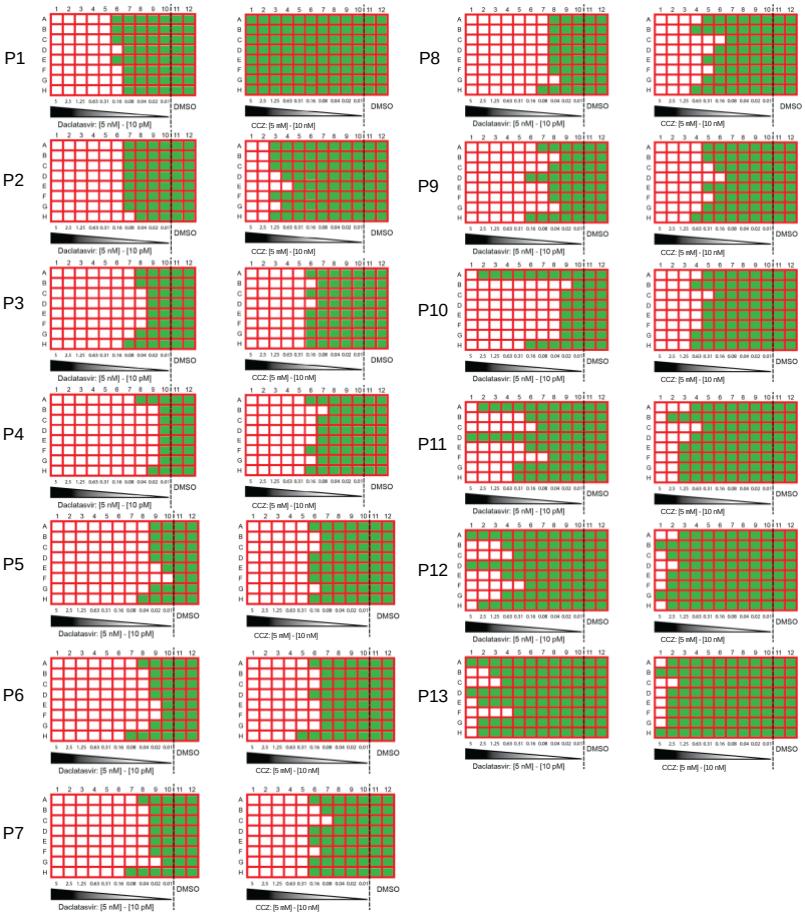
SUPPLEMENTAL INFORMATION

Chlorcyclizine Inhibits Viral Fusion of Hepatitis C Virus Entry by Directly Targeting HCV Envelope Glycoprotein 1

Zongyi Hu, Adam Rolt, Xin Hu, Christopher D. Ma, Derek J. Le, Seung Bum Park, Michael Houghton, Noel Southall, D. Eric Anderson, Daniel C. Talley, John R. Lloyd, Juan C. Marugan, and T. Jake Liang

A


P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13
DMSO
Daclatasvir: [5 nM] - [10 pM]
CCZ: [5 nM] - [10 nM]

B

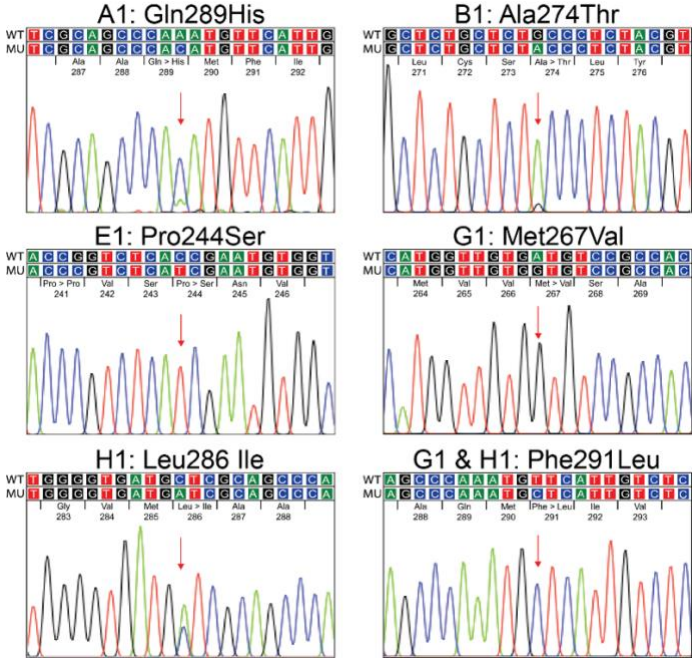

A1: Gln289His
B1: Ala274Thr
E1: Pro244Ser
G1: Met267Val
H1: Leu286 Ile
G1 & H1: Phe291Leu
WT
MU

Figure S1. Selection and detection of (*S*)-CCZ- and daclatasvir-resistant HCV RASs. Related to Figure 1

A, Generation of (*S*)-CCZ RASs. A concentration gradient of (*S*)-CCZ ranging from 10 nM to 5 μM or daclatasvir from 10 pM to 5 nM in 2-fold increments was used to select (*S*)-CCZ RASs. For the first passage (P1), Huh7.5.1 cells were infected first with HCV J6/JFH1 and then treated with various (*S*)-CCZ concentrations. For each passage, the infection was quantified by HCV core immunostaining and shown with color-coding (green: > 50% infection; white: < 50% infection). With extended passages, Huh7.5.1 cells became gradually infected at higher concentrations, indicative of a gradual emergence of HCV RASs. By the 13_{th} passage, most of the cells in column 1 (5 μM) or 2 (2.5 μM) showed infection, indicating the presence of a predominance of HCV (*S*)-CCZ-RASs. For daclatasvir, RASs were readily selected by passage 13 as well. **B,** Detection of putative (*S*)-CCZ RASs. DNA sequencing chromatograms show detection of each RAS. A1, B1, E1, G1, and H1 refer to five individual HCV clones that were analyzed in this study. F291L was detected in both G1 and H1 HCV clones in combination with M267V and L286I, respectively. WT: wild-type J6/JFH1 nucleotide sequence, MU: mutant nucleotide sequence.

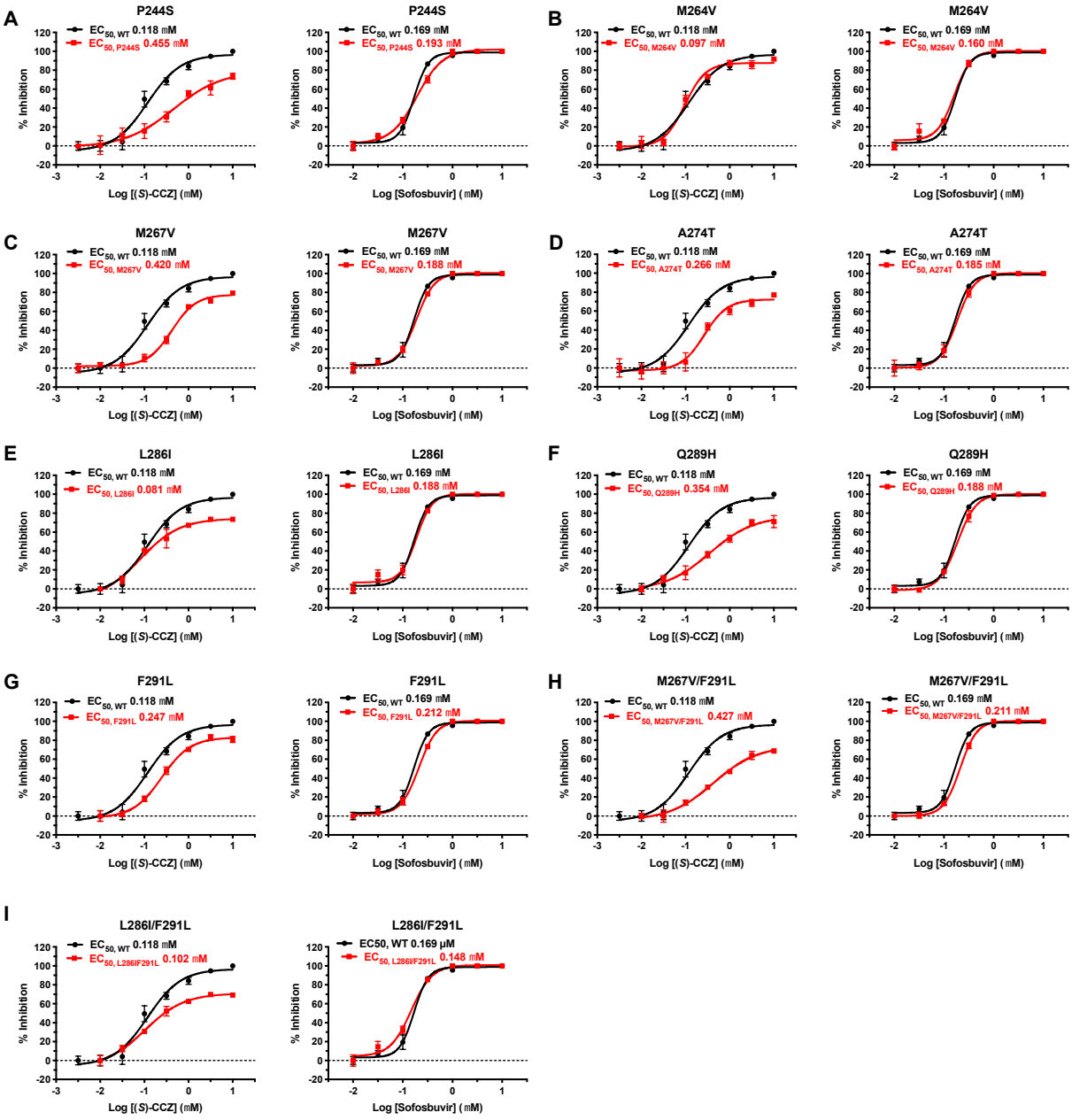


Figure S2. Dose response curves of (*S*)-CCZ and sofosbuvir against HCV-WT and HCV E1 mutants. Related to Figure 1.

Huh7.5.1 cells were infected with HCV-WT (black curve) or mutant HCV (red curve) and treated with (*S*)-CCZ and sofosbuvir at ½ log incremental doses ranging from 0.003-10 µM. After 72 h of incubation, HCV infection was detected by HCV core immunofluorescent staining. Positive foci (≥ 4 stained cells in each group) were counted and the data were normalized to the DMSO control. The computer software program, Prism 7, was used to plot the dose-response curves and calculation of EC_{50}. HCV-WT was used as a reference for each mutant virus. Data points were presented as mean values ± SEM (n=3). The results are representative of three separate experiments.

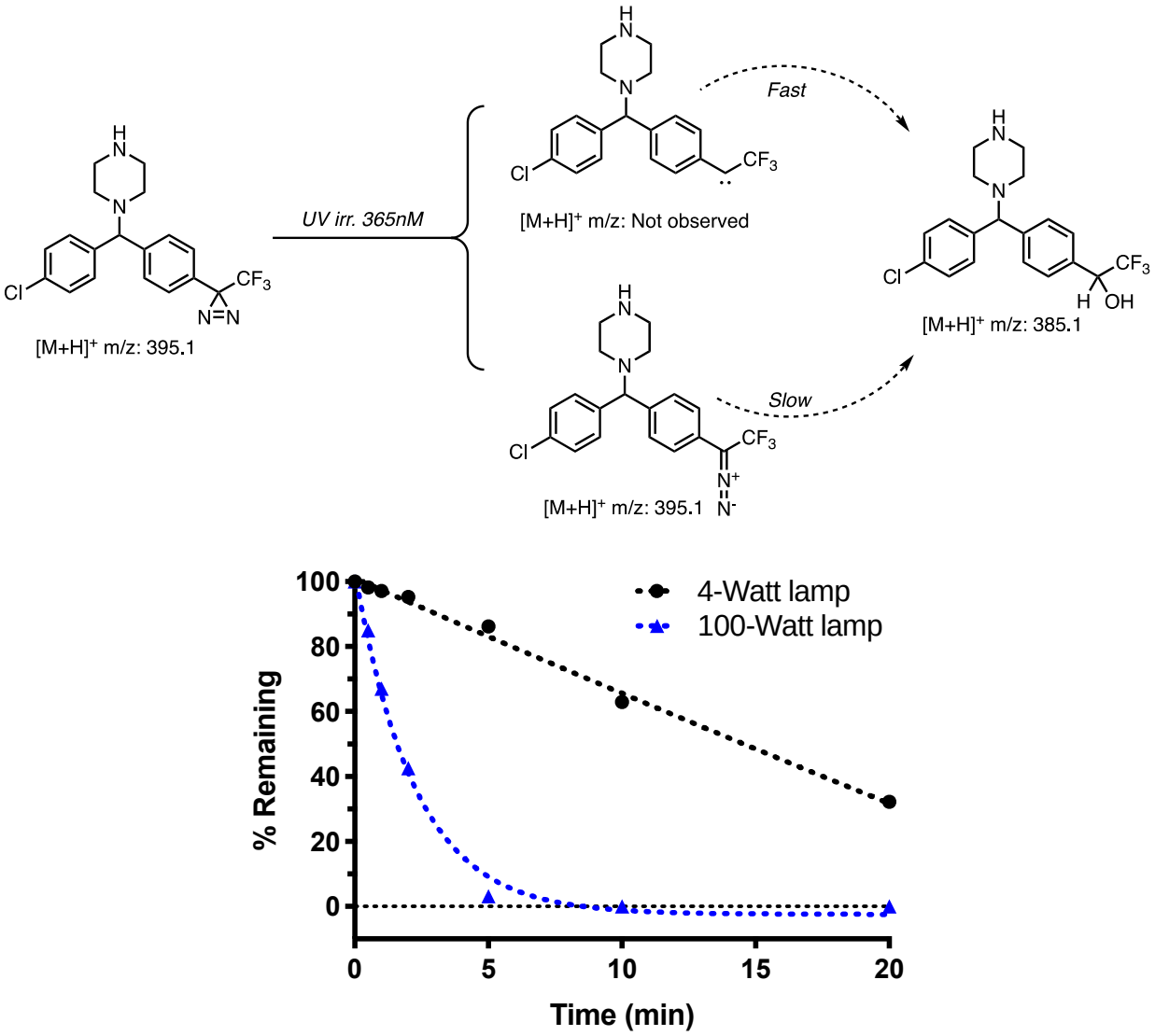
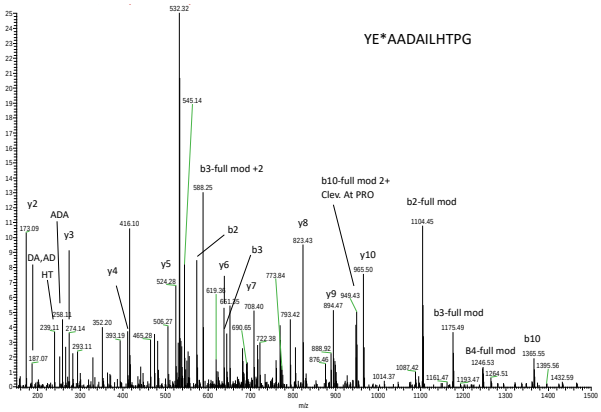
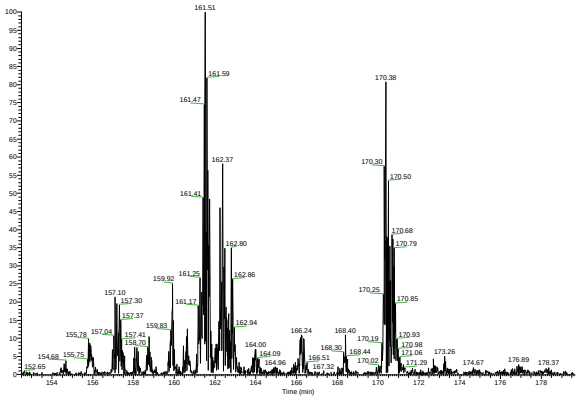


Figure S3. **Time course of CCZ-diazirine conversion under UV irradiation**. **Related to Figure 4.** CCZ-diazirine solution was exposed to UV irradiation with 100 W mercury lamp with a 365 nm bypass filter and monitored by mass spectrophotometry over time. Disappearance of CCZ-diazirine can be tracked by LCMS and underwent complete conversion to solvent insertion product in 5 min.

