Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

SUPPORTING MATERIALS

Small Molecule Peptidomimetic Trypsin Inhibitors: Validation Of An EKO Binding Mode, But With A Twist

Rui-Liang Lyu, Shaon Joy, Charles Packianathan, Arthur Laganowsky,* and Kevin Burgess*

Department of Chemistry, Texas A & M University, Box 30012, College Station, TX 77842-3012, USA

E-mail: burgess@tamu.edu

Contents

A. Abbreviations

abs/s	absorptions per second
Asc	ascorbate
BAEE	benzoyl-L-arginine ethyl ester
BAPNA	benzoylarginine-*p*-nitroaniline (CAS # 911-77-3, Sigma Aldrich)
Boc	tert-butyloxycarbonyl
bzam	benzylamine
bzmd	benzamidine
bzmdz	benzimidazole
CuAAC	copper(I)-catalyzed azide-alkyne cycloaddition
DCM	dichloromethane
DIPEA	N,N-diisopropyleythlamine
DMF	N,N-dimethylformamide
DMOPP	dimethyl 2-oxopropylphosphonate
DMSO	dimethyl sulfoxide
EDC·HCl	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	ethylenediaminetetraacetic acid
EKO	Exploring Key Orientations
ELS	evaporative light scattering
ESI-MS	electrospray ionization mass spectrometry
Fmoc	fluorenylmethyloxycarbonyl
HATU	hexafluorophosphate azabenzotriazole tetramethyl uranium
HOBt	hydroxybenzotriazole
HPLC	high performance liquid chromatography
K_d	disassociation constant
LC-MS	liquid chromatography–mass spectrometry
µW	microwave
Nap	naphthyl
NMR	nuclear magnetic resonance
Pd/C	palladium on activated carbon
PPI	protein-protein interaction
RMSD	root mean square deviation
rt	room temperature
Tf	trifluoromethanesulfonyl; triflic
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIPS	triisopropylsilane
Tris	tris(hydroxymethyl)aminomethane
Ts	*p*-toluenesulfonyl; tosyl

B. Materials and Instruments

Materials. Reagents were purchased at a high commercial quality (typically 97% or higher) and used without additional purification.

Proteins. Bovine pancreatic trypsin used for kinetic assays was purchased from Sigma Aldrich (T1426, ≥10,000 BAEE units/mg); that used for crystallography was from Alfa Aeser (J63688).

Reaction conditions and solvents. All reactions, unless otherwise stated, were carried out under inert atmosphere with dry solvents under anhydrous conditions. Glassware for anhydrous reactions was dried in an oven at 140 °C for minimum 6 hours prior to use. Dry solvents were obtained by passing the previously degassed solvents through activated alumina columns.

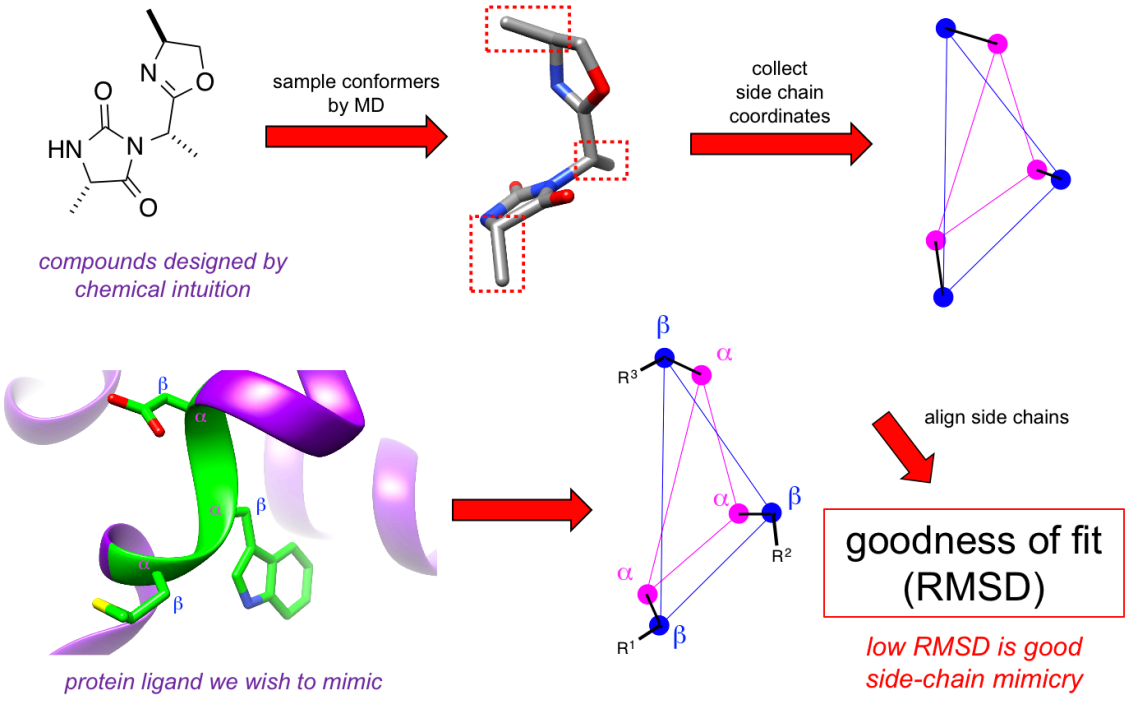
NMR. High field NMR spectra were acquired using either a Bruker Avance III 400 MHz or an Avance 500 MHz cryoprobe system; data were calibrated using residual non-deuterated solvent as an internal reference. The following abbreviations were used to explain the multiplicities: s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, dd = double doublet, dt = double triplet, dq = double quartet, m = multiplet, br = broad.

Other instruments. Both analytical and preparative HPLC uses Agilent's 1260 Infinity II LC systems with water/acetonitrile mobile phase. HPLC analyses uses a C18 column and UV and ELS detectors. Mass spec (ESI-MS) data were collected on triple-stage quadrupole instrument in a positive mode. Kinetic assays were performed using a BioTek Synergy H4 microplate reader.

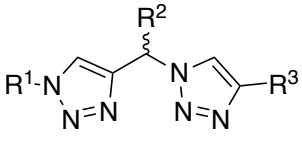
C. Molecular Modeling

Exploring Key Orientations (EKO)

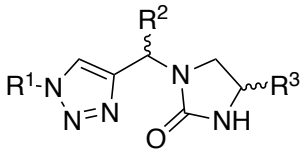
The procedure of using EKO to evaluate peptidomimetics designs is detailed in our previous publications.[1, 2] Specifically, hypothetical designs of peptidomimetic cores are installed with methyl side chains, and their possible conformers are sampled using a molecular dynamics routine (QMD). All conformers within 3.0 kcal/mol from the lowest energy conformer are deemed preferred conformers, and they are clustered into families by the Cα-Cβ coordinates of the amino acid side chains it carries. The conformer having lowest energy in each family was selected as a representative. For a protein-protein interaction (PPI) of interest, the interface residues of each protein are first identified according to the distance of the residues to their protein partners and their side chain orientations. The representative conformers of each family are then systematically aligned to every possible set of interface side chains of the proteins. The "goodness" of an alignment is then characterized by its Root Mean Square Deviation (RMSD).



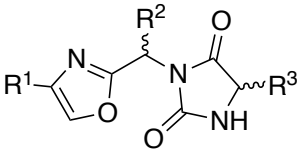
Examples of compounds evaluated by EKO



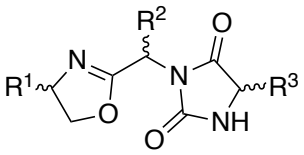
TT-mer



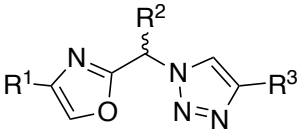
KB8



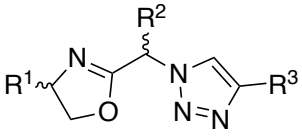
JT016



JT001



JT023



JT025

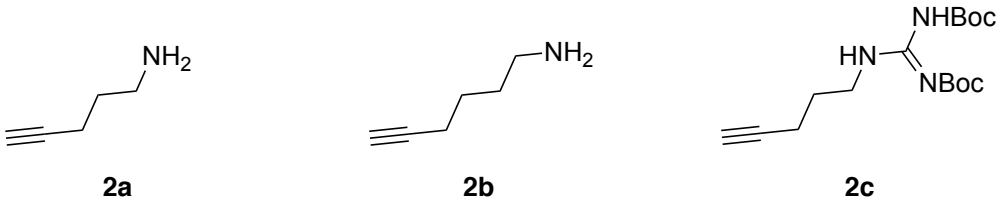
Calculations To Assess Relative Energies of Conformers

Molecular mechanics calculations using the Merck molecular force field (MMFF94) were performed by systematically varying the dihedral angles of the two rotatable bonds in the backbone in one degree increments. Energies was compared for the favorable and unfavorable conformers using a Python library RDKit (www.RDKit.org).

