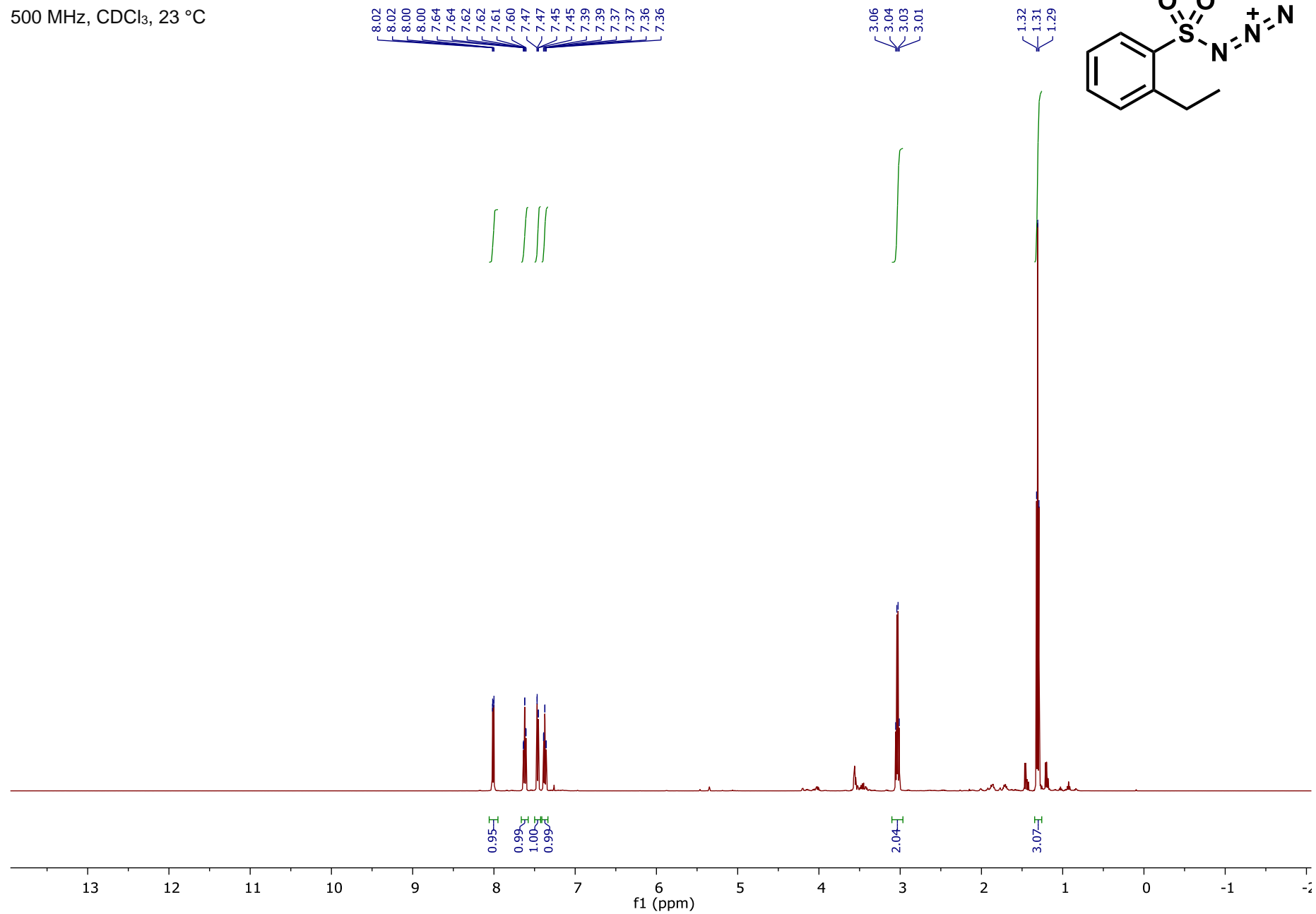
^{13}C NMR spectrum of 2-Ethylbenzenesulfonyl azide (4)

125 MHz, CDCl$_3$, 23 °C
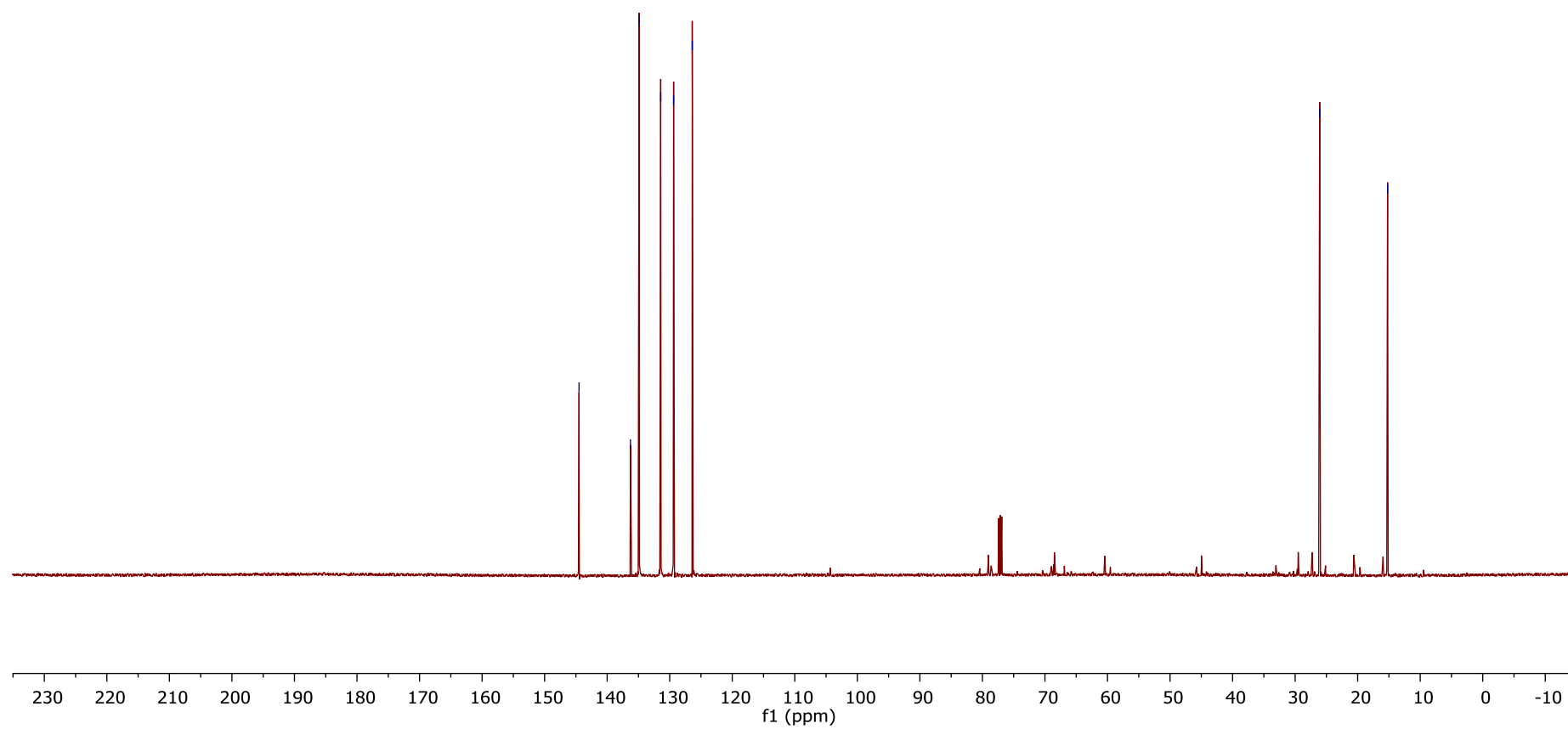