



Supplementary Materials for

Structural and mechanistic bases for a potent HIV-1 capsid inhibitor

Stephanie Bester^{1,†}, Guochao Wei^{1,†}, Haiyan Zhao^{2,†}, Daniel Adu-Ampratwum³, Naseer Iqbal²,
Valentine V. Courouble⁴, Ashwanth C. Francis⁵, Arun S. Annamalai¹, Parmit K. Singh^{6,7},
Nikoloz Shkriabai¹, Peter Van Blerkom², James Morrison¹, Eric M. Poeschla¹, Alan Engelman^{6,7},
Gregory B. Melikyan⁵, Patrick R. Griffin⁴, James R. Fuchs³, Francisco Asturias^{2,*}, Mamuka
Kvaratskhelia^{1,*}

*Correspondence to: (Mamuka Kvaratskhelia) mamuka.kvaratskhelia@cuanschutz.edu, Tel: +1 303 724 3862; (Francisco Asturias) francisco.asturias@cuanschutz.edu, Tel: +1 303.724.1624.

This PDF file includes:

Materials and Methods

Figs. S1-S53

Tables S1-S5

Caption for Movies S1-S10

¹H, ¹³C, and HSQC NMR spectra of GS-6207 and synthetic intermediates

Other Supplementary Materials for this manuscript includes the following:

Movies S1-S10

Materials and Methods

Cells

HEK293T, HEK293T/17, HeLa, and HeLa-derived TZM-bl cells (from NIH AIDS Reference and Reagent Program) were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM, Mediatech, Manassas VA) supplemented with 10% fetal bovine serum (FBS, Sigma, St. Louis, MO) and 100 U/ml penicillin-streptomycin (PS, Gemini Bio-Products, Sacramento, CA). The growth medium for HEK293T/17 was supplemented with 0.5 mg/ml G418 sulfate (Mediatech, Manassas VA). Jurkat, MT4 and Thp 1 cells were cultured in RPMI 1640 medium with 10% FBS and 1% PS at 37 °C. Thp 1 cells were differentiated by incubation with 100 ng/ml phorbol-12-myristate-13-acetate (PMA) for 48 h (27). All cell lines were cultured in a 37 °C incubator supplemented with 5% CO₂, and were passaged at regular intervals. PBMCs were obtained from whole blood of healthy donors under Colorado Multiple Institutional Review Board (COMIRB) approved protocol 15-0181. PBMCs were isolated by density gradient centrifugation using Ficoll-Paque PLUS (GE Healthcare) and cultured in RPMI-1640 with 10% FCS. PBMCs were initially activated with 3 µg/mL PHA-E (Sigma) and 20 U/mL IL-2 (R&D Systems) and additional 20 U/mL IL-2 was added every other day. Mycoplasma were tested routinely by using Mycoscope Mycoplasma PCR detection kit (Genlantis) and no contamination was detected.

Plasmids, Viral Vectors, and Reagents

Single amino acid substitutions in CA-coding sequence (E45A, N57S, M66I, Q67H, K70A, N74D and E180A) were introduced in pNL4-3E-R+luc (28) using QuikChange XL Site-directed mutagenesis KIT (Aligent). The expression plasmid pMD.G (29) encoding the vesicular stomatitis virus G envelope was used to produce pseudotype viruses. The pCypA-DsRed and Vpr-INmNG expressing plasmids have been described previously (5, 6, 30). Cyclosporin A (CsA) was obtained from Calbiochem (Burlington, MA), dissolved in DMSO at 50 mM and stored in aliquots at -20 °C. Cy5-conjugated anti-mouse antibody was from SouthernBiotech (Birmingham, AL). Phosphate buffered saline containing Mg²⁺/Ca²⁺ (dPBS) and Mg/Ca-free (PBS) was purchased from Corning (MediaTech, Manassas, VA). Hoechst33442 (#62249) was from ThermoScientific. The following reagents were obtained from the NIH AIDS Reference and Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl cells expressing CD4, CXCR4 and CCR5 from Drs. J.C. Kappes and X. Wu; anti-p24 antibody AG3.0 donated by Dr. J. Alan.

Virus Production

All viruses were produced from transfected HEK293T cells with pMD.G and WT or mutants based on pNL4-3E-R+luc by using X-treme Gene HP (Roche). Medium was replaced after 12 h post-transfection and virus-containing supernatant was collected after 60 h post-transfection. Supernatant was clarified by using a 0.45- μ m syringe filter and stored at -80 °C. The virus was quantified by p24 antigen ELISA (Zeptometrix).

DNA Extraction and q-PCR

Quantification of HIV-1 reverse transcription products, 2-LTR circles, and integrated proviruses were analyzed by q-PCR. The primers and qPCR conditions were previously described (31). Briefly, VSV-G pseudotyped HIV virions produced from HEK293T cells were quantified by HIV-1 P24 ELISA (ZeptoMetrix) and treated with 60 U/ml DNase I (Ambion) to remove the residual plasmid. 10^6 HEK293T cells/well in 6-well plate were infected with VSV-G pseudotyped HIV virions (2 μ g of p24). The cells were collected at 12 h post-infection. Total genomic DNA was isolated by using DNeasy Blood& Tissue kit (Qiagen). All the samples were normalized to GAPDH and $2^{-\Delta\Delta C_t}$ method was used for relative quantification analysis (32).

For cellular fractionation and DNA quantification, cells were washed 3 times and incubated with 200 μ l fractionation buffer (20 mM HEPES, pH 7.4, KCl 10 mM, 2 mM MgCl₂, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, PI cocktail) on ice for 15 min. Cell suspension was passed through a 27 gauge needle 10 times and left on ice for 20 min. Samples were centrifuged at 720 g for 5 min and the supernatant containing the cytoplasm, membrane and mitochondria was collected. The pellet was washed 6 times with 500 μ l fractionation buffer and passed through a 25 gauge needle 10 times following by 10 min centrifugation at 720 g for 10 min. Supernatant was discarded and the pellet which contained the nucleus was collected. DNA was extracted following the instruction of QIAamp DNA Mini Kit (Qiagen). Cytosolic viral DNA was normalized to mitochondrial DNA and nuclear viral DNA was normalized to chromosomal GAPDH DNA. $2^{-\Delta\Delta C_t}$ method was used for relative quantification analysis (32).

Antiviral and Cytotoxicity Assays

For assays of early steps of infection 5×10^4 cells/well were seeded in 24-well plate. After 4 h, indicated concentrations of the inhibitor or DMSO were added in the supernatant and incubated for 30 min. Subsequently cells were infected with VSV-G pseudotyped HIV viruses (250 ng of p24) and incubated at 37 °C for 3 h. Supernatants were removed and cells were washed with PBS once. Fresh complete medium containing indicated concentrations of the inhibitor or DMSO was added and cells were cultured for 48 h. Subsequently cells were lysed with 1 x reporter lysis buffer (Promega). Cell debris were removed by centrifugation, the luciferase activity in cellular extracts was determined using Luciferase Assay System (Promega).

For assays of late steps of HIV-1 replication, virions were produced from transfected HEK293T cells in the presence of indicated concentrations of GS-6207 or DMSO. The same volume of filtered virions was used to infect target cells in the absence of GS-6207.

For assays of the complete replication cycle, virions were produced from transfected HEK293T cells in the presence of indicated concentrations of GS-6207 or DMSO. The same volume of filtered virions was used to infect target cells in the presence of matching concentrations of GS-6207. The activity of luciferase was measured as noted above.

Cell viability was determined using the Vybrant MTT Cell Proliferation Assay Kit (Thermo Fisher Scientific). 1×10^5 PBMC/MT4 cells were plated per well in triplicate and treated with GS-6207 or DMSO at a final concentration of 0.5% DMSO for all wells in RPMI-1640 media without phenol red. 72 h later the microplate was centrifuged to pellet cells, then media was removed and replaced with 100 μ L fresh media. 10 μ L of 12 mM MTT stock solution was added to each well and cells were incubated for 2 h at 37 °C. 100 μ L SDS-HCl solution was added to each well and the plate was incubated for an additional 8 h at 37 °C. Wells were mixed gently and absorbance was read at 570 nm.

Immunoblotting

Proteins were separated with Bolt™ 4-12% Bis-Tris Plus Gels (Invitrogen) and transferred to nitrocellulose membranes. The following antibodies were used to detect Nup153 (ab24700, Abcam) and CPSF6 (ab175237, Abcam).

Pseudovirus Production and Characterization

Fluorescently labeled pseudoviruses were produced and characterized as described previously (5). Briefly, HEK293T/17 cells grown in 6-well culture plates were transfected with the following plasmids: pHIVeGFPΔEnv (2 μg), VSV-G (0.2 μg), Vpr-INmNG (0.5 μg) and, where indicated, CypA-DsRed (0.5 μg), using the JetPrime Transfection reagent (VWR, Radnor, PA). Six hours after transfection, the medium was replaced with 2 ml of fresh DMEM/10% FBS without phenol red, and the samples were incubated for additional 36 h at 37 °C, 5% CO₂. Viral supernatant was collected, filtered through a 0.45 μm filter and quantified for p24 content using AlphaLISA immunoassay kit (PerkinElmer, Waltham, MA) or RT activity (RTU) measured using the PERT protocol as described previously (6, 33). Multiplicity of infection (MOI) was determined in TZM-bl cells by measuring the %eGFP expressing cells at 48 h post-infection with VSV-G pseudotyped HIVeGFP virus.

Nuclear Import and HIV-1 Infectivity via Green Fluorescence Reporters

VSV-G pseudotyped HIVeGFP viruses labeled with INmNG were used to infect TZM-bl cells at MOI 1 (5x10⁴ cells/well) on poly-L-lysine treated 8-well chamber slides (LabTek). Prior to virus binding, cell nuclei were stained for 10 min with 2 μg/ml of Hoechst-33342. Infection was synchronized by binding virus to cells through spinoculation for 30 min at 1500×g, 12 °C. The cells were washed twice, and virus entry was synchronously initiated by adding pre-warmed DMEM with indicated concentrations of GS-6207 or DMSO to samples followed by culture in a CO₂ incubator. At 4 h post-infection (hpi), cells were examined on a Zeiss LSM880 confocal microscope for nuclear INmNG complexes using a 63x/1.4NA oil-immersion objective. Z-stack images spaced by 0.5 μm were acquired from 4 fields of view. The number of INmNG puncta in the nuclei of TZM-bl cells was determined off-line using ICY software. After image acquisition, the cells were moved back into a culture incubator for additional 20 h. At 24 hpi, the cells were fixed with 2% PFA for 7 min. The infected cells from 4 fields of view (~300-400 cells/field) were imaged using a 10x air objective. The fraction of eGFP expressing cells per field of view was determined by normalizing to the total number cells identified by Hoechst-33342 nuclear stain.

HIV-1 Capsid Core-Stability Measurements in Vitro and in Infected Cells

VSV-G pseudotyped HIV-1 particles produced in 293T/17 cells were co-labeled with INmNG and CypA-DsRed, as described (6). For *in vitro* core stability measurements, single virus

particles were bound to a poly-L-lysine treated coverglass and mildly permeabilized with saponin (100 µg/ml) for 1 min. Saponin was washed away, and the permeabilized virions were treated with indicated concentrations of GS-6207 and imaged at room temperature on a Zeiss LSM880 confocal microscope. Images were acquired from 4 neighboring fields of view at 1 min intervals. Cyclosporin A (CsA, 5 µM) was added at 31 min after permeabilization/onset of imaging to displace CypA-DsRed from permeabilized viral cores that have not disassembled during the imaging time window.

A live cell-based assay for HIV-1 core stability was performed, as previously described (5). Briefly, VSV-G pseudotyped HIVeGFP particles co-labeled with INmNG and CypA-DsRed were used to infect TZM-bl cells (MOI 0.008). Virus was allowed to bind to 5×10^4 TZM-bl cells adhered to a 8-well chamber slide (Lab-Tek) by spinoculation for 30 min at $1500 \times g$, 12 °C. Prior to virus binding, cell nuclei were stained for 10 min with 2 µg/ml of Hoechst-33342. The cells were washed twice, and virus entry was synchronously initiated by adding pre-warmed live-cell imaging buffer (Invitrogen) supplemented with 10% FBS and with the indicated concentration of GS-6207 or DMSO, followed by incubation in a CO₂ incubator maintained at 37 °C. At 3 hpi, the cells were imaged on a Zeiss LSM880 confocal microscope. Images were acquired from 4 neighboring fields of view 5 min before and after CsA (25 µM) addition in order to discriminate between intact virions and post-fusion cores. CsA treatment results in the quick displacement of CypA-DsRed from post-fusion cores in the cytoplasm, while intact virions in endosomes do not release this marker. The number of INmNG and CypA-DsRed puncta in pre- and post-CsA treatment images was determined using the spot detector plugin in ICY. The number of stable cores residing in the cytoplasm were determined by single particle-based detection of CypA-DsRed loss from INmNG puncta upon CsA addition.

Image Acquisition and Analysis

3D time-lapse live cell imaging was carried out on a Zeiss LSM880 laser scanning confocal microscope using a C-Apo 63x/1.4NA oil-immersion objective. Tile-scanning imaging was performed for 4 or 9 adjacent fields of view using line averaging along with stringent axial sampling. For visualizing the nuclear HIV-1 IN-complexes (nuclear import) or core-stability in the cytoplasm, Z-stacks spaced by 0.3 or 0.5 µm, respectively, were acquired. To analyze eGFP expressing infected cells, single Z-stack imaging was performed using a 0x air objective from 4 or

9 fields of view using tile-scanning. INmNG, CypA-DsRed and Hoechst 33342 fluorescence was excited using 488, 561 and 405 nm laser lines, respectively. 3D-image series were processed off-line using ICY image analysis software (<http://icy.bioimageanalysis.org/>).

In order to discriminate between cytoplasmic and nuclear IN-spots, an in-house protocol was created using the ICY protocols module. Briefly, the nuclear volume in three dimensions was detected using the Hoechst nuclei intensity by the HK-means and the connected components plugins. To avoid detection of nuclear membrane-associated IN spots, the obtained 3-dimensional region of interest (ROI) corresponding to the nuclear volume was shrunk by 0.5 μm in X-Y-Z using an ROI-erosion plugin. The IN complexes within the eroded ROI were considered intranuclear, while the remaining spots were deemed to reside in the cytoplasm. The number of INmNG complexes per nucleus was determined for >30 nuclei (from 4 random fields of view) for each independent experiment. The % of infection was determined by counting the fraction of eGFP expressing cells per field of view and normalizing to the total number of cells determined by Hoescht 33342 nuclear staining.

For *in vitro* and cellular core-stability measurements, the number of INmNG and CypA-DsRed puncta was detected, using the wavelet spot detector in ICY. The kinetics of CA disassembly *in vitro* was determined by normalizing the number of CypA-DsRed puncta to that at $t=0$ min and subtracting the background CsA-resistant CypA-DsRed puncta at $t=31$ min presumably corresponding to immature viruses. The total number of INmNG puncta in each field of view remained constant over the 30 min imaging period. The number of stable post-fusion HIV-1 cores in the cytoplasm of living cells was determined by estimating the number of CypA-DsRed puncta per cell that were lost after CsA treatment.

Integration Site Sequencing

The preparation of samples for integration site sequencing was performed as described previously with some modifications (34). Briefly, at day 0 HEK293T cells were seeded (2×10^5 cells/well) in a 6-well plate. On day 1, the cells were treated with different concentration of GS-6207 (0.5, 1 and 2 nM; three wells per each concentration) for 1 h. The control cells were treated with DMSO. One h later, cells were infected with VSV-G pseudotyped HIV luciferase virus (pNL4-3E-R+luc) (2000 ng of p24) and incubated for 2 h at 37 °C. 2 h later the medium was removed and replaced with fresh medium containing the matching concentration of the inhibitor.

Cells were harvested 48 h post infection and 30% cells were used to measure luciferase activity with Luciferase Assay System (Promega). The remainder of cells were cultured for additional 5 days in 150 mm cell culture dish in normal media without the inhibitor. The cells were harvested, genomic DNA was prepared using QIAamp DNA Mini Kit (Qiagen) and digested with MseI and BglII. Purified DNA fragments were ligated to adapters and used for U5-based PCR amplification of virus-cell junctions as described (13, 35). PCR amplicons were sequenced on the Illumina platform by Genewiz. Paired - FASTQ files were processed as described (36). The values for the random integration control (RIC), wild-type and CPSF6 knockout HEK293T cells were previously published (13). Statistical analysis was done as described (35).

Recombinant Proteins

WT and A92E HIV-1 CA proteins were expressed from pET3a-CA construct in *E. coli* and purified as described (37). The quadruple CA mutant (CA_{A14C/E45C/W184A/M185A}) for hexamer formation was expressed from pET11a in *E. coli* and purified as described (38). GST fused CypA protein with TEV protease cleavage site was engineered in the pET3a plasmid, expressed in *E. coli* and purified through two column chromatography steps using 5 ml GSTrap 4B and HiTrap Q High Performance 5 ml columns (GE Healthcare). After elution from the GST column, the eluate was treated with TEV protease at 4 °C for ~12 hs to cleave off the GST tag. The resulting CypA was further purified by applying to HiTrap Q column and collecting the flow-through fractions.

In vitro CA Tube Stability Assay

42.5 μM WT CA nanotubes were assembled in buffer containing 2M NaCl and 25 mM Tris-HCl, pH 8.0 overnight on ice. The assembled CA tubes were then mixed with GS-6207 at a 1:1 ratio. The control reactions contained 4% DMSO. The final reaction condition was 50 mM Tris-HCl (pH 8.0), 2 M NaCl, 4% DMSO in 40 μL total volume. The reactions were incubated on ice for 30 min and centrifuged at 21,000*g for 5 min at 4 °C. Supernatant was removed and CA pellets were re-suspended in a buffer containing 40 μL of 150 mM NaCl and 50 mM Tris-HCl (pH 8.0). Resuspended CA pellets in low salt were incubated on ice for 0, 1, 4, 24, 48 and 96 h. The samples from each time point were centrifuged for 5 min at 21,000*g at 4 °C and resulting pellets were analyzed by SDS-PAGE after visualizing with AcquaStain.

Stoichiometry Determination

To determine the GS-6207 to CA ratio needed to stabilize pre-formed curved CA lattice, increasing concentrations of the compound were added to 40 μM preassembled (2 M NaCl buffer at 37 °C for 1 h) WT CA tubes. The final reaction condition was 50 mM Tris-HCl (pH 7.5), 2 M NaCl, 4.3% DMSO in 27.7 μL total volume. The reactions were incubated at room temperature for 60 min and then diluted 10-fold with ice cold 50 mM Tris-HCl (pH 8.0) solution to the final NaCl concentration of 200 μM . After 5 min incubation the samples were centrifuged at 21,000*g for 5 min. The supernatant was removed and the pellet fractions were run on SDS-PAGE and visualized with BlueFast (Eton Bioscience Inc.). CA bands were quantitated by ImageJ software (NIH). The data points were plotted by piecewise linear regression in Origin software (OriginLab Corp.) and the stoichiometry for GS-6207 binding to CA hexamers was calculated as described (https://s3-us-west-2.amazonaws.com/oww-files-public/7/7b/FAQ_Stoichiometry_V03-2.pdf).

Effects of CypA on CA Tube Stability

WT CA tubes were assembled by incubation in 2 M NaCl containing buffer at 37 °C for 1 hr. The assembled CA tubes were mixed with GS-6207 at a 1:1 ratio (42.5 μM of each). In control experiments, DMSO was added to pre-assembled CA tubes. This was followed by addition of increasing concentrations of CypA (or buffer alone for the control samples). The final reaction condition was 50 mM Tris-HCl (pH 8.0), 2 M NaCl, 5% DMSO in 20 μL total volume. The reactions were incubated on ice for 30 min and centrifuged at 21,000*g for 5 min at 4°C. Supernatant was removed and pellet fractions were run on SDS-PAGE and visualized with AcquaStain (Bulldog Bio, Inc.).

Thermal Shift Assay

A QuantStudio 3 Real-Time PCR System (Applied Biosystems) was used to measure effects of GS-6207 on the thermal stability of cross-linked CA hexamers. The thermal shift assay (39, 40) is based on the denaturation of a protein in the presence or absence of a bound ligand where the exposure of hydrophobic residues can be detected with the fluorescent dye, Sypro orange (Invitrogen). The melting temperature (T_m) is calculated from this data. After incubation of cross-linked CA hexamers with GS-6207 for 10 min at 37 °C, Sypro orange dye was added. The mixture was subsequently heated in a QuantStudio 3 Real-Time PCR System from 35-80 °C in increments

of 0.05°C/s. Changes in protein thermostability were analyzed by using QuantStudio 3 software provided by the manufacturer. All assays were performed in triplicate.

X-ray Crystallography

The cross-linked CA_{A14C/E45C/W184A/M185A} hexamers were prepared as described (38). After assembly, the hexamers were purified by size exclusion chromatography using a GE Healthcare HiLoad 16/600 Superdex 200 pg column and a buffer containing 20 mM Tris-HCl, pH 8.0 and 150 mM NaCl. Non-reducing SDS-PAGE was utilized to detect the CA_{A14C/E45C/W184A/M185A} hexamers. Once the purified hexamer was diluted with the aforementioned buffer to reach 1 μM, GS-6207 was added over time to reach 1 μM. The solution containing pre-formed CA_{A14C/E45C/W184A/M185A} hexamer and GS-6207 was then concentrated to 10 mg/mL of hexamer with a 50 kDa molecular weight cutoff Amicon Ultra-15 Centrifugal concentrator. Crystals of CA_{A14C/E45C/W184A/M185A} hexamer in complex with GS-6207 were grown by hanging drop vapor diffusion at 4 °C with an equal volume of crystallization buffer. The crystallization buffer contained 3% PEG3350, 0.125 M NaI, 0.1 M sodium cacodylate pH 6.5, and 6% glycerol. The crystals were hexagonal rods and appeared within two weeks. The cryogenic solution for the crystals consisted of crystallization buffer with additional PEG3350 and Glycerol to reach 20 and 10%, respectively. Crystals were flash cooled in liquid nitrogen. The data was collected at the Advanced Light Source, Beamline 4.2.2 (Macromolecular Crystallography; MBC) at 100 K and a wavelength of 1.00003 Å. The data was processed and scaled using XDS (41). REFMAC (42) and MOLREP (43) within HKL3000 (44) were used to perform molecular replacement using PDB 6PU1 as a search model. The structure was refined using repeated cycles of model building and refinement via COOT (45) and phenix.refine (46), respectively. TLS Motion Determination (TLSMD) was utilized to analyze the structure's flexibility. This analysis provided TLS parameters in phenix.refine that assisted in refining the structure for anisotropic displacements (46). Ligands were individually oriented into structures based on the $F_o - F_c$ omit map density at 3 σ and refined with phenix.refine afterwards to ensure they fit the $2F_o - F_c$ density at 1 σ (46). Water molecules were originally added through the Find Water COOT program, but each water was then evaluated individually to ensure they fit the $2F_o - F_c$ density at 1 σ (47). Molprobit was used to assess the final model of each structure and ensure their quality (48). Final Ramachandran statistics are 98% favorable and 1.8% disallowed.

The coordinates are deposited in the Protein Data Bank under accession code 6VKV. The data collection and refinement statistics are given in Table S2.

Cryo-EM Analysis

An 80 μ L aliquot of CA tubes was prepared by incubating A92E CA (3 mg/ml) overnight at 4 °C in 1 M NaCl, 50 mM Tris pH 8.0. Pre-assembled A92E tubes were incubated with GS-6207 (42 μ M) on ice for half an hour, followed by dialysis with 250 mL of buffer containing 50 mM Tris, 150 mM NaCl, 0.01% NP-40 and 2% glycerol for 50 min at 4 °C. After dialysis, a 2.5 μ L aliquot of the inhibitor-stabilized tube suspension was applied to lacey carbon grids (Ted Pella) that had been plasma cleaned for 6 s on a Solarus Plasma cleaner (Gatan) using an Ar/O₂ gas mixture. Samples were vitrified in liquid ethane using a Vitrobot Mark IV (ThermoFisher) and stored in liquid nitrogen. Cryo-EM specimens were imaged on a Talos Arctica transmission electron microscope (ThermoFisher) operating at 200 keV using 28,000 x magnification. Images were recorded on a K3 Summit direct electron detector (Gatan). Details for cryo-EM data collection are listed in Table S5.

A total of 4246 micrographs were recorded. Motion correction (MotionCor2) and CTF parameter estimation (ctffind4) were done in Relion 3.0.4 (49). A total of 146,291 helical segments were extracted from manually selected, well-ordered helical tubes using Relion. Tube segments were subjected to multiple rounds of 2D clustering and further analysis was focused on classes with the best layer lines. After selection of a homogeneous subset of tube segments, an initial volume was calculated using Relion's refine 3D without applying helical symmetry and using a featureless cylindrical shell as initial reference. Initial helical parameters were estimated from the resulting un-symmetrized volume using tools in relion_helix_tool_box and checked visually by comparing a helical model (also generated using tools in relion_helix_tool_box) to the initial volume obtained without applying symmetry.

After finding initial helical parameters (twist = 150.2°, rise = 7.5Å), 208,225 tube segments were re-extracted into 400 x 400 pixel boxes with an overlap of ~80% (10 asymmetric unit shift) and segments were subjected to many rounds of 2D clustering. Segments having the best helical layer lines were selected, the micrographs which contributed to these segments (686) were identified and 42,758 segments were re-extracted. The extracted segments were subjected to further 2D clustering and a subset of 40,679 segments were selected for 3D refinement using

Relion, yielding a 3D map with 7.4 Å resolution. Two rounds of 3D classification were then performed, which generated two good classes whose particles were selected and further refined. Finally, 24,505 segments were identified and used to generate a 3D map with a resolution of 6.3Å based on the Fourier shell Correlation (FSC) value of 0.143 (final helical parameters: twist =150.32°, rise = 7.61Å). This final map was used to build an atomic model of the tube hexamer by rigid body fitting of the GS-6207 bound CA monomer model from X-ray crystallography.

For further validation of the map obtained by helical image analysis, we applied a non-helical approach (RASTR) (20) to calculate a volume from the same subset of data used for helical analysis. The position and rotation to the center of a single hexamer in the final helical volume was determined and used along the final refined helical rise and twist to extract a total of 245,050 unique “particles” from which surrounding tube density was subtracted by reprojection using Relion. The resulting stack of 245,050 particles, each extracted into a 192 x 192 pixel box centered around the chosen asymmetric unit for each subtracted volume, was imported into Cryosparc (50). 2D clustering and *ab-initio* heterogeneous reconstruction with 3 volumes were then used to select a subset of 96,392 (40% of initial dataset) particle images. These images, along with the cleanest reference from *ab-initio* volume determination were then used to obtain a volume through 3D Refinement with a nominal resolution of 6.08 Å (map deposited with EMDB code EMD-21423). The helical and RASTR maps were essentially superimposable, but the RASTR map showed somewhat better definition of the H4 helices in the NTDs of the CA hexamer and weaker density for disordered loops in the N-terminal portion of the CA monomer.

Hydrogen-Deuterium Exchange (HDX) Detected by Mass Spectrometry (MS)

Solution-phase amide HDX experiments were carried out with a fully automated system (CTC HTS PAL, LEAP Technologies, Carrboro, NC; housed inside a 4 °C cabinet) as described (51) with slight modifications. Peptides were identified using tandem MS (MS/MS) experiments performed on a QExactive (Thermo Fisher Scientific, San Jose, CA) over a 70 min gradient. Product ion spectra were acquired in a data-dependent mode and the five most abundant ions were selected for the product ion analysis per scan event. The MS/MS *.raw data files were converted to *.mgf files and then submitted to MASCOT (version 2.3 Matrix Science, London, UK) for peptide identification. The maximum number of missed cleavages was set at 4 with the mass

tolerance for precursor ions +/- 0.6 Da and for fragment ions +/- 8 ppm. Oxidation to methionine was selected for variable modification. Pepsin was used for digestion and no specific enzyme was selected in MASCOT during the search. Peptides included in the peptide set used for HDX detection had a MASCOT score of 20 or greater. The MS/MS MASCOT search was also performed against a decoy (reverse) sequence and false positives were ruled out if they did not pass a 1% false discovery rate. The MS/MS spectra of all the peptide ions from the MASCOT search were further manually inspected and only the unique charged ions with the highest MASCOT score were included in HDX peptide set.

CA tubes were initially formed overnight on ice at 4 °C and then incubated with GS-6207 or DMSO for up to 1 h at 4 °C. For the differential HDX experiments, we compared: 1) WT CA tubes +/- GS-6207 (high salt) in 1.5 M NaCl, 50 mM Tris, pH 8.0; 2) WT CA +/- GS-6207 (low salt) in 100 mM NaCl, 50 mM Tris, pH 8.0; and 3) CA(A92E) +/- GS-6207 (high salt) in 1 M NaCl, 50 mM Tris, pH 8.0. The reactions (5 µl) were mixed with 20 µl of D₂O-containing HDX buffer and incubated at 4 °C for 0 s, 10 s, 30 s, 60 s, 900 s or 3,600 s. Following on-exchange, unwanted forward- or back-exchange was minimized, and the protein was denatured by the addition of 25 µl of a quench solution (0.1 M Na Phosphate, 50 mM TCEP). Samples were then immediately passed through an immobilized pepsin column (prepared in house) at 50 µl min⁻¹ (0.1% v/v TFA, 4 °C) and the resulting peptides were trapped and desalted on a 2 mm × 10 mm C₈ trap column (Hypersil Gold, Thermo Fisher, Grand Island, NY). The bound peptides were then gradient-eluted (4-40 % CH₃CN v/v and 0.3 % v/v formic acid) across a 2.1 mm × 50 mm C₁₈ separation column (Hypersil Gold, Thermo Fisher, Grand Island, NY) for 5 min. Sample handling and peptide separation were conducted at 4 °C. The eluted peptides were then subjected to electrospray ionization directly coupled to a high resolution Orbitrap mass spectrometer (QExactive, Thermo Fisher Scientific, San Jose, CA). Each HDX experiment was carried out in triplicate with a single preparation of each protein-ligand complex. The intensity weighted mean m/z centroid value of each peptide envelope was calculated and subsequently converted into a percentage of deuterium incorporation. This is accomplished by determining the observed averages of the undeuterated and fully deuterated spectra using the conventional formula described (52). The fully deuterated control, 100% deuterium incorporation, was calculated theoretically, and corrections for back-exchange were made on the basis of an estimated 70% deuterium recovery and accounting for 80% final deuterium concentration in the sample (1:5 dilution in D₂O HDX

buffer). Statistical significance for the differential HDX data is determined by an unpaired *t*-test for each time point, a procedure that is integrated into the HDX Workbench software (53).

The HDX data from all overlapping peptides were consolidated to individual amino acid values using a residue averaging approach. Briefly, for each residue, the deuterium incorporation values and peptide lengths from all overlapping peptides were assembled. A weighting function was applied in which shorter peptides were weighted more heavily and longer peptides were weighted less. Each of the weighted deuterium incorporation values were then averaged incorporating this weighting function to produce a single value for each amino acid. The initial two residues of each peptide, as well as prolines, were omitted from the calculations. This approach is similar to that previously described (54).

HDX analyses were performed in triplicate. Deuterium uptake for each peptide is calculated as the average of %D for all on-exchange time points and the difference in average %D values between the unbound and bound samples is presented as a heat map with a color code given at the bottom of the figure (warm colors for deprotection and cool colors for protection). Peptides are colored by the software automatically to display significant differences, determined either by a >5% difference (less or more protection) in average deuterium uptake between the two states, or by using the results of unpaired *t*-tests at each time point (*p*-value < 0.05 for any two time points or a *p*-value < 0.01 for any single time point). Peptides with non-significant changes between the two states are colored grey. The exchange at the first two residues for any given peptide is not colored. Each peptide bar in the heat map view displays the average Δ %D values, associated standard deviation, and the charge state. Additionally, overlapping peptides with a similar protection trend covering the same region are used to rule out data ambiguity.

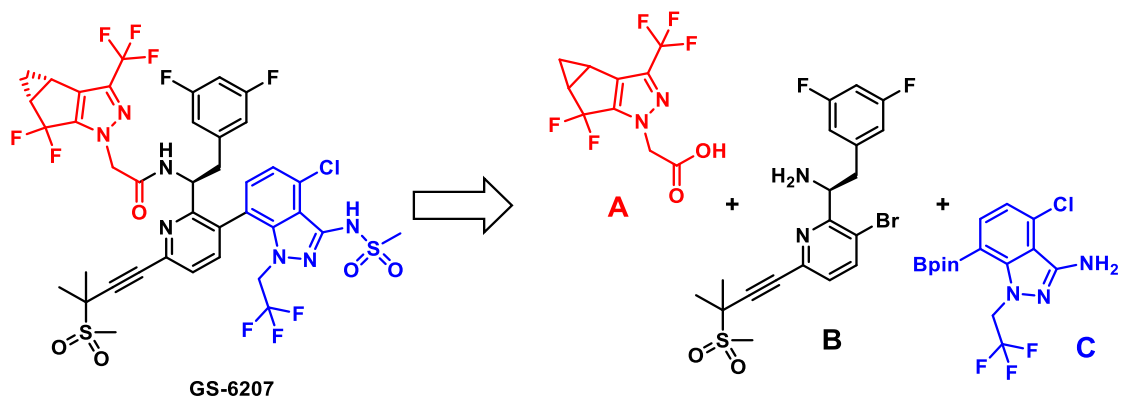
General Chemistry Procedures

Unless stated otherwise, reactions were monitored by thin layer chromatography (TLC) on silica gel plates (60 F₂₅₄), using hexanes and ethyl acetate (EtOAc) and visualized with ultraviolet light or *p*-anisaldehyde. Dichloromethane (CH₂Cl₂), tetrahydrofuran (THF), diethyl ether, toluene, and dimethylformamide (DMF) were purified with a Pure Solv. MD-6 solvent purification system. All other solvents and reagents were used as received from chemical vendors. Column chromatography was performed on a Teledyne ISCO CombiFlash® NextGen 300+ Instrument with RediSep® Flash Column and using hexanes and ethyl acetate as the mobile phase

unless otherwise stated. Solvents for chromatography are listed as percentage (v/v). ^1H and ^{13}C NMR spectra were recorded on a Bruker DPX 400 spectrometer in CDCl_3 , CD_3OD , or $\text{DMSO-}d_6$ solutions operating at 400 MHz for ^1H NMR and 100 MHz for ^{13}C NMR unless otherwise stated. Chemical shifts are reported in ppm on the δ scale relative to residual CHCl_3 ($\delta = 7.26$ for ^1H NMR and $\delta = 77.2$ for ^{13}C NMR) as an internal reference. Spin multiplicities are given as s (singlet), d (doublet), t (triplet), and m (multiplet) as well as b (broad). Coupling constants (J) are given in hertz (Hz). High resolution mass spectra were measured on a Thermo LTQ Orbitrap instrument. Trace analyses were obtained using a Shimadzu Prominence HPLC system.

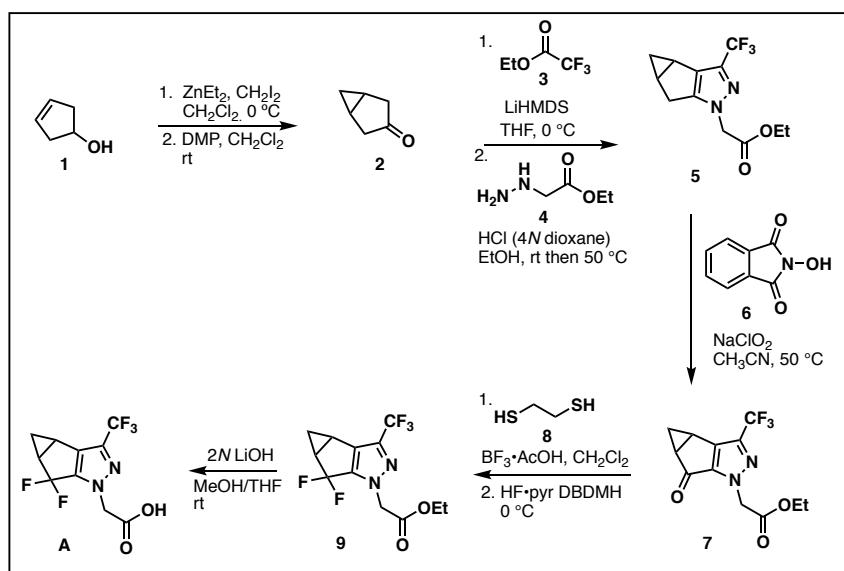
Synthesis of GS-6207

GS-6207 was prepared through slight modification of the reported procedure (55). The synthesis of this complex molecule can be simplified into the preparation of three basic subunits (see below), which are then coupled in a highly efficient manner to form the final, desired compound. Subunit **A** was generated as a racemic mixture in 8 steps from 3-cyclopentenol (Scheme S1). Subunit **B**, the central portion of the molecule, was prepared in 7 steps from 3,6-dibromo-2-methylpyridine (**10**), using a chiral sulfinyl imine intermediate (**12**)(56) to establish the desired configuration of the difluorobenzyl substituent (Scheme S2). Key to this synthesis was the use of 3,5-difluorobenzylzinc bromide (**14**) at 0 °C, resulting in an improved distereoselectivity (4:1) as compared to the addition of difluorobenzylmagnesium bromide. Subunit **C** was prepared in 5 steps from 1-bromo-4-chloro-2-fluorobenzene (**20**) as the pinacolato borane intermediate (subunit **C**) for subsequent coupling reaction with subunit **B** (Scheme S3). After successful coupling of subunits **B** and **C**, the **A** subunit could likewise be introduced using a standard peptide coupling method. A final deprotection of the sulfonimide group on subunit **C** to the corresponding sulfonamide provided a mixture of both diastereomers and atropisomers (Scheme S4). These isomers could be readily separated by silica gel column chromatography to provide the desired diastereomer (GS-6207) in a 19:1 mixture of atropisomers. These atropisomers were found to rapidly interconvert in solution at room temperature over a period of hours as monitored by HPLC analysis in both methanol and DMSO. Although the interconversion was slightly slower in methanol, both samples ultimately equilibrated to a 4:1 mixture over the course of 7 days.




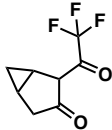
Retrosynthetic disconnection of GS-6207 into three simpler subunits for chemical synthesis.

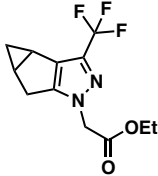
Scheme S1: Synthesis of Subunit A



Bicyclo [3.1.0] hexane-3-ol. To anhydrous CH_2Cl_2 (100 mL) in a 500-mL round bottom flask was added diethylzinc (99 mL, 89.87 mmol, 0.9 M in *n*-hexane) at 0 °C, followed by a solution of diiodomethane (7.2 mL, 89.87 mmol) in CH_2Cl_2 (40 mL), and the mixture was stirred for 30 min at this temperature. To this solution was added 3-cyclopentenol (4.2 g, 49.93 mmol) in CH_2Cl_2 (20 mL). The resulting solution was stirred at 0 °C for 3 h (monitored by TLC) followed by quenching with 1N HCl (30 mL) to obtain two clear layers. The two layers were separated, and the aqueous layer was extracted with CH_2Cl_2 (20 mL \times 2). The organic phases were combined, dried over anhydrous Na_2SO_4 and filtered. The resulting filtrate containing crude bicyclo[3.1.0]hexan-3-ol was used directly for the next step.

Bicyclo[3.1.0]hexan-3-one (2). To the crude bicyclo[3.1.0]hexane-3-ol from above in CH₂Cl₂ (~200 mL), was added Dess-Martin periodinane (31.8 g, 74.89 mmol), portion-
 wise at room temperature and the resulting mixture was stirred for 2 h. The reaction was quenched with (1:1) NaHCO₃:Na₂S₂O₃ (200 mL). 100 mL of water was added and the resulting mixture was stirred for approximately 1 h to obtain two clear layers, which were separated. The aqueous layer was extracted with CH₂Cl₂ (50 mL x 2). The organic phase was dried over anhydrous Na₂SO₄ and filtered. The solvent was removed via distillation at 1 atm and the crude product was subsequently purified by distillation (24 mmHg, bp: 50–54 °C) to give the title compound (2.6 g, 57 % yield over 2 steps). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 2.63–2.57 (m, 2H), 2.14–2.19 (d, *J* = 20 Hz, 2H), 1.60–1.52 (m, 2H), 0.89–0.94 (m, 1 H), -0.05–0.02 (m, 1 H); ¹³C-NMR (100 MHz, CDCl₃) δ: 41.3, 13.5, 12.4; HRMS (*m/z*): [M+Na]⁺ calculated 119.0467, found 119.0475.

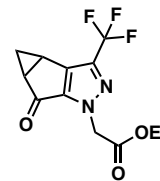
2-(2,2,2-trifluoroacetyl)bicyclo[3.1.0]hexan-3-one. A 1.0 M solution of lithium bis(trimethylsilyl)amide (LiHMDS) in tetrahydrofuran (6.24 mL, 6.24 mmol) in a 50-
 mL round-bottom flask was cooled in an ice bath with stirring under argon. A solution of **2** (500 mg, 5.201 mmol) in anhydrous tetrahydrofuran (12.5 mL) was then added dropwise via an addition funnel to the LiHMDS solution, and the whole mixture was stirred for 30 min with cooling under argon at 0 °C. The mixture was then treated dropwise with ethyl trifluoroacetate (**3**) (0.68 mL, 6.241 mmol) and stirred for a further 2 h at room temperature. After complete consumption of starting material, the mixture was partitioned between saturated NH₄Cl solution (20 mL) and EtOAc (20 mL). The layers were separated and the aqueous layer was extracted with EtOAc (10 mL x 2). The combined organic layers were dried over anhydrous Na₂SO₄ and the solvent removed by rotary evaporation to give the title compound as a brown oil which was used directly in the next step.

Ethyl 2-(3-(trifluoromethyl)-3b,4,4a,5-tetrahydro-1H-cyclopropa[3,4]cyclopenta[1,2-c]pyrazol-1-yl)acetate (5). The crude material from the previous step was charged in a 100-mL round-bottom flask and dissolved in EtOH (12.5 mL). 4*N* HCl in 1,4-dioxane (2.6 mL, 10.402 mmol) was added followed by ethyl hydrazinoacetate hydrochloride (**4**) (1.22 g, 7.802 mmol). The resulting mixture was heated at 70 °C


overnight. The EtOH was removed under vacuum, and the residue obtained was dissolved in EtOAc (50 mL). The whole mixture was transferred into a separatory funnel and washed with aqueous saturated NaHCO₃ solution (15 mL) and water (15 mL). The aqueous layers were combined and extracted (2 x 50 mL). The combined organic layers were dried with Na₂SO₄. After filtering, the solvent was removed and the residue was purified by flash column chromatography (15% EtOAc in hexanes) to obtain the title compound as a white solid (0.70 g, 44% over 2 steps). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 4.89, 4.82 (ABq, *J*_{AB} = 17.3, 2H), 4.22 (q, *J* = 7.3 Hz, 2H), 2.93 (dd, *J* = 16.5, 6.2 Hz, 1H), 2.75 (d, *J* = 6.2 Hz, 1H), 2.18–2.05 (m, 2H), 1.26 (t, *J* = 7.3 Hz, 3H), 1.19–1.09 (m, 1H), 0.38–0.32 (m, 1H); ¹³C-NMR (100 MHz, CDCl₃) δ: 166.9, 151.5, 135.9 (q, *J* = 40.0 Hz), 129.9, 121.5 (q, *J* = 268.5 Hz), 62.2, 52.3, 27.4, 22.0, 17.6, 14.2, 14.1; HRMS (m/z): [M+Na]⁺ calculated 297.0824, found 297.0821.

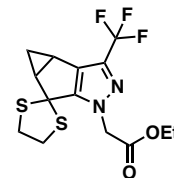
Ethyl 2-(5-oxo-3-(trifluoromethyl)-3b,4,4a,5-tetrahydro-1H-cyclopropa[3,4]cyclopenta[1,2-c]pyrazol-1-yl)acetate (7).

A 10-mL round-bottom flask was charged with **5** (0.684 g, 2.494 mmol) and a 2:1 mixture of CH₃CN/H₂O (25 mL). To this mixture was added *N*-hydroxyphthalimide (0.081 g, 0.499 mmol) and NaClO₂ (0.902 mg, 9.977 mmol). The resulting mixture was heated to 50 °C overnight. A second portion of NaClO₂ (0.902 g, 9.977 mmol) was then added and the mixture was heated at 50 °C until all starting material was consumed (as monitored by ¹H-NMR). After consumption of starting material, the temperature was lowered to 20 °C, aqueous sodium bisulfite (40% w/w, 20 mL) was added followed by EtOAc (30 mL) and the layers were separated. The resulting aqueous layer was then back extracted with EtOAc (5 mL x 2). The organic layers were combined and washed with saturated aqueous NaHCO₃, water and brine. The organic layer was concentrated under reduced pressure and the residue was used directly in the next reaction.



Ethyl 2-(3-(trifluoromethyl)-4,4a-dihydrospiro[cyclopropa[3,4]cyclopenta[1,2-c]pyrazole-5,2'-[1,3]dithiolane]-1(3bH)-yl)acetate (7a).

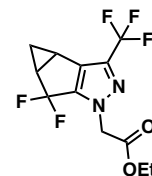
To the crude material (compound **7**) from the previous step in CH₂Cl₂ (25 mL) was added ethane-1,2-dithiol (0.52 mL, 6.235 mmol) in one portion followed by BF₃•2AcOH (0.87 mL, 6.24 mmol). The reaction was stirred at ambient temperature for 12 h and then cooled to 0 °C after which, it was quenched with saturated aqueous NaHCO₃ (20.0 mL). The organic layer was



separated, washed with brine (5.0 mL) and dried over Na₂SO₄. Solvents were removed *in vacuo* and the residue was used directly in the next reaction. HRMS (m/z): [M+Na]⁺ calculated 387.0419, found 387.0417.

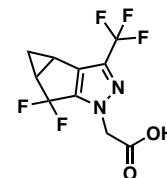
Ethyl 2-((3b*S*,4a*R*)-5,5-difluoro-3-(trifluoromethyl)-3b,4,4a,5-tetrahydro-1*H*-cyclopropa[3,4]cyclopenta[1,2-*c*]pyrazol-1-yl)acetate (9).

A solution of 1,3-dibromo-5,5 dimethylhydantoin (DBDMH) (2.17 g, 7.596 mmol) in CH₂Cl₂ (9.50 mL) was cooled to -78 °C in a Teflon bottle. HF•pyridine (3.4 mL, 37.98 mmol) was added dropwise. The mixture was stirred at -78 °C for 30 min. A solution of crude compound **7a** (0.1815 mol) in CH₂Cl₂ (3.20 mL) was added drop-wise via an addition funnel. The resulting mixture was stirred for 30 min and then slowly warmed to -30 °C, where the temperature was maintained for 1.5 h. The mixture was slowly poured into an aqueous saturated solution of NaHCO₃ (30 mL) and extracted with ethyl acetate (20 mL x 3). The combined organic layer was dried over Na₂SO₄. Solvents were removed *in vacuo* to afford the crude product, which was further purified by flash column chromatography (5% EtOAc in hexanes) to provide the title compound (0.48 g, 62% 3 steps). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 4.86 (s, 2H), 4.24 (q, *J* = 7.3 Hz, 2H), 2.55–2.40 (m, 2H), 1.45–1.35 (m, 1H), 1.29 (t, *J* = 7.2 Hz, 3H) 1.20–1.10 (m, 1H); ¹³C-NMR (100 MHz, CDCl₃) δ: 166.0, 143.4 (t, *J* = 29 Hz), 136.5 (q, *J* = 39 Hz), 132.8 (m), , 120.5 (q, *J* = 268 Hz), 120.1 (t, *J* = 243 Hz), 62.5, 52.2, 28.1, (dd, *J* = 36, 29 Hz), 23.4, 14.1, 12.4 (dd, *J* = 4.0, 2.2 Hz); HRMS (m/z): [M+Na]⁺ calculated 333.0633, found 333.0635.



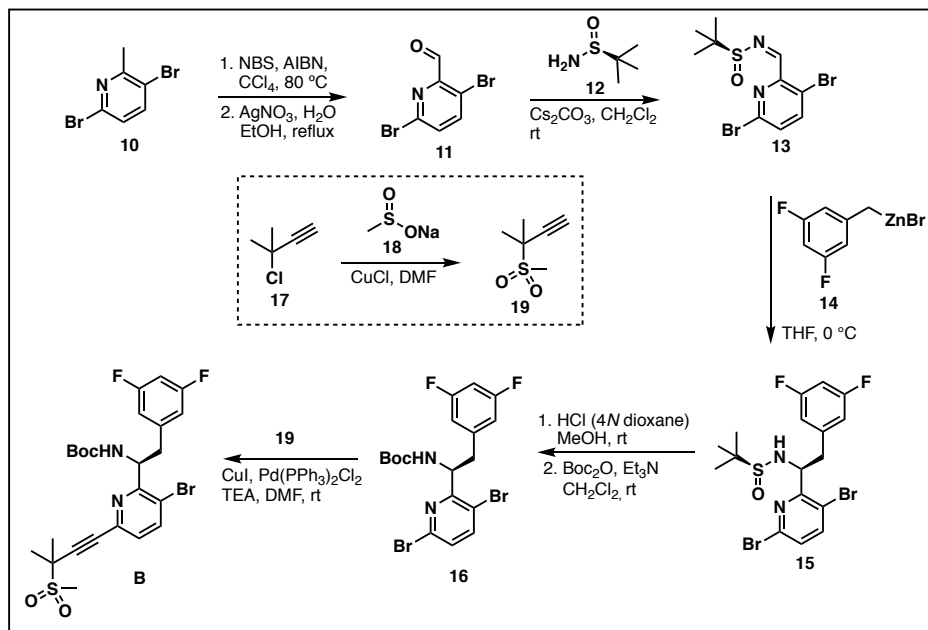
2-((3b*S*,4a*R*)-5,5-difluoro-3-(trifluoromethyl)-3b,4,4a,5-tetrahydro-1*H*-cyclopropa[3,4]cyclopenta[1,2-*c*]pyrazol-1-yl)acetic acid (A).

To a solution of **9** (480 mg, 1.547 mmol) in a 2:1 mixture of THF/MeOH (6.0 mL) contained in a 25-mL round-bottom flask was added 2*N* LiOH (1.6 mL, 3.095 mmol). The mixture was stirred for 1 h at ambient temperature and then concentrated to remove most of the THF and MeOH. The obtained aqueous phase was acidified with 1*N* HCl to adjust the pH to 2-3, and then extracted with EtOAc (15 mL x 2). The organic phase was dried over Na₂SO₄, filtered and concentrated *in vacuo* to provide subunit **A** (0.42 g, 96%) as a pale yellow solid. ¹H-NMR (400 MHz, CDCl₃) δ ppm: 4.94 (s, 2H), 2.55–2.40 (m, 2H), 1.44–1.34 (m, 1H), 1.19–1.09 (m, 1H); ¹³C-NMR (100 MHz, CDCl₃) δ: 170.8, 143.6 (t, *J* = 29 Hz), 136.6 (q, *J* = 39.9 Hz), 133.0 (m), 120.5



(q, $J = 268$ Hz), 120.1 (t, $J = 243$), 51.5, 28.1, (dd, $J = 36, 29$ Hz), 23.5, 12.5 (dd, $J = 4.0, 2.2$ Hz); HRMS (m/z): $[M+Na]^+$ calculated 305.0320, found 305.0323.

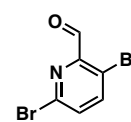
Scheme S2: Synthesis of Subunit B



3,6-dibromo-2-(dibromomethyl)pyridine (10a). To a solution of **10** (10.0 g, 39.85 mmol) in CCl_4 (100 mL) in a 250-mL round-bottom flask was added *N*-bromosuccinimide (21.3, 119.52 mmol) and 2,2'-azobis(2-methylpropionitrile) (1.31 g, 7.97 mmol). The mixture was heated at 80 °C overnight and then cooled to ambient temperature. The resulting solid was removed by filtration and the filtrate was concentrated under reduced pressure. The crude product was used directly in the next step. An analytical sample was obtained after flash column chromatography eluting with 0-10% EtOAc in hexanes. 1H -NMR (400 MHz, $CDCl_3$) δ ppm: 7.68 (d, $J = 8.4$ Hz, 1H), 7.32 (d, $J = 8.4$ Hz, 1H), 7.02 (s, 1H); ^{13}C -NMR (100 MHz, $CDCl_3$) δ : 155.6, 143.3, 140.8, 130.7, 115.8, 38.4; HRMS (m/z): $[M+H]^+$ calculated 409.7031, found 409.7046



3,6-dibromopicolinaldehyde (11). A solution of silver nitrate (7.6 g, 45 mmol) in water (24 mL) was added dropwise to a solution of **10a** (crude, 39.85 mmol) in refluxing EtOH (200 mL) in a 500-mL three-necked flask. The mixture was stirred at

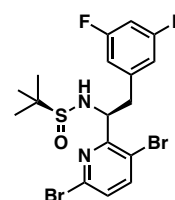


80 °C for 5 h. After this time, the mixture was cooled to room temperature, diluted with water (100 mL), extracted with EtOAc (3 x 20 mL). The combined organic layers were dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The product was obtained after flash column chromatography eluting with 10% EtOAc in hexanes (7.36 g, 70% yield over 2 steps). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 10.09 (s, 1H), 7.87 (d, *J* = 8.4 Hz, 1H), 7.53 (d, *J* = 8.4 Hz, 1H); ¹³C-NMR (100 MHz, CDCl₃) δ: 189.4, 149.1, 144.9, 141.1, 133.0, 120.3; HRMS (m/z): [M+H]⁺ calculated 265.8634, found 265.8637.

(*S*)-*N*-((3,6-dibromopyridin-2-yl)methylene)-2-methylpropane-2-sulfinamide (13). Cs₂CO₃ (0.148 g, 0.453 mmol) was added to a mixture of the crude aldehyde **11** (0.10 g, 0.378 mmol) and (*S*)-(+)-*tert*-butanesulfinamide (0.0503 g, 0.415 mmol) in CH₂Cl₂ (2 mL) in a 25-mL round-bottom flask. The mixture was stirred at room temperature for 1 h, followed by dilution with CH₂Cl₂ (20 mL). After washing with H₂O (20 mL), the organic layer was dried over Na₂SO₄ and concentrated. The crude material was purified by flash column chromatography (15% EtOAc in hexanes), leading to the desired product as a pale yellow solid (5.72 g, 61%). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 8.86 (s, 1H), 7.82 (d, *J* = 8.4 Hz, 1H), 7.44 (d, *J* = 8.4 Hz, 3H), 1.31 (s, 9H); ¹³C-NMR (100 MHz, CDCl₃) δ: 160.0, 149.6, 143.9, 141.1, 131.2, 121.9, 58.6, 22.8; HRMS (m/z): [M+Na]⁺ calculated 390.8909, found 390.8912.