Conformers generated in RDKit were then imported into Gaussian 16, then their energies were simulated in the gas phase (DFT at B3LYP/3-21g).

D. Synthesis

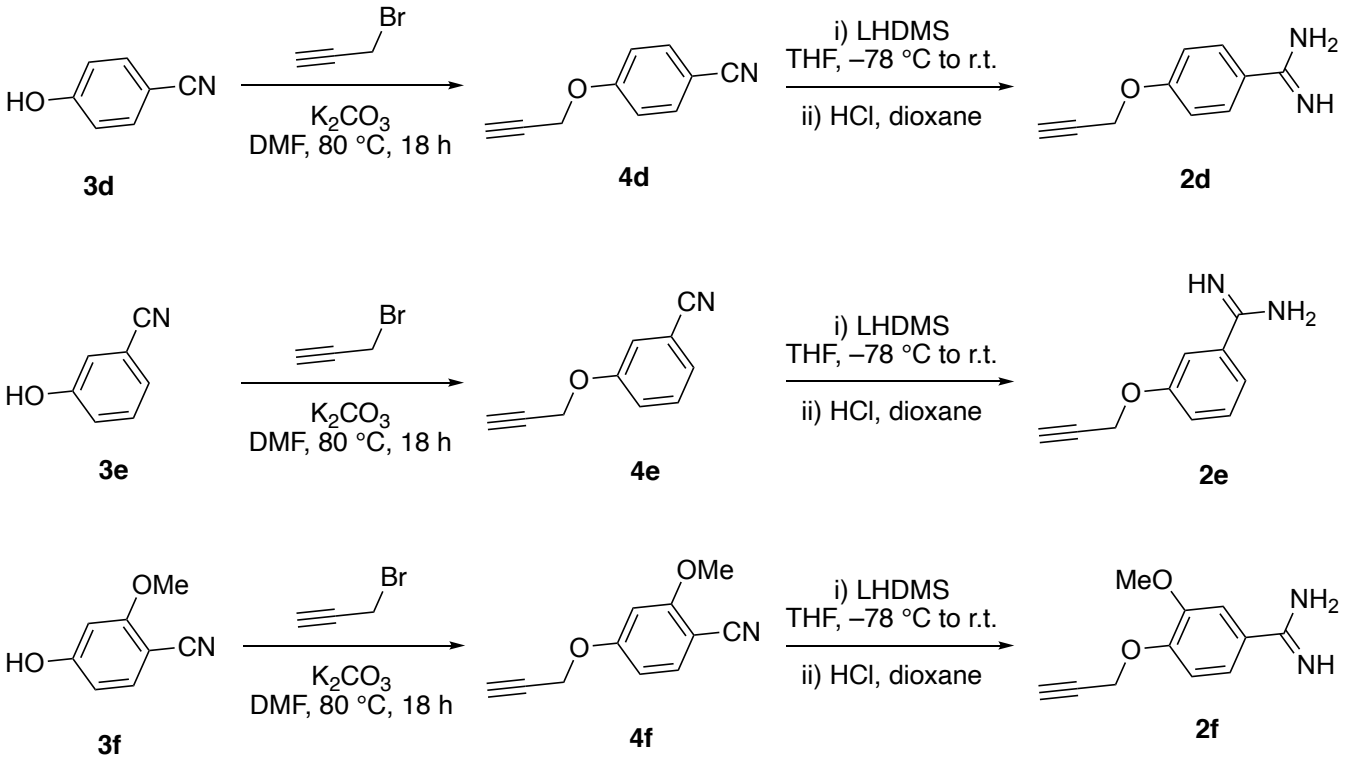
Synthesis of the alkynes



2a **2b** **2c**

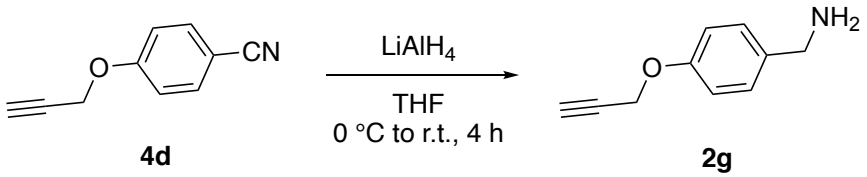
Pent-4-yn-1-amine (**2a**) and hex-5-yn-1-amine (**2b**) and were purchased from commercial sources.

Synthesis of N,N'-bisBoc-1-(pent-4-yn-1-yl)guanidine (**2c**): To a solution of the **2a** (12.26 mmol, 1.0 eq) in THF (100 mL), DIPEA (61.3 mmol, 3.0 eq) and N,N'-pyrazole-carboxamidine (36.78 mmol, 3.0 eq) were added and let to stir at room temperature until completion. The solvent was removed and the residue was diluted with DCM and extracted with NH_4Cl. Aqueous layer was further extracted with DCM (x2). The combined organic layers were washed with brine, 5% HCl and dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure. The product was sufficiently pure to use without further workup.



Synthesis of propynyloxy benzonitrile (**4d – f**): To a stirred solution of the corresponding cyano phenol (**3d – f**, 20.0 mmol, 1.0 eq) in 32 mL DMF was added potassium carbonate (22.2 mmol, 1.12 eq). The resulting suspension was stirred at 80 °C temperature for 30 mins, after which it was cooled down to room temperature. At this point, propargyl bromide (21.8 mmol, 1.09 eq) was added to the mixture and stirred at 80 °C for another 18 h. When the reagent consumption was complete, the mixture was diluted with DCM (20 mL), and the precipitate was filtered off. The residual solvent was removed under reduced pressure. The alkyne was sufficiently pure to use without further workup.

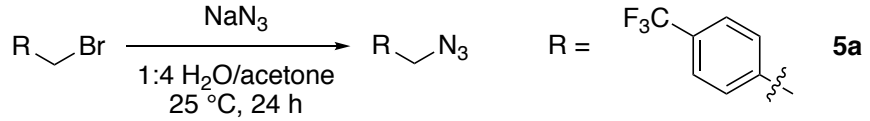
Synthesis of propynyloxybenzamidine (**2d – f**) The nitrile intermediate (**4d – f**) was suspended in anhydrous THF (~2 mL per mmol of intermediate) and cooled to -78 °C. LHDMS (1.0M in THF solution, 26.05 mmol, 5.0 eq) was added gradually to the stirred mixture at -78 °C. The mixture was then stirred for 16 h at room temperature, by which time TLC indicated complete reaction. The reaction mixture was cooled to 0 °C and quenched with a solution of 4.0 M HCl in dioxane (1 equiv/LHDMS). The solvent was immediately evaporated under reduced pressure and high vacuum was applied. EtOH was added to the dry residue, followed by addition of EtOH-HCl until the mixture became acidic to pH paper. The crude was purified by prep-HPLC to give the product amidine.

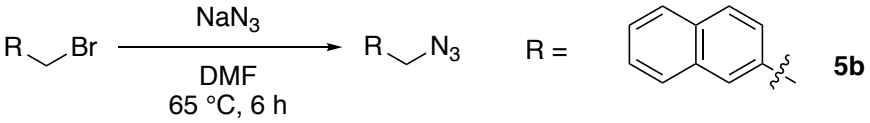


4d **2g**

Synthesis of 3-propynyloxy benzylamine (**2g**) To a solution of the **4d** in dry THF (20 mL) at 0 °C was added a solution of lithium aluminum hydride (1.0 M in THF, 18.0 mmol, 4.0 eq). The reaction mixture was stirred at room temperature for 4 h or until completion. The reaction mixture was then diluted with ether and cooled to 0 °C. Water was added slowly followed by 15% aqueous sodium hydroxide solution. The mixture was then warmed to room temperature and stirred for 15 mins. At this point, some anhydrous magnesium sulfate was added and stirred for another 15 mins. Finally, the mixture was filtered to remove the salts, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography on silica gel to afford 3-propynyloxybenzylamine as a pale yellow oil (0.86 g) at a yield of 76%.