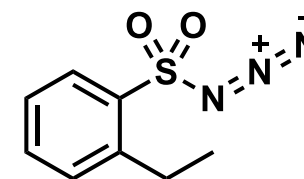
¹H NMR spectrum of 2-Phenyl-1-(*p*-toluenesulfonyl)aziridine (*rac*-3)

500 MHz, CDCl₃, 23 ° C
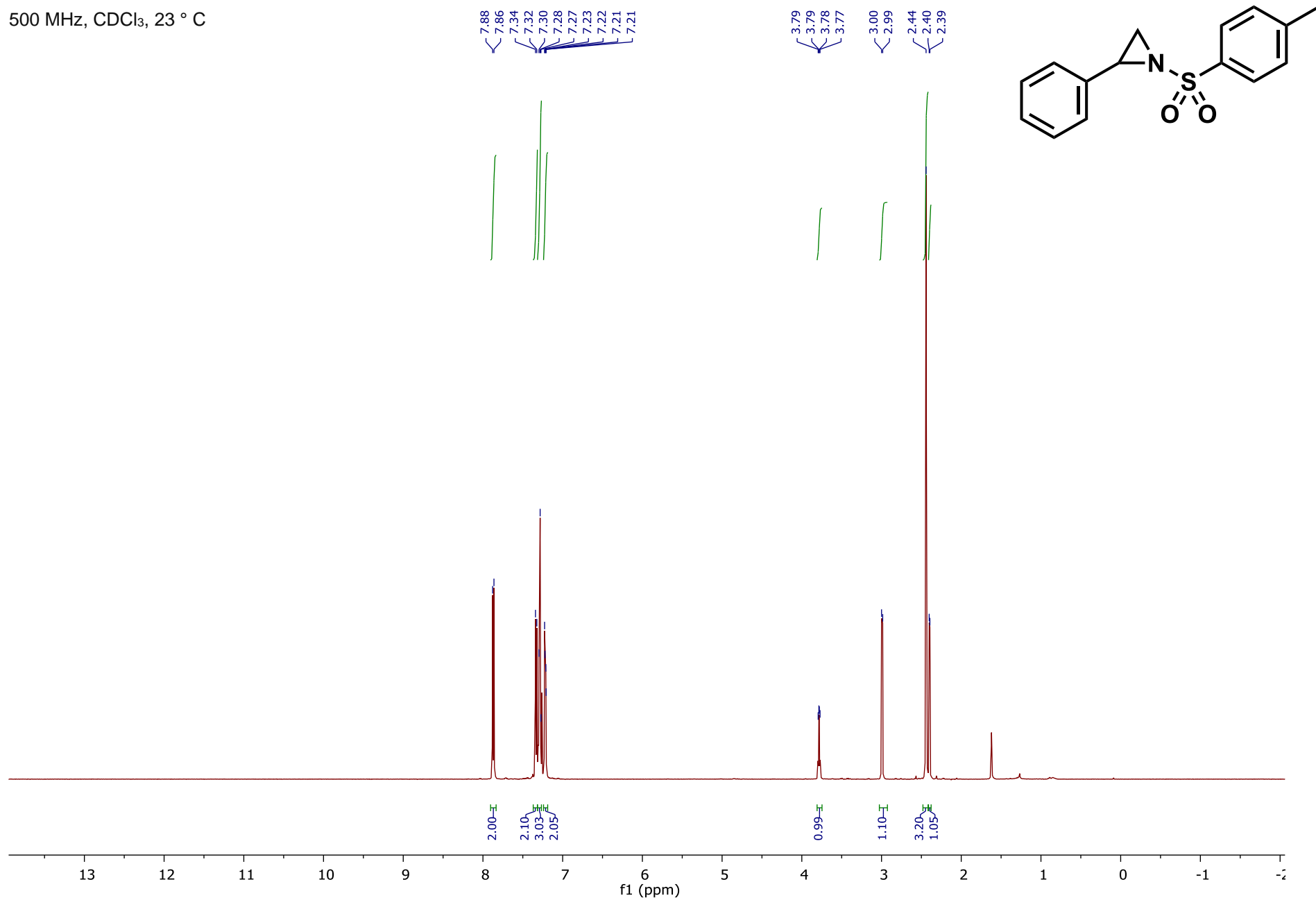
^{13}C NMR spectrum of 2-Phenyl-1-(p-toluenesulfonyl)aziridine (rac-3)

125 MHz, CDCl$_3$, 23 °C
