Supplementary Information

Biocatalytic Carbene Transfer Using Diazirines

Nicholas J. Porter[1], Emma Danelius[2,3], Tamir Gonen[2,3,4], & Frances H. Arnold[1]*

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, MC 210-41, Pasadena, California 91125, USA

[2]Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, CA 90095

[3]Department of Biological Chemistry, University of California Los Angeles, 615 Charles E.Young Drive South, Los Angeles, CA 90095

[4]Department of Physiology, University of California Los Angeles, 615 Charles E. Young Drive South, Los Angeles, CA 90095

* Author to whom correspondence should be addressed: frances@cheme.caltech.edu

Table of Contents:

Materials and Methods

General

Chemicals. Unless otherwise noted, chemicals and reagents used in synthetic reactions and buffers were obtained from commercial suppliers (Fischer Scientific, VWR, Millipore Sigma, Combiblocks, TCI America) and were used without further purification.

Instrumentation. In the synthesis of substrates and authentic standards, preparative flash chromatography was performed using a Biotage Isolera automated chromatography instrument using hand-packed silica gel columns (230–400 mesh, Silicycle Inc.). NMR spectra were recorded on a Bruker Avance 400 spectrometer equipped with a cryoprobe operating at 400 MHz and 101 MHz for ^{1}H and ^{13}C, respectively, or a Varian Mercury 300 spectrometer operating at 282 MHz for ^{19}F. NMR data were analyzed in MestReNova (MestreLab Research). Chemical shifts are reported in ppm (δ) with solvent resonance serving 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: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, m = multiplet, bs = broad singlet), coupling constant (Hz), and integration. High-resolution mass spectra were obtained at the California Institute of Technology Mass Spectral Facility. GC-FID data were collected on an Agilent 7820A GC system. GC-MS data were collected on a Shimadzu GCMS-QP2010 SE. For screening, HPLC-UV data were collected on an Agilent 1200 series HPLC equipped with an Agilent C18 column (Poroshell 120 ESC18, 4.6 x 50 mm, 2.7 μm packing). For chiral separations, either normal-phase HPLC-UV or chiral supercritical fluid chromatography were used. For normal-phase HPLC-UV, data were collected on a Hewlett Packard 1100 series HPLC equipped with a Daicel Chiralpak IB column (4.6 x 250 mm, 5 μm packing). For supercritical fluid chromatography, data were collected on a Thar Analytical SFC equipped with a Daicel Chiralpak OJ-H column (4.6 x 250 mm, 5 μm packing). UV-Vis data were collected using a Shimadzu UV-1800 Spectrophotometer.

Experimental Details

Materials. Oligonucleotides were obtained from Integrated DNA Technologies (IDT DNA). Unless otherwise stated, PCRs were run using the Phusion® High Fidelity PCR Kit (New England Biolabs). Gibson assembly mix[1] was prepared by combining isothermal master mix and enzymes T5 exonuclease, Phusion® DNA polymerase, and Taq DNA ligase (New England

Biolabs). For all biocatalytic reactions, a buffer control was also prepared in which the same procedure is applied to only the buffer used for the reaction.

Initial B–H insertion testing with BOR^{P*}. From an agar plate bearing BL21(DE3) cells transformed with *Rhodothermus marinus* cytochrome *c* variant BOR^{P*}, a single colony was selected with a sterile toothpick and inoculated into 5 mL of Lysogeny broth (LB) supplemented with 100 μg/mL carbenicillin (carb) and 20 μg/mL chloramphenicol (chlor). This culture was grown at 37 °C and 250 rpm in an Innova 4000 benchtop shaker overnight (18 hours). From this starter culture, 500 μL were used to inoculate 25 mL of Hyper Broth (HB) supplemented with 100 μg/mL carb and 20 μg/mL chloramphenicol (chlor) in a 125 mL Erlenmeyer flask. This expression culture was grown for ~2.5 hours, at which point the culture reached an OD_{600} of 1.0. The flask was transferred to ice for 15 minutes, after which the culture was supplemented with 0.02 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2 M 5-aminolevulinic acid (ALA) to induce expression. Expression was carried out at 22 °C and 220 rpm for 20 hours. The cells were then pelleted by centrifugation at 4000x*g* in an Allegra 25R tabletop centrifuge equipped with an S5700 swinging bucket rotor. The pellet was then resuspended in 2 mL of nitrogen-free M9 minimal medium (47.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.6 mM NaCl, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$; abbreviated as M9-N) and the OD_{600} was measured by 200-fold dilution. Samples were then prepared of cells with OD_{600} measurements of 60, 30 and 15 to test whether the observation of product changes with increasing amounts of biocatalyst.

Alongside these whole-cell suspensions, separate samples were prepared containing M9-N buffer alone, 100 μM porcine hemin (Sigma), 0.75 mg/mL bovine serum albumin (BSA), or 100 μM hemin supplemented with 0.75 mg/mL BSA. In separate GC vials, 380 μL of the sample solution were supplemented with 20 μL of 200 mM *N*-heterocyclic borane **1** and 200 mM CF_3-phenyl-3*H*-diazirine **4** in acetonitrile. Vials were sealed and incubated in an anaerobic Coy chamber shaking at 600 rpm for 24 hours. After the reactions were complete, 600 μL of 1 mM 1,2,3-trimethoxybenzene in 1:1 hexanes:ethyl acetate were added to each vial and samples were vortexed and then separated by centrifugation at 14,000x*g* for 5 minutes. From each sample, 200 μL of the organic layer were then transferred to a separate GC vail equipped with a 0.4-mL glass GC vial insert and sealed with screw caps. Samples were analyzed using a GC-MS equipped with an Agilent HP5-MS column (1 μL injection; split ratio = 50; 1 min @ 90 °C, ramp to 300 °C at 50 °C/min, hold at 300 °C for 3 min). The standard peak (RT = 2.33) and product peak (RT = 4.00) were integrated using built-in analysis software and the product was verified by comparison with previously synthesized material.[2] Results are presented in Table S1.

Table S1. Relative product/standard ratios of samples tested for B–H insertion of borane **1** using diazirine **4** as a carbene source. All values are normalized to the amount of product observed in the buffer control which is presumed to be approximately related to the amount of product generated by the GC-MS during analysis by thermolysis in the instrument. Of the samples tested, only porcine hemin displays significant activity over background and no change in the abundance of product **3** is observed with increasing OD_{600} of BOR^{P*}-harboring cells.

Sample		Normalized Product/Standard Ratio
M9-N buffer		1.0 ± 0.6
100 μM hemin		6.8 ± 0.4
0.75 mg/mL BSA		2.4 ± 0.3
100 μM hemin + 0.75 mg/mL BSA		2.8 ± 0.5
E. coli cells harboring BOR^{P*}	OD_{600} = 15	1.9 ± 0.7
	OD_{600} = 30	2.1 ± 0.8
	OD_{600} = 60	2.0 ± 0.5

Enzyme discovery. Glycerol stock plates bearing *E. coli* harboring genes for 60 $P450_{BM3}$ variants, 60 P411 variants, 46 globin variants, and 25 *Rhodothermus marinus* cytochrome *c* variants from lineages previously engineered for various non-natural reactions were used to inoculate starter cultures in 96-well deep-well plates containing 400 µL of LB-carb. Starter cultures were grown at 37 °C, 220 rpm, and 80% humidity in a Multitron Infors shaker overnight (16–18 hours). From these starter cultures, 50 µL were used to inoculate expression cultures (1 mL HB-carb) in fresh, sterile 96-well deep-well plates. For cytochrome *c* variants, the wells were additionally supplemented with 50 µg/mL of chloramphenicol. These cultures were grown at 37 °C, 220 rpm, and 80% humidity in a Multitron Infors shaker for 3 hours and then transferred onto ice for 30 minutes. Expression was then induced for the P450s, P411s, and globins using 0.5 mM IPTG and 1.0 mM ALA and for the cytochromes *c* using 0.02 mM IPTG and 0.2 mM ALA. Expression was carried out at 22 °C, 220 rpm overnight (18–20 hours). Plates were centrifuged at 5,000x*g* for 5 minutes in an Allegra 25R tabletop centrifuge (Beckman-Coulter) equipped with a S5700 swinging bucket rotor, and the supernatant was decanted. Pellets were brought into a Coy anaerobic chamber (~0–10 ppm O_2) along with M9-N sparged with argon, as well as 400 mM NHC-borane and 400 mM 3-trifluoromethyl-3-phenyl-diazirine, each in acetonitrile. Cell pellets were resuspended in 380 µL of M9-N by vortexing using a Vortex Genie 2 (VWR), and then cell suspensions were supplemented with 10 µL of each substrate ($[substrates]_{final}$ = 10 mM). Reactions were sealed with adhesive aluminum seals and incubated in the Coy chamber on an orbital shaker at 600 rpm for 20 hours. Plates were then removed from the Coy chamber and 610 µL of 0.67 mM 1,2,3-trimethoxybenzene in 4:6 hexanes:ethyl acetate were applied to each well. Plates were sealed, shaken, and vortexed, then centrifuged at 5,000x*g* for 5 minutes. Using a 12-channel 200 µL pipette, 200 µL of each organic layer were transferred to 400 µL GC vial inserts, which were subsequently transferred to GC vials and sealed with screw caps. Samples were analyzed using a GC-MS equipped with an Agilent HP5-MS column (1 µL injection; split ratio = 50; 1 min @ 90 °C, ramp to 300 °C at 50 °C/min, hold at 300 °C for 3 min). The standard peak (RT = 2.33) and product peak (RT = 4.00) were integrated using built-in analysis software.

Cloning, expression, and purification of *Ape*Pgb variants. The gene encoding *Aeropyrum pernix* protoglobin Y60G F145Q (*Ape*Pgb GQ) was previously cloned into the pET22b(+) vector (Novagen) with a C-terminal 6xHisTag.[3] For crystallography, the sequence of *Ape*Pgb C45G W59L Y60V V63R C102S F145Q I149L (GLVRSQL) was cloned into a pSUMO vector (kanamycin resistance) for the expression of a construct bearing an N-terminal 6xHisTag

and an Ulp1-cleavable SUMO tag followed by the *Ape*Pgb GLVRSQL sequence bearing no purification tag (pSUMO-GLVRSQL). Electrocompetent *E. cloni* EXPRESS BL21(DE3) cells (Lucigen) were transformed with this plasmid using a MicroPulser Electroporator (Bio-Rad). Following electroporation, transformed cells were immediately supplemented with 950 μL of SOC medium and plated on LB-carb (100 μg/mL) or LB-kan (50 μg/mL) agar plates with no outgrowth. Plates were incubated at 30 °C for 16–18 hours and then stored at 4 °C for up to two weeks. To prepare overnight cultures, 5 mL LB-carb were inoculated with single colonies and grown at 37 °C and 220 rpm for 14–18 hours. Overnight cultures were used to isolate plasmids using a Qiagen Miniprep kit, prepare 25% glycerol cell stocks for storage at -80 °C, and inoculate expression cultures. Sanger sequencing was performed on isolated plasmids by Laragen, Inc. (Culver City, CA) using the T7 promoter primer.

For protein expression, 250 mL or 1000 mL of appropriate media supplemented with 100 μg/mL carbenicillin or 50 μg/mL kanamycin in a 500-mL Erlenmeyer flask or a 2.8-L Fernbach flask, respectively, were inoculated with 1–5 mL of starter cultures. To express enzyme with *in vivo* heme occupancy levels (20–30%), Hyper Broth (HB-carb) was used. To express enzyme to which exogenous heme would be loaded in purification, Terrific Broth (TB-carb) was used. Cultures were then grown at 37 °C and 180 rpm. When the culture had grown to OD_{600} = 0.8–1.0, the flask was transferred to ice for 30–45 minutes. For cultures grown in HB, expression was induced with 0.5 m IPTG and 1.0 mM ALA. For cultures grown in TB, expression was induced with only 1.0 mM IPTG. Following induction, expression was carried out at 22 °C and 180 rpm for 18–20 hours. Cells were then pelleted by centrifugation at 5,000x*g* in an Allegra 25R tabletop centrifuge equipped with a TS-5.1-500 swinging bucket rotor and stored in 50-mL conical tubes at -20 °C for at least 24 hours prior to purification.

For purification of protein for small-scale reactions, cell pellets were removed from -20 °C and thawed in room temperature water. Once the pellet was thawed, the cells were resuspended in binding buffer (50 mM tris(hydroxymethyl)aminomethane (Tris, pH 8.0), 150 mM NaCl, 20 mM imidazole; ~5 mL/g of cells) either by pipetting or vortexing and then lysed by sonication (QSonica Q500 sonicator, 25% amplitude, 33% duty cycle, 5 minutes). Lysate was clarified by centrifugation at 13,000x*g* in an Avanti J-25 floor centrifuge equipped with a JA-12 fixed angle rotor followed by filtration (0.45 μm syringe filter). The protein was then purified using an Äkta Purifier equipped with a 5-mL HisTrap HP column (Cytiva), eluting with a gradient of 20–500 mM imidazole using elution buffer (50 mM Tris buffer (pH 8.0), 150 mM NaCl, 500 mM imidazole). Fractions containing the protein on interest were pooled and concentrated 10-

fold by centrifugal filtration (Amicon Ultra-15 30 kDa MWCO). Protein concentration was determined using the Pierce BCA Protein Assay kit. For protein expressed in TB with no ALA present, the heme concentration was determined using the hemochromagen assay.[3] Generally, the heme concentration was 1–15% of the protein concentration. Free heme (porcine, Sigma; 10 mM stock in 0.1 M NaOH) was then used to supplement the protein with additional heme equivalent to the difference between the determined protein and heme concentrations. Upon the addition of the heme solution, the protein solution quickly turned from brown to dark red, indicating heme binding to the enzyme. The protein solution was then transferred into dialysis tubing and dialyzed at room temperature against 4 L of storage buffer (50 mM potassium phosphate buffer (pH 8.0), 150 mM NaCl) for 16–20 hours. Protein was then concentrated to 1–2 mM (23–46 mg/mL), flash frozen by dripping into liquid nitrogen, and stored in a conical tube at -80 °C for future use.

For the purification of protein for crystallography, cell pellets from HB expression of pSUMO-GLVRSQL were removed from -20 °C and thawed in room temperature water. Once thawed, cell pellets were resuspended in lysis buffer (50 mM potassium phosphate (pH 8.0), 150 mM NaCl, 200 µM porcine hemin, 2 mM $MgCl_2$, 0.75 mg/mL lysozyme, ~0.1 mg/mL DNaseI). Prior to pellet resuspension, this buffer is murky due to partial binding of the hemin by lysozyme or DNaseI, but this does not impede lysis performance. Cell suspension was incubated at 37 °C for 2 hours with intermittent mixing by inversion, after which cell debris was pelleted by centrifugation at 13,000x*g* in an Avanti J-25 floor centrifuge equipped with a JA-12 fixed angle rotor. The dark red lysate was then subjected to HisTrap purification using the same buffers as for the purification of enzyme for reactions above. Following isolation of pSUMO-GLVRSQL, the protein was concentrated to 2.5 mL in a 30,000 MWCO centrifugal concentrator and passed over a PD-10 column (Cytiva) to exchange the enzyme into proteolysis buffer (50 mM tris(hydroxymethyl)aminomethane (Tris, pH 8.0), 150 mM NaCl, 5 mM DL-dithiothreitol (DTT)). The protein was then supplemented with Ulp1 protease and digested overnight on a benchtop rocker (6/min) at 4 °C. The protein was then passed over Ni-NTA resin (Qiagen) equilibrated with storage buffer (50 mM potassium phosphate (pH 8.0), 150 mM NaCl) to remove the cleaved SUMO tag. The desired red protein was then concentrated to >20 mg/mL using a 10,000 MWCO centrifugal concentrator and stored in 40 µL aliquots at -80 °C prior to use.

Site-saturation mutagenesis. Site-saturation mutagenesis (SSM) was performed using primers bearing NNK degenerate codons and was utilized for rounds 1–6 of directed evolution

(DE) in this work. Oligonucleotides containing NNK codons for the desired positions as well as within the ampicillin resistance cassette of the pET22b(+) plasmid were used to amplify libraries in two pieces containing overlaps in the *Ape*Pgb gene at the site of mutation and in the β-lactamase gene of the ampicillin resistance cassette. Using this method, incorrect assembly results in a product that fails to confer ampicillin/carbenicillin resistance. The oligonucleotides used to generate variants reported in this work are presented in Table S2. Following a 1-hour digestion with DpnI (New England Biolabs) at 37 °C, PCR products were applied alongside a 1 kb DNA ladder (New England Biolabs) to 1% agarose gels containing 1X SYBR Gold nucleic acid stain (Thermo Fisher) and were visualized using a blue transilluminator. PCR products were excised from the gel and the DNA fragments were extracted using the Zymoclean Gel DNA recovery kit. Isothermal Gibson assembly (50 °C, 1 hour) was used to prepare the assembled library. Gibson products were then purified using the QIAquick PCR Purification Kit (QIAGEN) and used to transform 25–50 µL aliquots of electrocompetent *E. cloni* EXPRESS BL21(DE3) cells. Transformed cells were supplemented with 950 µL SOC medium and immediately plated on LB-carbenicillin (100 mg/mL) agar plates. Plates were incubated at 30 °C for 16–18 hours and then stored at 4 °C for up to two weeks.

Table S2. Oligonucleotides used for site-saturation mutagenesis (SSM). Mutated codons are shown in blue. The ampR forward and reverse oligonucleotides were used in each round of SSM to amplify the plasmid in two separate pieces for subsequent Gibson assembly. Fragments containing mutations were amplified using the forward primer containing the codon library and the ampR reverse primer while the backbone fragment was amplified using the reverse primer for each library and the ampR forward primer. Varied nucleobases are represented as N (A, T, C, G) and K (G or T).

Mutations relative to *Ape*Pgb GQ	Direction	Sequence (5' ⇒ 3')
C45X	forward	GACGTAATGTACTTGAAAAAGGCG**NNK**GACGTTCTGAAAG
C45X	reverse	CGCCTTTTTCAAGTACATTACGTCCTTTTCACCTAACATGAC
W59X	forward	TTGATGAGATCCTTGACTTG**NNK**GTTGGTTGGGTAGCATC
W59X	reverse	CAAGTCAAGGATCTCATCAACTTGATCTTTCAGAACGTC
Y60X	forward	TTGACTTGTGG**NNK**GGTTGGGTAGCATCAAATGAGC
Y60X	reverse	CCACAAGTCAAGGATCTCATCAACTTGATC
V63X	forward	CTTGACTTGCTGGTTGGTTGG**NNK**GCATCAAATGAG
V63X	reverse	CCAACCAACCAGCAAGTCAAGGATCTCATCAACTTGATC
C102X	forward	CCTGGATTCTGGACACTACC**NNK**CGCGACTATAAC
C102X	reverse	GGTAGTGTCCAGAATCCAGGCTCCAAAG
I149X	forward	GTTATCTTATCGCACAGATCTATCCT**NNK**ACCGCCACTATCAAG
I149X	reverse	AGGATAGATCTGTGCGATAAGATAACGAAGTGGGATATGGG
ampR	forward	CCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACC GCTTTTTTGC
ampR	reverse	CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATG GTTATGGCAG

Error-prone mutagenesis. Error-prone mutagenesis was used in round 7 of DE in this work. Oligonucleotides binding at either end of the protoglobin encoding sequence were used to amplify the gene:

- forward: 5'–CTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG–3'
- reverse: 5'–GCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAG–3'

PCRs were run using Taq polymerase (New England Biolabs) in the presence of 200, 300, 400, and 500 μM $MnCl_2$. The pET22b(+) backbone was also amplified using the reverse complementary primers with Phusion DNA polymerase (New England Biolabs). PCR products were isolated and assembled in the same manner as SSM libraries, and the resultant Gibson products were similarly purified and used to transform electrocompetent *E. coli* cells and plated on LB-carb/agar.

Recombination of beneficial mutations. Following the identification of potentially beneficial mutations by error-prone mutagenesis, oligonucleotides were designed to create a library containing all combinations of these mutations as well as the residue present at each position in the parent enzyme (Table S3). Due to the proximity of these sites in the nucleotide sequence, two rounds of Splicing by Overlap Extension (SOE) PCR were required to ensure generated fragments were long enough to be reasonably purified. Briefly, in the first round, primers corresponding to each end of the protoglobin-encoding gene as well as the forward and reverse primers installing mutations at positions 79 and 165 were used in PCRs to amplify short fragments comprising the entire gene with short overlaps between each fragment. These fragments were then mixed and subjected to another PCR to stitch together the overlaps between these fragments, generating a library of genes encoding the full protoglobin with variation at some of the desired sites. This procedure was then repeated, installing mutations at sites 60, 61, 123, and 175, using the PCR product of the first round as the template to produce a gene library bearing variation at all desired positions. The final gene library encoding the combinatorial library was then ligated into the pET22b(+) backbone using Gibson assembly, and the resultant Gibson products were purified and used to transform electrocompetent *E. coli* cells and plated on LB-carb/agar.

Table S3. Oligonucleotides used for recombination of mutants discovered through error-prone mutagenesis. Mutated codons are shown in blue. If other varied codons are present on the same primer, they are bolded and underlined. Primers were used in two iterative rounds of SOE PCR to generate a library containing all 64 possible combinations of these mutants. Varied nucleobases are represented as Y (C or T), K (G or T), S (G or C), and R (A or G).

Mutation site in *Ape*Pgb	Amino Acids Screened	Sequence (5' ⇒ 3')
60	**Val** / Ala	GATCCTTGACTTGCTG**GYT**__**GKT**__TGGCGGGCATCAAATG
61	**Gly** / Val	GATCCTTGACTTGCTG__**GYT**__**GKT**__TGGCGGGCATCAAATG
79	**Gly** / Arg	CAATCCGGATACA**SGA**GAGCCTATTAAGGAATAC
123	**Ser** / Pro	CTTCGTCATCACCGT**YCA**AAGAAAGGGGTCACAG
165	**Glu** / Gly	GAAAGGTGGCTCTCCG**GRA**GACATCGAAGGGATGTAC
175	**Phe** / Leu	GATGTACAACGCTTGG**YTC**AAGTCTGTAGTTTTAC

Library screening. From agar plates bearing BL21(DE3) cells transformed with *Ape*Pgb variant libraries, isolated single colonies were picked using sterile toothpicks and used to inoculate starter cultures (0.4 mL LB-carb) in sterile 96-well deep-well plates. Four wells on each plate were inoculated with only a sterile toothpick (sterile control) and an additional four wells on each plate were inoculated with the parent variant (parent controls). Variants were screened in either half-plate (44 picked colonies; SSM libraries only) or full-plate (88 picked colonies) arrays. These starter culture plates were sealed with air-permeable tape and grown at 37 °C, 250 rpm, and 80% humidity in a Multitron Infors shaker overnight (16–18 hours). From these starter cultures, 50 µL were used to inoculate expression cultures (1 mL HB-carb) in fresh sterile 96-well deep-well plates. Starter culture plates were then stored at 4 °C. Expression cultures were grown at 37 °C, 250 rpm, and 80% humidity for three hours, then cooled on ice for 45 minutes. Protein expression was induced by adding 0.5 mM IPTG and 1.0 mM ALA and carried out at 22 °C and 220 rpm overnight (18–20 hours). Cells were pelleted (5,000x*g*, 5 minutes), the supernatant was removed, and pellets were used for screening the desired reaction as described below.