Synthesis of the organic azides

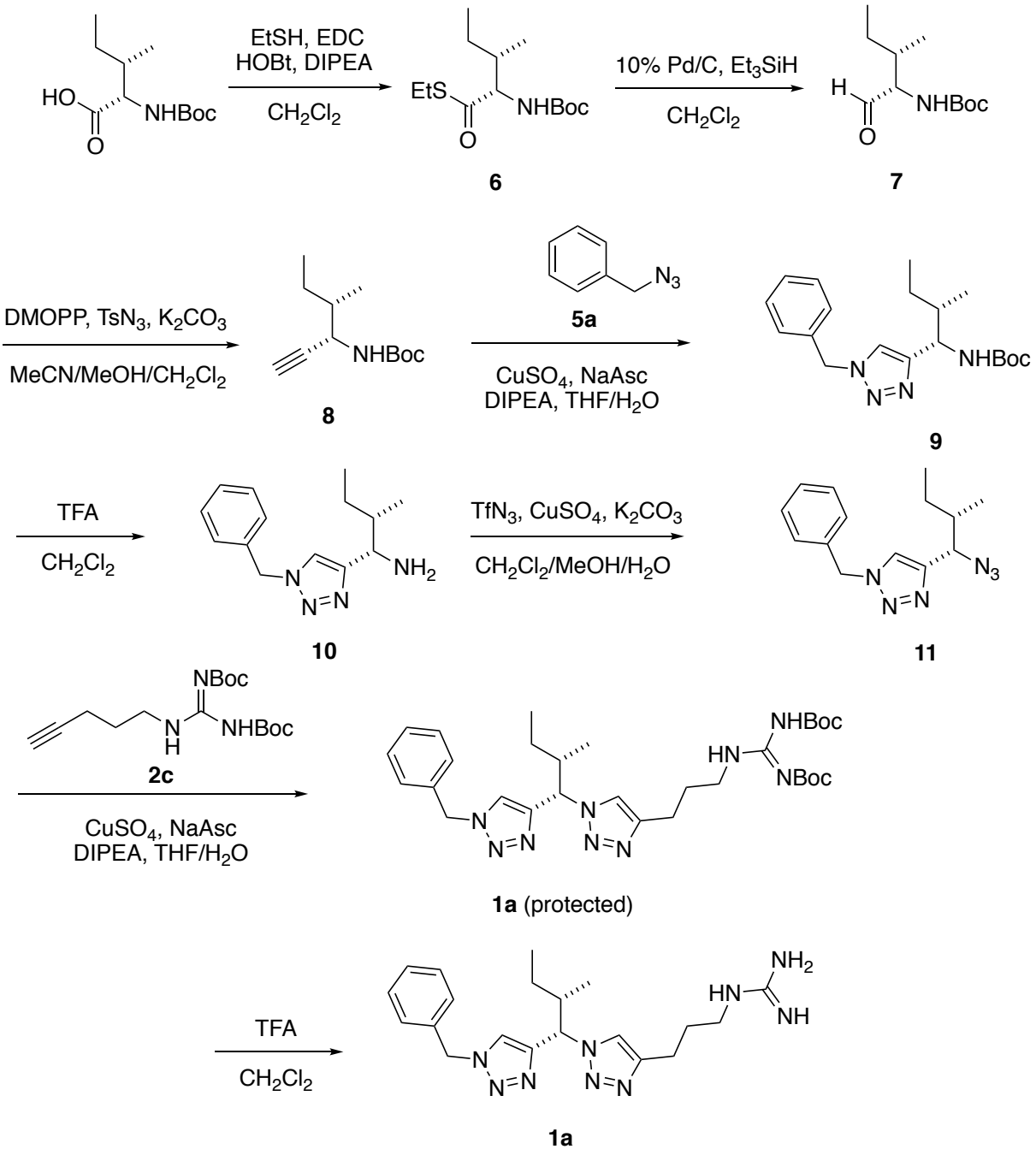




General Procedure A: To a stirred solution of the corresponding bromide (5 mmol, 1.0 eq) in a 100 mL water/acetone mixture (1:4) was added sodium azide (7.5 mmol, 1.5 eq). The resulting suspension was stirred at room temperature for 24 hours, and periodically monitored by TLC. When the reagent consumption was complete, the mixture was diluted with DCM (20 mL), and the organic phase was separated. The aqueous phase was extracted with DCM (3 X 20 mL), and the combined organic extracts were dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure. The azide **5a** was obtained in 73% yield and was sufficiently pure to use without further workup.

General Procedure B: To a stirred solution of the corresponding bromide (5 mmol, 1.0 eq) in anhydrous DMF (10 mL) in N_2 atmosphere was added sodium azide (7.5 mmol, 1.5 eq) at 25 °C. The reaction mixture was then stirred for 6 h at 65 °C. After reaction completion, the mixture was quenched with water and extracted with DCM (3 X 10 mL). Combined organic extracts were dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure. The azide **5b** was obtained in 86% yield and was sufficiently pure to use without further workup.

Solution phase synthesis of TT-mers



In this section, we use the compound **1a** (R^1 = 2'-Nap, R^2 = Ile, R^3 = Arg) as an example to illustrate the solution phase synthesis of TT-mers, but the approach is applicable to all compounds in this series.

Synthesis of tosyl azide. 7.6 g tosyl chloride (40 mmol) and 3.1 g sodium azide (50 mmol, 1.2 eq) were dissolved in 100 mL 1:1 acetone/H_2O and stirred overnight at 25 °C. With acetone removed under vacuum, the aqueous portion was extracted with 3 x 50 mL ethyl acetate. The organic phase was dried over anhydrous $MgSO_4$ and evaporated under vacuum. The product is a colorless liquid at 25 °C and a colorless solid at –20°C.

Synthesis of the thioester (**6**). To a solution of 4.8 g Boc-L-isoleucine (20 mmol, 1 eq) in 100 mL DCM were added 5.8 g EDC·HCl (1.5 eq), 5.1 g HOBt (1.5 eq), 5.3 mL DIPEA (3 eq), and 1.8 mL ethanethiol (1.2 eq). The reaction mixture was stirred overnight at 25 °C. The reaction mixture was then concentrated under vacuum, redissolved in ethyl acetate, and washed with 3 x 100 mL 10% citric acid, 3 x 100 mL saturated $NaHCO_3$, 100 mL saturated NaCl. The resulting solution was dried over anhydrous $MgSO_4$ and evaporated under vacuum. The resulting product was a light yellowish sticky liquid and was used in the next step without further purification.

Synthesis of the aldehyde (**7**). This was performed as described by Fukuyama et al.[3] Specifically, to a 50 mL Schlenk flask were added 10 mL dichloromethane solution of ~20 mmol **6** and 1.06 g 10% Pd/C (0.05 eq). 10 mL Et_3SiH (3 eq) was added to this suspension drop wise over a period of 30 min, then the mixture was gently stirred for another 2 h. The reaction mixture was filtered through Celite, concentrated under vacuum and used in the next step immediately, without further purification.

Synthesis of the alkyne (**8**). This was performed as described by Bestmann *et al.*[4, 5] Specifically, 3.3 mL dimethyl 2-oxopropylphosphonate (DMOPP, 1.2 eq), 3.7 mL tosyl azide (1.2 eq), and 8.3 g K_2CO_3 were dissolved in 150 mL acetonitrile and reacted for 2 h. **7** from the last step was dissolved in 60 mL MeOH and added to the mixture. The mixture was then stirred at 25 °C overnight. Following the reaction, with the organic solvents evaporated, the mixture was suspended in 100 mL water and extracted by 3 x 50 mL ether. The crude was concentrated and purified by column chromatography with 1:9 ethyl acetate/hexane as mobile phase. The alkyne showed up on TLC as a bright spot with R_f = 0.4 when stained with potassium permanganate. Product collected was a light yellow sticky liquid with an yield of 2.0 g (46% over three steps).

Installation of the R^1 group through CuAAC reaction. 1.0 eq **8**, 1.0 eq **5d**, 0.2 eq $CuSO_4·5H_2O$, 1,0 eq sodium ascorbate, and 10.0 eq DIPEA was dissolved in 5.0 mL H_2O. Small amount of THF may be added if the compounds were not soluble. The reaction was stirred under nitrogen for 12 h, before the THF portion was evaporated (if present) and the product was extracted by ethyl acetate. The product **9** is typically pure enough without further purification.

Deprotection of the Boc group. **9** was dissolved in ~5 mL 1:1 TFA/DCM and stirred for 1 h. The solution was concentrated with a rotary evaporator, and redissolved in DCM and evaporated multiple times for complete removal of TFA. The crude product **10** was directly used in the next step.