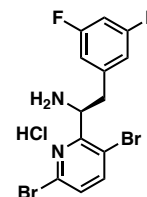
(*S*)-*N*-((*S*)-1-(3,6-dibromopyridin-2-yl)-2-(3,5-difluorophenyl)ethyl)-2-methylpropane-2-sulfinamide (15). To a 500-mL round-bottom flask containing a solution of **13** (5.72 g, 15.506 mmol) in THF (80 mL) at 0 °C was added (3,5-difluorobenzyl)zinc bromide (**14**) (0.5 M in THF, 62 mL, 31.01 mmol) dropwise via an addition funnel. The reaction was warmed to room temperature and stirring was maintained at that temperature for 1 h. The reaction was quenched with saturated aqueous NH₄Cl (100 mL) and the mixture was extracted with EtOAc (2 x 50 mL). The combined organic layers were dried over Na₂SO₄, filtered, and concentrated. The crude residue was purified by flash column chromatography (30% EtOAc in hexanes) to afford the title compound as a single diastereomer (6.62 g, 86%). ¹H-NMR (300 MHz, CDCl₃) δ ppm: 7.67 (d, *J* = 8.4 Hz, 1H), 7.28 (d, *J* = 8.4 Hz, 1H), 6.78–6.61 (m, 3H), 4.99 (dt, *J* = 9.1, 4.8 Hz, 1H), 4.81 (d, *J* = 9.1 Hz, 1H), 3.83 (dd, *J* = 13.6,



4.8 Hz, 1H), 2.86 (dd, $J = 13.6, 9.1$ Hz, 1H), 1.14 (s, 9H); HRMS (m/z): $[M+Na]^+$ calculated 518.9346, found 518.9346.

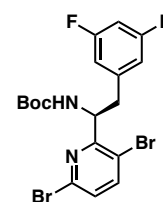
(S)-1-(3,6-dibromopyridin-2-yl)-2-(3,5-difluorophenyl)ethan-1-amine hydrochloride (15a).

To a stirred solution of (15a) (6.62 g, 13.341 mmol) in MeOH (27 mL) in 100-mL round-bottom flask was added 4*N* HCl in dioxane (6.60 mL) and the resulting mixture stirred at room temperature for 3 h at which point the reaction was deemed complete by TLC analysis. The mixture was concentrated to obtain the crude residue as white solid, which was used directly in the next step. HRMS (m/z): $[M+H]^+$ calculated 392.9231, found 392.9230.



***tert*-Butyl (S)-1-(3,6-dibromopyridin-2-yl)-2-(3,5-difluorophenyl)ethylcarbamate (16).** A

500-mL round bottom flask was charged with crude material from the reaction above and CH₂Cl₂ (66 mL). To this solution was added triethylamine (3.7 mL, 26.682 mmol) and di-*tert*-butyl dicarbonate (3.5 g, 16.009 mmol). The mixture was stirred at room temperature for 2 h (based on TLC analysis), then diluted with water and extracted with CH₂Cl₂ (15 mL x 2). The organic layers were combined and washed with brine, dried with Na₂SO₄ and concentrated. The crude material was purified by flash column chromatography to provide the title compound (6.06 g, 92% over 2 steps). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 7.66 (d, $J = 8.3$, 1H), 7.27 (d, $J = 8.3$, 1H), 6.78–6.55 (m, 3H), 5.66–5.35 (m, 2H), 3.15–2.83 (m, 2H), 1.40 (s, 9H); ¹³C-NMR (100 MHz, CDCl₃) δ: 162.9 (dd, $J = 249.4, 13.4$ Hz), 159.1, 155.0, 142.9, 140.2, 128.5, 119.1, 112.4 (d, $J = 25.0$ Hz), 112.4 (d, $J = 11.6$ Hz), 102.3 (t, $J = 25.3$ Hz), 80.1, 54.2, 41.6, 28.4; HRMS (m/z): $[M+Na]^+$ calculated 514.9575, found 514.9584.



3-Methyl-3-(methylsulfonyl)but-1-yne (19). To a stirred suspension of sodium

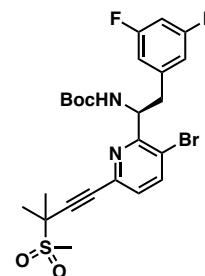
methanesulfinate (13) (3.00 g, 29.25 mmol) and copper(I) chloride (0.24 g, 2.4 mmol) in DMF (50 mL) in a 100-mL round-bottom flask was added 14 (2.74 mL, 24.44 mmol) dropwise. The resulting mixture was heated to 40 °C and stirred for 16 h. When the reaction was completed as determined by TLC analysis the mixture was cooled to room temperature and diluted with EtOAc. The solution was washed with water and brine. The organic layer was collected, dried over Na₂SO₄, and then filtered. The solution was concentrated under vacuum and purified by flash



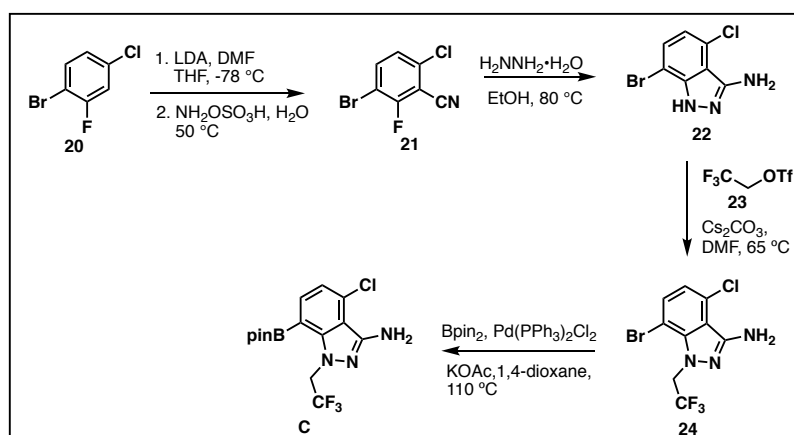
column chromatography (3% EtOAc in hexanes) to provide the title compound (1.46 g, 41%). ¹H-NMR (400 MHz, CDCl₃) δ 3.04: (s, 3H), 2.58 (s, 1H), 1.67 (s, 6H); ¹³C-NMR (100 MHz, CDCl₃) δ 82.1, 75.2, 57.3, 34.9, 22.8; HRMS (m/z): [M+Na]⁺ calculated 169.0294, found 169.0297.

tert-Butyl (S)-(1-(3-bromo-6-(3-methyl-3-(methylsulfonyl)but-1-yn-1-yl)pyridin-2-yl)-2-(3,5-difluorophenyl)ethyl)carbamate (B).

A 25-mL round-bottom flask was charged with **16** (0.40 g, 0.813 mmol), **19** (0.143 g, 0.975 mmol), DMF (4 mL) and triethyl amine (0.34 mL, 2.44 mmol). Bis(triphenylphosphine)palladium(II) dichloride (33 mg, 0.0406 mmol) and copper(I) iodide (7.74 mg, 0.0406 mmol) were added. After stirring for 2 h at room temperature, the mixture was diluted with water (20 mL) and extracted with EtOAc (20 mL x 3). The organic layers were washed with brine, dried with Na₂SO₄ and concentrated. The crude compound was purified by flash column chromatography to provide subunit **B** (0.34 g, 74%). ¹H-NMR (400 MHz, CDCl₃) δ: 7.81 (d, *J* = 8.2 Hz, 1H), 7.23 (d, *J* = 8.2 Hz, 1H), 6.77–6.55 (m, 3H), 5.77–5.42 (m, 2H), 3.21–3.00 (m, 4H), 2.99–2.88 (m, 1H), 1.80 (s, 6H), 1.79 (br s, 6H), 1.41 (br s, 9H); ¹³C-NMR (100 MHz, CDCl₃) δ: 158.3, 155.0, 141.1, 127.7, 112.8 (d, *J* = 24.5 Hz), 112.5 (d, *J* = 11.8 Hz), 102.2 (t, *J* = 25.3 Hz) 84.7, 80.0, 58.1, 54.4, 41.7, 35.4, 28.5, 22.8, 22.6; HRMS (m/z): [M+H]⁺ calculated 557.0916, found 557.0920.



Scheme S3: Synthesis of Subunit C

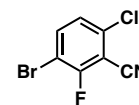


3-bromo-6-chloro-2-fluorobenzaldehyde (20a). A solution of lithium diisopropylamide (LDA) (1.2 mL, 1.25 mmol, 2M in THF) in THF (6.25 mL) in a 25-mL round-bottom flask was cooled to -78 °C. **20** (0.209 g, 0.9998 mmol) in THF



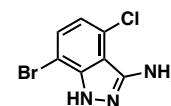
(1 mL) was added dropwise via an addition funnel, forming a yellow solution. After 10 min, DMF (0.26 mL, 16.3 mmol) was added dropwise via a syringe. The mixture was allowed to warm up to -20 °C over the course of 1 h, at which point TLC analysis indicated that all the starting material had been consumed. The reaction was then quenched with saturated aqueous solution of NH₄Cl, diluted with EtOAc (100 mL), and washed with water and brine. The organic layers were dried over Na₂SO₄, filtered, and concentrated under reduced pressure to yield the crude product as light yellow solid. An analytical sample was obtained after flash column chromatography eluting with 10% EtOAc in hexanes. ¹H-NMR (400 MHz, CDCl₃) δ ppm: 10.09 (s, 1H), 7.87 (dd, *J* = 8.8, 7.2, 1H), 7.53 (dd, *J* = 8.6, 1.3 Hz, 1H); ¹³C-NMR (100 MHz, CDCl₃) δ: 186.1, 159.7 (d, *J* = 263.8 Hz), 138.2, 136.1, 127.6 (d, *J* = 4.4 Hz), 122.8 (d, *J* = 11.2 Hz), 109.3 (d, *J* = 21.7 Hz); HRMS (m/z): [M+H]⁺ calculated 238.9092, found 238.9095.

3-bromo-6-chloro-2-fluorobenzonitrile (21). A 250-mL round-bottom flask containing a suspension of crude compound **20a** (crude, 23.873 mmol) and *O*-sulfonic acid hydroxylamine (5.3996 g, 47.746 mmol) in water (120 mL) was heated at 50 °C



overnight. The suspension was cooled and extracted with EtOAc (50 mL x 3), dried over Na₂SO₄, filtered and concentrated under reduced pressure. The crude material was purified by flash column chromatography (10% EtOAc in hexanes) to give **21** (7.02 g, 63% yield over 2 steps). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 10.43 (s, 1H), 7.74 (dd, *J* = 8.8, 7.2, 1H), 7.24 (dd, *J* = 8.6, 1.3 Hz, 1H); ¹³C-NMR (100 MHz, CDCl₃) δ: 160.5 (d, *J* = 261.7 Hz), 138.2 (d, *J* = 2.1 Hz), 137.1, 126.4 (d, *J* = 4.1 Hz), 110.5, 108.3 (d, *J* = 20.4 Hz), 104.8 (d, *J* = 18.9 Hz); HRMS (m/z): [M+H]⁺ calculated 238.9092, found 238.9095.

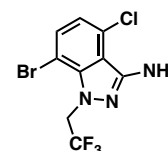
7-bromo-4-chloro-1*H*-indazol-3-amine (22). To a solution of **21** (2.0 g, 8.53 mmol) in anhydrous EtOH (10.0 mL) in a 50-mL round-bottom flask was added NH₂NH₂•H₂O (2.00 mL, 34.12 mmol) dropwise via syringe with stirring at room



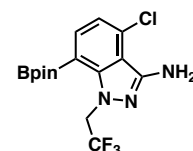
temperature. The mixture was heated at 90 °C for 2 h, cooled to room temperature, and concentrated in vacuo to remove most of the EtOH. The crude product was dissolved in EtOAc

(30 mL) and washed successively with water (30 mL) and saturated aqueous NaCl (30 mL). The organic layers were dried with Na₂SO₄, filtered, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (3% MeOH in CH₂Cl₂) to give the title compound (1.31 g, 62%). ¹H-NMR (400 MHz, DMSO-*d*₆) δ ppm: 12.21 (br s, 1H), 7.42 (d, *J* = 7.8 Hz 1H), 6.85 (d, *J* = 7.8 Hz, 3H), 5.30 (br s, 2H); ¹³C-NMR (100 MHz, DMSO-*d*₆) δ: 149.1, 141.1, 129.5, 125.2, 119.1, 111.9, 101.0; HRMS (m/z): [M+H]⁺ calculated 247.9408, found 247.9407.

7-bromo-4-chloro-1-(2,2,2-trifluoroethyl)-1H-indazol-3-amine (24). To a 50-mL round-bottom flask was added **22** (0.20 g, 0.81 mmol, 1.00 equiv), *N,N*-dimethylformamide (3.0 mL), and Cs₂CO₃ (0.53 g, 1.62 mmol, 2.00 equiv). The resulting mixture was stirred for 20 min prior to the addition of **23** (0.19 mL, 1.46 mmol, 1.80 equiv). The mixture was then stirred overnight at 65 °C. It was then cooled to room temperature, quenched by the addition of 50 mL of water/ice, and extracted with EtOAc (3 x 15 mL). The combined organic layers were washed with 15 mL of H₂O, 15 mL of brine, dried over anhydrous Na₂SO₄, and concentrated *in vacuo*. The crude material was purified by flash column chromatography (15% EtOAc in hexanes) to provide the title compound (0.15 g, 54%). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 7.45 (d, *J* = 8.1 Hz, 1H), 6.89 (d, *J* = 8.1 Hz, 1H), 5.21 (q, *J* = 8.0 Hz, 2H), 3.45 (br s, NH₂); ¹H-NMR (400 MHz, DMSO-*d*₆) δ ppm: 7.49 (d, *J* = 8.0 Hz 1H), 7.01 (d, *J* = 8.0 Hz, 1H), 5.72 (br s, 2H), 5.34 (q, *J* = 8.7, 2H); ¹³C-NMR (100 MHz, DMSO-*d*₆) δ: 149.0, 139.3, 133.1, 126.0, 124.2 (q, *J* = 282.5 Hz), 121.0, 114.5, 100.4, 49.2 (q, *J* = 33.0 Hz); HRMS (m/z): [M+H]⁺ calculated 329.9438, found 329.9438.

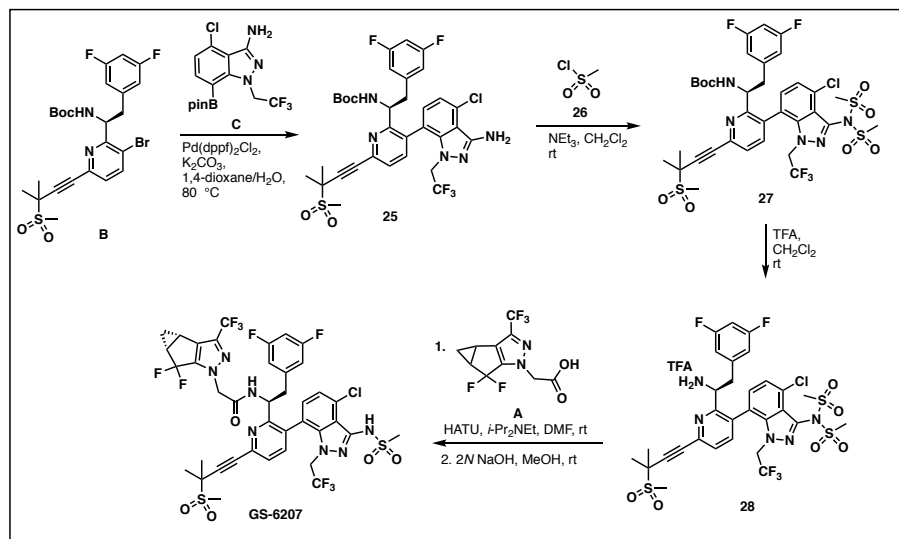


4-chloro-1-(2,2,2-trifluoroethyl)-7-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1H-indazol-3-amine (C). A 25-mL round-bottom flask was charged with **24** (145.00 mg, 0.44 mmol, 1.00 equiv.), 1,4-dioxane (4.40 mL), KOAc (173.30 mg, 1.77 mmol, 4.00 equiv), 4,4,5,5-tetramethyl-2-(tetramethyl-1,3,2-dioxaborolan-2-yl)-1,3,2-dioxaborolane (336.20 mg, 1.32 mmol, 3.00 equiv.), and Pd(PPh₃)₂Cl₂ (31.00 mg, 0.044 mmol, 0.1 equiv.). The mixture was stirred overnight at 110 °C, then cooled to room temperature, quenched by the addition of 5 mL of water/ice, and extracted with EtOAc (2 x 5 mL). The combined organic layers were washed with 10 mL of H₂O, 10 mL of brine, dried over anhydrous



Na₂SO₄ and concentrated *in vacuo*. The residue was applied onto a flash column (RediSep®) and eluted with 15% EtOAc in hexanes to afford subunit **C** (86.2 mg, 52%). ¹H-NMR (400 MHz, CDCl₃) δ ppm: 7.94 (d, *J* = 7.7 Hz, 1H), 7.08 (d, *J* = 7.7 Hz, 1H), 5.49 (m, 2H), 2.37 (br s, NH₂), 1.38 (12H, s); ¹³C-NMR (100 MHz, CDCl₃) δ: 148.4, 146.0, 139.4, 131.3, 124.38 (q, *J* = 280.9 Hz), 120.0, 112.7, 85.0, 84.9, 50.7 (q, *J* = 34.0 Hz), 24.9; HRMS (m/z): [M+Na]⁺ calculated 398.1025, found 398.1030.

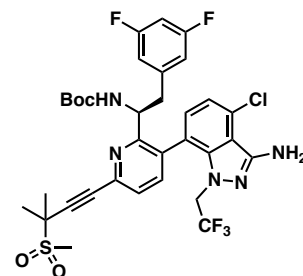
Scheme S4: Coupling of subunits **A**, **B**, and **C**



tert-butyl (S)-(1-(3-(3-(3-amino-4-chloro-1-(2,2,2-trifluoroethyl)-1H-indazol-7-yl)-6-(3-methyl-3-(methylsulfonyl)but-1-yn-1-yl)pyridin-2-yl)-2-(3,5-difluorophenyl)ethyl)carbamate (25).

Subunit **B** (114 mg, 0.205 mmol), subunit **C** (100.0 mg, 0.266 mmol), [1,1'-bis(diphenylphosphino)-ferrocene]dichloropalladium(II) (17.0 mg, 0.0205 mmol), and potassium carbonate (85.0 mg, 1.35 mmol) were charged in a 10-mL round-bottom flask and placed under argon. 1,4-Dioxane (2.0 mL) and water (0.3 mL) were added. The mixture was

heated to 110 °C and allowed to stir overnight. Upon completion (TLC analysis), the reaction mixture was cooled to room temperature and partitioned between EtOAc (10 mL) and water (10 mL). The aqueous layer was removed, and the organic layer was concentrated under vacuum. The resulting residue was purified by flash column chromatography to provide the title compound (101 mg, 78%). ¹H-NMR (400 MHz, CDCl₃) δ 7.52–7.47 (m, 2H), 7.11–7.00 (m, 1H), 6.89 (d *J* =



7.5Hz, 1H), 6.61 (tt, $J = 9.2, 2.3$ Hz, 1H), 6.21–6.15 (m, 2H), 5.56 (d, $J = 9.5$ Hz, 1H), 4.80–4.56 (m, 3H), 4.19–4.36 (m, 1H), 3.94–3.80 (m, 1H), 3.18 (s, 3H), 3.02–2.82 (m, 2H), 1.85 (s, 6H), 1.38 (bs, 9H) (*atropisomer not included*); HRMS (m/z): $[M+Na]^+$ calculated 748.1754, found 748.1755.

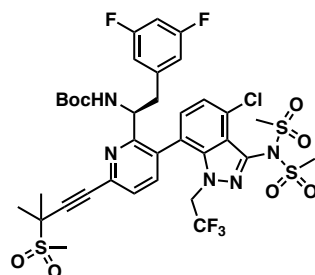
tert-butyl

(*S*)-(2-(3,5-difluorophenyl)-1-(3-(4-methyl-3-(*N*-

(methylsulfonyl)methylsulfonamido)-1-(2,2,2-trifluoroethyl)-1*H*-indazol-7-yl)-6-(3-methyl-

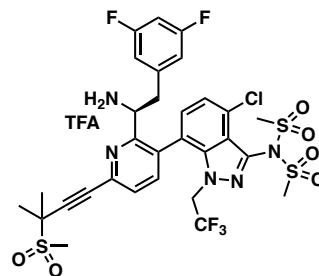
3-(methylsulfonyl)but-1-yn-1-yl)pyridin-2-yl)ethyl)carbamate (**27**).

25 (101 mg, 0.139 mmol) was dissolved in CH_2Cl_2 (1.5 mL) and triethylamine (68 μ L, 0.0486 mmol) was then added. To the resulting mixture was added methanesulfonyl chloride (**26**) with stirring at room temperature. The reaction mixture was stirred at room temperature for 1 h until the reaction was complete. Upon completion (TLC analysis), the reaction mixture was diluted with EtOAc (10 mL) and washed successively with saturated aqueous $NaHCO_3$, water, and brine. The organic layer was dried over Na_2SO_4 , filtered, and concentrated. The residue was purified by flash column chromatography (30% EtOAc in hexanes) to afford the title compound (99.5 mg, 81%). 1H -NMR (400 MHz, $CDCl_3$) δ 7.54–7.49 (br, 2H), 7.11 (br, 1H), 6.61 (tt, $J = 9.0, 2.3$ Hz, 1H), 6.21 (br 2H), 6.10 (br, 1H), 5.39 (br, 1H), 4.64–4.44 (m, 2H), 4.21–4.06 (m, 1H), 3.64 (s, 3H), 3.55 (s, 3H), 3.18 (s, 3H), 3.14 (s, 3H), 3.09–2.22 (m, 1H), 2.90–2.82 (m, 1H), 1.85 (s, 6H), 1.38 (bs, 6H) (*atropisomer not included*); HRMS (m/z): $[M+Na]^+$ calculated 904.1307, found 904.1307.

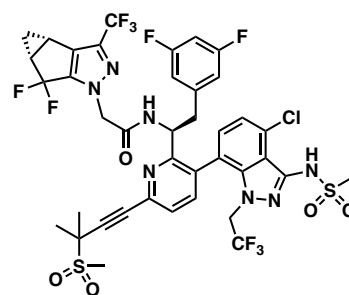


(*S*)-*N*-(7-(2-(1-amino-2-(3,5-difluorophenyl)ethyl)-6-(3-methyl-3-(methylsulfonyl)but-1-yn-1-yl)pyridin-3-yl)-4-methyl-1-(2,2,2-trifluoroethyl)-1*H*-indazol-3-yl)-*N*-(methylsulfonyl)methanesulfonamide (**28**). **27** (99.5 mg, 0.113

mmol) was charged in 10-mL round-bottom flask and a mixture of CH_2Cl_2 /TFA (2:1, 1.4 mL) was added and stirred for 1 h. The mixture was then concentrated to afford the title compound as trifluoroacetic acid salt, which was used in the next step without purification.



***N*-((*S*)-1-(3-(4-chloro-3-(methylsulfonamido)-1-(2,2,2-trifluoroethyl)-1*H*-indazol-7-yl)-6-(3-methyl-3-(methylsulfonyl)but-1-yn-1-yl)pyridin-2-yl)-2-(3,5-difluorophenyl)ethyl)-2-((3*bS*,4*aR*)-5,5-difluoro-3-(trifluoromethyl)-3*b*,4,4*a*,5-tetrahydro-1*H*-cyclopropa[3,4]cyclopenta[1,2-*c*]pyrazol-1-yl)acetamide (GS-6207).**



The crude **28** (crude, 0.113 mmol), subunit **A** (38.181 mg, 0.135 mmol), and diisopropylethylamine (40 μ L, 0.225 mmol) were charged in a 10-mL round-bottom flask and dissolved in DMF (1.0 mL). 1-Hydroxybenzotriazole (47 mg, 0.124 mmol) was added. The reaction mixture was stirred at ambient temperature for 30 min, diluted with water (10 mL) and extracted with EtOAc (10 mL x 3). The organic layers were washed with brine, dried, and concentrated. The residue was dissolved in EtOH (1.0 mL) and 0.25 mL of 2*N* aq. NaOH was added. The resulting mixture was stirred at room temperature for 30 min and then diluted with water (10 mL) and acidified with 1 *N* HCl. After extraction with EtOAc, the organic layer was washed with water, brine, and dried over Na₂SO₄. The residue was then purified by RP-Silica gel chromatography (30% H₂O in MeOH with 0.1% TFA). Fractions containing each diastereomer were combined and back extracted with EtOAc, dried and concentrated to give 55 mg of the title compound (HPLC trace analysis 19:1) along with its diastereomer (yield not determined). **GS-6207**: ¹H-NMR (400 MHz, CD₃OD) δ 7.74 (d, *J* = 8.1 Hz, 1H) 7.69 (d, *J* = 8.1 Hz, 1H), 7.20 (dd, *J* = 7.8, 1.4 Hz, 1H), 6.78 (tt, *J* = 9.2, 2.3 Hz, 1H), 6.51 (d, *J* = 7.6, 6.3 Hz, 1H), 6.36–6.26 (m, 2H), 4.94–4.70 (m), 4.70–4.48 (m, 1H), 4.07–3.87 (m, 1H), 3.24 (s, 6H), 3.17–2.85 (m, 2H), 2.60–2.43 (m, 2H), 1.83 (s, 6H), 1.50–1.38 (m, 1H), 1.16–1.05 (m, 1H); ¹³C-NMR (100 MHz, CD₃OD) δ : 166.9, 166.8, 165.8, 165.6 (d, *J* = 7.2 Hz), 163.3, 163.2 162.9 (dd, *J* = 248.5, 13.3 Hz), 159.8, 159.8 (d, *J* = 5.9 Hz), 144.5, 142.6 (t, *J* = 124.5 Hz), 141.1, 140.8, 132.5, 129.0, 127.9, 123.6, 120.5 (q, *J* = 268.1 Hz), 120.1 (t, *J* = 243.0 Hz), 113.4 (d, *J* = 24.2 Hz) 113.2 (d, *J* = 13.1 Hz), 103.1(t, *J* = 26.4 Hz), 89.4, 86.3, 79.6, 59.4, 54.1, 53.9 (d, *J* = 20.5 Hz), 42.0, 41.8, 35.7, 29.3, 28.7 28.9 (dd, *J* = 34.4, 29.5 Hz), , 23.8, 23.7 13.1; HRMS (*m/z*): [M+Na]⁺ calculated 990.1327, found 990.1302.

¹H NMR spectra of GS-6207 in CDCl₃ and DMSO-*d*₆, as well as an HSQC spectrum of GS-6207 in DMSO-*d*₆ have been included in the Supplementary Materials. An additional ¹H NMR spectra of GS-6207 after standing in solution for 4 days in deuterated methanol is also included to

Stability of GS-6207 samples at room temperature in MeOH or DMSO

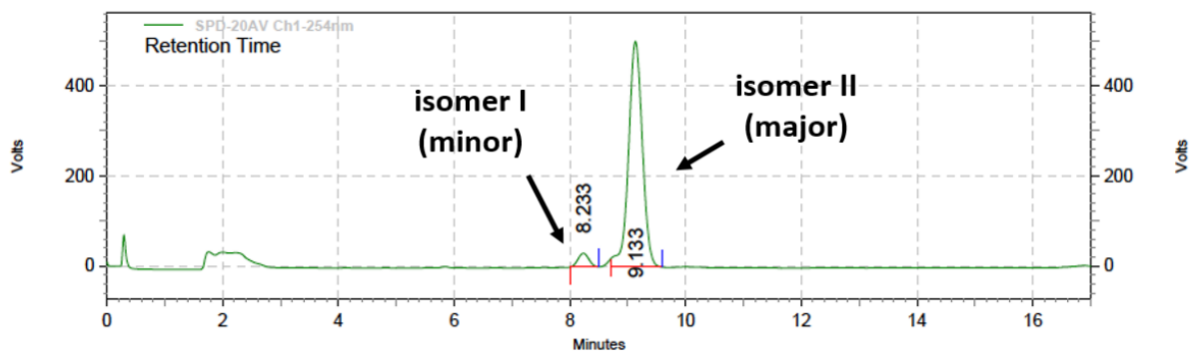
time (h)	MeOH rt		DMSO rt	
	% isomer I	% isomer II	% isomer I	% isomer II
0	5.0	95.0	5.0	95.0
1	6.0	94.0	7.0	93.0
2	8.0	92.0	9.5	90.5
24	14.0	86.0	17.0	83.0
48	16.1	83.9	17.0	83.0
120	15.8	84.2	18.1	81.9

Stability of GS-6207 samples through freeze/thaw (f/t) cycles in MeOH or DMSO

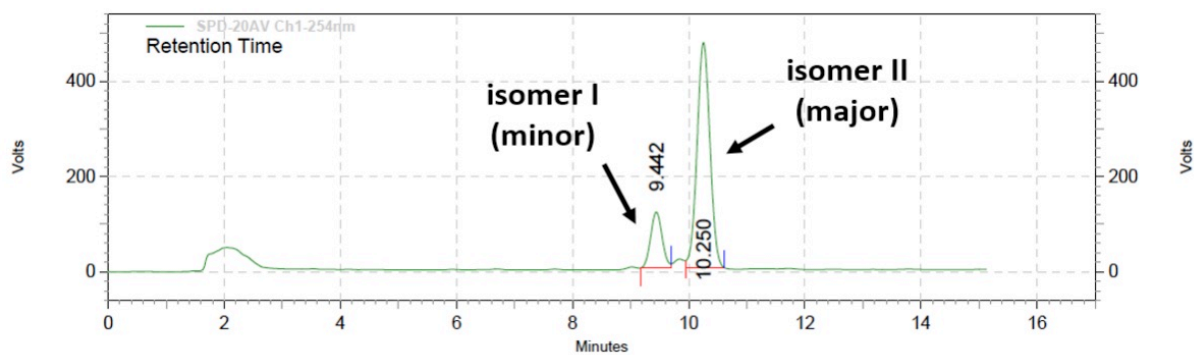
time (h)	MeOH f/t		DMSO f/t	
	% isomer I	% isomer II	% isomer I	% isomer II
0	5.0	95.0	5.0	95.0
24	6.0	94.0	7.0	93.0
48	9.3	90.7	17.0	83.0
120	10.0	90.0	18.0	82.0
144	11.5	88.5	20.0	80.0

Representative HPLC traces

GS-6207 in DMSO (freeze/thaw); time = 0 h



GS-6207 in DMSO (freeze/thaw); time = 144 h



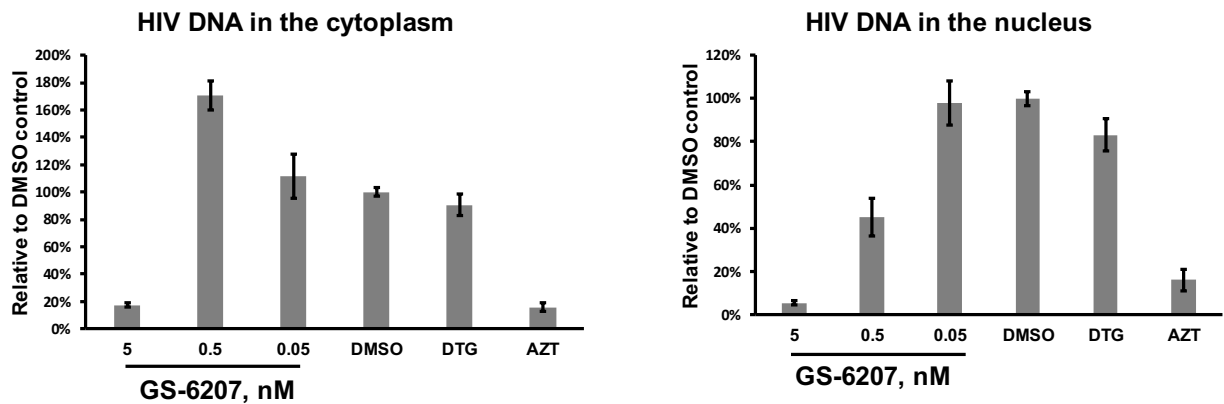


Fig. S1

Effects of GS-6207 on HIV-1 DNA levels in the cytoplasm and nuclei of infected cells. At a pharmacologically relevant concentration (5 nM), GS-6207 decreased HIV-1 DNA levels in both the cytoplasm ($p = 0.000638$) and nucleus ($p = 0.00017$). In control experiments the reverse transcriptase inhibitor AZT yielded similarly reduced levels of viral DNA in the cytoplasm and nucleus. At a lower concentration (0.5 nM), GS-6207 increased HIV-1 DNA levels in the cytoplasm ($p = 0.00741$) and decreased nuclear viral DNA levels ($p = 0.00256$) compared with the DMSO control. For comparison, DTG resulted in similar or slightly reduced viral DNA levels in the nucleus compared with the DMSO control, whereas 2-LTR circles were increased in the presence vs absence of DTG (Fig. 1C). 0.05 nM GS-6207 behaved largely as a negative control DMSO in all assays (see above and Fig 1).

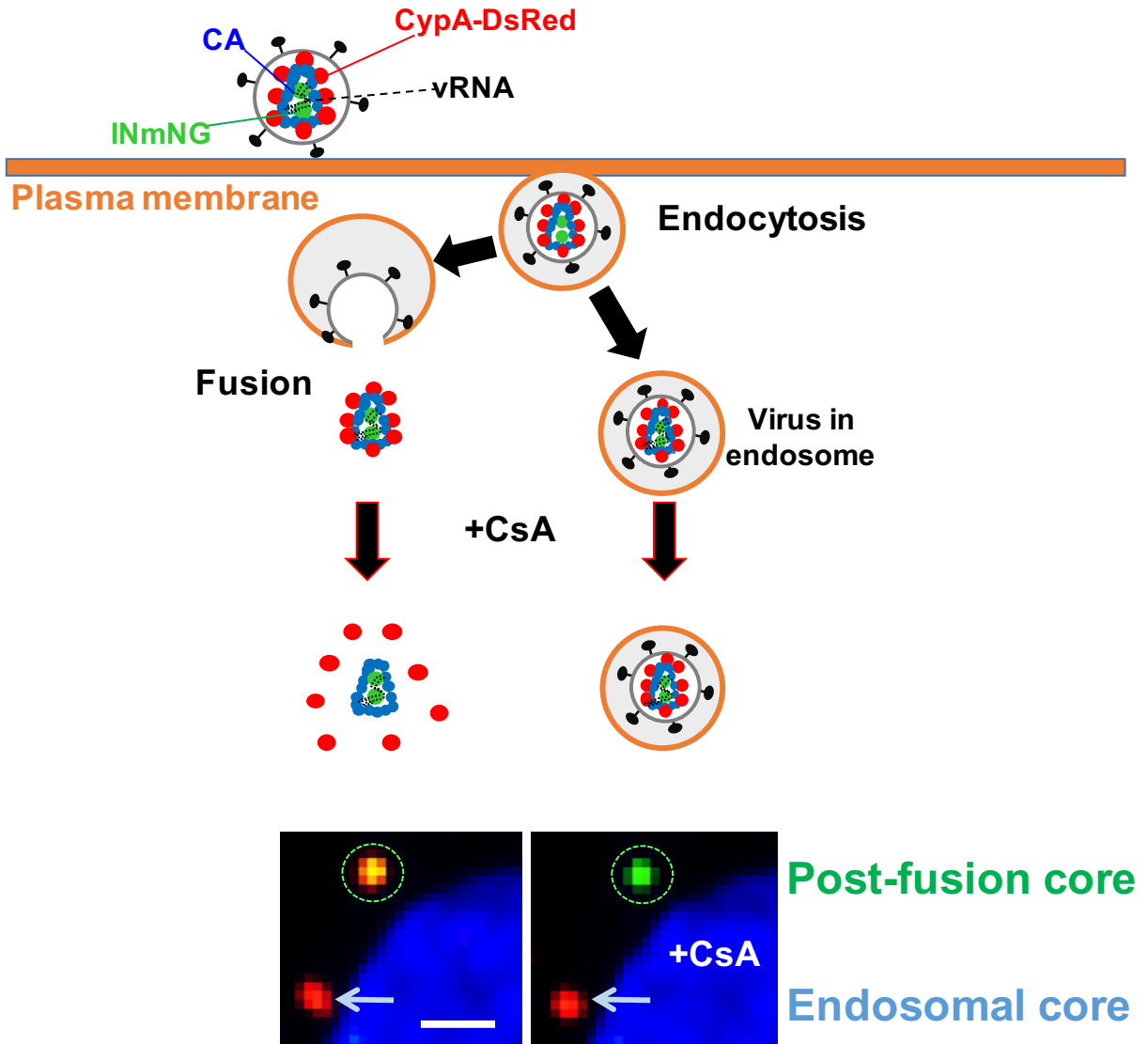


Fig. S2

Effects of GS-6207 on stability of HIV-1 cores in the cytoplasm. 50,000 TZM-bl cells were infected with VSV-G pseudotyped HIV-1 fluorescently labeled with INmNG (green) and CypA-DsRed/CA marker (red) at MOI 0.05 in the presence or absence of GS-6207 (see Fig 1D). At 3 hpi, post-fusion cytoplasmic CypA-DsRed-positive HIV-1 cores were identified by the addition of 25 μ M CsA, which displaces CypA-DsRed from the CA shell of the virus core. Images of an infected cell in the absence of GS-6207 showing CsA-induced loss of CypA-DsRed from the cytoplasmic post-fusion core (circle). The virus trapped in an endosome does not respond to CsA treatment (arrow). Numbers of post-fusion HIV-1 cores per cell determined by quantifying the number of CypA-DsRed puncta in the cytoplasm that disappeared after CsA treatment is reported in Fig. 1D.

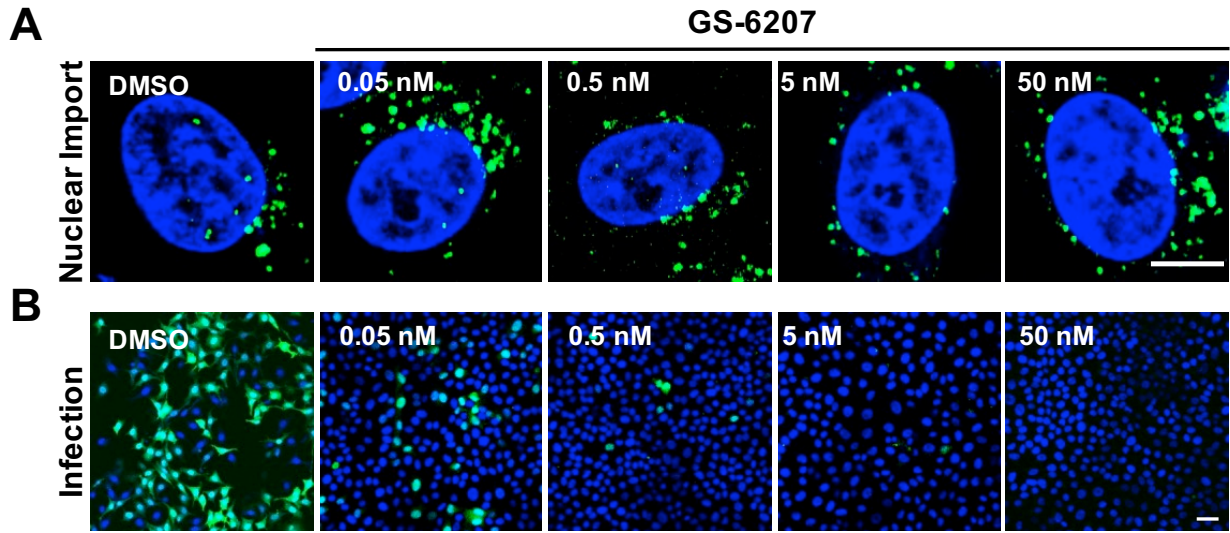


Fig. S3

Effects of GS-6207 on HIV-1 nuclear import and infectivity. HIV-1 nuclear import and infectivity were measured in TZM-bl cells at 4 and 24 hpi, respectively. (A) Representative images of single HIV-1 INmNG complexes (green) in the nuclei of cells treated with indicated concentrations of GS-6207 or DMSO. (B) Representative images of HIV-1 infection in the presence of GS-6207 from the same plates as shown in (A). Scale bars in (A) 5 μm and (B) 100 μm . Quantitative results are shown in Fig. 1E and F.

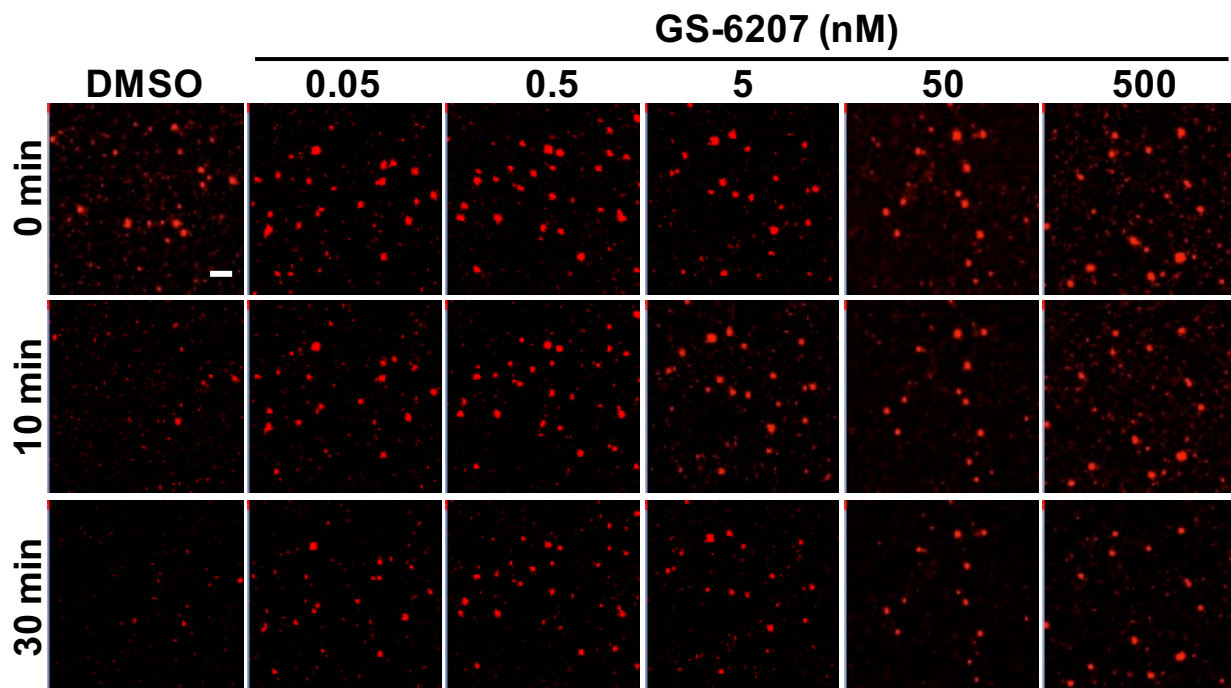


Fig. S4

Effects of GS-6207 on stability of HIV-1 cores *in vitro*. VSV-G pseudotyped HIV-1 particles fluorescently labeled with INmNG (not shown for clarity) and CypA-DsRed/CA marker (red) were bound to a poly-lysine-coated coverlip. Representative images show CypA-DsRed labeled HIV-1 CA (red puncta) after 0, 10 and 30 min of membrane permeabilization with saponin at room temperature in the presence of indicated concentrations of GS-6207 or DMSO. Scale bar is 2 μm . Quantitative results are shown in Fig. 1G.

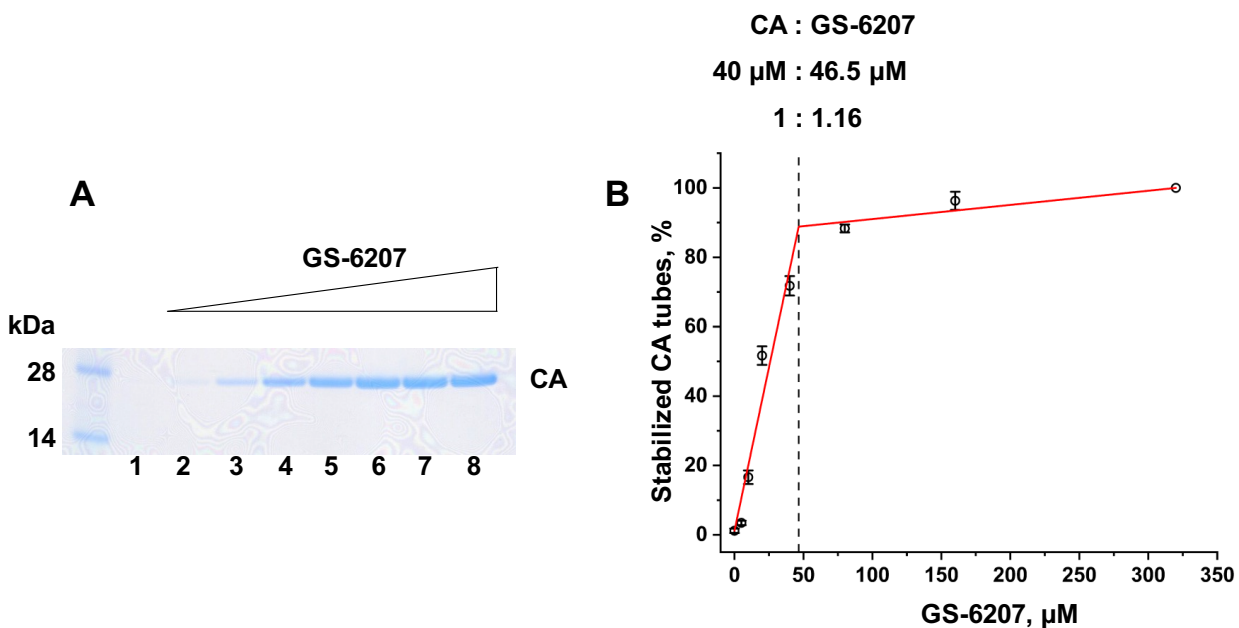


Fig. S5

Quantitative analysis of GS-6207 mediated stabilization of CA tubes in 150 mM NaCl. (A) Representative SDS-PAGE image. Pre-formed CA tubes (40 μ M) were incubated with the following concentrations of GS-6207: lane 1: no inhibitor, lane 2: 5 μ M, lane 3: 10 μ M, lane 4: 20 μ M, lane 5: 40 μ M, lane 6: 80 μ M, lane 7: 160 μ M, lane 8: 320 μ M. (B) Quantification of results in (A). The error bars indicate SDs from five independent SDS-PAGE analyses. The stoichiometry for CA:GS-6207 was determined using piecewise linear regression in Origin software (see Materials and Methods).

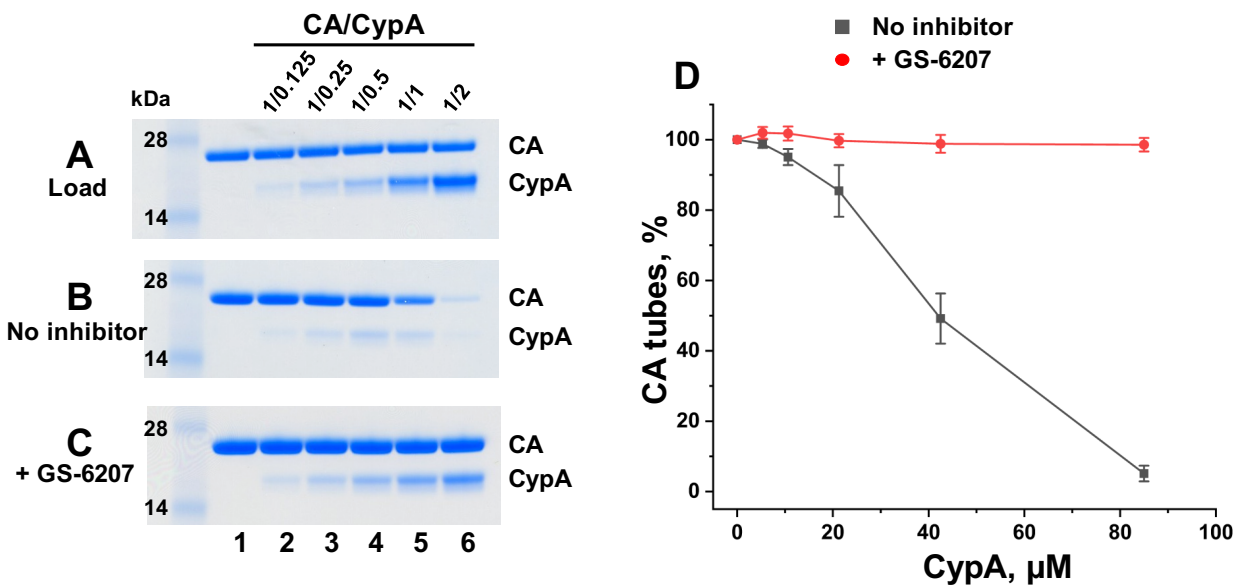


Fig. S6

Effects of CypA on stability of pre-formed CA tubes in the absence and presence of GS-6207.

(A) Loads of CA and CypA. The fixed concentration of CA (40 μ M) was used in all reactions. The CA/CypA ratio for each reaction is indicated. (B) CA tubes were preformed, DMSO control was added, and then the mixture was exposed to increasing concentrations of CypA. Representative results of the pelleted fractions are shown. (C) Same as in panel B, except 40 μ M GS-6207 instead of DMSO was added to preformed CA tubes. (D) Quantification of the results in (B) and (C). Error bars indicate SDs from three independent experiments.

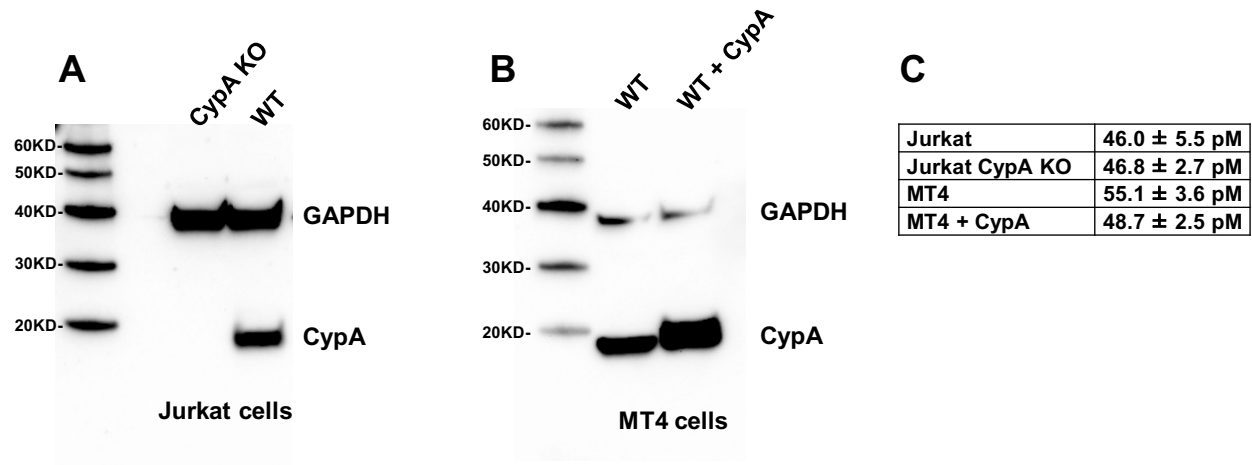


Fig. S7

Effects of CypA on antiviral activities of GS-6207. (A) Immunoblotting of CypA and GAPDH in WT and CypA knock-out (KO) Jurkat cells. (B) Immunoblotting of CypA and GAPDH in WT and CypA overexpressing MT4 cells. (C) Antiviral activities of GS-6207 in the indicated cells (average ± SD for n = 3 independent experiments).

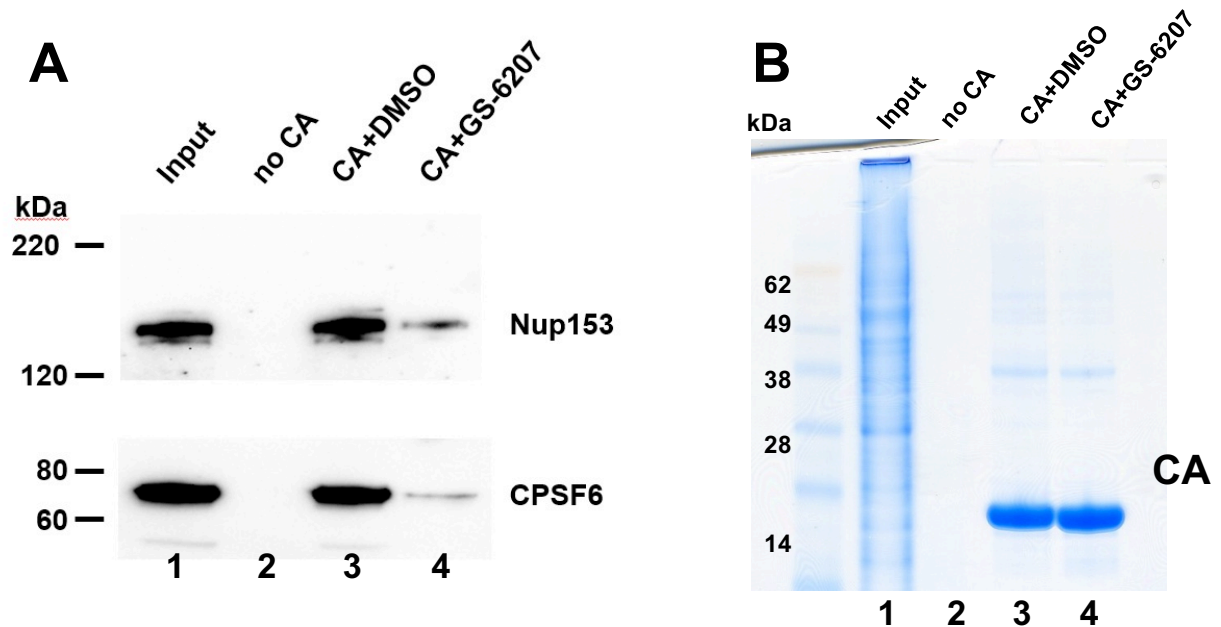


Fig. S8

Effects of GS-6207 on binding of cellular Nup153 and CPSF6 to CA tubes. Representative immunoblotting (A) and Coomassie stained gel (B) images. Lanes 1: cellular lysates; lanes 2: cellular lysates pelleted in 2 M NaCl containing buffer in the absence of CA tubes; lanes 3: preformed WT CA tubes (42.5 μ M) were incubated with DMSO and then cellular lysates pre-equilibrated in 2 M NaCl containing buffer were added to the mixture. Pelleted fractions were recovered through centrifugation and analyzed by immunoblotting (A) or Coomassie staining (B); lanes 4: preformed WT CA tubes (42.5 μ M) were incubated with 85 μ M GS-6207 and then cellular lysates pre-equilibrated in 2 M NaCl containing buffer were added to the mixture. Pelleted fractions were recovered through centrifugation and analyzed by immunoblotting (A) or Coomassie staining (B). 2 M NaCl was maintained in all buffers throughout all reactions, which insured that similar levels of CA tubes were formed with DMSO (lanes 3) and GS-6207 (lanes 4) thus allowing for accurate comparison of Nup153 and CPSF6 levels pelleted in the absence vs presence of the CA inhibitor.