Screening for B–H insertion. Immediately after centrifugation, cell pellets were transferred into a Coy anaerobic chamber (~0–10 ppm O_2) along with nitrogen-free M9 minimal medium (47.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.6 mM NaCl, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$; abbreviated as M9-N) sparged with argon, and substrate mix (200 mM 3-CF_3-3-phenyl-diazirine and 200 mM *N*-heterocyclic carbene borane in acetonitrile). Pellets

were then resuspended in 380 µL of M9-N buffer using a Vortex Genie 2 benchtop vortexer (VWR). Cell suspensions were supplemented with 20 µL of substrate mix ($[substrates]_{final}$ = 10 mM; 5% acetonitrile). The plates were sealed with an adhesive aluminum foil seal and incubated in the Coy chamber on an orbital shaker at 600 rpm overnight (20–24 hours). For work-up and analysis, the plates were removed from the Coy chamber and 600 µL of 1 mM 1,2,3-trimethoxybenzene in 1:1 hexanes:ethyl acetate were added to each well. The plates were vigorously shaken, vortexed, and then centrifuged at 5,000x*g* for 10 minutes. Using a 12-channel 200 µL pipet, 200 µL of the organic layer were removed from each well and transferred to a separate 0.4-mL glass GC vial insert. These inserts were loaded into Agilent GC vials and sealed with screw caps. Samples were analyzed using a GC-MS equipped with an Agilent HP5-MS column (1 µL injection; split ratio = 50; 1 min @ 90 °C, ramp to 300 °C at 50 °C/min, hold at 300 °C for 3 min). The standard peak (RT = 2.33) and product peak (RT = 4.00) were integrated using built-in analysis software, and the product was verified by comparison with previously synthesized material.[4] Wells displaying increases in yield were validated by resuspending cells from a 50-mL expression culture to OD_{600} = 60 in M9-N and screening this cell suspension alongside the parent in an identical manner.

Screening for benzyl acrylate cyclopropanation. Immediately after centrifugation, cell pellets were transferred into a Coy anaerobic chamber (~0–10 ppm O_2) along with nitrogen-free M9 minimal medium (47.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.6 mM NaCl, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$; abbreviated as M9-N) sparged with argon and substrate mix (200 mM 3-phenyl-3-*H*-diazirine and 400 mM benzyl acrylate in acetonitrile). Pellets were then resuspended in 380 µL of M9-N buffer using a Vortex Genie 2 benchtop vortexer (VWR). Cell suspensions were supplemented with 20 µL of substrate mix ($[diazirine]_{final}$ = 10 mM; $[acrylate]_{final}$ = 10 mM, 5% acetonitrile). The plates were sealed with an adhesive aluminum foil seal and incubated in the Coy chamber on an orbital shaker at 600 rpm overnight (20–24 hours). For work-up and analysis, the plates were removed from the Coy chamber and 600 µL of 1.67 mM 4-ethyl-anisole in acetonitrile were added to each well. The plates were vigorously shaken, vortexed, and then centrifuged at 5,000x*g* for 10 minutes. From each well, 200 µL of the supernatant were transferred to individual wells of an AcroPrep™ filter plate (PALL) fitted to a polypropylene 96-well microtiter plate (Agilent) and the samples were filtered by centrifugation at 1,500x*g* and plates were sealed with NAL-96 pierceable film (USA Scientific). Samples were analyzed by HPLC-UV using an isocratic method (68% acetonitrile in ddH_2O; 2-5 µL injections). Peak areas for the ethyl anisole

standard (RT = 0.90 min), as well as the *cis*- and *trans*-cyclopropane products (RT_{cis} = 1.42 min, RT_{trans} = 1.65 min) were obtained using the built-in ChemStation analysis software (Agilent).

Following screening, wells displaying improved product yield relative to parent controls were streaked from the corresponding starter culture onto LB-carb agar plates. A single colony was picked and grown in 5 mL LB-carb overnight (220 rpm, 37 °C). These overnight cultures were used in flask protein expression (50 mL HB-carb) and small-scale biocatalytic reactions to verify enhanced activity, as well as for preparation of variant plasmids using the Qiagen Miniprep Kit for Sanger sequencing by Laragen, Inc. Small-scale reactions were set up in 2-mL GC vials. *E. coli* expressing the appropriate *Ape*Pgb variant (380 µL, suspended in M9-N to an OD_{600} of 30 or 60) were added to the vials, and these vials were transferred into a Coy anaerobic chamber (~0–10 ppm O_2). To each vial, the appropriate substrate mix for the desired reaction was used in screening. Following reaction incubation in the Coy chamber, these reactions were worked up and analyzed as in screening. Variants displaying increased yield in these validation reactions were carried forward in evolution.

Comparison of diazo- and diazirine-based carbene transfer activity. From agar plates bearing BL21 (DE3) *E. coli* cells transformed with plasmids encoding *Ape*Pgb Y60Q F145Q (GQ), Y60V F145Q (VQ), and W59L Y60V F145Q (LVQ), single colonies were picked and grown in 5 mL LB-carb overnight (220 rpm, 37 °C). These variants were expressed in 50 mL HB-carb and small-scale biocatalytic reactions were set up using whole cell suspensions at OD_{600} = 60. Reactions were set up in triplicate with substrate concentrations of 10 mM borane **1** and 10 mM carbene donor (either diazo **2** or diazirine **4**). After 24 hours, the reaction was extracted with 1 mM 1,2,3-trimethoxybenzene in 1:1 EtOAc:hexanes and analyzed on an Agilent 7820A GC-FID system equipped with an Agilent HP-5 GC column (1 µL injection; 2 min @ 90 °C, ramp to 320 °C at 30 °C/min, hold at 320 °C for 2 min). Yields of organoborane **3** were determined using an experimentally determined calibration curve and correcting for dilution of the product upon extraction from the whole cell suspension.

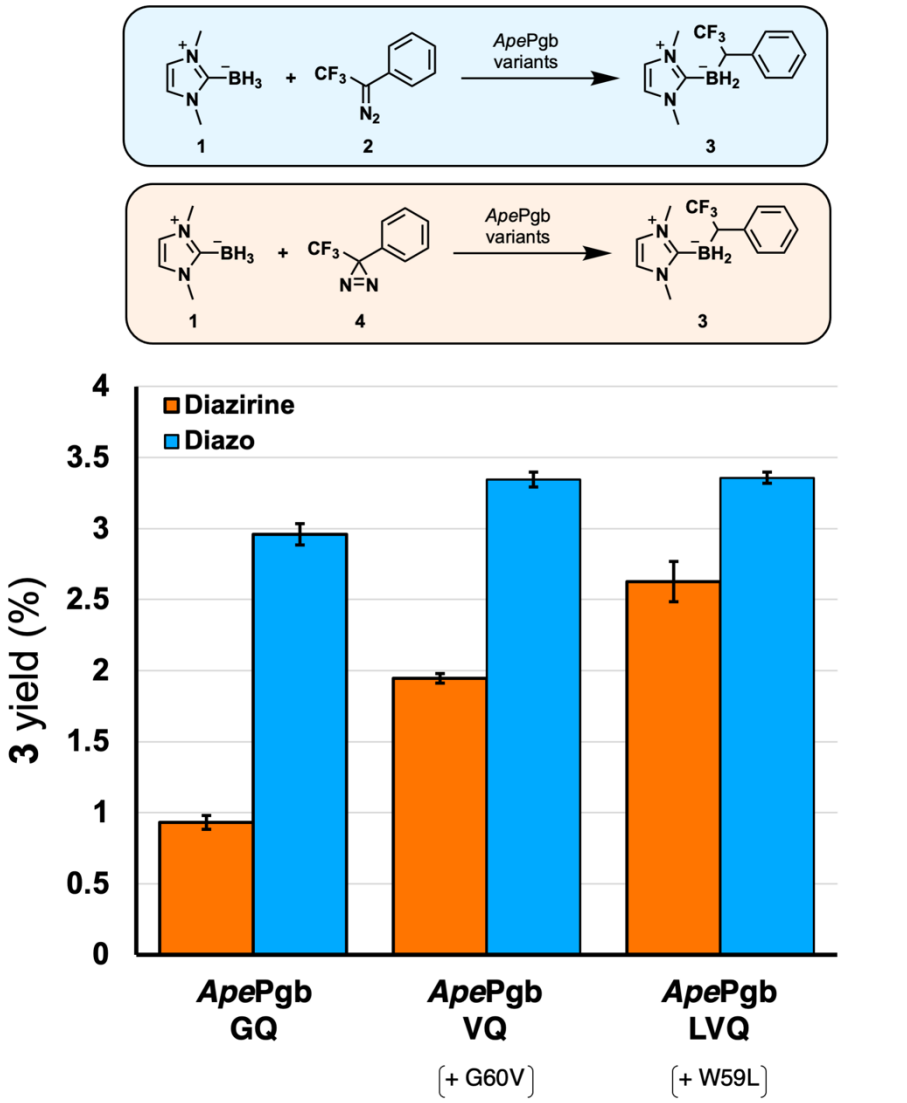


Figure S1. Yield of organoborane **3** using either diazo compound **2** (blue) or diazirine **4** (orange) as a carbene source in reactions using whole cell suspensions of each variant at OD_{600} = 60. Over the course of directed evolution, activity using **2** exhibits minimal improvement (13% increase in yield between *Ape*Pgb GQ and LVQ) while activity using **4** increases 3-fold between these same variants.

Hemochromagen assay. The hemochromagen assay was performed as previously described.[4] Briefly, 500 µL of lysate or protein solution were added to a plastic cuvette. A 500 µL aliquot of solution I (0.2 M NaOH, 40% pyridine, 0.5 mM potassium ferricyanide) was then added, and the spectrum of this oxidized sample was collected across 380–800 nm. Sodium dithionite (10 µL of 0.5 M solution in 1 M KPi at pH 8.0) was added, and the reduced spectrum was collected across 380–800 nm. The pyridine hemochromagen concentration was determined using its Q bands, using the previously determined extinction coefficient 23.98 mM^{-1} cm^{-1} for (557 $nm_{reduced}$ – 540 $nm_{oxidized}$).

Purified protein reactions. Small-scale reactions were set up in 2-mL GC vials. Purified *Ape*Pgb GLAVRSQLL was thawed and diluted to the desired concentration with storage buffer (50 mM potassium phosphate buffer, pH 8.0, 150 mM NaCl), added to vials, and brought into the Coy chamber (~0–10 ppm O_2). To each vial, sodium dithionite was added ($[Na_2S_2O_4]_{final}$ = 2 mM), as well as the desired diazirine and carbene acceptor substrate at final concentrations of 10 mM and 20 mM, respectively, using 5% acetonitrile as a cosolvent. For reactions run in darkness, all lights were shut off and doors to other lit rooms were shut prior to the addition of reactants. Vials were then sealed in darkness and placed into an opaque black plastic box. Reactions were incubated in the Coy chamber on an orbital shaker (600 rpm) for 24 hours. For work-up and analysis, the vials were removed from the Coy chamber and 600 µL of 1 mM 1,2,3-trimethoxybenzene in 1:1 hexanes:ethyl acetate were added to each well. The combined aqueous and organic layers were transferred to 1.7-mL Eppendorf tubes, and the extraction was carried out using a Retsch MM 301 mixing mill (1 min @ 30 Hz). Samples were resolved by centrifugation (5 minutes at 20,000x*g*), and 200 µL of the organic layer were removed from each well and transferred to a separate 0.4-mL glass GC vial insert. These inserts were loaded into Agilent GC vials and sealed with screw caps. Samples were analyzed using GC-FID equipped with an Agilent HP5 GC column (1 µL injection; 2 min @ 90 °C, ramp to 320 °C at 30 °C/min, hold at 320 °C for 2 min). Peak integrations were collected using Agilent ChemStation software and yields were determined using calibration curves generated using synthesized authentic standards.

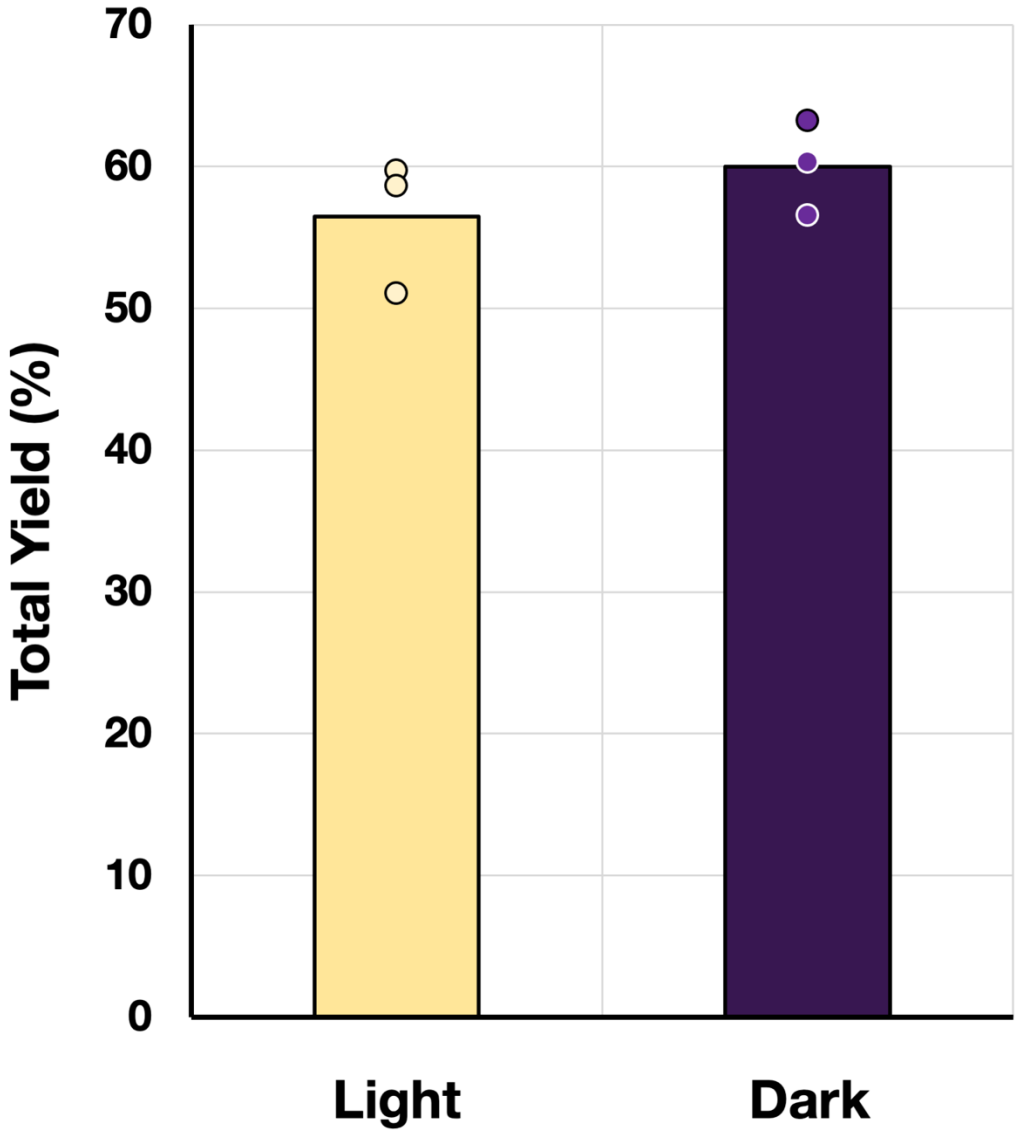


Figure S2. Product yield of *Ape*Pgb GLAVRSQLL for benzyl acrylate cyclopropanation using diazirine **6a** in ambient light and a rigorously dark control. Dark reaction was set up by adding substrates to enzyme solution with lights off and reactions were stored in an opaque black plastic box. Yields were obtained by GC-FID using calibration curves of authentic standards.

Diazo detection assay. Purified *Ape*Pgb GLAVRSQLL or hemin (porcine; 10 mM in 0.1 M NaOH) were thawed and diluted to 100 μM in storage buffer (50 mM KPi at pH 8.0, 150 mM NaCl). For heat-treated protein controls, the protein sample was incubated at 95 °C for 15 minutes resulting in a precipitated suspension that was used without clarification. For each sample, 360 μL were added to a 2-mL GC vial and these vials were brought into a Coy anaerobic chamber (~0–10 ppm O_2) along with separate GC vials containing aliquots of 400 mM 3-phenyl-3*H*-diazirine and 400 mM dibenzylcyclooctyne amine (DBCO-NH_2) in acetonitrile as well as 40 mM dithionite in 1 M potassium phosphate buffer (pH 8.0). A 20-μL aliquot of the dithionite stock was added to the GC vial containing enzyme, followed by 10 μL of diazirine solution and 10 μL of DBCO-NH_2 solution ($[enzyme]_{final}$ = 50 μM, $[dithionite]_{final}$ = 2 mM, $[diazirine]_{final}$ = 10 mM, $[DBCO\text{-}NH_2]_{final}$ = 10 mM). The reactions were sealed with a screw cap and incubated in the Coy chamber overnight (24 hours). The vials were removed from the Coy chamber, and 400 μL of acetonitrile were added to each vial. The samples were transferred to 1.7-mL Eppendorf tubes and vortexed prior to centrifugation at 14,000x*g* in an Eppendorf MiniSpin® plus benchtop centrifuge. A 200-μL aliquot of each reaction was transferred to a separate 400-μL GC vial insert and analyzed on an Agilent 1260 HPLC-MS equipped with an Agilent C18 column (Poroshell 120 ESC18, 4.6 x 50 mm, 2.7 μm packing) using acetonitrile/H_2O (0.1% v/v acetic acid): 5% to 95% acetonitrile over 3 min; 1 mL/min. Peaks for diazo adduct were compared to those observed for authentic standard using the same method.

Interpretation of results presented in Figure S2. Samples tested in this experiment are referred to as follows:

A. Buffer. Buffer (50 mM potassium phosphate (pH 8.0), 150 mM NaCl, and 2 mM dithionite)

B. Heme. Buffer + 50 μM porcine hemin

C. ΔGLAVRSQLL. Buffer + 50 μM *Ape*Pgb GLAVRSQLL that had been denatured by incubation at 95 °C for 15 minutes

D. GLAVRSQLL. Buffer + 50 μM *Ape*Pgb GLAVRSQLL as purified

In the buffer control **A**, the product was detected, likely corresponding to the slow, noncatalytic rearrangement of the diazirine to the corresponding diazo compound. In the presence of free heme (**B**), the total peak area was approximately 10-fold larger, suggesting that free heme can accelerate the rate of diazirine isomerization in solution. Relatively to the peak observed with free heme, the same concentration of folded enzyme (**D**) displayed

approximately 2-fold more of the corresponding cycloaddition product. This suggests that the macromolecular structure of the enzyme accelerates this isomerization to a degree greater than free heme alone. This is supported by the observation that denatured enzyme samples (**C**) exhibit similar product peak areas as the free heme controls. This means that aggregated polypeptides with no ordered structure around the heme cofactor do not accelerate this isomerization rate. In the folded enzyme samples, we make no assertions about whether cycloaddition product **S2** is being formed inside or outside of the enzyme active site, only that these products are observed to a greater degree when the macromolecular structure is intact. This is consistent with the enzyme catalyzing the isomerization of the diazirine, proceeding through the diazo isomer or an equivalent intermediate, as a mode of activation to the corresponding carbene.

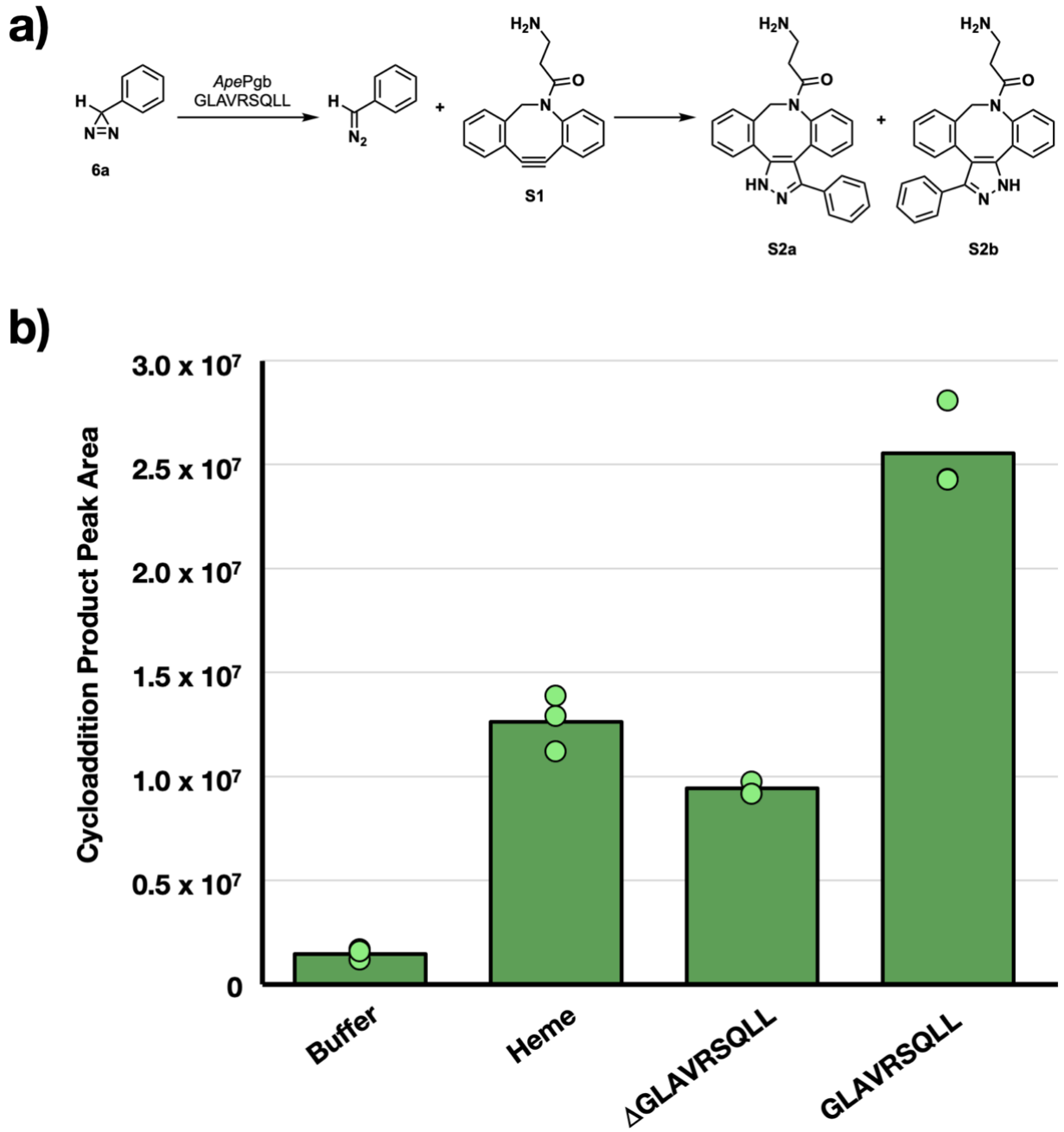


Figure S3. (a) Reaction of dibenzocyclooctyne amine (DBCO-NH_2) with phenyldiazomethane to form regioisomeric pyrazole products that can be quantified by LC-MS. **(b)** Sum of LC-MS peak areas for both regioisomers under various conditions. **Buffer** = 50 mM potassium phosphate buffer (pH 8.0), 150 mM NaCl, 2 mM dithionite, 10 mM diazirine **6a**, 10 mM DBCO-NH_2, 7.5% acetonitrile. **Heme** = Buffer + 50 μM porcine hemin. **ΔGLAVRSQLL** = 50 μM *Ape*Pgb GLAVRSQLL in buffer incubated at 95 °C for 15 minutes (unclarified). **GLAVRSQLL** = 50 μM *Ape*Pgb GLAVRSQLL in buffer.