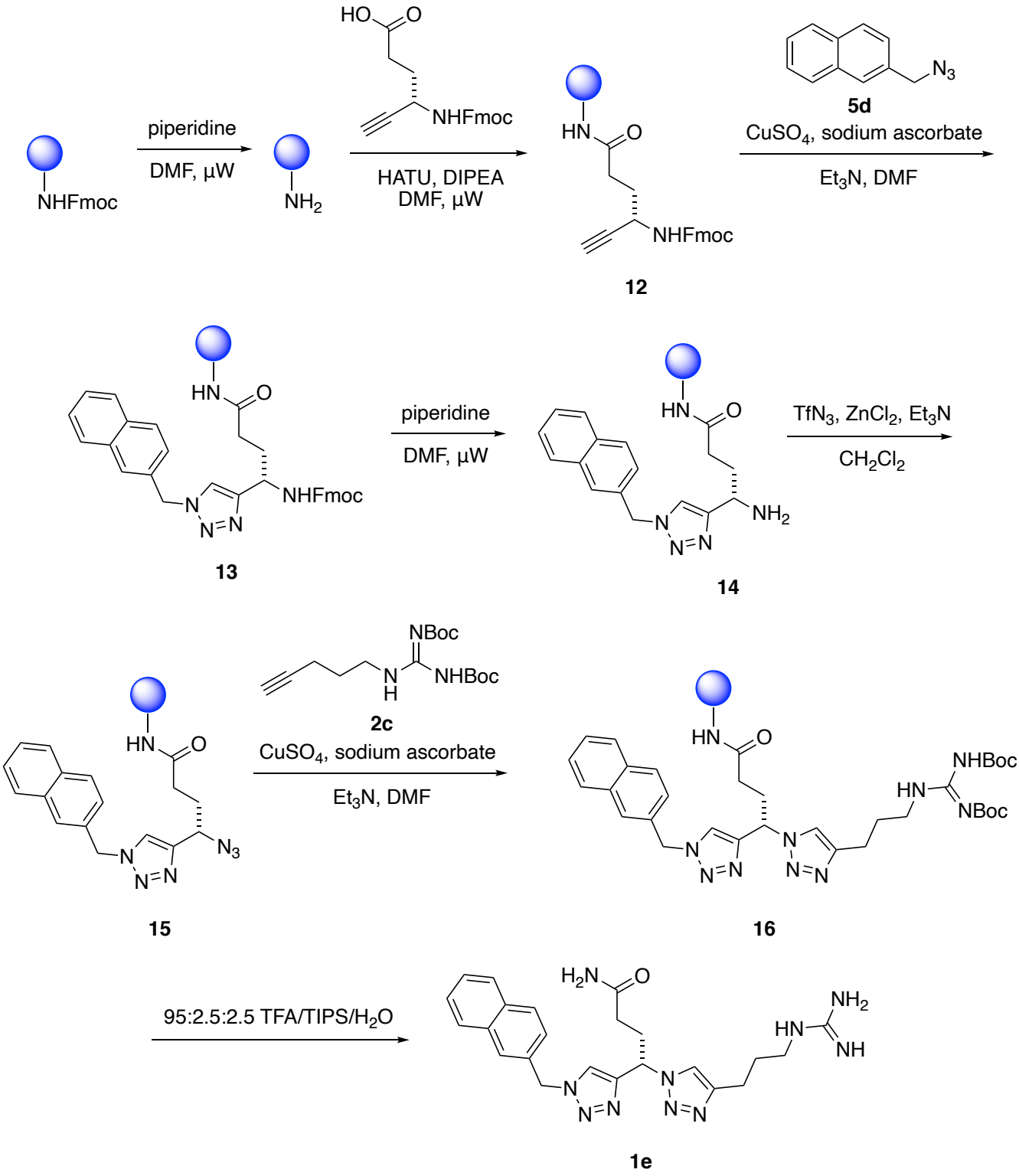
Synthesis of the azide (**11**). Triflic azide (TfN_3) was prepared by reacting triflic anhydride with sodium azide in 1:1 DCM/water. Following the completion of the reaction, the organic phase was separated, and the aqueous one was extracted with DCM for 2 – 3 times. To the combined organic phase was then added the methanol solution of **10**, as well as an aqueous solution 0.2 eq. $CuSO_4·5H_2O$ and 3 eq K_2CO_3. The reaction was stirred at room temperature overnight and purified by column chromatography. Product collected was a yellowish oil with an yield of 0.65 g (53% over three steps).

Installation of the R^3 group through CuAAC reaction. Same procedure as described before.

Final deprotection and purification. The compound may have side chains protected by Boc or Alloc groups. Boc groups were deprotected as described in earlier sections. For Alloc groups, the compound was dissolved in ~5 mL DCM along with 0.2 eq $Pd(PPh_3)_4$ and 15 eq Ph_3SiH and reacted under microwave at 38 °C for 5 min. This process may be repeated one more time to ensure complete deprotection. Finally, the product was purified by reverse phase HPLC.

Counterion exchange. The product was a trifluoroacetate salt at this point. We found this counterion alone could reduce activities of some serine proteases to a certain extent, hence the compound needed to be switched to a chloride salt for accurate measurement of its inhibitory activity. This was completed by re-dissolving the compound in 0.1 mM HCl water solution and lyophilizing. Additions of small portions of acetonitrile may be required to dissolve the compound. This process was performed three times to ensure complete removal of all trifluoroacetate counterions.

Solid phase synthesis of TT-mers



In this section, we use the compound **1e** (R^1 = 2'-Nap, R^2 = Gln, R^3 = Arg) as an example to illustrate the solid phase synthesis of TT-mers, but the approach is applicable to all compounds in this series.

Deprotection of Fmoc groups. 0.30 g TentaGel S RAM resin (0.22 mol/g loading capacity) was weighed in a

fritted syringe. The beads were swelled sequentially with MeOH (2 mL, 5 min), DCM (2 mL, 10 min) and DMF (2 mL, 15 min) on a shaker. To the syringe was added 2 mL 20% piperidine in DMF, and the reaction mixture was stirred under microwave (100 W, 75 °C) for 3 min. The solution was drained, and the beads were washed with 2 mL DMF five times. Removal of Fmoc groups was confirmed by ninhydrin test.

Amide coupling. 0.093 g Fmoc-Glu-yne (0.266 mmol, 4 eq.), 0.190 g HATU (0.500 mmol, 0.35 M), and 0.1 mL DIPEA (0.57 mmol, 8.7 eq.) were dissolved in 2 mL DMF. This solution was added to the syringe, which was then stirred under microwave (100 W, 75 °C) for 10 min. The solvent was drained, and beads were washed as described in the previous step. Completion of reaction was confirmed by ninhydrin test.

Installation of the R^1 group through CuAAC reaction. 0.061 g **5d** (0.33 mmol, 5 eq.), 0.017 g $CuSO_4 \cdot 5H_2O$ (0.068 mmol, 1 eq.), 0.013 g sodium ascorbate (0.066 mmol, 1 eq.), and 0.1 mL DIPEA (0.57 mmol, 8.7 eq.) were dissolved in 2 mL DMF. This solution was added to the syringe and shaken for 12 h. The solution was drained, and the beads were washed with 2 mL 0.05 M EDTA solution three times, and 2 mL DMF five times.

On-bead azide synthesis. 0.9 mL Tf_2O (5.3 mmol, 8.1 eq.) and 1.7 g NaN_3 (26.0 mmol, 40 eq.) were dissolved in 20 mL 1:1 DCM/water mixture and stirred rigorously for 4 h. The organic phase was separated and the aqueous phase was washed with DCM three times. The organic phases were combined, washed by saturated Na_2CO_3 and brine, dried over anhydrous $MgSO_4$, filtered, and concentrated to approximately 2 mL by blowing with argon. At the meantime, Fmoc group on the amine was deprotected as described as before. 0.018 mg $ZnCl_2$ (0.13 mmol, 0.2 eq.) and 0.4 mL Et_3N were dissolve in 0.5 mL MeOH, which was added to the TfN_3 solution prepared before. The reaction mixture was added to the syringe and shaken for 12 h. The solution was drained and beads were washed in the same way as described in the *azide-alkyne click reaction* step.

Installation of the R^3 group through CuAAC reaction. This is performed in the same way as described in the R^1 group installation step.

Cleavage and purification. 9.5 mL TFA was mixed with 0.25 mL TIPS and 0.25 mL water. 2 mL of this solution was added to the syringe and shaken for 4 h. The solution was collected by pushing through the frit, and concentrated by blowing with argon. This crude mixture was diluted in 1:1 acetonitrile/water, purified by reverse phase preparative HPLC, and lyophilized to obtain final product as a yellow-colored oil.