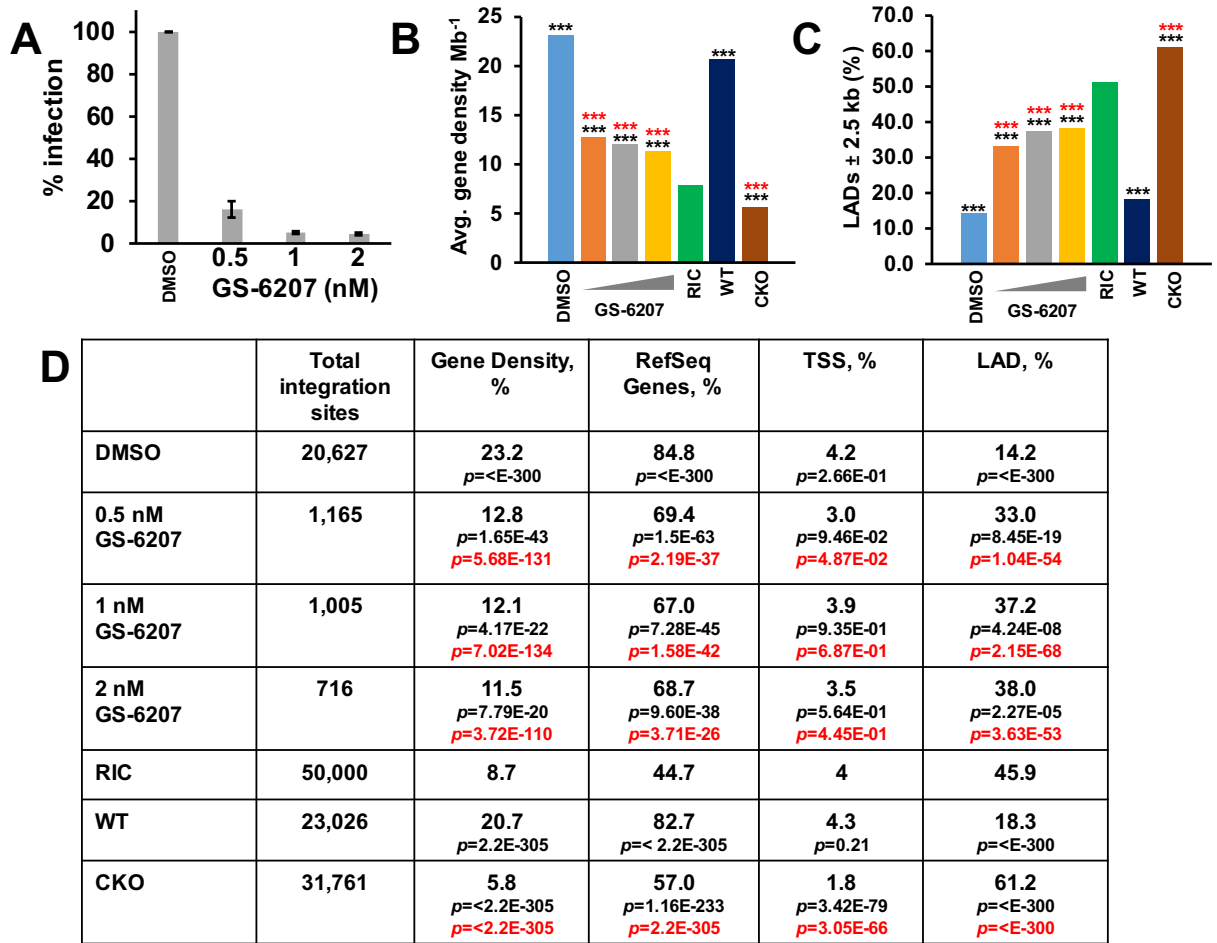


Fig. S9

Effects of GS-6207 on HIV-1 integration site selectivity in chromatin. (A) Infection (luciferase activity) results for the samples used for Illumina sequencing. HEK293T cells were incubated with indicated concentrations of GS-6207 or DMSO and then infected with VSV-G pseudotyped HIV-1. (B and C) Distribution of HIV-1 integration sites with respect to average gene density per Mb (B), and lamina-associated domains (LADs) (C). The triangle indicates infections conducted in the presence of GS-6207 (0.5, 1.0, and 2.0 nM). Random integration control (RIC) and historical values from WT and matched CPSF6 knockout (CKO) HEK293T cell lines are shown (13). P values (***) $p < 0.0001$ were calculated by Fisher's exact test or by Wilcoxon-Mann-Whitney rank sum test (for gene density). Black asterisks, differences as compared to RIC; red asterisks, differences compared to matched DMSO (for drug-treated samples) or WT (for CKO) controls. (D) Distribution of HIV-1 integration sites with respect to average gene density per Mb, Refseq genes, transcriptional start sites (TSSs) and LADs. RIC values and reported values from WT and

matched CKO HEK293T cell lines are shown. *P* values compare differences with RIC (black) and matched DMSO or WT controls (in red).

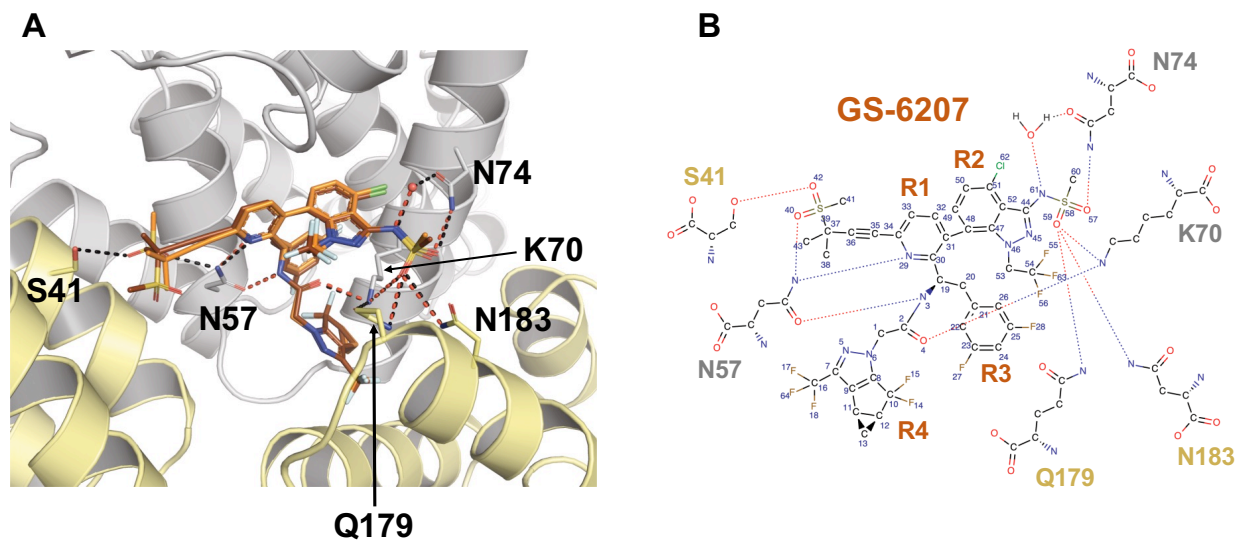


Fig. S10

Interactions of two GS-6207 conformations with a CA hexamer. (A) X-ray crystal structure reveals two conformations of GS-6207 (conformation 1: orange; conformation 2: chocolate) bound to the CA_{A14C/E45C/W184A/M185A} hexamer at the binding pocket formed by adjacent CA subunits CA1 (light grey) and CA2 (pale yellow) (PDB ID: 6VKV). Hydrogen bond interactions of conformation 1 and 2 are denoted by black and red dashed lines, respectively. In general, the two conformations form very similar interactions with the binding pocket. The only noticeable differences between them are seen in positioning of the 2-(methanesulfonyl)-2-methylpropane group, which establishes hydrogen bonding with S41 through O42, and N57 through O40 only in conformation 1 but not 2. The 2-(methanesulfonyl)-2-methylpropane group is located between polar side chains and a hydrophobic pocket, which affords this group sufficient flexibility to interact with the binding pocket in either conformation. For simplicity, in other figures we show only conformation 1. Comparison of our crystal structure with the published docked GS-6207 conformation (2) reveals a difference in the position of the trifluoroethyl moiety on the indazole ring. While this group on the docked compound (2) appears to point toward the difluorobenzyl and cyclopenta-pyrazole rings, our crystal structure shows that this moiety is positioned nearer to the pyridine ring. (B) Chemical structure of GS-6207 highlighting the key interactions of conformation 1 of GS-6207 with CA hexamer residues through colored dashed lines and the major ring systems labeled R1-R4. Two ring systems R3 and R4 establish extensive Van der Waals interactions with CA1-NTD and CA2-CTD. The difluorobenzyl group (R3) docks deep into a hydrophobic pocket of CA1-

NTD, which includes residues L56, V59, M66, L69, K70, I73 and Y130. The cyclopenta-pyrazole ring (R4) engages with CA1-NTD and CA2-CTD residues Q63, M66, Q67, K70, Y169, L172, and R173. R1 and R2 also provide interactions with the protein. The pyridinium ring (R1) engages with CA1-NTD residues N53, G106, and T107, and the 2-(methanesulfonyl)-2-methylpropane group of R1 interacts with CA2-NTD residues I37 and P38 as well as CA1-NTD residue T54. The indazole ring (R2) forms strong Van der Waals interactions with several CA1-NTD residues including K70, I73, A105, T107, and Y130. In addition, GS-6207 establishes a hydrogen bonding network with both CA subunits. A significant number of these interactions were previously predicted by docking GS-CA1 and GS-6207 into the CA hexamer binding pocket (2). The following three residues of the CA1-NTD engage with the inhibitor: the side chain of N57 interacts with N3 positioned between R1 and R3, N29 on R1, and O40 on the sulfonyl group of R1; K70 engages with O4 between R3 and R4, and O59 on the sulfonamide group of R2; N74 interacts with O57 of the sulfonamide group. The previously published CA structure with docked GS-6207 predicted the interaction between K70 and C=O of amide group of GS-6207 that we observe in the crystal structure (2). CA2 provides additional hydrogen bonding to the inhibitor including S41 of the CA2-NTD contacting with O42 of the sulfonyl group of R1, and both Q179 and N183 of the CA2-CTD engage with O59 of the sulfonamide group.

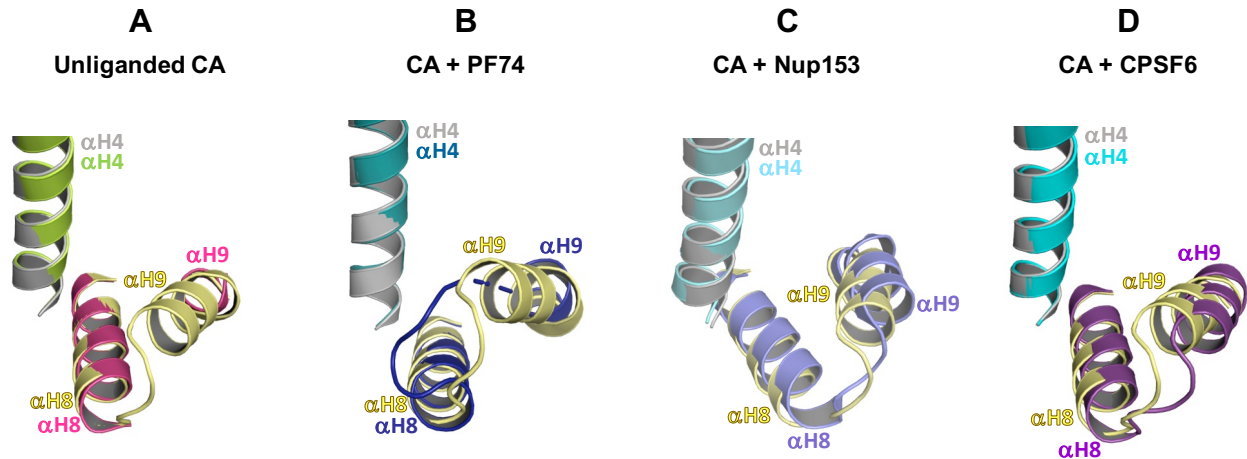


Fig. S11

Conformational variability in α H9 of CA_{A14C/E45C/W184A/M185A} hexamers in the absence and presence of different ligands. (A) Comparison of α H4, α H8, and α H9 of GS-6207 bound (grey/pale yellow, PDB ID: 6VKV) vs unliganded (lime/warm pink, PDB ID: 3H47) CA hexamers. The majority of α H9 residues (179-187) in unliganded CA hexamer lacked defined electron density and are not shown. (B) Comparison of α H4, α H8, and α H9 of CA hexamer structures in complex with GS-6207 (grey/pale yellow) vs PF74 (dark teal/indigo, PDB ID: 4U0E). Several α H9 residues (181-185) in the CA-PF74 structure lacked defined electron density and are not shown. (C) Comparison of α H4, α H8, and α H9 of CA hexamer structures in complex with GS-6207 (grey/pale yellow) vs the Nup153 peptide (aquamarine/slate, PDB ID: 5TSV). α H9 has a defined electron density in the presence of bound Nup153 peptide. However, α H9 is positioned closer to α H4 in the presence of GS-6207 compared with the Nup153 peptide. (D) Comparison of α H4, α H8, and α H9 of CA hexamer structures in complex with GS-6207 (grey/pale yellow) vs the CPSF6 peptide (cyan/purple, PDB ID: 4WYM). The α H9 helix has a defined electron density in the presence of bound CPSF6 peptide. However, α H9 is positioned closer to α H4 in the presence of GS-6207 compared with the CPSF6 peptide.

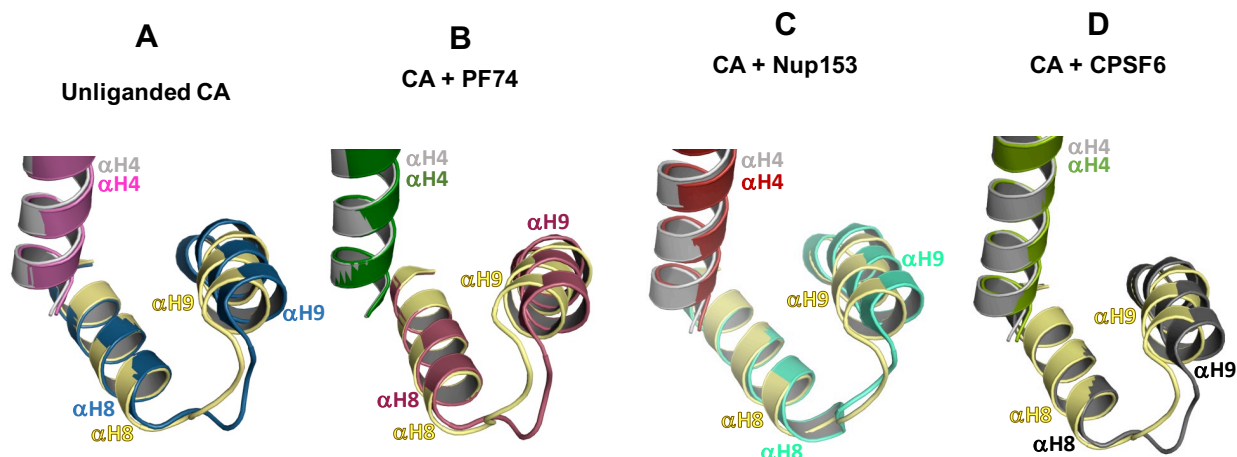


Fig. S12

Comparison of α H9 of the GS-6207 bound cross-linked CA hexamer with native CA hexamers in the absence and presence of different ligands. (A) Comparison of α H4, α H8, and α H9 of GS-6207 bound $CA_{A14C/E45C/W184A/M185A}$ hexamer (grey/pale yellow, PDB ID: 6VKV) vs unliganded native CA hexamer (pink/Prussian blue, PDB ID: 4XFX). α H9 is positioned closer to α H4 in the presence of GS-6207 compared to in its absence. (B) Comparison of α H4, α H8, and α H9 of GS-6207 bound $CA_{A14C/E45C/W184A/M185A}$ hexamer (grey/pale yellow, PDB ID: 6VKV) vs PF74 bound native CA hexamer (forest/raspberry, PDB ID: 4XFZ). α H9 is positioned closer to α H4 in the presence of GS-6207 compared with PF74. (C) Comparison of α H4, α H8, and α H9 of GS-6207 bound $CA_{A14C/E45C/W184A/M185A}$ hexamer (grey/pale yellow, PDB ID: 6VKV) vs the Nup153 peptide bound to native CA hexamer (red/mint, PDB ID: 6AYA). α H9 is positioned closer to α H4 in the presence of GS-6207 compared with the Nup153 peptide. (D) Comparison of α H4, α H8, and α H9 of GS-6207 bound $CA_{A14C/E45C/W184A/M185A}$ hexamers (grey/pale yellow, PDB ID: 6VKV) vs the CPSF6 peptide bound to native CA hexamer (green/slate, PDB ID: 6AY9). α H9 is positioned closer to α H4 in the presence of GS-6207 compared with the CPSF6 peptide.

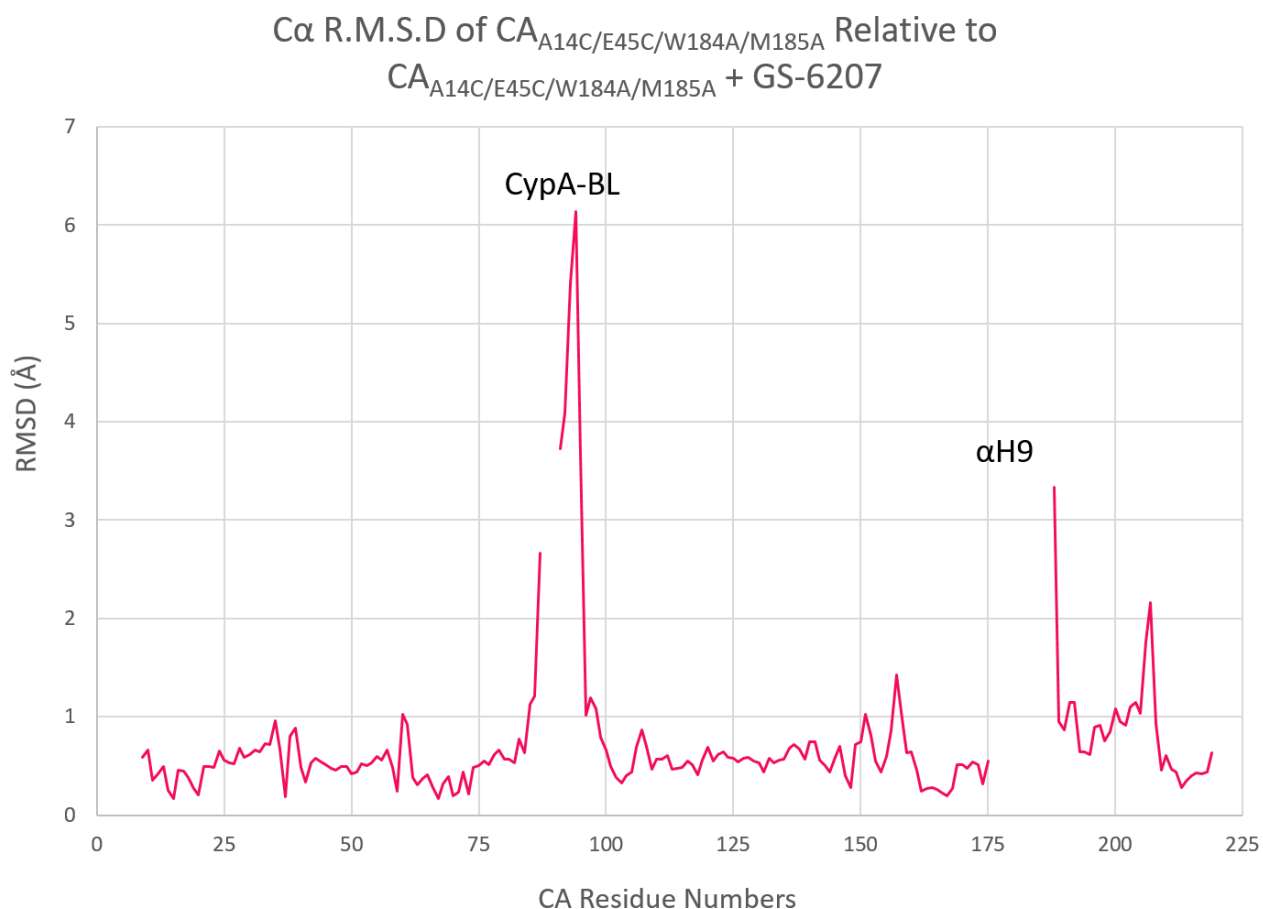


Fig. S13

Structural comparison of unliganded cross-linked CA hexamer with cross-linked CA hexamer + GS-6207 through alpha carbon root mean square deviation (RMSD). RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of CA_{A14C/E45C/W184A/M185A} (PDB ID: 3H47) alpha carbons measured against CA_{A14C/E45C/W184A/M185A} + GS-6207 is shown (PDB ID: 6VKV). The unliganded CA hexamer lacks defined electron density for a region consisting of the majority of α H9 residues (176-187). Therefore, these regions could not be compared by RMSD and are shown as blank. However, the observation that the majority of α H9 is disordered in the absence of the inhibitor, whereas the entire α H9 is very well ordered in the presence of GS-6207, is telling. A few amino acids (88-90) in the CypA binding loop (CypA-BL; residues 85-93) also lack electron density in the absence of the inhibitor and are shown as blank.

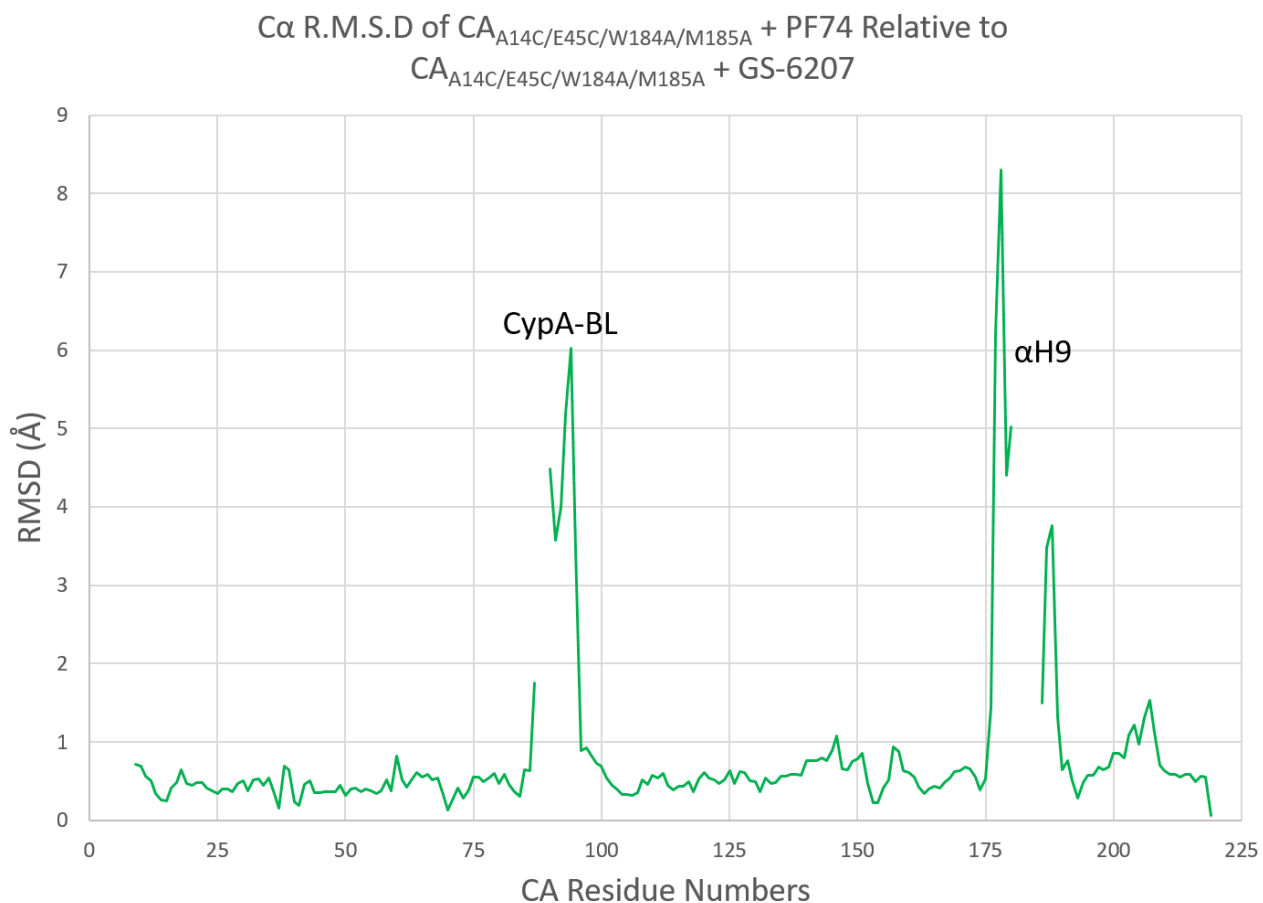


Fig. S14

RMSD-based structural comparison of cross-linked CA hexamer + PF-74 with cross-linked CA hexamer + GS-6207. RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of CA_{A14C/E45C/W184A/M185A} + PF-74 (PDB ID: 4U0E) alpha carbons measured against CA_{A14C/E45C/W184A/M185A} + GS-6207 is shown (PDB ID: 6VKV). Several α H9 residues (181-185) lack defined electron density in the CA hexamer + PF-74 structure and cannot be compared with CA hexamer + GS-6207 by RMSD. However, the observation that α H9 is partly disordered in the presence of PF-74, whereas the entire α H9 very well ordered in the presence of GS-6207, is telling. In addition, the segments of α H9 that yield defined density in the CA + PF74 structure show marked difference compared with CA + GS-6207. Two residues (88-89) within the CypA-BL in the CA hexamer + PF-74 structure also lack defined electron density and therefore are shown as blank.

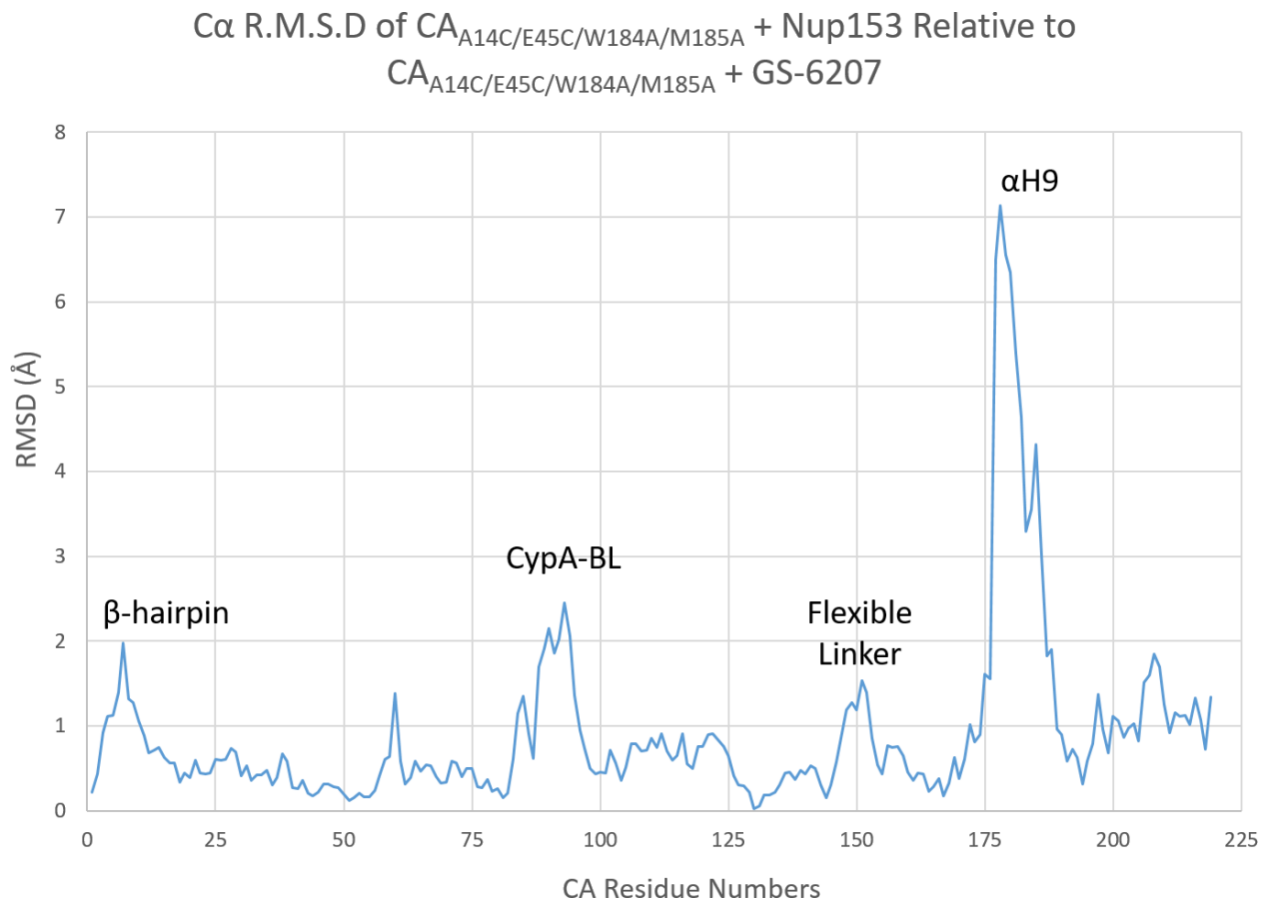


Fig. S15

RMSD-based structural comparison of cross-linked CA hexamer + Nup153 to cross-linked CA hexamer + GS-6207. RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of CA_{A14C/E45C/W184A/M185A} + Nup153 peptide (PDB ID: 5TSV) alpha carbons measured against CA_{A14C/E45C/W184A/M185A} + GS-6207 is shown (PDB ID: 6VKV). The most pronounced difference is observed at α H9 (residues 179-192). There are also minor differences at the flexible β -hairpin at the beginning of the NTD, the CypA-BL, and the flexible linker between the NTD and CTD.

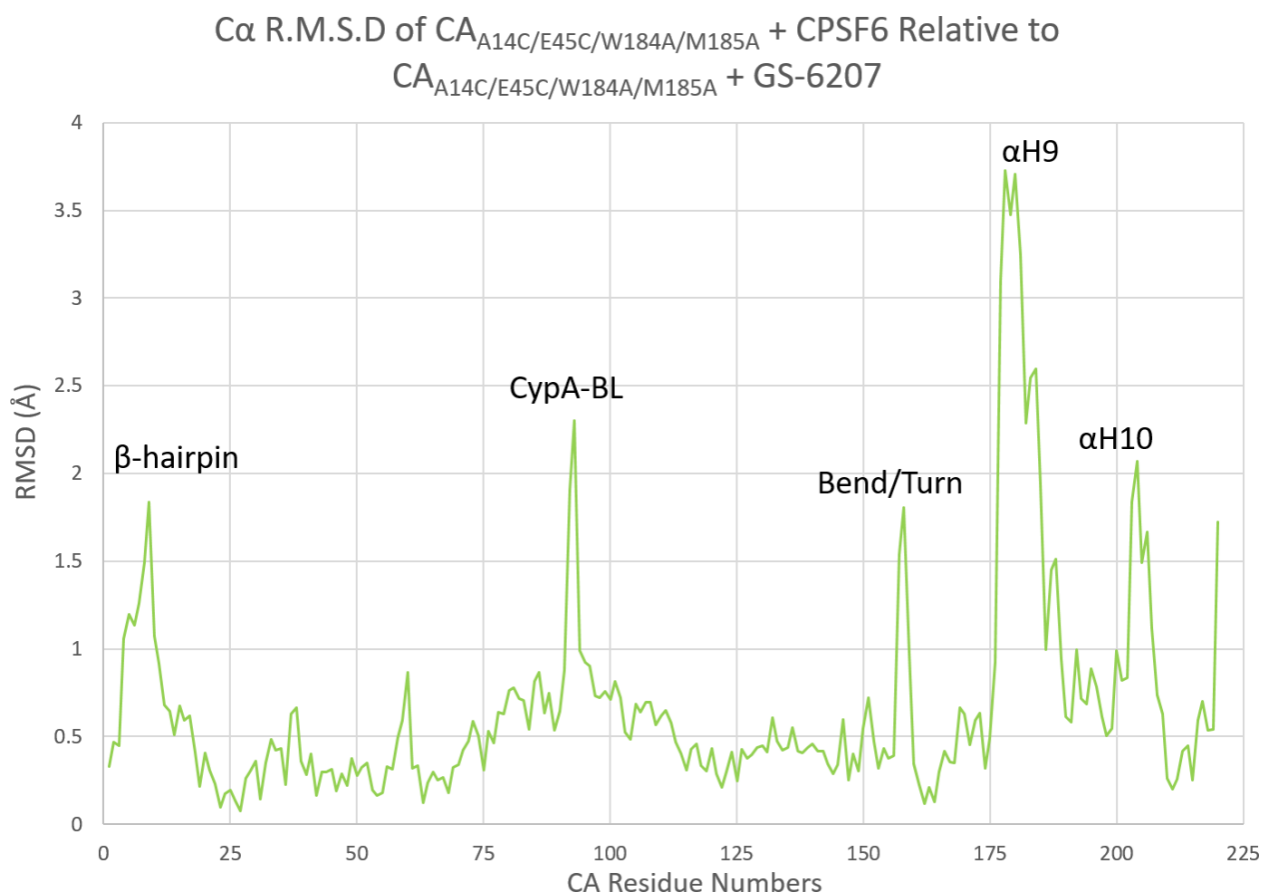


Fig. S16

RMSD-based structural comparison of cross-linked CA hexamer + CPSF6 to cross-linked CA hexamer + GS-6207. RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of $\text{CA}_{\text{A14C/E45C/W184A/M185A}}$ + CPSF6 peptide (PDB ID: 4WYM) alpha carbons measured against $\text{CA}_{\text{A14C/E45C/W184A/M185A}}$ + GS-6207 is shown (PDB ID: 6VKV). The most pronounced difference is observed at α H9 (179-192). There are a few additional minor differences at the flexible β -hairpin at the beginning of the NTD, the CypA-BL, α H10 (196-205), and a flexible region (157-158) that can be a turn or bend depending on the structure. The relevance of the structural deviation between the terminal few residues of these structures may be incidental.

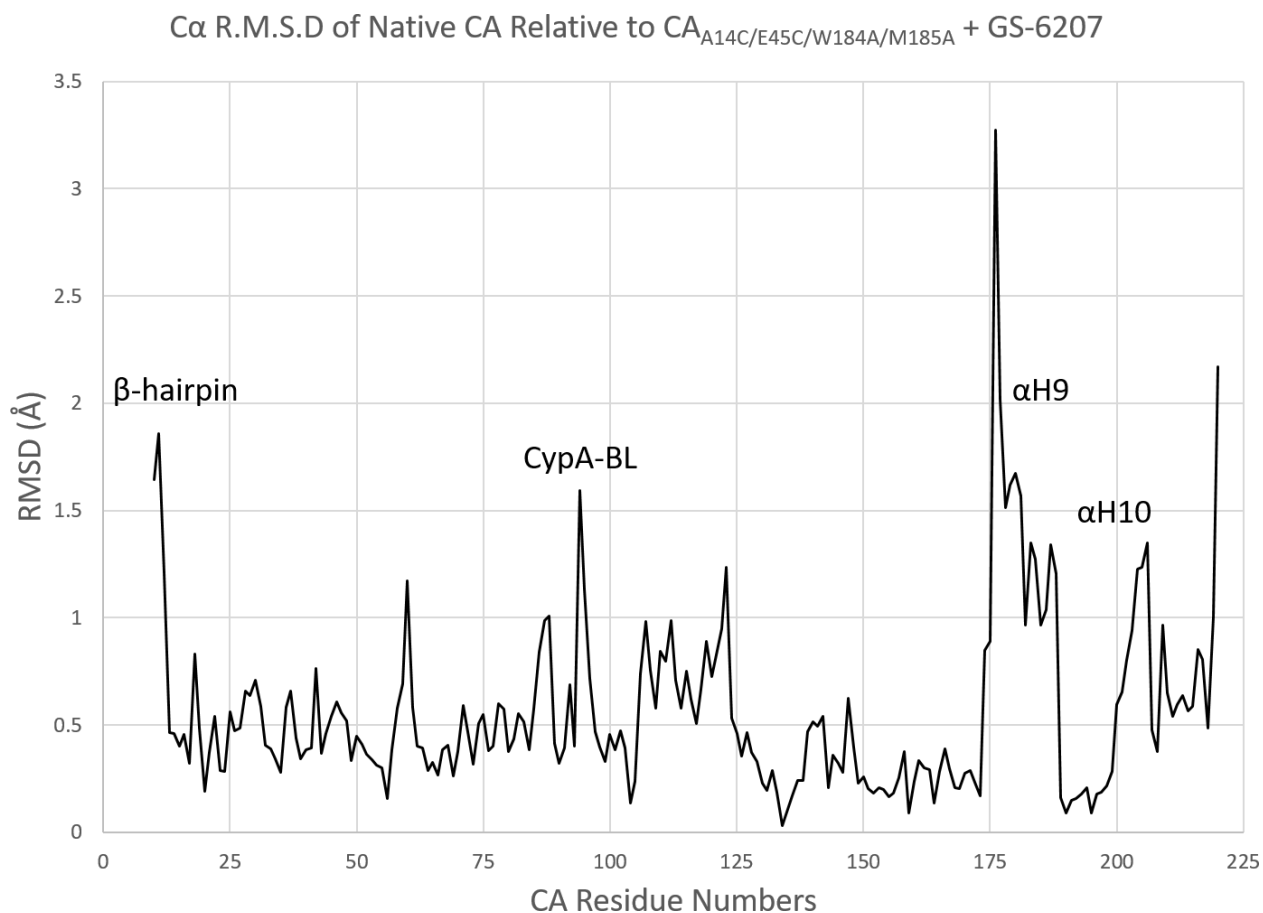


Fig. S17

RMSD-based structural comparison of native CA hexamer to cross-linked CA hexamer + GS-6207. RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of native CA (PDB ID: 4XFX) alpha carbons measured against CA_{A14C/E45C/W184A/M185A} + GS-6207 is shown (PDB ID: 6VKV). The most pronounced difference is observed at α H9 (179-192). There are other differences at the CypA-BL, α H10 (196-205), and to an extent the flexible β -hairpin at the beginning of the NTD. A few residues in the β -hairpin of the native CA hexamer structure lacked defined electron density, which means they are not present and cannot be fully compared. Additionally, a few residues at the end of the CTD appear to show increased variability between the structures, though this difference may be incidental.

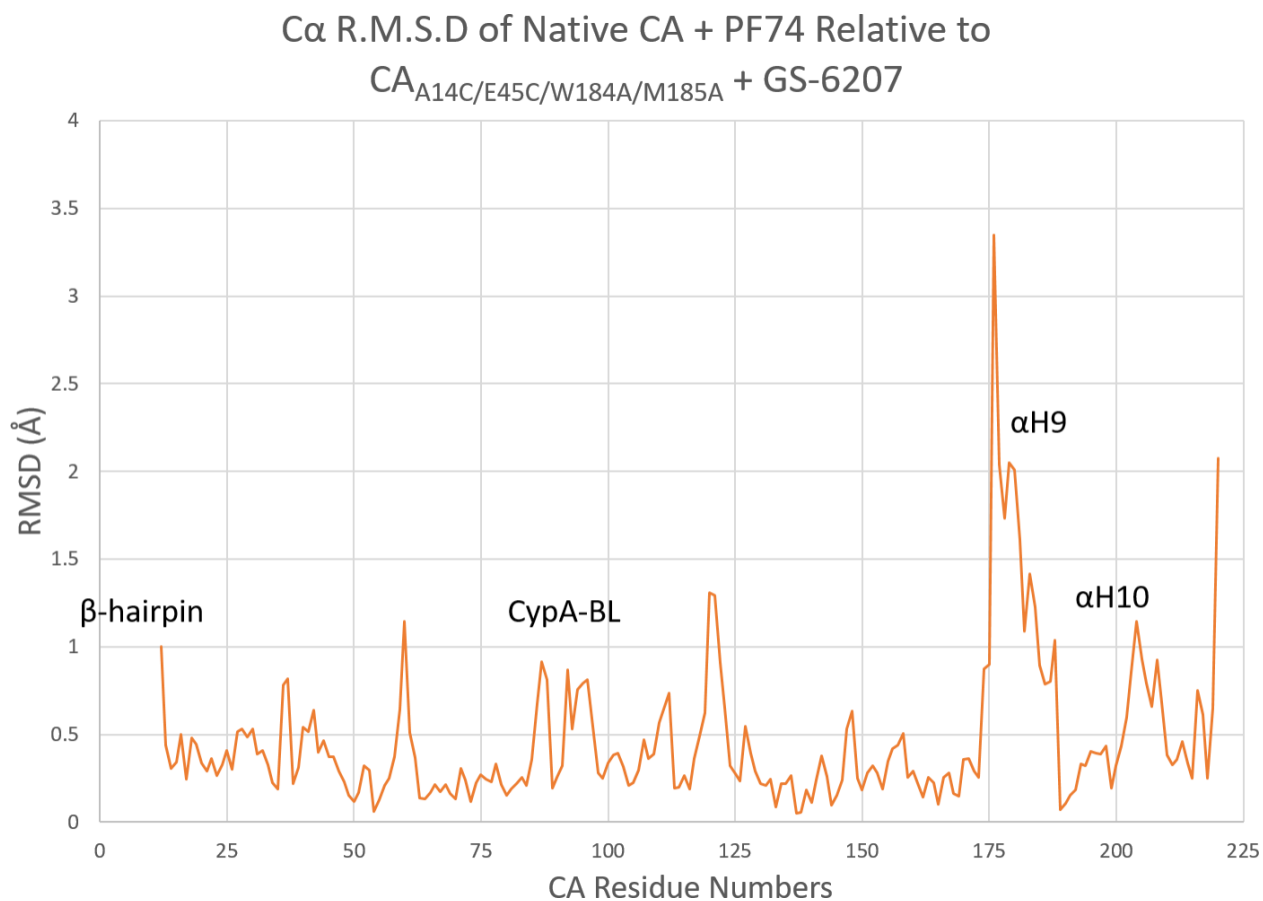


Fig. S18

RMSD-based structural comparison of native CA hexamer + PF-74 to cross-linked CA hexamer + GS-6207. RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of native CA + PF-74 (PDB ID: 4XFZ) alpha carbons measured against CA_{A14C/E45C/W184A/M185A} + GS-6207 is shown (PDB ID: 6VKV). The most marked difference is observed at α H9 (179-192). There are minor differences that center around residues 60 and 120 within the NTD, as well as α H10 (196-205). A few residues in the β -hairpin lacked defined electron density, which means they could not be fully compared. Additionally, a few residues at the end of the CTD appear to show increased variability between the structures. The last few residues of a structure can show increased flexibility due to physiological irrelevant reasons such as crystal packing. Therefore, it is possible this difference is incidental.

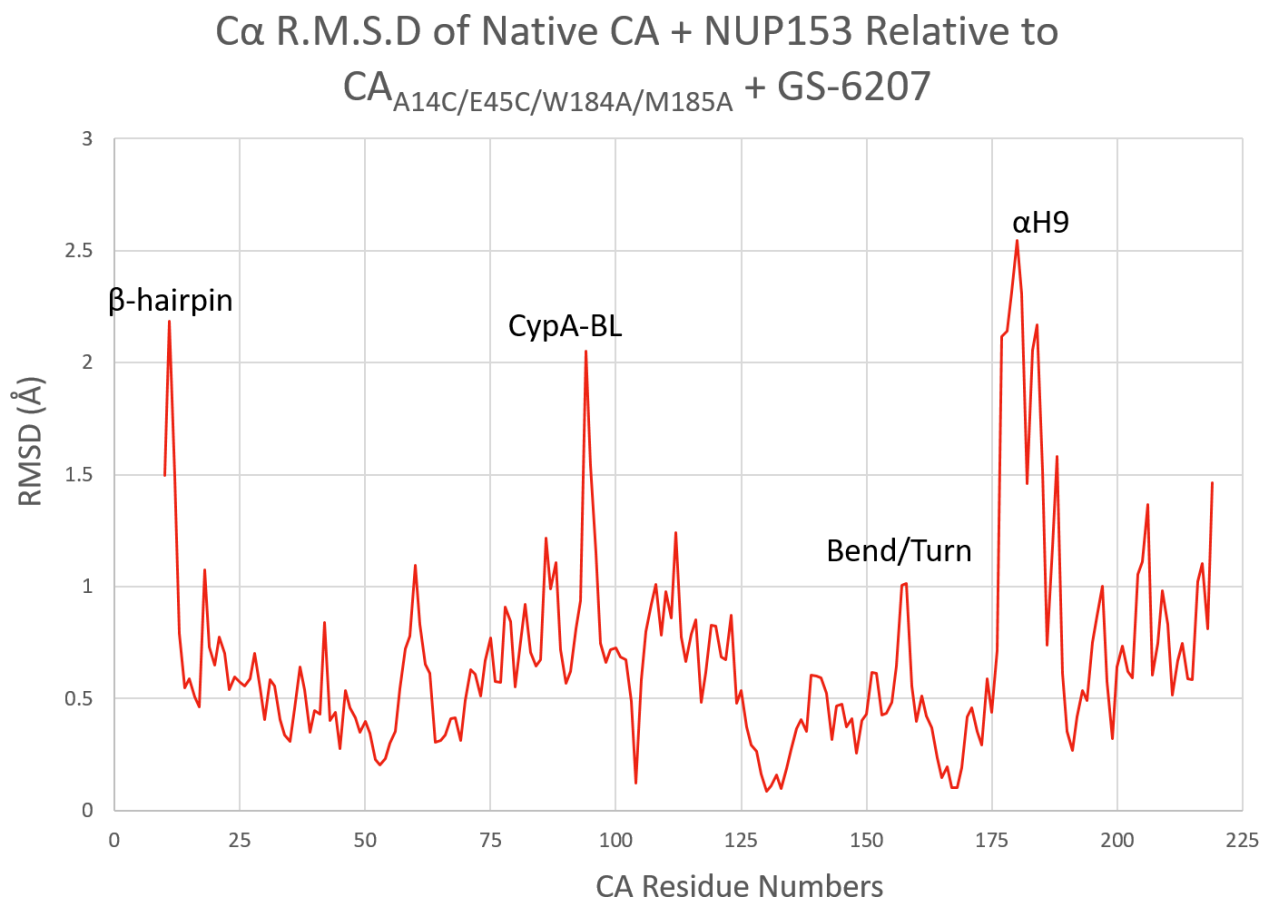


Fig. S19

RMSD-based structural comparison of native CA hexamer + Nup153 to cross-linked CA hexamer + GS-6207. RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of native CA + Nup153 peptide (PDB ID: 6AYA) alpha carbons measured against CA_{A14C/E45C/W184A/M185A} + GS-6207 is shown (PDB ID: 6VKV). The most marked difference is observed at α H9 (179-192). There are other differences at the flexible β -hairpin at the beginning of the NTD, the CypA-BL, and a flexible region (157-158) that can be a turn or bend depending on the structure. A few residues in the β -hairpin lacked defined electron density, which means they are not present and cannot be fully compared.

α R.M.S.D of Native CA + CPSF6 Relative to
 $CA_{A14C/E45C/W184A/M185A} + GS-6207$

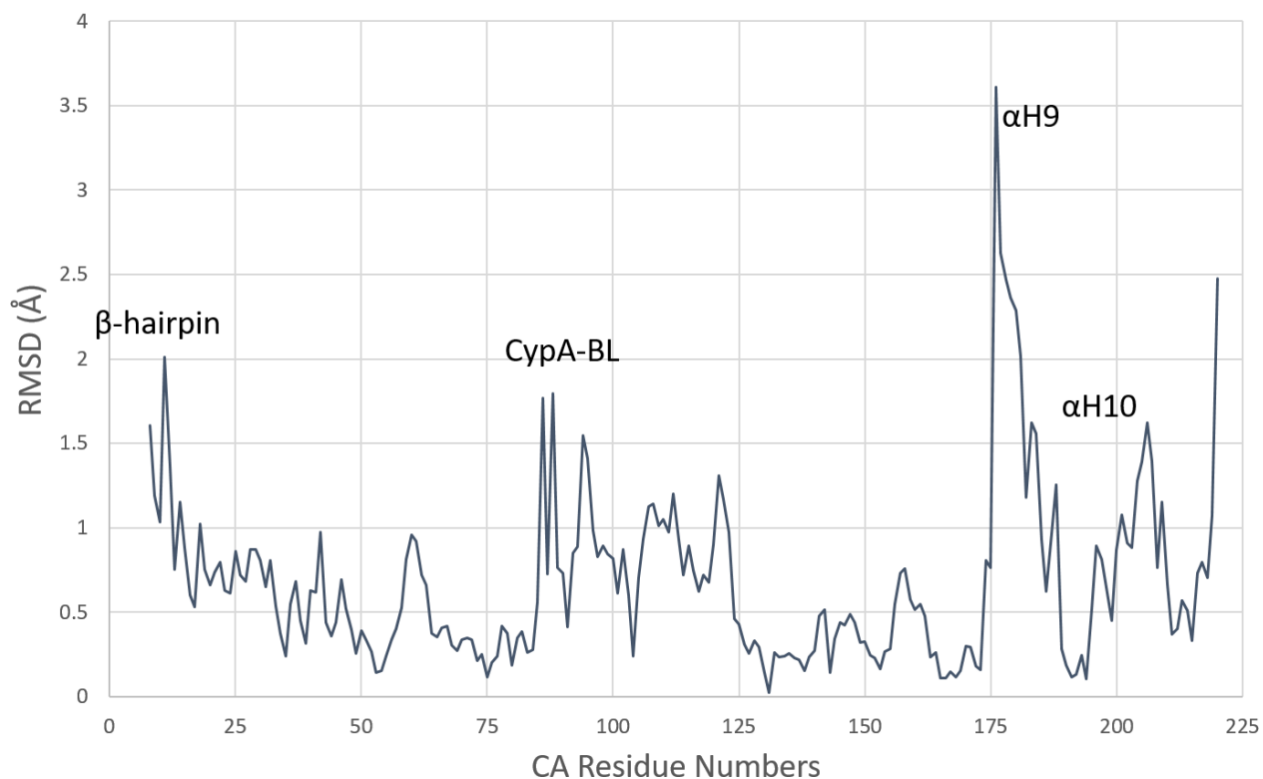


Fig. S20

RMSD-based structural comparison of native CA hexamer + CPSF6 to cross-linked CA hexamer + GS-6207. RMSDs were measured using Superpose (57) in CCP4. A line graph of the RMSD of native CA + CPSF6 peptide (PDB ID: 6AY9) alpha carbons measured against $CA_{A14C/E45C/W184A/M185A} + GS-6207$ (PDB ID: 6VKV) is shown. The most marked difference is observed at α H9 (179-192). There are other differences at the flexible β -hairpin at the beginning of the NTD, the CypA-BL, and α H10 (196-205). A few residues in the β -hairpin lacked defined electron density and thus could not be fully compared. Additionally, a few residues at the end of the CTD appear to show increased variability between the structures. The last few residues of a structure can show increase flexibility due to unconstrained movement. Therefore, it is possible this difference is incidental.

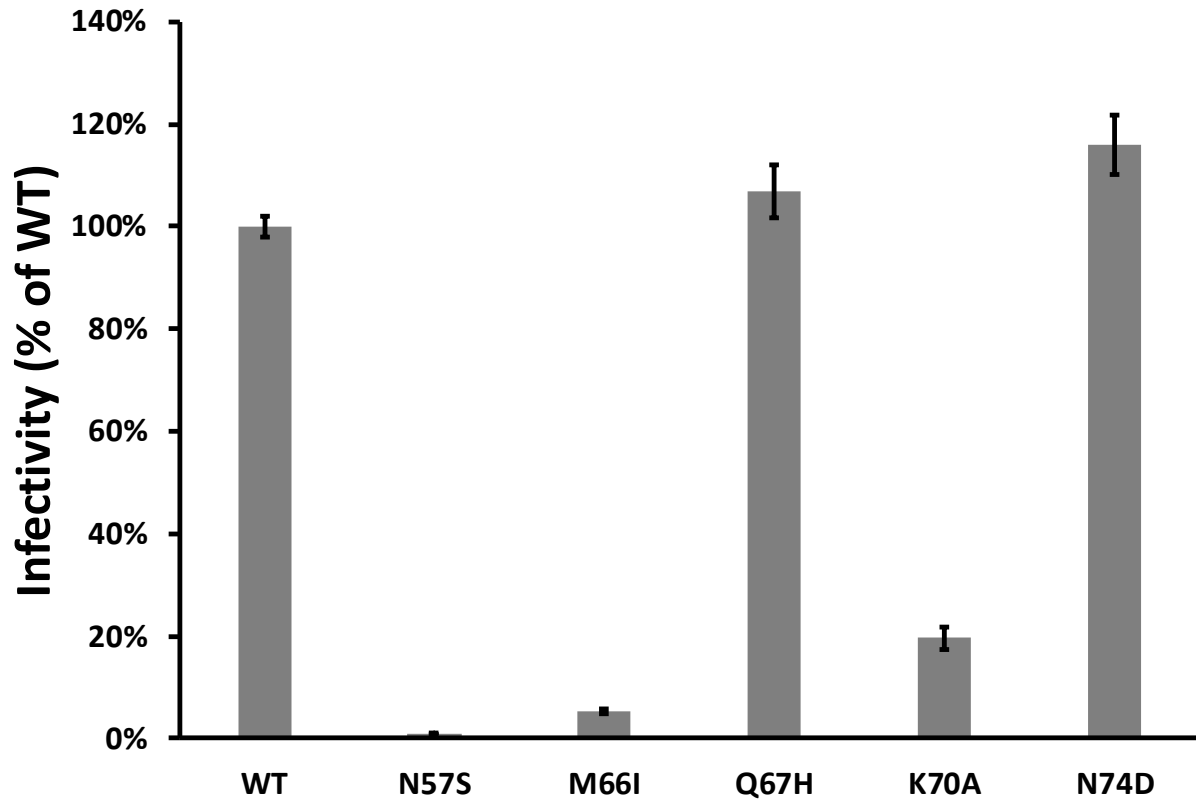


Fig. S21

Infectivity of the indicated drug-resistant mutant viruses. Results (average \pm SD of $n = 3$ independent experiments) are plotted versus the WT, which was set to 100%.