A


YE*AADAILHTPG
b3-full mod +2
b10-full mod 2+
Clev. At PRO
b2-full mod
b2
b3
y2
ADA
y3
DA,AD
HT
y4
y5
y6
y7
y8
y9
y10
b3-full mod
B4-full mod
b10
m/z

B


Time (min)

C

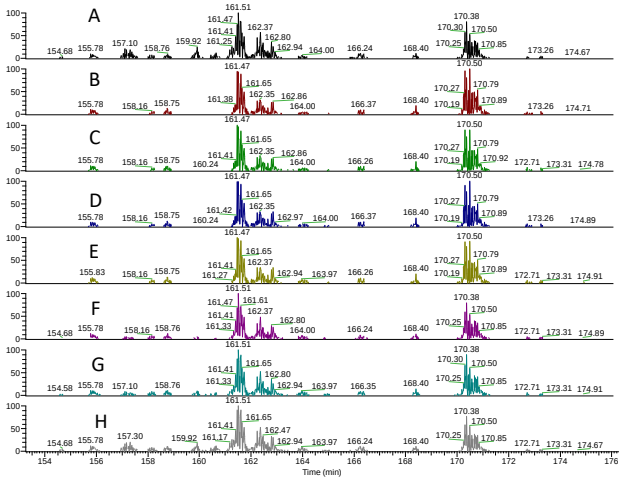

A
B
C
D
E
F
G
H
Time (min)

Figure S4. Analysis of the modified peptide YEAADAILHTPG. Related to Figure 5

A, A fragmentation spectrum of the modified peptide YE*AADAILHTPG. Many dominant peaks were labeled using an independent determination from protein prospector. The 416.10 peak is associated with modification of the first and possibly second residue but not others. **B,** The total ion current of the fragmentation spectra for the modified peptide YEAADAILHTPG. The ion current was collected periodically for M/Z 690.32. Nearly all features are thought to originate from the CCZ-biotin labeled peptide YEAADAILHTPG with various sites of modification giving rise to different elution times. The gradient used was 0.068%/min so the elution concentration difference is on the order of 1-2% acetonitrile. In principle each amino acid can be modified at different locations. **C,** Comparative ion currents of fragmentation spectra. Total intensities for (A) all MS2 at 690.324 and then the unmodified fragments of YEAADAILHTPG; y2 (H), y3(G), y4(F), y5(E), y6(D), y7(C), y8(B). Qualitatively as intensity "falls off" going upward through the time traces. The majority of species are unmodified from the aspartic acid residue on. This is consistent with the estimates of site occupancy made by MaxQuant. Taken further up to the N-terminus of the peptide (to y11) showed little intensity change. This suggests the majority of intensity is for versions where the first residue was modified.

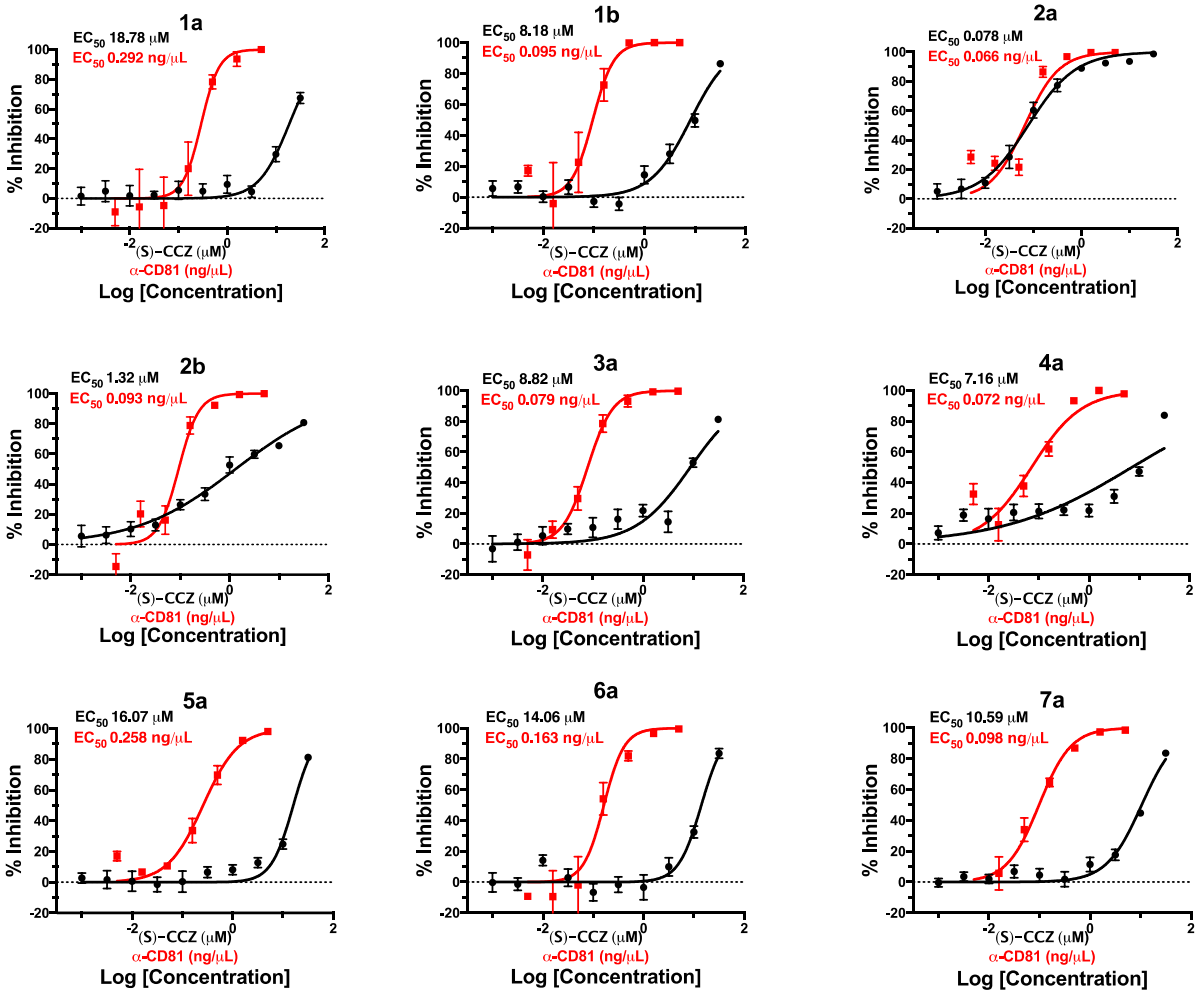


Figure S5. Dose-response curves of (*S*)-CCZ against various HCV-RLuc chimeric genotypes. Related to Figure 1.

Huh7.5.1 cells were infected against various HCV-RLuc chimeric genotypes (1a, 1b, 2a, 2b, 3a, 4a, 5a, 6a, and 7a) in the presence of (*S*)-CCZ. After a 48 h incubation, the cells were harvested for the luciferase assay (black circles) to assess luminescence produced by the virus, and % inhibition was calculated. In a parallel plate, various concentrations of anti-CD81 antibody was incubated during infection and % inhibition was calculated (red squares). The EC_{50} values were calculated using the computer software program, Prism 7. HCV-RLuc (genotype 2a) was used as the reference and the data points were presented as mean values ± SEM (n = 6). The results are representative of three separate experiments.

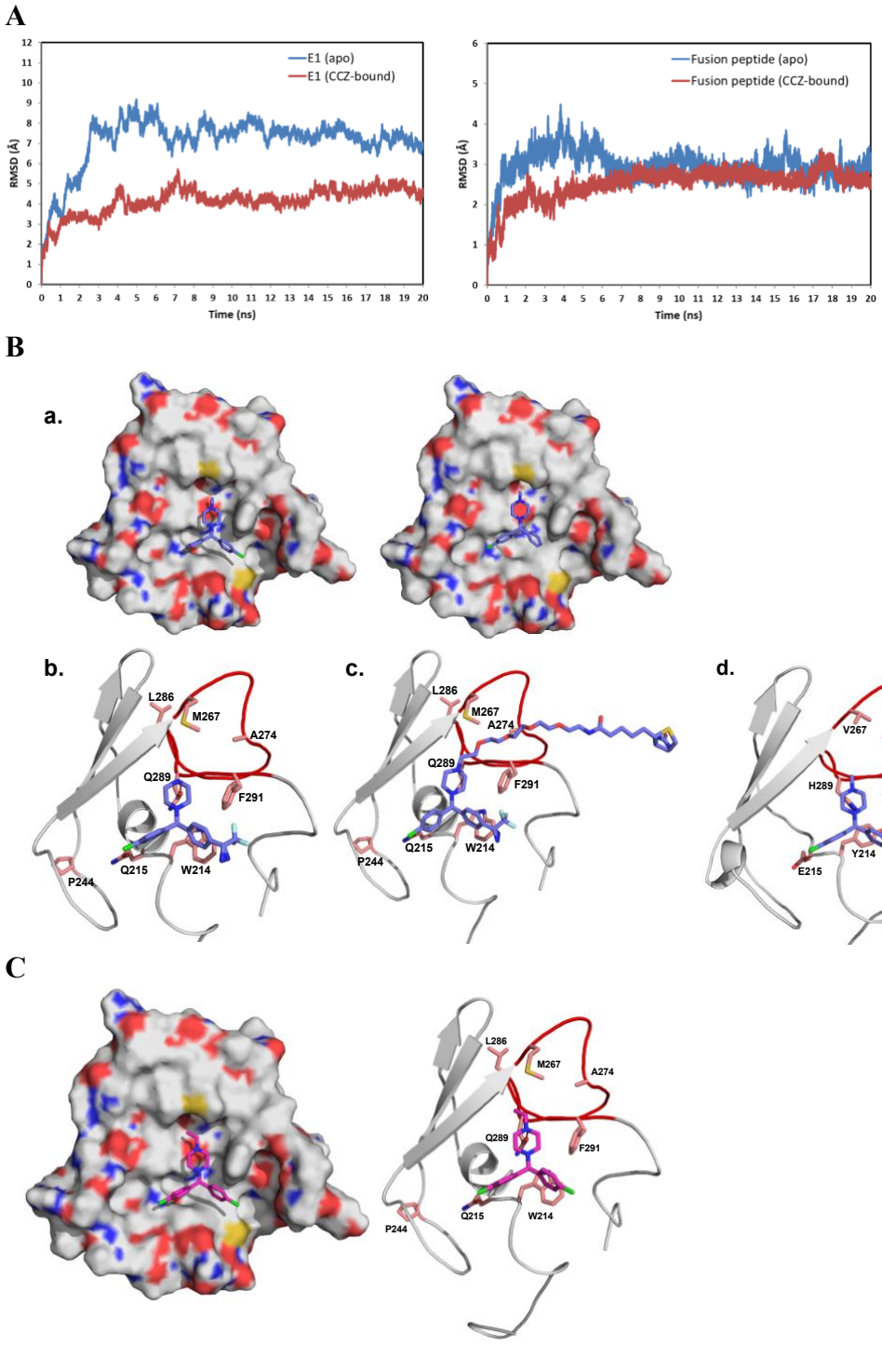

A
E1 (apo)
E1 (CCZ-bound)
RMSD (Å)
Time (ns)
Fusion peptide (apo)
Fusion peptide (CCZ-bound)
RMSD (Å)
Time (ns)
B
a.
b.
L286
M267
A274
Q289
F291
Q215
W214
P244
c.
L286
M267
A274
Q289
F291
Q215
W214
P244
d.
V267
T274
H289
L291
Y214
E215
C
L286
M267
A274
Q289
F291
Q215
W214
P244

Figure S6. Molecular modelling of CCZ binding to HCV E1 protein. Related to Figure 6.
A, Comparison of root mean-square deviation (RMSD) of E1 and the fusion loop in the apo form and CCZ-bound complex. RMSD of the backbone atoms for the E1 structural model with respect to the initial structure over 20 ns of MD simulations showed that the E1 protein structure is highly dynamic at the terminal loop regions and is significantly stabilized upon the small molecule CCZ binding at the hydrophobic pocket (left panel). The putative fusion peptide of E1 reaches a stable conformation to form a hydrophobic pocket and accommodate the CCZ binding (right panel). **B,** Molecular modeling of the hydrophobic pocket of the E1 protein with (*R*)-CCZ, CCZ-diazirine, CCZ-DB and (*S*)-CCZ-resistant HCV mutants. a) (*R*)-CCZ and (*S*)-CCZ have similar EC_{50}'s, indicating similar affinities to the E1 protein. Although enantiomerically different, (*R*)-CCZ can flip in the hydrophobic pocket with interactions with the E1 protein similar to (*S*)-CCZ. b) CCZ-diazirine was used as a precursor of CCZ-DB. The model shows that CCZ-diazirine has similar affinities to the residues that (*S*)-CCZ interacts with. c)The model shows CCZ-DB interacting with the key residues in the hydrophobic pocket with similar affinity to (*S*)-CCZ. Only the *S*-configuration of the CCZ-diazirine and CCZ-DB is shown here and the diazirine cross-links with the closest residue W214, whereas *R*-configuration, not shown here, allows interaction with the Q215. The long polyethylene glycol-biotin chain extends to the top right of the hydrophobic pocket of the fusion peptide and is probably exposed outward on the surface of the E1 protein, allowing interaction of the biotin with the Neutravidin beads. d) The mutations, W214Y, Q215E, P244S, M267V, A274T, L286I, Q289H, and F291L, confer resistance to (*S*)-CCZ. The mutations present in the model disrupts the original interaction of each residue with (*S*)-CCZ. Presented is a hypothetical model of all the different mutations with resulting loss of interactions between (*S*)-CCZ and the wild-type E1 structure. **C,** Molecular modeling of dichloro-ethyl-CCZ interaction with E1 protein. The dichloro-ethyl-CCZ analog is more active than (*S*)-CCZ by approximately 10-fold in the anti-HCV EC_{50}. The additional chlorine group enhances the compound's interaction with the bottom of the hydrophobic pocket and the additional carbon of the ethyl group allows a stronger interaction with the M267 in comparison to (*S*)-CCZ illustrated in Figure 6A.