Switching of the counterion. This was performed in the same way as described in the solution phase synthesis.

E. Trypsin Binding Assay with DL-BAPNA as Substrate

Theoretical background

Kinetic assays were performed as described in literature.[6, 7] Specifically, a non-UV active substrate, benzoyl-L-arginine-*p*-nitroaniline (L-BAPNA) is incubated with trypsin (or other trypsin family proteases). Hydrolysis of this substrate produces *p*-nitroaniline, which absorbs strongly around 410 nm. Thus, the reaction rate can be calculated by the increase in UV absorption over time at this wavelength.

However, BAPNA is only commercially available as a racemic mixture, where only the L-isomer is hydrolyzed by trypsin family proteases, and the D-isomer is a competitive inhibitor.

Consider the case where a substrate (L-BAPNA in this case) is incubated with a catalyst (trypsin) and two competitive inhibitors (D-BAPNA and the inhibitor whose K_d we want to measure). The Michaelis-Menten equation for this system is

$$v = -\frac{\mathrm{d}[\mathrm{S}]}{\mathrm{d}t} = \frac{v_{\max}[\mathrm{S}]}{K_\mathrm{M}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{d}} + \frac{[\mathrm{I}']}{K_\mathrm{d}'}\right) + [\mathrm{S}]}$$

where $[\mathrm{S}]$ is the concentration of L-BAPNA; $[\mathrm{I}]$ and K_d are the concentration and disassociation constant to trypsin of the inhibitor of interest; $[\mathrm{I}']$ and K_d' are those of D-BAPNA; $v_{\max}$ is the reaction rate achieved at saturating substrate concentration; and K_M is Michaelis-Menten constant, i.e. the substrate concentration at which reaction rate is half of $v_{\max}$. Taking the reciprocal of each side of the equation,

$$\begin{aligned}
\frac{1}{v} &= \frac{K_\mathrm{M}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{d}} + \frac{[\mathrm{I}']}{K_\mathrm{d}'}\right) + [\mathrm{S}]}{v_{\max}[\mathrm{S}]} \\
&= \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{d}}\right) + \frac{K_\mathrm{M}\frac{[\mathrm{I}']}{K_\mathrm{d}'} + [\mathrm{S}]}{v_{\max}[\mathrm{S}]} \\
&= \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{d}}\right) + \frac{K_\mathrm{M} + K_\mathrm{d}'}{v_{\max}K_\mathrm{d}'} \\
&= \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{d}}\right) + \mathrm{const}
\end{aligned}$$

Here we used the fact that in a racemic mixture, $[\mathrm{S}] = [\mathrm{I}']$. By cancelling out these terms we see the second term of the equation is a constant regardless of substrate concentration and the inhibitor of interest. The plot $1/v - [\mathrm{I}]$ is therefore a straight line.

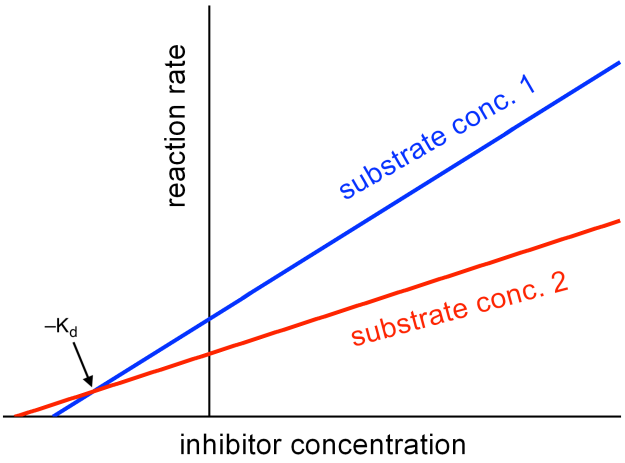
Consider performing this reaction at two substrate concentrations $[\mathrm{S}]_1$ and $[\mathrm{S}]_2$. This gives two initial reaction rates v_1 and v_2, which are described by

$$\begin{aligned}
\frac{1}{v_1} &= \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]_1}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{d}}\right) + c \\
\frac{1}{v_2} &= \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]_2}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{d}}\right) + c
\end{aligned}$$

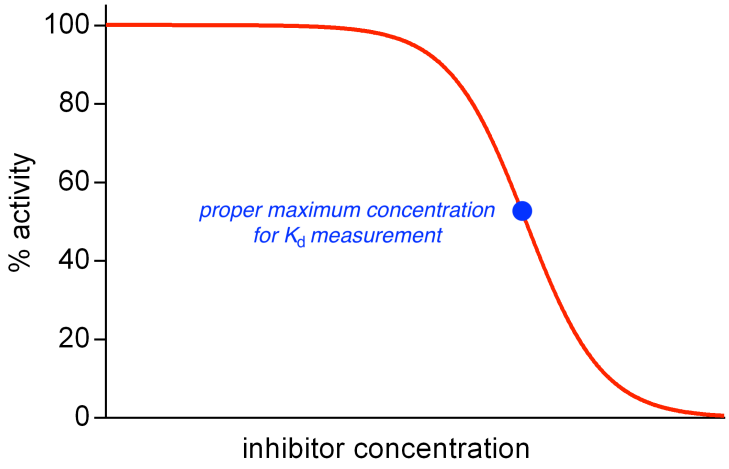
Solving this system of equations, we see these two lines intersect at a hypothetical negative inhibitor concentration $[\mathrm{I}] = -K_\mathrm{d}$. In practice, we perform the reaction at three different substrate concentrations, and

take the negative of the average of the three intersections as our experimental K_d.

Notice that this fact holds regardless of the unit used to express the reaction rate. Therefore, we directly used the change in the *reads of UV detector* over time (absorptions per second, abs/s), instead of having to convert this to the rate of change in actual *product concentration* over time through a calibration curve.



K_d measurement



A typical dose response curve and the proper maximum concentration for K_d measurement

Experimental procedures

Buffer. 6.1 g Tris and 1.5 g $CaCl_2 \cdot 2H_2O$ are dissolved in 500 mL deionized water; this solution is then adjusted to pH 8.0 by dropwise addition of 6 M HCl.

Enzyme stock solutions. 4.8 mg bovine pancreatic trypsin (10,000 BAEE units/mg) are dissolved in 20 mL 1.2 mM HCl; 28 mg bovine thrombin (88 BAEE units/mg) are dissolved in 5 mL 0.1 M NaCl solution. The two stock solutions are mixed gently and kept on ice.

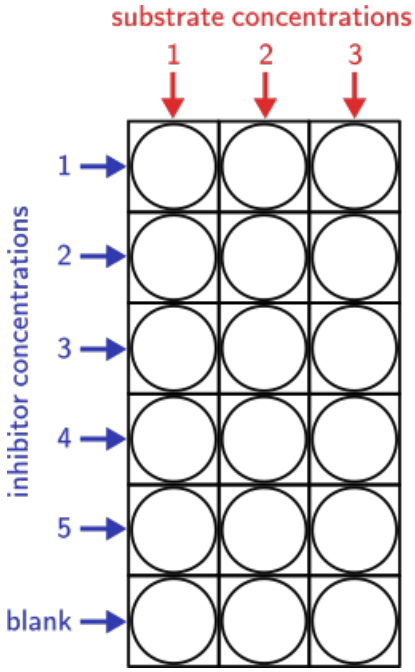
Substrate stock solutions. 43 mg DL-BAPNA hydrochloride are dissolved in 5 mL DMSO to give 20 mM stock solution. This is subsequently diluted to 10, 5, 3, 2, and 1 mM stock solutions.

Inhibitor stock solutions. Inhibitors of interest are dissolved in DMSO at various concentrations. We also choose

a known serine protease inhibitor, benzamidine hydrochloride, as a reference inhibitor. 1.57 mg of this compound are dissolved in 10 mL DMSO to give 1 mM stock solution.

Dose response experiment is first performed to find the proper inhibitor concentration range for K_d measurement. The inhibitor stock solution is diluted repeatedly to give a series of various concentrations. The reaction is set up using a typical 96-well plate. To each well is added 140 μL buffer, 20 μL enzyme stock solution, and 20 μL inhibitor solution. The plate is shaken at 37 °C for 2 min, then to each well is added 20 μL substrate solution. The concentration of substrate solution should be determined by to the activity of the enzyme. In this work we used 3 mM BAPNA for trypsin (which gives 150 μM of L-BAPNA in the reaction mixture) and 20 mM BAPNA for thrombin (1 mM of L-BAPNA in reaction mixture). The plate is immediately placed in the plate reader, and absorption at 410 nm of each well is monitored for 5 min, with 10 s interval. The dose response curve (figure above for example) is plotted and the inhibitor concentration that reduces reaction rate roughly by half is selected as the maximum concentration to use in K_d measurement. Inhibitor concentrations too larger or smaller than this may either make the reaction to be too slow, or nor distinguishable from the uninhibited rate; either way the error in K_d measurement is increased.