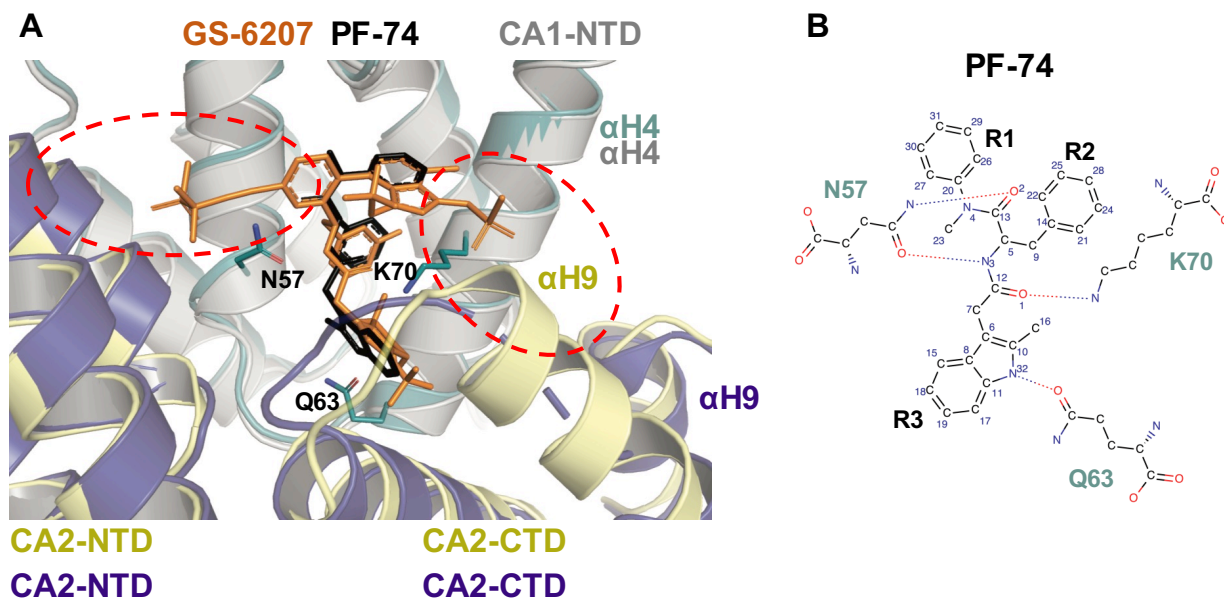


Fig. S22

Comparison of the binding sites of GS-6207 and PF-74 within the binding pocket of the CA_{A14C/E45C/W184A/M185A} hexamer. (A) Superimposition of the x-ray crystal structures of CA hexamers in complex with GS-6207 (CA1: light grey; CA2: pale yellow; GS-6207: orange; PDB ID: 6VKV) and PF-74 (CA1: dark teal; CA2: indigo; PF-74: black; PDB ID: 4U0E). Key CA residues interacting with PF-74 are indicated. Red dashed ovals indicate the bridging interactions of GS-6207 with CA2-NTD and CA2-CTD, which are lacking in the presence of PF-74. In particular, GS-6207 directly binds and stabilizes a distinct conformation of α H9 (pale yellow). In sharp contrast, PF-74 has no interaction with α H9 (indigo) and the majority of α H9 residues in the CA + PF-74 structure are not resolved well. (B) Chemical structure of PF-74 highlighting the key interactions with CA hexamer residues through colored dashed lines and denoting the major PF-74 rings through labels R1-R3.

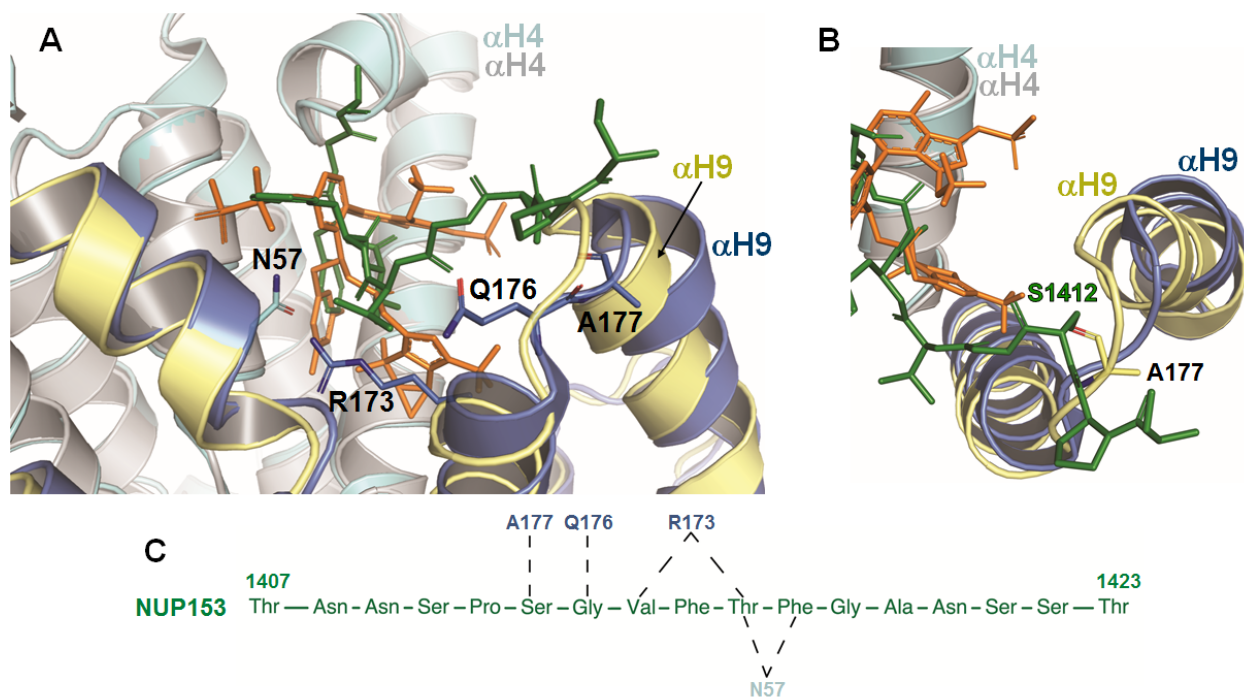


Fig. S23

Comparison of the binding sites of GS-6207 and the Nup153 peptide within the binding pocket of the CA hexamer. (A) Superimposition of the x-ray crystal structures of CA + GS-6207 (CA1: light grey; CA2: pale yellow; GS-6207: orange; PDB ID: 6VKV) and CA + Nup153 peptide (1407-1413) (CA1: aquamarine; CA2: navy; Nup153: green; PDB ID: 5TSV). Key CA residues interacting with Nup153 are indicated. The α H9 helices in GS-6207 + CA and Nup153₁₄₀₇₋₁₄₂₃ + CA (PDB ID: 5TSV) are indicated. The differential positioning of α H9 helices in the two structures is immediately evident, whereas all other helices superimpose very well. (B) There is a steric clash between Nup153 residue S1412 with the CA main chain of A177 in the context of a more closed conformation of α H9 adapted in the presence of GS-6207. This is indicated by comparing the positions of α H9 (pale yellow and navy) in the presence of GS-6207 and Nup153, respectively. (C) Diagram of key interactions between Nup153₁₄₀₇₋₁₄₂₃ and CA hexamer residues through black dashed lines.

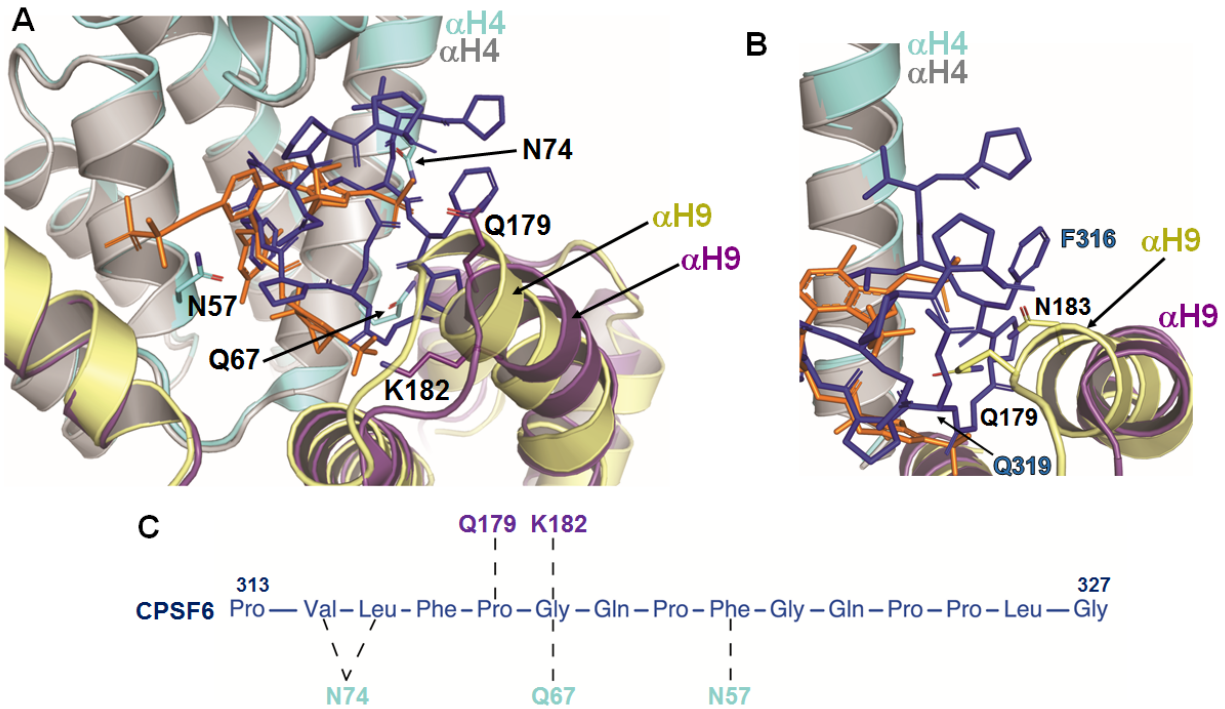


Fig. S24

Comparison of the binding sites of GS-6207 and CPSF6₃₁₃₋₃₂₇ within the binding pocket of the CA hexamer. (A) Superimposition of the x-ray crystal structures of CA + GS-6207 (CA1: light grey; CA2: pale yellow; GS-6207: orange; PDB ID: 6VKV) and CA + CPSF6₃₁₃₋₃₂₇ (CA1: cyan; CA2: purple; CPSF6₃₁₃₋₃₂₇: blue; PDB ID 4WYM). Key CA residues interacting with CPSF6₃₁₃₋₃₂₇ are indicated. The α H9 helices in GS-6207 + CA and CPSF6₃₁₃₋₃₂₇ + CA are indicated. The differential positioning of α H9 helices in the two structures is immediately evident, whereas all other helices superimpose closely. (B) There are steric clashes between CPSF6 residues F316 and Q319 with the CA sidechains of N183 and Q179 in the context of a more closed conformation of α H9 adapted in the presence of GS-6207. This become apparent by comparing the positions of α H9 (pale yellow and purple) in the presence of GS-6207 and CPSF6, respectively. (C) Diagram of key interactions between CPSF6₃₁₃₋₃₂₇ and CA hexamer residues through black dashed lines.

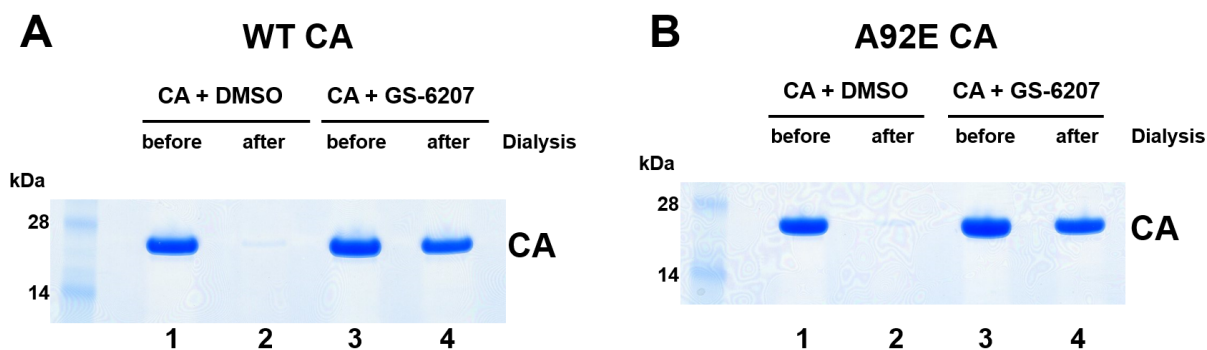
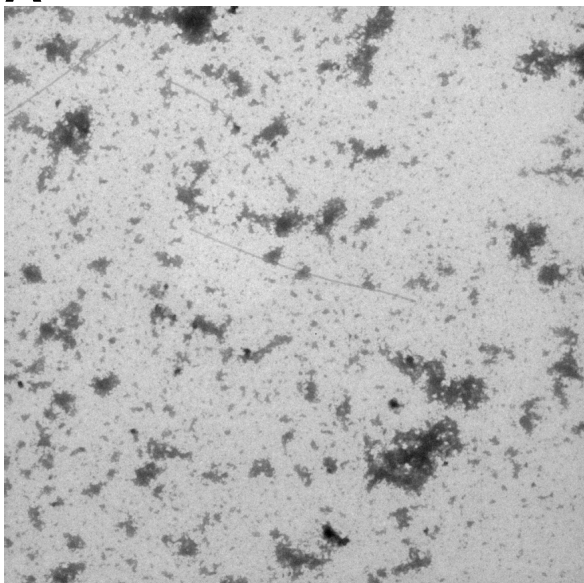


Fig. S25

Stabilization of preformed CA tubes by GS-6207. Representative SDS-PAGE images for WT (A) and A92E mutant (B) CA proteins show that GS-6207 but not DMSO stabilizes preformed CA tubes. The pelleted fractions of CA proteins recovered from the following reactions are shown: lanes 1, DMSO was added to WT (A) and A92E (B) CA tubes pre-formed in 2 and 1 M NaCl, respectively, and then reactions were pelleted; lanes 2, DMSO was added to WT (A) and A92E (B) CA tubes pre-formed in 2 and 1 M NaCl, respectively, and then reactions were dialyzed extensively against a 150 mM NaCl buffer; lanes 3, GS-6207 was added to WT (A) and A92E (B) CA tubes pre-formed in 2 and 1 M NaCl, respectively, and then reactions were pelleted; lanes 4, GS-6207 was added to WT (A) and A92E (B) CA tubes pre-formed in 2 and 1 M NaCl, respectively, and then reactions were dialyzed extensively against a 150 mM NaCl buffer. Remarkably, even though no GS-6207 was included in the dialysis buffer, CA + GS-6207 samples yielded substantial amounts of CA protein in pelleted fractions. These findings indicate a very tight binding of the inhibitor to preformed tubes and are consistent with a very slow dissociation rate of related GS-CA1 from stabilized CA hexamers (1) (also see Fig. S26).

A CA + DMSO in 150 mM NaCl



B CA + GS-6207 in 150 mM NaCl

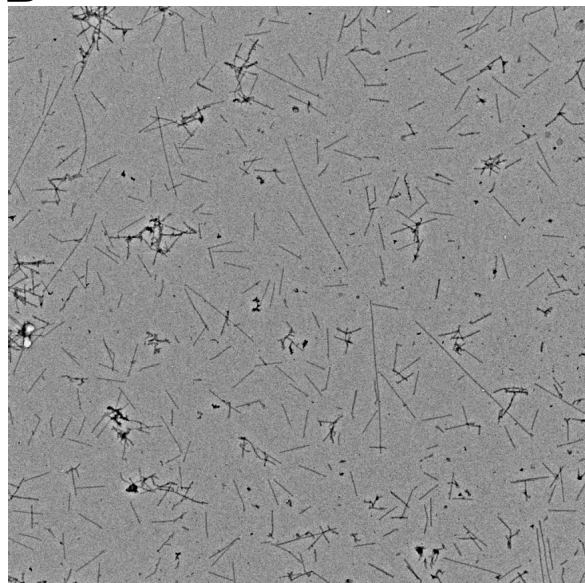


Fig. S26

EM images of stained CA tubes showing the effect of GS-6207 on tube stability in 150 mM NaCl buffer. (A) In control experiments, DMSO was added to A92E CA tubes pre-formed in 1 M NaCl, followed by dialysis against a 150 mM NaCl buffer. Very few tubes remained. (B) GS-6207 was added to A92E CA tubes pre-formed in 1 M NaCl, followed by dialysis against a 150 mM NaCl buffer. No GS-6207 was included in the dialyzes buffer (also see Fig. S25). A large number of stabilized tubes is observed. Quantitation of the total length of tubes + DMSO vs tubes + GS-6207 in low magnification images ($\sim 500 \mu\text{m}^2$) after dialysis yielded average values of $\sim 15 \mu\text{m}$ and $\sim 335 \mu\text{m}$, respectively. These results indicate that only $\sim 4\%$ of drug free tubes survive dialysis into a low salt buffer.

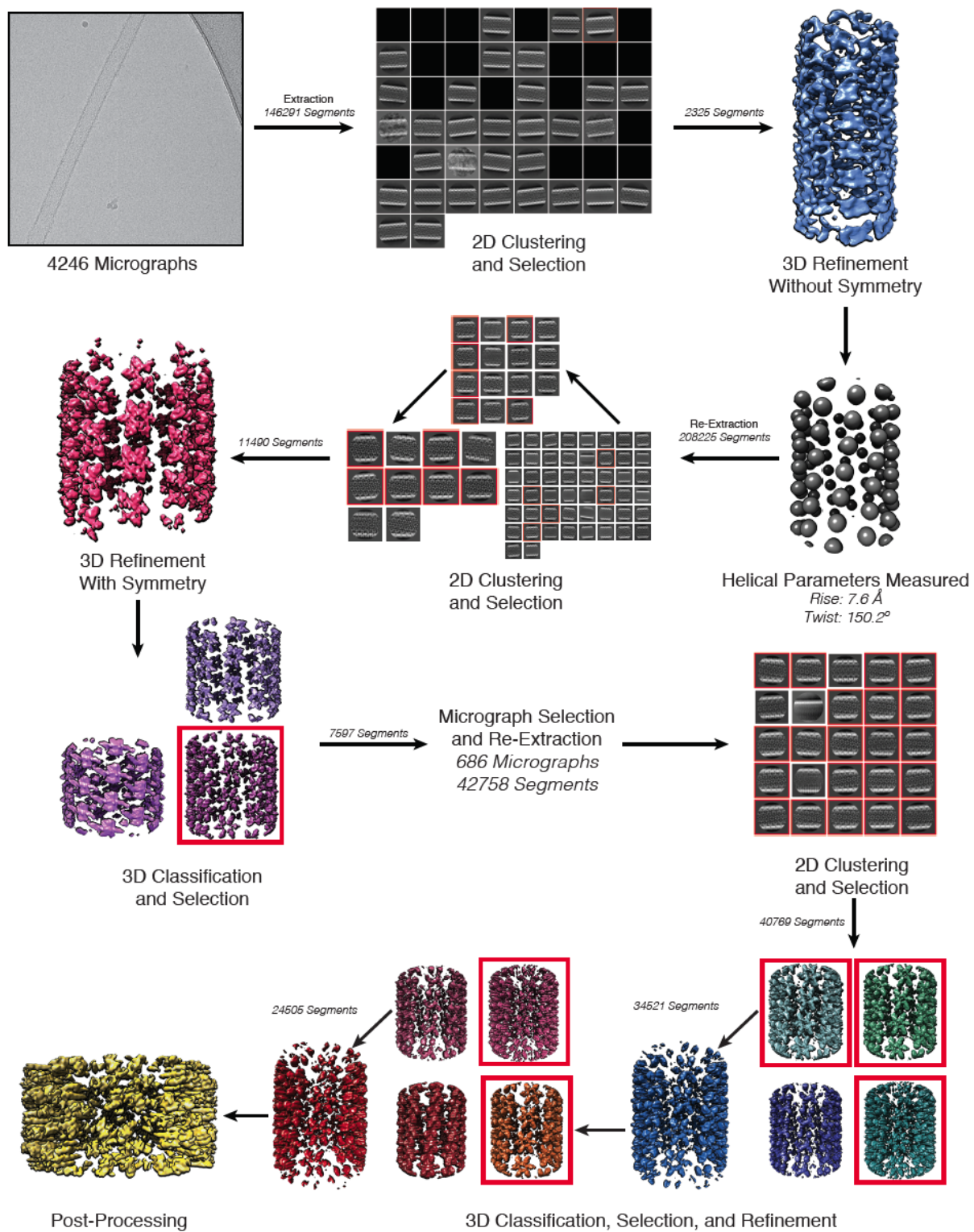


Fig. S27

Graphical summary of the sequence of steps involved in calculation of a cryo-EM map of GS-6207 stabilized CA helical tubes. After interactive tube selection from micrographs, tube segments were extracted and subject to 2D clustering. Images in the best 2D classes (judged by quality of layer lines) were then used for 3D reconstruction without applying helical symmetry. Initial helical parameters were estimated from initial non-symmetrized volumes and segments were re-extracted using as shift a multiple of the helical rise. This was followed by new 2D clustering, 3D helical reconstruction and 3D classification. Micrographs that contributed segments included in the best 3D volumes were re-extracted and the new set of segments was again run through 2D clustering, 3D reconstruction and 3D classification. After identifying a final set of homogeneous images, helical parameters were refined to obtain a final volume.

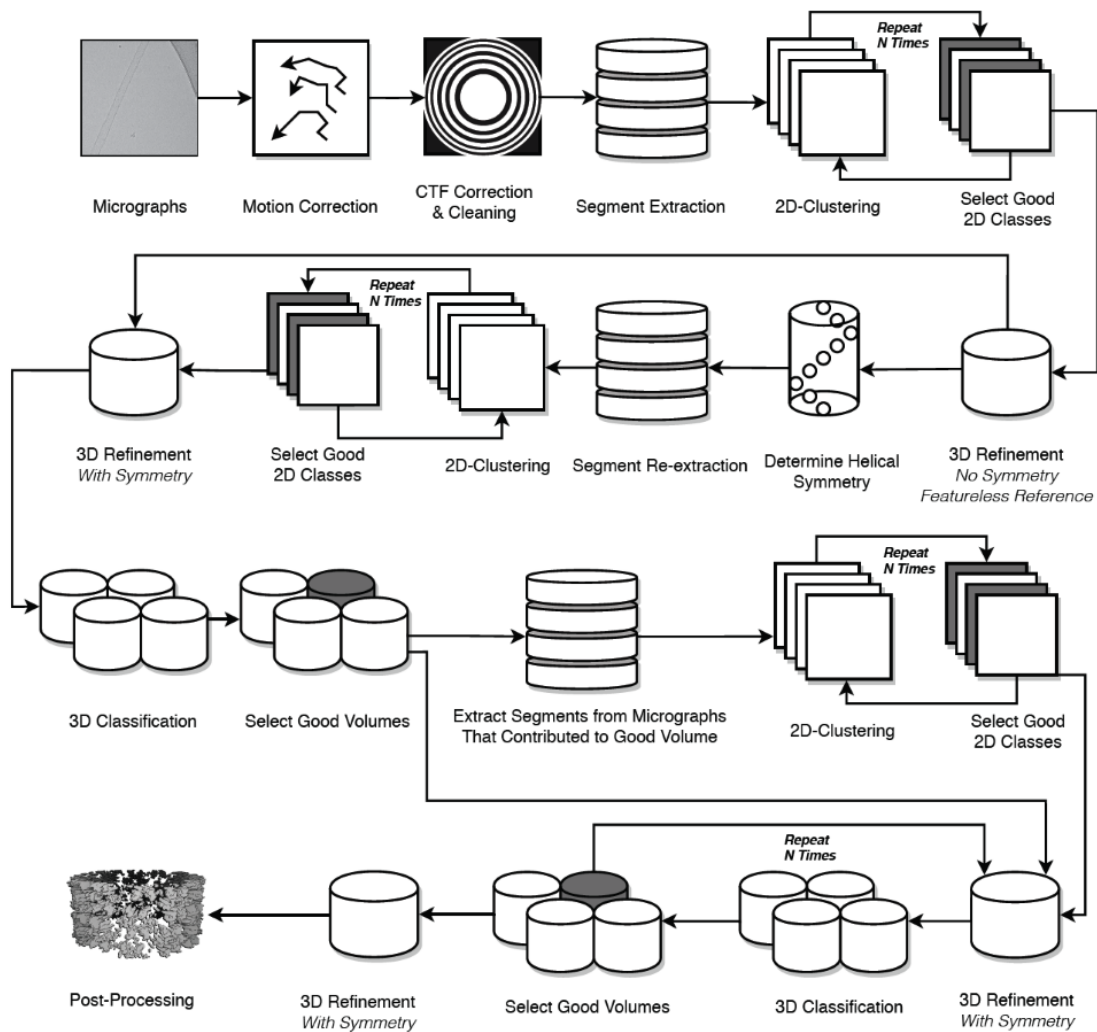


Fig. S28

Flowchart for helical image analysis of GS-6207 stabilized CA tube segments. The flowchart summarizes the steps involved in image screening, tube segment extraction, 2D clustering and 3D classification used to arrive at the final refined helical map.

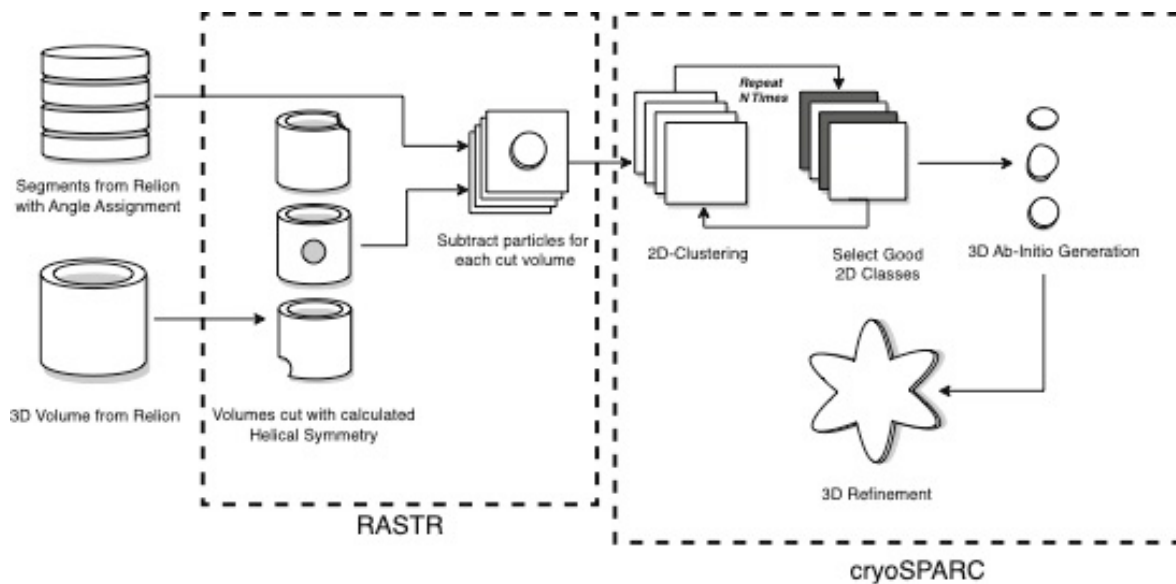


Fig. S29

Flowchart for RASTR image analysis of GS-6207 stabilized CA tube segments. The flowchart summarizes the steps involved in generation and analysis of GS-6207 stabilized tube patches from tube homogeneous tube segments, and the use of those tube patches to determine a map of a tube hexamer without the need to consider helical symmetry.

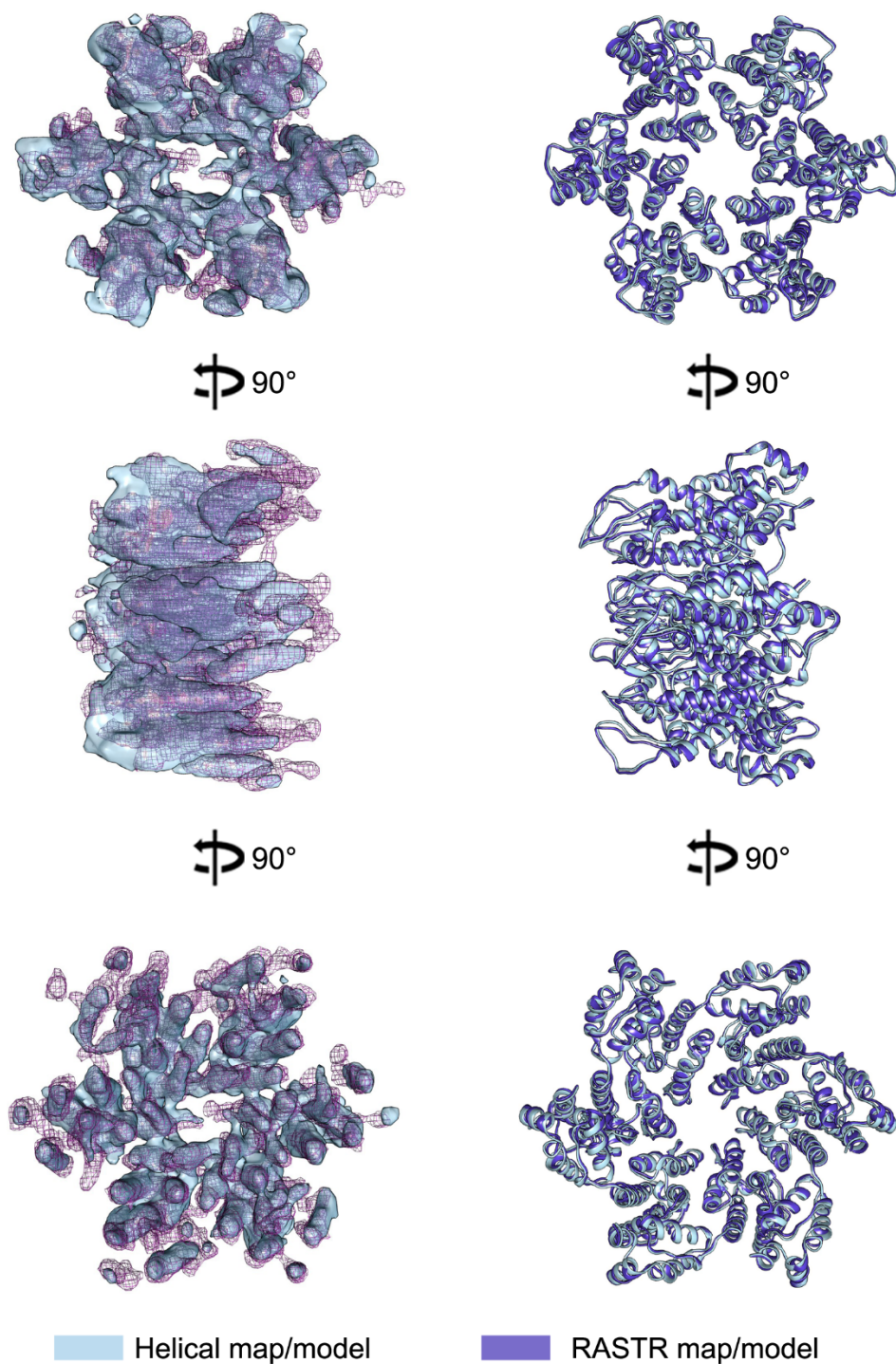
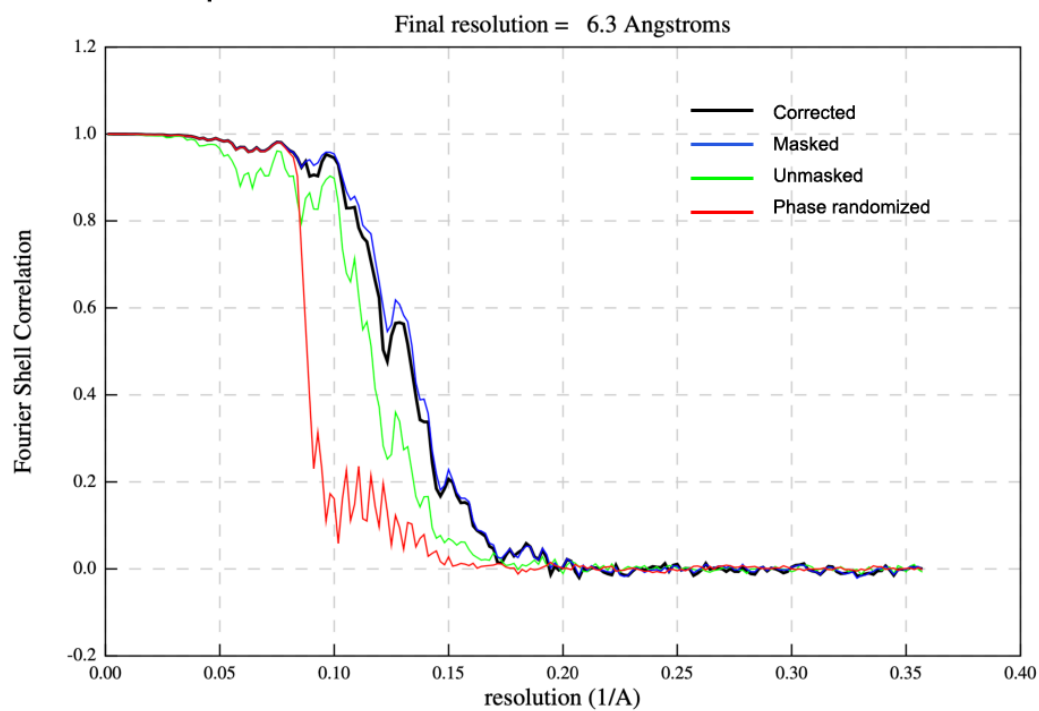


Fig. S30

Comparison between helical and RASTR cryo-EM maps and corresponding models. The position of corresponding alpha helices in cryo-EM maps of a hexamer in a GS-6207 stabilized

CA tube calculated using helical processing or RASTR (which does not depend on helical parameter determination or symmetry) are essentially superimposable. As expected, the same is true about models derived from each map. Density corresponding to poorly-ordered loops in the N-terminal portion is easier to see in the helical map, but the position of all helices is the same in both maps.

Helical map FSC



RASTR map FSC

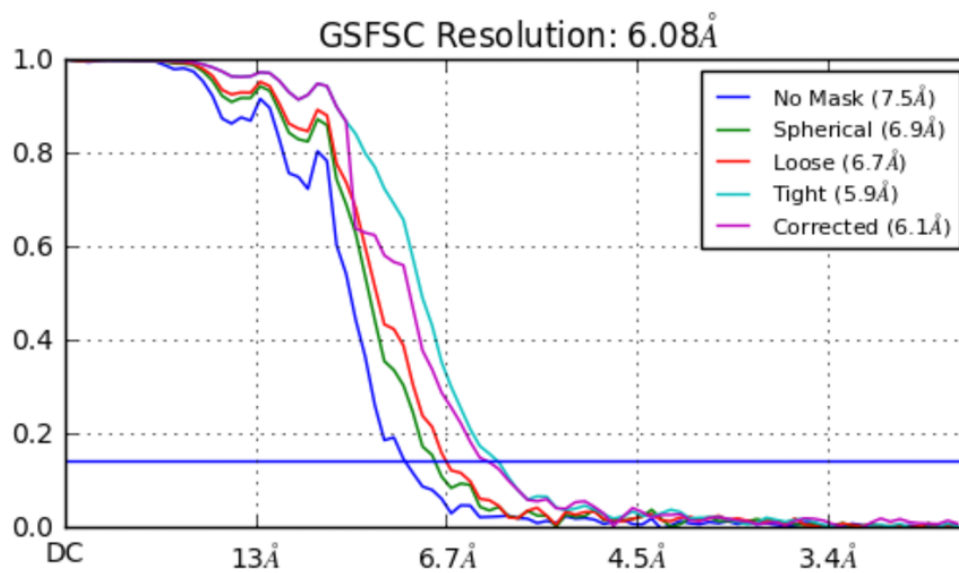
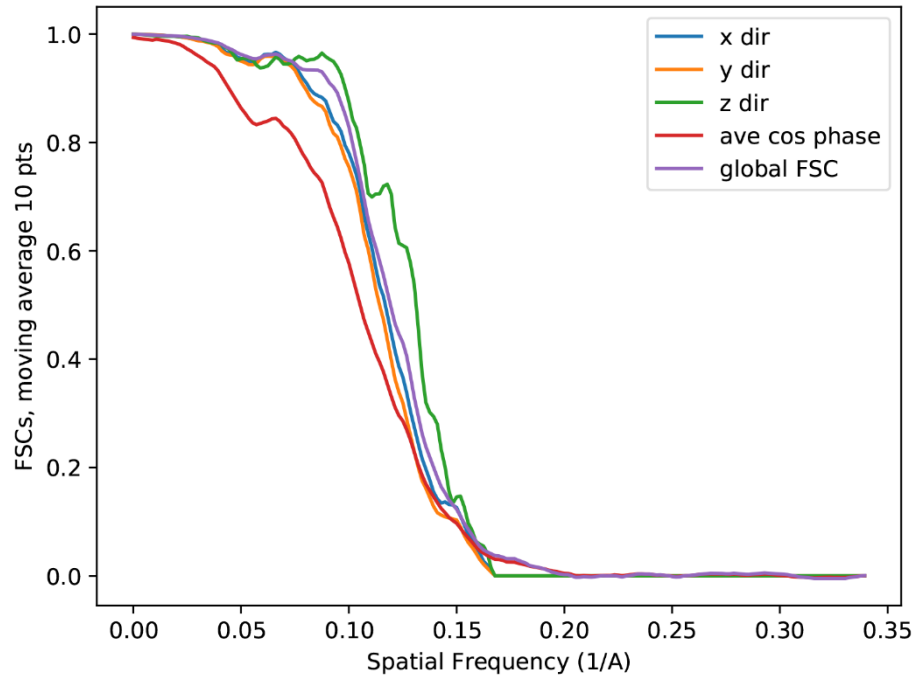


Fig. S31

Gold standard FSC plots for helical and RASTR maps.

3D FSC of Helical map (EMD-21424)



3D FSC of RASTR map (EMD-21423)

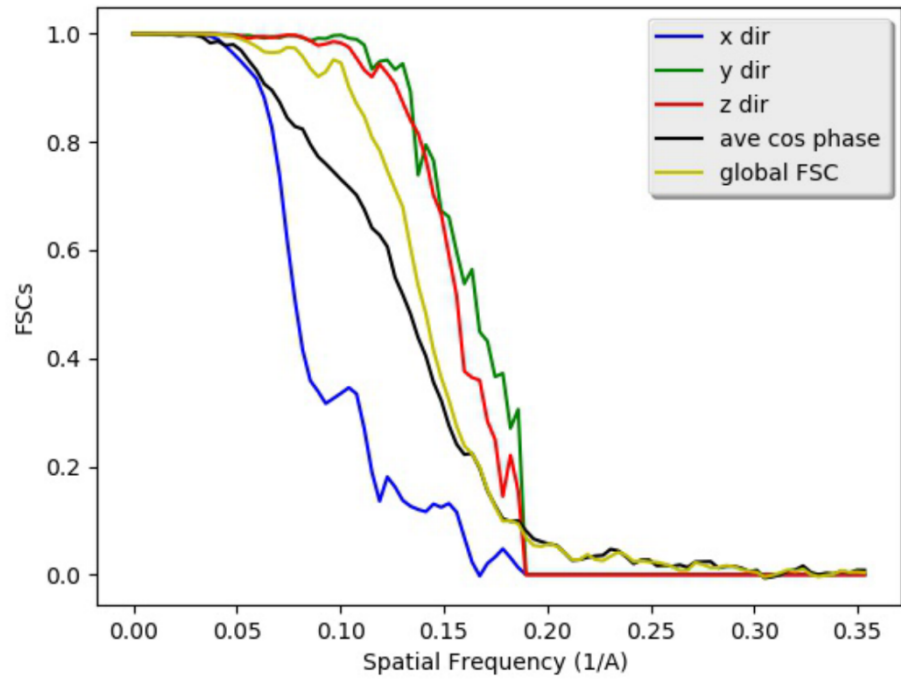


Fig. S32

3D FSC plots for helical and RASTR maps.

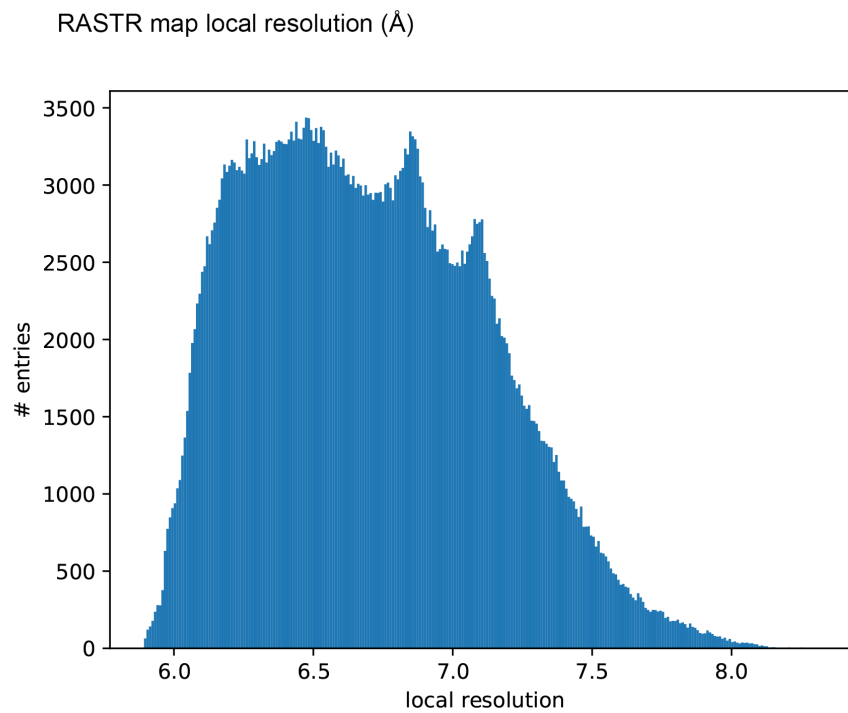
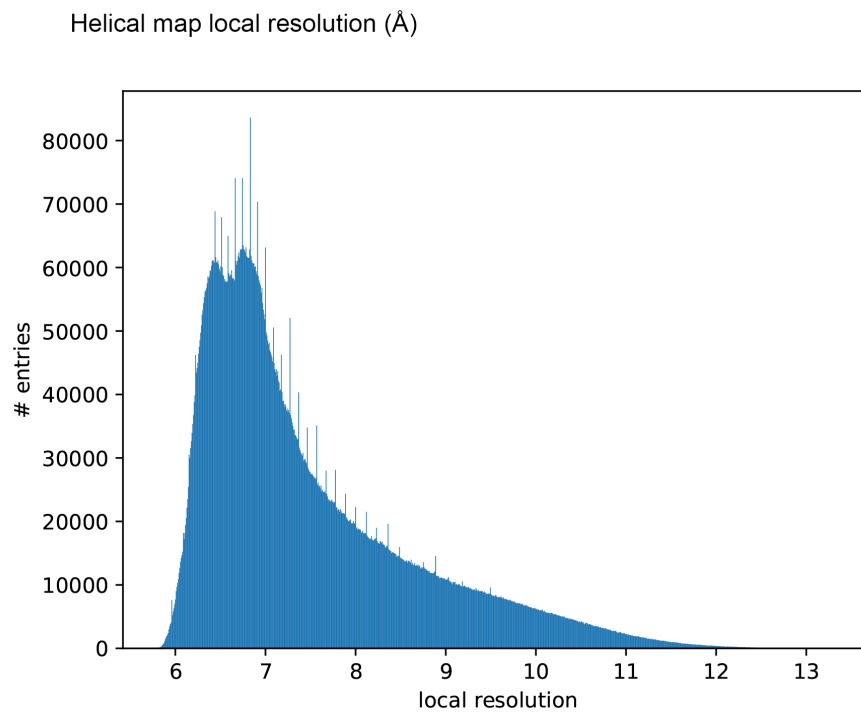


Fig. S33

Local resolution histograms for helical and RASTR maps.

RASTR map to model (6VWS) FSC

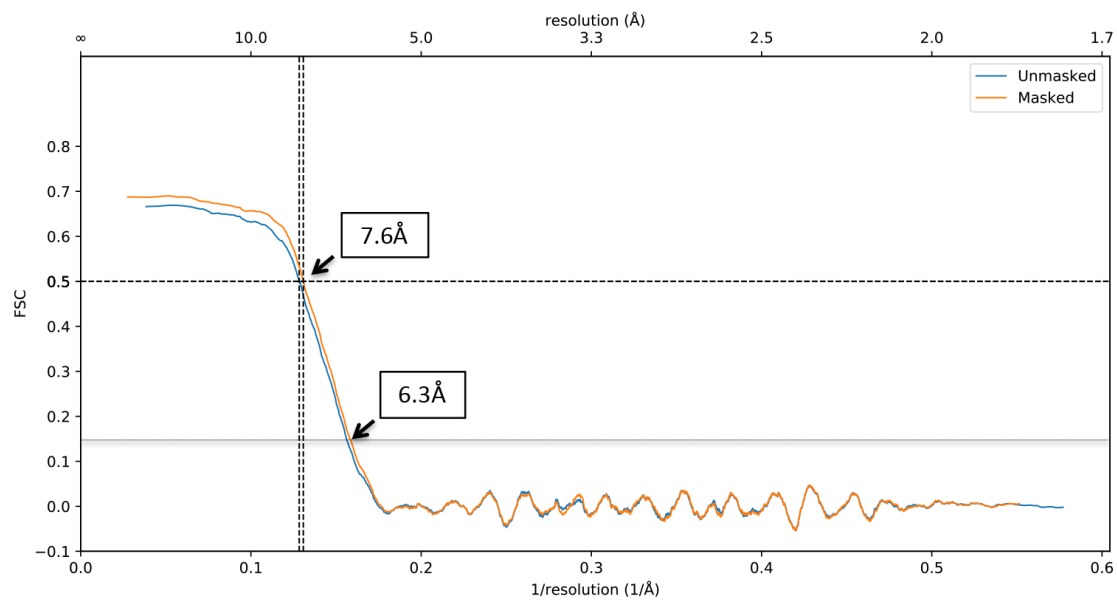


Fig. S34

Map-to-model FSC between RASTR map and model 6VWS.

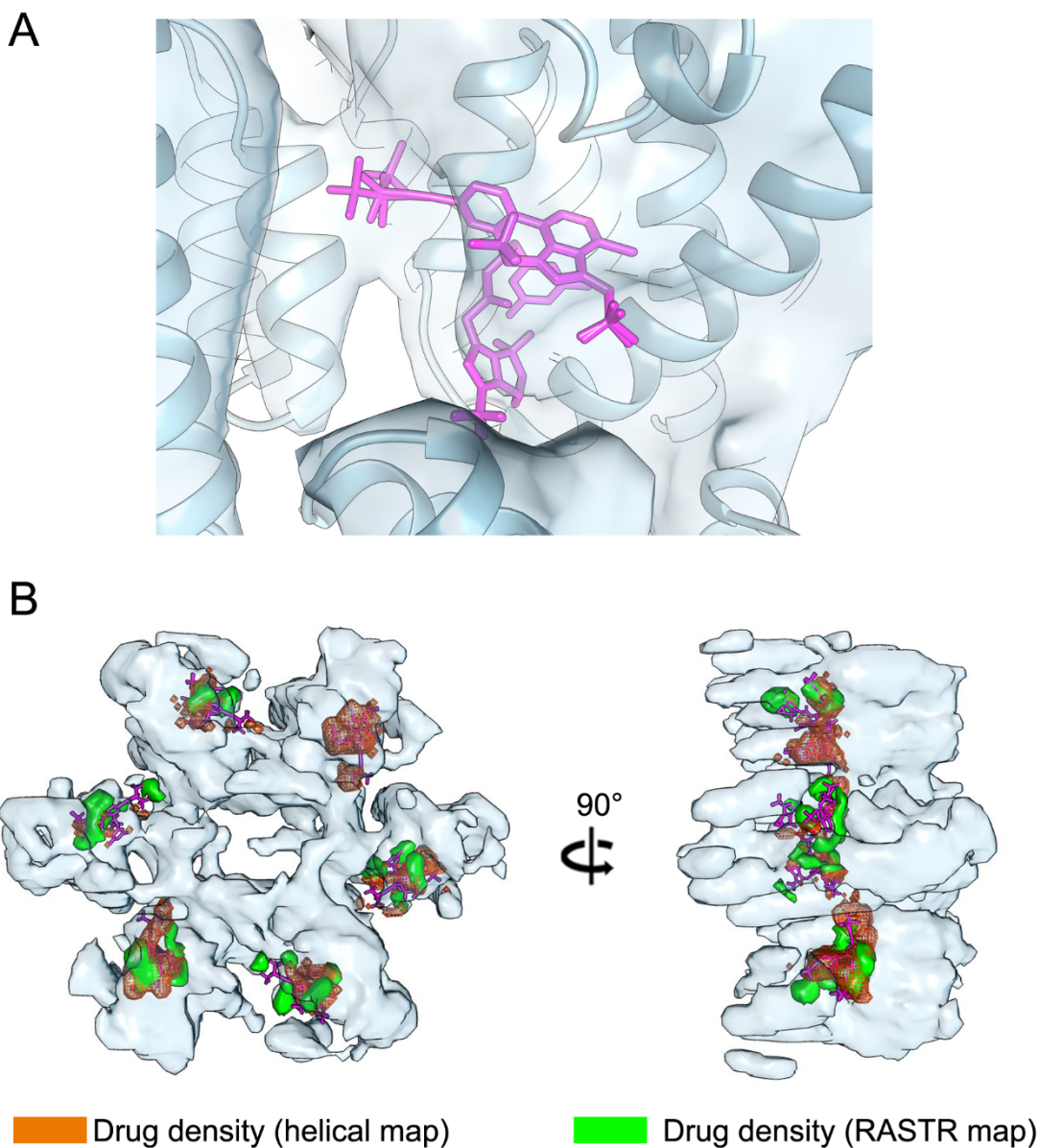


Fig. S35.

GS-6207 density in helical and RASTR cryo-EM maps. **A.** A close up view of the helical cryo-EM map showing density corresponding to a bound GS-6207. **B.** Top and side views of the cryo-EM hexamer map with GS-6207 density segmented to highlight its distribution (segmented GS-6207 density in orange and green for the helical and RASTR maps, respectively). Calculation of a straight difference map was not possible because we could not obtain a cryo-EM map of an inhibitor-free tube with the same helical parameters as only ~4% of drug free tubes survived in

low salt (Fig. S26A). Calculation of a difference map between the cryo-EM map and a molecular map based on the CA portion of the atomic model was complicated by the fact that various portions of the hexamer structure show comparatively low density due to mobility. Therefore, the best way to visualize inhibitor density was to segment the cryo-EM map based on the atomic model to highlight density in the map that overlaps with the inhibitor positions. The appearance of drug density is not exactly the same for every monomer, but it must be noted that the hexamer derived from the helical map was not symmetrized in any way. A limited resolution (6.1-6.3Å) of our cryo-EM maps did now allow us to examine drug interactions with curved CA hexamers in more detail, but the resolution was sufficient to reveal the substantial effects of drug binding on the hexamer structure.

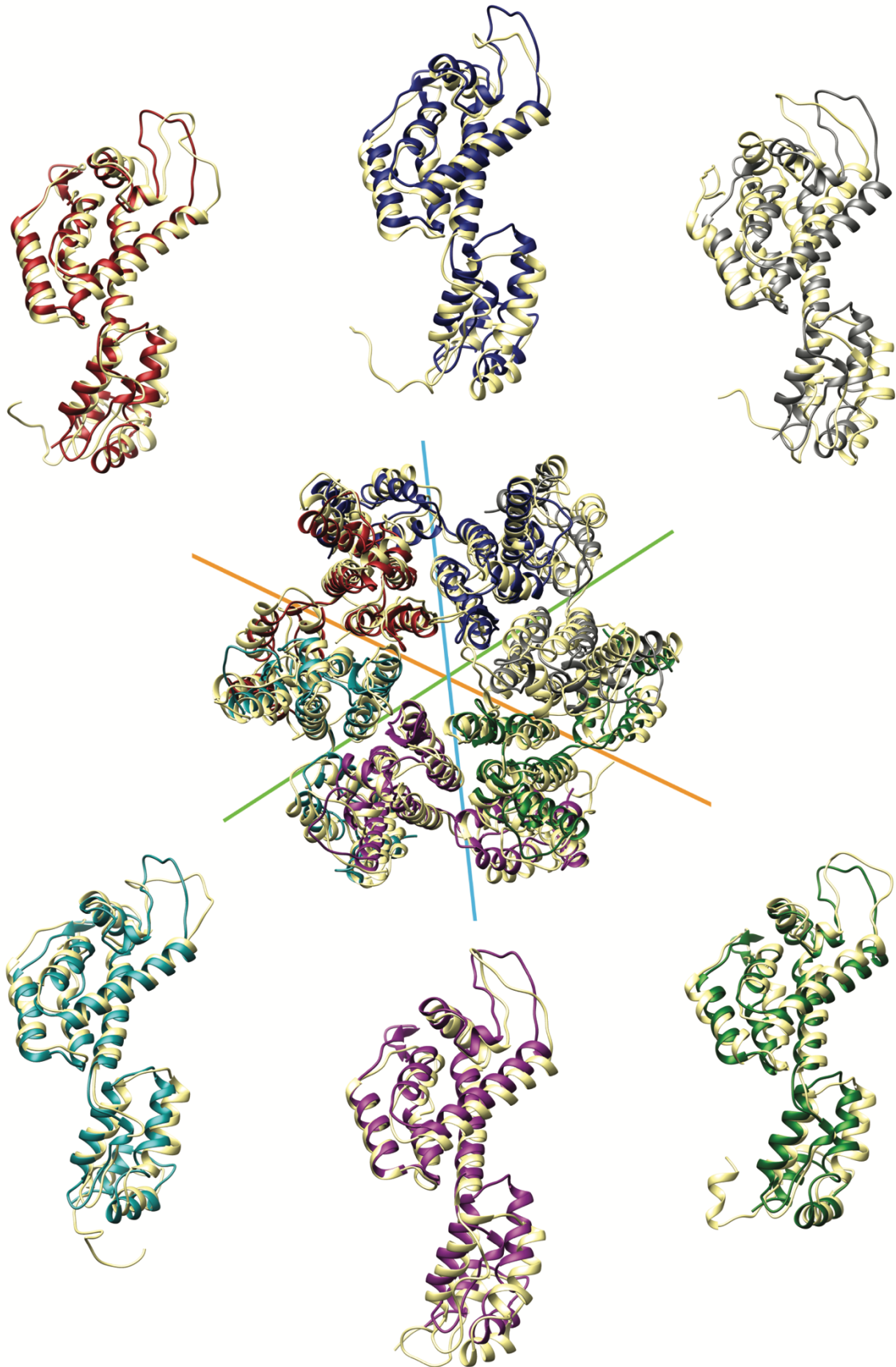


Fig. S36

Comparison of cryo-EM structures of CA hexamers derived from CA_{A92E} tubes in the absence and presence of GS-6207. An overlay of CA hexamers derived from cryo-EM structures of CA_{A92E} tubes of in the absence (light yellow; PDB ID: 3J34) and presence of GS-6207 (each monomer colored differently; PDB ID: 6VWS). The cyan, green and orange lines indicate the three helical directions of CA tube + GS-6207. Surrounding the overlaid CA hexamers are close-up views of the individual monomers (from the overlaid cryo-EM structures), which highlight the reasonably similar positioning of the NTDs and large differences in the positioning of the CTDs between the cryo-EM structures.