Methods S1 (Related to STAR Methods)

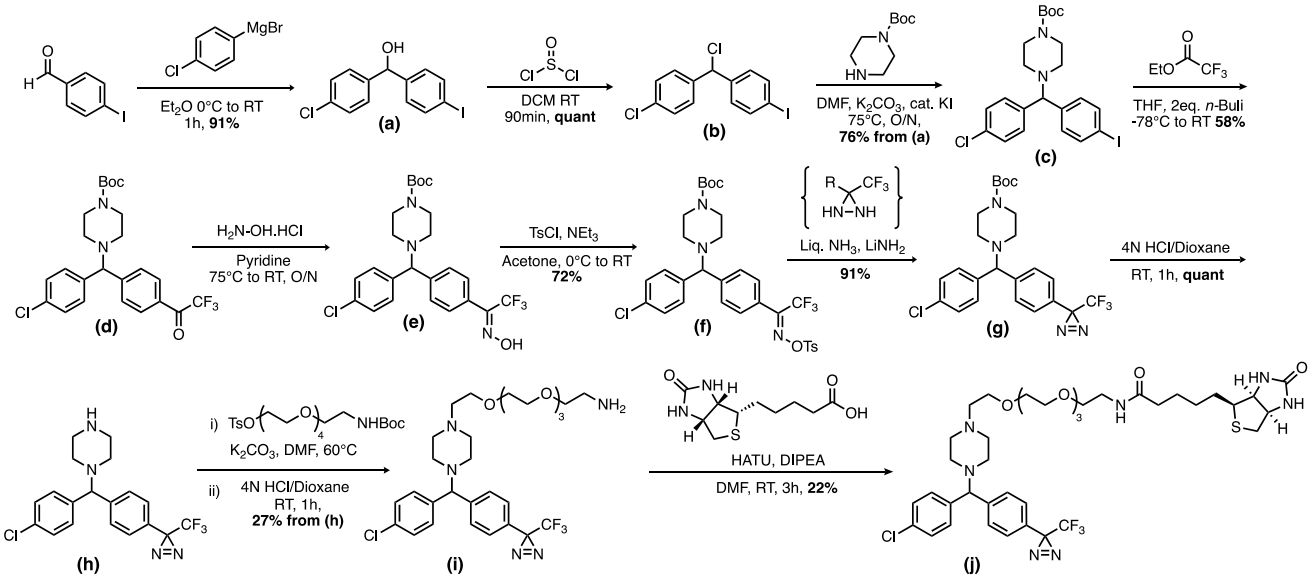
Synthetic methods general. All air or moisture sensitive reactions were performed under positive pressure of nitrogen with oven-or flame dried glassware where stated. Anhydrous solvents were purchased from Sigma-Aldrich (St. Louis, MO). 'Normal phase chromatography', column chromatography, or gradient column chromatography refers to automatic purification using a 'Tyledyne ISCO CombiFlash® Rf+' system with pre-loaded silica gel 'Redisep RF flash columns' of an appropriate size. Preparative purification was performed on a Waters semi-preparative HPLC system (Waters Corp., Milford, MA). The column used was a Phenomenex Luna C18 (5 micron, 30 x 75 mm; Phenomenex, Inc., Torrance, CA) at a flow rate of 45.0 mL/min. The mobile phase consisted of acetonitrile and water (each containing 0.1% trifluoroacetic acid). A gradient of 10% to 50% acetonitrile over 8 min was used during the purification. Fraction collection was triggered by UV detection at 220 nm. Analytical analysis was performed on an Agilent LC/MS (Agilent Technologies, Santa Clara, CA). A Phenomenex Luna C18 column (3 micron, 3 x 75 mm) was used at a temperature of 50°C. Purity determination was performed using an Agilent diode array detector for both Method 1 and Method 2.

Method 1: A 7-min gradient of 4% to 100% acetonitrile (containing 0.025% trifluoroacetic acid) in water (containing 0.05% trifluoroacetic acid) was used with an 8-min run time at a flow rate of 1.0 mL/min.
Method 2: A 3-min gradient of 4% to 100% acetonitrile (containing 0.025% trifluoroacetic acid) in water (containing 0.05% trifluoroacetic acid) was used with a 4.5-min run time at a flow rate of 1.0 mL/min. Retention time (RT) and m/z for M + 1 is reported, occasionally M + Na is observed exclusively.
Mass determination was performed using an Agilent 6130 mass spectrometer with electrospray ionization in the positive mode.

^{1}H and ^{13}C NMR spectra were recorded on Varian 400 MHz spectrometers (Agilent Technologies, Santa Clara, CA). Signals are described as singlets (s), doublets (d), triplets (t), quadruplets (q), septuplets (sept), doublet of doublets (dd) and doublet of triplets (dt). Spectra were analyzed in MestReNova.

It was noted that the piperazine compounds tend to decompose under normal phase silica gel chromatography, the addition of 0.2% of triethylamine v/v into the hexanes mobile phase was sufficient to stop the this. In such circumstances where the products undergo normal phase chromatographic purification, this additive was used as a precaution in every case, as stated.

Synthesis of CCZ-Diazirine-Biotin.



(4-chlorophenyl)(4-iodophenyl)methanol (**a**)

A flame-dried 500 mL round-bottom flask was charged with 4-iodobenzaldehyde (5 g, 21.55 mmol) and Et_2O (170 mL) under nitrogen. The mixture was cooled to 0°C and (4-chlorophenyl)magnesium bromide (22.63 mL, 22.63 mmol) was added dropwise. The reaction was stirred at 0°C for 1h. H_2O (25 mL) was added to quench the reaction, the mixture was poured into a separatory funnel, diluted with brine (100 mL) and EtOAc. Organics were separated and the aqueous layer was extracted twice with EtOAc (2 x 150 mL). The combined organic phases were dried over magnesium sulfate, and concentrated under reduced pressure to give the crude product, which is purified by gradient column chromatography 15-30% EtOAc in hexanes to yield 1-(4-chlorophenyl)(4-iodophenyl)methanol, **a** as a white solid. (6.78 g, 19.68 mmol, 91% yield).

LCMS RT (Method 2) 3.61 min, (no m/z in +ve mode)
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 – 7.60 (m, 2H), 7.36 – 7.23 (m, 4H), 7.13 – 7.07 (m, 2H), 5.75 (s, 1H), 2.25 (s, 1H).
^{13}C NMR (400 MHz, $CDCl_3$) δ 142.98, 141.68, 137.64, 133.60, 128.74, 128.37, 127.83, 93.35, 77.32, 77.00, 76.68, 75.06.
IR (neat) 3332, 1903, 1586, 1481.

1-chloro-4-(chloro(4-iodophenyl)methyl)benzene (**b**)
1-(4-chlorophenyl)(4-iodophenyl)methanol **(a)** (4g, 11.61 mmol) was dissolved in DCM (20 mL) in an oven-dried 50 mL round-bottom flask at room temperature, to this was added N,N-dimethylformamide (0.1 mL, 1.291 mmol) followed by dropwise addition of thionyl chloride (2.54 mL, 34.8 mmol), evolution of gas is noted. The reaction was stirred for 90 min and was found to be complete by TLC. The majority of volatiles were then removed under rotary evaporation, followed by further evaporation under high vacuum to give the title compound -chloro-4-(chloro(4-iodophenyl)methyl)benzene **(b)** as a sand-coloured solid (crude product 4.57 g, 12.59 mmol). Crude product was taken through to the next step.

LCMS RT (Method 2) 4.02min, (no m/z in +ve mode),
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.71 – 7.66 (m, 2H), 7.35 – 7.29 (m, 4H), 7.14 – 7.11 (m, 2H), 6.02 (s, 1H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 140.29, 139.00, 137.75, 134.21, 129.53, 129.04, 128.82, 94.05, 77.32, 77.00, 76.68, 62.66. IR (neat) 1901, 1582, 1483, 1398.
Tert-butyl4-((4-chlorophenyl)(4-iodophenyl)methyl)piperazine-1- carboxylate (**c**)
To a stirring solution of 1-chloro-4-(chloro(4-iodophenyl)methyl)benzene (**b**), (4.5 g, 12.40 mmol) in N,N-dimethylformamide (35 mL) in a 100 mL round- bottom flask was added potassium carbonate (3.43 g, 24.79 mmol), potassium iodide (2.264 g, 13.64 mmol) and tert-butyl piperazine-1-carboxylate (6.93 g, 37.2 mmol). The reaction vessel is fitted with a reflux condenser and stirred at 75°C for 16 h. The reaction is poured into a separatory funnel, diluted with brine EtOAc (150 mL) which is extracted with brine (3 x 100 mL). The organic phase was dried over magnesium sulfate, and concentrated under reduced pressure to give the crude product, which is purified by gradient column chromatography 0-35% gradient EtOAc in hexanes (containing 0.2% triethylamine v/v) to yield tert-butyl 4-((4-chlorophenyl)(4- iodophenyl)methyl)piperazine -1-carboxylate as a white solid. (4.86 g, 9.48 mmol, 76 % yield for 2 steps).

LCMS RT (Method 2) 3.65min (m/z 513.1)
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.61 (d, *J* = 8.4 Hz, 2H), 7.30 (d, *J* = 8.6 Hz, 2H), 7.25 (d, *J* = 8.8 Hz, 2H), 7.12 (d, *J* = 8.3 Hz, 2H), 4.16 (s, 1H), 3.41 (s-br, 4H), 2.30 (s-br, 4H), 1.43 (s, 9H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 154.71, 141.65, 140.29, 137.77, 132.99, 129.69, 129.03, 128.86, 92.63, 79.61, 74.75, 51.56, 43.94 (weak, broad), 28.38.
IR (neat) 2975, 2812, 1680, 1482, 1402.

Tert-butyl4-((4-chlorophenyl)(4-(2,2,2- trifluoroacetyl)phenyl)methyl)piperazine-1-carboxylate (**d**)
To a solution of tert-butyl 4-((4-chlorophenyl)(4-iodophenyl)methyl) piperazine-1-carboxylate **(c)** (2 g, 3.90 mmol) in THF (40 mL) in a flame-dried 250mL round bottom flask at -78°C was added n-butyllithium (5.00 mL of 1.6 M solution in hexanes, 8.00 mmol) dropwise and the reaction was stirred for 20 min at -78°C. To this was added ethyl 2,2,2- trifluoroacetate (0.928 mL, 7.80 mmol) and the reaction is stirred for 20 min at -78°C then allowed to warm to

room temperature, and stirring is continued for 30 min. Reaction was quenched with sat.$NH_4Cl_{(aq)}$ (20 mL), poured into a separatory funnel and diluted with EtOAc, which was washed with sat. $NaCl_{(aq)}$ organics dried over magnesium sulfate, evaporated under reduced pressure and purified by gradient column chromatography 10-25% gradient EtOAc in hexanes (containing 0.2% triethylamine v/v). Yields tert-butyl 4-((4- chlorophenyl)(4-(2,2,2-trifluoroacetyl)phenyl)methyl) piperazine-1- carboxylate **(d)** as an off-white/yellow solid (1.122 g, 2.116 mmol, 58% yield).

LCMS RT (Method 2) 3.133 (m/z 501.2 (hydrate))
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.03 – 7.97 (m, 2H), 7.59 (dd, *J* = 8.6, 2.1 Hz, 2H), 7.35 – 7.26 (m, 4H), 4.31 (s, 1H), 3.43 (m, 4H), 2.34 (d, *J* = 6.9 Hz, 4H), 1.43 (s, 9H)
^{13}C NMR (101 MHz, $CDCl_3$) δ 154.68, 150.42, 139.24, 133.50, 130.73, 129.19, 129.11, 129.04, 128.94, 128.88, 128.29, 79.72, 75.06, 51.62, 44.47, 28.38.
^{19}F NMR (376 MHz, $CDCl_3$) δ -71.44.
HRMS (ESI) *m/z* calcd for [M (hydrate) + H] 501.1767, found 501.1731.
IR (neat) 2975, 1668, 1420, 1246.

Tert-butyl4-((4-chlorophenyl)(4-(2,2,2-trifluoro-1- (hydroxyimino)ethyl)phenyl)methyl) piperazine-1-carboxylate (e)
tert-butyl4-((4-chlorophenyl)(4-(2,2,2-trifluoroacetyl)phenyl)methyl) piperazine-1-carboxylate, **(d)** (976 mg, 2.021 mmol) was dissolved in pyridine 20 mL, hydroxylamine hydrochloride (211 mg, 3.03 mmol) is added, the reaction was stirred at 70^{0}C for 3 h, then stirred at room temperature overnight. The reaction was subjected to rotary evaporation to remove the pyridine, resultant crude oil was then purified directly by gradient column chromatography 0-30% gradient EtOAc in hexanes (containing 0.2% triethylamine v/v) to yield tert-butyl 4-((4-chlorophenyl)(4-(2,2,2-trifluoro-1-(hydroxyimino)ethyl)phenyl)methyl)piperazine-1-carboxylate, **(e)** as a white solid. NMRs display approx. 1:0.8 mixture of conformational isomers. (*viz. syn/anti* oximes). Major isomer reported.

LCMS RT (Method 2) 3.36 min (498.1 m/z)
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.49 – 7.22 (m, 8H), 4.23 (d, *J* = 4.1 Hz, 1H), 3.42 (d, *J* = 5.3 Hz, 4H), 2.33 (t, *J* = 5.0 Hz, 4H), 2.22 (d, *J* = 4.6 Hz, 1H), 1.43 (s, 9H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 154.91, 154.89, 148.98, 144.09, 143.75, 140.00, 139.97, 136.93, 133.11, 129.57, 129.24, 129.17, 129.13, 129.02, 128.89, 128.70, 127.79, 127.75, 127.65, 125.41, 124.08, 117.03, 79.94, 75.18, 51.67, 51.64, 44.62, 28.37.
^{19}F NMR (376 MHz, $CDCl_3$) -62.30, -66.19.
HRMS (ESI) *m/z* calcd for [M + H] 498.1766, found 498.1790.
IR (neat) 2977, 1662, 1431, 1247, 1164.

Tert-butyl(Z)-4-((4-chlorophenyl)(4-(2,2,2-trifluoro-1 ((tosyloxy)imino)ethyl)phenyl)methyl) piperazine-1-carboxylate (f)
To a solution of (Z)-tert-butyl 4-((4-chlorophenyl)(4-(2,2,2-trifluoro-1 (hydroxyimino)ethyl)phenyl) methyl)piperazine-1-carboxylate (**e**) (1.2 g, 2.410 mmol) in acetone (15 mL) at 0°C was added triethylamine (1.008 mL, 7.23 mmol). Following this, 4-methylbenzene-1-sulfonyl chloride (0.551 g, 2.89 mmol) was added and the reaction mixture was stirred at room temperature until determined complete as monitored by TLC (approx. 1 h). Upon completion the volatiles were removed by rotary evaporation and the crude transparent red oil was purified directly by gradient column chromatography 0-40% gradient EtOAc in hexanes over 10 CV (containing 0.2% triethylamine v/v) which gave (Z)-tert-butyl4-((4-chlorophenyl)(4-(2,2,2-trifluoro-1((tosyloxy)imino)ethyl) phenyl)methyl)piperazine-1-carboxylate, **(f)** (1.132 g, 1.735 mmol, 72 % yield) as a transparent colourless oil.

LCMS RT (Method 2) 3.83 min (652.2m/z)
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.87 (d, *J* = 8.4 Hz, 2H), 7.50 (d, *J* = 8.4 Hz, 2H), 7.40 – 7.27 (m, 6H), 4.26 (s, 1H), 3.43 (d, *J* = 5.2 Hz, 4H), 2.47 (s, 3H), 2.37 – 2.27 (m, 4H), 1.44 (d, *J* = 1.0 Hz, 9H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 154.69, 153.14, 146.12, 146.06, 139.59, 133.29, 131.09, 129.82, 129.27, 129.25, 129.00, 128.02, 123.34, 120.94, 79.62, 51.61, 28.37, 21.76.
^{19}F NMR (376 MHz, $CDCl_3$) δ -61.50, -66.46.

Tert-butyl 4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3- yl)phenyl)methyl) piperazine-1-carboxylate (g)

To condensed neat liquid ammonia (5 mL) at -78°C was added tert-butyl (Z)-4- ((4-chlorophenyl)(4-(2,2,2-trifluoro-1 ((tosyloxy)imino)ethyl)phenyl)methyl) piperazine-1-carboxylate **(f)** (400 mg, 0.613 mmol) dissolved in 3mL of Et_2O. To this was added lithium amide (70.4 mg, 3.07 mmol). The vessel was sealed, and the reaction was allowed to warm to room temperature and stirred overnight, the following day, the reaction was cooled to -78°C, opened, and allowed to warm to room temperature under a stream of nitrogen. White solid resulted after ether and ammonia had evaporated which was taken up in solvent DCM and purified directly by gradient column chromatography 0-80% gradient EtOAc in hexanes over 15 CV (containing 0.2% triethylamine v/v) yielding tert-butyl 4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3- yl)phenyl)methyl) piperazine-1-carboxylate, **(g)** (0.760 g, 1.535 mmol, 91% yield).