K_d measurement is performed in a similar manner as dose response experiment. The inhibitor stock solution is diluted to the concentration selected in the dose response experiment, which is then further diluted to 1/5, 2/5, 3/5, and 4/5 of this concentration, respectively. Each run of K_d measurement is performed in a 3 × 6 section of the 96-well plate. To each well is added 140 μL buffer, 20 μL enzyme stock solution, 20 μL inhibitor, and 20 μL buffer; each row in the 3 × 6 section is added a different inhibitor concentration, and each column a different substrate concentration (figure below). The reaction is monitored for 5 min similar as before. The measurement is repeated for at least two times for each inhibitor, and K_d is calculated as described in the "Theoretical background" section.



Each data point was collected three times in 1 9 x 6 well format. The plots on which the K_d data were obtained are based on the average of those three readings per experiment.

F. X-Ray Crystallography

Crystallization of trypsin bound 1l

Bovine pancreatic trypsin was purchased from the Alfa Aser (cat # J6368803). The crystals of trypsin bound **1l** were grown from a 30 mg/mL solution of trypsin dissolved in the MES buffer at pH 6.0 containing 4 mM **1l** and 1 mM $CaCl_2$. The mixture was incubated at room temperature for 20 minutes before crystallization. Crystals were obtained by mixing 2.0 µL of protein solution with an equal volume of reservoir solution consisting 1.9 M $(NH_4)_2SO_4$, 50 mM MES pH 6.0, 2 mM **1l**. Crystals obtained at 293 K using hanging drop Linbro 24 well plates (HR3-110, Hampton Research, Aliso Viejo, CA) were transferred to a cryoprotectant solution (reservoir solution containing 20% glycerol) and flash-frozen in liquid nitrogen for data collection.

X-ray data collection, processing and refinement

Crystallographic data were collected on a Rigaku 3HR X-ray source equipped with an R-Axis IV++ detector at a crystal-to-detector distance of 250 mm with 0.5° rotation angle (720 image frames) about φ using a RIGAKU CCD detector under standard cryogenic conditions (100 K) at the home R Axis IV++ source. Diffraction data were indexed, integrated, and scaled with the XDS program.[8] Crystals diffracted to 2.38 Å and belonged to space group C_2 with cell dimensions as follows: a = 189.27 Å, b =70.32 Å, c = 137.74 Å and α = β = 90.0° and γ = 133.0°. Evaluation of crystal packing parameters indicated that the lattice can accommodate four molecules in the asymmetric unit with a solvent content of 63.0% and a Matthews coefficient of 3.34 Å Da^{-1}).[9] The structure was determined by molecular replacement method using Phaser.[10] The atomic coordinates of the trypsin (PDB ID 1s0r) were used as the starting model. Structure refinement of each data set was performed with Phenix[11] and REFMAC5[12] implemented with the CCP4 suite.[13] Approximately 5% of the reflections were used for the test set. The refined model required manual adjustment to improve the fit to the experimental electron density using the program COOT.[14] At this stage restrained refinement was carried out, and the final R-factor and R_{free} converged to 20.7% and 24.7% for the trypsin bound **1l**. A Ramachandran plot calculated using PROCHECK[15] indicated that 95.5% of the residues are in the most favored region, 4.41% of the residues in the additionally allowed regions and none of the residues in the disallowed region. Final data collection, processing, refinement statistics, and Protein Data Bank accession codes are given in the following table. Figures were prepared using PyMOL.[16]

data-collection and refinement statistics for Trypsin bound **1l**; values in parentheses are for the highest resolution bin

Data collection	
X-ray source	R-Axis IV++
Wavelength (Å)	1.54
Detector	CCD
Resolution (Å)	40.0 – 2.38 (2.46 – 2.38)
Space group	C121
Unit cell parameters (Å, °)	a = 189.27 b = 70.32 c = 137.74 α = β = 90.0 γ = 133.0
Observed reflections	162059 (10581)
Unique reflections	52631 (3703)
Multiplicity	3.07 (2.85)
<I> / σ(I)	13.40 (4.0)
Completeness (%)	98.51 (93.60)
CC (1/2) (%)	99.7 (92.8)
∇R_{merge} (%)	6.6 (26.9)
ΔR_{meas} (%)	8.0 (33.0)
B-factor Wilson plot (Å^2)	27.52
Refinement	
R_{work}/R_{free} (%)	0.2074 (0.2717) / 0.2477 (0.3487)
Number of non-hydrogen atoms	7296
Number of protein atoms	6516
Number of ligand atoms	164
Number of water molecules	616
Protein residues	892
Mean B value (Å^2)	29.70
RMS deviation from ideal	
Bond lengths (Å)	0.008
Bond angles (°)	0.97
Ramachandran plot	
Most favored regions (%)	95.59
Additionally, allowed regions (%)	4.41
Outliers (%)	0.0
PDB code	7jwx

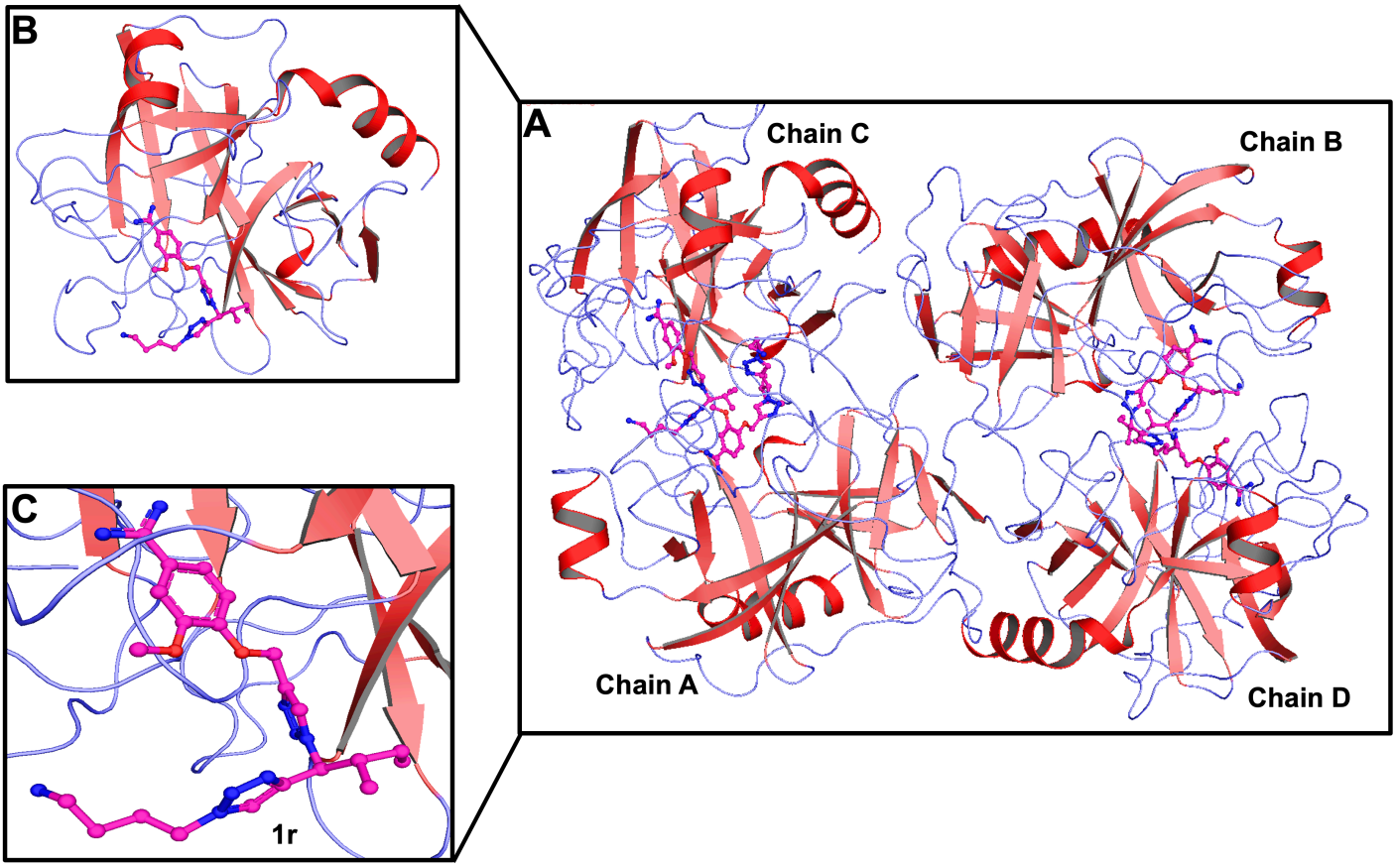
$\nabla R_{\mathrm{merge}} = \frac{\sum_{hkl}\sum_i |I_{i,hkl} - \bar{I}_{hkl}|}{\sum_{hkl}\sum_i |I_{i,hkl}|}$ where $I_{i,hkl}$ is the observed intensity and $\bar{I}_{hkl}$ is the average intensity over symmetry equivalent measurements.