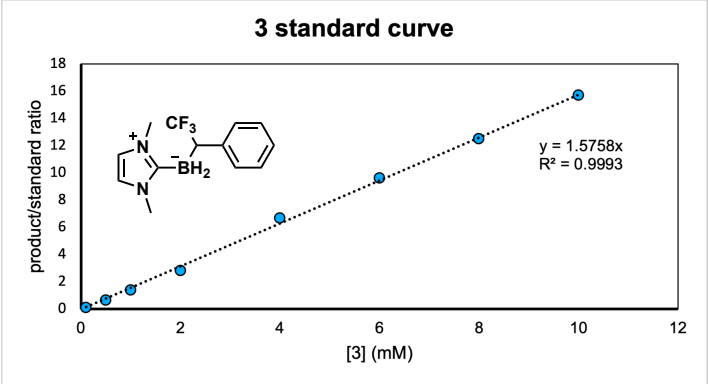
Calibration curve generation. For each tested reaction, synthetic standards were dissolved to 100 mM in 1:1 hexanes:ethyl acetate. Calibration samples were then prepared at 0.5, 1.0, 2.0, 4.0, 6.0, 8.0, and 10.0 mM in 1:1 hexanes:ethyl acetate, maintaining a standard concentration of 1 mM 1,2,3-trimethoxybenzene. Analysis was performed on an Agilent 7820A GC system equipped with an Agilent HP-5 GC column (1 µL injection; 2 min @ 90 °C, ramp to 320 °C at 30 °C/min, hold at 320 °C for 2 min). Retention times are reported for each reaction in Table S3.

Table S4. Retention times of compounds analyzed by GC-FID in calibration curve generation.

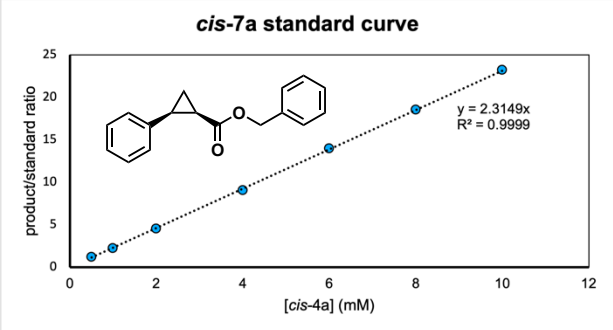
Compound	GC-FID Retention Time (min)
1,2,3-trimethoxy-benzene (standard)	4.82
3	7.76
cis-**7a**	7.68
trans-**7a**	7.90
cis-**7b**	7.63
trans-**7b**	7.86
cis-**7c**	8.25
trans-**7c**	8.53
cis-**9**	6.24
trans-**9**	6.61
11	6.65
13	6.35

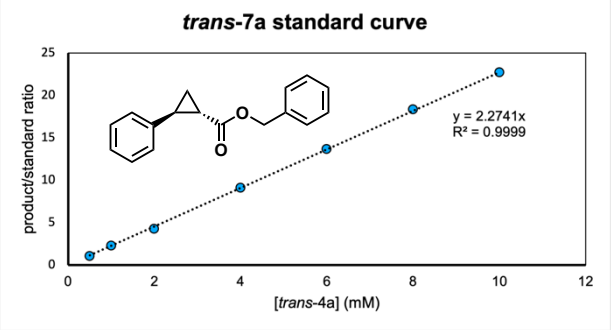
Gas Chromatography – Flame Ionization Detection (GC-FID) Standard Curves

3. *(1,2-dimethyl-1H-imidazol-3-ium-2-yl)(2,2,2-trifluoro-1-phenylethyl)dihydroborate*

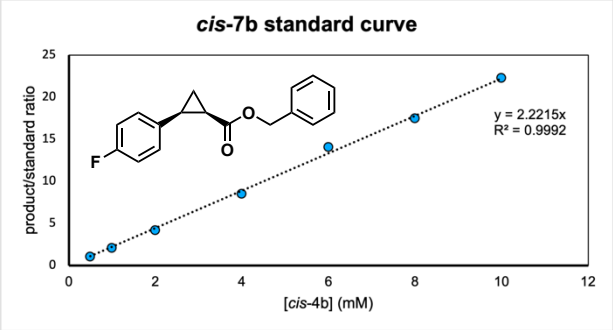


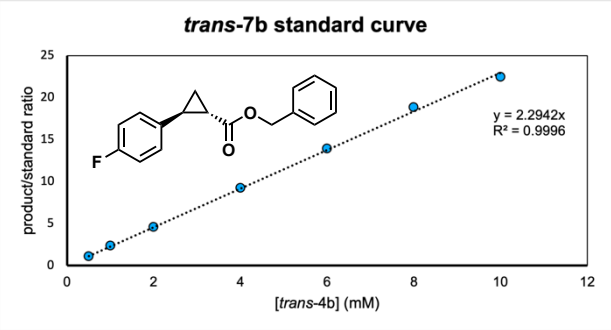
7a. *benzyl 2-phenylcyclopropane-1-carboxylate.*



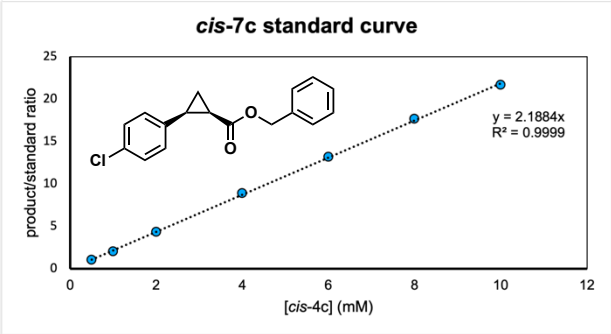


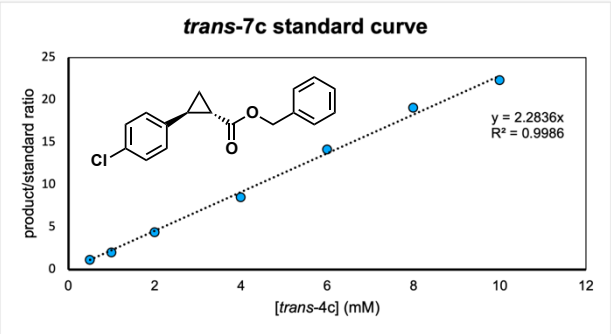
7b. *benzyl 2-(4-fluorophenyl)cyclopropane-1-carboxylate.*



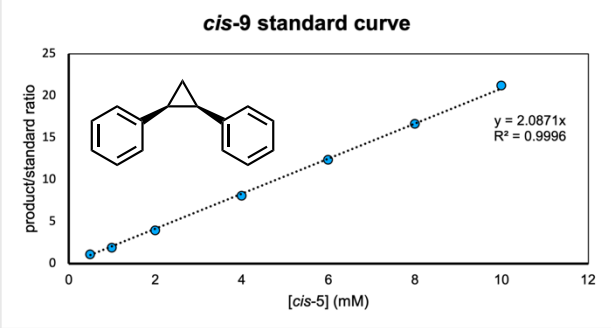


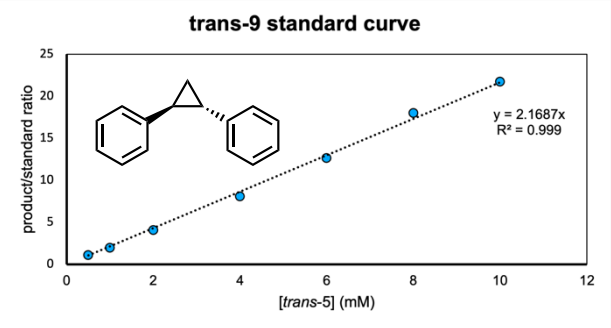
7c. *benzyl 2-(4-chlorophenyl)cyclopropane-1-carboxylate.*



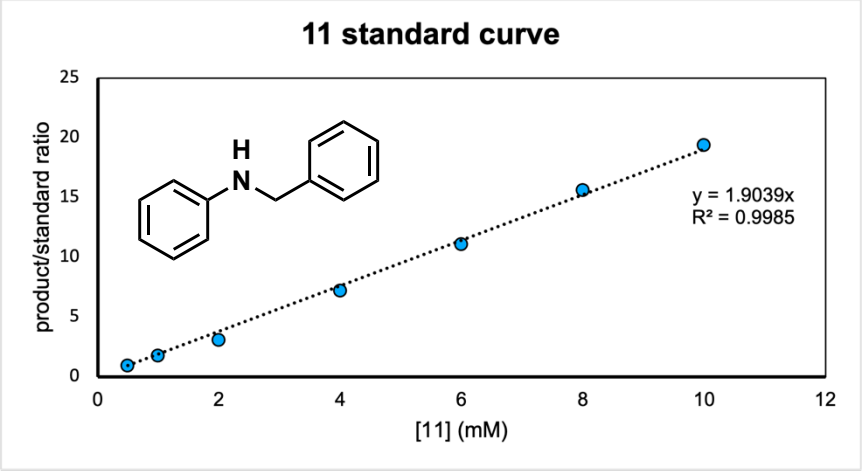


9. *1,2-diphenylcyclopropane.*

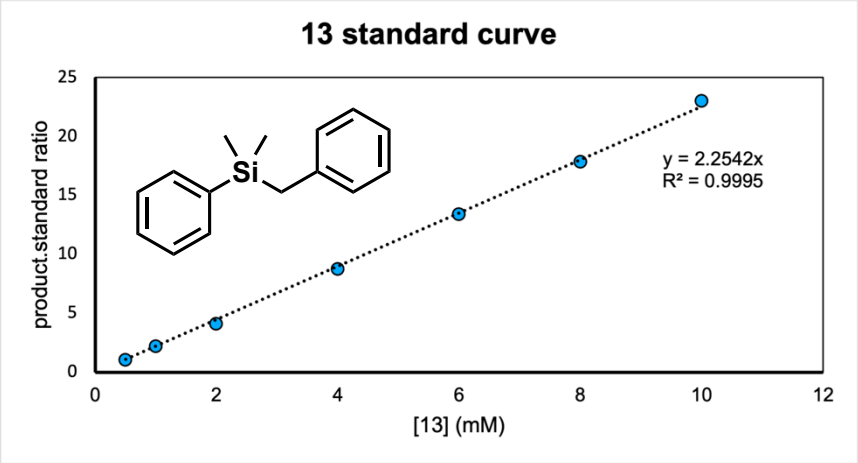




11. *N-benzyl-aniline.*

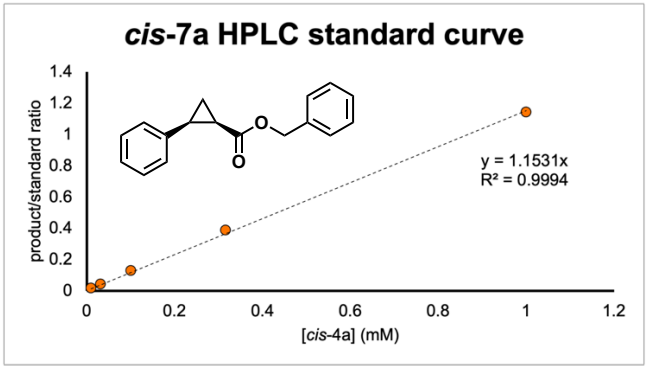


13. *benzyldimethyl(phenyl)silane.*



High-performance liquid chromatography (HPLC) standard curve (7a)

For analysis of *cis*-product yield across the lineage, HPLC-UV was used. A standard curve was generated varying the concentration of *cis*-**7a** from 10 μM to 1 mM using 1 mM 4-ethyl-anisole as a standard in 38% storage buffer and 62% acetonitrile. The standard displayed a retention time of 1.14 min and *cis*-**7a** displayed a retention time of 1.36 min.



Chiral analysis. For product **7a**, chiral analysis was performed by HPLC-UV with a chiral stationary phase, using a Hewlett Packard 1100 series HPLC instrument equipped with a Daicel Chiralpak IB column with hexanes/2-propanol as a mobile phase (1 mL/min). For products **7b** and **7c**, supercritical fluid chromatography was used on a Thar Analytical SFC using a Daicel Chiralpak OJ-H column with 5% *i*-PrOH and 7% *i*-PrOH in hexanes (70% liquid CO_2; 3.5 mL/min) serving as a mobile phase for the analysis of **7b** and **7c**, respectively.

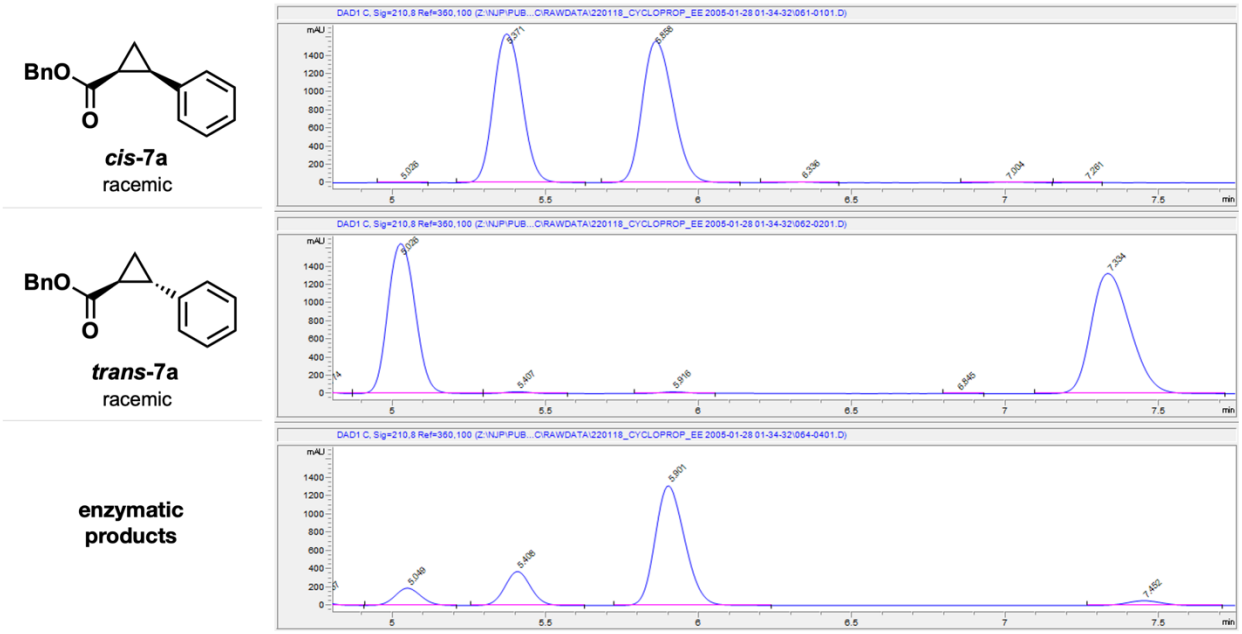


Figure S4. Chiral separation of cyclopropane **7a** enantiomers. Separation was performed 1100 series HPLC equipped with a Daicel Chiralpak IB column (4.6 x 250 mm, 5 μm packing) using 1.5% *i*-PrOH in hexanes as a mobile phase at a flow rate of 1.25 mL/min.

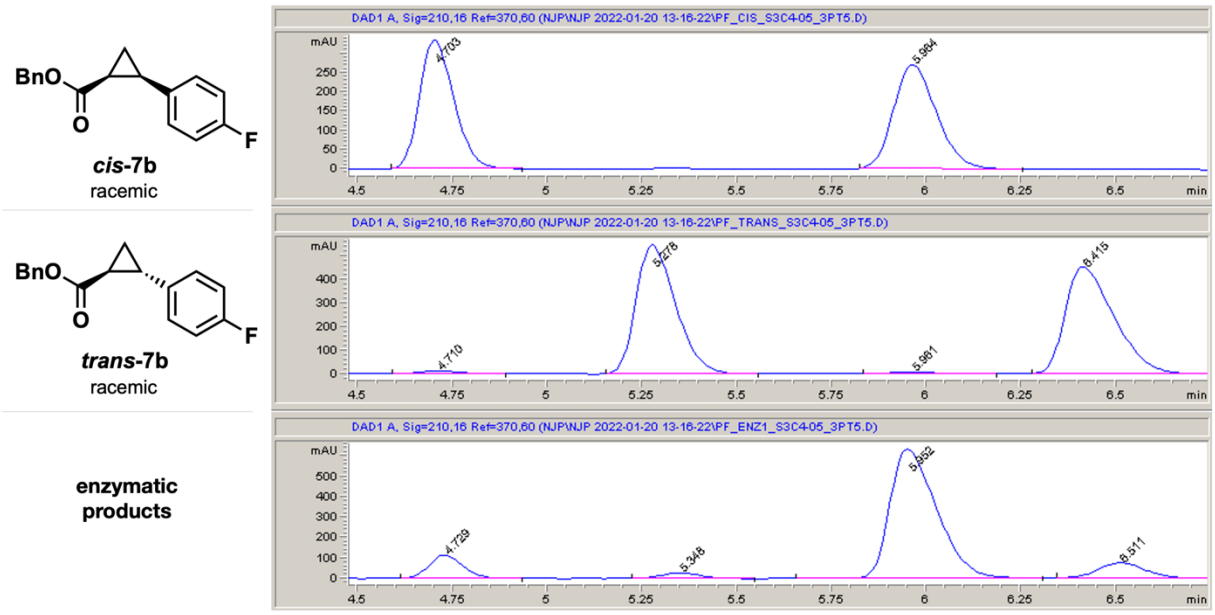


Figure S5. Chiral separation of cyclopropane **7b** enantiomers. Separation was performed on a Thar Analytical SFC equipped with a Daicel Chiralpak OJ-H column (4.6 x 250 mm, 5 μm packing) using 5% *i*-PrOH in hexanes as a mobile phase (70% liquid CO_2) at a flow rate of 3.5 mL/min.

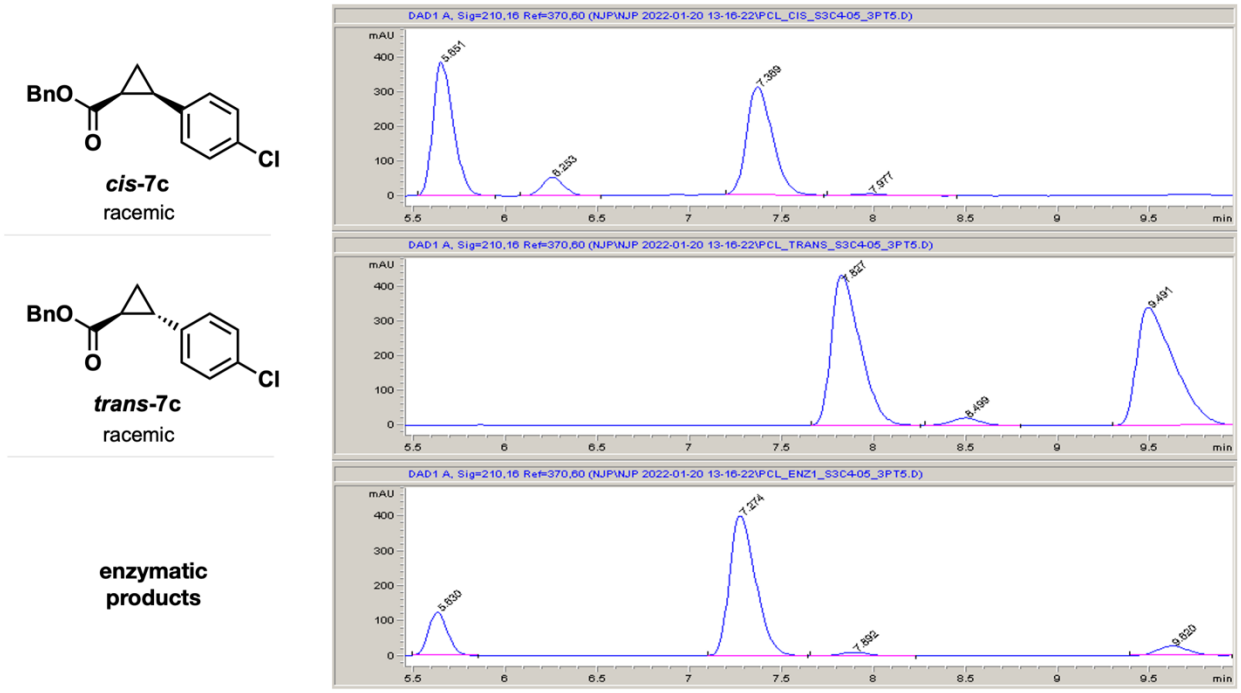
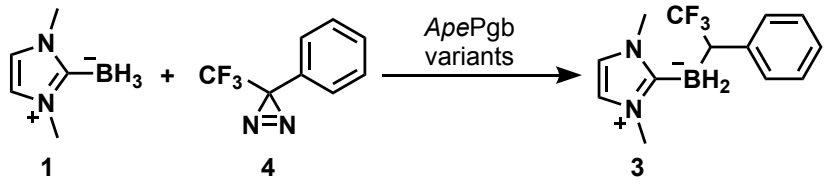


Figure S6. Chiral separation of cyclopropane **7c** enantiomers. Separation was performed on a Thar Analytical SFC equipped with a Daicel Chiralpak OJ-H column (4.6 x 250 mm, 5 μm packing) using 7% *i*-PrOH in hexanes as a mobile phase (70% liquid CO_2) at a flow rate of 3.5 mL/min.

Protein Engineering Strategy

Rounds 1 & 2.

Targeted Reaction: B–H insertion into NHC borane **1**

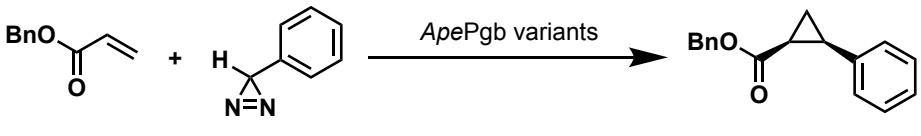


1 **4** **3**

In the preliminary rounds of evolution targeting B–H insertion, the primary goal was to verify that the underlying enzymatic activity with a diazirine carbene donor could be improved by directed evolution. The parent enzyme, *Aeropyrum pernix* protoglobin Y60G F145Q (*Ape*Pgb GQ) bears two mutations relative to the wild-type enzyme. These positions, 60 and 145, along with position 59 had been previously shown to have effects on enzymatic activity for unactivated alkene cyclopropanation.[2] In the first round of evolution, these three positions along with axial ligand H120 were targeted for site-saturation mutagenesis (SSM), as these sites were likely to be responsive to evolution. Site-saturation libraries were screened in whole-cell cultures, analyzing 88 variants per library in addition to four parent controls and four sterile controls. Beneficial variants were identified by GC-MS screening for the formation of the organoborane product **3.** Of these sites, only positions 59 and 60 yielded beneficial mutations. The most beneficial mutation, G60V, was kept and then position 59 was screened in the presence of this mutation, yielding a further beneficial mutation W59L. With this new variant, *Ape*Pgb LVQ, we then proceeded to target an asymmetric reaction using a diazirine bearing no electron-withdrawing group.