LCMS RT (Method 2) 3.82 min (m/z 495.1)
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.43 – 7.37 (m, 2H), 7.31 – 7.21 (m, 4H), 7.12 – 7.06 (m, 2H), 4.21 (s, 1H), 3.44 – 3.32 (m, 4H), 2.28 (s, 4H), 1.42 (s, 9H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 154.69, 143.75, 140.04, 133.12, 129.07, 128.91, 128.13, 128.12, 126.81, 126.80, 123.39, 120.66, 79.62, 74.80, 51.57, 36.61, 28.36, 24.65.
^{19}F NMR (376 MHz, $CDCl_3$ δ -65.23.

1-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3yl)phenyl)methyl) piperazine (h)

To tert-butyl 4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin- 3yl)phenyl)methyl) piperazine-1-carboxylate **(g)** (550 mg, 1.111 mmol) at 0°C in the dark was added a commercially available solution of HCl in Dioxane (4 M, 9 mL, 36.0 mmol) followed by 1mL of methanol. The reaction was allowed to warm to room temperature was stirred for 3 h. On completion as determined by LCMS, the solvent was removed by rotary evaporation under reduced pressure in the dark. Diazirinyl CCZ·HCl **(h)** is isolated as a yellow solid in quantitative yield and was used directly in the next step.

LCMS RT (Method 2) 3.09 min (m/z 395.0)
^{1}H NMR (400 MHz, $CDCl_3$) δ 9.67 (s, 2H), 7.97 (s, 2H), 7.83 (s, 2H), 7.39 (s, 2H), 7.25 (d, *J* = 9.2 Hz, 4H), 5.69 (s, 1H), 3.84 (s, 4H), 3.52 (s, 4H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 135.61, 131.36, 130.48, 129.97, 128.87, 127.74, 123.18, 120.45, 74.28, 48.46, 41.20, 28.35, 27.94.
^{19}F NMR (376 MHz, $CDCl_3$) δ -65.04.
HRMS (ESI) *m/z* calcd for [M + H] 395.1245, found 395.1226.
IR (neat) 3378, 2933, 2490, 1641, 1438.

14-(4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3- yl)phenyl)methyl)piperazin-1-yl)-3,6,9,12-tetraoxatetradecan-1-amine (i)

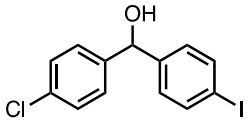
To a solution of 1-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3- yl)phenyl)methyl)piperazine (200 mg, 0.507 mmol) in DMF (2.533 mL) with K_2CO_3 (280 mg, 2.086 mmol) was added 2,2-dimethyl-4-oxo- 3,8,11,14,17-pentaoxa-5-azanonadecan-19-yl 4-methylbenzenesulfonate (398 mg, 0.810 mmol) and the reaction was stirred at 60°C and monitored by LCMS. After 6 h, almost full consumption of starting material was noted, with the concomitant production of Boc-(**i**) (LCMS RT (Method 2) 3.46 min (m/z 714.2)). At this point, the solvent was removed by rotary evaporation under reduced pressure. The crude is solubilised in ~ 1 mL of anhydrous dioxane and 0.5 mL of MeOH, followed by the addition of 4 mL of 4N HCl in dioxane at room temperature, disappearance of the peak for Boc-**i** was noted after 30 min by LCMS. Excess HCl was removed from the reaction mixture by bubbling with nitrogen for 20 min, followed by evaporation of volatiles under reduced pressure and purification by gradient reverse phase chromatography and lyophilisation of pure fractions to yield 14- (4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3- yl)phenyl)methyl)piperazin-1-yl)-3,6,9,12-tetraoxatetradecan-1-amine **(i)** an oil, as a TFA salt. (83 mg, 0.135 mmol, 26.7 % yield).

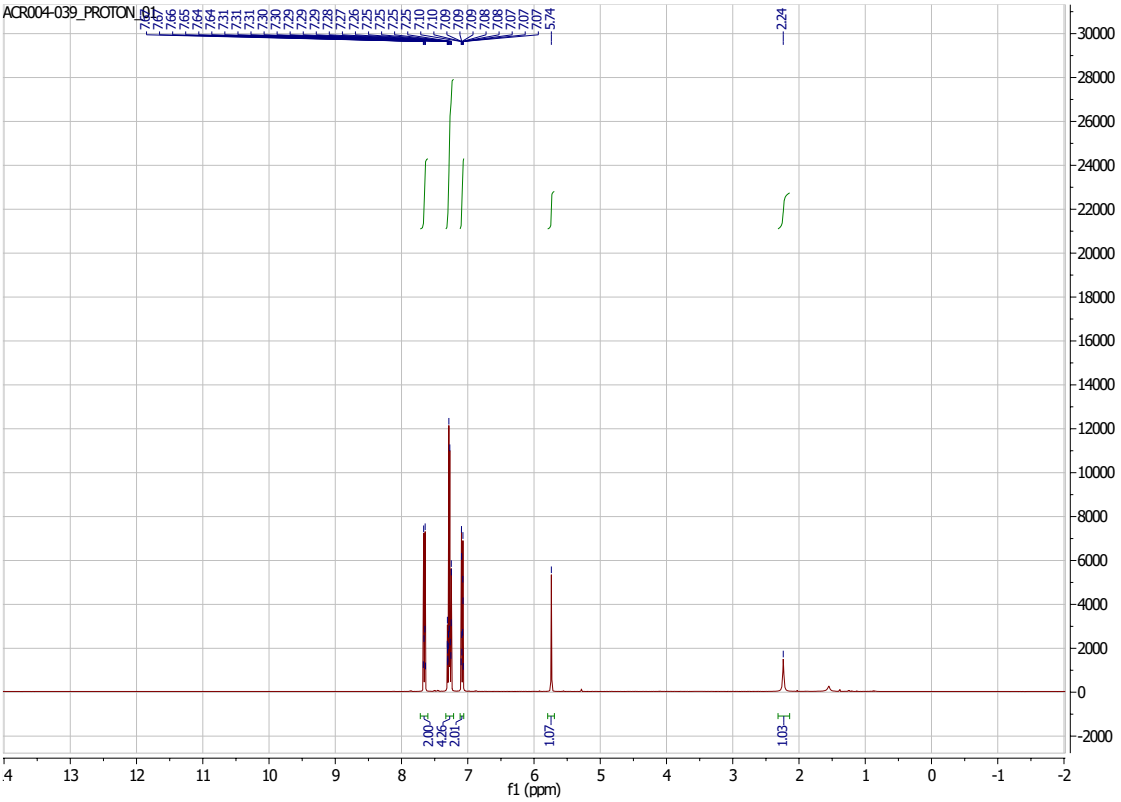
LCMS RT (Method 2) 2.95 min (m/z 614.2)
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.30 (s, 3H), 7.95 (d, *J* = 7.9 Hz, 2H), 7.82 (d, *J* = 7.3 Hz, 2H), 7.35 (d, *J* = 8.1 Hz, 2H), 7.20 (d, *J* = 7.9 Hz, 2H), 5.73 (s, 1H), 3.98 (s, 4H), 3.82 (t, *J* = 4.6 Hz, 2H), 3.74 – 3.52 (m, 18H), 3.36 (s, 2H), 3.19 (s, 2H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 135.63, 130.31, 130.07, 129.94, 129.04, 127.70, 123.19, 120.47, 70.03, 69.88, 69.49, 66.47, 64.68, 61.61, 56.76, 48.38, 39.73.
^{19}F NMR (376 MHz, $CDCl_3$) δ -65.25.
HRMS (ESI) *m/z* calcd for [M + H] 614.2720, found 614.2729.

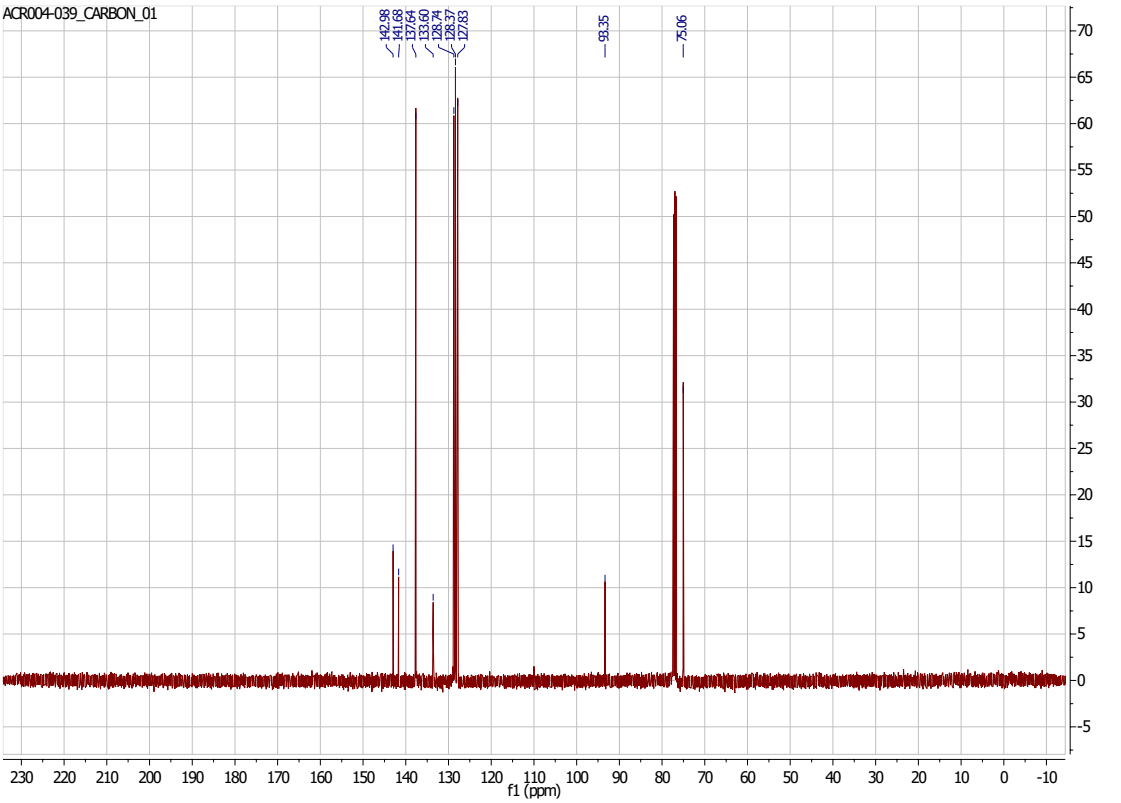
N-(14-(4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3- yl)phenyl)methyl)piperazin-1-yl)-3,6,9,12-tetraoxatetradecyl)-5- ((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4- yl)pentanamide (j)
To a solution of Biotin (9.01 mg, 0.037 mmol) in DMF (307 μL) at room temperature under an atmosphere of nitrogen was added HATU (14.03 mg, 0.037 mmol) and diisopropylethylamine (6.44 μL, 0.037 mmol). The reaction was allowed 15 min stirring for pre-activation then to this was added 14-(4-((4- chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl) methyl)piperazin-1-yl)-3,6,9,12-tetraoxa tetradecan-1-amine hydrochloride (20 mg, 0.031 mmol), followed by 3 further equivalents of diisopropylethylamine. The reaction was stirred for 2 h and monitored by LCMS. Following completion the reaction mixture was purified directly by gradient reverse phase preparative HPLC and lyophilized to yield N-(14-(4-((4- chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl)methyl)piperazin-1-yl)-3,6,9,12-tetraoxatetradecyl)-5- ((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide (15.8 mg, 0.019 mmol, 40.8 % yield) diazirine-CCZ-Biotin a white solid as its TFA salt.

LCMS RT (Method 2) 3.14 min (m/z 840.2)
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.56 (d, *J* = 8.4 Hz, 2H), 7.44 (d, *J* = 8.5 Hz, 2H), 7.33 (d, *J* = 8.5 Hz, 2H), 7.18 (d, *J* = 8.2 Hz, 2H), 6.94 (s, 1H), 6.75 (s, 1H), 4.72 (s, 1H), 4.60 (dd, *J* = 7.9, 4.7 Hz, 1H), 4.40 (dd, *J* = 7.9, 4.5 Hz, 1H), 3.88 – 3.80 (m, 2H), 3.69 – 3.49 (m, 19H), 3.43 (d, *J* = 4.6 Hz, 2H), 3.37 – 3.31 (m, 2H), 3.25 – 3.16 (m, 1H), 3.08 (s, 4H), 2.95 (dd, *J* = 13.1, 4.9 Hz, 1H), 2.77 (d, *J* = 13.1 Hz, 1H), 2.28 (td, *J* = 7.2, 3.4 Hz, 2H), 1.68 (tt, *J* = 13.5, 7.5 Hz, 4H), 1.45 (p, *J* = 7.1 Hz, 2H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 174.84, 164.79, 160.78, 139.18, 135.62, 132.97, 129.70, 129.16, 128.17, 127.51, 116.88, 114.00, 74.48, 70.22, 70.19, 70.08, 69.84, 69.41, 64.93, 62.34, 60.80, 56.51, 55.17, 51.34, 48.49, 40.26, 39.44, 35.24, 27.67, 25.35.
^{19}F NMR (376 MHz, $CDCl_3$) δ -65.20, -75.96.
HRMS (ESI) *m/z* calcd for [M + H] 840.3496, found 840.3521

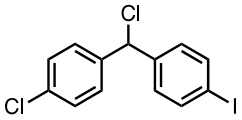
(4-chlorophenyl)(4-iodophenyl)methanol, (**a**)