$\Delta R_{\mathrm{work}} = \frac{\sum_{hkl} ||F_{\mathrm{obs}}| - |F_{\mathrm{cal}}||}{\sum_{hkl} |F_{\mathrm{cal}}|}$ and R_{free} is calculated for a random chosen 5% of reflections which were not used for structure refinement and R work is calculated for the remaining reflections.

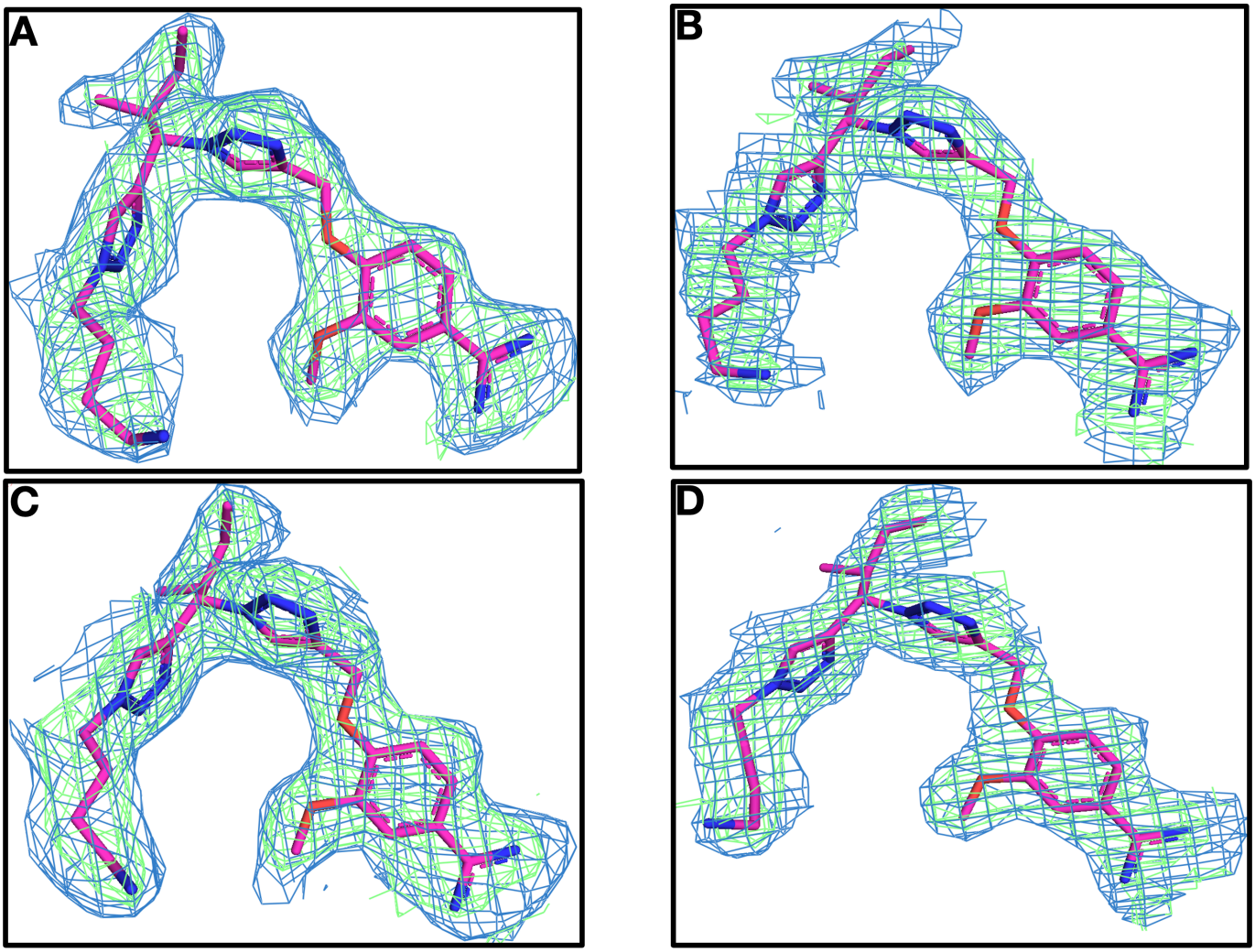
Refinement using Phenix[11]

Ramachandran Plot calculated by PROCHECK[15]

Cartoon diagram representation of trypsin-bound **1l** colored in red (α-helix), salmon (β-sheets) and light blue (γ-turns). **A** The overall structure of trypsin-bound **1l** consists of four molecules in the asymmetric molecules. **B** The expanded view of the chains A bound **1l**. **C** Compound **1l** represented by ball-and-stick and colored magenta (carbon), blue (nitrogen), red (oxygen).

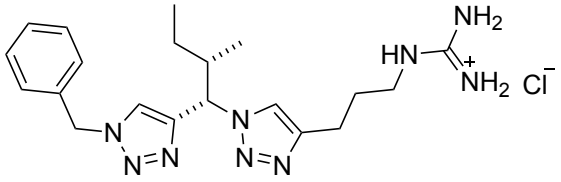


The 2Fo-Fc electron density maps of the trypsin bound **1l** are shown in (**A**) Chain A, (**B**) in Chain B, (**C**) in Chain C and (**D**) in Chain D contoured at 1.0 σ (light blue). The **1l** electron densities are outlined with the Fo- Fc unbiased OMIT map contoured at 2.5 (light green).