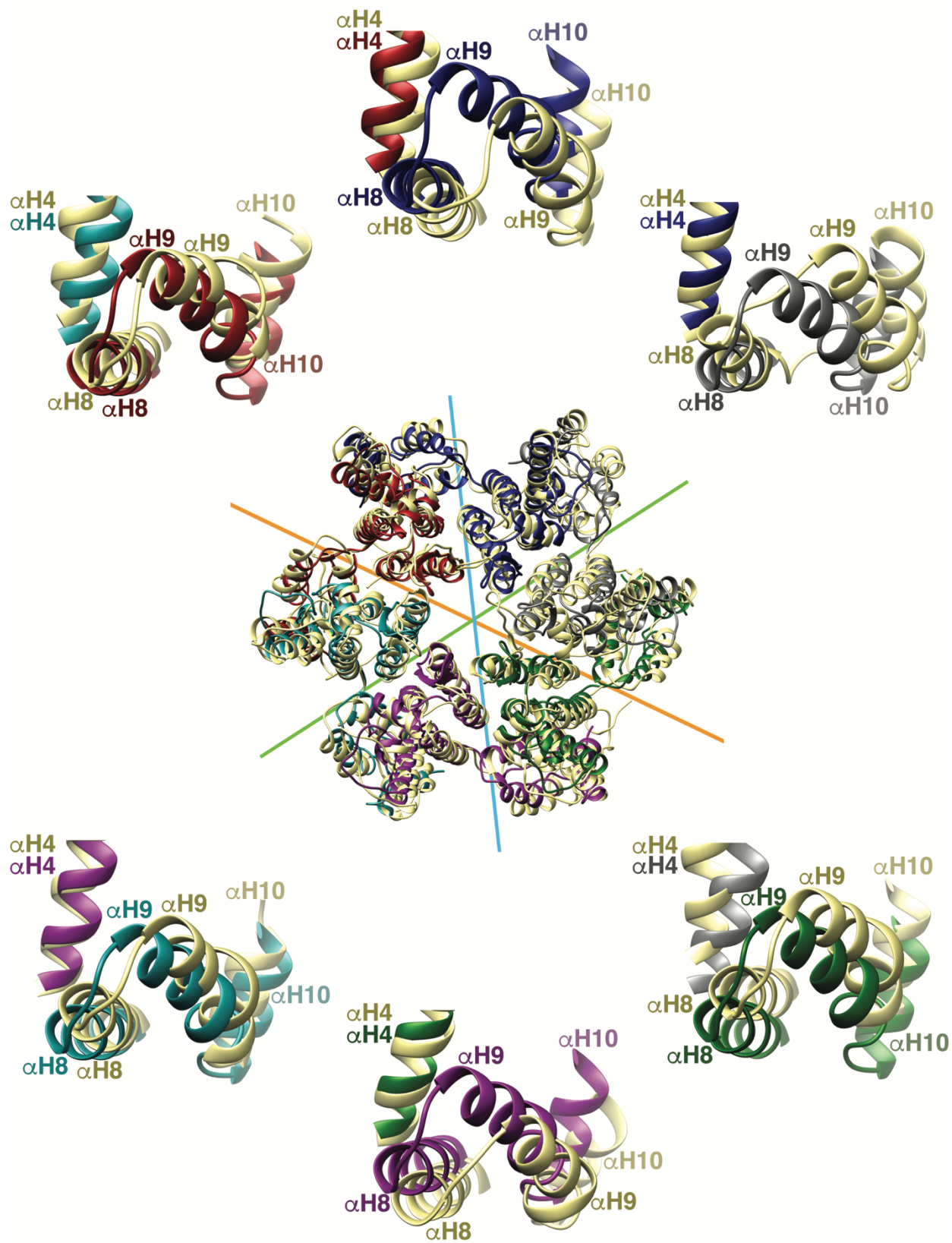


Fig. S37

Comparison of NTD α H4 and CTD α H8, α H9, and α H10 of the cryo-EM structures of CA hexamers derived from CA_{A92E} tubes in the absence and presence of GS-6207. Six close-up views of overlaid NTD α H4 and CTD α H8, α H9, and α H10 of CA hexamers derived from cryo-EM structures of CA_{A92E} tubes of in the absence (light yellow; PDB ID: 3J34) and presence of GS-6207 (each monomer colored differently; PDB ID: 6VWS) surrounding the overlaid CA hexamers. The cyan, green and orange lines indicate the three helical directions of CA tube + GS-6207. The overlaid CA hexamer structures demonstrate reasonably similar α H4 positions in the NTDs and large differences in the CTDs (represented by α H8, α H9, and α H10). Note, NTD α H4 and CTD α H9 from adjacent subunits are positioned closer together in the presence of the inhibitor compared with the inhibitor free structure.

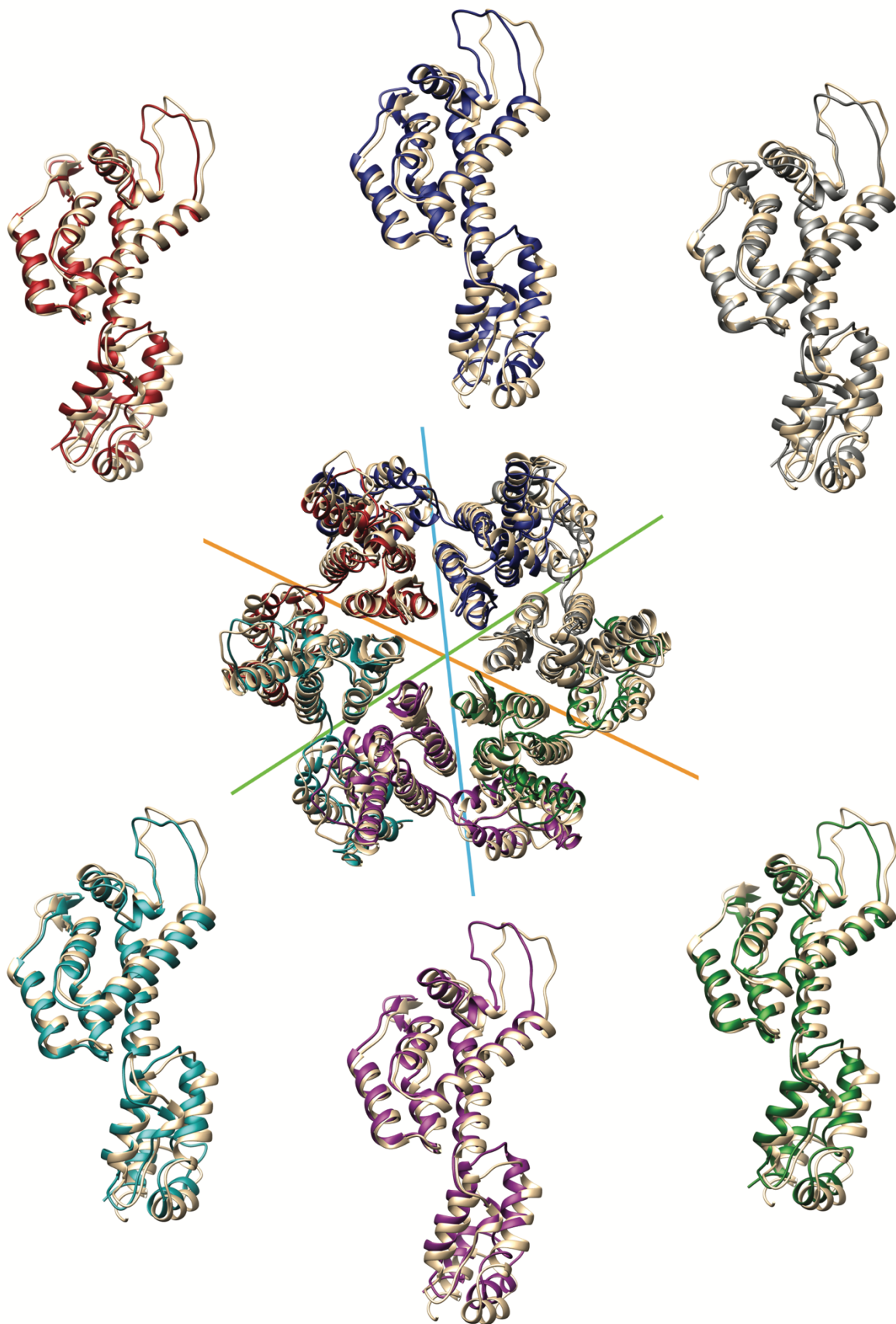


Fig. S38

Comparison of our cryo-EM structure of GS-6207 bound curved hexamer with the cryo-ET structure of a native hexamer from HIV-1 particles. The cryo-EM structure of GS-6207 bound hexamer (each monomer colored differently; PDB ID: 6VWS) was overlaid with cryo-ET structure of a native hexamer from HIV-1 particles (tan; PDB ID: 5MCX). The cyan, green and orange lines indicate the three helical directions of CA tube + GS-6207. Surrounding the overlaid CA hexamers are close-up views of the individual monomers (from the overlaid CA hexamers), which highlight the reasonably similar positioning of the NTDs and the large differences in the positioning of the CTDs between the two CA hexamer structures.

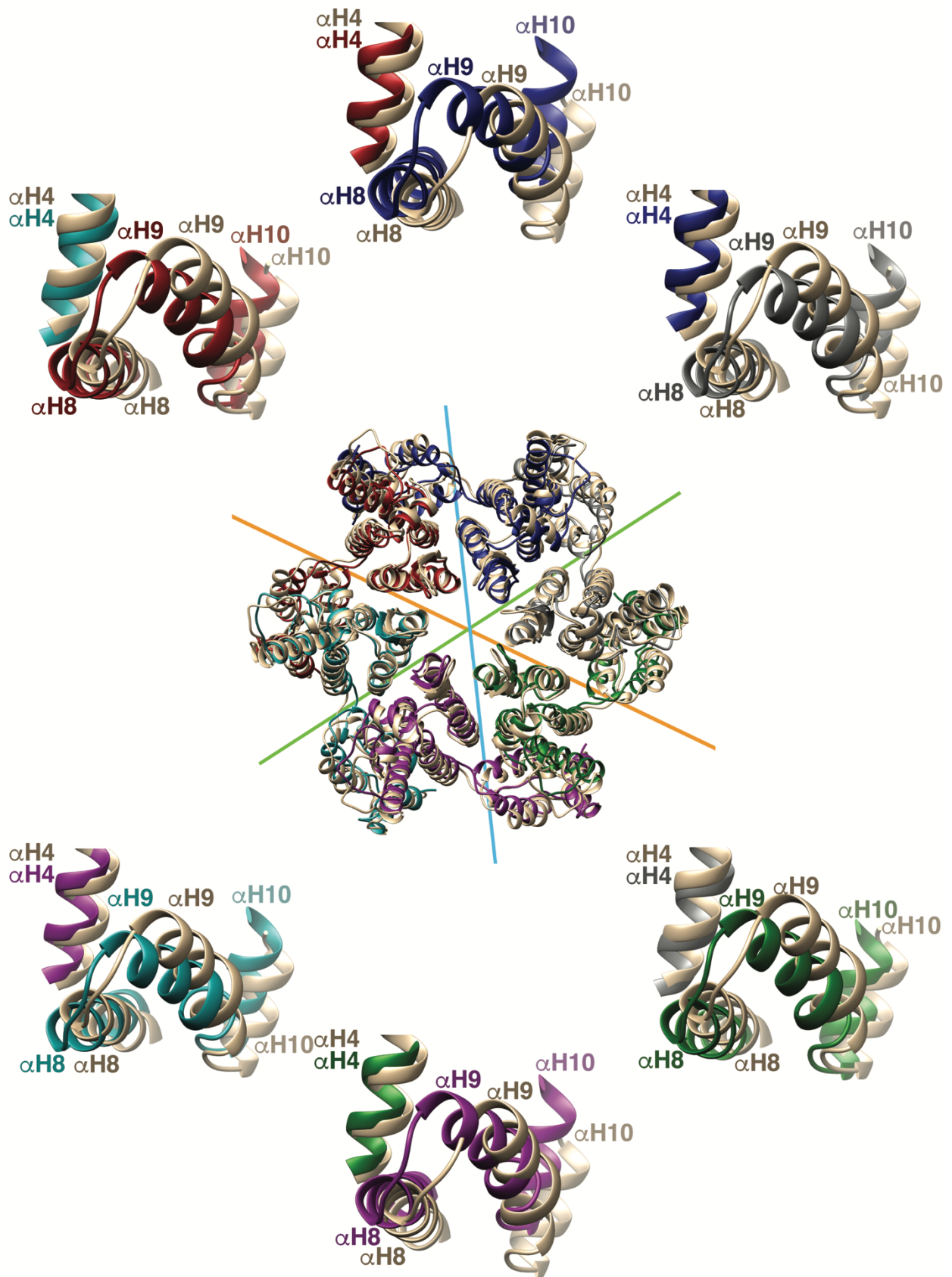


Fig. S39

Comparison of NTD α H4 and CTD α H8, α H9, and α H10 of our cryo-EM structure of GS-6207 bound curved hexamer with the cryo-ET structure of a native hexamer from HIV-1 particles. Six close-up views of overlaid NTD α H4 and CTD α H8, α H9, and α H10 of our cryo-EM structure of GS-6207 bound hexamer (each monomer colored differently; PDB ID: 6VWS) with cryo-ET structure of a native hexamer from HIV-1 particles (tan; PDB ID: 5MCX) surrounding the overlaid CA hexamers. The cyan, green and orange lines indicate the three helical directions of CA tube + GS-6207. Note, NTD α H4 and CTD α H9 from adjacent subunits are positioned closer together in the presence of the inhibitor compared with the structure of inhibitor free native hexamer from HIV-1 particles.

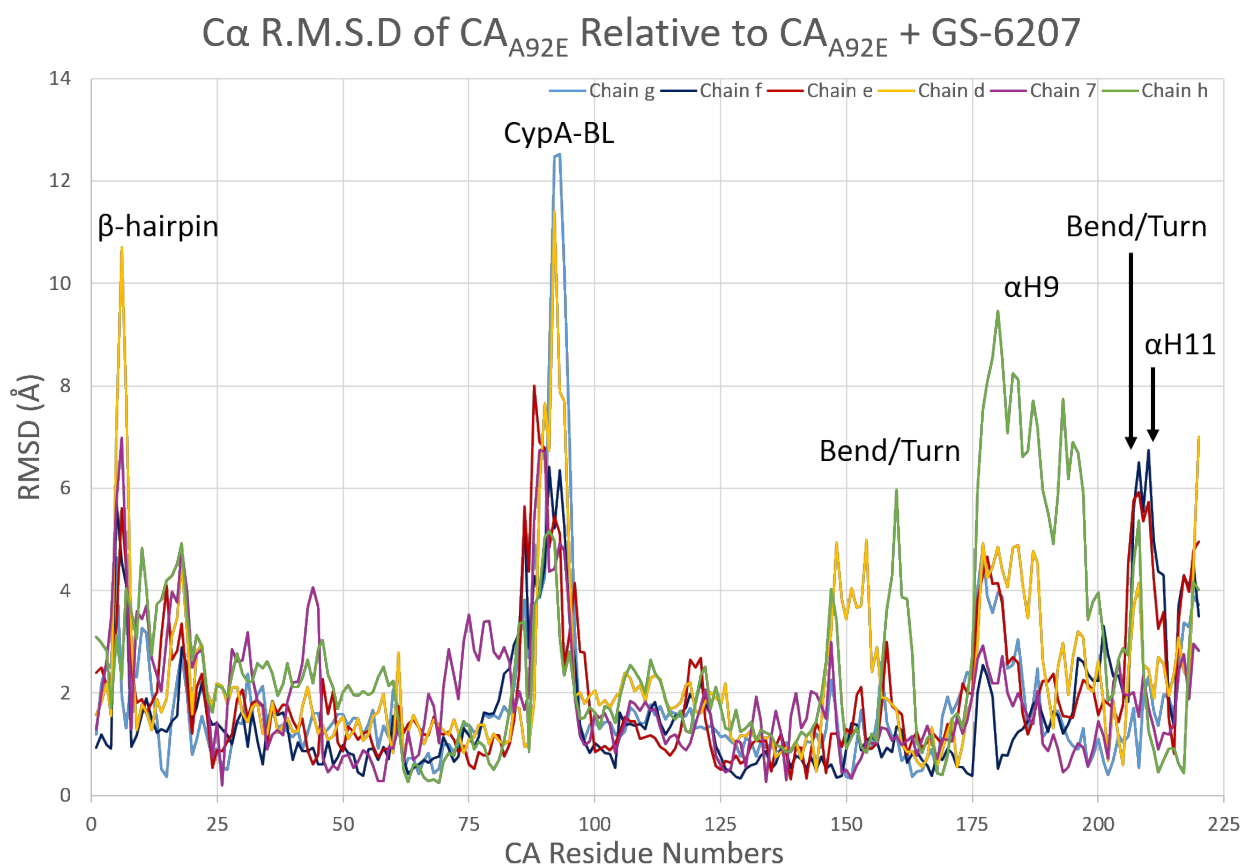


Fig. S40

RMSD-based structural comparison of CA monomers within CA_{A92E} tubes +/- GS-6207. RMSDs were measured using Superpose (57) in CCP4. Line graphs of the RMSD of the alpha carbons of the 6 monomers in CA_{A92E} tubes (PDB ID: 3J34) measured against a monomer from CA_{A92E} tubes + GS-6207 are shown (PDB ID: 6VWS). Each colored line represents a different monomer in CA_{A92E} tubes. One of the most pronounced differences is observed at α H9 (residues 179-192, in particular for chain h). There are other differences at the N-terminal flexible β -hairpin (particularly for chain d), CypA-BL (notably with chains g and d), α H11 (211-217), and two flexible regions (157-158 and 206-207; chains d and h most notable) that can be a turn or bend depending on the structure. In addition, a few residues at the end of the CTD appear to show increased variability between the structures. The terminal residues of a structure can show increased flexibility due to unconstrained movement. Consequently, it is possible this difference is incidental.

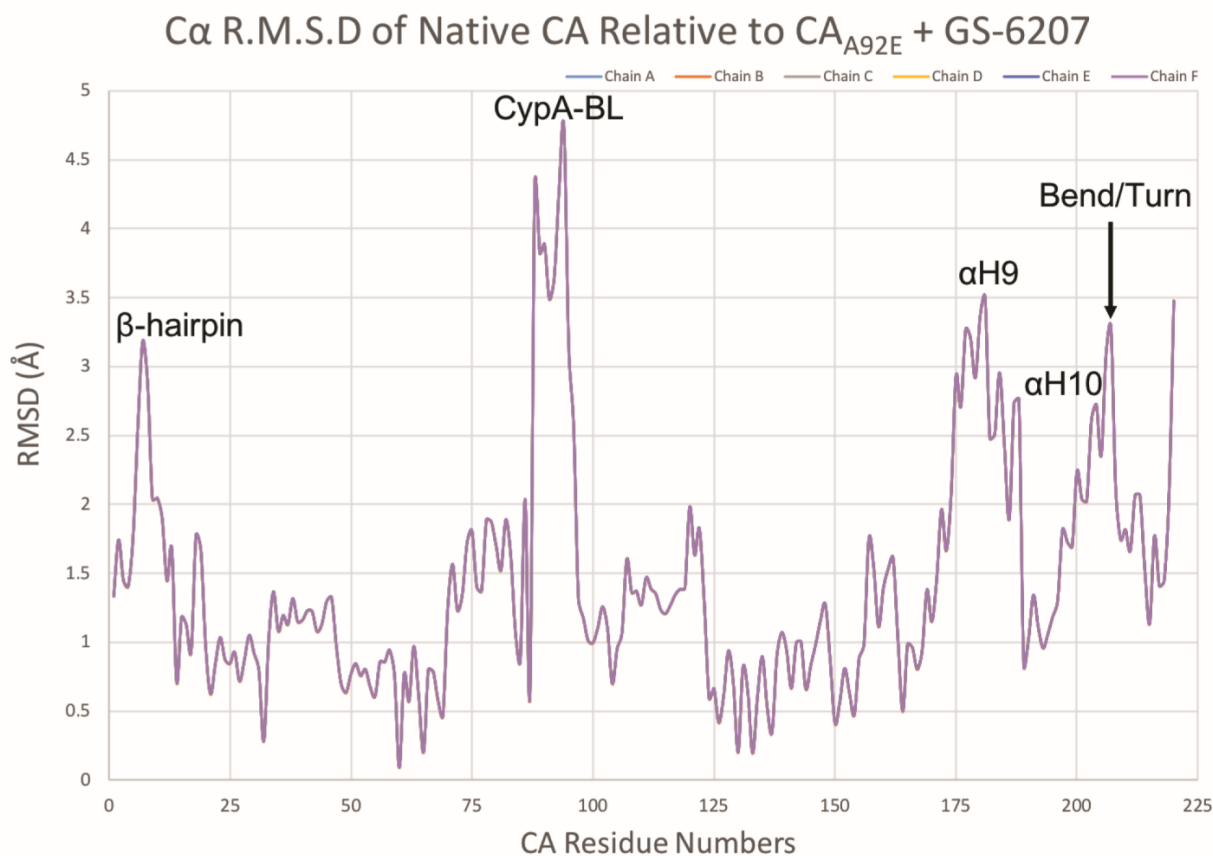


Fig. S41

RMSD-based structural comparison of CA monomers from CA_{A92E} tubes + GS-6207 derived from cryo-EM and native CA hexamers derived from cryo-ET analysis of HIV-1 particles. RMSDs were measured using Superpose (57) in CCP4. Line graphs of the RMSD of the alpha carbons of 6 monomers from a native CA hexamer in intact virus particles (PDB ID: 5MCX) when measured against a monomer from CA_{A92E} tubes + GS-6207 (PDB ID: 6VWS). Each colored line represents a different monomer in native CA hexamer. All six of the colored lines overlapped exactly, therefore only one colored line is discernable. One of the most pronounced differences in structure is observed at α H9 (179-192). There are other differences at the CypA-BL, α H10 (196-205), a flexible region (206-207) that can be a turn or bend depending on the structure, and the flexible β -hairpin at the beginning of the NTD. In addition, a few residues at the end of the CTD appear to show increased variability between the structures. The terminal residues of a structure can show increased flexibility due to unconstrained movement. Consequently, it is possible this difference is incidental.

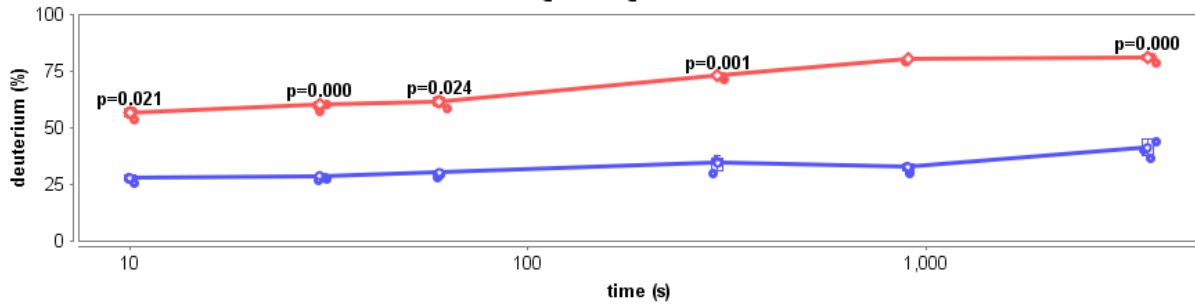
A

WT CA \pm GS-6207 at 1.5 M NaCl

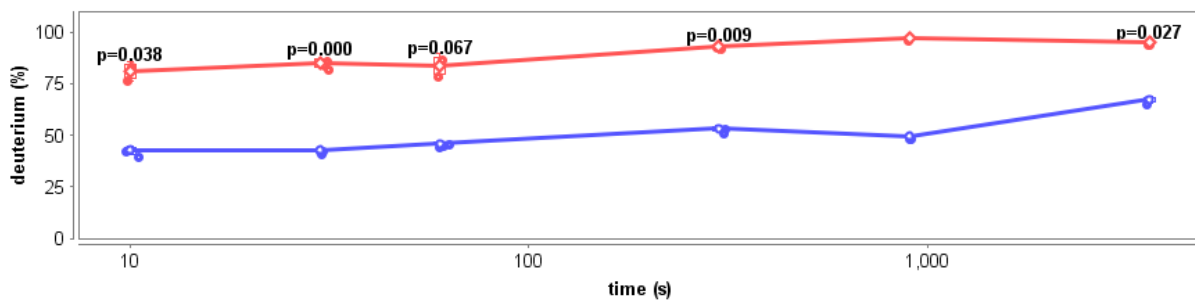
● CA + DMSO

● CA + GS-6207

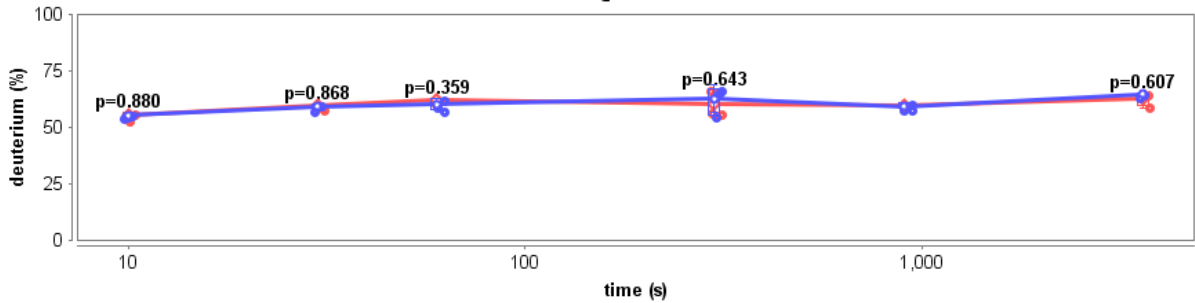
1 NTVGGHQAAMQML+2 57-69



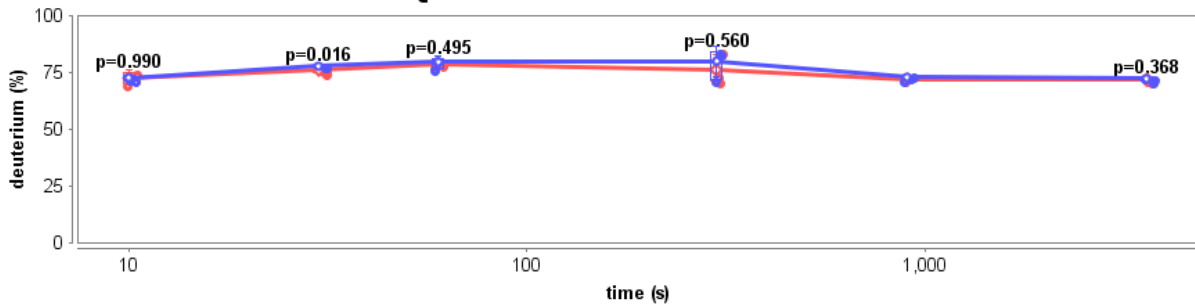
1 YKTLRAEQASQEVKNWMTE+4 169-187



1 WDRLHPVHAGPIAPGQMREPRGSD+3 80-103



1 CQGVGGPGHKARVL+3 218-231



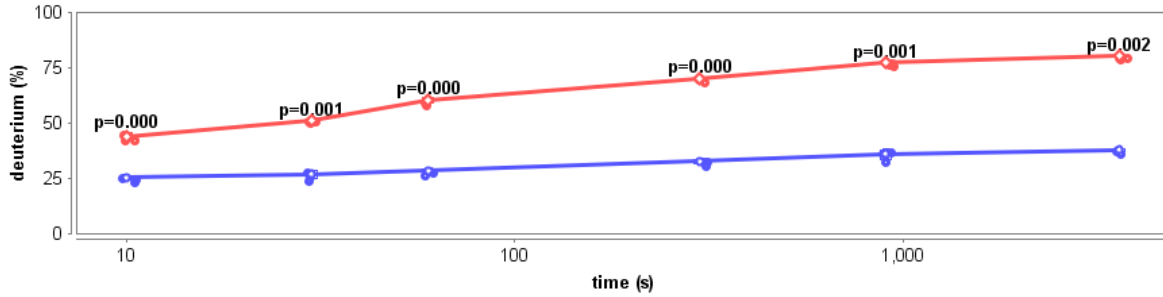
B

WT CA \pm GS-6207 at 100 mM NaCl

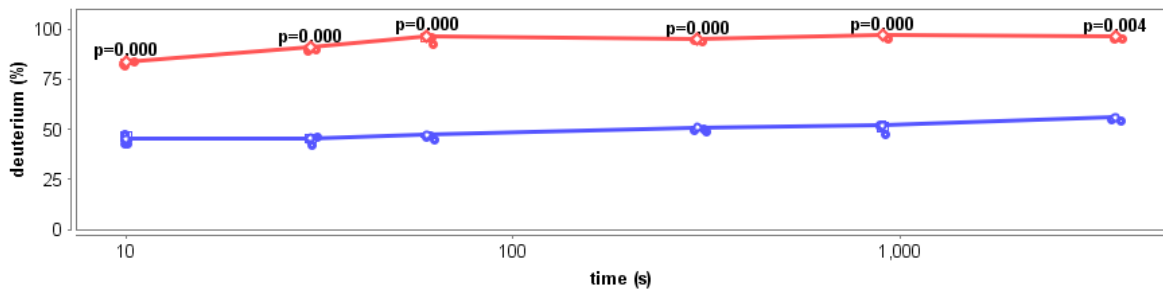
—●— CA + DMSO

—●— CA + GS-6207

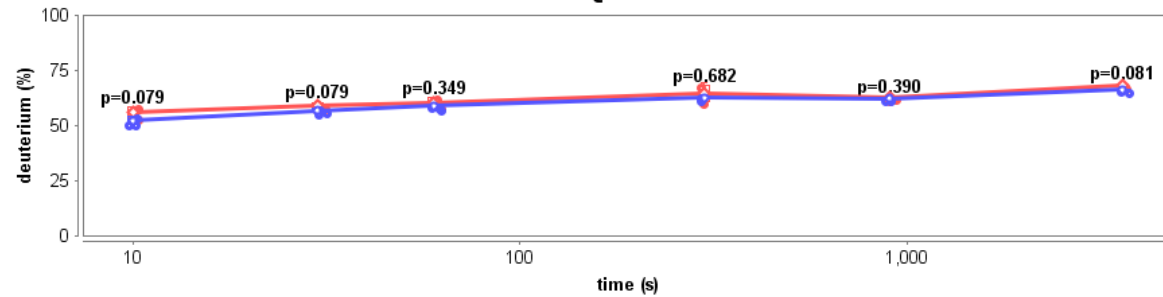
1 NTVGGHQAAMQML+2 57-69



1 YKTLRAEQASQEVKNWMT+4 169-187



1 WDRLHPVHAGPIAPGQMREPRGSD+3 80-103



1 CQGVGGPGHKARVL+3 218-231

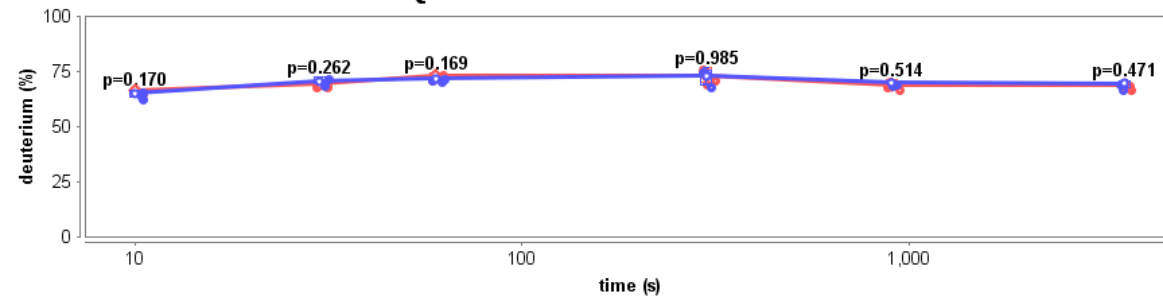


Fig. S42

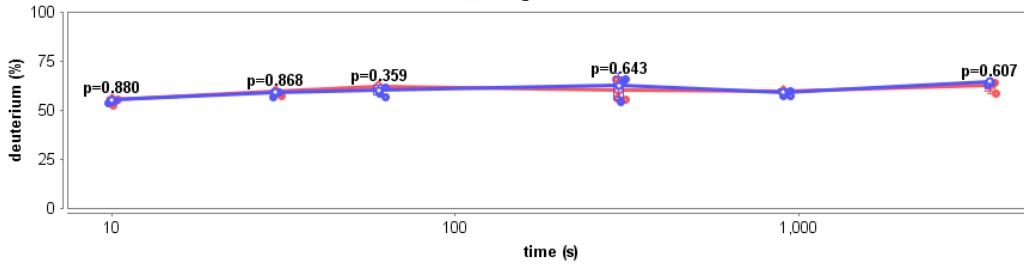
Representative deuterium build-up plots. Differential HDX-MS results with WT CA tubes bound to GS-6207 in (A) high (1.5 M NaCl) and (B) low (100 mM NaCl) salt conditions. The HDX results for select peptides that demonstrate strong (57-69, 169-187) vs no noticeable (80-103, 218-231) protections in the presence vs absence of GS-6207 are shown.

A

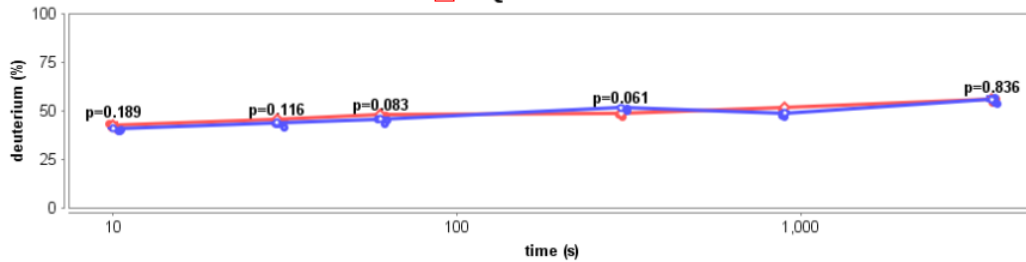
WT CA \pm GS-6207 at 1.5 M NaCl

● CA + DMSO ● CA + GS-6207

1 WDRLHPVHAGPIAPGQMREPRGSD+3 80-103



1 EAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTTSTL+4 76-111

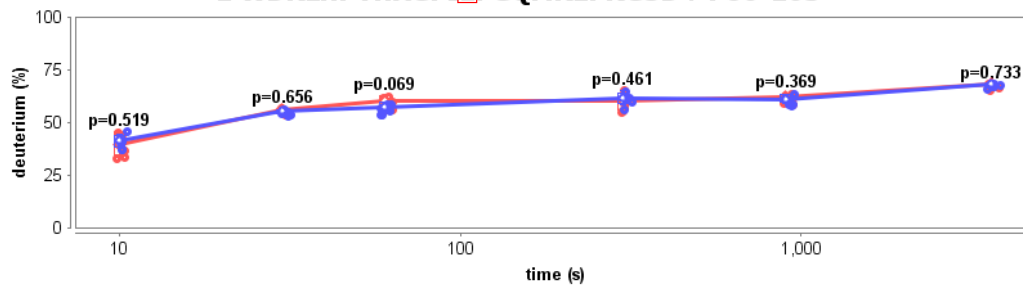


B

A92E CA \pm GS-6207 at 1 M NaCl

● CA + DMSO ● CA + GS-6207

1 WDRLHPVHAGPIEPGQMREPRGSD+4 80-103



1 EAAEWDRLHPVHAGPIEPGQMREPRGSDIAGTTSTL+4 76-111

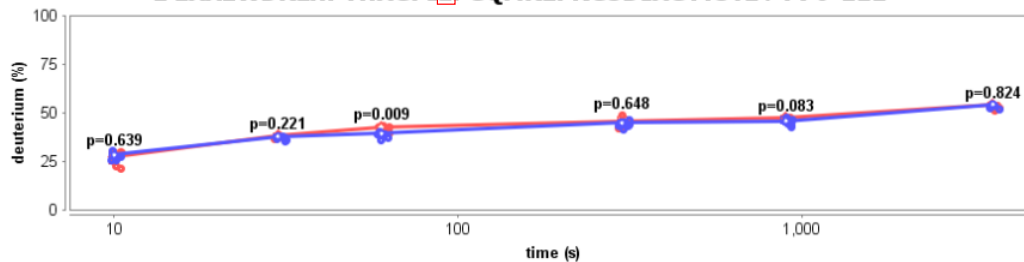
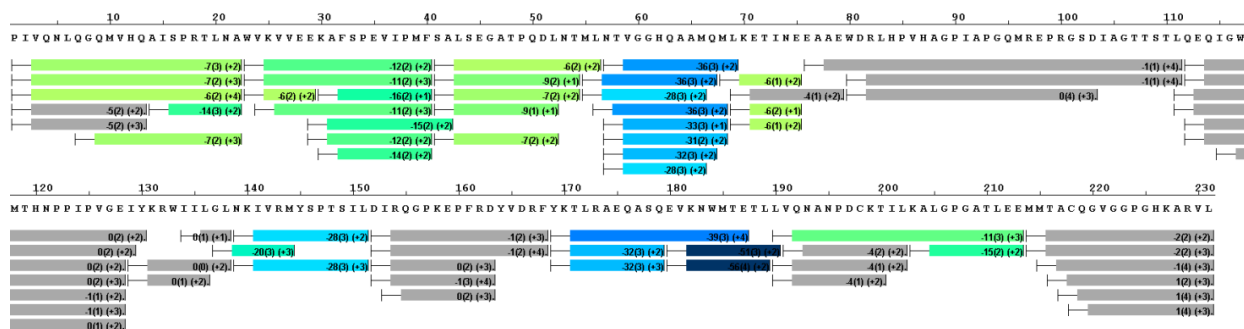


Fig. S43

Representative deuterium build-up plots showing comparable dynamics in the differential HDX-MS experiment with WT (A) or A92E mutant (B) CA tubes bound to GS-6207. The peptides containing CA residue 92 (denoted by red box) did not exhibit any protections upon addition of GS-6207 to WT (A) or A92E (B) CA tubes.

A
WT CA ± GS-6207
at 1.5 M NaCl



B
WT CA ± GS-6207
at 100 mM NaCl

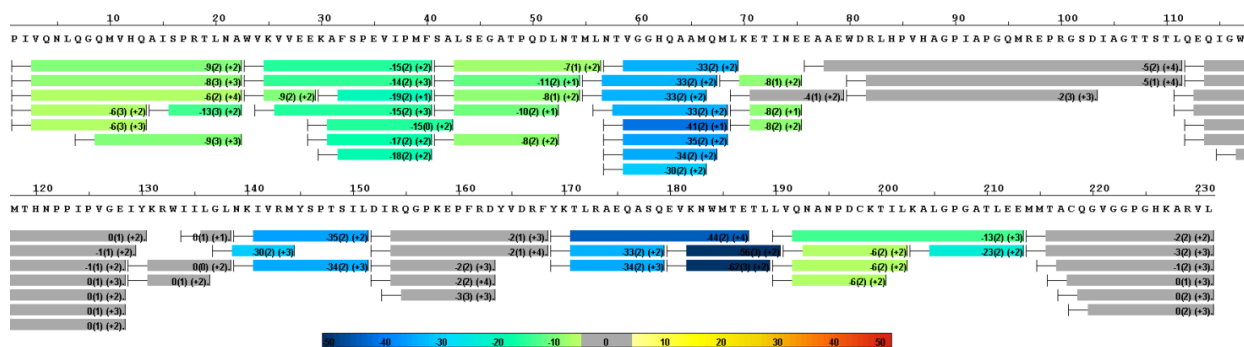


Fig. S44

Results from differential HDX-MS comparing WT CA tubes binding to GS-6207 in high (A) and low (B) salt buffers.

Sequence	Charge	Start	End	1.5 M NaCl	100 mM NaCl
PIVQNLQGMVHQ	2	1	13	-5 (2)*	-6 (3)
PIVQNLQGMVHQ	3	1	13	-5 (2)*	-6 (3)
PIVQNLQGMVHQAISPRTLNA	2	1	22	-7 (3)	-9 (2)
PIVQNLQGMVHQAISPRTLNA	3	1	22	-7 (2)	-8 (3)
PIVQNLQGMVHQAISPRTLNA	4	1	22	-6 (2)	-6 (2)
QGQMVHQAISPRTLNA	3	7	22	-7 (2)	-9 (3)
AISPRTLNA	2	14	22	-14 (3)	-13 (3)
WVKVVEE	2	23	29	-6 (2)	-9 (2)
WVKVVEEKAFSPEVIPMF	2	23	40	-12 (2)	-15 (2)
WVKVVEEKAFSPEVIPMF	3	23	40	-11 (2)	-14 (2)
YKVVVEEKAFSPEVIPMF	3	24	40	-11 (2)	-15 (2)
EKAFSPEVIPMF	2	29	40	-12 (2)	-17 (2)
EKAFSPEVIPMFSA	2	29	42	-15 (2)	-15 (0)
KAFSPEVIPMF	1	30	40	-16 (2)	-19 (2)
KAFSPEVIPMF	2	30	40	-14 (2)	-18 (2)
SALSEGATPQDL	1	41	52	-9 (1)	-10 (2)
SALSEGATPQDL	2	41	52	-7 (2)	-8 (2)
SALSEGATPQDLNT	1	41	54	-9 (2)	-11 (2)
SALSEGATPQDLNT	2	41	54	-7 (2)	-8 (1)
SALSEGATPQDLNTML	2	41	56	-6 (2)	-7 (1)
MLNTVGGHQAAM	2	55	66	-28 (3)	-33 (2)
MLNTVGGHQAAMQ	2	55	67	-36 (3)	-33 (2)
LNTVGGHQAAMQM	2	56	68	-36 (3)	-33 (2)
NTVGGHQAAM	2	57	66	-28 (3)	-30 (2)
NTVGGHQAAMQ	2	57	67	-32 (3)	-34 (2)
NTVGGHQAAMQM	1	57	68	-33 (3)	-41 (2)
NTVGGHQAAMQM	2	57	68	-31 (2)	-35 (2)
NTVGGHQAAMQML	2	57	69	-36 (3)	-33 (2)
MLKETINE	2	68	75	-6 (1)	-8 (1)
LKETINE	1	69	75	-6 (2)	-8 (2)
LKETINE	2	69	75	-6 (1)	-8 (2)
LKETINEEAAE	2	69	79	-4 (1)*	-4 (1)*
EAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTTSTL	4	76	111	-1 (1)*	-5 (2)*
WDRLHPVHAGPIAPGQMREPRGSD	3	80	103	0 (4)*	-2 (3)*
WDRLHPVHAGPIAPGQMREPRGSDIAGTTSTL	4	80	111	-1 (1)*	-5 (1)*
LQEQIGWMTHNPPIPVGE	2	111	128	0 (2)*	-1 (1)*
LQEQIGWMTHNPPIPVGE	3	111	128	0 (2)*	0 (1)*
QEQIGWMTHNPPIPVGE	2	112	128	-1 (1)*	0 (1)*
QEQIGWMTHNPPIPVGE	3	112	128	-1 (1)*	0 (1)*
QEQIGWMTHNPPIPVGEI	2	112	129	0 (2)*	-1 (1)*
QEQIGWMTHNPPIPVGEIY	2	112	130	0 (2)*	0 (1)*
IGWMTHNPPIPVGE	2	115	128	0 (1)*	0 (1)*
IYKRWIL	2	129	136	0 (1)*	0 (1)*
IYKRWILGL	2	129	138	0 (0)*	0 (0)*
IILGL	1	134	138	0 (1)*	0 (1)*
GLNKIVRM	3	137	144	-20 (3)	-30 (2)
NKIVRMYSPTSIL	2	139	151	-28 (3)	-35 (2)
NKIVRMYSPTSIL	3	139	151	-28 (3)	-34 (2)
DIRQGPKEPFRD	3	152	163	0 (2)*	-2 (2)*
DIRQGPKEPFRD	4	152	163	-1 (3)*	-2 (2)*
DIRQGPKEPFRDYVDRF	3	152	168	-1 (2)*	-2 (1)*
DIRQGPKEPFRDYVDRF	4	152	168	-1 (2)*	-2 (1)*
IRQGPKEPFRD	3	153	163	0 (2)*	-3 (3)*
YKTLRAEQASQ	2	169	179	-32 (3)	-33 (2)
YKTLRAEQASQ	3	169	179	-32 (3)	-34 (2)
YKTLRAEQASQEVKNWMTTE	4	169	187	-39 (3)	-44 (2)
EVKNWMTTEL	2	180	189	-56 (4)	-62 (3)
EVKNWMTETLL	2	180	190	-51 (3)	-56 (3)
LVQANPDCKT	2	190	200	-4 (1)*	-6 (2)
LVQANPDCKTIL	2	190	202	-4 (1)*	-6 (2)
LVQANPDCKTILKALGPGATLEE	3	190	213	-11 (3)	-13 (2)
VQANPDCKTIL	2	191	202	-4 (2)*	-6 (2)
KALGPGATLEE	2	203	213	-15 (2)	-23 (2)
MMTACQVGGPGHKARVL	2	214	231	-2 (2)*	-2 (2)*
MMTACQVGGPGHKARVL	3	214	231	-2 (2)*	-3 (2)*
MTACQVGGPGHKARVL	3	215	231	-1 (4)*	-1 (2)*
TACQVGGPGHKARVL	3	216	231	1 (2)*	0 (1)*
ACQVGGPGHKARVL	3	217	231	1 (4)*	0 (2)*
CQVGGPGHKARVL	3	218	231	1 (4)*	0 (2)*

Fig. S45

Results from differential HDX-MS comparing WT CA binding to GS-6207 in high and low salt buffers. Individual CA peptide sequences and their protection patterns observed by HDX-MS experiments are shown.

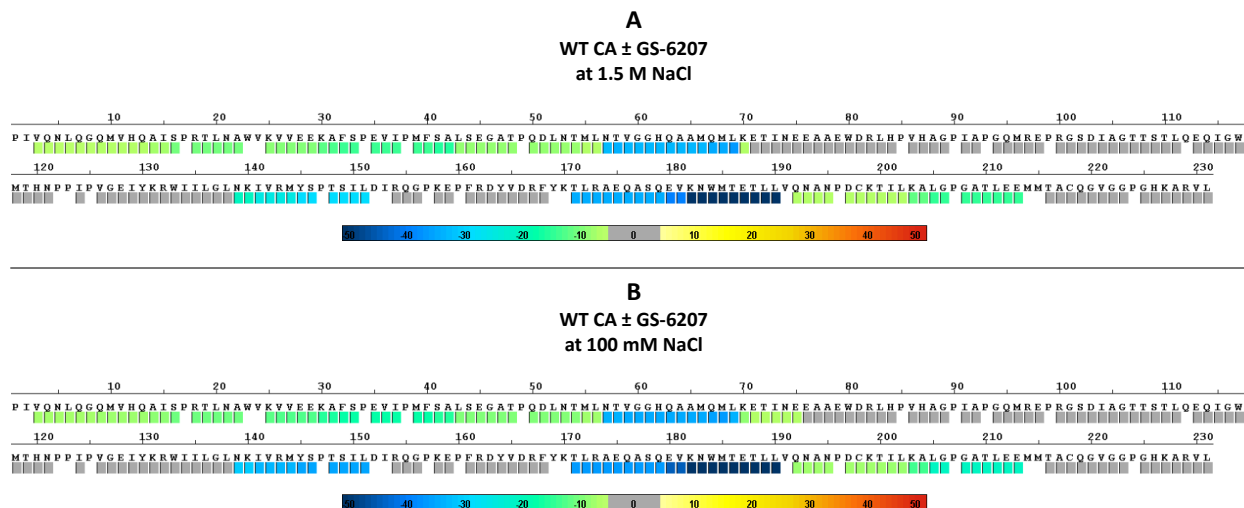


Fig. S46

Consolidated results from differential HDX-MS comparing WT CA tubes binding to GS-6207 in high (A) and low (B) salt buffers.

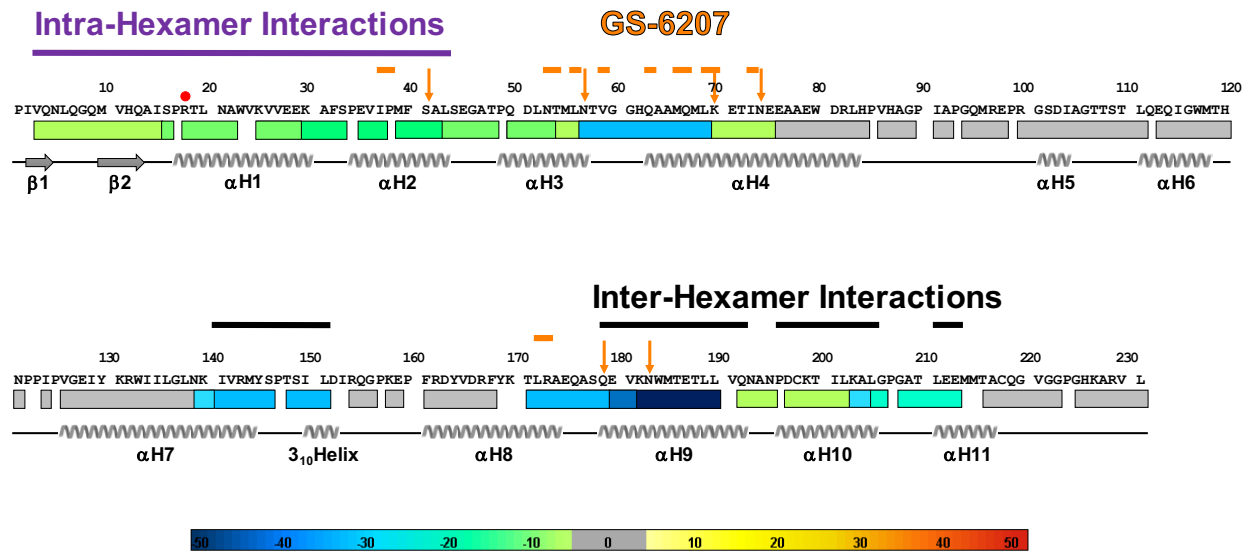


Fig. S47

Interpretation of the HDX results in the context GS-6207 mediated stabilization of CA tubes.

The primary structure of WT CA is shown with HDX protections indicated underneath. The purple and black bars above the sequence highlight CA regions that mediate intra- and inter-hexamer interactions, respectively. The CA regions interacting with GS-6207 through Van der Waals and hydrogen bonding interactions are indicated by orange bars and arrows, respectively. The principal IP6 binding residues R18 is indicated by a red circle. The secondary structure of WT CA is displayed underneath the HDX protections.

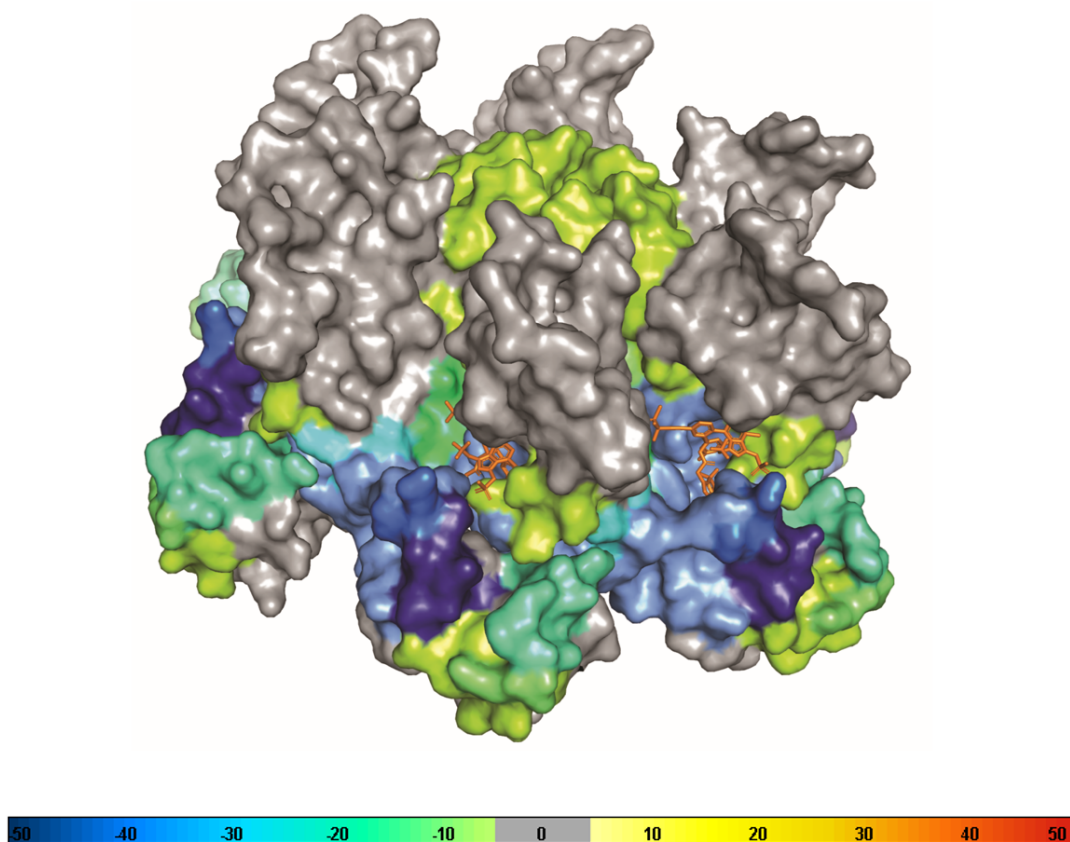


Fig. S48

HDX protections highlighted in the context of the co-crystal structure of GS-6207 bound CA hexamer. Surface and cartoon representation of CA_{A14C/E45C/W184A/M185A} in complex with GS-6207 (orange; PDB ID: 6VKV) with color coding of the HDX protection pattern. Strong HDX protections are seen at the direct drug binding site.