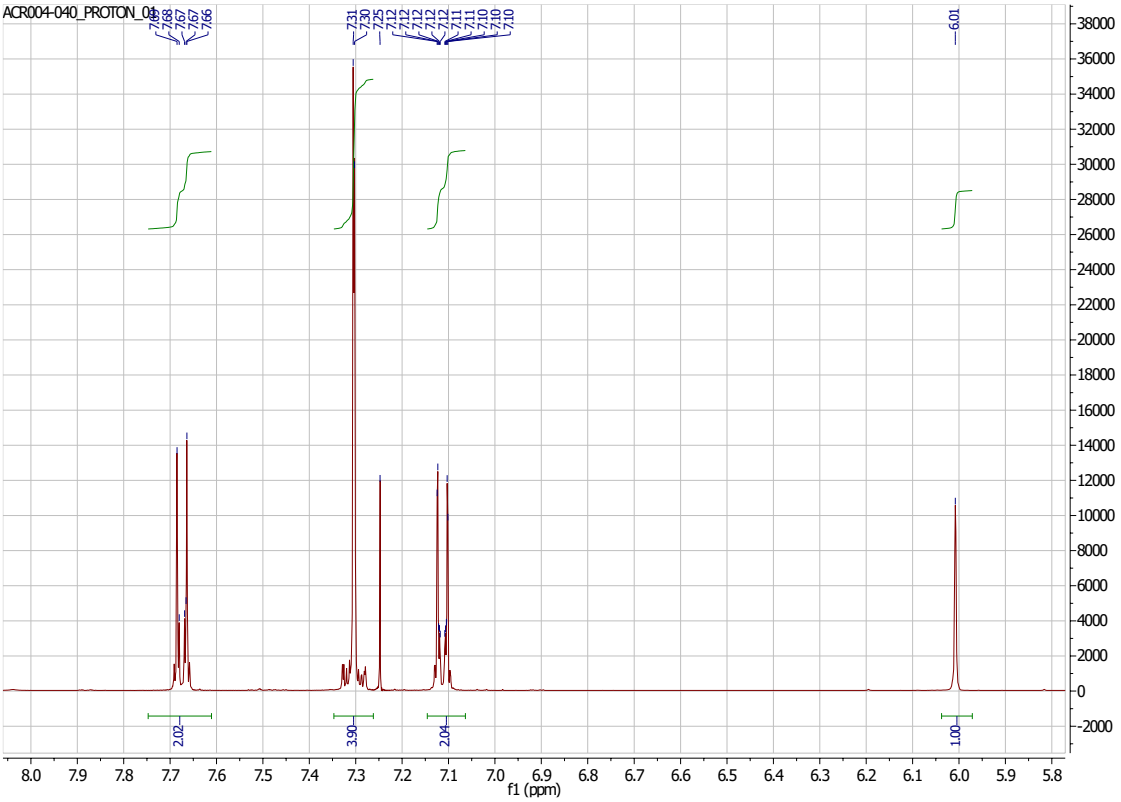


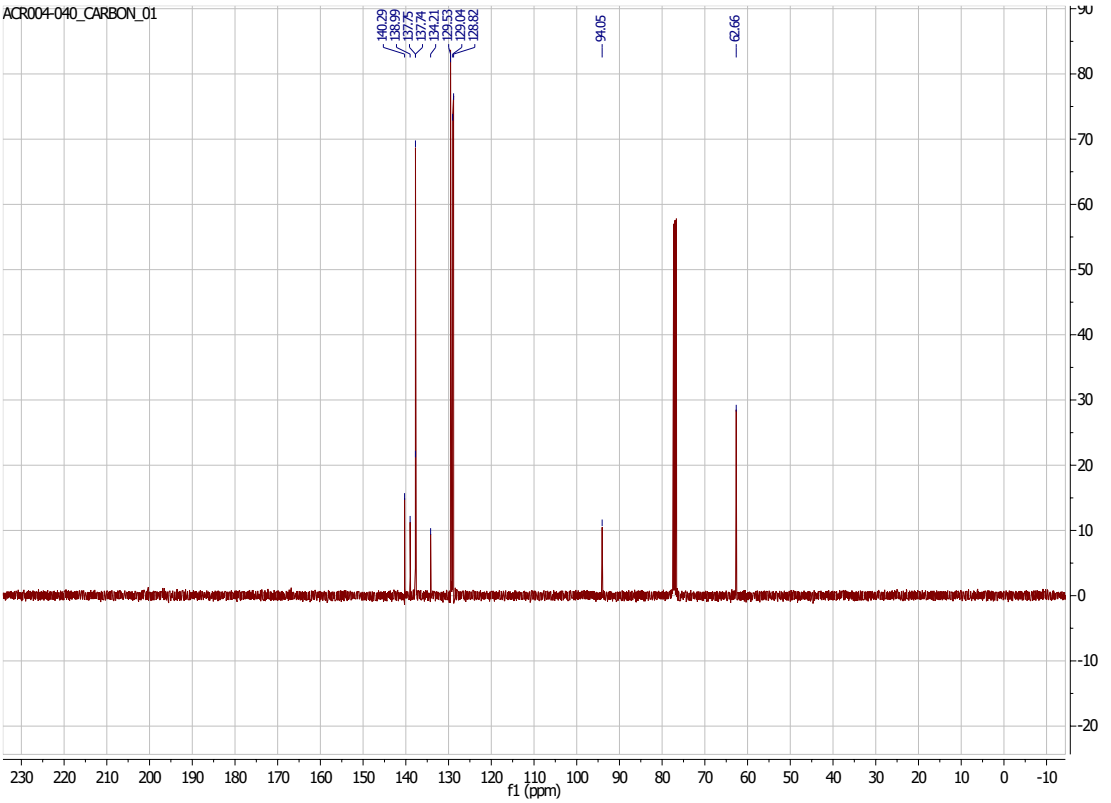




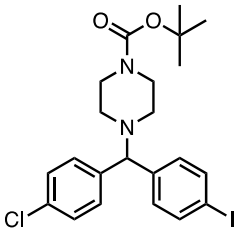
1-chloro-4-(chloro(4-iodophenyl)methyl)benzene, (**b**)

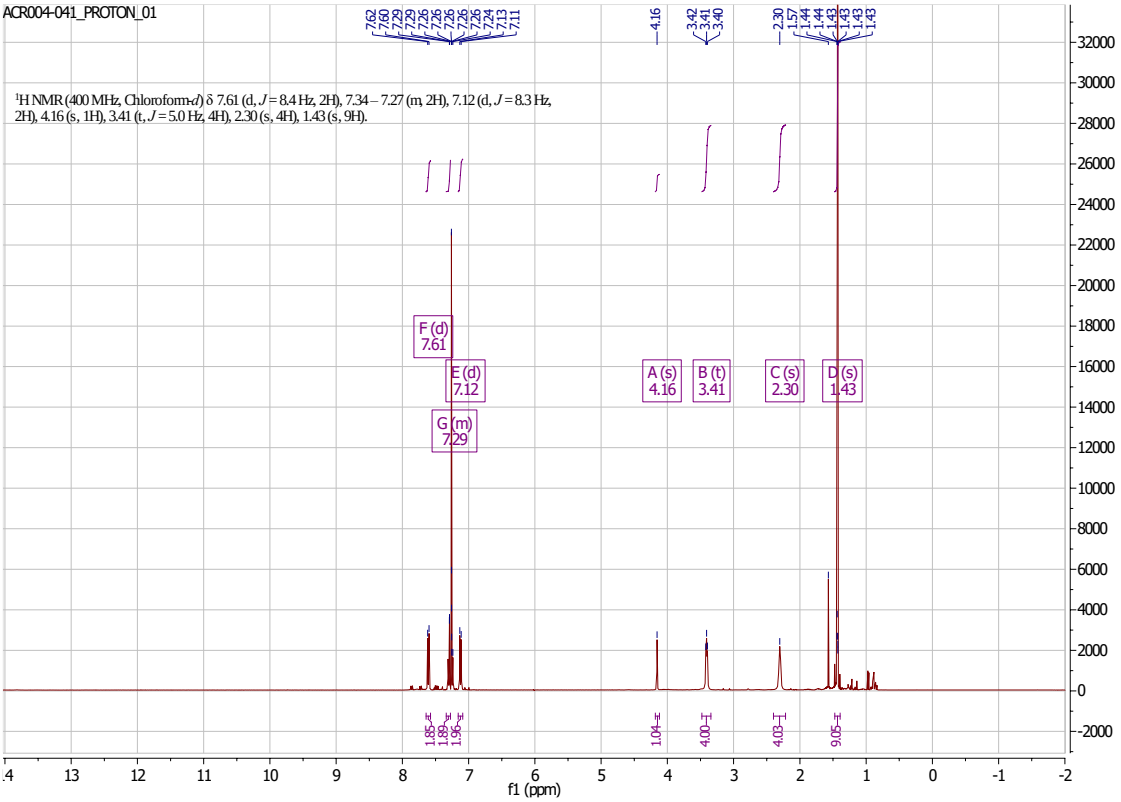


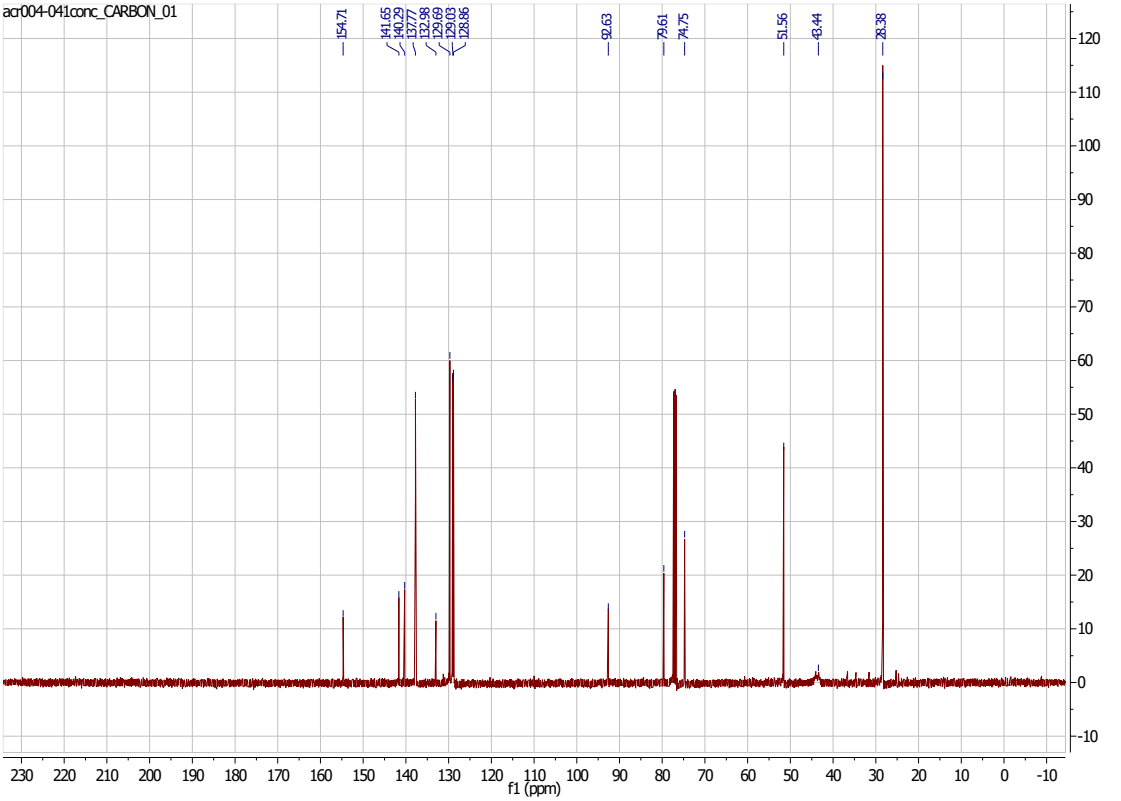




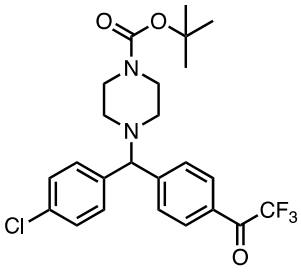
tert-butyl 4-((4-chlorophenyl)(4-iodophenyl) methyl)piperazine -1-carboxylate, (**c**)

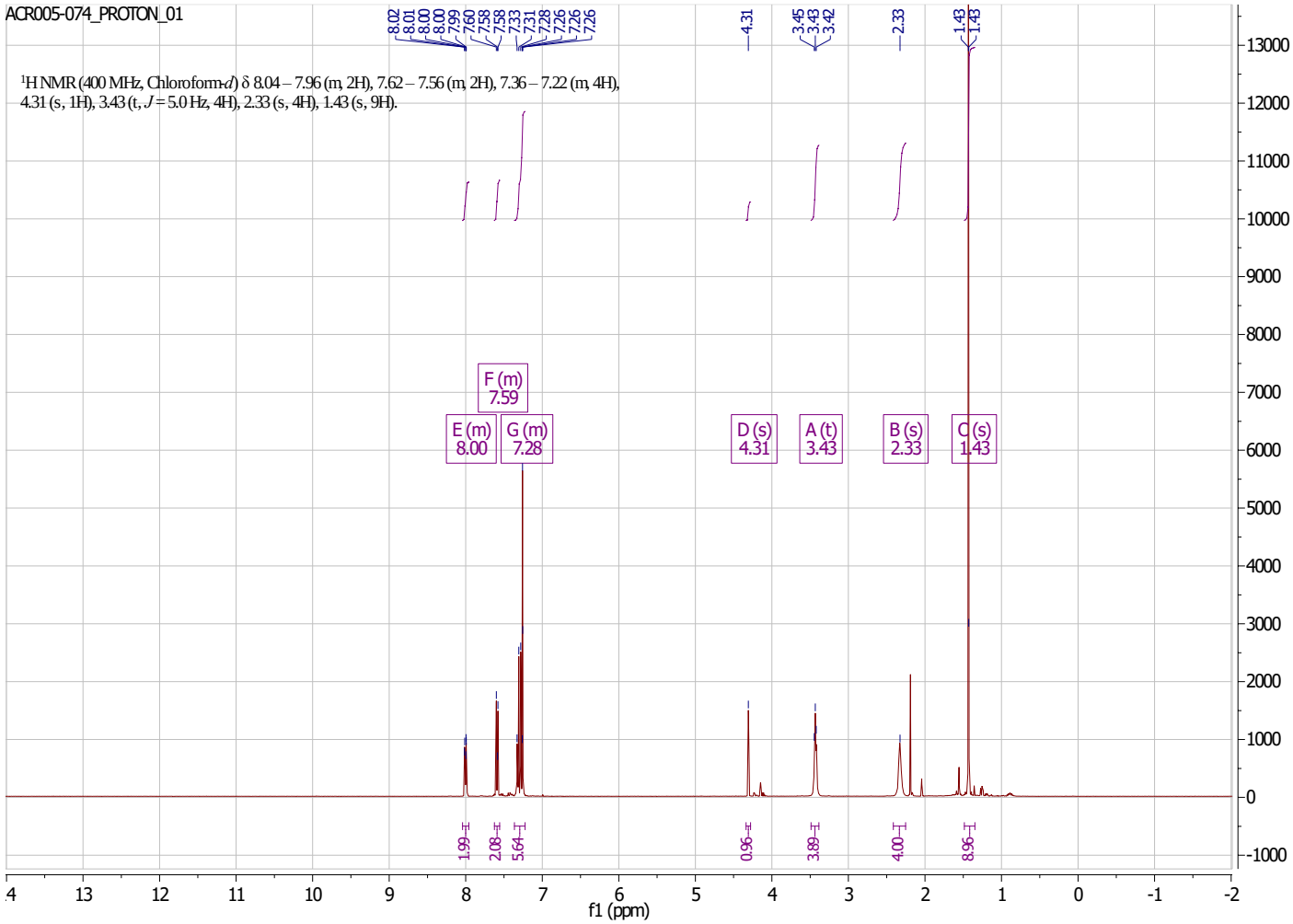


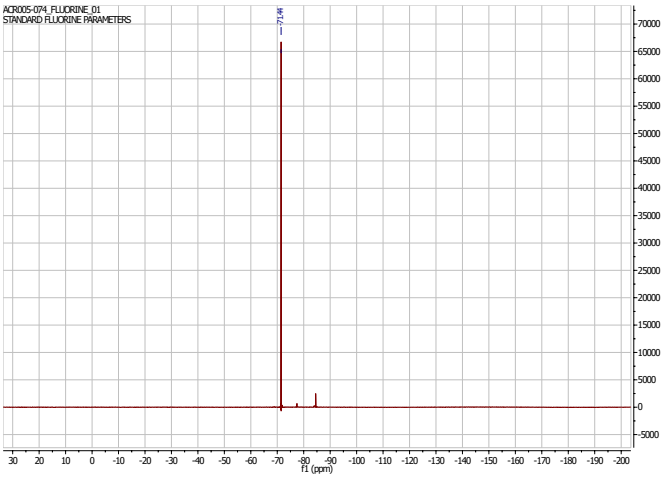


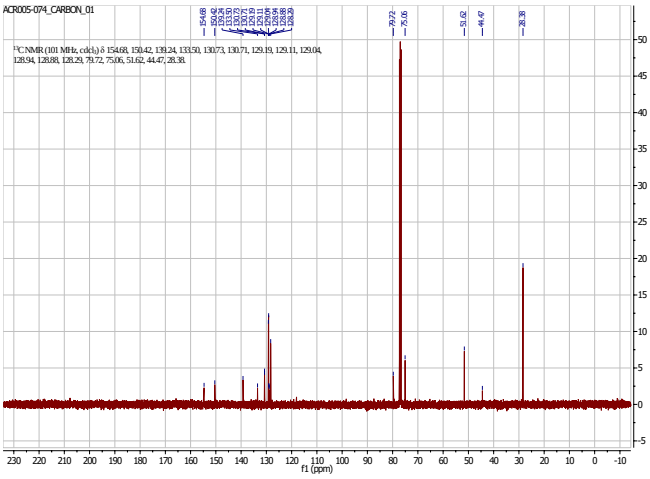


tert-butyl 4-((4-chlorophenyl)(4-(2,2,2-trifluoroacetyl)phenyl) methyl)piperazine-1-carboxylate, (**d**)