Rounds 3–7.

Targeted Reaction: Cyclopropanation of benzyl acrylate **5** using diazirine **6a**



5 **6a** **7a**

When designing the protein engineering strategy for an asymmetric reaction with diazirine **6a**, we considered the screening burden and throughput required. This led to the selection of cyclopropanation of **5** as a target reaction due to the speed with which variants

could be screened. Using HPLC, the activity of each variant could be screened in approximately 2.5 minutes per sample, giving separation of the alkene and diazirine substrates as well as the product diastereomers. Given the throughput of this method, we opted for SSM at positions near the active site or at sites of interest in a homology model of *Ape*Pgb LVQ generated using RaptorX.[5] Other regions of interest beyond the active site were the dimer interface or near the sharp turn between the A and B helices. For each site screened, SSM primers were used to generate libraries at the corresponding positions. For each library, 44 variants were screened in half of a 96-well deep-well plate along with two parent controls and two sterile controls. Over the course of rounds 3–6, positions 11, 45, 56, 62, 63, 73, 89, 90, 93, 102, 112, 114, 119, 148, 149, 152, and 185 were randomized and screened, sequentially fixing mutations V63R, I149L, C45G, and C102S to yield variant *Ape*Pgb GLVRSQL. Notably, in round 6 only position 102 was screened to remove the cysteine residue at this position predicted to form a disulfide with C45, which was removed in the previous round.

With most sites in regions of interest having been screened, we then turned to error prone mutagenesis to produce libraries. Using 300 μM $MnCl_2$ in PCR samples amplified using *Taq* polymerase, an error prone library was generated and screened in four 96-well deep-well plates (88 variants, four parent controls, four sterile controls). This yielded two variants each displaying multiple mutations that displayed an ~2-fold boost in activity. Relative to *Ape*Pgb GLVRSQL, one variant bore mutations G61V, G79R, and F175L while the second bore V60A, S123P, and E165G. These mutations were recombined in a single library which was then screened in a single 96-well deep-well plate. This yielded variant *Ape*Pgb GLAVRSQLL displaying an approximately 5-fold boost relative to the previous parent.

MicroED Structure Determination

Crystallization screening and optimization. Protein was removed from -80 °C, thawed, and diluted in storage buffer (50 mM potassium phosphate (pH 8.0), 150 mM NaCl) to the required concentration as necessary. Using a Crystal Gryphon robot (Art Robbins Instruments), 40 mg/mL and 20 mg/mL solutions of *Ape*Pgb GLVRSQL bearing no HisTag were screened against the Index (Hampton Research), PEGRx (Hampton Research), JCSG+ (Molecular Dimensions), and Wizard (Rigaku) sparse matrix crystallization screens using 200 µL:200 µL protein:precipitant drops against a 30 µL reservoir. Trays were sealed and stored at room temperature. Crystals appeared in well B8 of the Wizard Screen (0.4 M sodium phosphate monobasic / 1.6 M potassium phosphate dibasic, 0.1 M imidazole (pH 8.0), 0.2 M NaCl) after 24 hours. Crystallization was optimized by drop ratio variation with this condition, producing optimal crystals with 0.5 µL 20 mg/mL *Ape*Pgb GLVRSQL and 2.5 µL precipitant solution. Large crystal clusters appeared after 1-3 days. Individual thin plates were isolated from these clusters, cryoprotected by immersion in either 25% glycerol or 20% ethylene glycol in reservoir solution, then mounted in nylon loops, cooled, and stored in liquid nitrogen prior to data collection. These samples were subjected to synchrotron irradiation at SSRL 12-2. However, these crystals exhibited little to no diffraction under these conditions.

MicroED grid preparation. Quantifoil R 2/2, 200 copper mesh grids were glow discharged for 30 s at 15 mA current with negative polarity. The grids were transferred to a Leica GP2 vitrification robot with the sample chamber set to 90% relative humidity and 4 °C. The crystal drop was diluted with 15 µL of mother liquor and the needle-like crystals were broken into smaller pieces by gently pipetting into the drop. A 2 µL aliquot of this protein crystal solution was pipetted onto the carbon side of the grid in the vitrification chamber. Following 20 s of incubation, the grid was blotted from the back for 30 s, plunged into liquid ethane and transferred to liquid nitrogen for storage.

MicroED data collection. The grids were loaded into a Thermo Fisher Scientific Titan Krios G3i transmission electron microscope (TEM), operating at an accelerating voltage of 300 kV under cryogenic conditions. The grids were screened for crystals at low-magnification using the Thermo Scientific EPU-D software. The identified crystals which appeared as thin rectangular sheets on the grid were screened for initial diffraction on a CetaD complementary metal-oxide-semiconductor (CMOS) camera. For data collection, the crystals were brought to eucentric height and followed by single 1 s exposure in diffraction mode with the stage at the

starting tilt angle. MicroED data were collected using a Falcon 4 direct electron detector in counting mode as a movie with continuous rotation of the stage at a rate of 0.15° s^{-1}, and the selected area aperture of 100 µm inserted. Frames were read out every 0.5 s giving MRC datasets of 840 images, corresponding to a 60° wedge from each crystal. The total wedge that was collected over several datasets corresponded to between +65° and -65°.

MicroED data processing. The MRC files were converted to individual frames in SMV format using the freely available MicroED software (https://cryoem.ucla.edu/). The reflections were indexed and integrated in XDS.[6] The datasets were scaled in AIMLESS[7] and phased by molecular replacement in Phaser[8] using a structure of *Ape*Pgb GLVRSQL predicted by AlphaFold2 through the ColabFold environment[9] as a search model. Models were refined in phenix.refine[10] using electron-scattering factors.

Table S5. Data collection and refinement statistics for the structure of *Ape*Pgb GLVRSQL

*Ape*Pgb GLVRSQL	
Unit Cell	
Space group	*P*1
a, *b*, *c* (Å)	46.2, 58.3, 80.7
α, β, γ (°)	104.1, 98.6, 90.1
Data Collection	
Wavelength (Å)	0.0197
Acceleration voltage (kV)	300
Temperature (°C)	-196
Resolution (Å)	24.26 – 2.10
Total no. of reflections	33825
R_{merge}	0.18
$CC_{1/2}$	97.2
I/σ(I)	6.96 (2.4)
Completeness (%)	74
Wilson B-factor (Å^2)	8.58
Refinement	
No. of reflections used in refinement/test set	33825 / 1632
R_{work}	0.197
R_{free}	0.221
No. of nonhydrogen atoms	6547
Protein residues	195
Average *B*-factor (Å^2)	18.2
Root-mean-square deviation from ideal geometry	
bonds (Å)	0.01
angles (Å)	1.5
Ramachandran plot (%)	
favored	95.88
allowed	3.85
outliers	0.27
Rotamer outliers (%)	2.6
PDB accession code	7UTE

Structure Analysis

General analysis of mutated sites. In the presented structure of *Ape*Pgb GLVRSQL, the mutations relative to the wild-type sequence are C45G, W59L, Y60V, V63R, C102S, F145Q, and I149L. All mutations are observed throughout the structure except for V63R in monomers B, C, and D where the density for this position and the neighboring positions is weak. In all monomers, G45 adopts a standard α-helical geometry beginning the final turn of the A helix. The side chain of L59 packs against those of F93 and T152 with closest interatomic distances of 3.3–3.6 Å and 3.5–3.7 Å, respectively across all monomers, while the backbone peptide nitrogen terminates the B helix by donating a hydrogen bond to the peptide oxygen of L56 in a 3_{10} helical turn. The mutation Y60V begins the disrupted loop formed by residues 60-70, a region that adopts a rigid helical conformation in previously determined structures of the homologous *M. acetovorans* protoglobin.[11] The most notable change at this position is the presence of a flipped peptide relative to the standard helical geometry in which the backbone carbonyl now accepts a hydrogen bond from either the Nε or Nη of R90 (C=O---N distance = 2.6–3.1 Å). The peptide nitrogen of this residue donates a hydrogen bond to the backbone carbonyl of L56 while the side chain forms the distal surface of the enzyme active site while packing against the E helix. Density for the R63 mutation can only be reasonably modeled for monomer A in which the side chain is oriented out into solvent and the backbone makes no clear interactions. However, under the presumption that this mutation is responsible for the disruption of the C-terminal portion of the B helix, this causes the side chain of the previous residue, W62, to instead pack between the side chains of P81, K83, and L86 near the beginning of the E helix with interatomic distances of 3.5–3.7 Å. Density for this indole side chain is clearly observed. The S102 mutation terminates the E helix with the side chain generally oriented out towards solvent. Notably, positions 45 and 102, both Cys in the wild-type enzyme, are located close to one another in space (5.5–5.6 Å Cα–Cα separation) suggesting that they may exist as a disulfide when the natural enzyme is under certain environmental conditions. The Q145 side chain packs near the heme cofactor while orienting towards a tunnel into the active site between the B and G helices. This side chain extends off of another 3_{10} helical turn between A144 and Y147 caused by interruption of the standard α-helical geometry by the side chain of P148. Mutation L149 follows this proline residue and its side chain packs against the porphyrin ring (3.3–3.8 Å distance depending on rotameric state) as well as F93 (3.4–3.8 Å) and Q145 (3.3–3.5 Å). Most mutations installed over evolution are near the active site and affect the internal

surface, except for G45 and S102 which remove Cys residues that may form a disulfide between the A and E helices.

B helix disruption and active site access. For analysis of changes around the active site, we have focused on monomer A in the presented MicroED structure of *Ape*Pgb GLVRSQL since this is the only monomer in which residues 60-70 can be reasonably modeled into the density (Figure S7a) which disrupts and truncates the B helix in the protoglobin fold (Figure S7b). In natural protoglobins, which bind diatomic gases,[12] these ligands access the active site through tunnels defined at the interfaces of the B/E helices and B/G helices. In the modeled conformation of this loop, it extends over the B/E helical interface and results in a broad tunnel at the B/G interface (Figure S8a). This likely helps larger molecules, such as a diazirine or other substrates, access the active site to interact with the heme cofactor and undergo catalysis. Given the disordered conformation of this loop in the other three monomers in this structure, it is reasonable that this region is flexible in solution and thus can adopt different conformations depending upon the molecule bound in the active site. The importance of this region in catalysis is underscored by the fact that two mutations were acquired in this loop, V60A and G61V, during the final round of evolution to generate *Ape*Pgb GLAVRSQLL, the primary variant studied in this work. These mutations, along with F175L at the dimer interface, resulted in a 4-fold boost in cyclopropanation activity. In the structure of GLVRSQL, sites 60 and 61 are located at the end of the truncated B helix near the B/E interface and their side chains constitute a portion of the internal active site surface (Figure S8b), meaning they can potentially play a role in both access and binding of substrates.

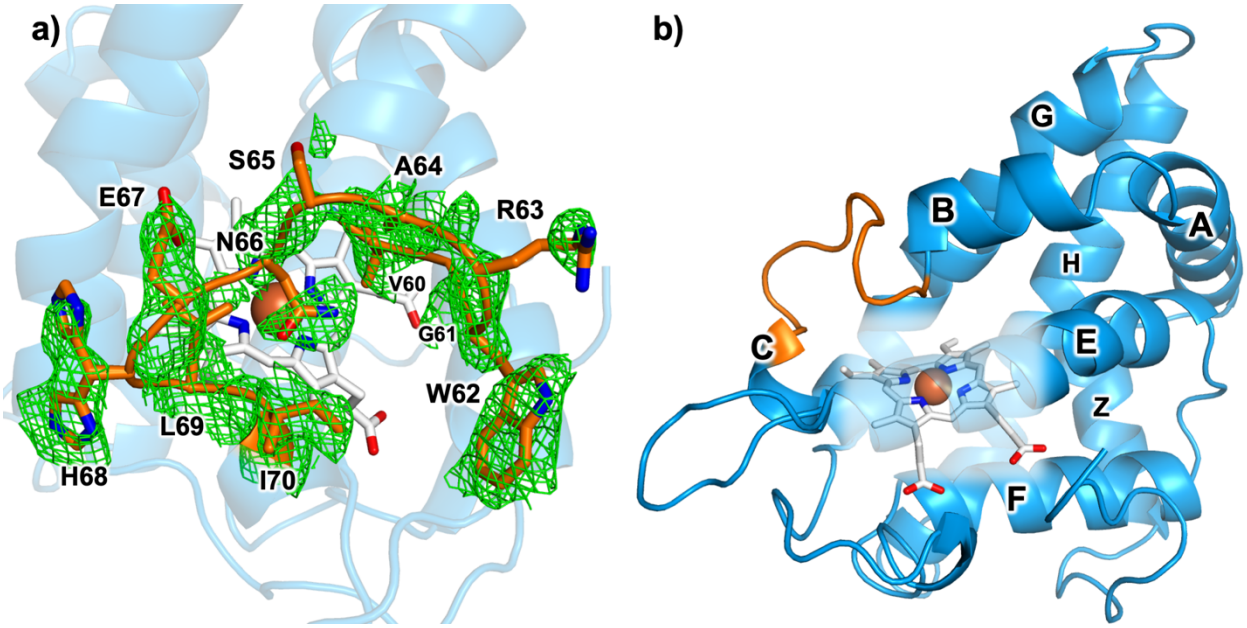


Figure S7. a) Polder omit map (green; 1.6σ) of the MicroED density for residues 60-70 (orange) in monomer A of the *Ape*Pgb GLVRSQL structure (blue). **b)** Structure of *Ape*Pgb GLVRSQL monomer A with helices labeled according to convention for the protoglobin fold. Helix B is observed to be truncated relative to other protoglobins as residues 60-70 are disrupted into a flexible loop.

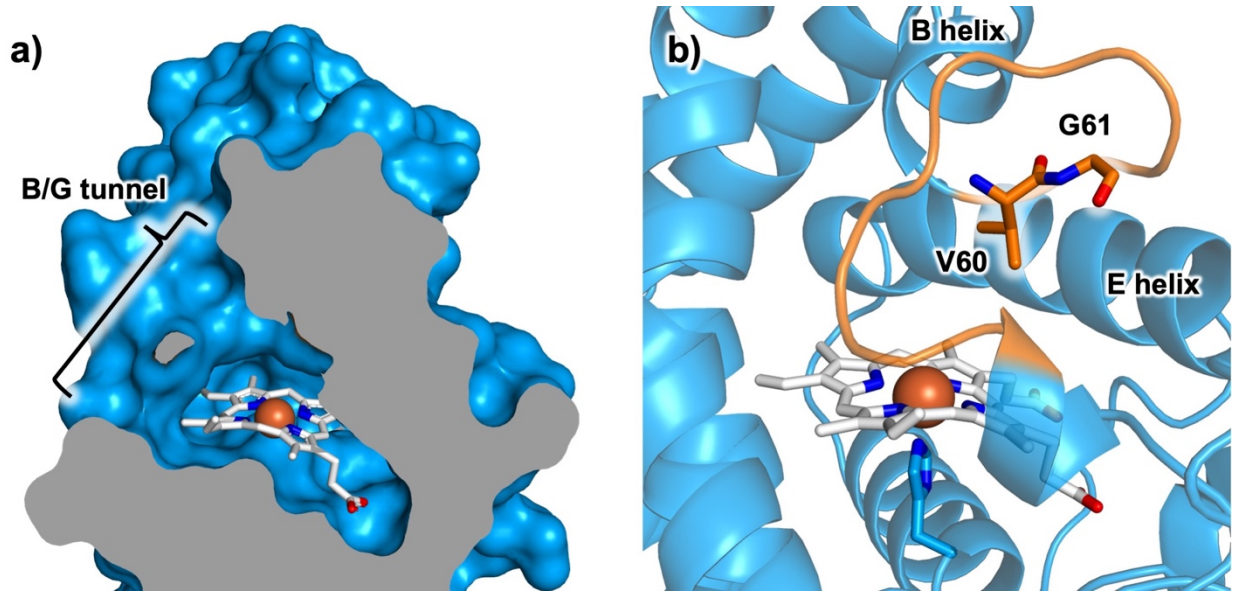
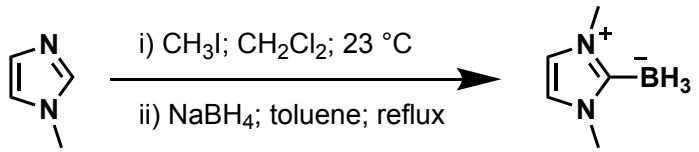


Figure S8. a) Sliced surface of ApePgb GLVRSQL (blue; monomer A) showing the broad tunnel between the B/G helices facilitating access to the active site heme (white). Protein interior is occluded in dark gray for clarity. **b)** Location of positions 60 and 61 which were mutated in the subsequent round of evolution for produce *Ape*Pgb GLAVRSQLL (V60A & G61V), showing their position at the B/E helical interface in the active site distal to the heme cofactor.

Synthesis of Authentic Standards and Substrates

***1,3-Dimethylimidazoyl-2-ylidene borane* (1).**



The NHC-BH_3 substrate was prepared as previously reported.[13] Under argon, *N*-methyl imidazole (4.11 g, 50 mmol) was dissolved in 10 mL of CH_2Cl_2 and was added to a three-neck round-bottom flask equipped with a stir bar and fitted with a septum, a reflux condenser, and connection to a Schlenk line on separate necks. While stirring at 400 rpm, iodomethane (3.75 mL, 60 mmol) was added dropwise over the course of 5 minutes and the reaction was stirred for 1 hour. The solvent was removed *in vacuo* yielding the crude product 1,3-dimethyl-1H-imidazol-3-ium iodide as a pale yellow powder. The flask was then reconnected to the condenser and Schlenk line, as well as a hose venting to the back of the fume hood. Sodium borohydride (2.27 g, 60 mmol) was added to the flask, and the third neck was sealed with a rubber septum. Under a slow flow of argon, 50 mL of anhydrous toluene were applied using a syringe. The reaction was then heated to 125 °C and stirred at 400 rpm for 20 hours. The contents of the flask were then decanted into a 200-mL round-bottom flask, and the residue in the reaction flask was washed twice with refluxing toluene. The washes and reaction were pooled, and the toluene was removed *in vacuo* to yield the crude product as a white powder. The product was isolated by recrystallization in ddH_2O to yield the product 1,3-dimethylimidazoyl-2-ylidene borane (1.54 g; 28% yield).

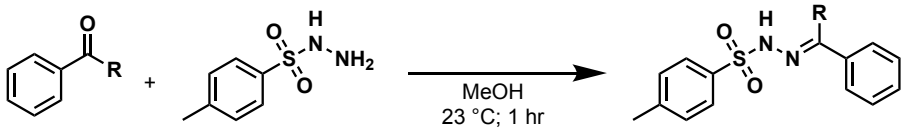
NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 6.79 (s, 2H), 3.72 (s, 6H), 1.36 – 0.62 (dd, *J* = 172.7, 86.3 Hz, 3H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 171.88, 120.03, 36.05.

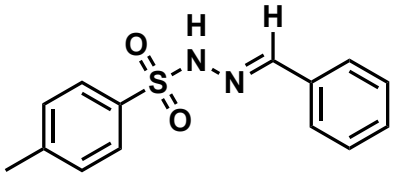
^{11}B NMR (128 MHz, $CDCl_3$, δ): -37.50 (q, *J* = 86.5 Hz).

Tosylhydrazones.



Tosylhydrazones were synthesized as previously described.[14] In a 40-mL reaction vial equipped with a stir bar, tosylhydrazide (930 mg, 5 mmol) was dissolved in 10 mL of methanol. The vial was then cooled in an ice bath for 5 minutes, after which the carbonyl compound (5 mmol) was added dropwise and the reaction was stirred on ice for 3 hours. When a white precipitate had formed, the vial was heated to dissolve the product and then stored at -20 °C for recrystallization. Solvent was then filtered off, and the crystalline product was dried *in vacuo.*

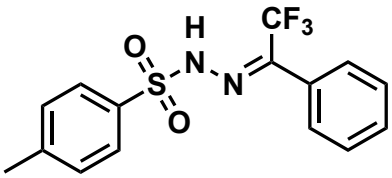
Benzaldehyde tosylhydrazone



^{1}H NMR (400 MHz, $CDCl_3$, δ): 8.08 (s, 1H), 7.91 – 7.86 (m, 2H), 7.77 (s, 1H), 7.60 – 7.55 (m, 2H), 7.38 – 7.33 (m, 3H), 7.33 – 7.29 (m, 2H), 2.40 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 147.9, 144.3, 135.2, 133.1, 130.5, 129.7, 128.7, 128.0, 127.4, 21.6.

Trifluoroacetophenone tosylhydrazone

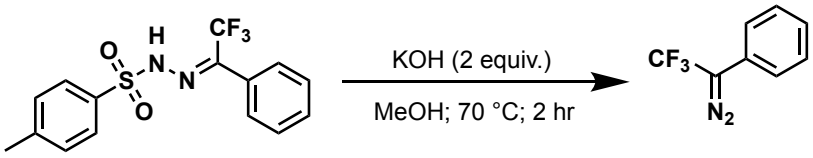


^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.94 (s, 1H), 7.81 (d, J = 8.3 Hz, 2H), 7.56–7.49 (m, 3H), 7.37–7.33 (m, 2H), 7.24 (m, 2H), 2.46 (3, 3H)

^{13}C NMR (101 MHz, $CDCl_3$, δ): 145.2, 141.7 (q, J_{CF} = 35.9 Hz), 134.7, 130.2, 130.0, 128.3, 128.2, 125.4, 120.1 (q, J_{CF} = 275 Hz), 21.9

^{19}F NMR (282 MHz, $CDCl_3$, δ): -68.3 (s)

1-Trifluoromethylphenyldiazomethane (2).



Trifluoromethylphenyldiazomethane was prepared according to previously described procedure.[15] Briefly, trifluoroacetophenone tosylhydrazone (1.71 g; 5 mmol) was added toa round bottom flask and dissolved in a solution of KOH (561 mg; 10 mmol) in MeOH (25 mL; 0.4 M). A reflux condenser was fitted to the flask and the reaction mixture was refluxed for 2 hours, resulting in a deep red solution. The reaction was cooled to room temperature and diluted with

75 mL of deionized water. The product was extracted with pentane (3 x 25 mL), the extract was dried with sodium sulfate, and the pentane was remove *in vacuo*. The diazo compound was then isolated by flash chromatography using 100% pentane as the mobile phase. The diazo compound-containing fractions were pooled and concentrated to yield a red oil. The diazo compound was stored at -20 °C prior to use.