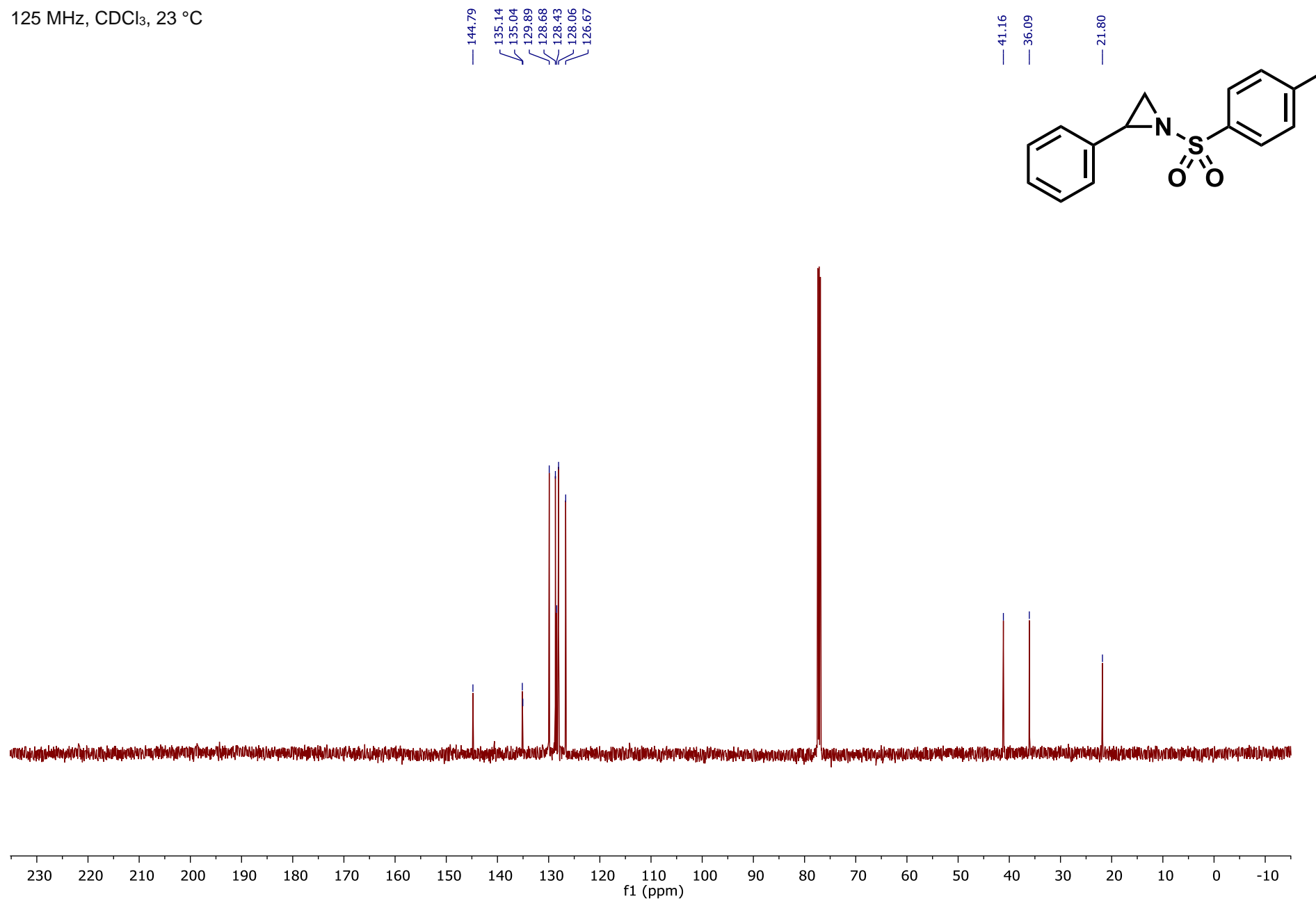
^{1}H NMR spectrum of 3-Methylbenzo[*d*]isothiazole 1,1-dioxide (S2)

400 MHz, CDCl$_3$, 23 °C
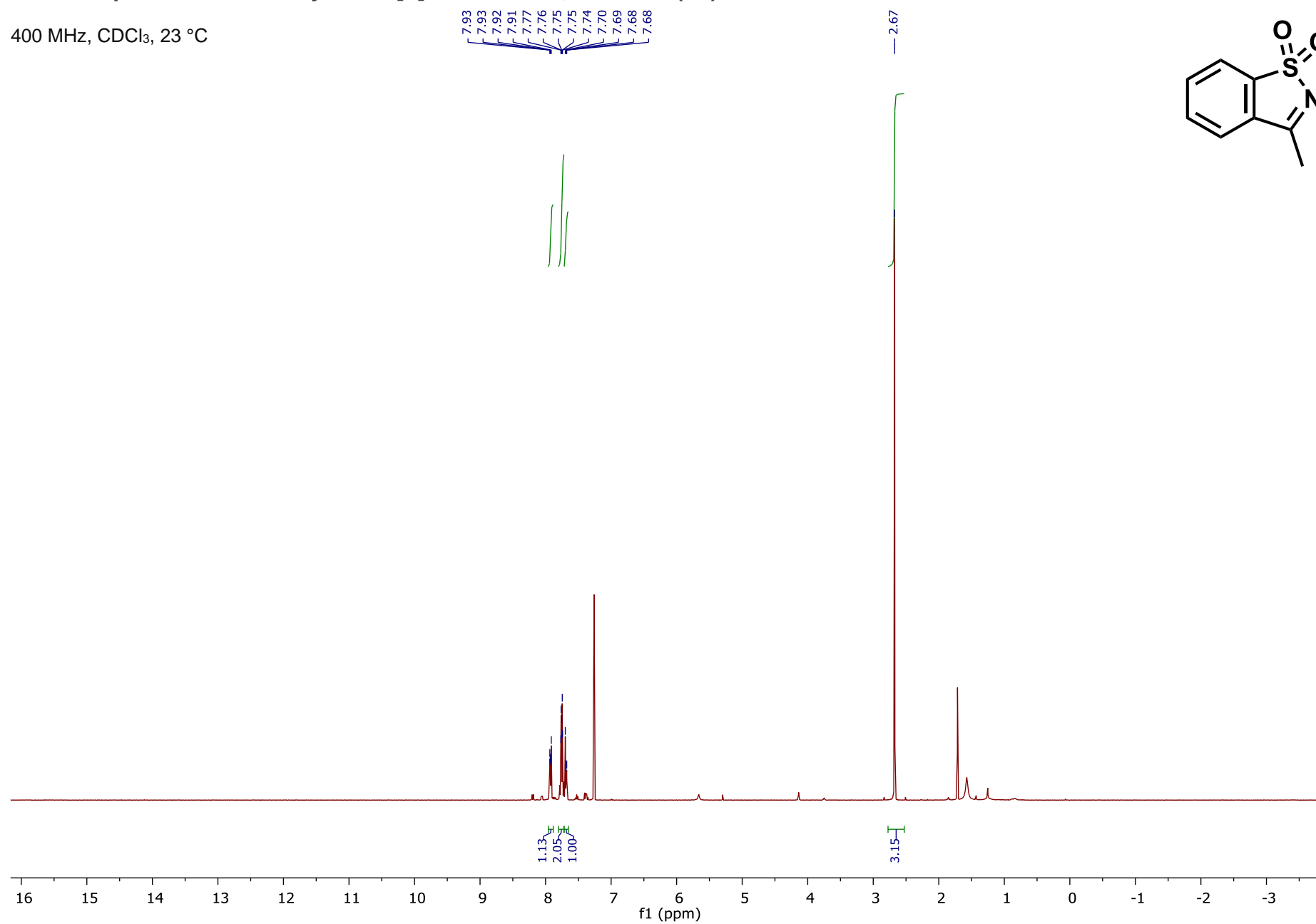