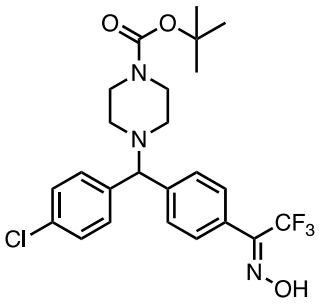


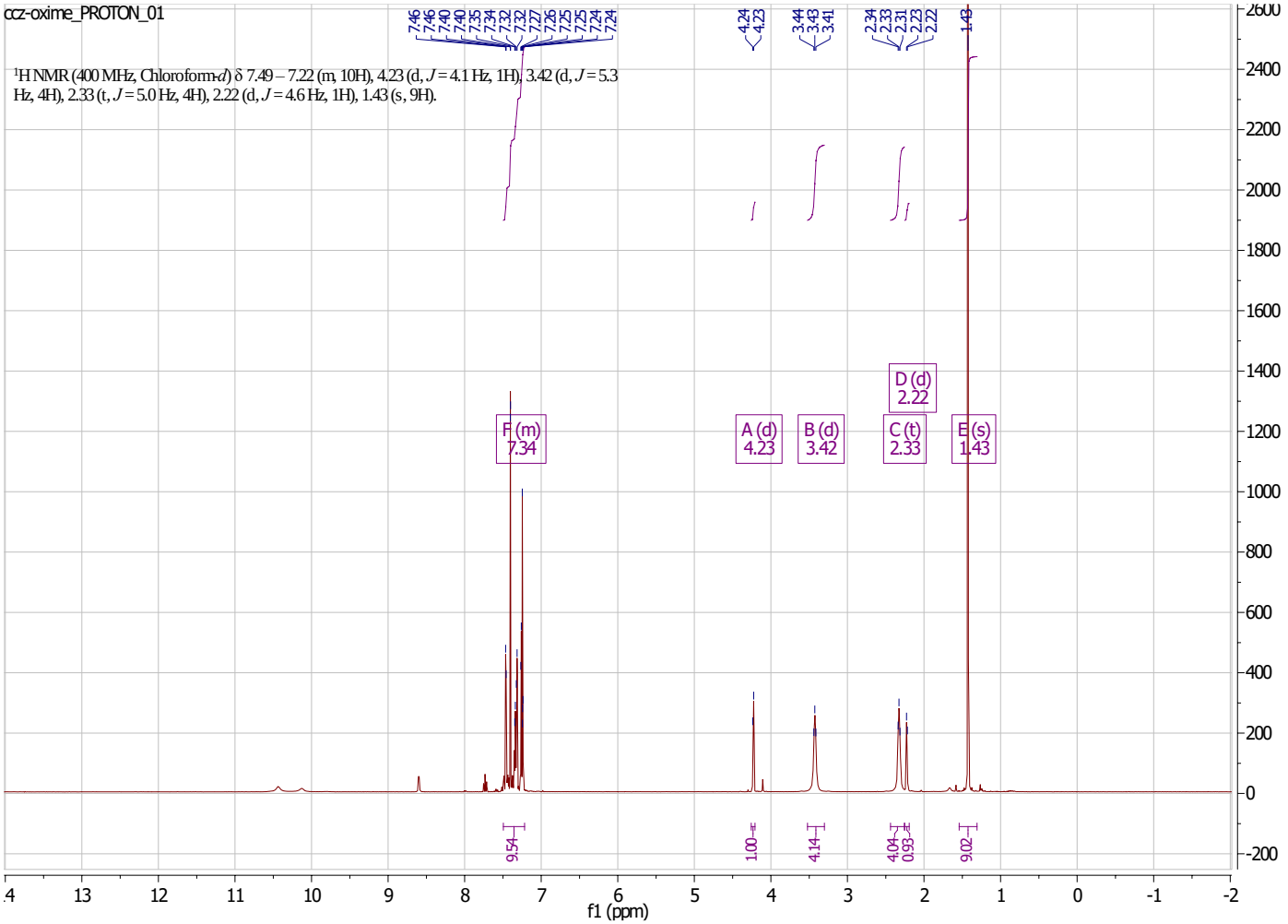


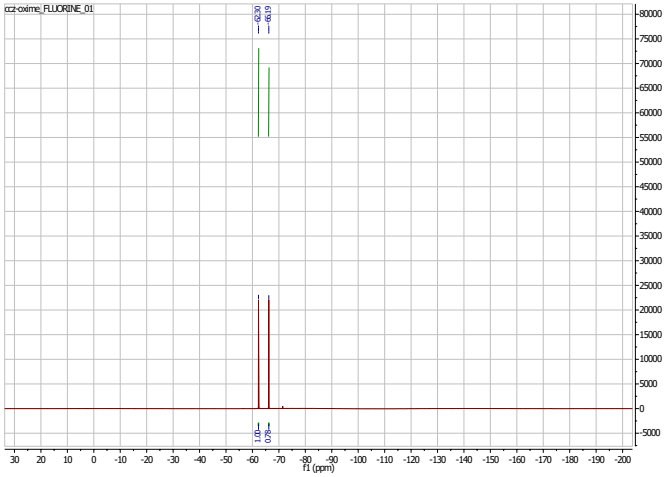


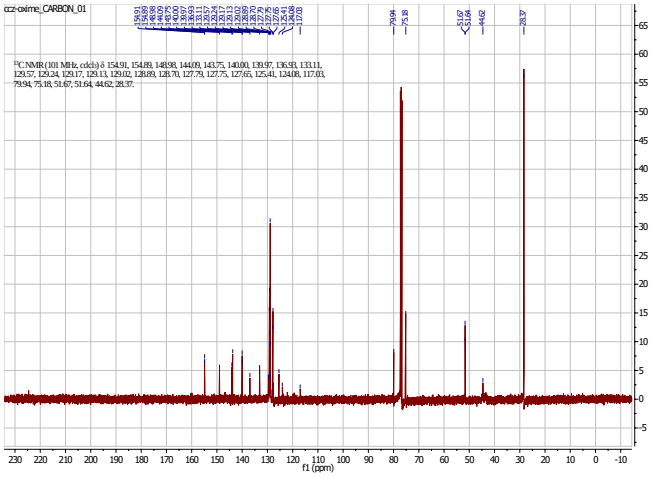


tert-butyl (Z)-4-((4-chlorophenyl)(4-(2,2,2-trifluoro-1-(hydroxyimino)ethyl)phenyl)methyl)piperazine-1-carboxylate, (**e**)

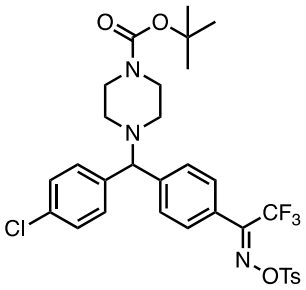


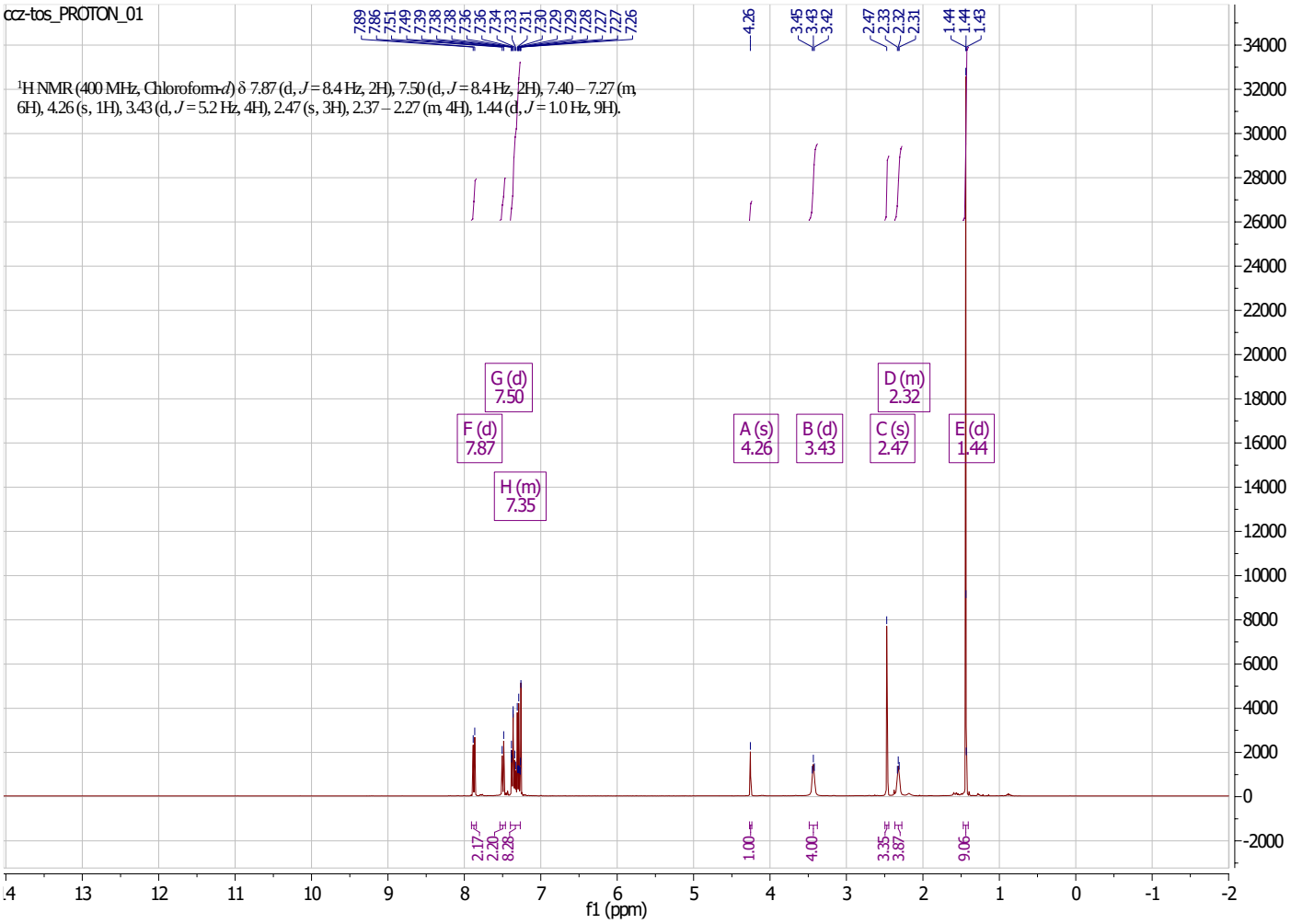


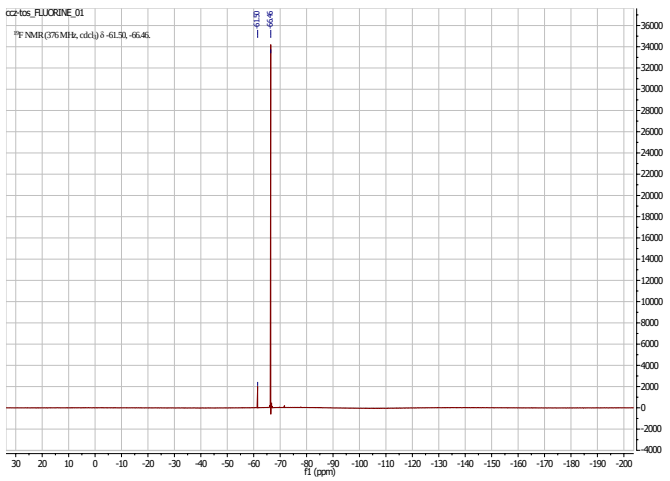


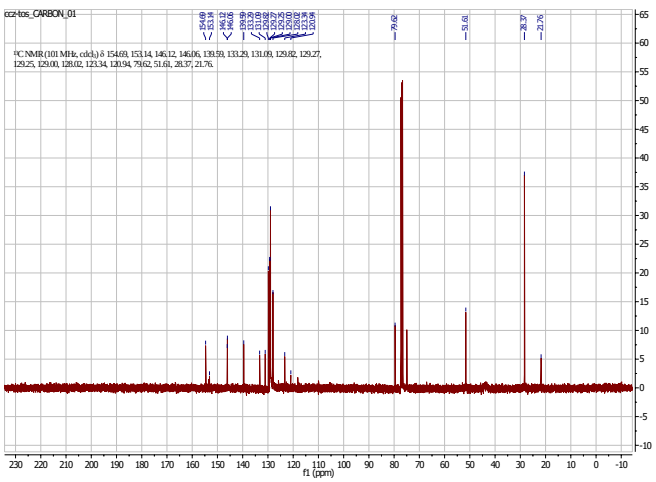


tert-butyl (Z)-4-((4-chlorophenyl)(4-(2,2,2-trifluoro-1-((tosyloxy)imino)ethyl)phenyl)methyl)piperazine-1-carboxylate, (**f**)