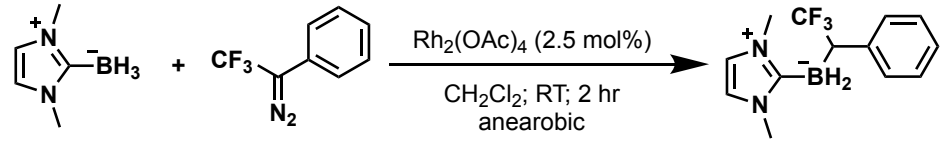
NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.40 (m, 2H), 7.2 (m, 1H), 7.1 (m, 2H)

^{13}C NMR (101 MHz, $CDCl_3$, δ): 129.6, 125.9 (q, J_{CF} = 273), 126.1, 123.6, 122.4 ($C=N_2$ not observed)

^{19}F NMR (282 MHz, $CDCl_3$, δ): -57.4 (s)

(1,2-dimethyl-1*H*-imidazol-3-ium-2-yl)(2,2,2-trifluoro-1-phenylethyl)dihydroborate (3)



Organoborane **3** was prepared using the previously published synthesis.[2] Briefly, a 4 mL vial was equipped with borane **1** (109 mg; 1 mmol), dirhodium acetate ($Rh_2(OAc)_4$; 11 mg; 25 µmol), and a stir bar. The vial was sealed with a septum, then evacuated and backfilled with argon three times. Under argon atmosphere, 2 mL of anhydrous CH_2Cl_2 were added and solids were dissolved, washing walls of vial to suspend rhodium dispersed when backfilled. Using a syringe pump, a CH_2Cl_2 solution (1 mL) of trifluoromethylphenyldiazomethane **2** (186 mg; 1 mmol) was added dropwise over the course of two hours. The solvent was then removed *in vacuo* and the crude reaction mixture was purified by flash chromatography using 40% ethyl acetate in hexanes as a mobile phase. Product-containing fractions were pooled and solvent was removed *in vacuo* to yield the purified product as a white solid.

NMR Spectroscopy:

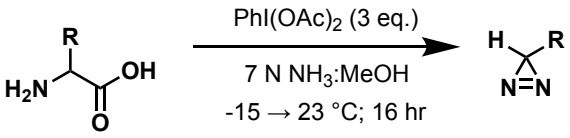
^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.21–7.06 (m, 5H), 6.8 (s; 2H), 3.53 (s; 6H), 2.82–2.61 (m; 1H), 2.2–1.4 (m; 2H)

^{13}C NMR (101 MHz, $CDCl_3$, δ): 143.6, 131.6 (q, J_{CF} = 273 Hz), 128.2, 128.1, 125.0, 120.6, 36.0; boron-binding carbons are not observed as these are weakened due to geminal coupling

^{11}B NMR (128 MHz, $CDCl_3$, δ): -26.7 (t, J = 89.7 Hz)

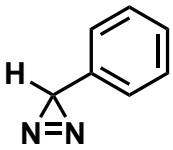
^{19}F NMR (282 MHz, $CDCl_3$, δ): -61.8 (d, J = 12.7 Hz)

Diazirines (6a–e).



Synthesis of diazirines **6a–e** was adapted from the previously reported method for the direct transformation of α-amino acids into terminal diazirines.[16] Under air, 60 mL of 7 N ammonia in methanol were transferred, along with a stir bar, to a 100-mL round-bottom flask in a salt ice bath (3:1 ice:NaCl; -15 °C). When reactants were not being added, the flask was sealed loosely with a rubber septum. The phenylglycine reactant (20 mmol) was added to the flask while stirring at 240 rpm. Phenyliodonium diacetate (PIDA; 19.3 g, 60 mmol, 3 eq.) was then added to the flask, which was then sealed with a rubber septum and fitted with an empty balloon to accommodate the eliminated carbon dioxide. The reaction was stirred for 16 hours, allowing the ice to melt and come to room temperature. Upon completion, the cloudy yellow/orange reaction was vacuum-filtered into a 250-mL separatory funnel and extracted three times with pentane. Using thin-layer chromatography with pentane as a mobile phase, the main constituents of this extract were iodobenzene (R_f = ~0.9), the product diazirine (R_f = 0.6-0.8), and the side product nitrile (R_f = 0-0.1). The pentane was removed *in vacuo* and the resulting yellow oil was purified by flash chromatography using 100% pentane as a mobile phase. Diazirine-containing fractions were pooled and concentrated *in vacuo*, yielding the product diazirine as a pale yellow or colorless oil (20–40% yield)

***3-Phenyl-3H-diazirine* (6a).** (20% yield)

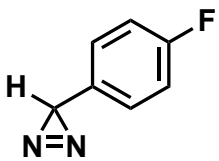


NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.37 – 7.29 (m, 3H), 6.96 (m, 2H), 2.05 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 136.38, 128.32, 128.01, 125.14, 23.38.

***3-(para-Fluorophenyl)-3H-diazirine* (6b).** (27% yield)



NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.06 – 6.97 (m, 2H), 6.92 – 6.84 (m, 2H), 2.05 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 164.0, 161.6, 161.0, 132.3, 132.2, 126.9, 126.8.

^{19}F NMR (282 MHz, $CDCl_3$, δ): -113.4 – -113.6 (m).

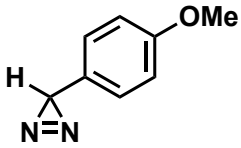
***3-(para-Chlorophenyl)-3H-diazirine* (6c).** (31% yield)

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.32 – 7.26 (m, 2H), 6.87 – 6.81 (m, 2H), 2.04 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 135.0, 134.2, 128.7, 126.5, 23.1.

***3-(para-Methoxy)-3H-diazirine* (6d).** (5% yield)

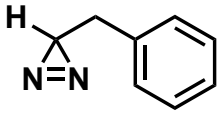


NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 6.90 – 6.78 (m, 4H), 3.79 (s, 1H), 2.01 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 159.8, 128.6, 126.4, 114.0, 55.5, 23.2.

***3-Benzyl-3H-diazirine* (6e).** (57% yield)

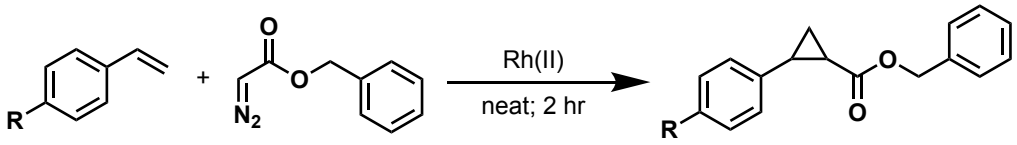


NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.38 – 7.31 (m, 2H), 7.29 – 7.24 (m, 3H), 2.55 (d, J = 4.4 Hz, 2H), 1.13 (q, J = 3.2, 2.0 Hz, 1H).

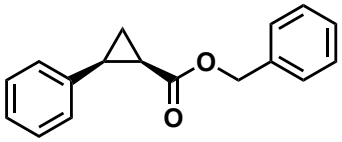
^{13}C NMR (101 MHz, $CDCl_3$, δ): 136.2, 129.1, 128.9, 127.1, 36.74, 22.1, 14.2.

Benzyl acrylate cyclopropanation standards (7a–c).



Racemic standards for cyclopropane product **7a–c** were synthesized using a previously reported procedure for synthesis of the related ethyl ester.[3] Briefly, rhodium acetate dimer (10–15 μmol, 4.4-6.6 mg) was added to a 2-dram vial equipped with a stir bar. The desired styrene (10 mmol) was added, washing the walls of the vial, then the vial was sealed with a septum and purged with argon. Using a syringe pump, 1 mL of 1 M benzyldiazoacetate in CH_2Cl_2 was then added dropwise over 2 hours. Solvent was removed *in vacuo* and the cyclopropane product was isolated by flash chromatography using a shallow gradient of 0–10% ethyl acetate in hexane.

***Benzyl cis-2-phenylcyclopropane-1-carboxylate* (*cis*-7a).**

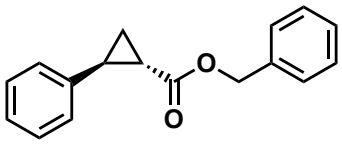


NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.32 – 7.28 (m, 3H), 7.26 (m, 4H), 7.25 – 7.19 (m, 1H), 7.13 – 7.07 (m, 2H), 4.94 (d, *J* = 12.3 Hz, 1H), 4.84 (d, *J* = 12.3 Hz, 1H), 2.62 (td, *J* = 9.0, 7.6 Hz, 1H), 2.17 (ddd, J = 9.3, 7.8, 5.6 Hz, 1H), 1.76 (dt, *J* = 7.5, 5.4 Hz, 1H), 1.37 (ddd, *J* = 8.7, 7.8, 5.1 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 171.0, 136.5, 136.1, 129.4, 128.5, 128.3, 128.1, 126.9, 66.2, 25.9, 21.9, 11.5.

***Benzyl trans-2-phenylcyclopropane-1-carboxylate* (*trans*-7a).**

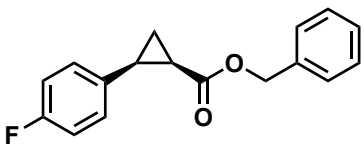


NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.29 – 7.20 (m, 5H), 7.19 – 7.14 (m, 2H), 7.12 – 7.06 (m, 1H), 7.01 – 6.96 (m, 2H), 5.05 (d, *J* = 1.8 Hz, 2H), 2.46 (ddd, *J* = 9.2, 6.5, 4.1 Hz, 1H), 1.86 (ddd, *J* = 8.4, 5.3, 4.2 Hz, 1H), 1.53 (ddd, *J* = 9.2, 5.3, 4.5 Hz, 1H), 1.23 (ddd, *J* = 8.4, 6.5, 4.6 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 173.4, 140.1, 136.1, 128.7, 128.6, 128.4, 128.4, 126.7, 126.3, 66.7, 26.5, 24.3, 17.4.

***Benzyl cis-2-(4-fluorophenyl)cyclopropane-1-carboxylate* (*cis-7b*).**



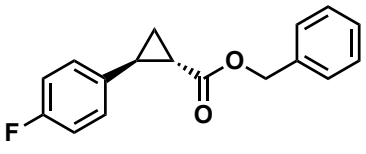
NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.31 – 7.27 (m, 3H), 7.20 – 7.15 (m, 2H), 7.14 – 7.08 (m, 2H), 6.93 – 6.86 (m, 2H), 4.93 (d, *J* = 12.3 Hz, 1H), 4.85 (d, *J* = 12.2 Hz, 1H), 2.60 – 2.50 (m, 1H), 2.13 (ddd, *J* = 9.3, 7.8, 5.6 Hz, 1H), 1.69 (dt, *J* = 7.5, 5.4 Hz, 1H), 1.36 (ddd, *J* = 8.7, 7.8, 5.1 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 170.9, 163.1, 160.6, 136.0, 132.2, 132.1, 130.94 130.9, 128.5, 128.4, 128.2, 115.1, 114.9, 66.4, 25.1, 21.9, 11.7.

^{19}F NMR (282 MHz, $CDCl_3$, δ): -116.02 – -116.17 (m).

***Benzyl trans-2-(4-fluorophenyl)cyclopropane-1-carboxylate* (*trans*-7b).**



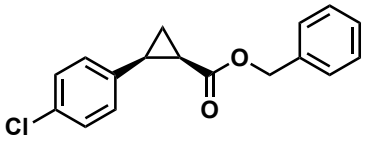
NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.40 – 7.32 (m, 5H), 7.06 (ddt, J = 8.3, 5.3, 2.6 Hz, 2H), 6.99 – 6.93 (m, 2H), 5.16 (d, *J* = 2.4 Hz, 2H), 2.55 (ddd, *J* = 9.3, 6.5, 4.1 Hz, 1H), 1.91 (ddd, *J* = 8.4, 5.3, 4.2 Hz, 1H), 1.62 (ddd, *J* = 9.2, 5.2, 4.6 Hz, 1H), 1.29 (ddd, *J* = 8.4, 6.6, 4.6 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 173.3, 163.0, 160.6, 136.0, 135.7, 135.7, 128.8, 128.5, 128.4, 128.0, 127.9, 115.6, 115.3, 66.8, 25.8, 24.1, 17.2.

^{19}F NMR (282 MHz, $CDCl_3$, δ): -116.28 (tt, *J* = 9.9, 5.5 Hz).

***Benzyl cis-2-(4-chlorophenyl)cyclopropane-1-carboxylate* (*cis*-7c).**

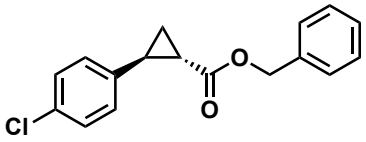


NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.31 – 7.27 (m, 3H), 7.19 – 7.12 (m, 4H), 7.12 – 7.07 (m, 2H), 4.93 (d, *J* = 12.2 Hz, 1H), 4.86 (d, *J* = 12.2 Hz, 1H), 2.54 (tdd, *J* = 8.6, 7.1, 0.8 Hz, 1H), 2.15 (ddd, *J* = 9.2, 7.8, 5.6 Hz, 1H), 1.70 (dt, *J* = 7.5, 5.4 Hz, 1H), 1.36 (ddd, *J* = 8.7, 7.9, 5.2 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 170.8, 136.0, 135.0, 132.6, 130.7, 128.6, 128.4, 128.3, 66.4, 25.2, 22.0, 11.6.

***Benzyl trans-2-(4-chlorophenyl)cyclopropane-1-carboxylate* (*trans*-7c).**

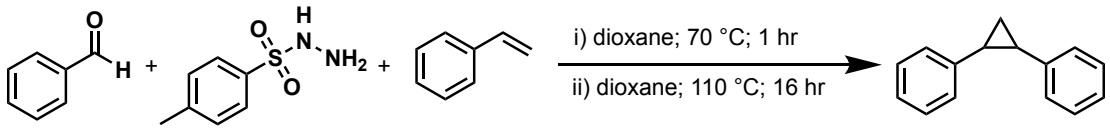


NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.39 – 7.31 (m, 5H), 7.26 – 7.21 (m, 2H), 7.05 – 6.99 (m, 2H), 5.16 (d, *J* = 2.4 Hz, 2H), 2.53 (ddd, *J* = 9.2, 6.5, 4.1 Hz, 1H), 1.93 (ddd, *J* = 8.4, 5.3, 4.1 Hz, 1H), 1.64 (ddd, *J* = 9.2, 5.3, 4.7 Hz, 1H), 1.30 (ddd, *J* = 8.4, 6.5, 4.7 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 173.1, 138.6, 136.0, 132.4, 128.7, 128.7, 128.5, 128.4, 127.7, 66.8, 25.9, 24.3, 17.4.

***1,2-Diphenylcyclopropane* (9).**



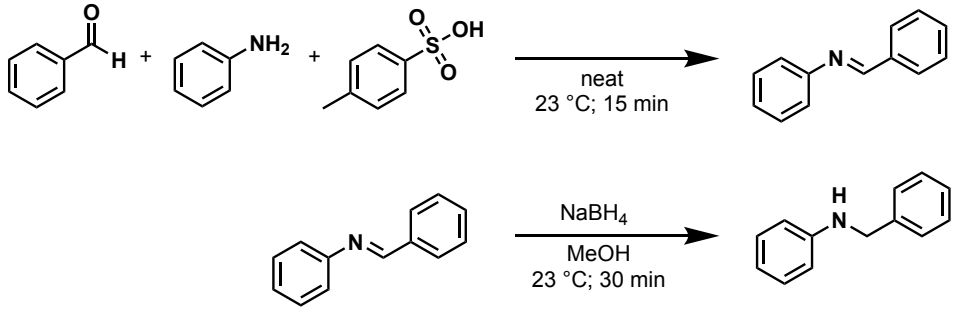
Diphenylcyclopropane **9** was synthesized as previously reported.[17] In a 50-mL round-bottom flask, benzaldehyde (300 µL, 3 mmol) and *p*-toulenesulfonyl hydrazide (560 mg, 3 mmol) were dissolved in 10 mL of dioxane, and the solution was stirred at 300 rpm and heated to 70 °C. After 1 hour, styrene (625 µL, 6 mmol) dissolved in 20 mL of dioxane was added and the flask was fitted with a reflux condenser topped with a septum and balloon. The reaction was heated to 110 °C and stirred at 300 rpm for 16 hours. The reaction was then cooled, transferred to an Erlenmeyer flask, diluted in 120 mL of ethyl acetate, and filtered over celite. The filtrate was dried with Na_2SO_4 and evaporated *in vacuo* to yield a crude yellow oil. The product was isolated by flash chromatography using hexane as a mobile phase. The *cis*- and *trans*- isomers could not be isolated from one another. The ratio of these products was determined by ^{1}H NMR.

NMR Spectroscopy: measurements are consistent with previously measured spectra[7]

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.33 – 7.27 (m, 4H *trans*), 7.22 – 7.18 (m, 2H *trans*), 7.18 – 7.13 (m, 4H *trans*), 7.12 – 7.07 (m, 4H *cis*), 7.07 – 7.01 (m, 2H *cis*), 6.94 (m, 4H *cis*), 2.49 (ddd, *J* = 8.4, 6.4, 1.7 Hz, 2H *cis*), 2.18 (td, *J* = 7.3, 1.8 Hz, 2H *trans*), 1.51 – 1.41 (m, 1H *cis* + 2H *trans*), 1.38 (m, 1H *cis*).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 142.7, 138.5, 129.1, 128.8, 128.5, 127.8, 125.9, 125.9, 125.7, 28.2, 24.4, 18.4, 11.5.

***N-Benzylaniline* (11).**



In a 20-mL glass vial equipped with a stir bar, aniline (1.83 mL, 20 mmol), benzaldehyde (2.04 mL, 20 mmol), and *p*-toluenesulfonic acid (170 mg, 1 mmol) were mixed with stirring until the

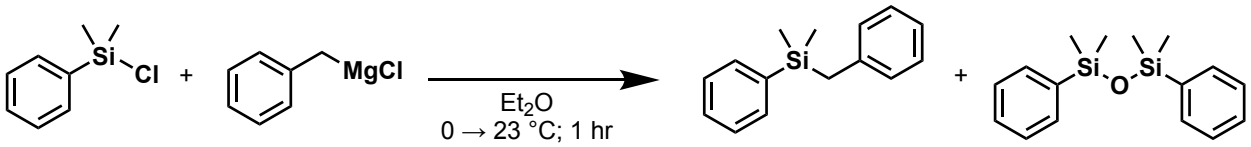
reaction had solidified (~15 minutes). The crude benzylidene aniline product was then isolated by recrystallization in ethanol and used without further purification. In an 8-mL vial equipped with a stir bar, benzylidene anline (360 mg, 2 mmol) was dissolved in 2 mL of methanol. Sodium borohydride (150 mg, 4 mmol) was then added, and the reaction was stirred for 30 minutes. The reaction was then quenched with 1 M HCl. The pH was adjusted to ~14 using 1 M NaOH, and the aqueous layer was then extracted three times with diethyl ether. The organic extracts were pooled, dried with Na_2SO_4, and the solvent was removed *in vacuo*. The product *N*-benzyl-aniline was purified by flash chromatography using a gradient of 0–10% ethyl acetate in hexane.

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.40 – 7.31 (m, 4H), 7.31 – 7.25 (m, 2H), 7.22 – 7.15 (m, 2H), 6.78 – 6.72 (m, 1H), 6.70 – 6.65 (m, 2H), 4.34 (s, 2H).

^{13}C NMR (101 MHz, $CDCl_3$, δ): 129.4, 128.8, 127.8, 127.5, 118.2, 113.5, 48.8.

***Benzyldimethyl(phenyl)silane* (13).**



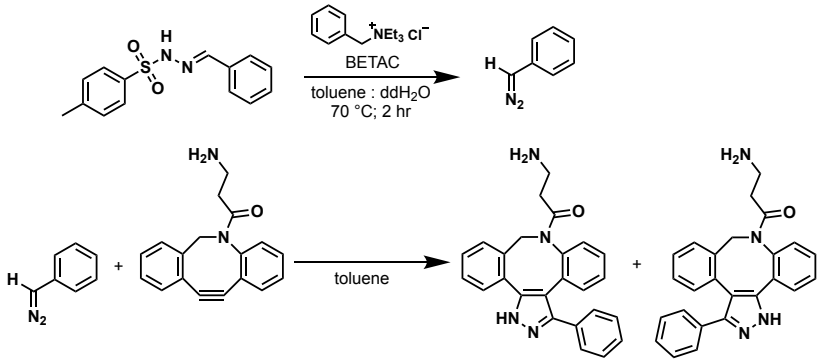
Under argon, 6 mL of 1 M benzylmagnesium chloride in diethyl ether were added to a 40-mL vial equipped with a stir bar. The vial was cooled in an ice bath and, while stirring at 250 rpm, chloro(dimethyl)phenylsilane (840 µL, 5 mmol) was added dropwise. The vial was then removed from the ice bath and stirred for one hour at room temperature. The reaction was quenched with 1 M HCl and transferred to a 50-mL separatory funnel. The aqueous layer was washed three times with diethyl ether, and the organic layers were pooled, dried with Na_2SO_4, and the solvent was removed *in vacuo*. The product was purified by flash chromatography using a gradient of 0–10% ethyl acetate in hexane. By ^{1}H NMR, the desired silane product could not be fully separated from the siloxane byproduct and was isolated in a mixture of 56% silane and 44% siloxane.

NMR Spectroscopy:

^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.50 – 7.45 (m, 2H), 7.41 – 7.33 (m, 3H), 7.19 (m, 2H), 7.08 (m, 1H), 6.95 (m, 2H), 2.32 (s, 2H), 0.26 (s, 6H)

^{13}C NMR (101 MHz, $CDCl_3$, δ): 140.0, 139.8, 138.6, 133.9, 133.2, 129.4, 129.2, 128.5, 128.2, 127.9, 127.9, 124.2, 26.3, 1.0, -3.3.

Dibenzylcyclooctyne amine cycloaddition product standards (S2).



In an 8-mL vial equipped with a stir bar, benzaldehyde tosylhydrazone (99.3 mg, 360 µmol) and benzyl triethylamine chloride (10 mg) were dissolved in 1 mL of 14% NaOH in ddH_2O. This aqueous layer was topped with 1 mL of toluene, and the reaction was stirred at 70 °C for two hours to generate phenyldiazomethane. When the organic layer had turned dark red, the reaction was quickly cooled to room temperature in a water bath. Using a 200-µL pipette fitted with a plastic tip, the organic layer was added dropwise to an 8-mL vial containing dibenzylcyclooctyne amine (50 mg, 180 µmol) until the red color persisted. The excess diazo was quenched with 1 M HCl (5 mL). The solution was then transferred to a 50-mL separatory funnel, and an additional 15 mL of 1 M HCl were added. The acidic aqueous layer was washed twice with pentane, and then the pH was adjusted to 14 using 1 M NaOH, which was verified using a pH strip. Upon transition to basic pH, the solution turned cloudy white. The basic aqueous layer was then extracted three times with diethyl ether. The organic layers were pooled, dried over Na_2SO_4, and filtered into a round-bottom flask. The solvent was removed *in vacuo*, yielding the product as a yellow powder. By ^1H NMR, the regioisomeric ratio (rr) was determined to be 45:55.