¹H NMR spectrum of 3-Methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide (*rac*-5)

500 MHz, CDCl₃, 23 °C
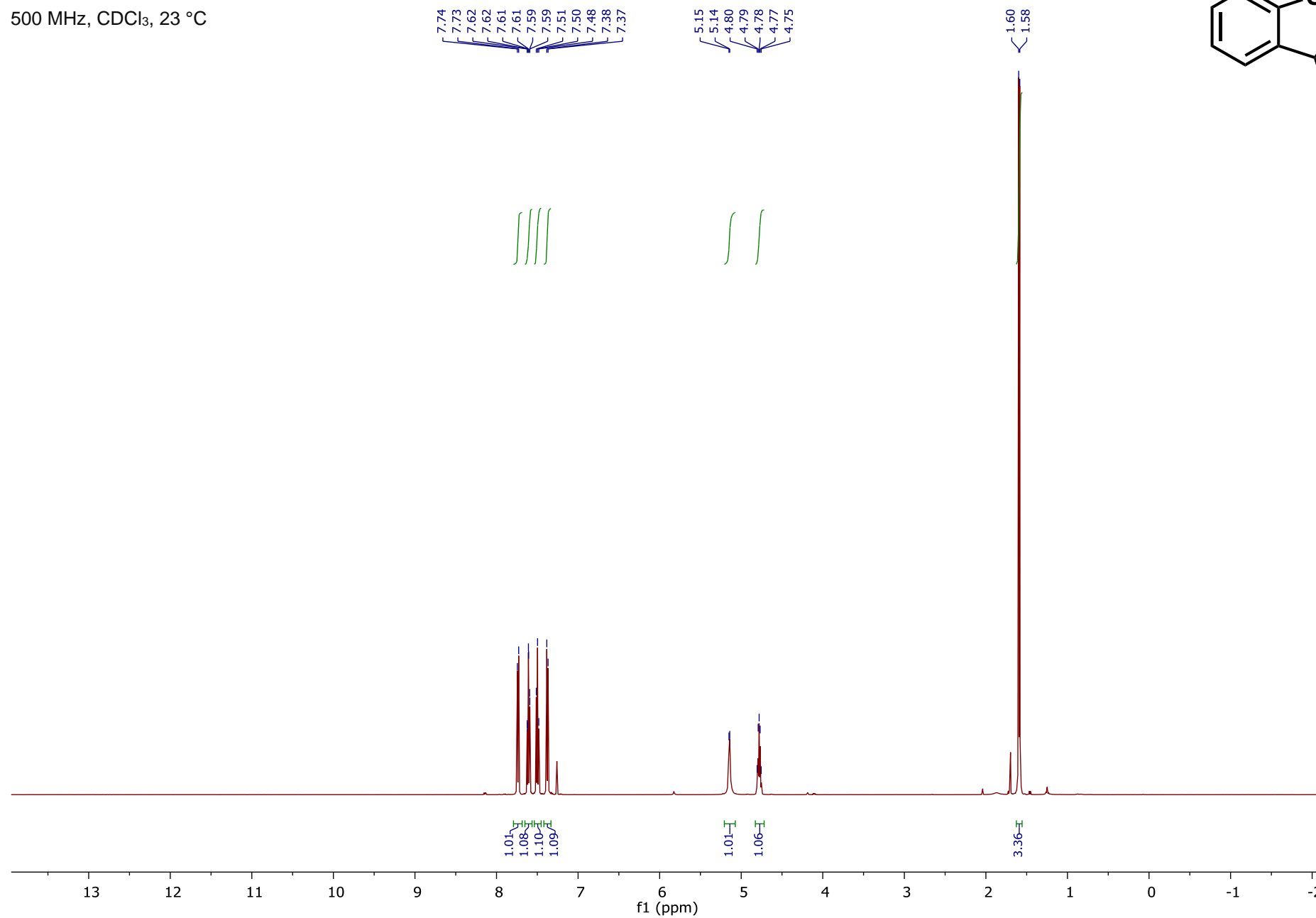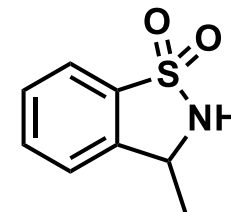
^{13}C NMR spectrum of 3-Methyl-2,3-dihydrobenzo[*d*]isothiazole 1,1-dioxide (*rac*-5)