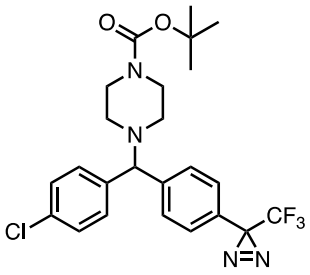


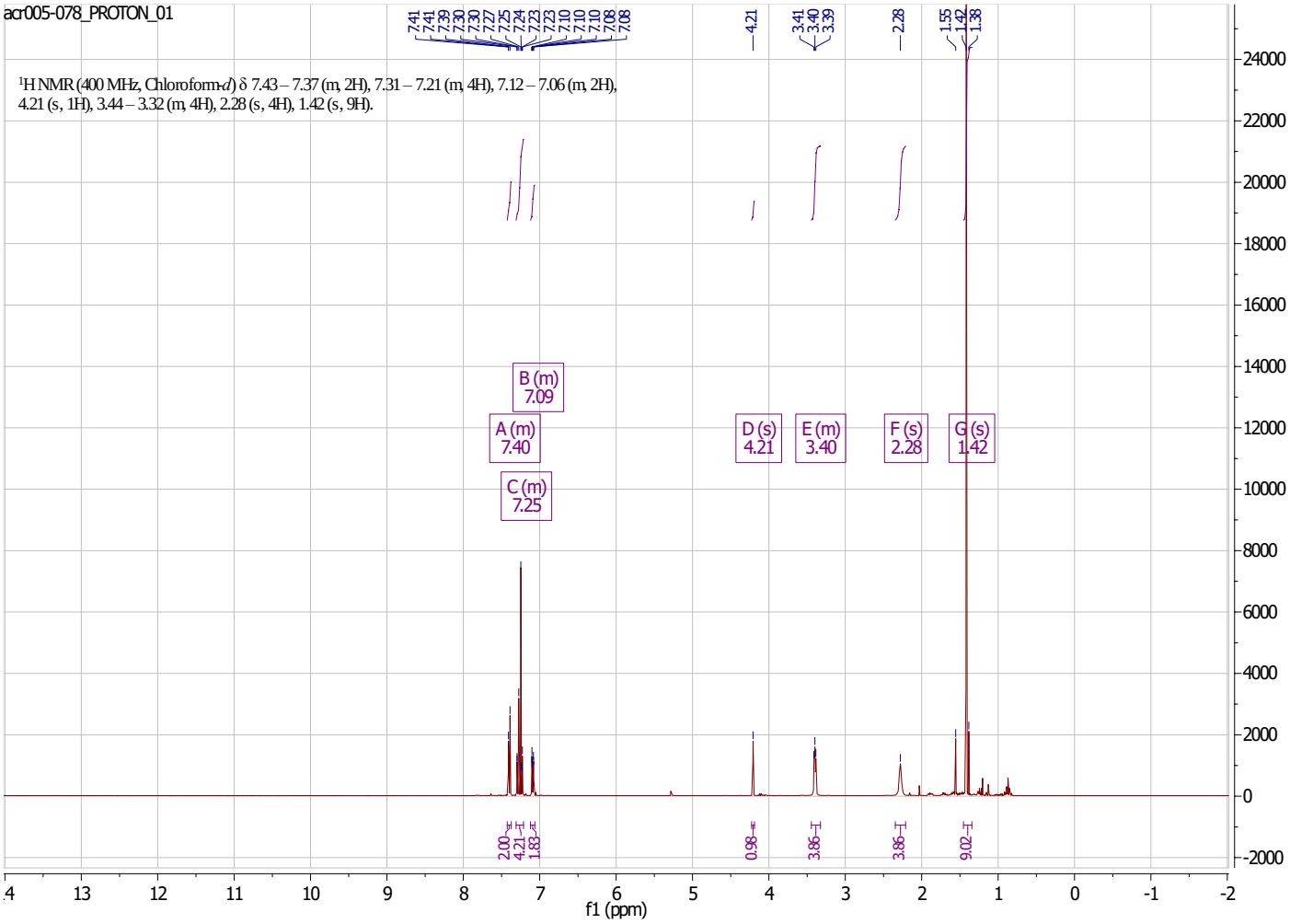


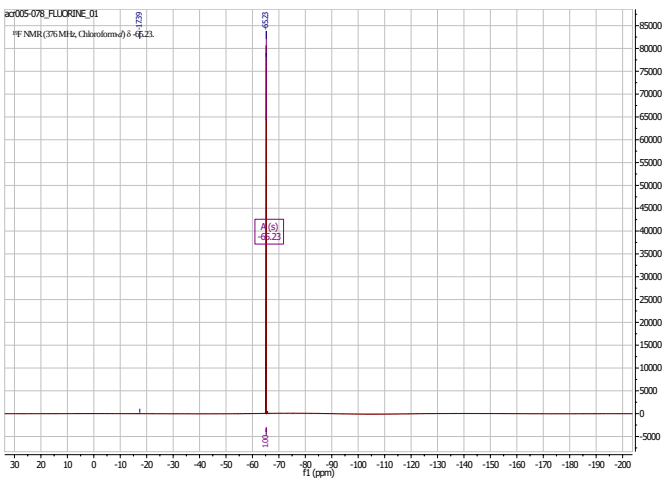


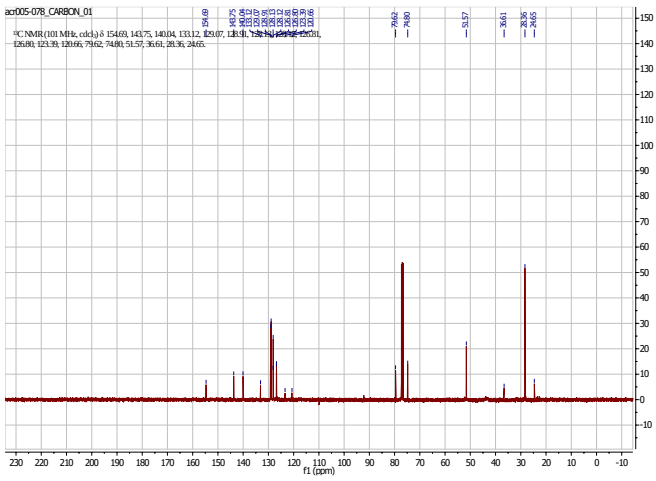


tert-butyl 4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl)methyl)piperazine-1-Carboxylate, (**g**)

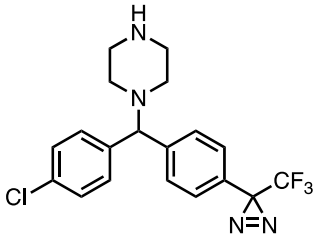


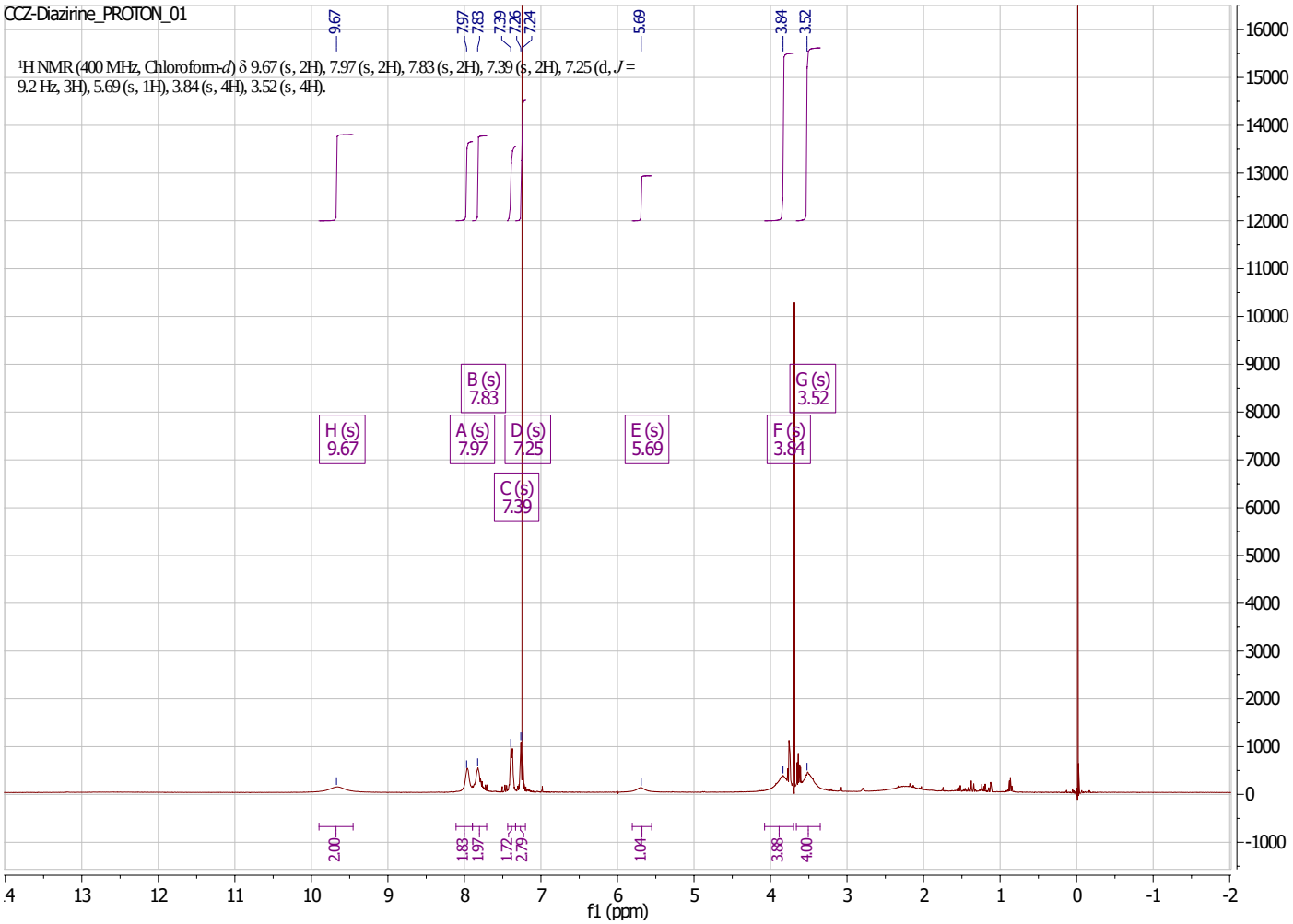


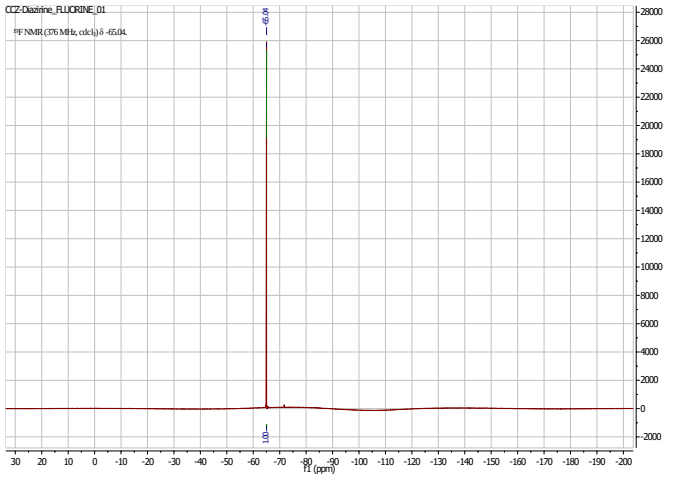


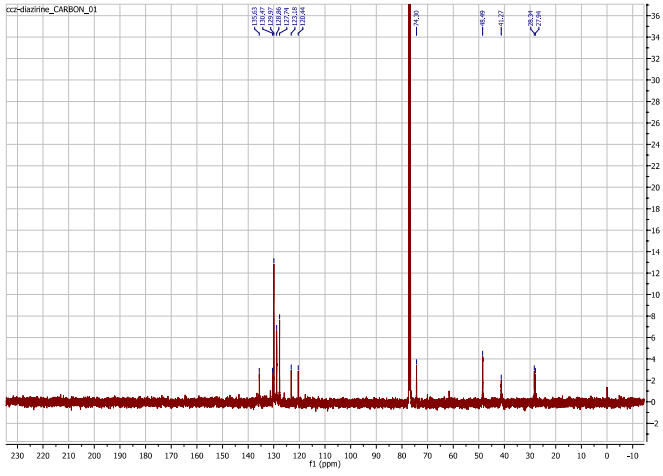


1-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl)methyl)piperazine (**h**)

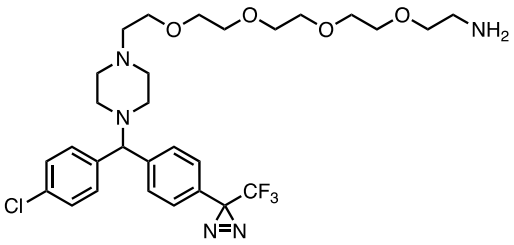


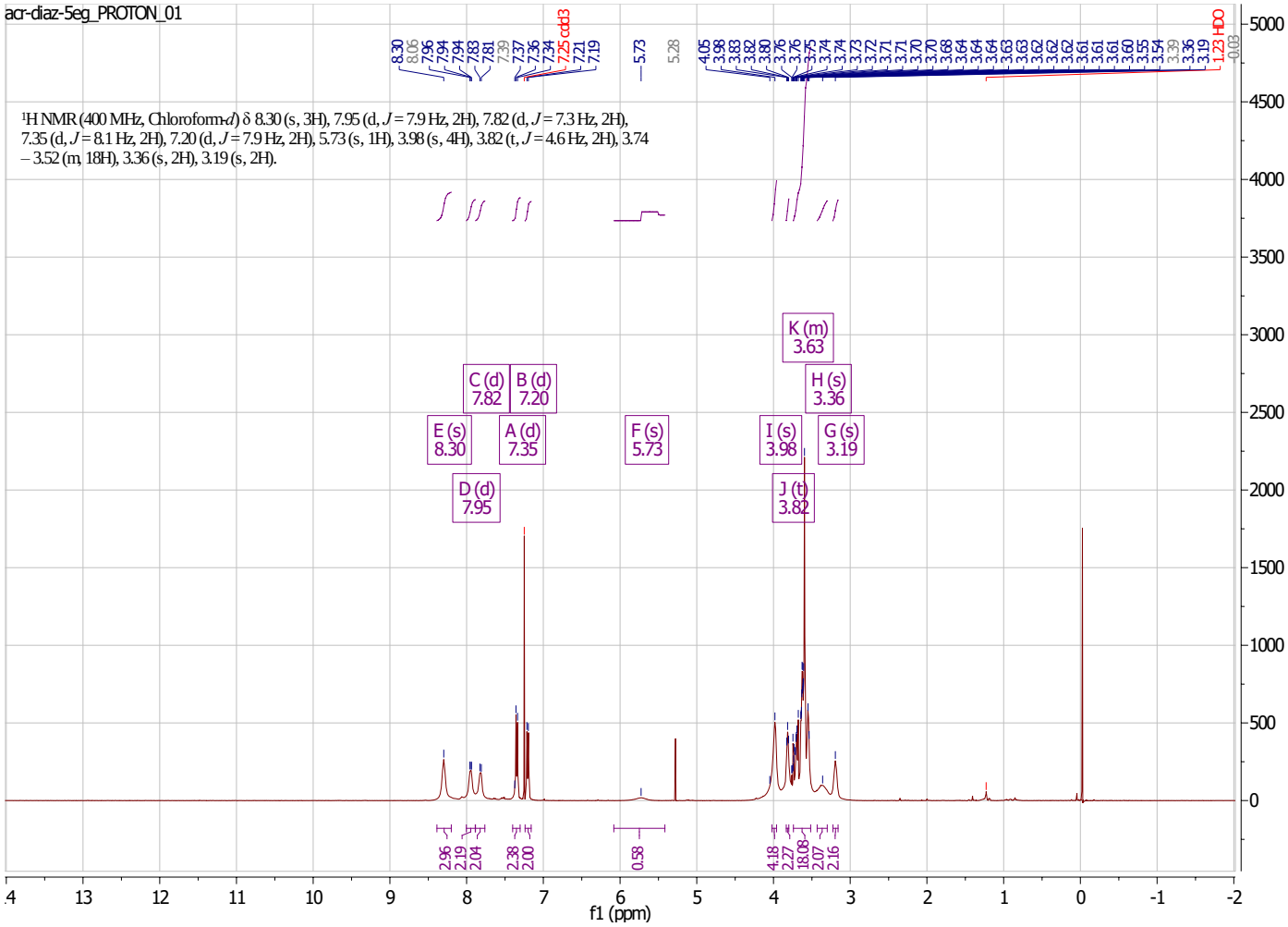


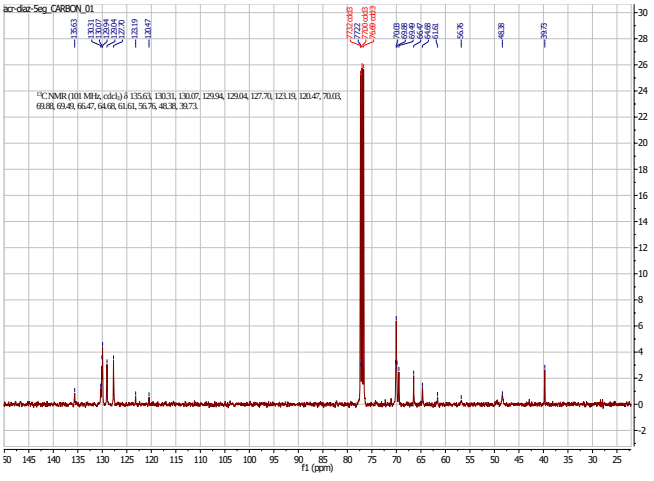


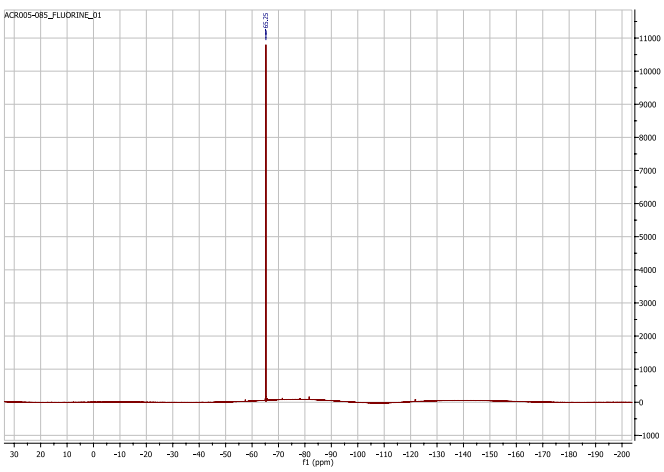


14-(4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl)methyl)piperazin-1-yl)-3,6,9,12-tetraoxatetradecan-1-amine, (**i**)