NMR Spectroscopy:

^1H NMR (400 MHz, $CDCl_3$, δ): 7.56 – 6.94 (m, 13 H), 5.92 (d, *J* = 15.9 Hz, 1H), 4.42 (d, *J* = 16.0 Hz, 1H major isomer), 4.40 (d, *J* = 16.0 Hz, 1H minor isomer), 2.80 – 2.65 (m, 2H), 2.26 – 2.15 (m, 1H), 1.97 – 1.84 (m, 2H)

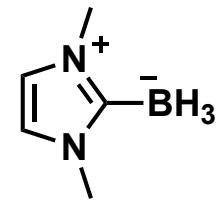
^{13}C NMR (101 MHz, $CDCl_3$, δ): 171.9, 171.5, 148.0, 146.7, 144.2, 143.5, 140.7, 140.2, 134.6, 134.3, 133.9, 133.7, 133.1, 132.7, 131.8, 131.6, 131.1, 130.3, 130.1, 130.0, 129.8, 129.5, 129.2, 128.9, 128.8, 128.7, 128.6, 128.5, 128.4, 128.3, 128.3, 128.2, 127.9, 127.5, 127.5, 127.2, 127.1, 121.8, 121.7, 119.1, 118.9, 117.5, 114.8, 111.2, 77.4, 52.8, 52.7, 37.9, 37.1, 37.0, 35.8, 26.4, 24.9, 1.2.

High Resolution Mass Spectrometry:

Calculated Mass: 394.17881 ($C_{25}H_{22}N_4O$). Observed Mass: 394.18045

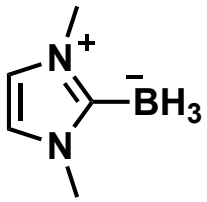
^{1}H NMR spectrum for *N*-heterocyclic carbene borane (**1**)

400 MHz, $CDCl_3$, 23 °C



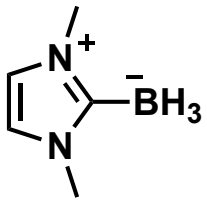
^{13}C NMR spectrum for *N*-heterocyclic carbene borane (**1**)

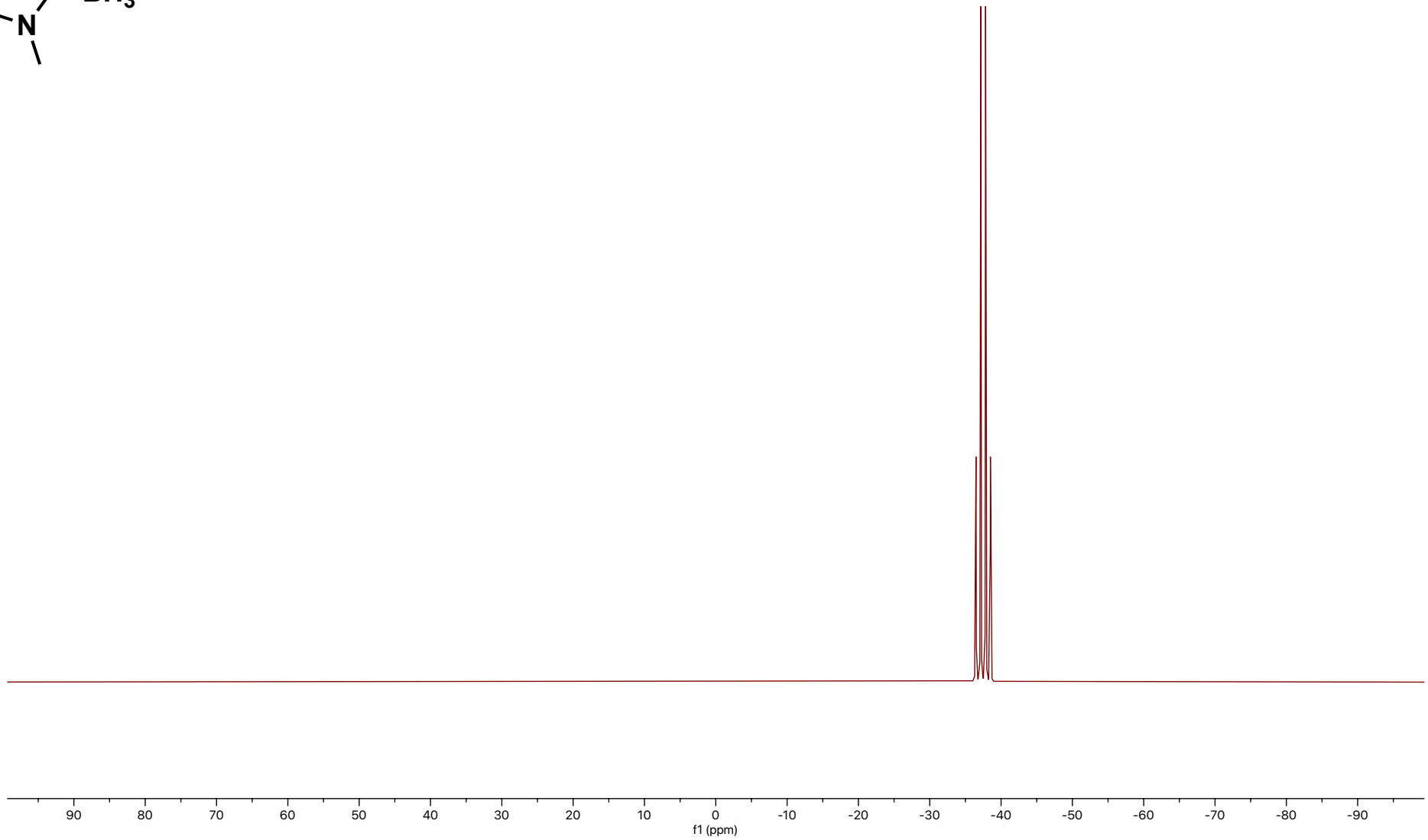
101 MHz, $CDCl_3$, 23 °C



^{11}B NMR spectrum for *N*-heterocyclic carbene borane (**1**)

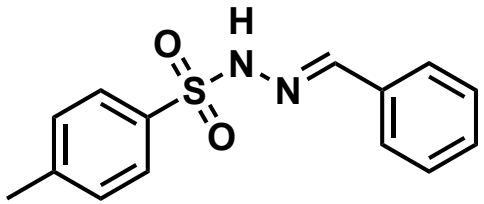
128 MHz, $CDCl_3$, 23 °C





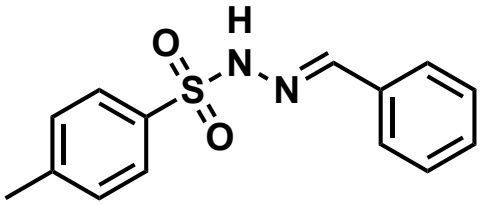
^{1}H NMR spectrum for benzaldehyde tosylhydrazone

400 MHz, $CDCl_3$, 23 °C



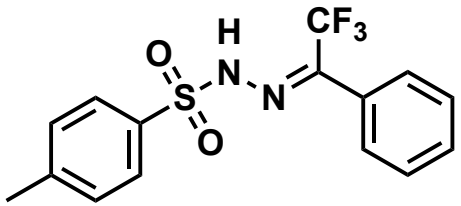
^{13}C NMR spectrum for benzaldehyde tosylhydrazone

101 MHz, $CDCl_3$, 23 °C



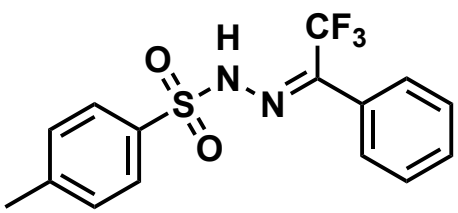
^{1}H NMR spectrum for trifluoroacetophenone tosylhydrazone

400 MHz, $CDCl_3$, 23 °C



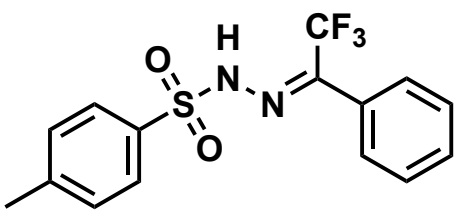
^{13}C NMR spectrum for trifluoroacetophenone tosylhydrazone

101 MHz, $CDCl_3$, 23 °C



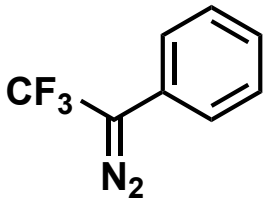
^{19}F NMR spectrum for trifluoroacetophenone tosylhydrazone

282 MHz, $CDCl_3$, 23 °C



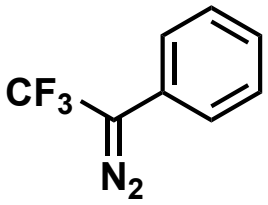
^{1}H NMR spectrum for 1-trifluoromethylphenyldiazomethane (**2**)

400 MHz, $CDCl_3$, 23 °C



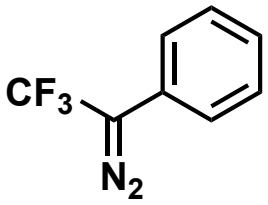
^{13}C NMR spectrum for 1-trifluoromethylphenyldiazomethane (**2**)

101 MHz, $CDCl_3$, 23 °C



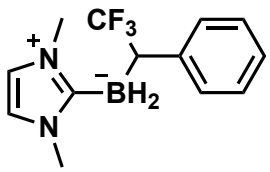
^{19}F NMR spectrum for 1-trifluoromethylphenyldiazomethane (**2**)

282 MHz, $CDCl_3$, 23 °C



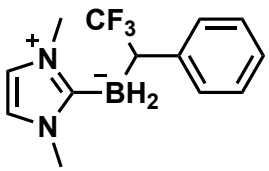
^{1}H NMR spectrum for (1,2-dimethyl-1*H*-imidazol-3-ium-2-yl)(2,2,2-trifluoro-1-phenylethyl)dihydroborate (**3**)

400 MHz, $CDCl_3$, 23 °C



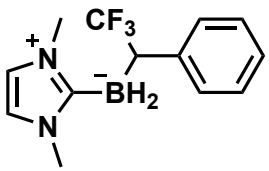
^{13}C NMR spectrum for (1,2-dimethyl-1*H*-imidazol-3-ium-2-yl)(2,2,2-trifluoro-1-phenylethyl)dihydroborate (**3**)

101 MHz, $CDCl_3$, 23 °C



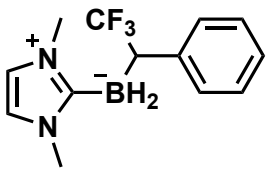
^{11}B NMR spectrum for (1,2-dimethyl-1*H*-imidazol-3-ium-2-yl)(2,2,2-trifluoro-1-phenylethyl)dihydroborate (**3**)

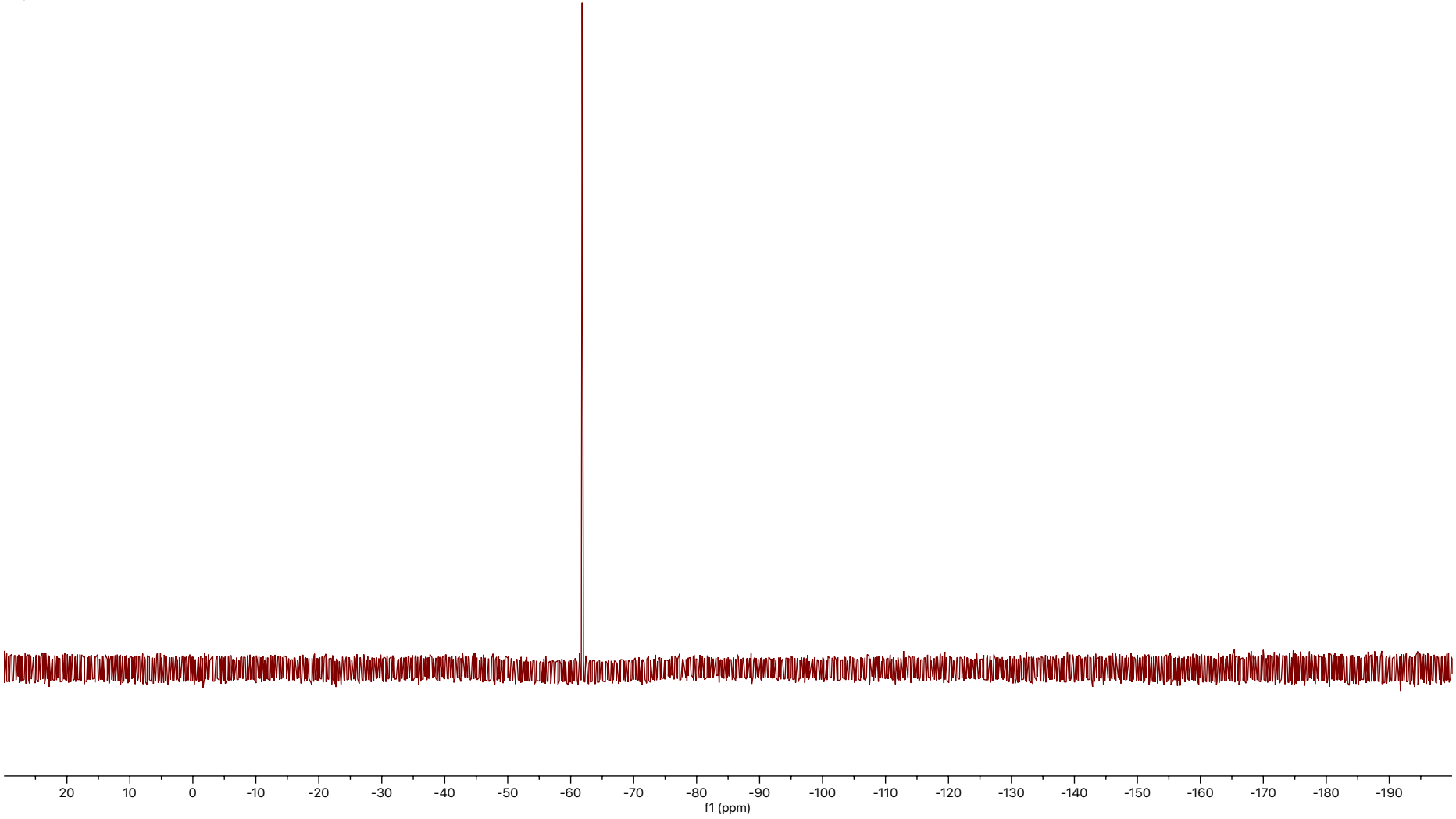
128 MHz, $CDCl_3$, 23 °C



^{19}F NMR spectrum for (1,2-dimethyl-1*H*-imidazol-3-ium-2-yl)(2,2,2-trifluoro-1-phenylethyl)dihydroborate (**3**)

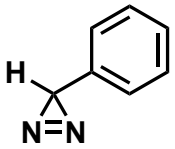
282 MHz, $CDCl_3$, 23 °C





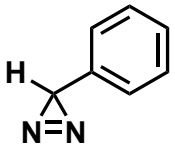
^{1}H NMR spectrum for 3-phenyl-*3H*-diazirine (**6a**)

400 MHz, $CDCl_3$, 23 °C



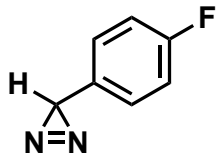
^{13}C NMR spectrum for 3-phenyl-*3H*-diazirine (**6a**)

101 MHz, $CDCl_3$, 23 °C



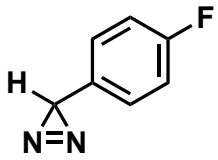
^{1}H NMR spectrum for 3-(4-fluorophenyl)-*3H*-diazirine (**6b**)

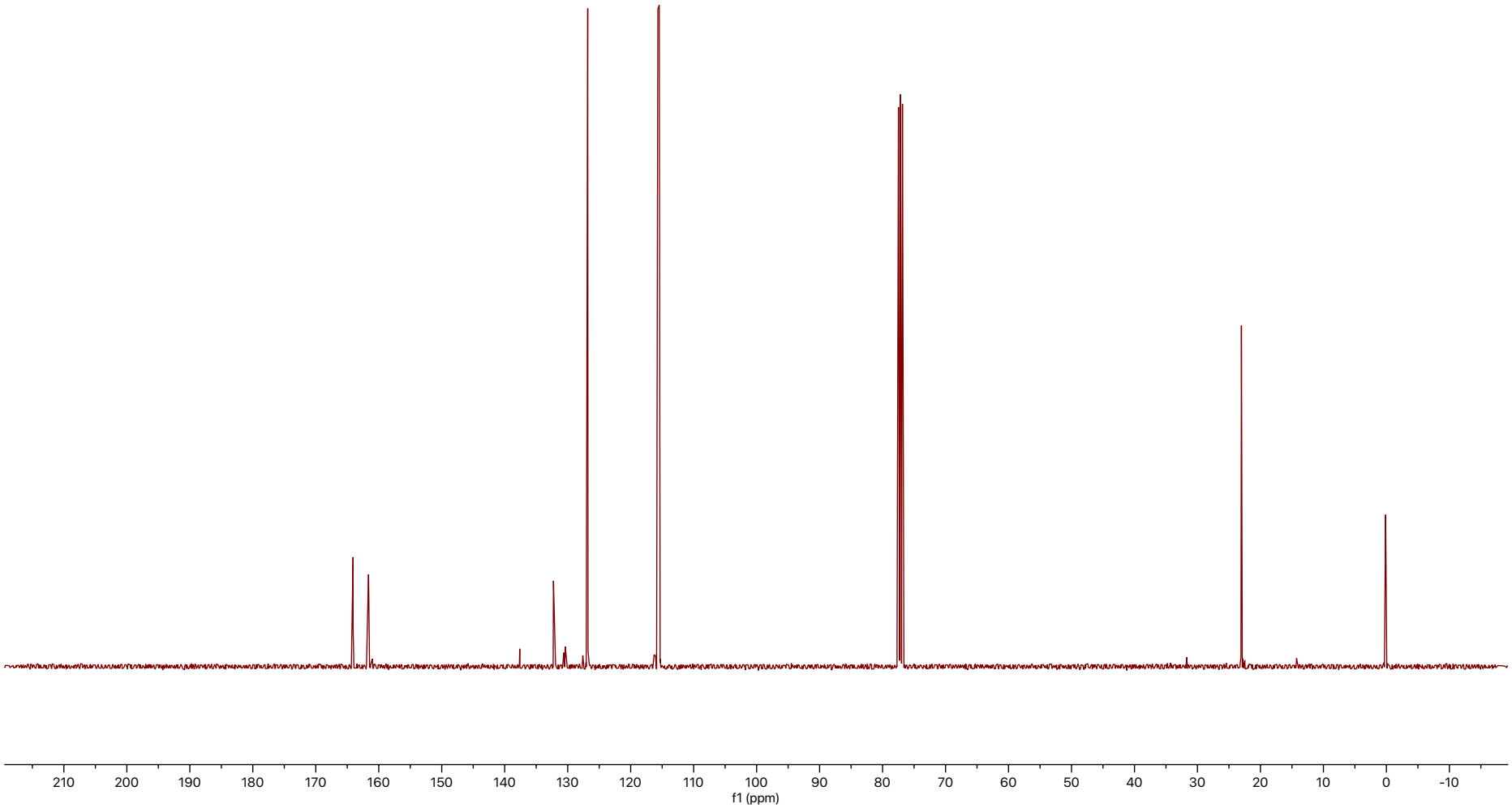
400 MHz, $CDCl_3$, 23 °C



^{13}C NMR spectrum for 3-(4-fluorophenyl)-*3H*-diazirine (**6b**)

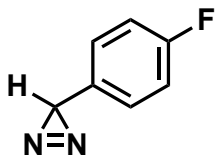
101 MHz, $CDCl_3$, 23 °C





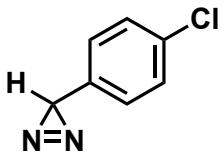
^{19}F NMR spectrum for 3-(4-fluorophenyl)-*3H*-diazirine (**6b**)

282 MHz, $CDCl_3$, 23 °C



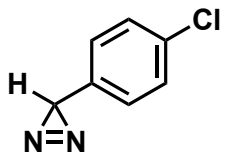
^{1}H NMR spectrum for 3-(4-chlorophenyl)-*3H*-diazirine (**6c**)

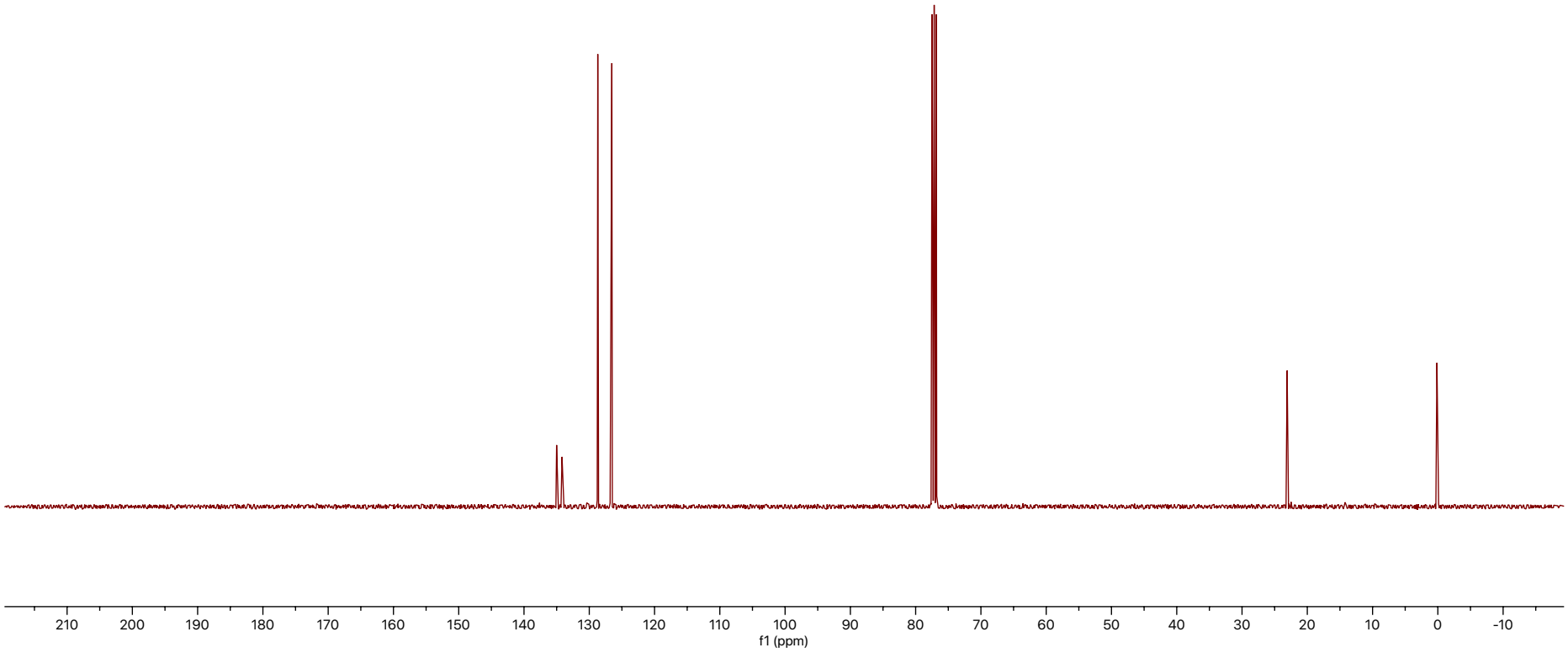
400 MHz, $CDCl_3$, 23 °C



^{13}C NMR spectrum for 3-(4-chlorophenyl)-*3H*-diazirine (**6c**)