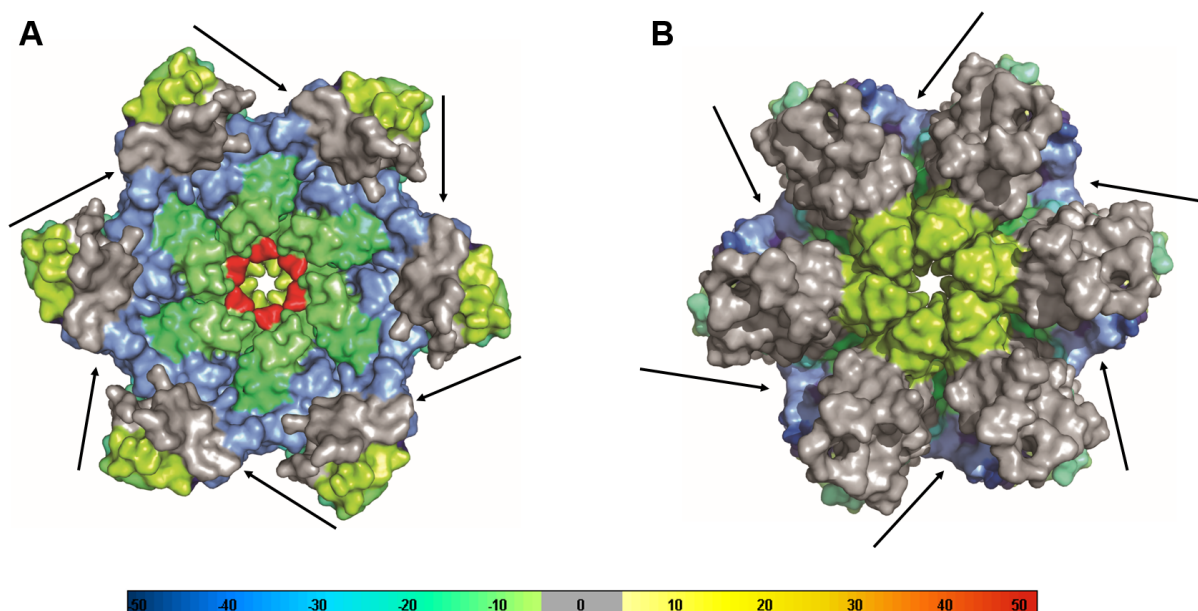


Fig. S49

Overview of HDX protections at intra-hexamer interaction sites. The top (A) and bottom (B) views of color coded CA_{A14C/E45C/W184A/M185A} hexamer. GS-6207 binding sites are indicated by arrows. The red region of the surface (A) indicates R18, which is a vital CA residue for binding to IP6.

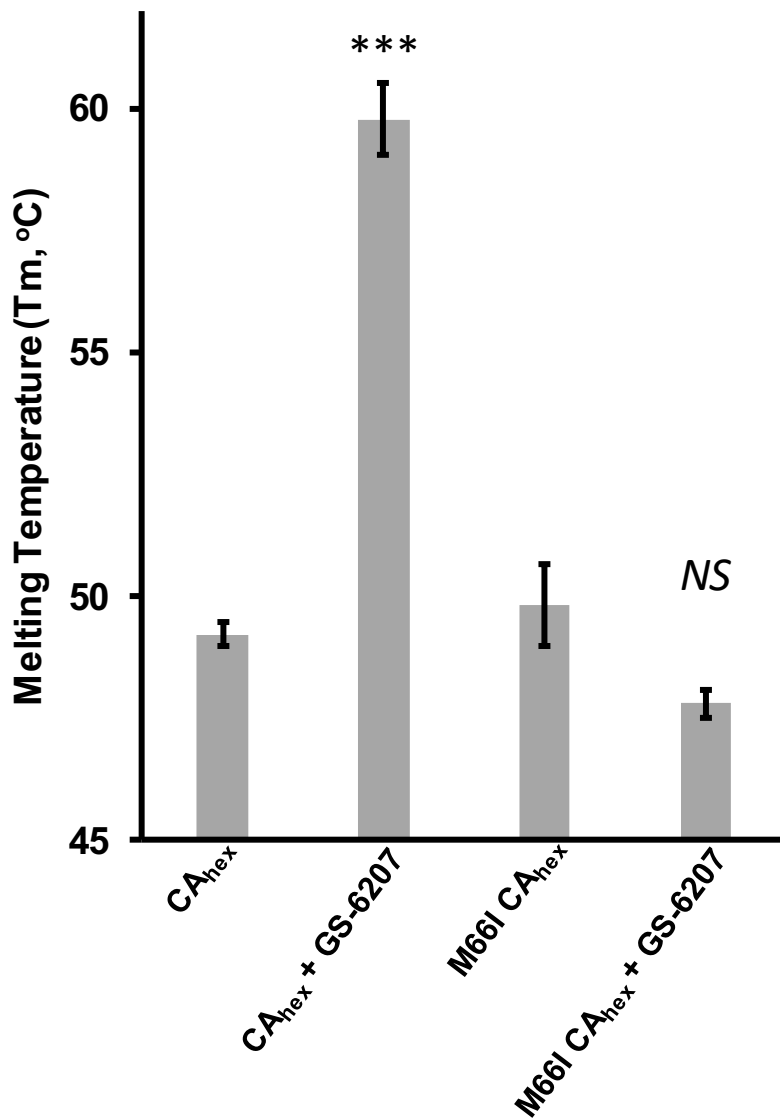


Fig. S50

Effect of GS-6207 on the thermal stability of CA hexamers. Addition of GS-6207 substantially increases (by ~10 °C, $p = 0.00078$, denoted as ***) the melting temperature (T_m) of cross-linked CA hexamers (CA_{hex}). The cross-linked hexamers containing the M66I resistance substitution (M66I CA_{hex}) were used as a control. CA_{hex} and M66I CA_{hex} yielded comparable T_m s, indicating that the M66I substitution did not grossly interfere with the architecture of the hexamers. However, the stability of M66I CA_{hex} did not increase in the presence of GS-6207 (NS, statistically not significant) due likely to the inability of the inhibitor to bind the mutant CA (*I*). The mean values from three independent experimental measurements \pm SD are shown.

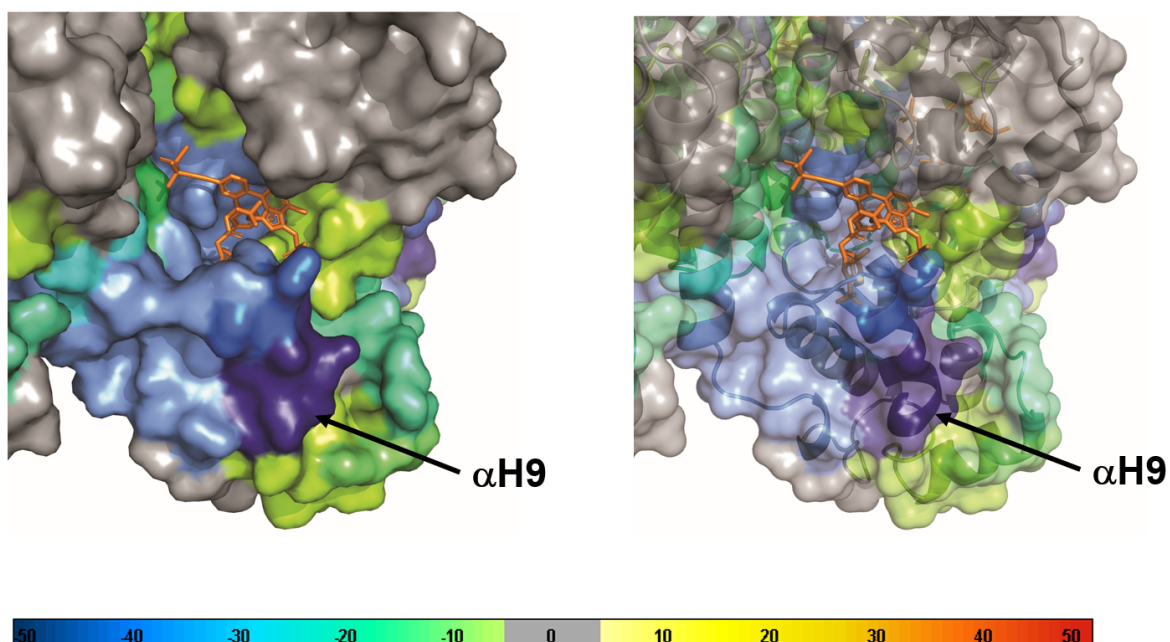


Fig. S51

Overview of HDX protections at inter-hexamer interaction sites. Color coded $CA_{A14C}/E45C/W184A/M185A$ hexamer indicates that the strongest GS-6207 induced protections are seen at α H9 (indigo).

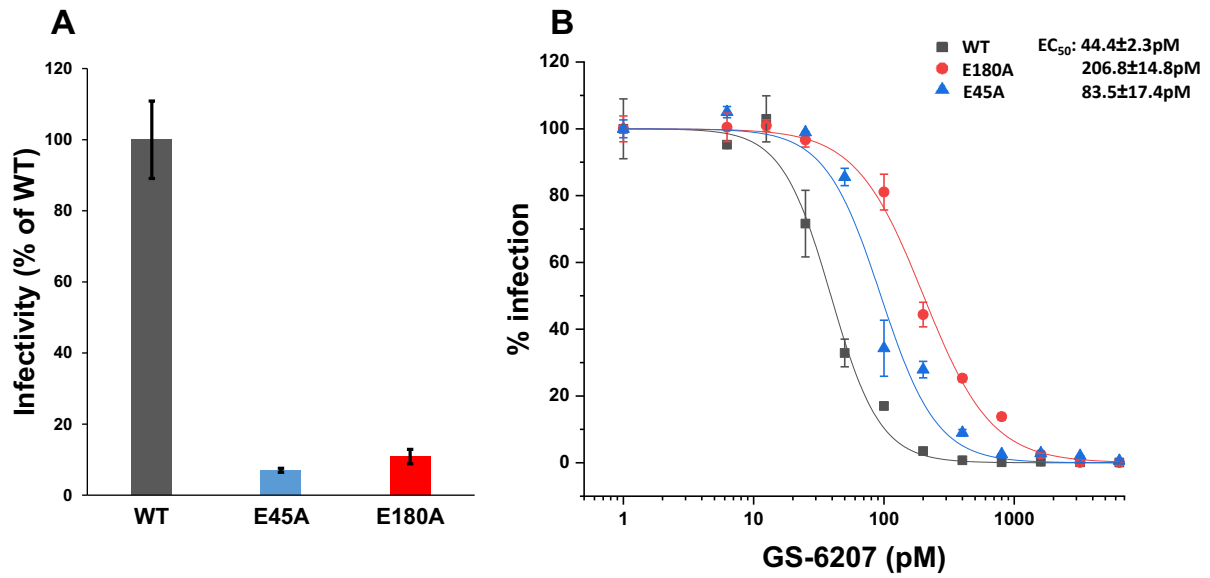


Fig. S52

Effects of E45A and E180A CA substitution on HIV-1 infectivity (A) and GS-6207 potency (B). Error bars indicate SDs from three independent experiments.

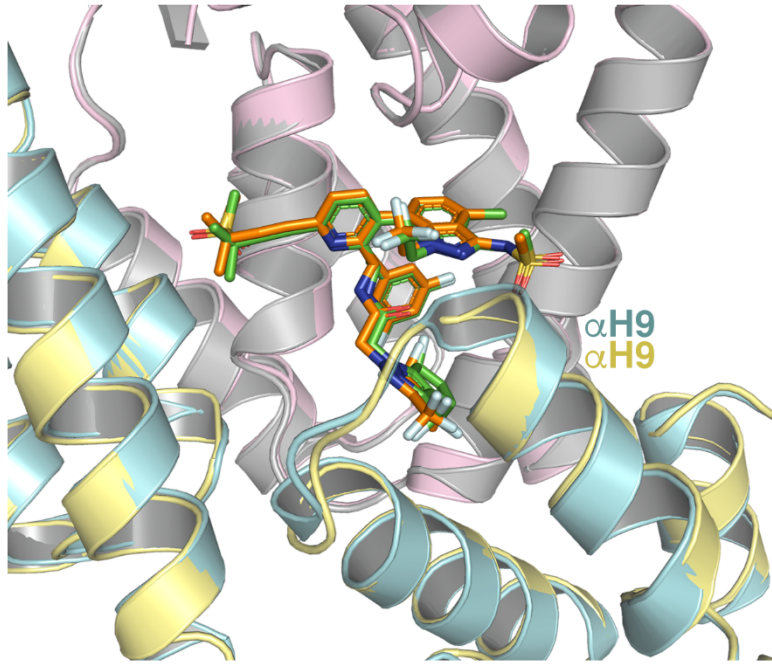
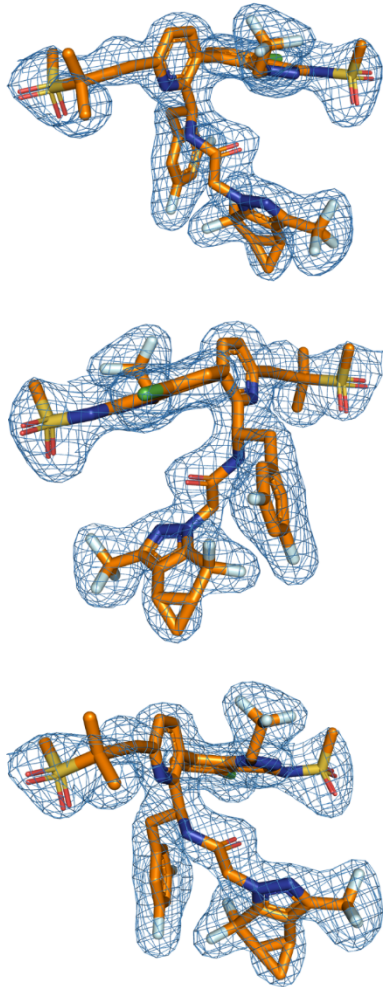
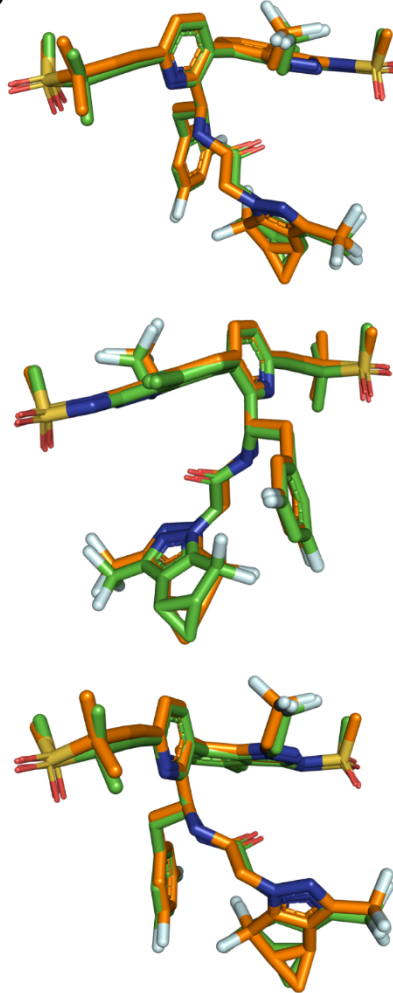
A**B****C**

Fig. 53.

Comparison of our (PDB ID 6VKV) and recently published (PDB ID 6V2F) x-ray structures of GS-6207 bound to the CA_{A14C/E45C/W184A/M185A} hexamer. A) Superimposing our CA_{A14C/E45C/W184A/M185A} hexamer (pale yellow/grey) + GS-6207 (orange) structure (PDB ID 6VKV) with the recently published CA_{A14C/E45C/W184A/M185A} hexamer (pale pink/light cyan) + GS-6207 (green) structure (PDB ID 6V2F) (26) reveals defined density for α H9 in the both structures. These helices, which we believe play a major role in the mechanism of GS-6207, from two different structures superimpose exceptionally well. This similarity is unlikely to be trivial as this region is known to be quite flexible and crystal structures of cross-linked CA_{A14C/E45C/W184A/M185A} hexamers often do not have defined density for α H9 as the mutations, W184A and M185A, intentionally disrupt α H9- α H9 interactions at the dimer interface. Additionally, the two structures are in different space groups, P6 vs P3, and were crystallized in different crystal conditions. B) Three rotated views of $2F_o - F_c$ density scaled to 1σ (light blue mesh) for GS-6207 (orange) bound to CA_{A14C/E45C/W184A/M185A} hexamer (for clarity the protein structure is not shown; PDB ID 6VKV). C) Three rotated close-up views of GS-6207 from our structure superimposed onto the published structure (PDB ID 6VKV and 6V2F, respectively) show very similar positioning of the inhibitor within the binding pocket as well as similar positioning of the moieties within the compound itself.

Table S1.
Antiviral activities of GS-6207.

HIV-1 replication	Cells	EC ₅₀ values of GS-6207 (pM)
Complete cycle	PBMC	12.0 ± 1.4
	MT4	46.1 ± 6.3
Early steps	Jurkat	46.0 ± 5.5
	MT4	55.1 ± 3.6
	Thp1 (undifferentiated)	26.6 ± 2.7
	Thp1 (differentiated)	24.1 ± 4.3
	Hela	44.6 ± 7.0
Late steps	HEK293T	313.9 ± 33.0

Averages ± SD for minimally three independent experiments.

Table S2.

Summary of X-ray data collection and refinement statistics.

CA Hexamer + GS-6207 PDB ID: 6VKV	
Data collection	
X-ray Source	ALS 4.2.2
Software	XDS
Wavelength	1.00003 Å
Space group	P6
Unit cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	160.05, 160.05, 57.536
α , β , γ (°)	90, 90, 120
Resolution (Å)	46.71 – 2.22 (2.29 - 2.22)
No. total reflections	446184 (32922)
No. unique reflections	41892 (3828)
<i>R</i> _{merge}	0.296 (2.412)
<i>R</i> _{pim}	0.101 (0.970)
CC1/2	0.994
<i>I</i> / σ <i>I</i>	8.0 (1.0)
Completeness (%)	100.0 (99.9)
Multiplicity	10.7 (8.6)
Refinement	
Resolution (Å)	44.26 – 2.22 (2.299 – 2.22)
No. reflections used in refinement	41745 (4045)
No. reflections used for <i>R</i> _{free}	2051 (189)
<i>R</i> _{work} (%)	23.14 (40.52)
<i>R</i> _{free} (%)	26.85 (41.91)
No. non-hydrogen atoms	5783
Protein	5143
Ligand/ion	403
Water	237
Wilson B-factor	29.55
Average B-factors	40.90
Protein	41.77
Ligands/ions	31.87
Waters	37.48
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	0.97
Ramachandran	
Favored (%)	98.2
Allowed (%)	1.8
Outliers (%)	0
Rotamer outliers (%)	0.18
Clashscore	6.91

Data were collected from a single crystal. Values in parentheses are from the highest-resolution shell.

Table S3.

Hydrogen bonding interactions of CA with GS-6207, PF-74, CPSF6 and Nup153.

CA Residues	GS-6207 Atoms	PF-74 Atoms	CPSF6 Residues	Nup153 Residues
S41	O42	-	-	-
N57	O40 N3 N29	O2 N3	F321	T1416 F1417
Q63	-	N32	-	-
Q67	-	-	G318	-
K70	O4 O59	O1	-	-
N74	O57	-	V314 L315	-
R173	-	-	P317*	V1414 T1416
Q176	-	-	-	G1413
A177	-	-	-	S1412
Q179	O59	-	G318	-
N183	O59	-	-	-

Table S4.

Antiviral activities of GS-6207 with respect to indicated mutant viruses.

HIV-1	GS-6207 EC₅₀ (pM)	Fold-change
WT	45 ± 7	1
N57S	2694 ± 90	~60
M66I	>3.8 × 10 ⁶	>8.4 × 10 ⁴
Q67H	436 ± 20	~10
K70A	2045 ± 45	~45
N74D	617 ± 41	~14

Results (average ± SD) are from n = 3 independent experiments.

Table S5.
Cryo-EM data collection, refinement, and validation statistics.

	HIV capsid Helical EMD-21424	RASTR of HIV capsid Helical 6VWS EMD-21423
Data collection and Processing (for each dataset)		
Microscope	Arctica	Arctica
Voltage (keV)	200	200
Camera	K3	K3
Magnification	28,000	28,000
Pixel size at detector (Å/pixel)	1.4	1.4
Total electron exposure (e-/Å ²)	47.36	47.36
Exposure rate (e-/pixel/sec)	30.94	30.94
Number of frames collected during exposure	40	40
Defocus range (µm)	1.5-2.5	1.5-2.5
Phase plate (if used)		
- phase shift range (in degrees)		
- number of images per phase plate position		
Automation software (EPU, SerialEM or manual)		
Tilt angle (if grid was tilted)		
Energy filter slit width (if used)		
Micrographs collected (no.)	4,264	4,264
Micrographs used (no.)	608	608
Total extracted particles (no.)	208,225	245,050
For each reconstruction:		
Refined particles (no.)	24,505	96,392
Final particles (no.)	24,505	96,392
helical symmetry parameters (twist/rise)	150.32°/ 7.618 Å	
Estimated error of translations/rotations (if available)		
Resolution (global, Å)	6.29	6.08
FSC 0.5 (unmasked/masked)	8.60/7.46	9.26/7.26
FSC 0.143 (unmasked/masked)	7.27/6.29	7.68/6.08
Resolution range (local, Å)		
Resolution range due to anisotropy (Å)		
Map sharpening <i>B</i> factor (Å ²) / (<i>B</i> factor Range)	-377	-109.4
Map sharpening methods	Relion (Postprocess)	Relion (Postprocess)
Model composition (for each model)		
Protein		1320
Ligands		
RNA/DNA		
Model Refinement (for each model)		
Refinement package		PHENIX
- real or reciprocal space		Real space
- resolution cutoff		6.1
Model-Map scores		
-CC		0.481
- Average FSC		
<i>B</i> factors (Å ²)		204.64
Protein residues		
Ligands		

RNA/DNA	
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.016
Bond angles (°)	1.681
Validation (for each model)	
MolProbity score	1.62
CaBLAM outliers	2.31
Clashscore	2.75
Poor rotamers (%)	
C-beta deviations	
EMRinger score (if better than 4 Å resolution)	
Ramachandran plot	
Favored (%)	98.93
Outliers (%)	1.07

Movie S1.

Side view of conformational morph of CA hexamer structures derived from cryo-EM of CA tubes transitioning from GS-6207 bound A92E CA hexamer (PDB ID: 6VWS) to unliganded A92E CA hexamer (PDB ID: 3J34). The conformational morphing was achieved through Chimera (47) by superimposing respective hexamers. It is particularly noticeable in the bottom right how the gold CTD moves away from the gray NTD in the absence of the inhibitor in the unliganded A92E CA hexamer structure (PDB ID: 3J34).

Movie S2.

Top view of movie S1 showing conformational morph of CA hexamer structures derived from cryo-EM of CA tubes transitioning from GS-6207 bound A92E CA hexamer (PDB ID: 6VWS) to unliganded A92E CA hexamer (PDB ID: 3J34).

Movie S3.

Side view of conformational morph of CA hexamers transitioning from GS-6207 bound CA hexamer derived from our cryo-EM of CA tubes + GS-6207 (PDB ID: 6VWS) to the CA hexamer structure derived from cryo-ET of native HIV-1 particles (PDB ID: 5MCX). The conformational morphing was achieved through Chimera (47) by superimposing respective hexamers. It is particularly noticeable in the bottom right how the gold CTD moves away from the gray NTD in the absence of the inhibitor in the native CA shell of HIV-1 particles (PDB ID: 5MCX).

Movie S4.

Top view of movie S3 showing conformational morph of CA hexamers transitioning from GS-6207 bound CA hexamer derived from our cryo-EM of CA tubes + GS-6207 (PDB ID: 6VWS) to CA hexamer structure derived from cryo-ET of native HIV-1 particles (PDB ID: 5MCX).

Movie S5.

Side view of conformational morph of CA hexamers transitioning from a flat surface seen in the crystal structure (PDB ID: 6VKV) to the curved topology seen in our cryo-EM structure

of CA tubes + GS-6207 (PDB ID: 6VWS). The conformational morphing was achieved through Chimera (58) by superimposing respective hexamers. It is particularly noticeable in the bottom right how gold CTD remains closely associated with gray NTD in the presence of GS-6207 as CA hexamer changes its conformation from a flat (PDB ID: 6VKV) to the curved (PDB ID: 6VWS) surface. GS-6207 is not shown because the program allows morphing of only proteins but not the inhibitor.

Movie S6.

Top view of movie S5 showing conformational morph of CA hexamers transitioning from a flat surface seen in the crystal structure (PDB ID: 6VKV) to the curved topology seen in our cryo-EM structure of CA tubes + GS-6207 (PDB ID: 6VWS).

Movie S7.

Side view of conformational morph of CA hexamers transitioning from a flat surface seen in the crystal structure (PDB ID: 6VKV) to the curved topology seen in the cryo-EM structure of unliganded CA tubes (PDB ID: 3J34). The conformational morphing was achieved through Chimera (58) by superimposing respective hexamers. It is particularly noticeable in the bottom right how gold CTD moves away from gray NTD in the absence of the inhibitor as CA hexamer changes its conformation from a flat (PDB ID: 6VKV) to the curved (PDB ID: 3J34) surface.

Movie S8.

Top view of movie S7 showing conformational morph of CA hexamers transitioning from a flat surface seen in the crystal structure (PDB ID: 6VKV) to the curved topology seen in the cryo-EM structure of unliganded tube (PDB ID: 3J34).

Movie S9.

Side view of conformational morph of CA hexamers transitioning from a flat surface seen in the crystal structure (PDB ID: 6VKV) to the curved topology seen in the cryo-ET structure of native HIV-1 particles (PDB ID: 5MCX). The conformational morphing was achieved through Chimera (58) by superimposing respective hexamers. It is particularly noticeable in the bottom right how gold CTD moves away from gray NTD in the absence of the inhibitor as CA

hexamer changes its conformation from a flat (PDB ID: 6VKV) surface to the curved topology (PDB ID: 5MCX) of the native CA shell of HIV-1 particles.

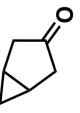
Movie S10.

Top view of movie S9 showing conformational morph of CA hexamers transitioning from a flat surface seen in the crystal structure (PDB ID: 6VKV) to the curved topology seen in the cryo-ET structure of native HIV-1 particles (PDB ID: 5MCX).

¹H, ¹³C, and HSQC NMR spectra of GS-6207 and synthetic intermediates

¹H, ¹³C, and HSQC NMR spectra were recorded on a Bruker DPX 400 spectrometer in CDCl₃, CD₃OD, or DMSO-*d*₆ solutions operating at 400 MHz for ¹H NMR and 100 MHz for ¹³C NMR unless otherwise stated. Spectral data is provided for intermediates **2**, **5**, **9**, **10a**, **11**, **13**, **15a**, **16**, **19**, **20a**, **21**, **22**, **24**, **25**, **27**, subunits A-C, and **GS-6207**.

GS-6207 spectra include ¹H NMR spectra of freshly prepared samples in CDCl₃, DMSO-*d*₆, and CD₃OD; a ¹³C spectrum in CD₃OD; an HSQC spectrum in DMSO-*d*₆; and ¹H NMR spectra of samples of GS-6207 allowed to stand in CD₃OD for 4 days in both CD₃OD and DMSO-*d*₆.

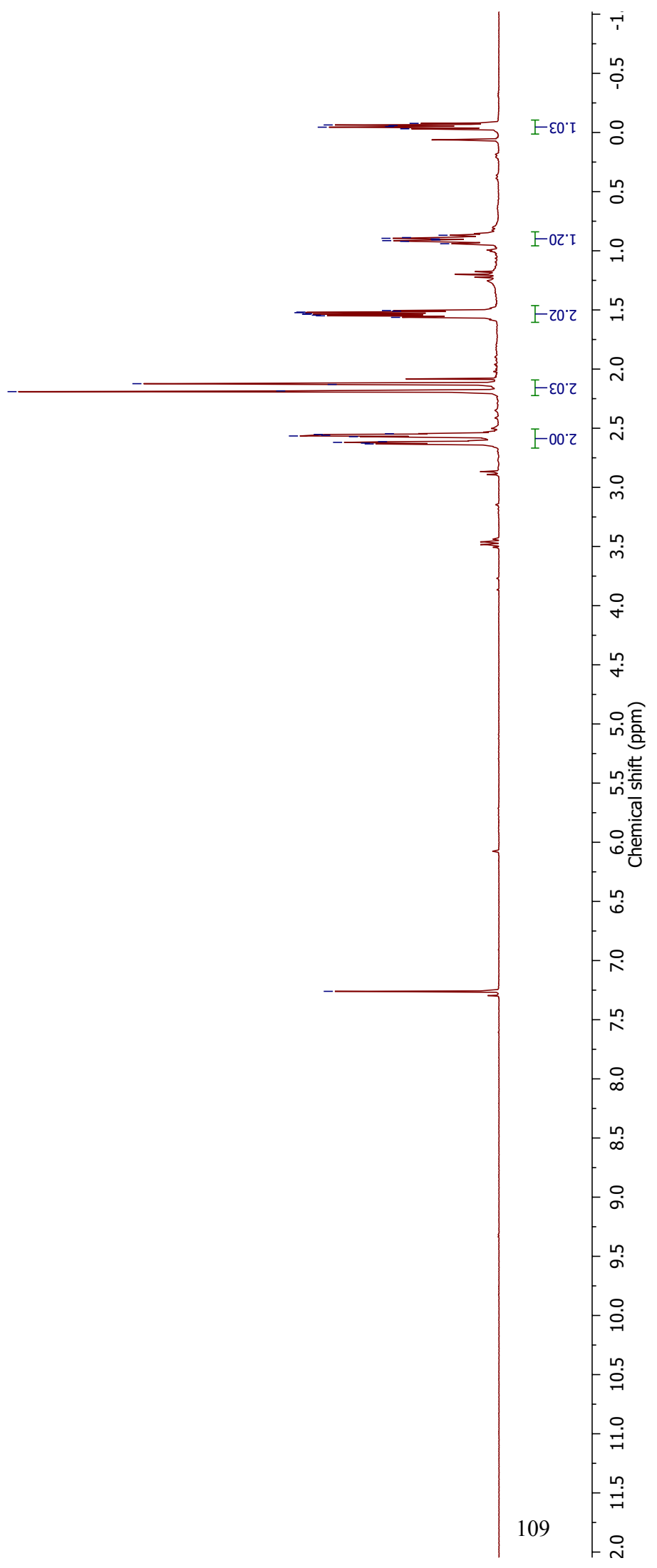


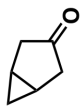
2

¹H NMR (400 MHz, CDCl₃)

— 7.26 CDCl₃

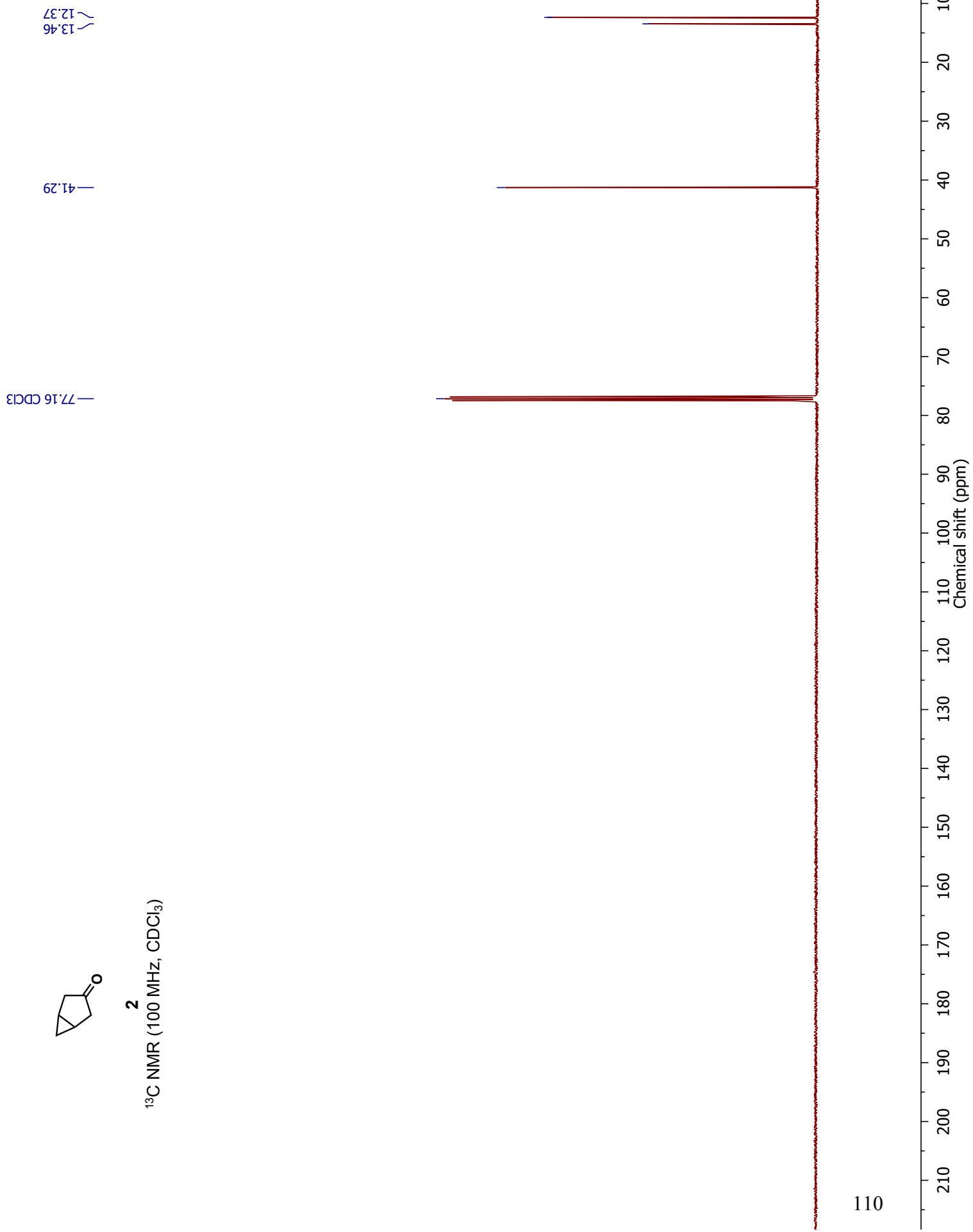
2.63
2.62
2.61
2.57
2.57
2.56
2.55
2.55
2.19
2.18
2.13
2.12
1.56
1.55
1.54
1.54
1.53
1.52
1.52
1.51
1.50
0.92
0.91
0.89
0.89
-0.03
-0.04
-0.05
-0.06

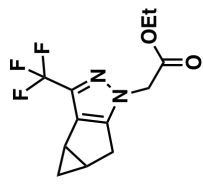




2

¹³C NMR (100 MHz, CDCl₃)

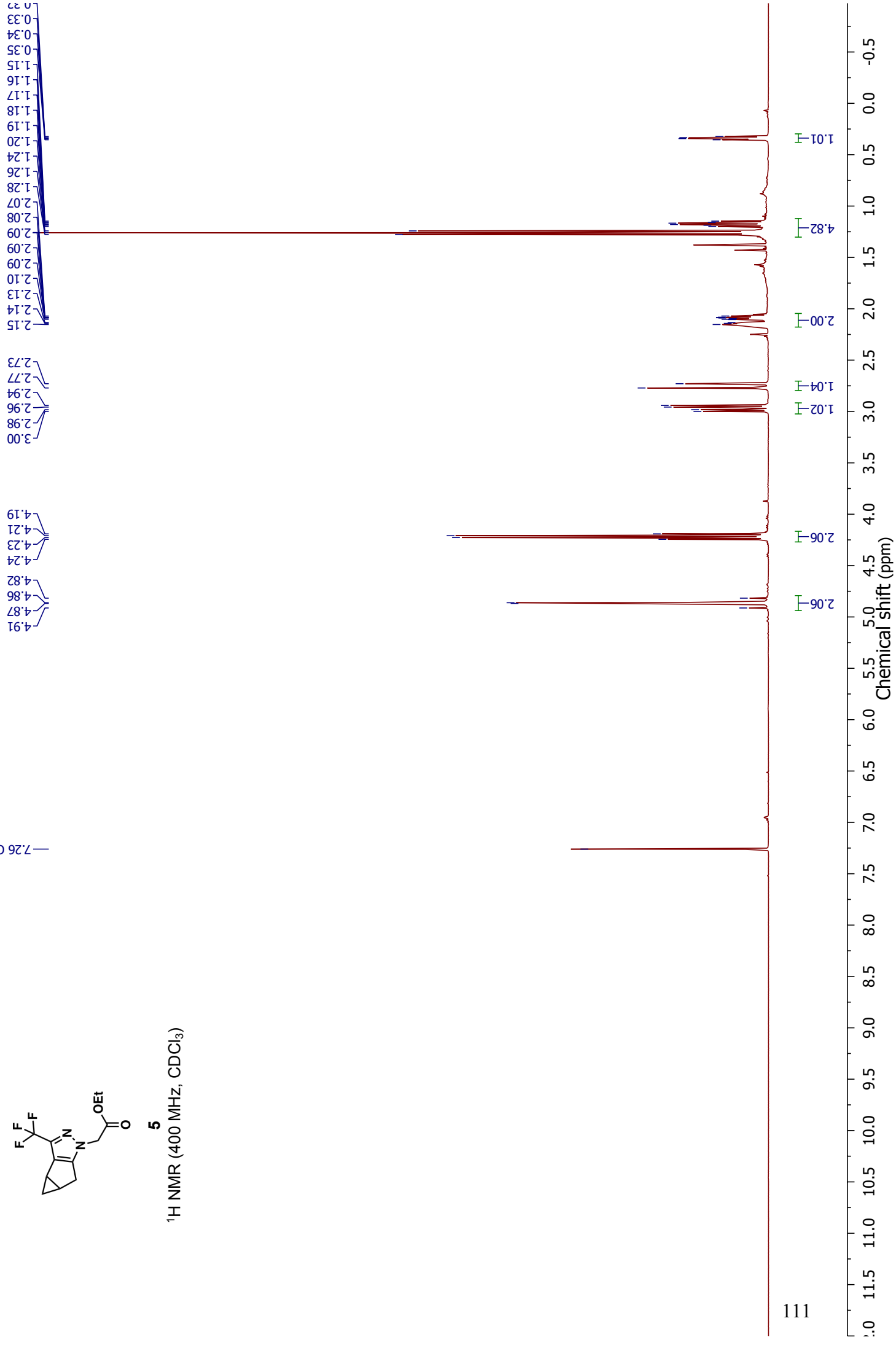


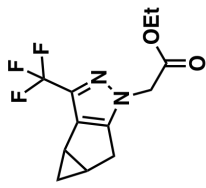


5

¹H NMR (400 MHz, CDCl₃)

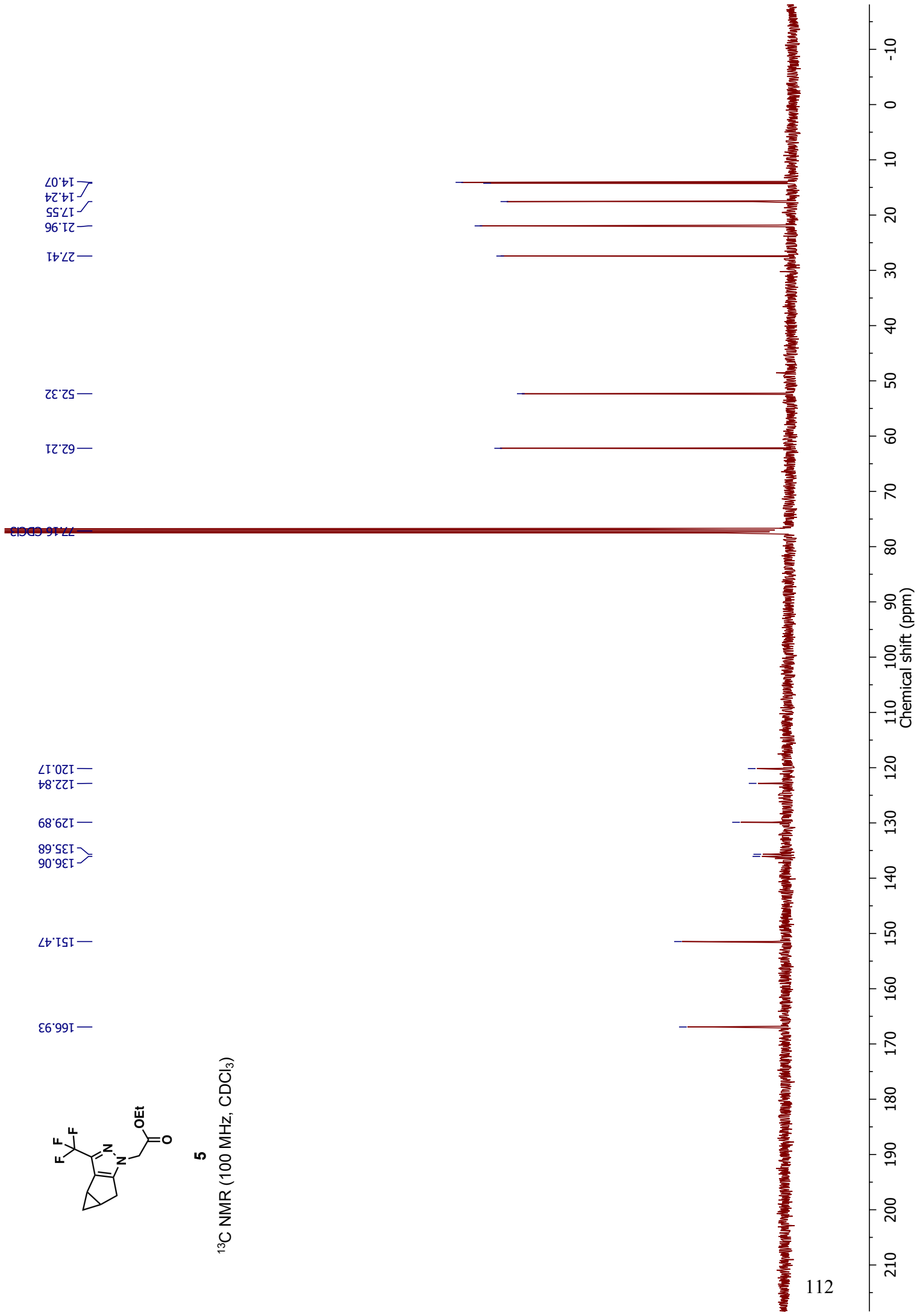
— 7.26 CDCl₃

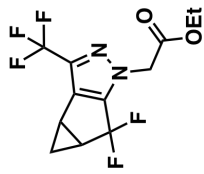




5

^{13}C NMR (100 MHz, CDCl_3)

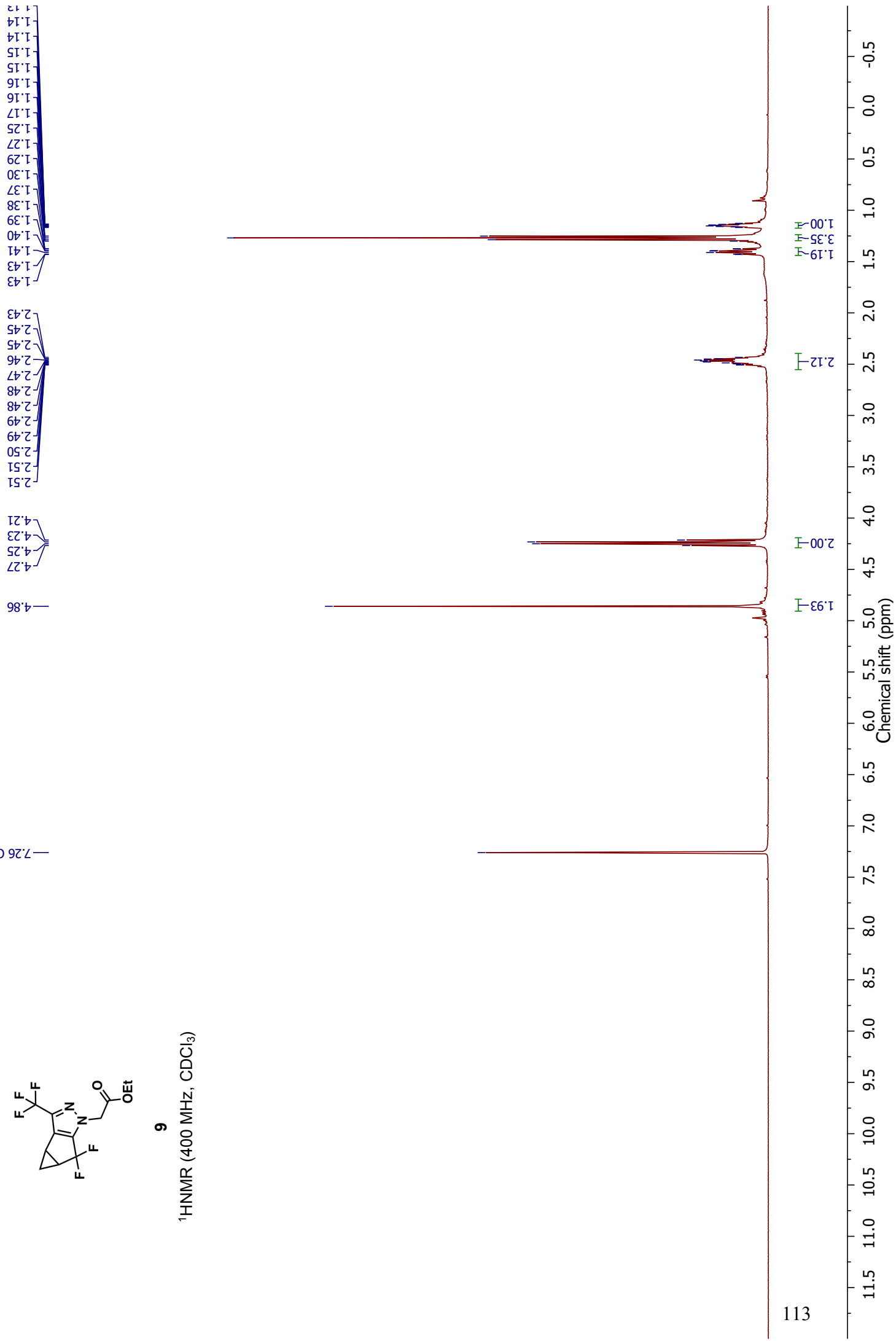


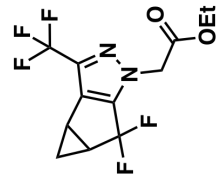


9

¹HNMR (400 MHz, CDCl₃)

— 7.26 CDCl₃





9

¹³C NMR (100 MHz, CDCl₃)

12.41
12.44
12.48
14.12
23.43
27.81
28.10
28.15
28.44

52.20

62.48

77.16 CDCl₃

117.74
119.34
120.17
122.02
122.60

132.84
136.28
136.67

143.40

166.03

Chemical shift (ppm)

-10

0

10

20

30

40

50

60

70

80

90

100

110

120

130

140

150

160

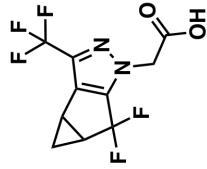
170

180

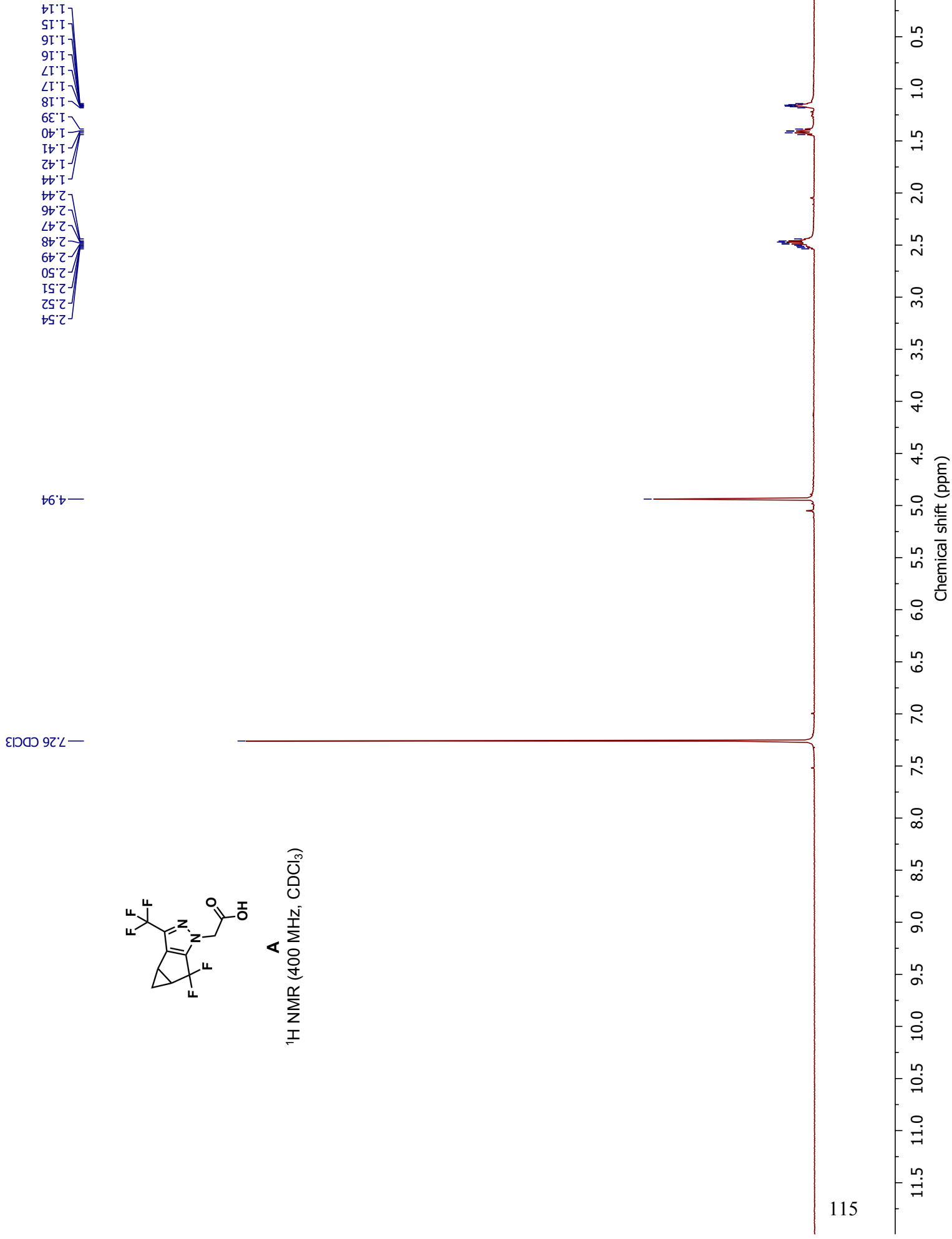
190

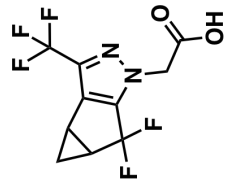
200

210

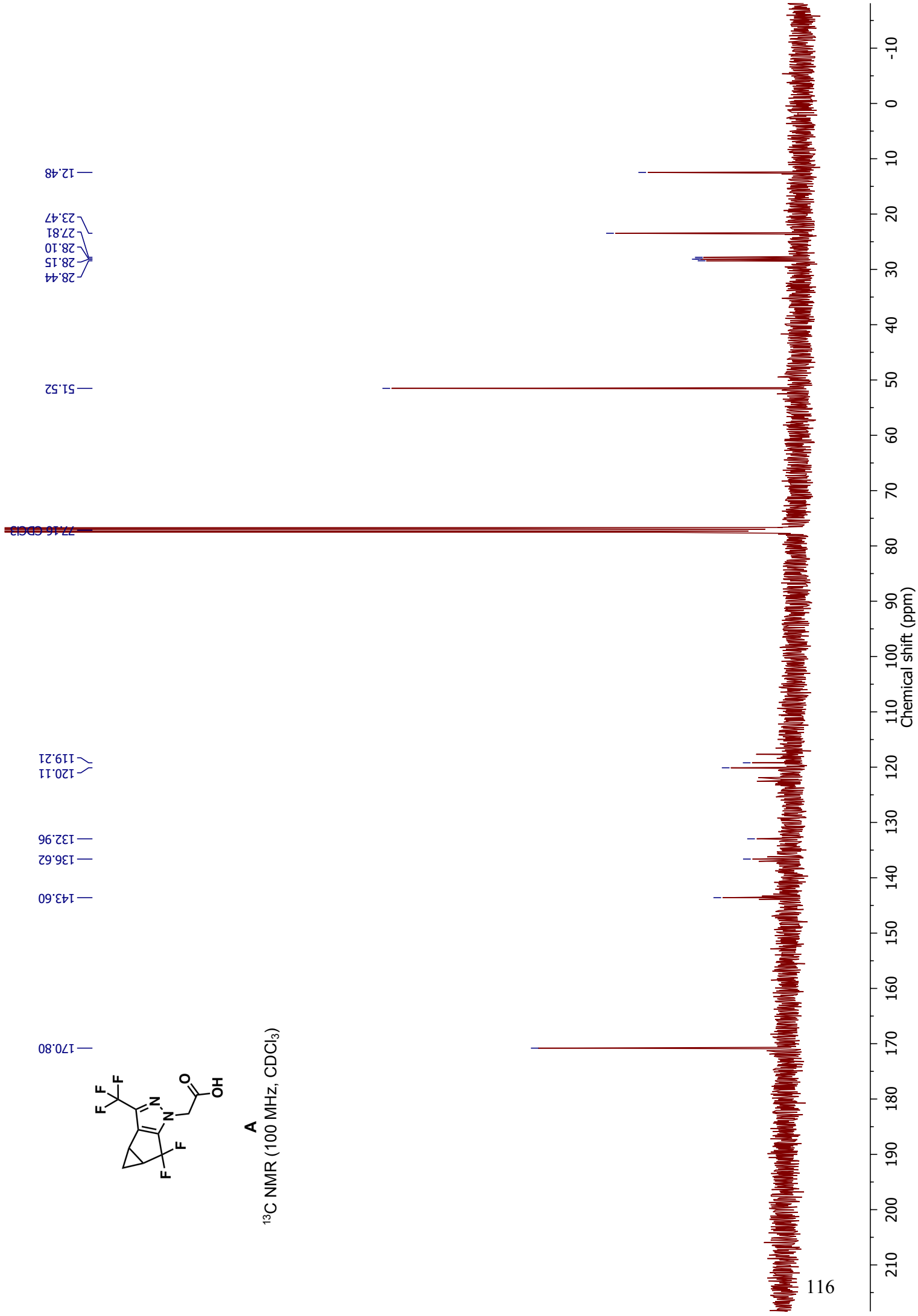


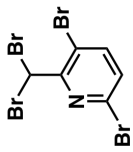
A
¹H NMR (400 MHz, CDCl₃)