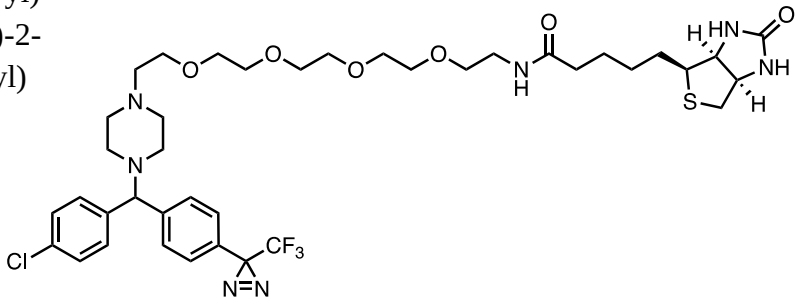


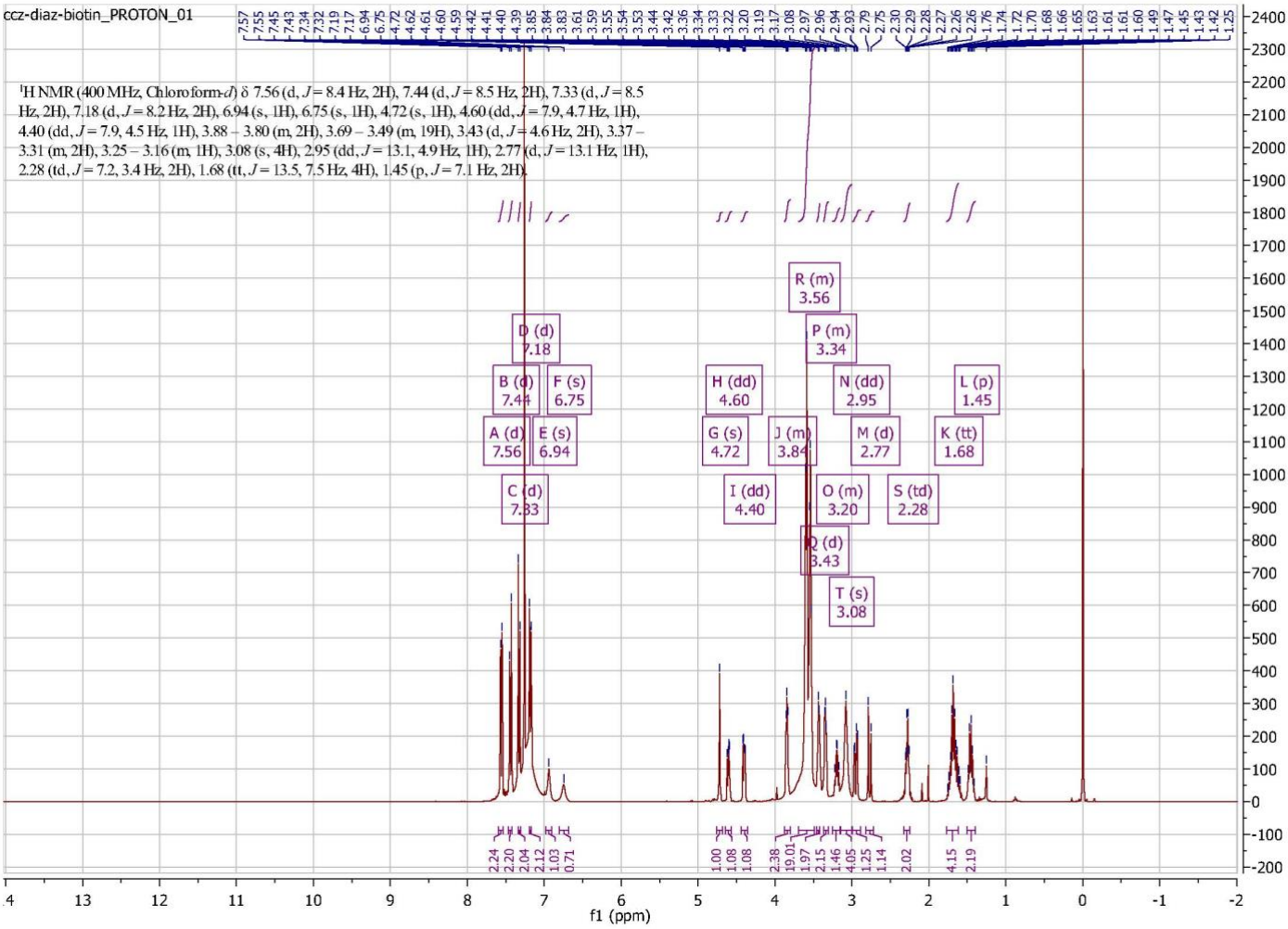


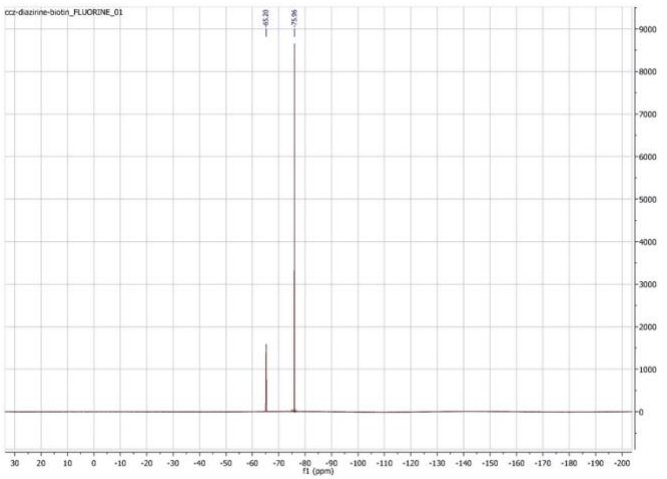


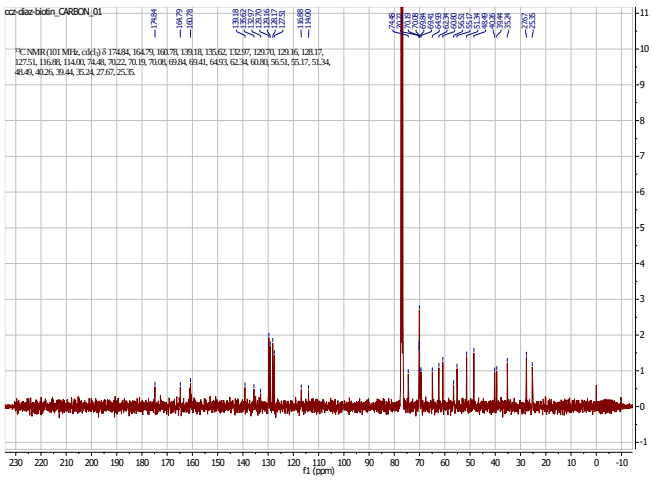


N-(14-(4-((4-chlorophenyl)(4-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl)methyl)piperazin-1-yl)-3,6,9,12-tetraoxatetradecyl)-5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl) pentanamide, (**j**)









Mass Spectrometric analysis of CCZ-diazirine-biotin cross-linked E1 glycoprotein. Static MS/MS of compound CCZ-diazirine-biotin was obtained using a Waters (Milford, MA) Xevo-G2 qTof mass spectrometer. The collision energy was 25 eV and the argon collision gas pressure was 0.1 mBar. The parent ion 840.3 daltons was generated using positive ion ESI at 2.8 KV capillary voltage. There were only 2 daughter ions generated. The daughter ion at 812 da was generated by loss of N_2 from the diazo- group and therefore would not be useful as a tracer ion since that portion of the molecule is consumed in the cross-linking step. The daughter ion at 529.3 da which contains the biotin group was useful as a tracer ion since it originates from the unreacted segment of the molecule and is very prominent in the MS/MS spectrum.

To confirm that the strategy would work with a cross-linked peptide we next created an evaporated film of a mixture of the CCZ-diazirine-biotin and [Glu_1]-Fibrinopeptide B (10 µg each) at room temperature in the dark with low light used for handling. This dried film was then illuminated with UV light for 2 min as described for the protein sample. The sample was then dissolved in 0.3 mL of 0.2% acetic acid 5% acetonitrile, washed three times with two volumes of ethyl acetate, briefly treated with a stream of nitrogen and applied serially to a C8 and C18 STAGE tips (Rappsilber et al., 2007) each containing four 10 gauge cores of C8 and C18 respectively (and worked up as described below). While excess reagent byproducts were also isolated this way along with unreacted [Glu_1]-Fibrinopeptide B enough cross-linked peptide was created that we could obtain fragmentation spectra. We found the cross-linked peptide apparently had a strong signal at 532.32 M/Z along with fragmentation peaks which would result from the loss of this fragment. Based solely on the success of use of this ion as a reporter ion for a cross-linked form combined with use of a "neutral loss" related to the reporter ion's departure to correctly predict fragmentation peaks of analyses we adopted this approach for analysis of cross-linked peptides. We did not further explore the difference in the details of the initial analysis and this peptide model test.

Next another sample was made in the same way using a pre-digested (trypsin) preparation of chicken lysozyme peptides. In addition to being used to confirm the MaxQuant based search strategy (further described below) these samples were also used to check the effects of typical digestion conditions needed for processing the target sample. For reasons unknown to us our samples were sensitive to the treatment level (overnight digestion etc.) normally used for reduction/alkylation and endoprotease digestion which were tested in mock reactions using the film material. We found that a sufficient signal survived after a 2 mM DTT (15 min at 30^0C in 1% sodium dodecanoate, 20 mM phosphate pH 7.8) reduction, 6 mM chloroacetamide (30 min at room temperature in the dark) alkylation, followed by dilution with 9 volumes and a brief digestion (2 h). This was used in the analysis of a sample of the protein and cross-linking agent described above using alpha-lytic protease (20:1, Sigma) as cleavage agent based on earlier digestion checks (although these were performed for longer times and gave better coverage). After digestion in a 300 µL volume the sample was acidified with sparring acetic acid until

initial cloudiness was seen. After 15 min, two volumes of ethyl acetate were added and the protonated dodecanoaic acid was extracted from the liquid phase of the sample and nearly all of the ethyl acetate was removed. Next nitrogen was blown down onto the solution surface to blow off additional ethyl acetate at room temperature for 15 min. The sample was then rapidly chilled in ice water and applied serially to fresh over large C8/C18 stage tips (as above) loading and washing the tips in a chilled centrifuge at 300G. After washing with 3 x 200 µL 0.5% acetic acid 50 mM ammonium acetate the columns were reversed in order in the stack (C18 on top) and 200 µL of 0.5% acetic acid 40% acetonitrile was used to begin elution of the sample into a fresh vial. This was followed by 200 µL 0.5% acetic acid 80% acetonitrile. Two aliquots were created. One sample was stored immediately at -80^{0}C while the other sample was then evaporated nearly to dryness using dry nitrogen gas at room temperature. After the addition of 50 µL of 0.4% acetic acid 5% acetonitrile and brief mixing the sample was kept on dry ice or stirred in a -80^{0}C freezer except for the brief times when the sample was used for subsequent injections.

For analysis of the UV cross linked protein samples by mass spectrometry all experiments involved the injection of 8 µL of sample onto a non-vented 75 micron x 50 cm integrated column-emitter EASY-Spray column mounted on a Thermo nLC-1000 UHPLC system with a temperature setting of 30^{0}C with a working flow rate of 100 nanoliter/min. A gradient from 5 to 32 % in 250 min was followed by a sharper transition to 95% in 10 min and a return to 5% for 40 min. Using a Thermo Lumos mass spectrometer an initial data-dependent analysis experiment was run using four separate mass windows (Davis et al., 2001) (420-528, 527-690, 689-1100, 1095-2000) from each window analytes above 2E3 in signal with charge states 1-6 were targeted unless seen 3 times in the last 3 seconds in which case they were ignored for 3 seconds (multiple sites of cross-linking were accommodated by softening the triggers and durations of sanction normally used in dynamic exclusion). To allow for greater speed and sensitivity the ion trap was used for the fragmentation data but using HCD activation of 30 with stepped collision energy (+/- 7%) and a first mass measurement at 110 and AGC target of 1E4 with maximum injection time of 35 milliseconds.

The resulting data was analyzed using MaxQuant 1.6.016(Cox and Mann, 2008) with default settings and a permissive model for alpha-lytic protease (allowing cleavage after G,A,S,T,V,L,I,E,M,F,Y) based on reported observations (Meyer et al., 2014) and an allowance of up to 10 missed sites to accommodate the limited level of digestion possible in the sample. A custom modification was used which described the modification expected and had been validated using the peptide film samples. The modified peptide would have a mass change of 811.335 Daltons as calculated from the formula $C_{39}H_{53}ClF_3N_5O_6S$ a potential “neutral loss” (though in fact the lost fragment was positively charged this still worked in tests) of 531.3090 Daltons as calculated from the formula $C_{24}H_{45}N_5O_6S$ and a singly charged diagnostic peak at the mass of the protonated version of this “neutral loss” species (M/Z 532.3163). Because the cross-link can form in essentially any residue of a protein searching data exhaustively is very time consuming.

Initial searches were performed using a data base limited to the known protein and common lab contaminants and in separate searches modification of subsets of the 19 non CYS amino acids (although these are explored later as described below). This search suggested 3 different peptides were modified (YEAA, YEAADA, and YEAADAILHTPG). To create a more realistic search space the cross-linked data was analyzed without consideration of the cross-linked modification against the human proteome and a sub-proteome containing all proteins identified in this sample (even remotely) was created for subsequent searches.

Using the sub-proteome search and a modification model including all amino acids in the larger peptide two peptides were detected (YEAADAILHTPG and ADAILHTPG) with Andromeda scores of 139 and 115 respectively. Indicated posterior error probabilities were 3E-163 and 0.0004. However, it is not clear how correct it is to apply these PEP values in this context. Although no other sites of modification were detected for the target protein they were reported for other proteins thought to be in the sample in most cases without the corresponding detection of the unmodified peptide and especially rich in the protein Titin, suggestive that a false discovery rate was higher than that suggested by the PEP values reported for these peptides. Although it is possible that CCZ-diazirine-biotin could react with other proteins in the sample (some of the better "other protein" cross-links where to proteins which were abundant in the sample) one very conservative approach is to view the best PEP for the modification from a different protein than the target protein as coming from a falsely reported cross-link (5.8 E -5) as establishing a rough correction for some over estimate of the PEP approach. The reported PEP for YEAADAILHTPG (1.72 E -14) would then weather this step. This view together with the fact that this peptide was the best modified peptide suggested for the sample and had such a large Andromeda score (which would climb to a value of 315 in the follow up experiment) strongly support the view that the modification is present. No other observations of the modification somewhere else on the protein were detected. This cannot exclude the possibility that there are other modification sites present but such sites would have to be missed through an unlikely combination of efficient cleavage leading to a very small peptide form or a very large peptide which becomes undetectable.

In the data-dependent-analysis experiment both peptides were observed multiple times over an extended time during the latter part of the gradient with multiple distinct chromatographic peaks as would be expected if there were multiple structural isomers created through the reagents modification of multiple sites on the peptide (this was also observed for peptides in the film test samples). To further explore possible isomers a directed experiment was performed using the same gradient conditions as in the first experiment collecting data using a target list which included the +3 form of the modified ADAILHTPG peptide and the +2 and +3 forms of the modified YEAADAILHTPG peptide. Once again MaxQuant was used to analyze this data. This data was searched using MaxQuant which provided the following intensity, localization

probabilities and elution orders (based on retention times). The best Andromeda score was now 315. This data, since it near constantly sampled and fragmented each of the targeted peptides, also allowed some information about the distribution of the modification within the peptide to be suggested.

Mod location	Rel. Intensity	Localization	Elution Order
Y*EAADAILHTPG	46%	.996	7
YE*AADAILHTPG	14%	.89	6
YEA*ADAILHTPG	0.8%	.856	4
YEAA*DAILHTPG	9%	.987	3
YEAAD*AILHTPG	4%	.981	5
YEAADAIL*HTPG	22%	.999	2
YEAADAILH*TPG	4%	.864	1

We made no attempt to define within any of the residue sites where the modification was located. Coverage of the protein in the digestion conditions was not comprehensive. For example, the modified peptide was the most N-terminal peptide detected and is bracketed by two unobserved stretches.