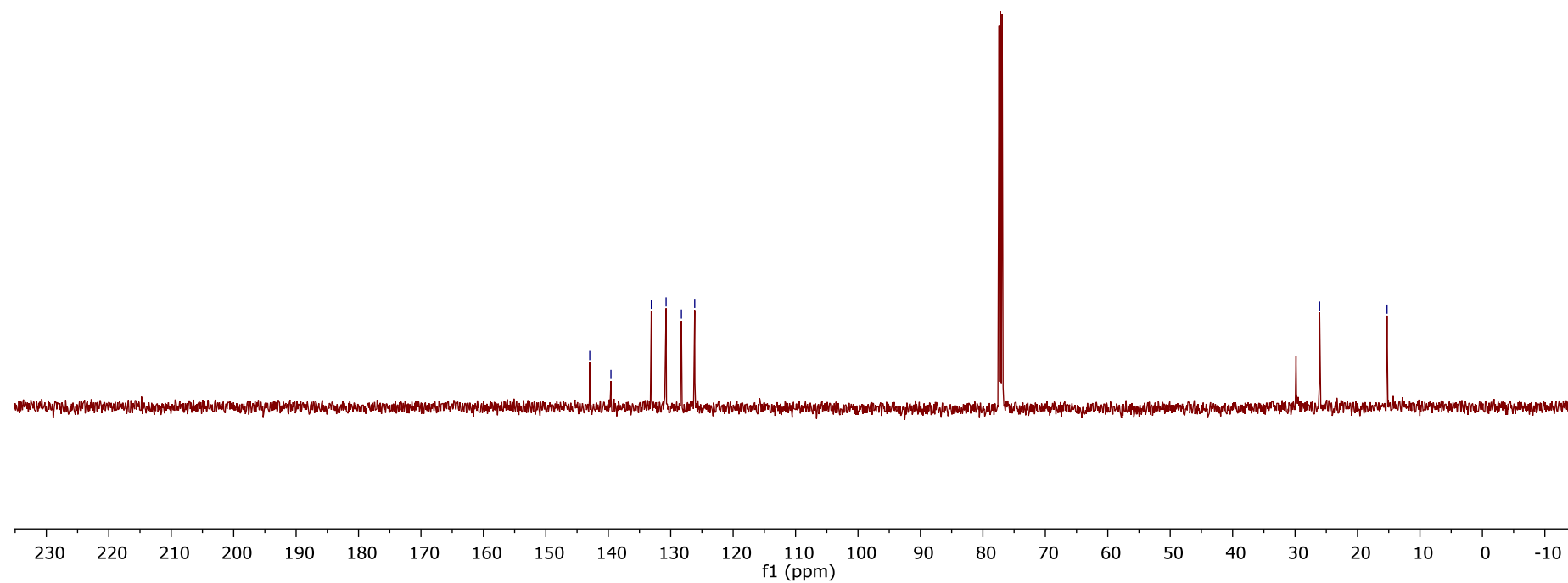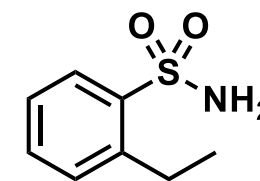
125 MHz, CDCl$_3$, 23 °C
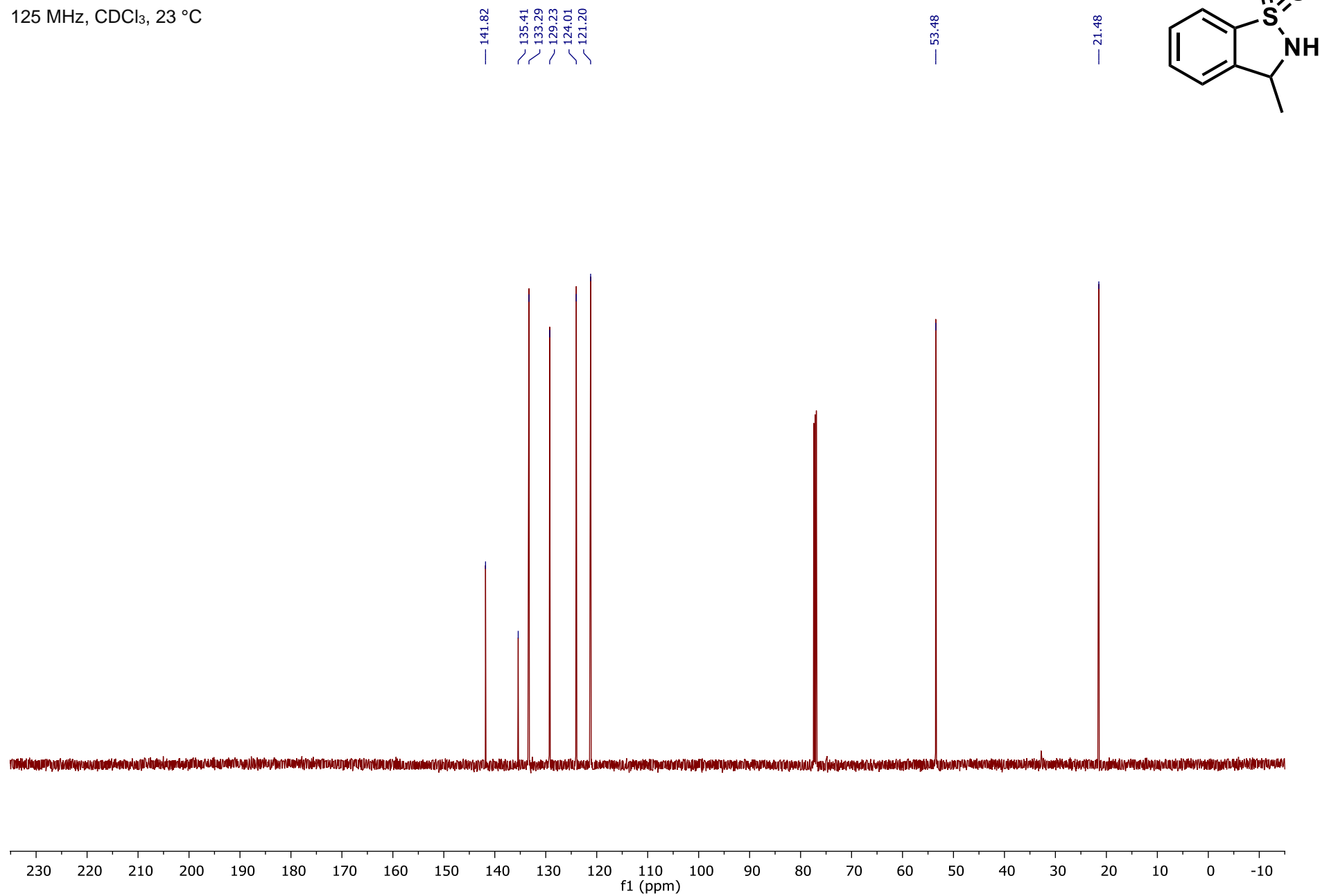
¹H NMR spectrum of 2-Ethylbenzenesulfonamide (6)

500 MHz, CDCl₃, 23 °C

^{13}C NMR spectrum of 2-Ethylbenzenesulfonamide (6)

125 MHz, CDCl$_3$, 23 °C