A
¹³C NMR (100 MHz, CDCl₃)

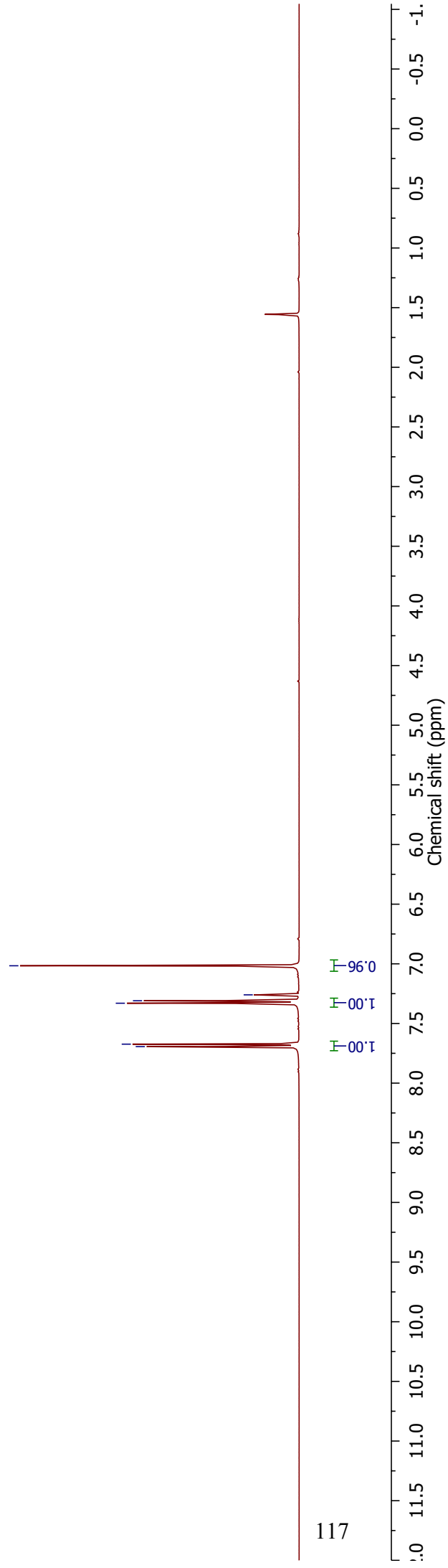


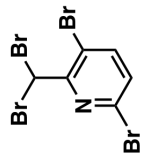


3,6-Dibromo-2-(dibromomethyl)pyridine (10a)

¹H NMR (400 MHz, CDCl₃)

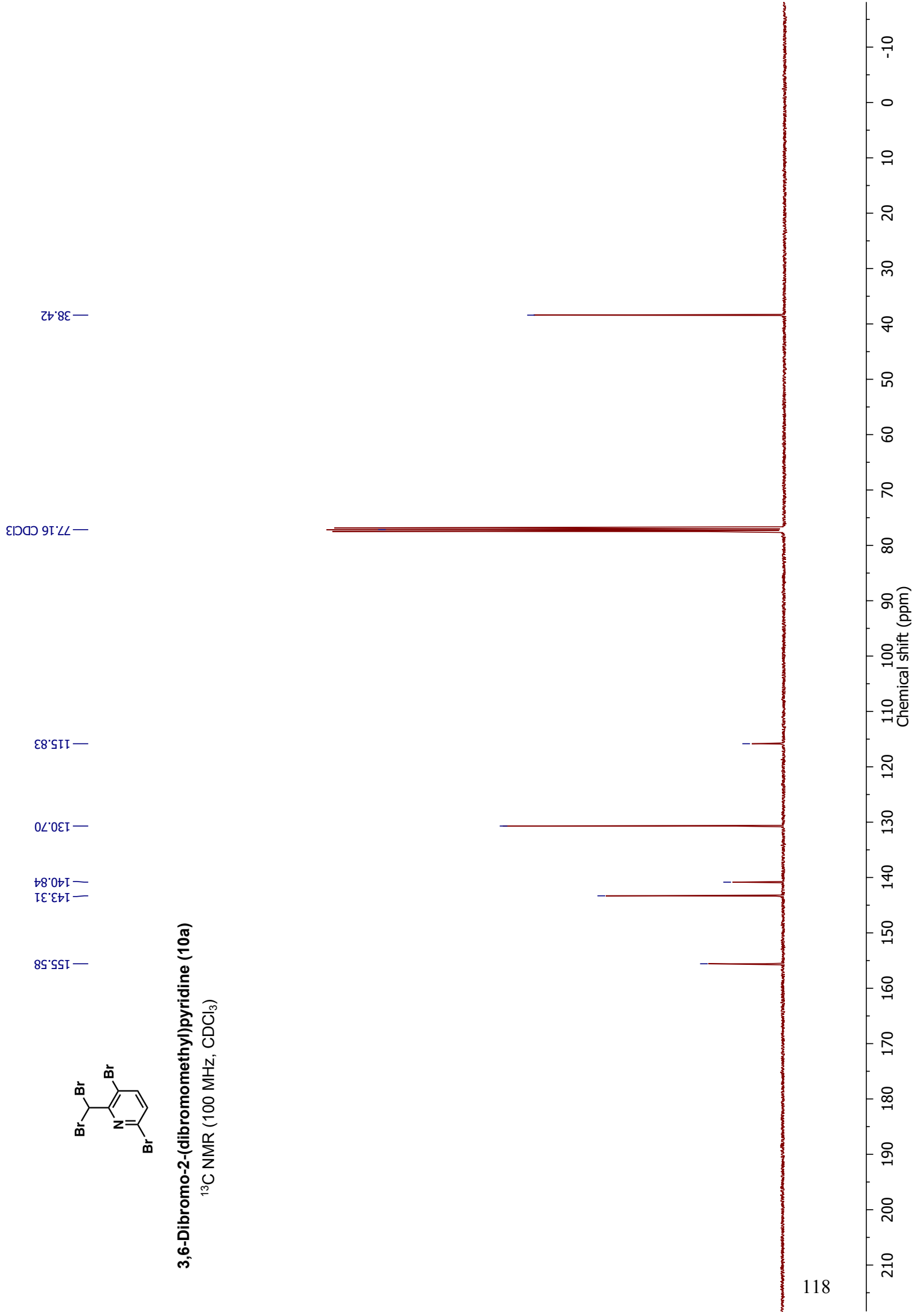
7.69
7.33
7.31
7.26
7.02
CDCl₃

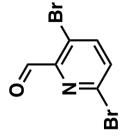




3,6-Dibromo-2-(dibromomethyl)pyridine (10a)

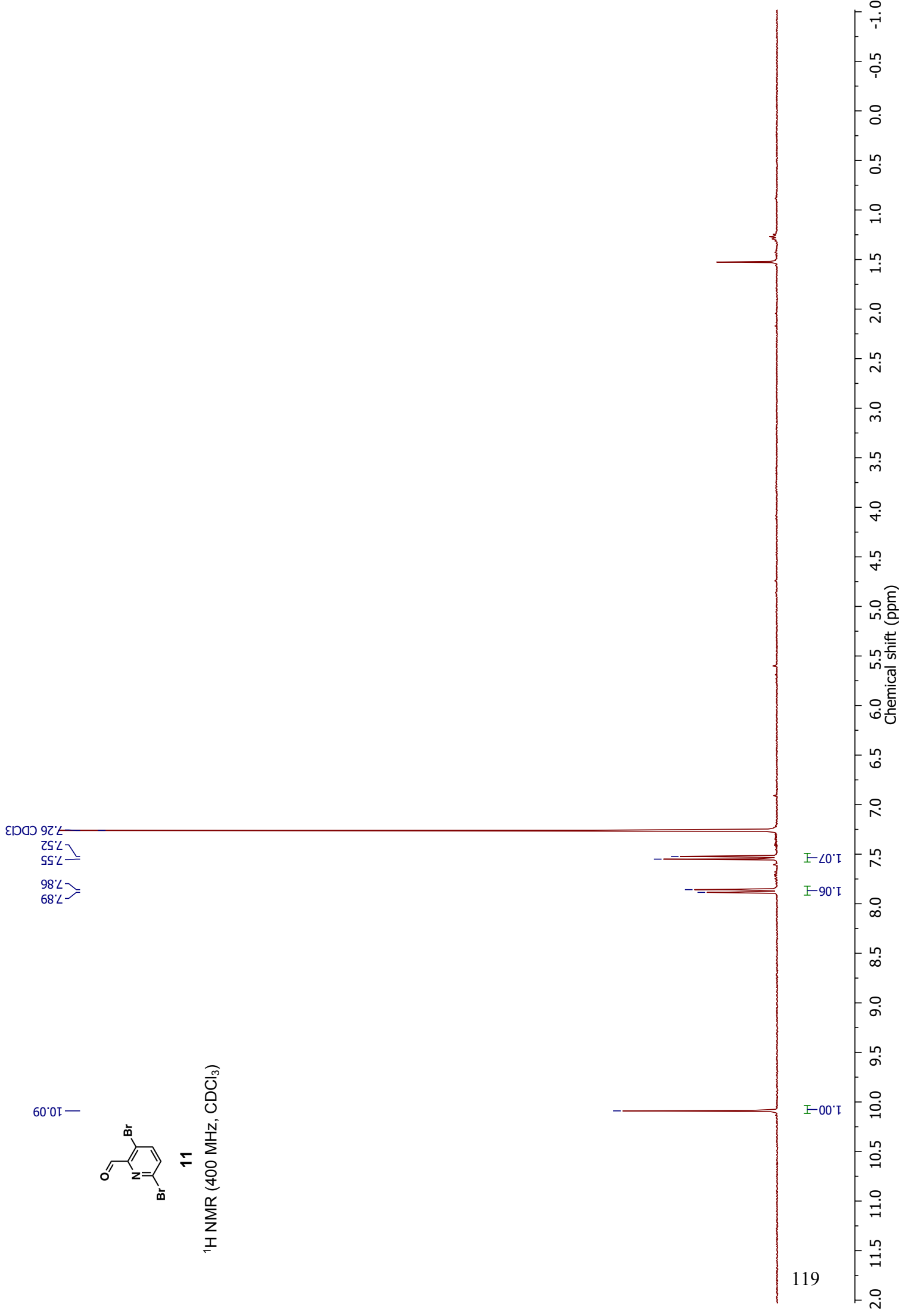
¹³C NMR (100 MHz, CDCl₃)

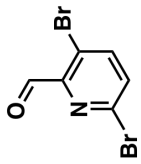




11

¹H NMR (400 MHz, CDCl₃)





11

¹³C NMR (100 MHz, CDCl₃)

189.40

149.07

144.94

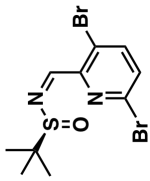
141.11

132.96

120.26

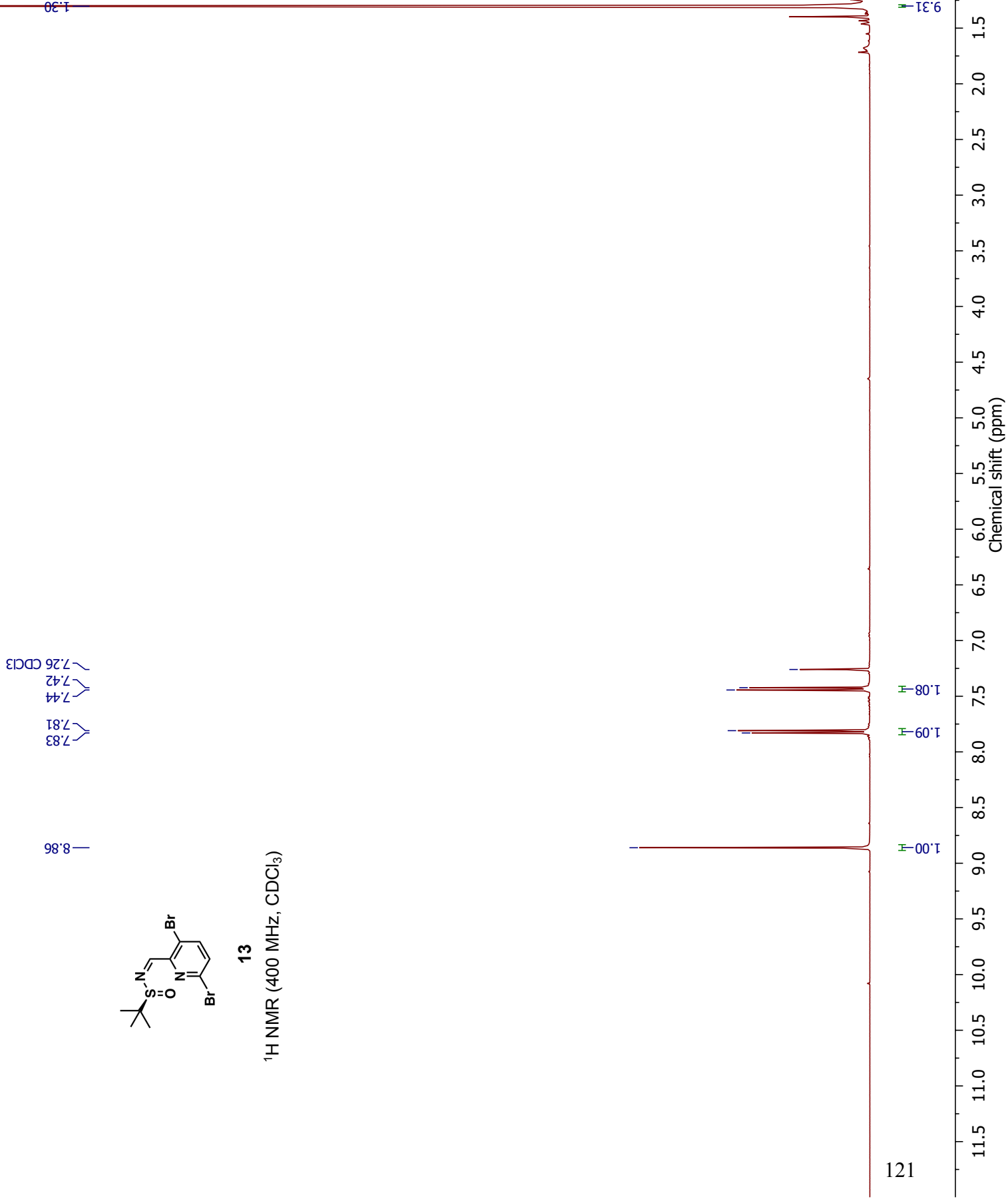
77.16 CDCl₃

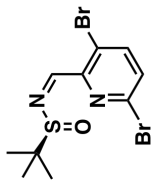
Chemical shift (ppm)



13

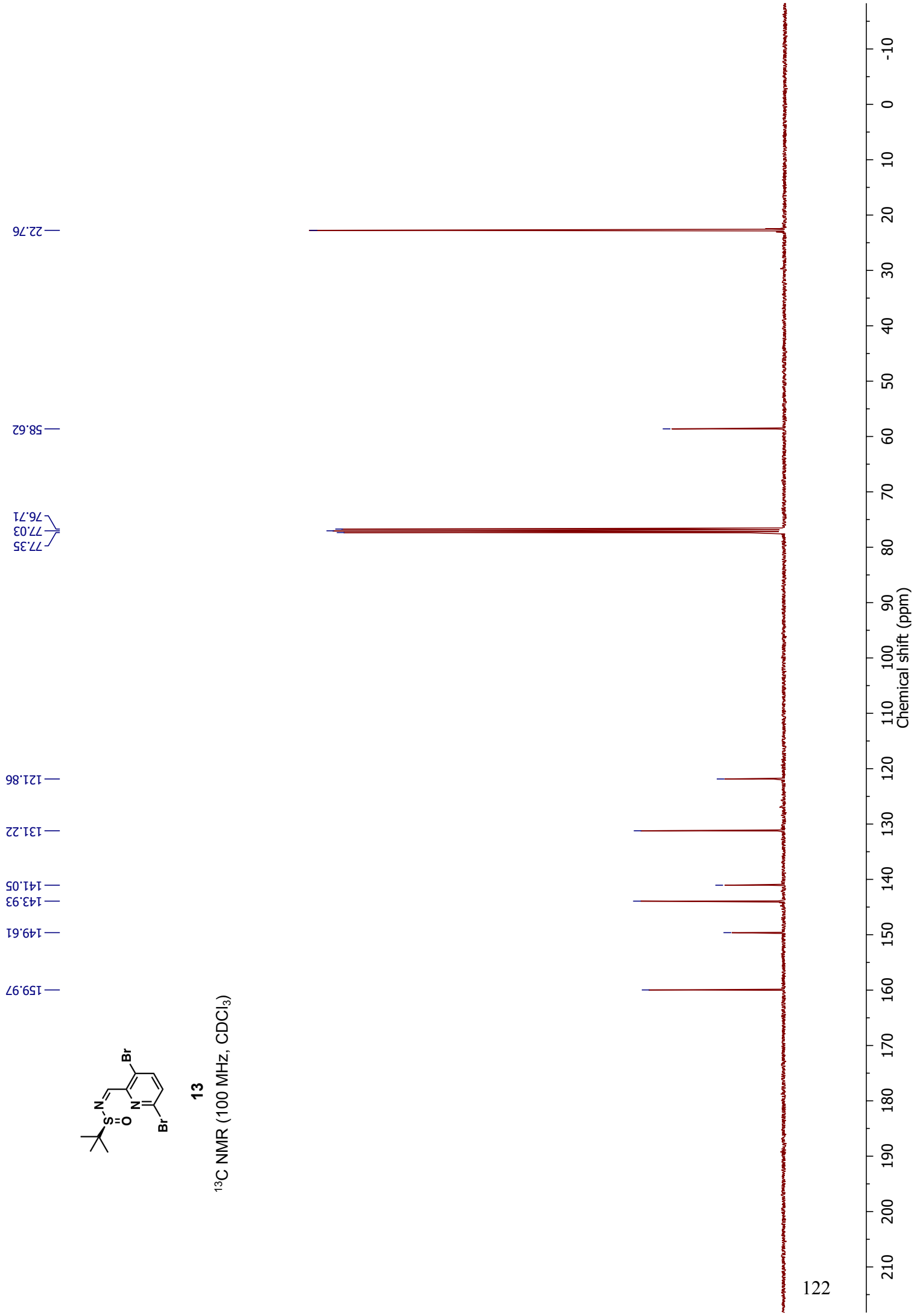
¹H NMR (400 MHz, CDCl₃)

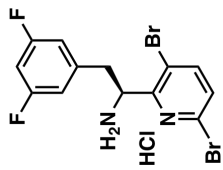




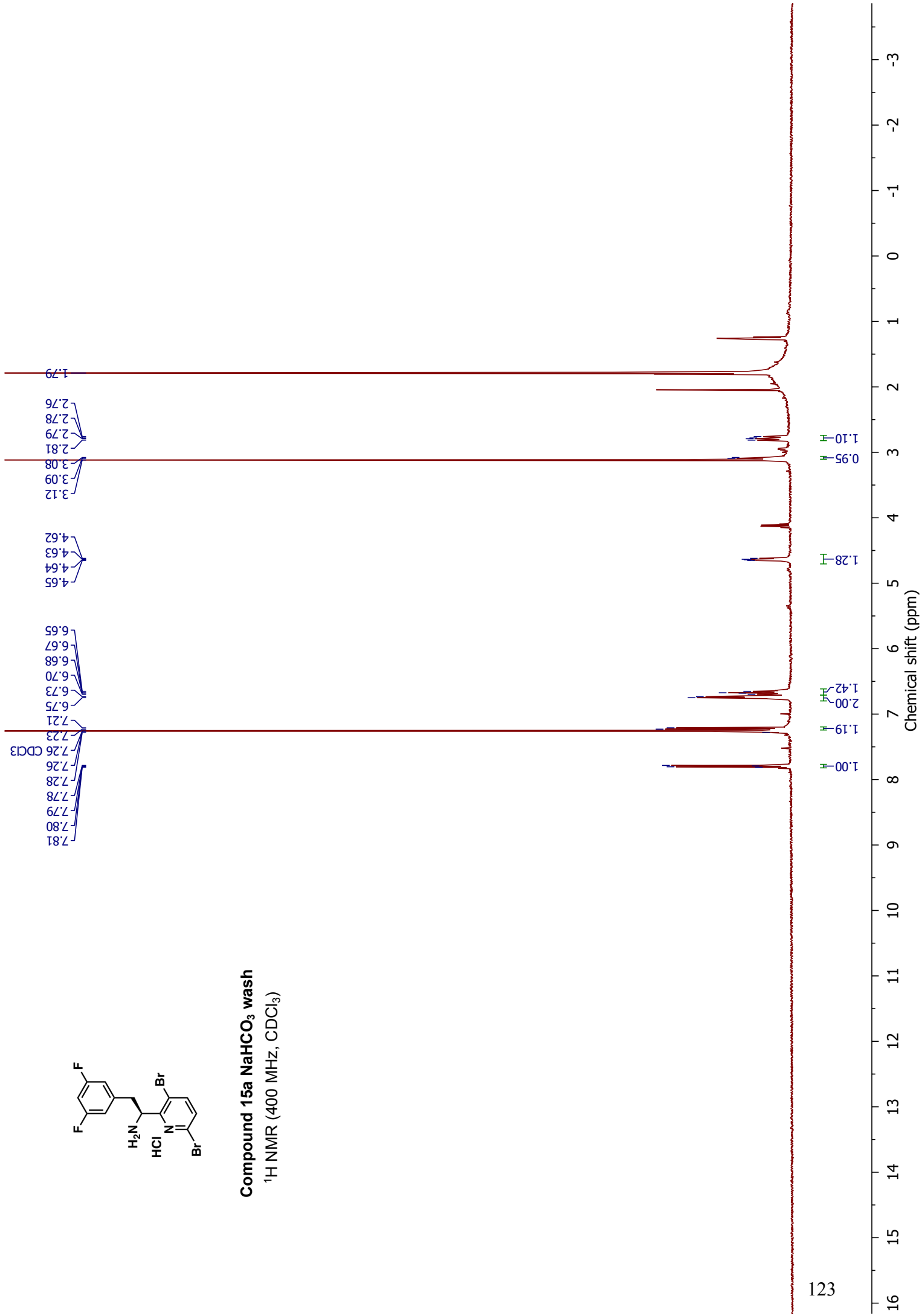
13

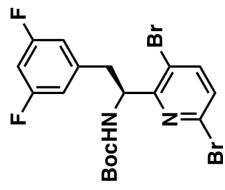
¹³C NMR (100 MHz, CDCl₃)





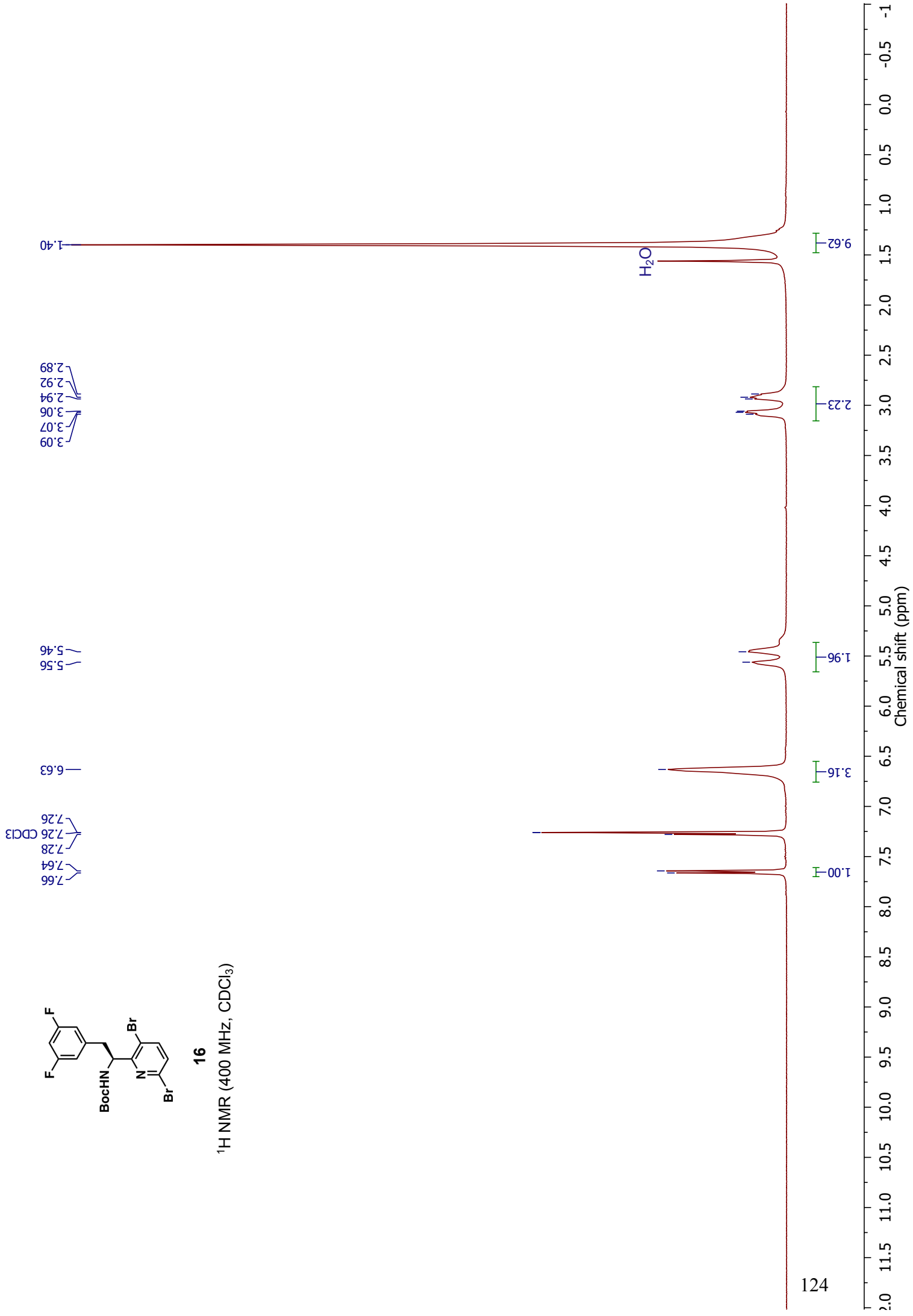
Compound 15a NaHCO₃ wash
¹H NMR (400 MHz, CDCl₃)

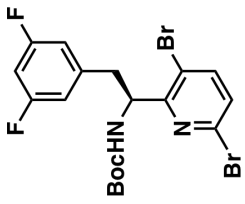




16

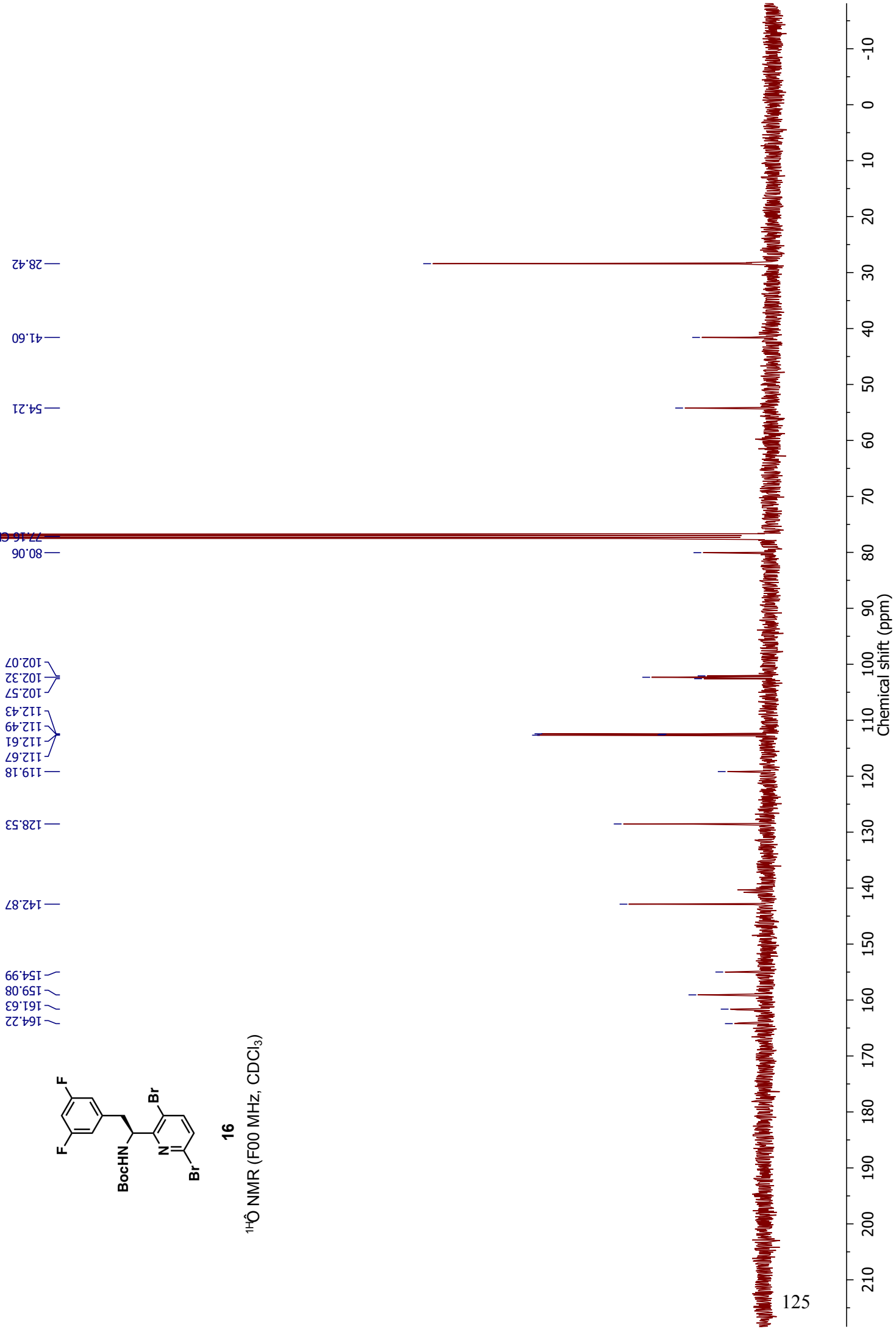
¹H NMR (400 MHz, CDCl₃)

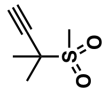




16

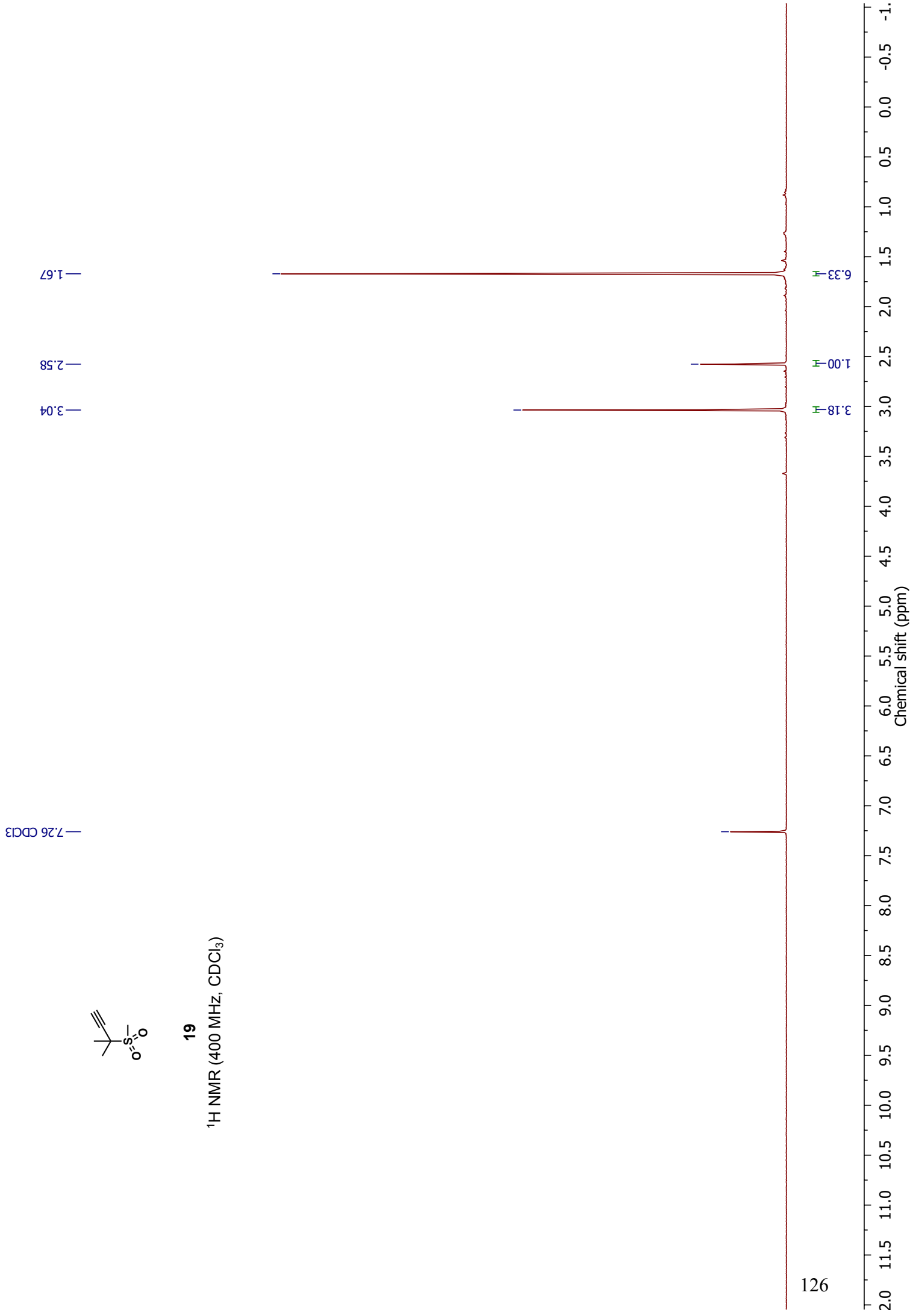
¹H NMR (F00 MHz, CDCl₃)

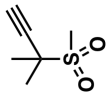




19

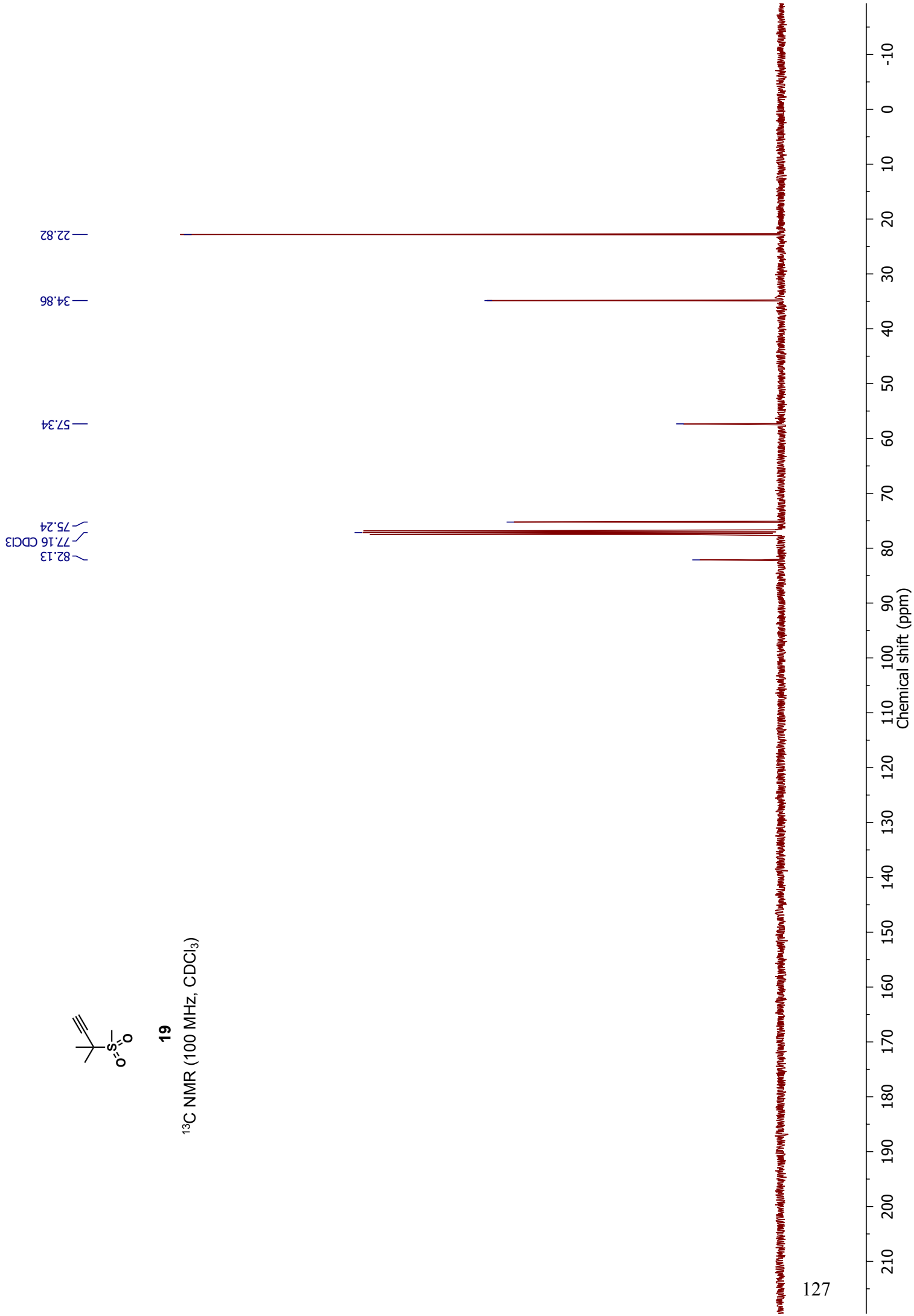
¹H NMR (400 MHz, CDCl₃)

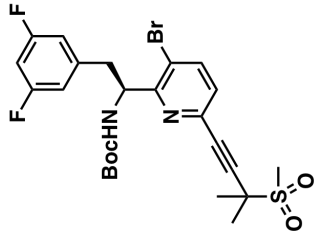




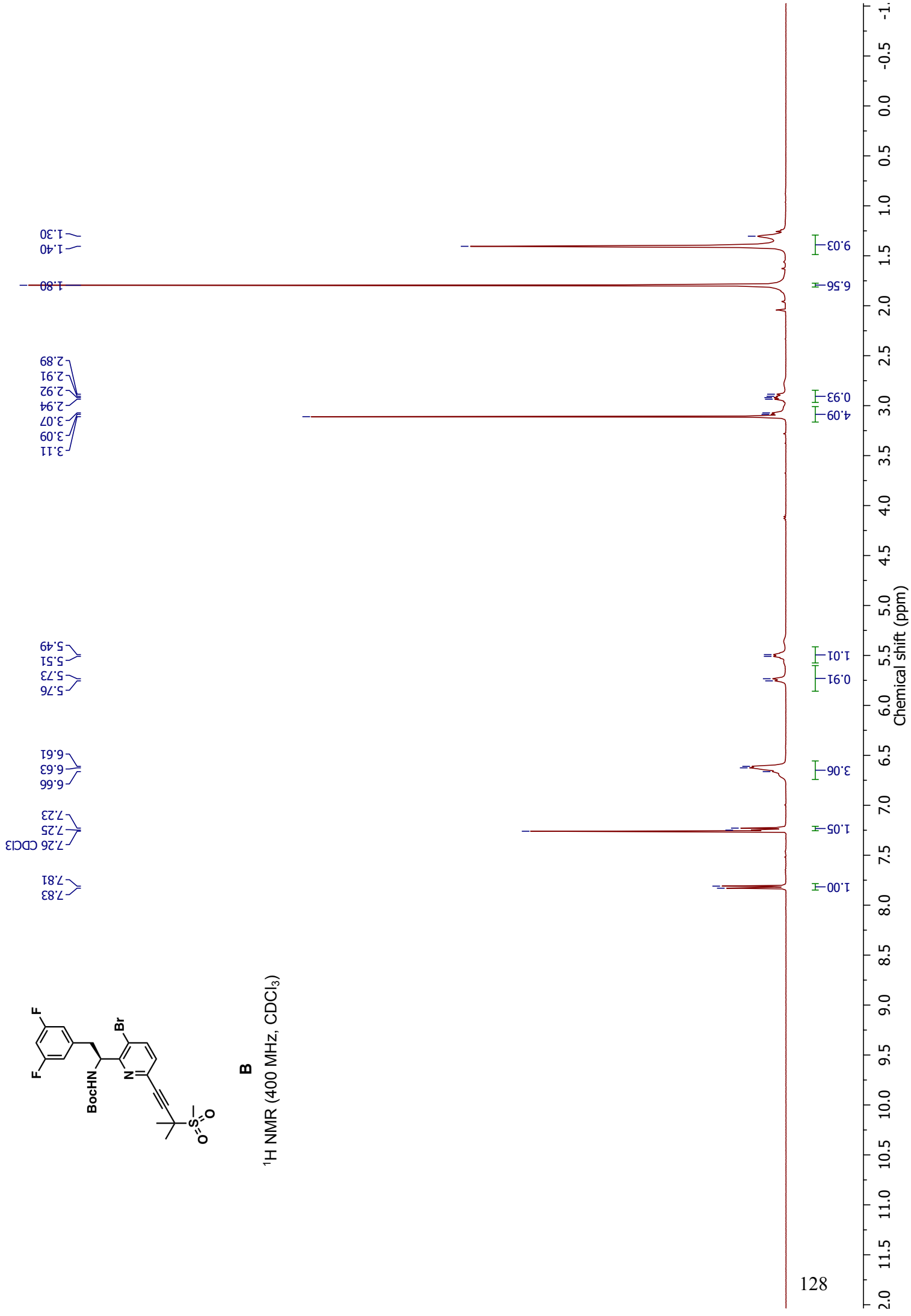
19

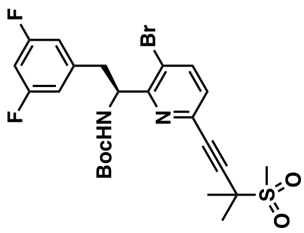
¹³C NMR (100 MHz, CDCl₃)



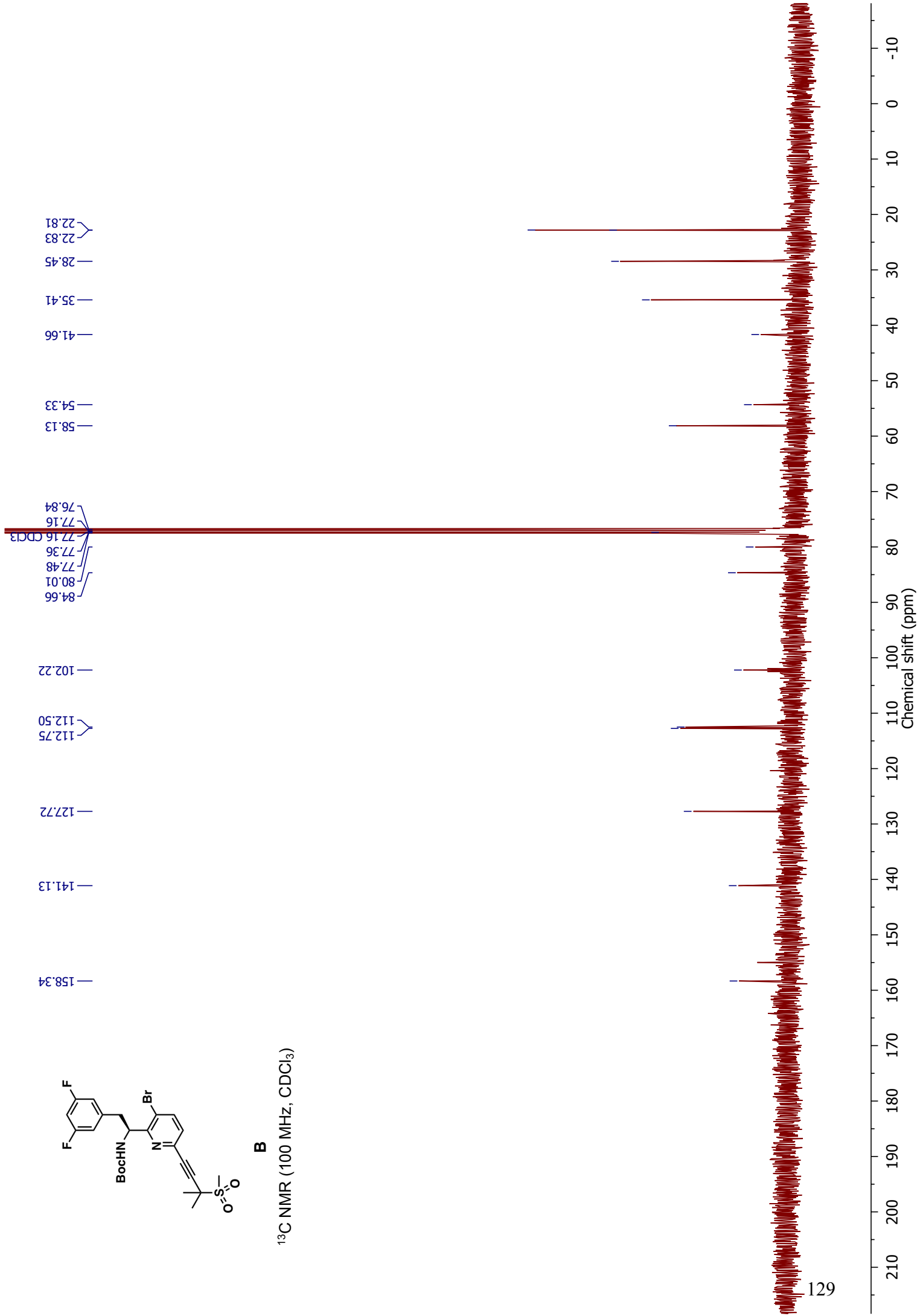


B
¹H NMR (400 MHz, CDCl₃)



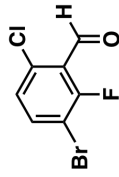


B
¹³C NMR (100 MHz, CDCl₃)



7.18
7.19
7.21
7.21
7.26
7.67
7.69
7.69
7.71

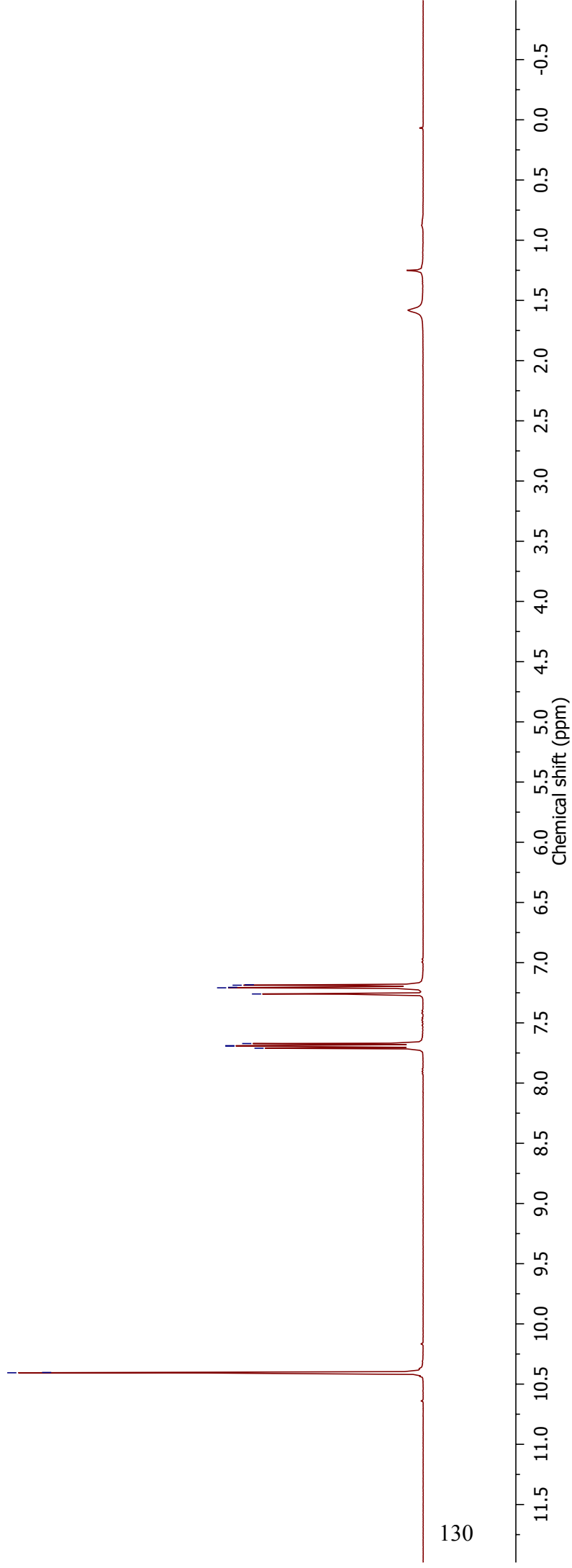
CDCl₃



3-Bromo-6-chloro-2-fluorobenzaldehyde (20a)

¹H NMR (400 MHz, CDCl₃)

10.41
10.40



186.05

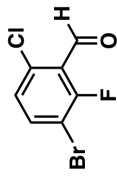
160.99
158.36

138.16
136.12

127.64
127.60
122.88
122.77

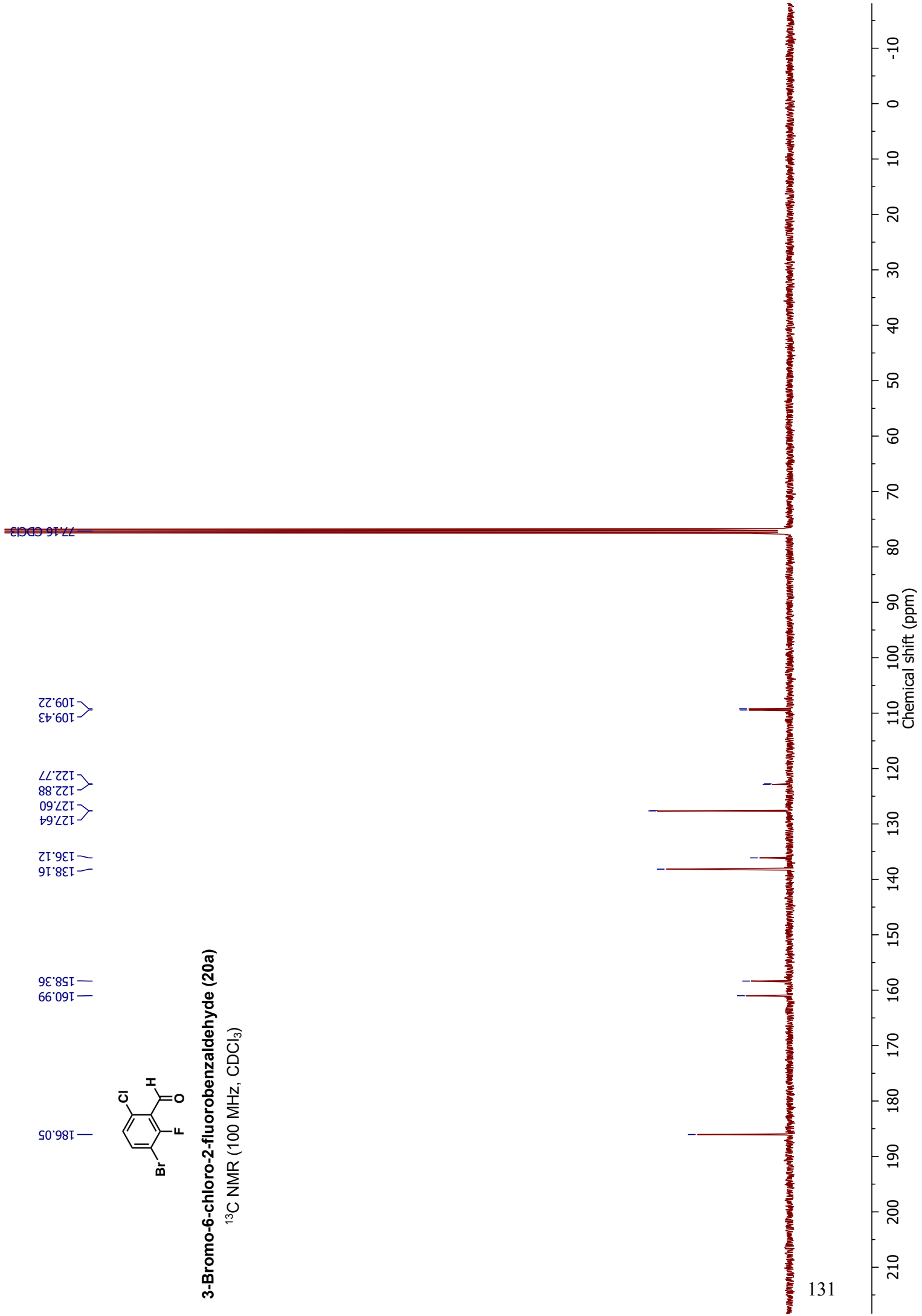
109.43
109.22

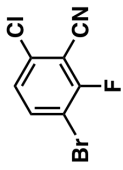
77.16 CDCl₃



3-Bromo-6-chloro-2-fluorobenzaldehyde (20a)

¹³C NMR (100 MHz, CDCl₃)