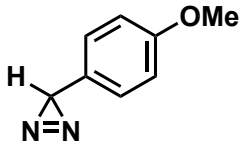
101 MHz, $CDCl_3$, 23 °C





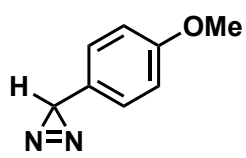
^{1}H NMR spectrum for 3-(4-methoxyphenyl)-*3H*-diazirine (**6d**)

400 MHz, $CDCl_3$, 23 °C



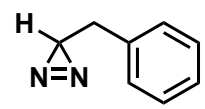
^{13}C NMR spectrum for 3-(4-methoxyphenyl)-*3H*-diazirine (**6d**)

101 MHz, $CDCl_3$, 23 °C



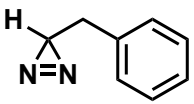
^{1}H NMR spectrum for 3-benzyl-*3H*-diazirine (**6e**)

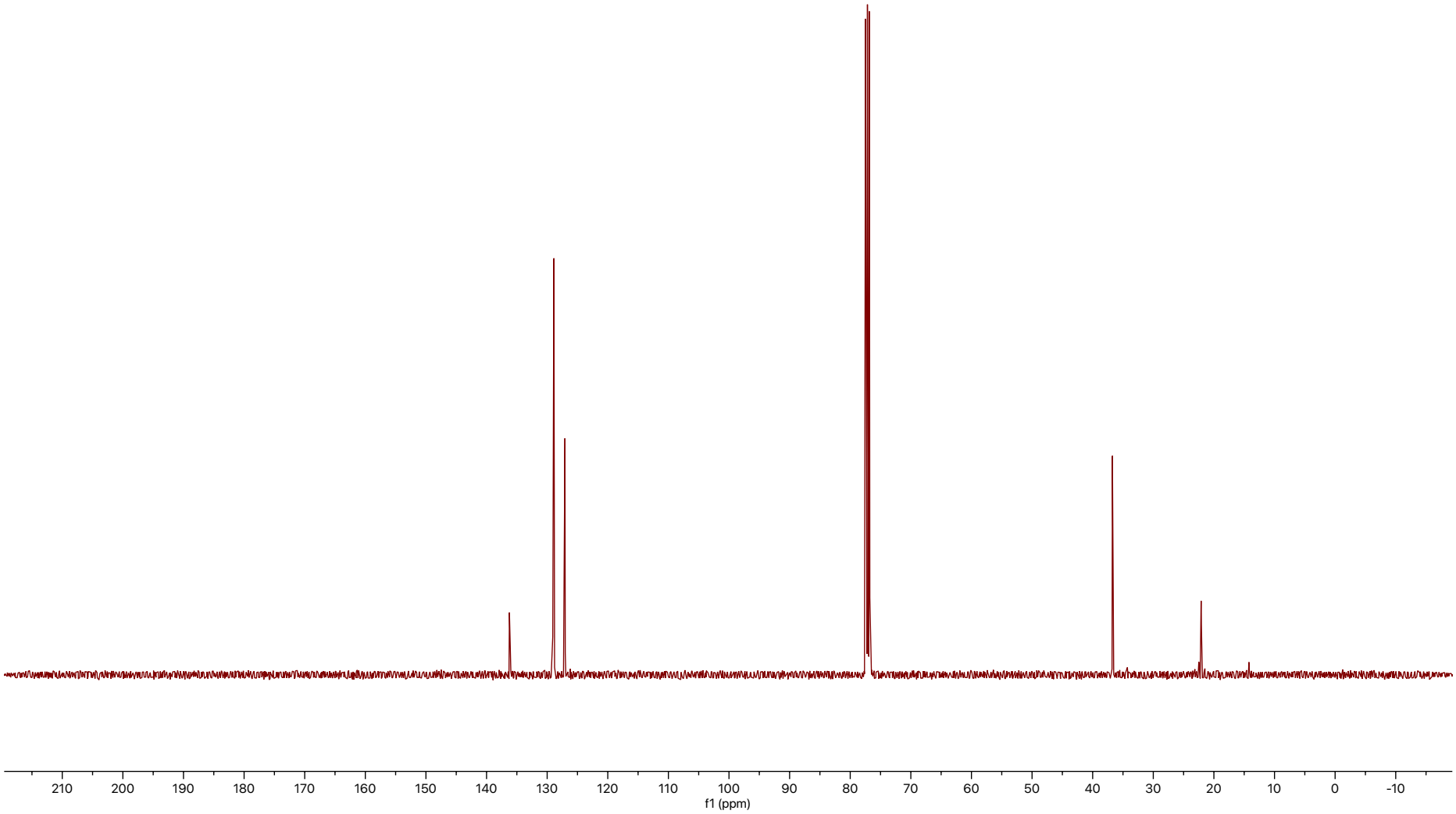
400 MHz, $CDCl_3$, 23 °C



^{13}C NMR spectrum for 3-benzyl-*3H*-diazirine (**6e**)

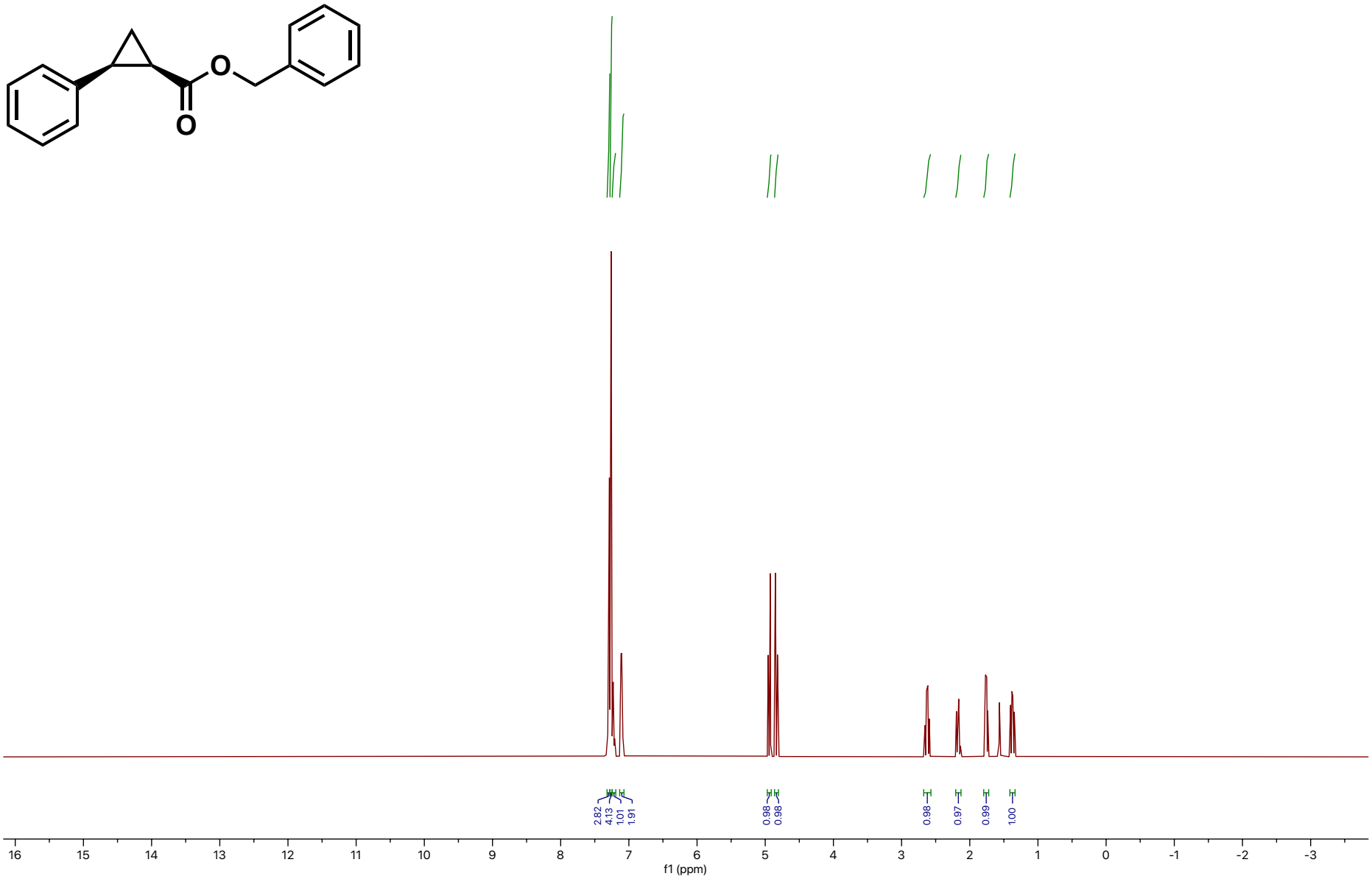
101 MHz, $CDCl_3$, 23 °C





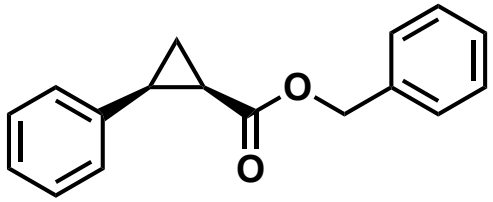
^{1}H NMR spectrum for benzyl *cis*-2-phenylcyclopropane-1-carboxylate (***cis-7a***)

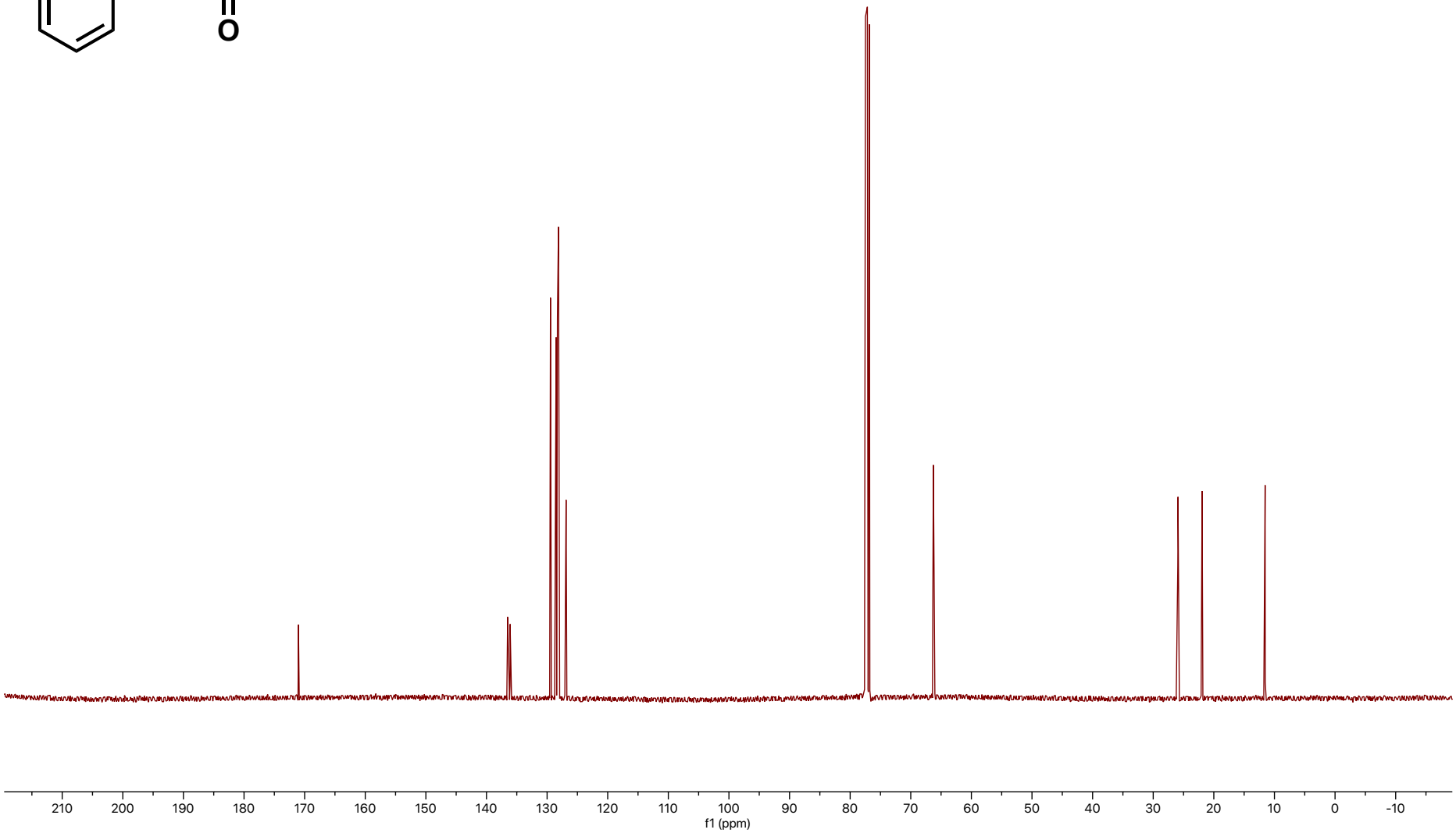
400 MHz, $CDCl_3$, 23 °C



^{13}C NMR spectrum for benzyl *cis*-2-phenylcyclopropane-1-carboxylate (***cis*-7a**)

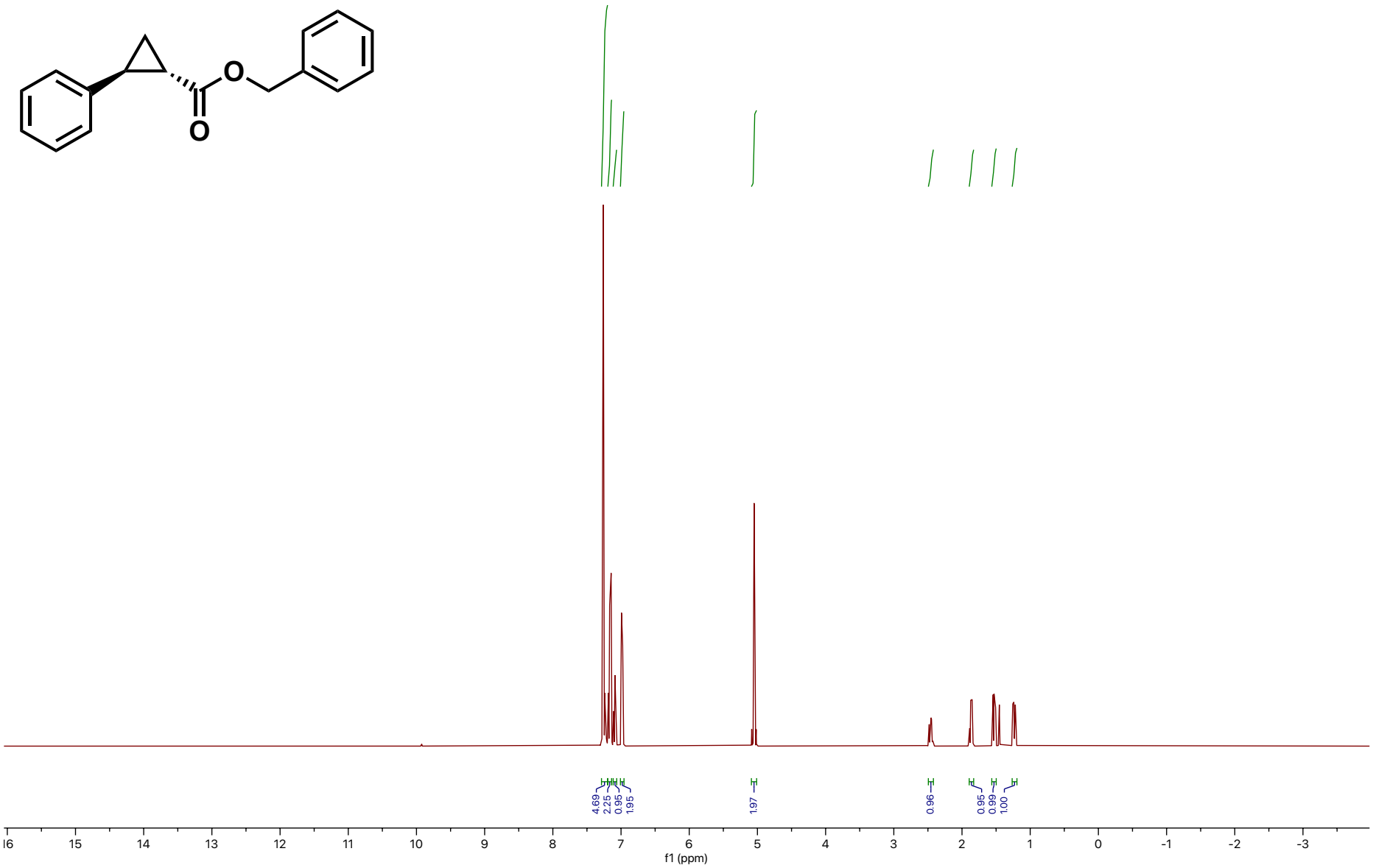
101 MHz, $CDCl_3$, 23 °C





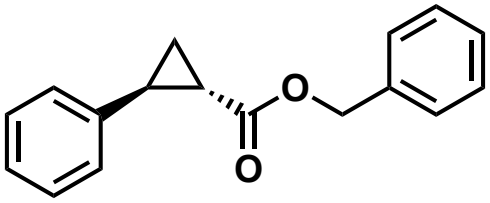
^{1}H NMR spectrum for benzyl *trans*-2-phenylcyclopropane-1-carboxylate (***trans-7a***)

400 MHz, $CDCl_3$, 23 °C



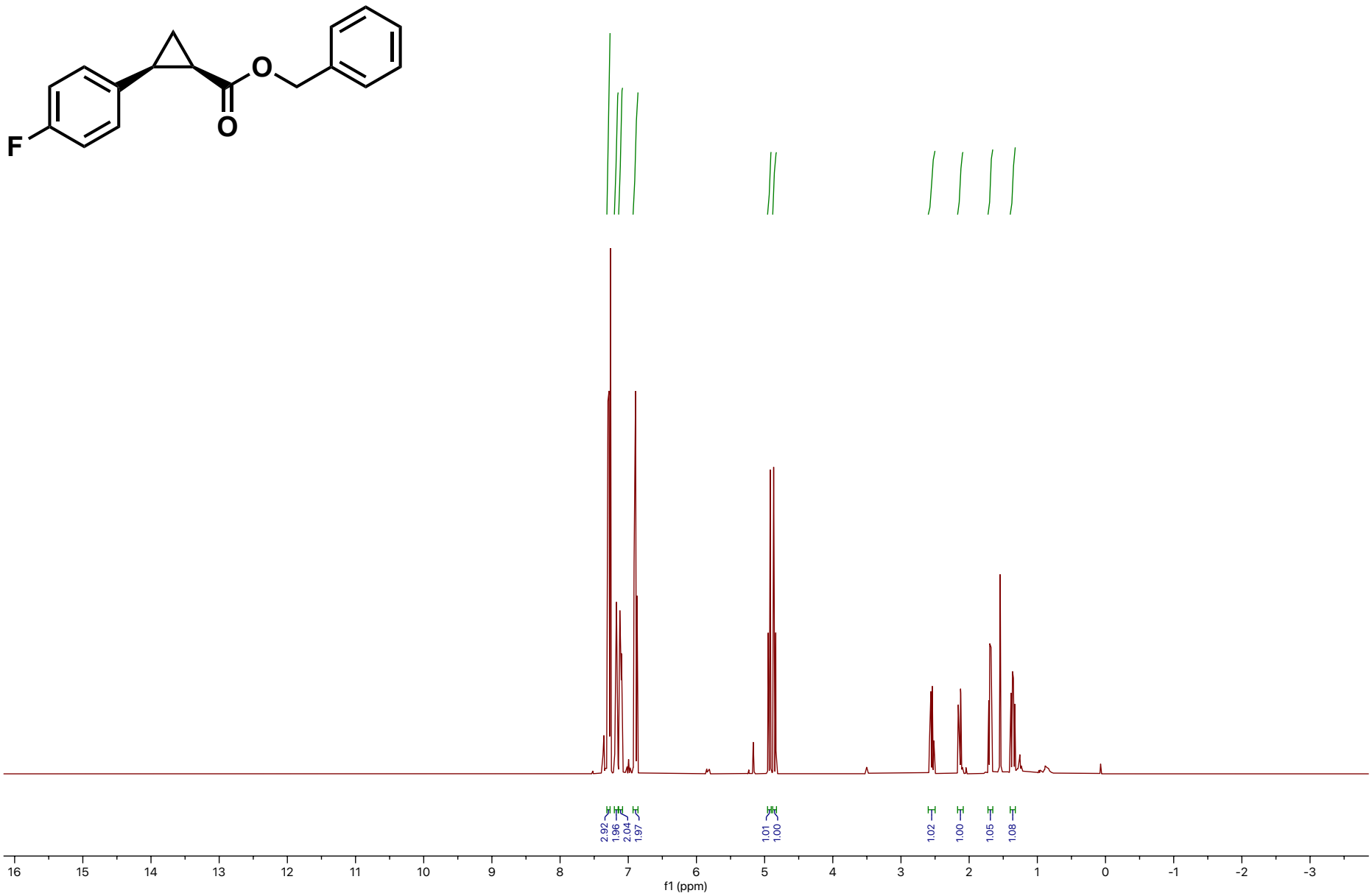
^{13}C NMR spectrum for benzyl *trans*-2-phenylcyclopropane-1-carboxylate (***trans-7a***)

101 MHz, $CDCl_3$, 23 °C



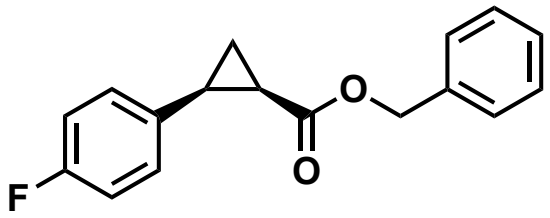
^{1}H NMR spectrum for benzyl *cis*-2-(4-fluorophenyl)cyclopropane-1-carboxylate (***cis*-7b**)

400 MHz, $CDCl_3$, 23 °C



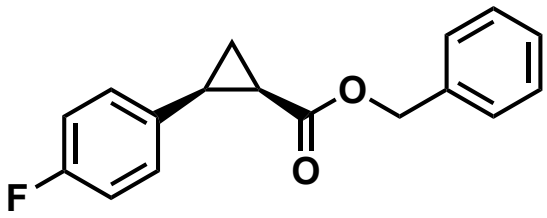
^{13}C NMR spectrum for benzyl *cis*-2-(4-fluorophenyl)cyclopropane-1-carboxylate (***cis*-7b**)

101 MHz, $CDCl_3$, 23 °C



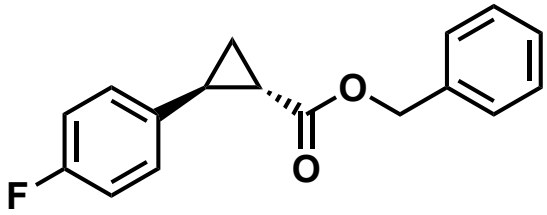
^{19}F NMR spectrum for benzyl *cis*-2-(4-fluorophenyl)cyclopropane-1-carboxylate (***cis*-7b**)

282 MHz, $CDCl_3$, 23°C



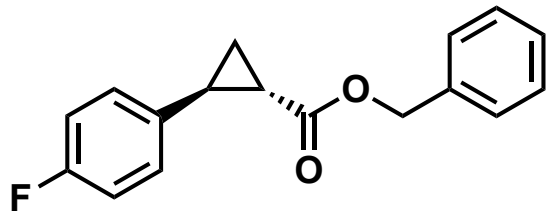
^{1}H NMR spectrum for benzyl *trans*-2-(4-fluorophenyl)cyclopropane-1-carboxylate (***trans*-7b**)

400 MHz, $CDCl_3$, 23 °C



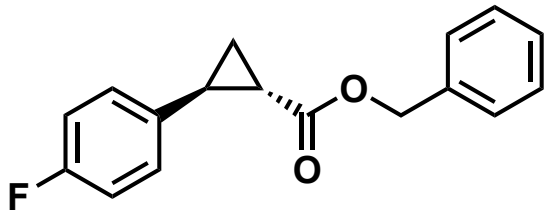
^{13}C NMR spectrum for benzyl *trans*-2-(4-fluorophenyl)cyclopropane-1-carboxylate (***trans*-7b**)

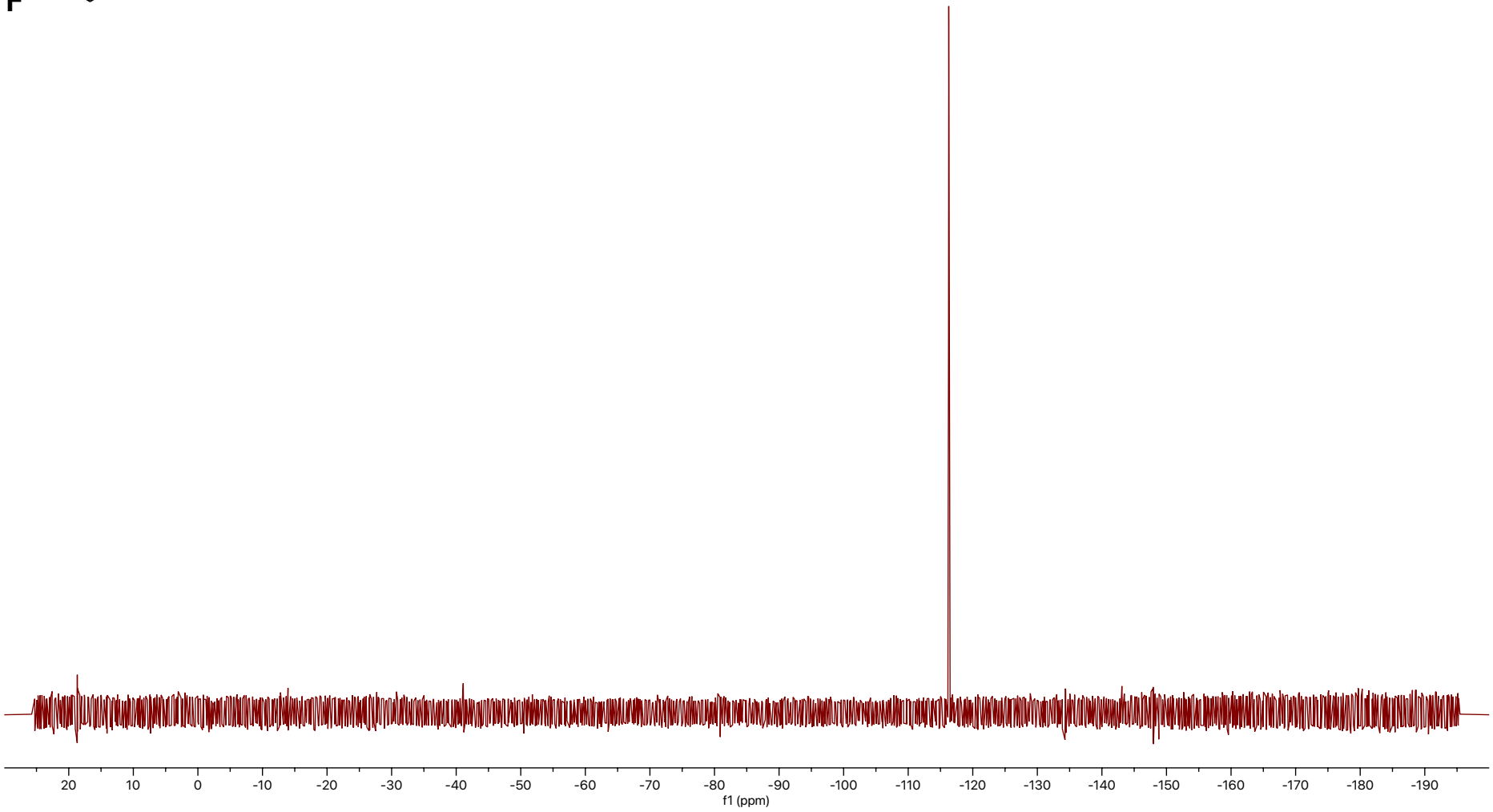
101 MHz, $CDCl_3$, 23 °C



^{19}F NMR spectrum for benzyl *trans*-2-(4-fluorophenyl)cyclopropane-1-carboxylate (***trans*-7b**)

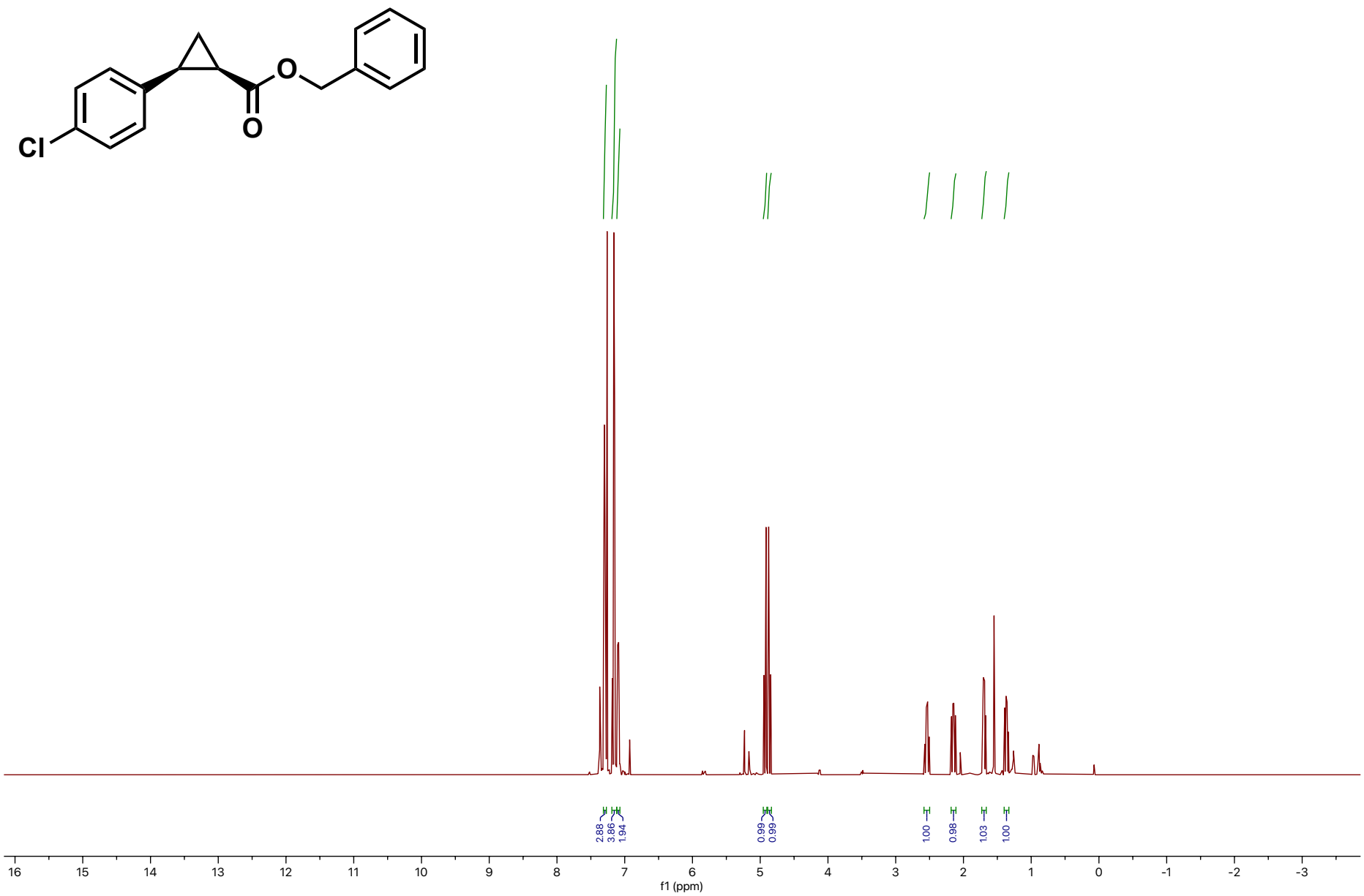
282 MHz, $CDCl_3$, 23 °C





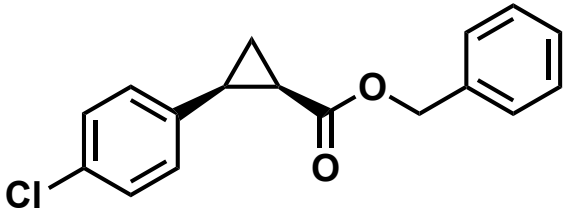
^{1}H NMR spectrum for benzyl *cis*-2-(4-chlorophenyl)cyclopropane-1-carboxylate (***cis*-7c**)

400 MHz, $CDCl_3$, 23 °C



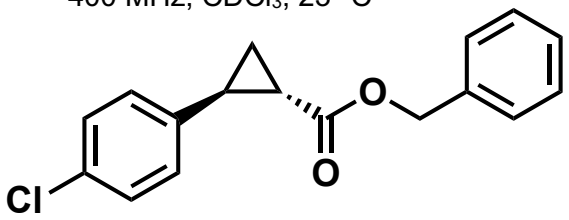
^{13}C NMR spectrum for benzyl *cis*-2-(4-chlorophenyl)cyclopropane-1-carboxylate (***cis*-7c**)

101 MHz, $CDCl_3$, 23 °C



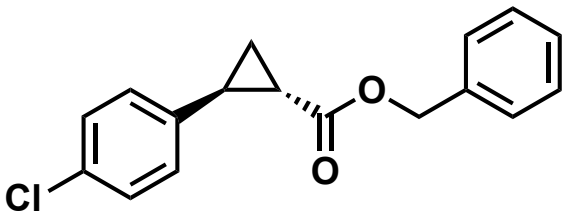
^{1}H NMR spectrum for benzyl *trans*-2-(4-chlorophenyl)cyclopropane-1-carboxylate (***trans-7c***)

400 MHz, $CDCl_3$, 23 °C



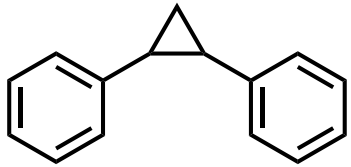
^{13}C NMR spectrum for benzyl *trans*-2-(4-chlorophenyl)cyclopropane-1-carboxylate (***trans-7c***)

101 MHz, $CDCl_3$, 23 °C



^{1}H NMR spectrum for 1,2-diphenylcyclopropane (**9**)

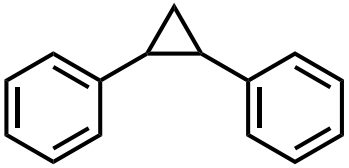
400 MHz, $CDCl_3$, 23 °C



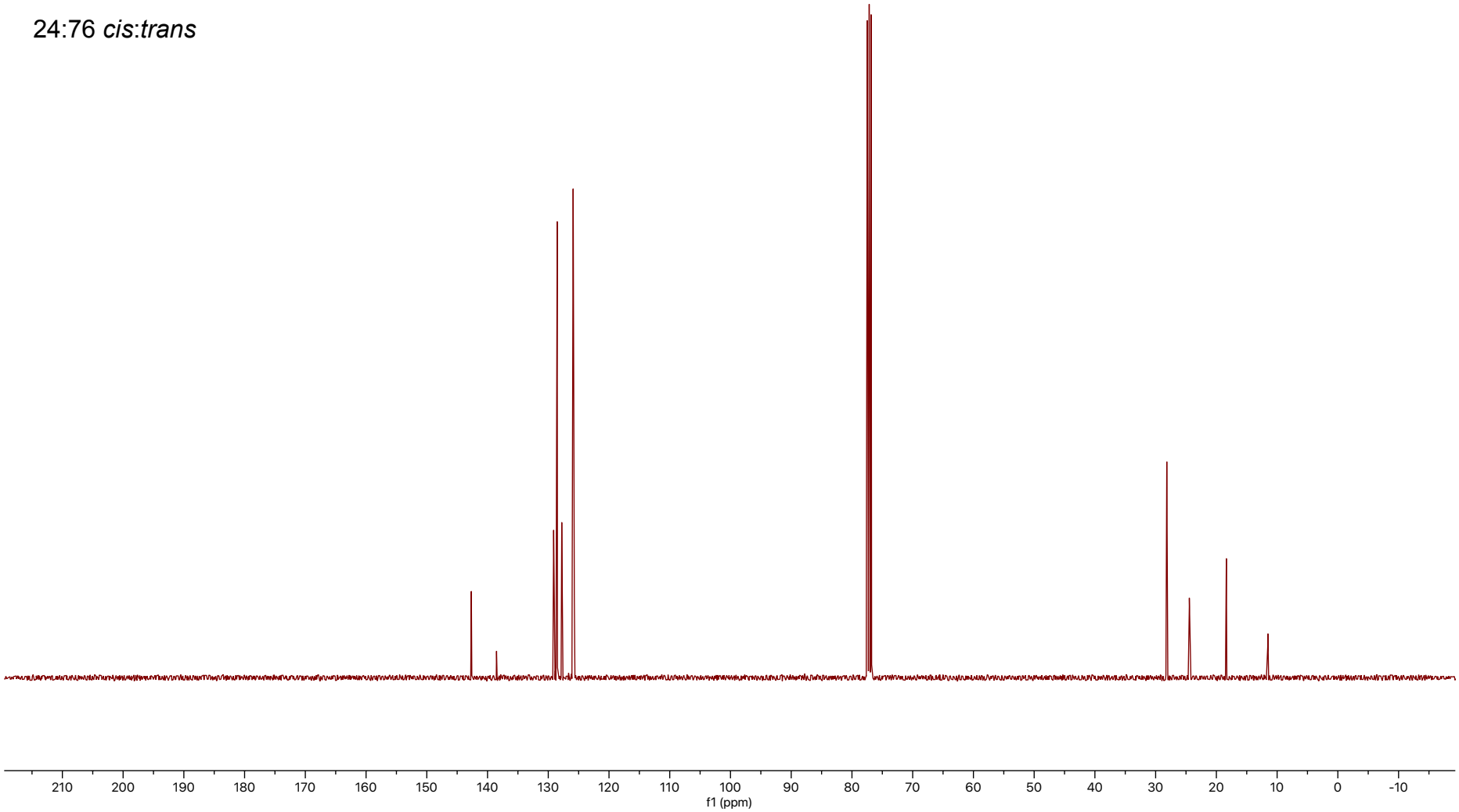
24:76 *cis*:*trans*

^{13}C NMR spectrum for 1,2-diphenylcyclopropane (**9**)

101 MHz, $CDCl_3$, 23 °C

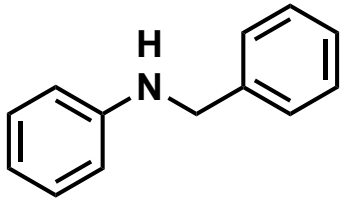


24:76 *cis:trans*



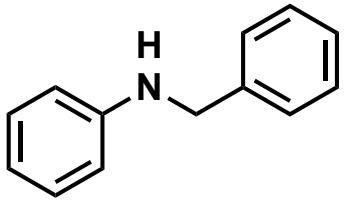
^{1}H NMR spectrum for *N*-benzylaniline (**11**)

400 MHz, $CDCl_3$, 23 °C



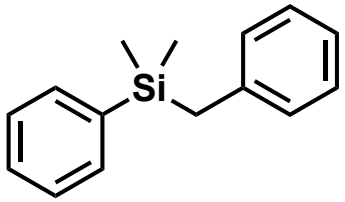
^{13}C NMR spectrum for *N*-benzylaniline (**11**)

101 MHz, $CDCl_3$, 23 °C



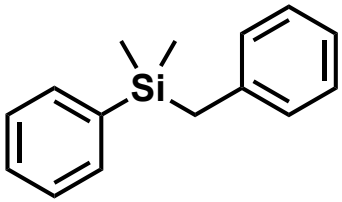
^{1}H NMR spectrum for benzyldimethyl(phenyl)silane (**13**)

400 MHz, $CDCl_3$, 23 °C



^{13}C NMR spectrum for benzyldimethyl(phenyl)silane (**13**)

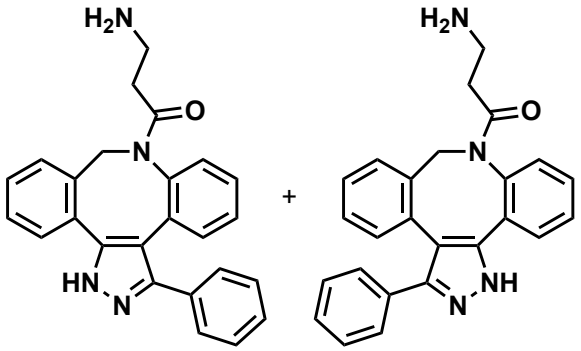
101 MHz, $CDCl_3$, 23 °C



^{1}H NMR spectrum for 3-amino-1-(3-phenyl-1,9-dihydro-8*H*-dibenzo[*b*,*f*]pyrazolo[4,3-*d*]azocin-8-yl)propan-1-one (**S2a**) +

3-amino-1-(3-phenyl-1,8-dihydro-9*H*-dibenzo[*b*,*f*]pyrazolo[3,4-*d*]azocin-9-yl)propan-1-one (**S2b**)

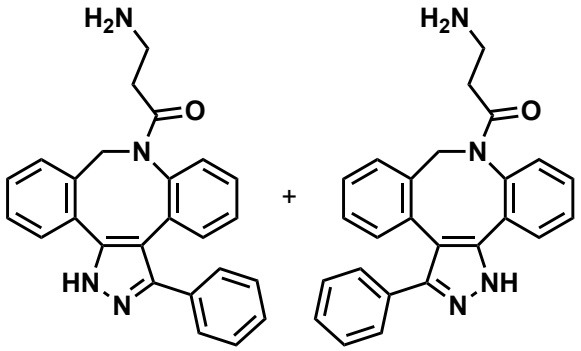
400 MHz, $CDCl_3$, 23 °C



^{13}C NMR spectrum for 3-amino-1-(3-phenyl-1,9-dihydro-8*H*-dibenzo[*b*,*f*]pyrazolo[4,3-*d*]azocin-8-yl)propan-1-one (**S2a**) +

3-amino-1-(3-phenyl-1,8-dihydro-9*H*-dibenzo[*b*,*f*]pyrazolo[3,4-*d*]azocin-9-yl)propan-1-one (**S2b**)

101 MHz, $CDCl_3$, 23 °C



+

References

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.
2. Kan, S. B. J.; Huang, X.; Gumulya, Y.; Chen, K.; Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **2017**, *552*, 132-136.
3. Knight, A. M.; Kan, S. B. J.; Lewis, R. D.; Brandenberg, O. F.; Chen, K.; Arnold F. H. Diverse Engineered Heme Proteins Enable Stereodivergent Cyclopropanation of Unactivated Alkenes. *ACS Cent. Sci.* **2018**, *4*, 372-377.
4. Barr, I.; Guo, F. Pyridine hemochromagen assay for determining the concentration of heme in purified protein solutions. *Bio-protocol* **2015**, *5*, e1594.
5. Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **2012**, *7*, 1511-1522.
6. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 133-144.
7. Evans, P. R., Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204-1214.
8. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
9. Mirdita, M.; Schütze, K.; Morikawi, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold – Making protein folding accessible to all. *bioRxiv*, DOI: https://doi.org/10.1101/2021.08.15.456425.
10. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H. Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68*, 352-367.
11. Pesce, A.; Tilleman, L.; Donné, J.; Aste, E.; Ascenzi, P.; Ciaccio, C.; Coletta, M.; Moens, L.; Viappiani, C.; Dewilde, S.; Bolognesi, M.; Nardini, M. Structure and Haem-Distal Site Plasticity in *Methanosarcina acetovorans* Protoglobin. *PLoS One*, **2013**, *8* (6), e66144.
12. Pesce, A.; Bolognesi, M.; Nardini, M. Protoglobin: structure and ligand-binding properties. *Adv. Microb. Physiol.* **2013**, *63*, 79-93.

13. Gardner, S.; Kawamoto, T.; Curran, D. P. 1,3-dimethylimidazoyl-2-ylidene borane. *Org. Synth.* **2015**, *92*, 342-355.
14. Barluenga, J.; Quiñones, N.; Tomás-Gamase, M.; Cabal, M.-P. Intermolecular Metal-Free Cyclopropanation of Alkenes Using Tosylhydrazones. *Eur. J. Org. Chem.* **2012**, *12*, 2312-2317.
15. Emer, E.; Twilton, J.; Tredwell, M.; Calderwood, S.; Collier, T. L.; Liégault, B.; Taillefer, M.; Gouverneur, V. Diversity-oriented approach to CF_3CHF-, CF_3Br-, CF_3CF_2, $(CF_3)_2CH$-, and $CF_3(SCF_3)CH$-substituted arenes from 1-(diazo-2,2,2-trifluoroethyl)arenes. *Org. Lett.* **2014**, *16*, 6004-6007
16. Glachet, T.; Marzag, H.; Rosa, N. S.; Colell, J. F. P.; Zhang, G.; Warren, W. S.; Franck, X.; Theis, T.; Reboul, V. Iodonitrene in Action: Direct Transformation of Amino Acids into Terminal Diazirines and $^{15}N_2$-Diazirines and Their Application as Hyperpolarized Markers. *J. Am. Chem. Soc.* **2019**, *141* (34), 13689-13696.
17. Wang, Y.; Wen, X.; Cui, X.; Wojtas, L.; Zhang, X. P. Asymmetric Radical Cyclopropanation of Alkenes with In Situ-Generated Donor-Substituted Diazo Reagents via Co(II)-Based Metalloradical Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (3), 1049-1052.