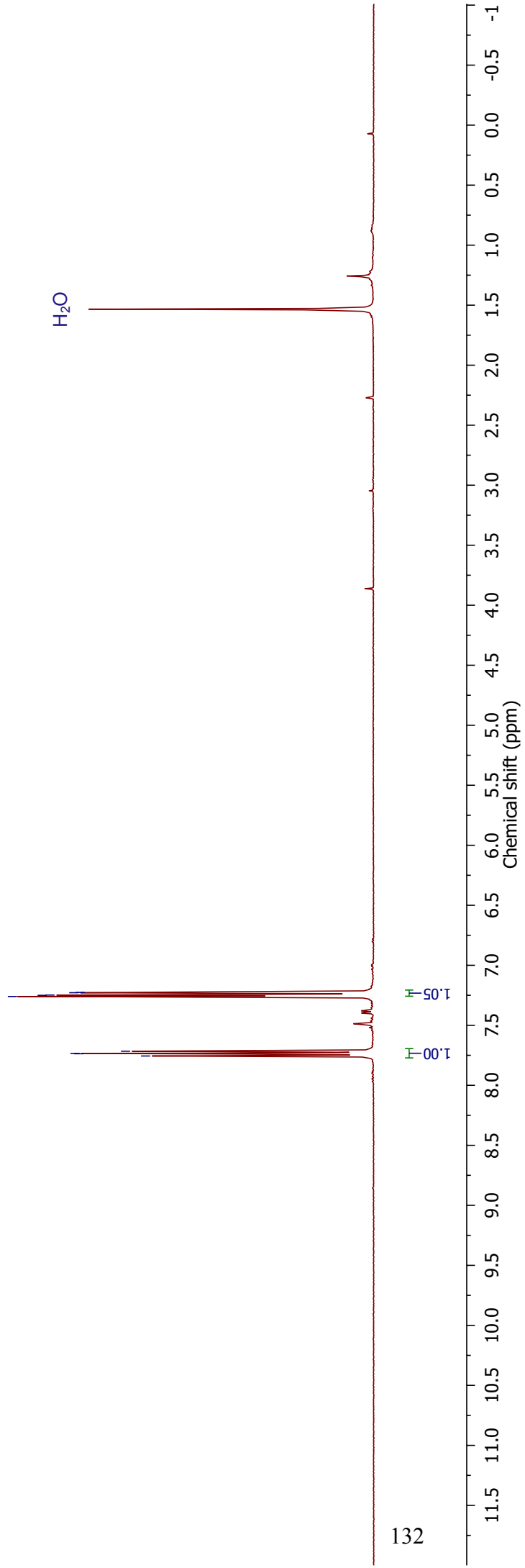


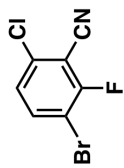


21

¹H NMR (400 MHz, CDCl₃)

7.76
7.74
7.73
7.72
7.26
7.25
7.25
7.23
7.23
CDCl₃

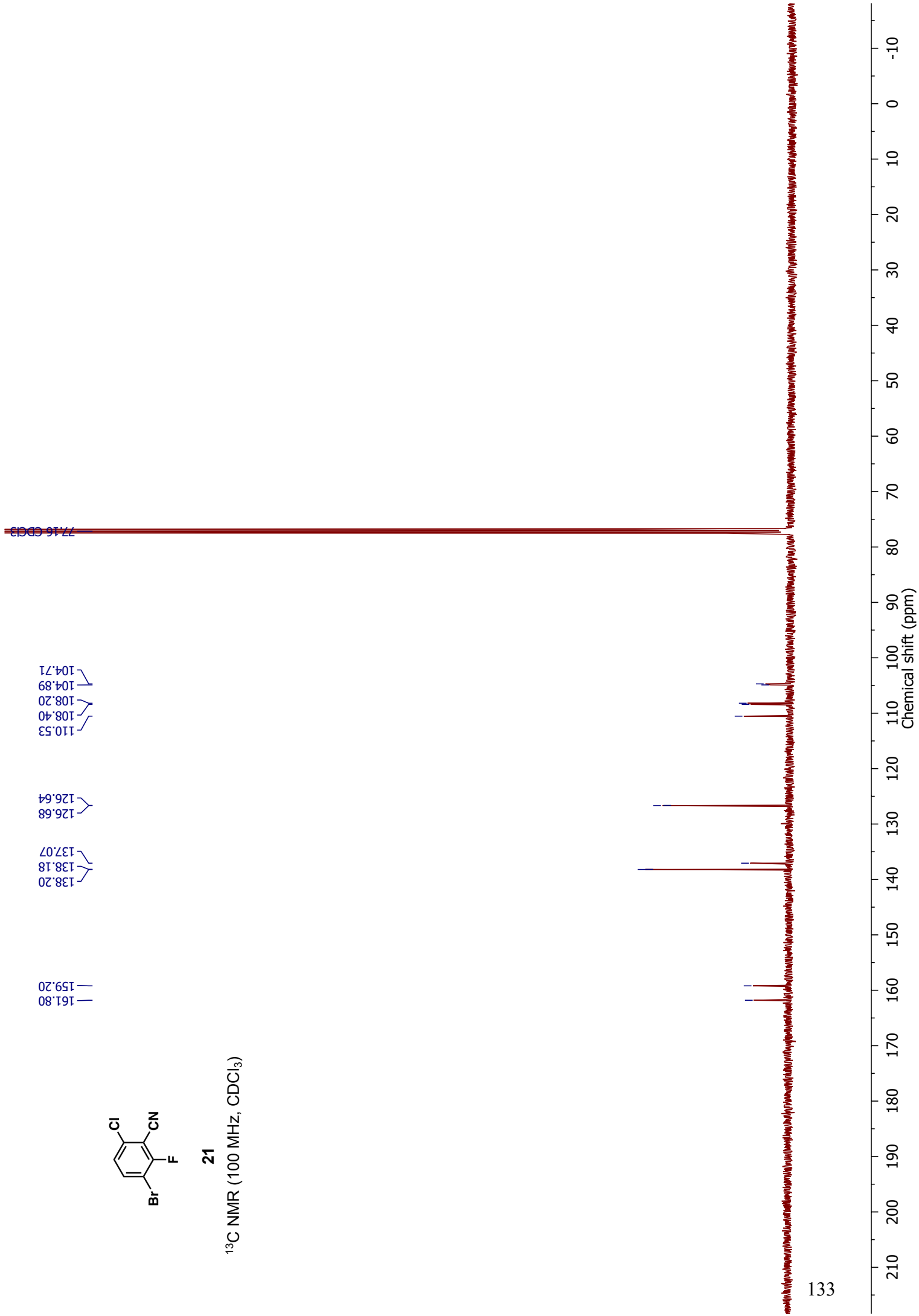


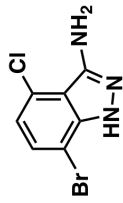


21

^{13}C NMR (100 MHz, CDCl_3)

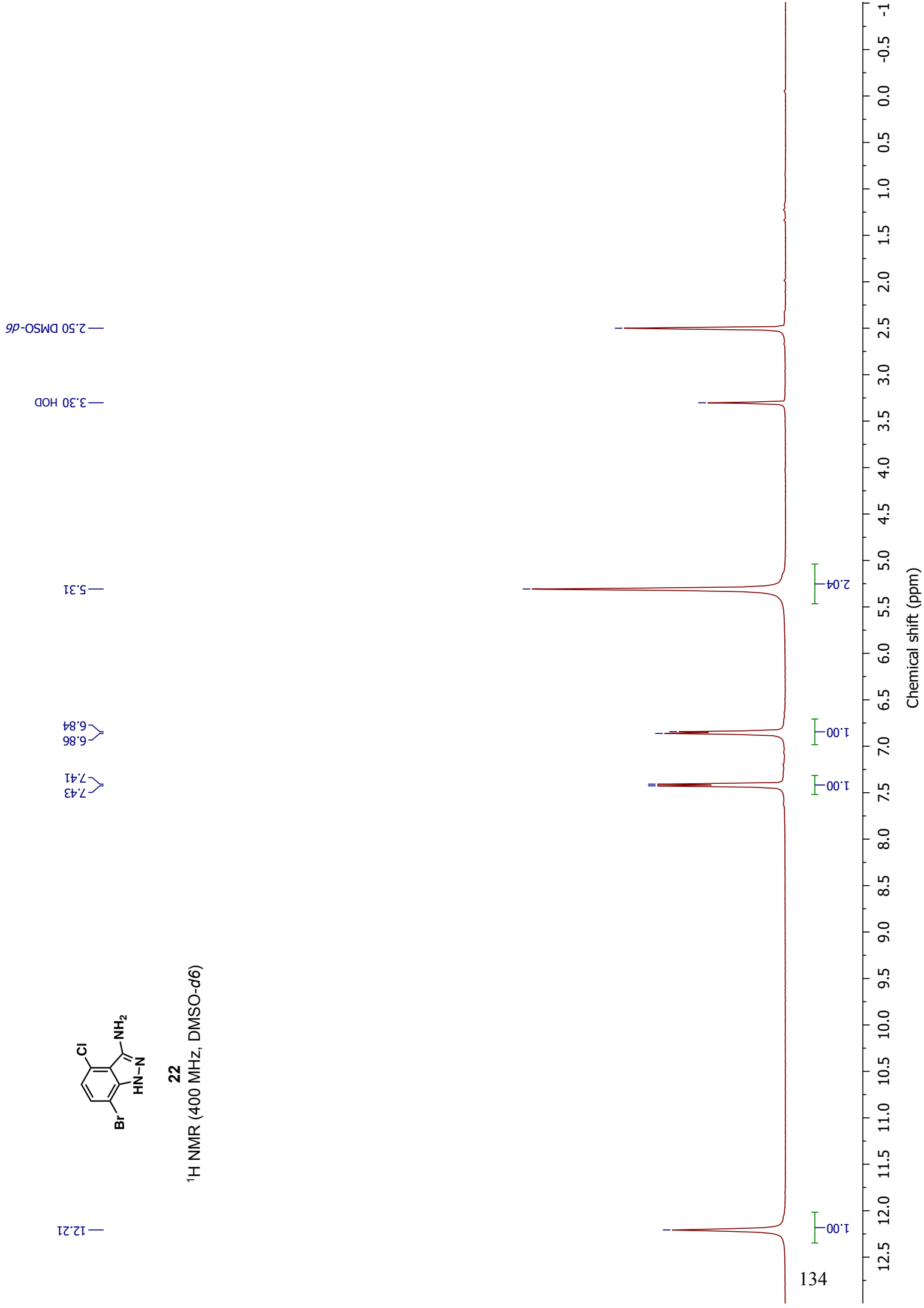
159.20
161.80
138.20
138.18
137.07
126.68
110.53
108.40
108.20
104.89
104.71

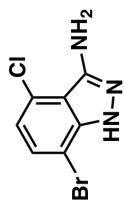




22

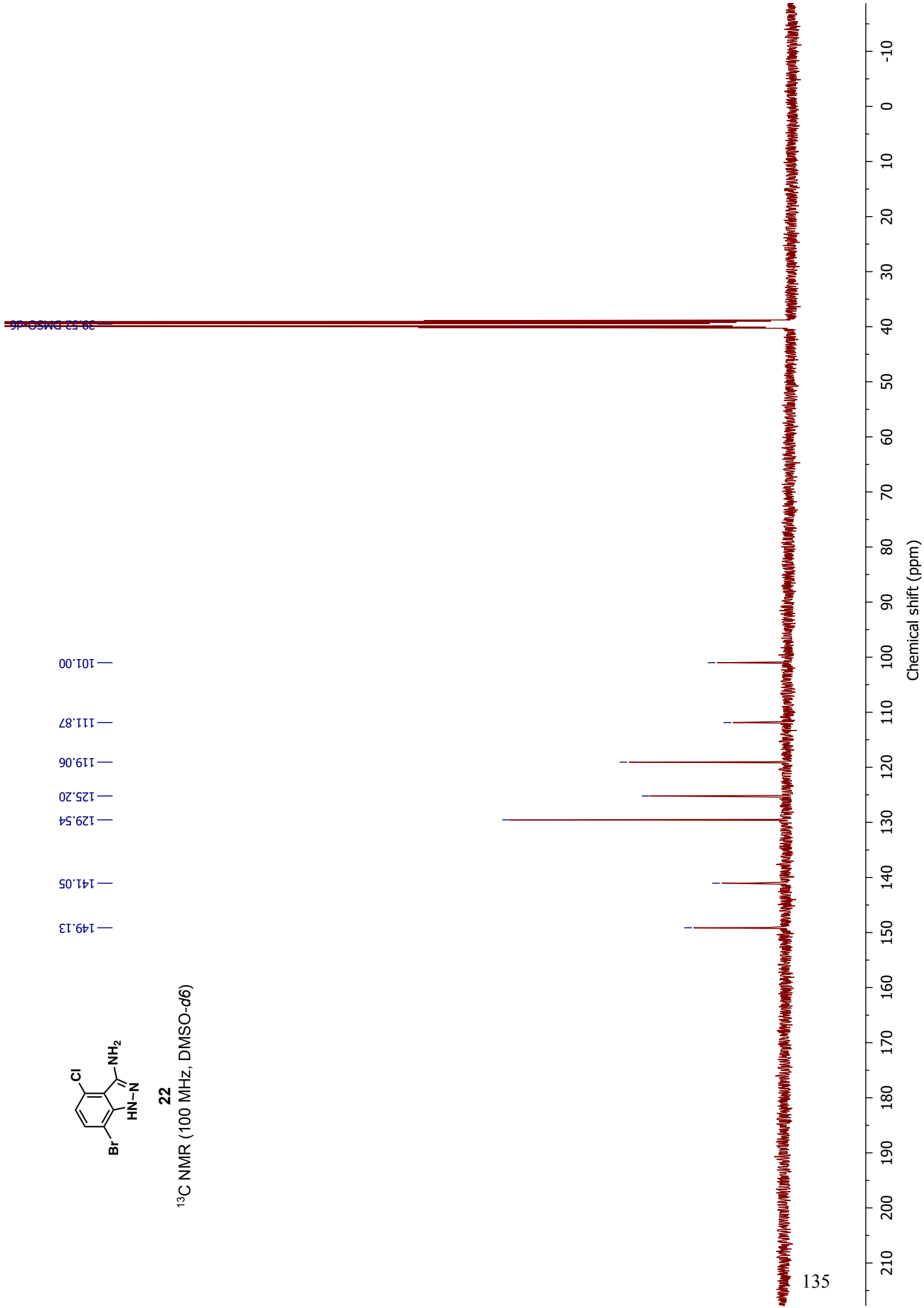
¹H NMR (400 MHz, DMSO-d₆)

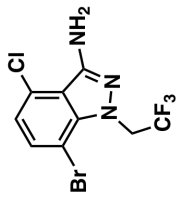




22

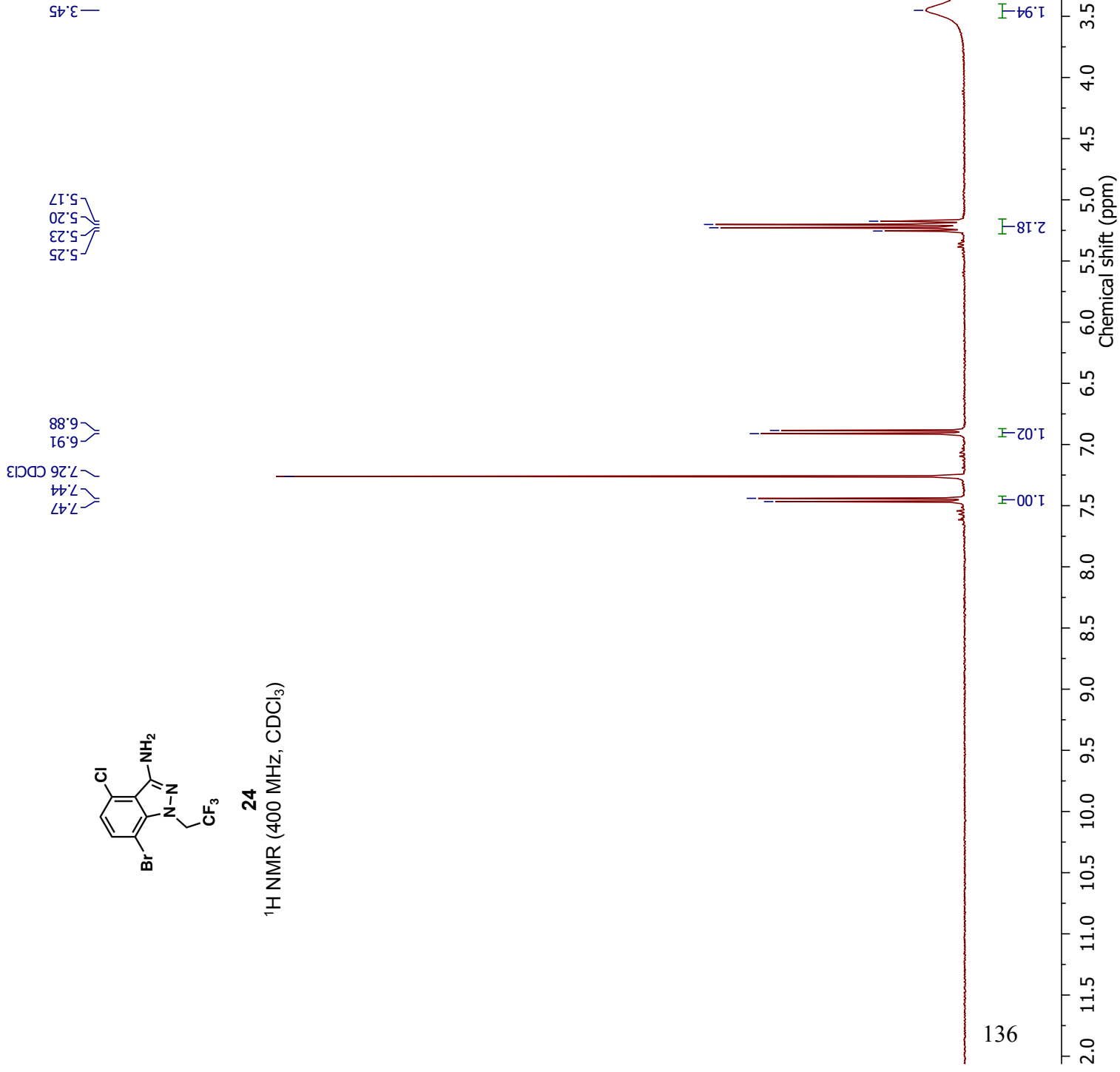
^{13}C NMR (100 MHz, DMSO- d_6)

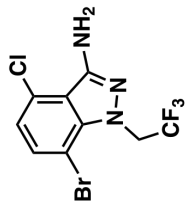




24

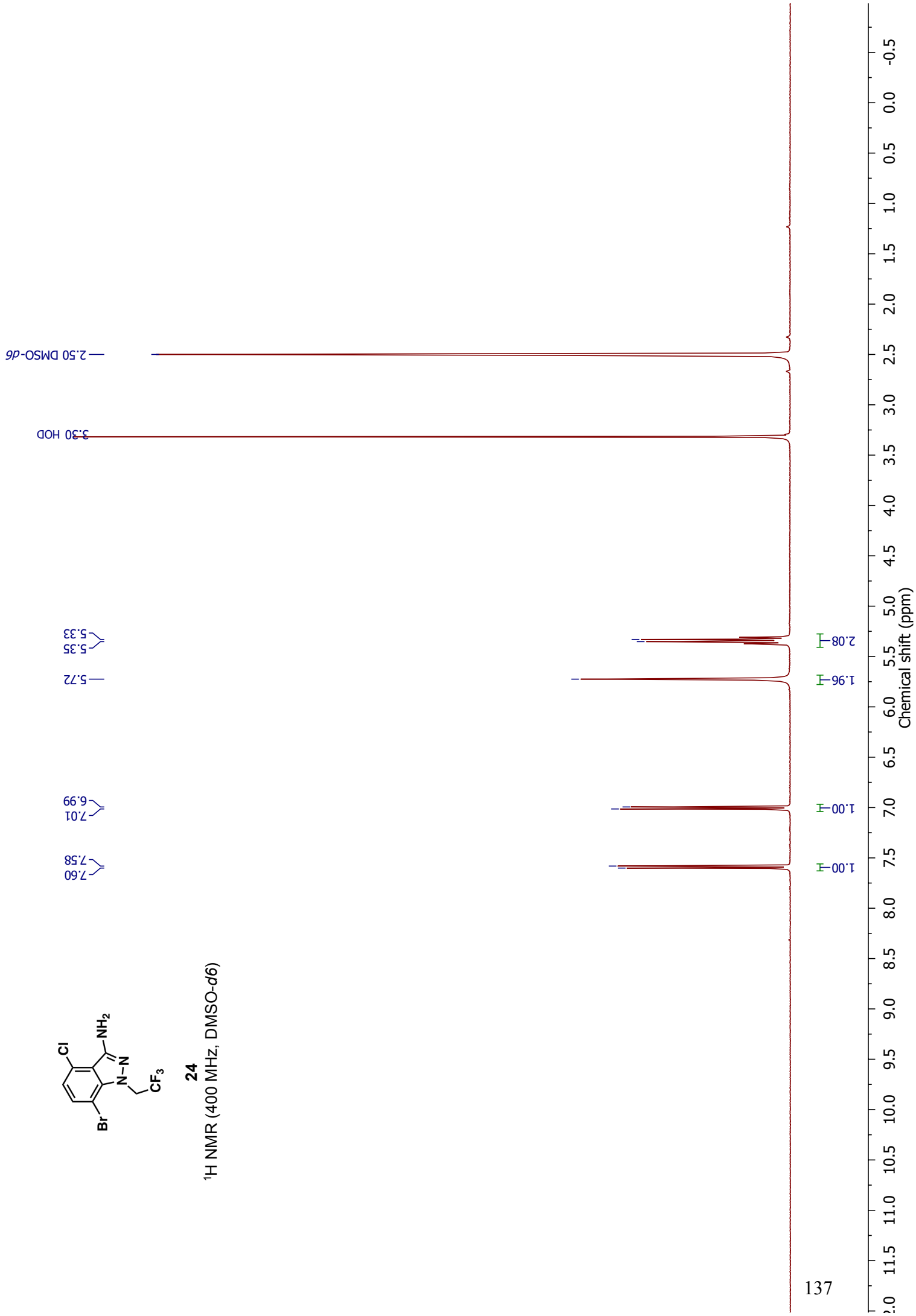
¹H NMR (400 MHz, CDCl₃)

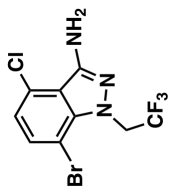




24

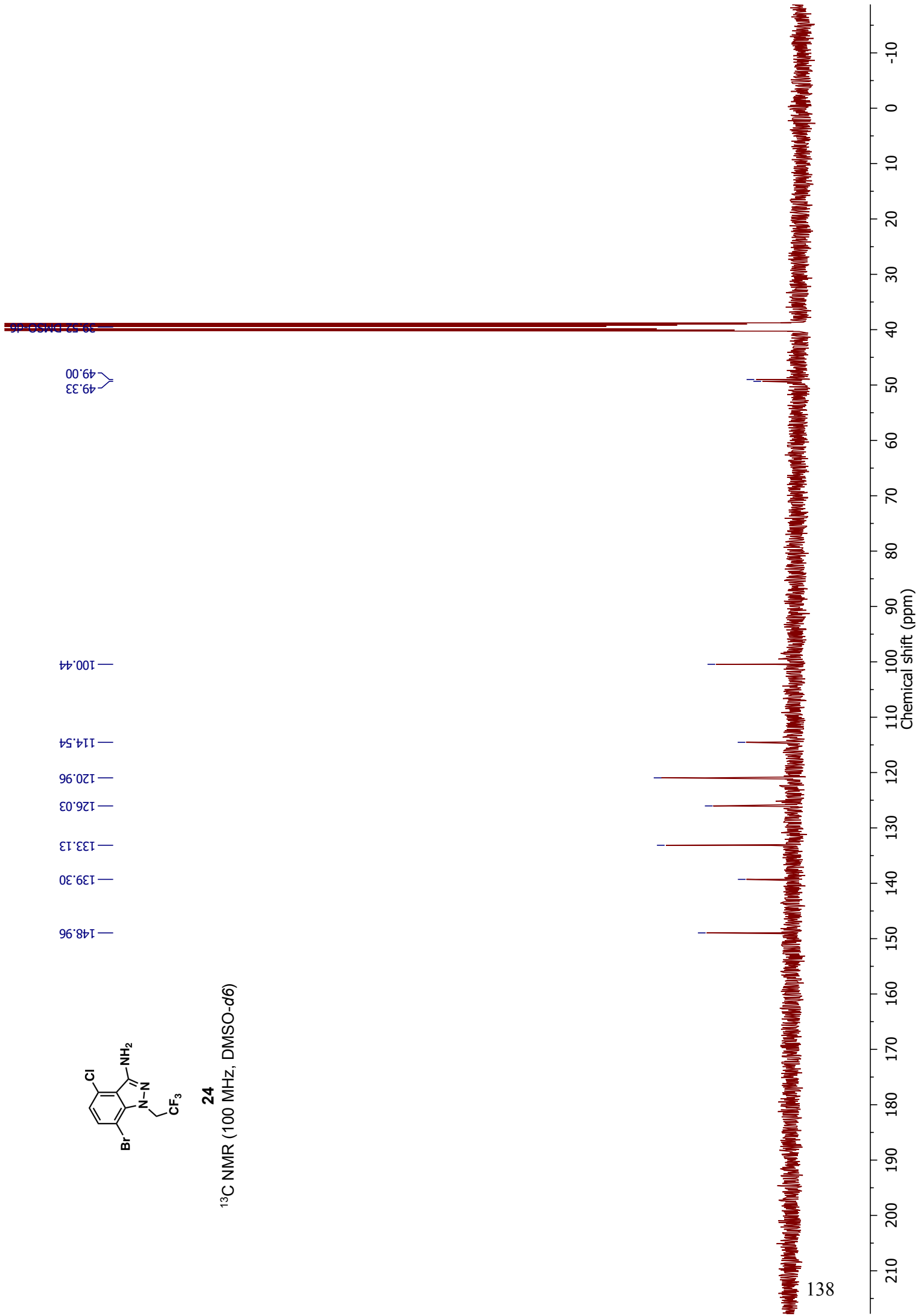
¹H NMR (400 MHz, DMSO-d₆)

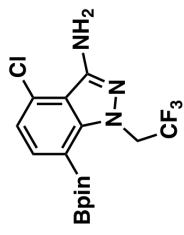




24

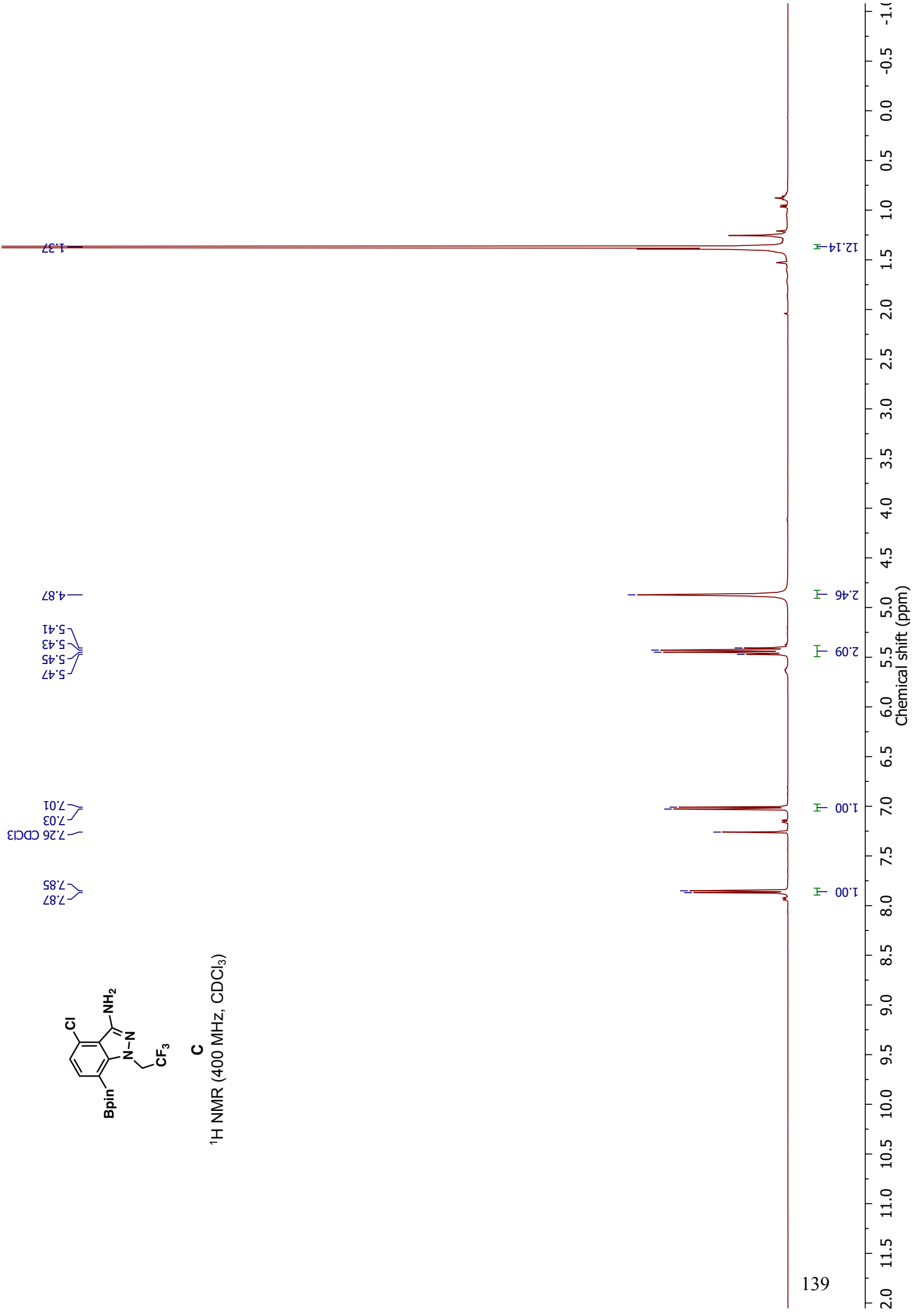
¹³C NMR (100 MHz, DMSO-*d*₆)

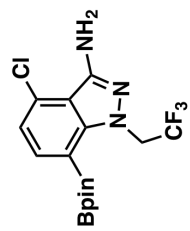




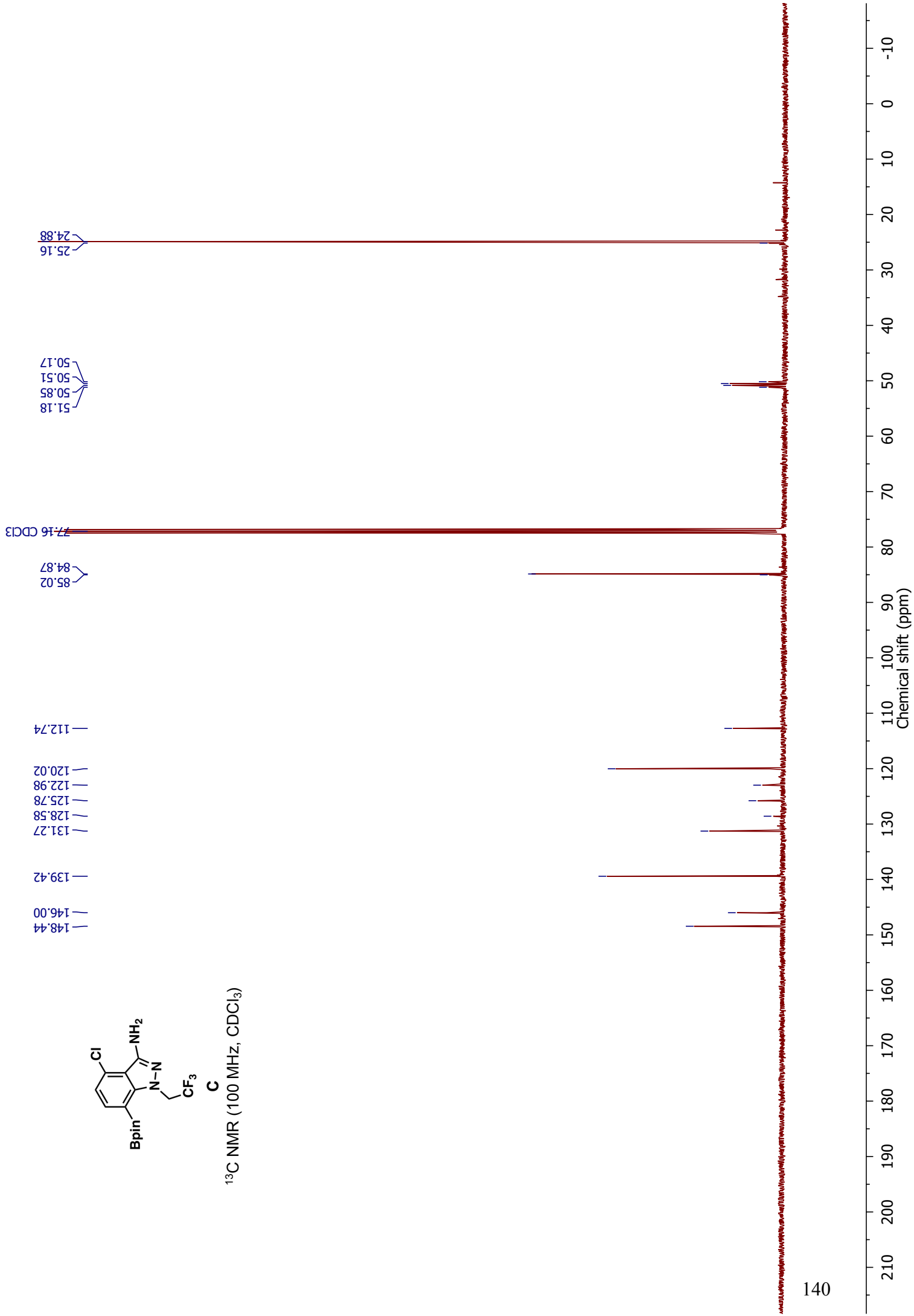
C

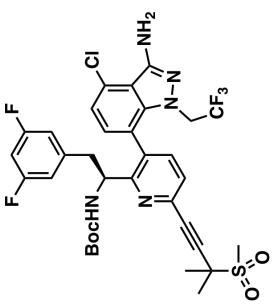
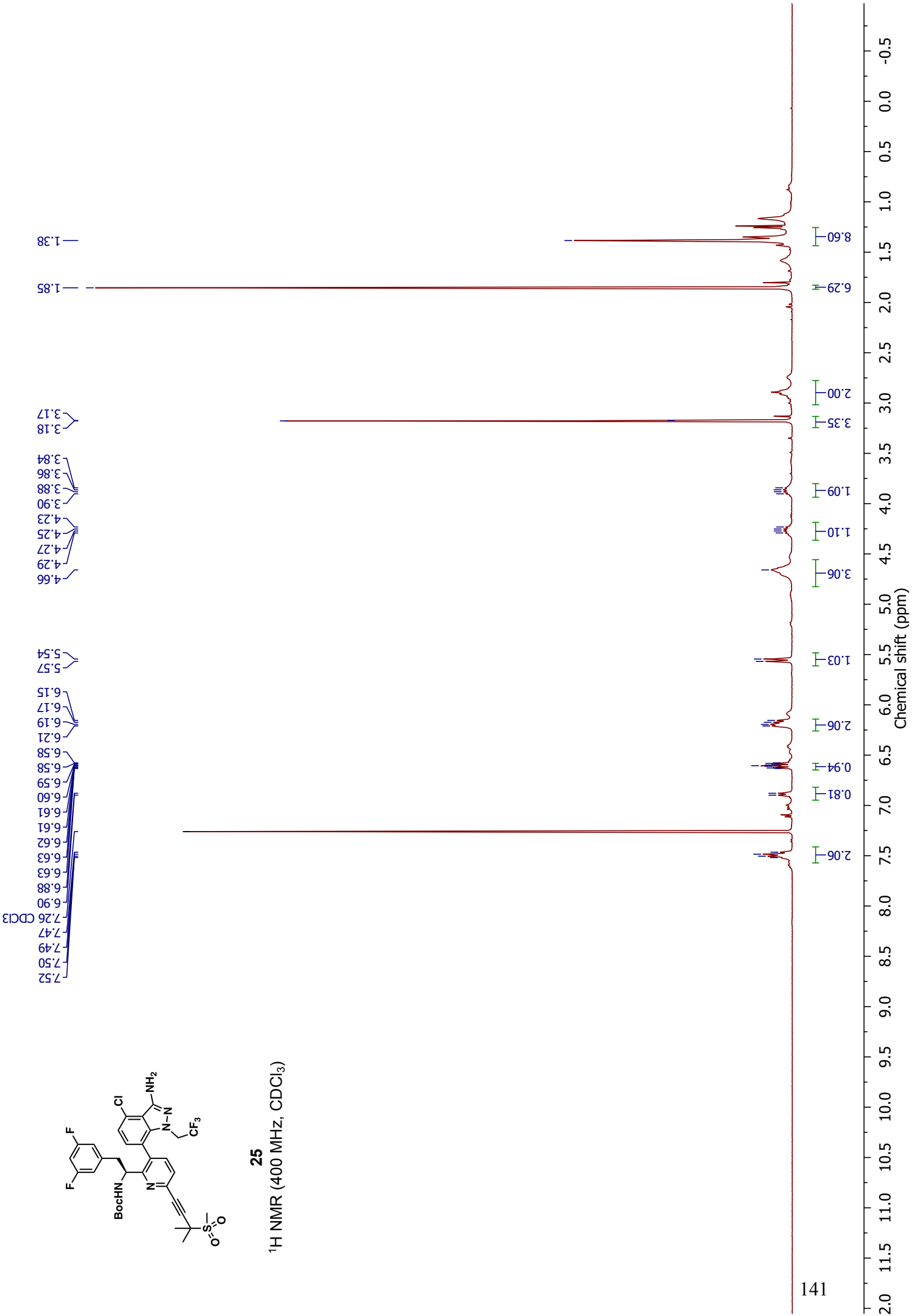
¹H NMR (400 MHz, CDCl₃)



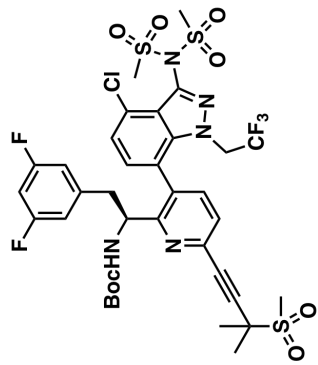


¹³C NMR (100 MHz, CDCl₃)



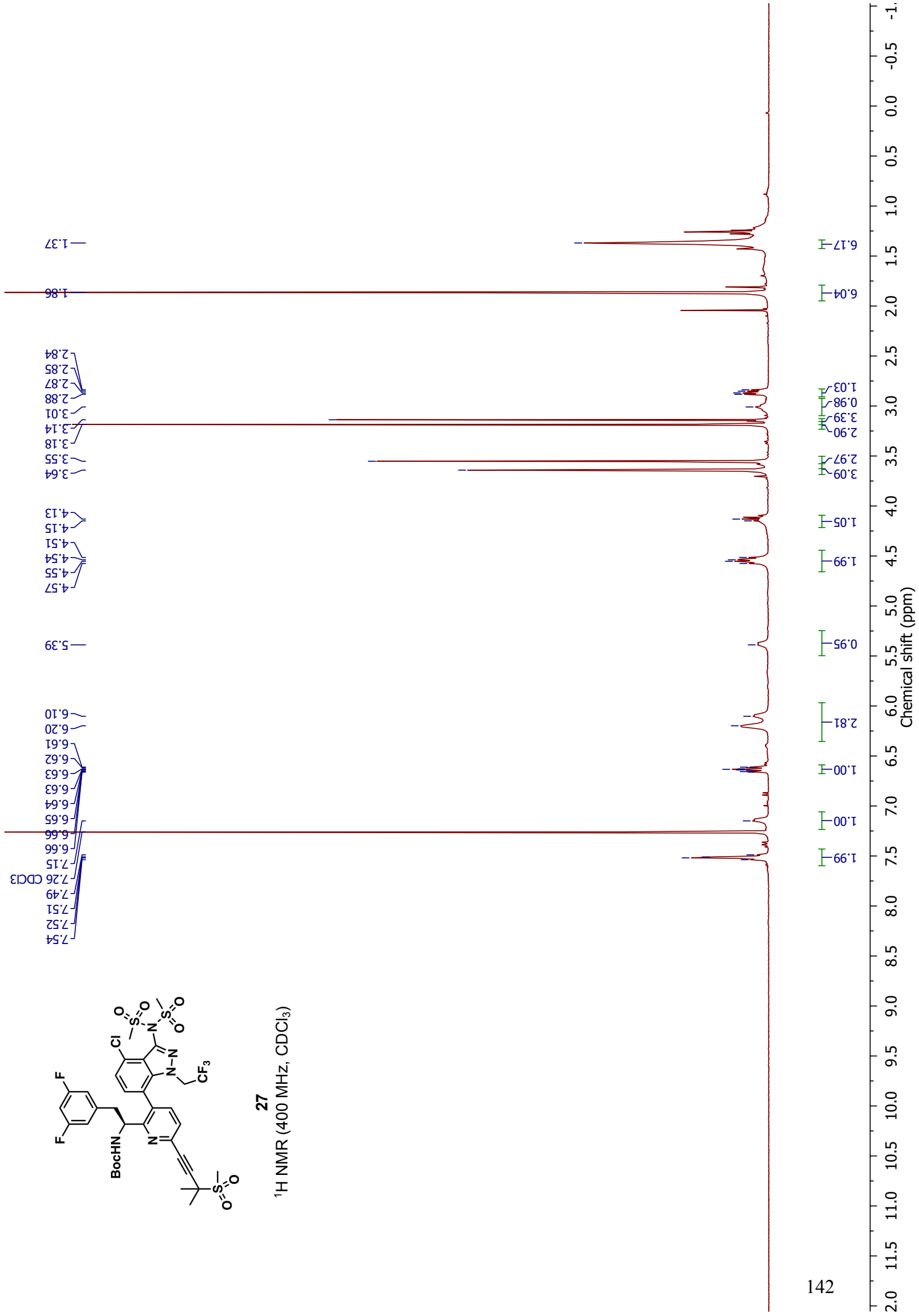


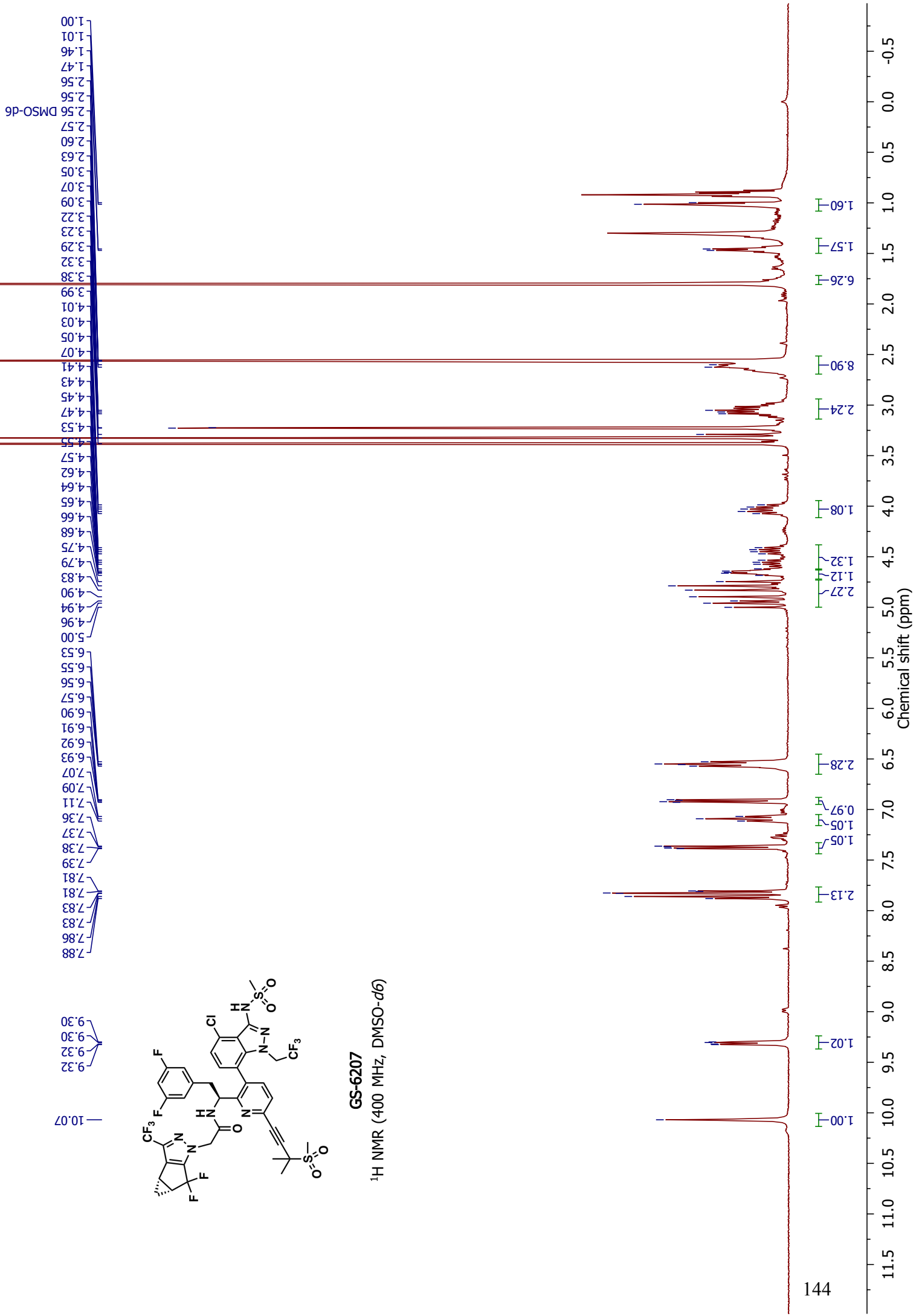
25
¹H NMR (400 MHz, CDCl₃)

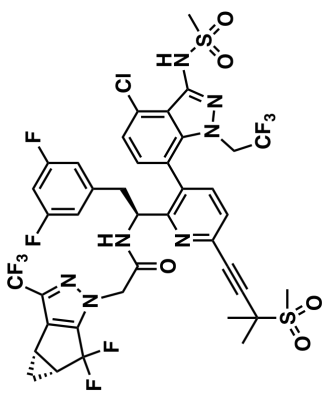


27

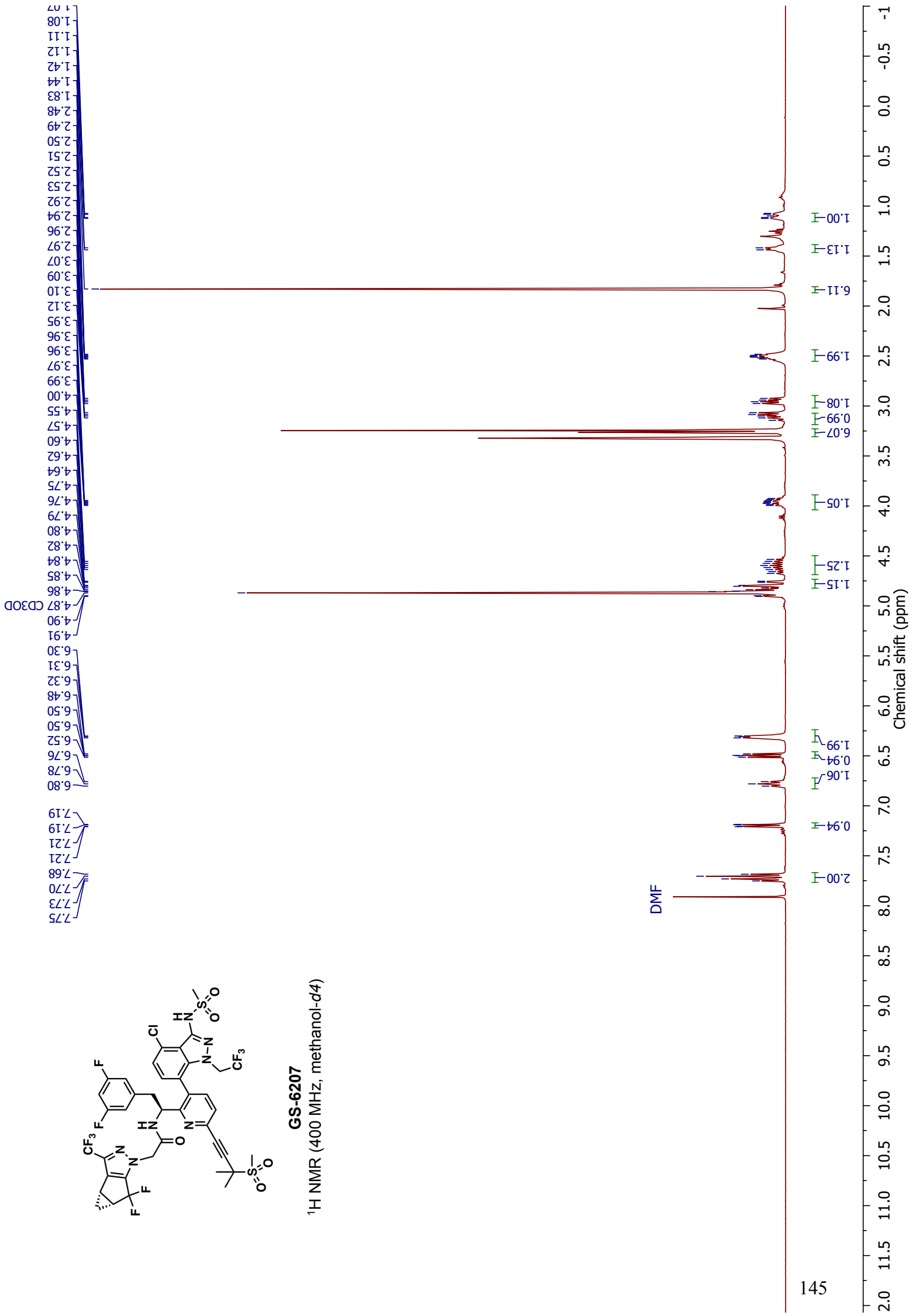
¹H NMR (400 MHz, CDCl₃)

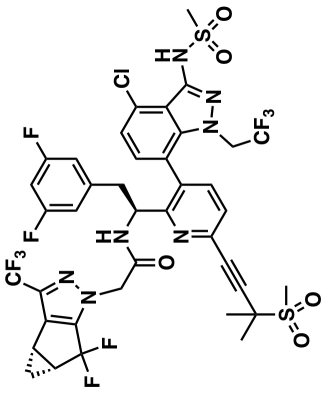






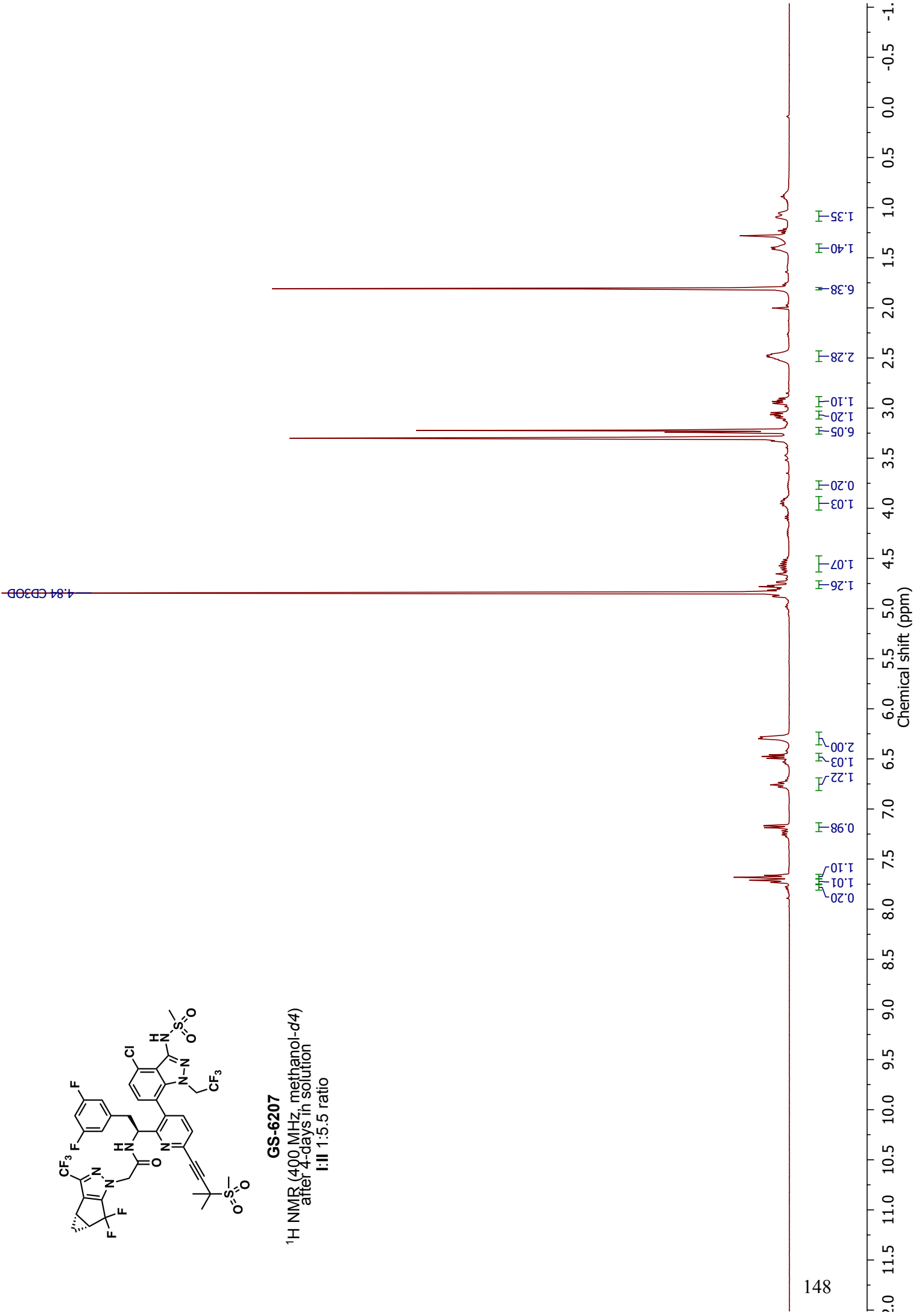
GS-6207
¹H NMR (400 MHz, methanol-d₄)

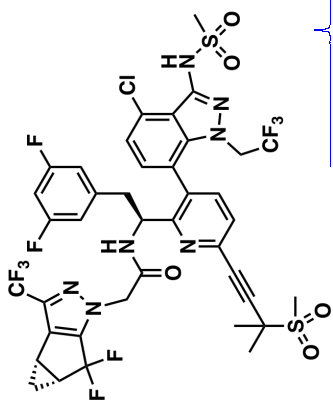




GS-6207

¹H NMR (400 MHz, methanol-d₄)
after 4-days in solution
I:II 1:5.5 ratio





GS-6207
 HSQC (400 MHz, DMSO-*d*₆)