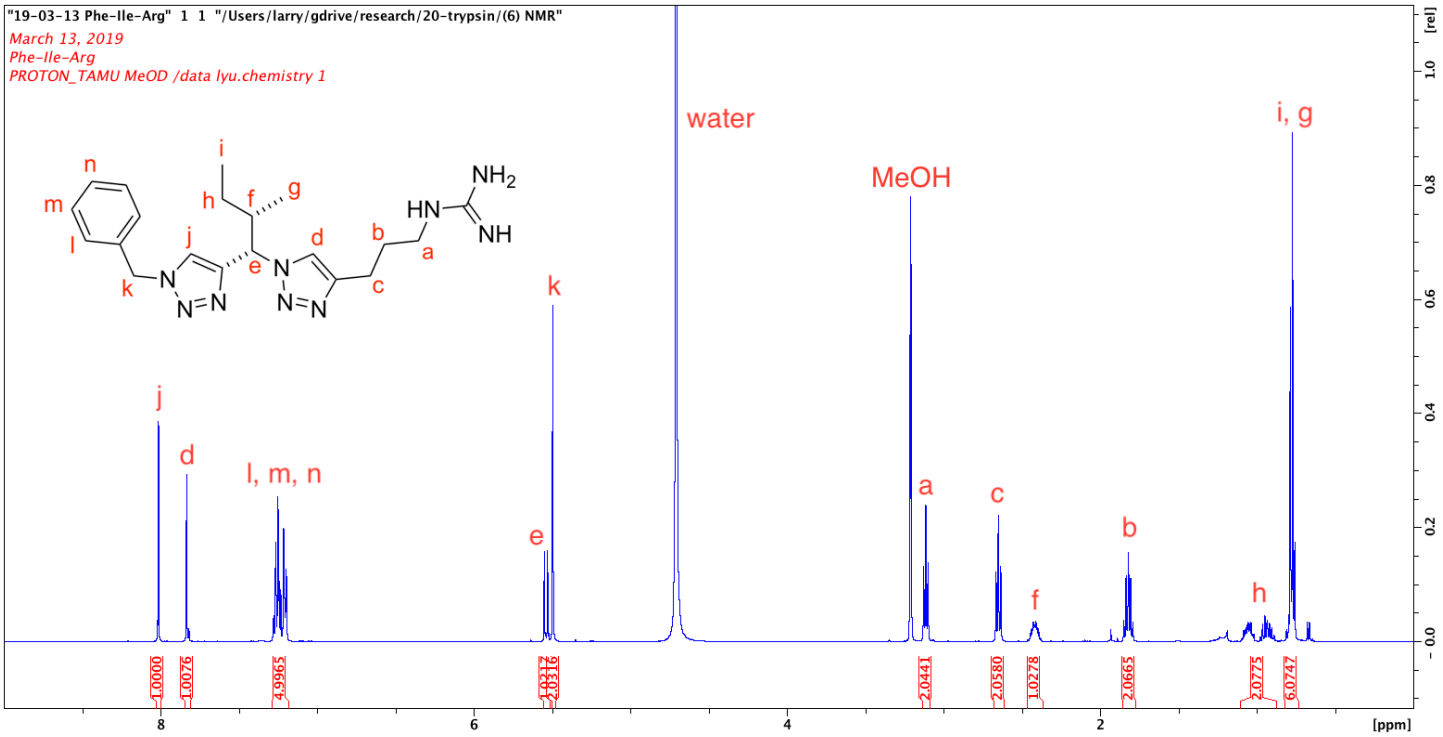


G. Characterization Data

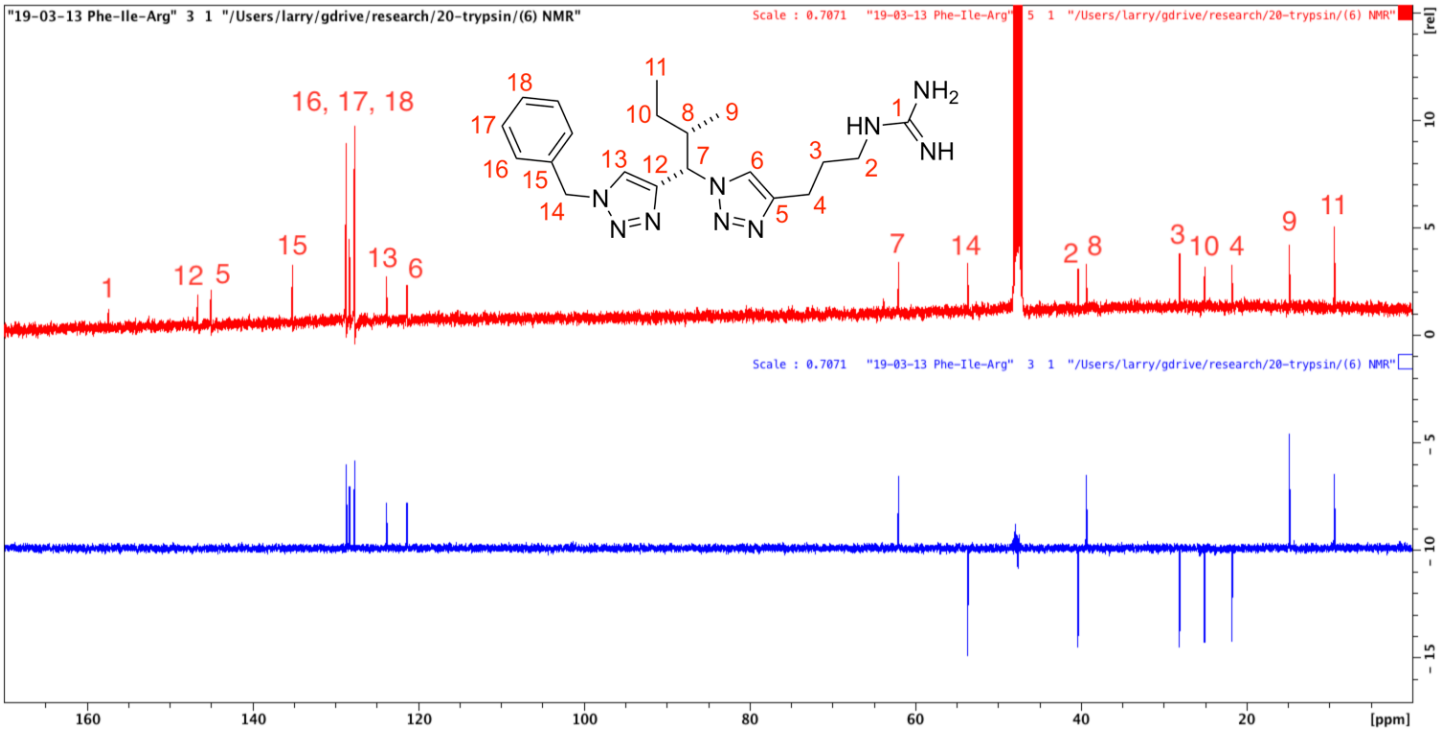
1a



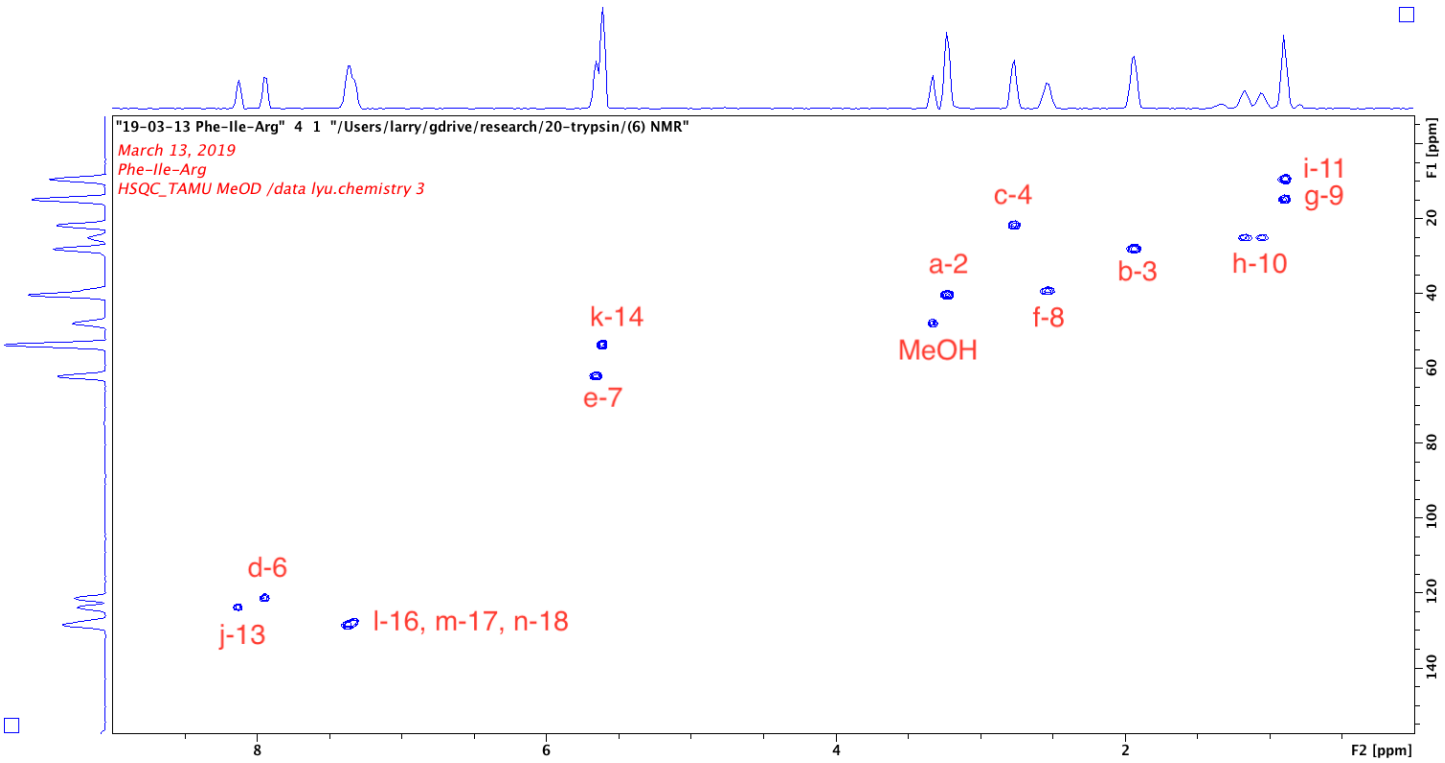
[1]H NMR (d_4-methanol, 500 MHz). δ 8.01 (s, 1H), 7.83 (s, 1H), 7.29-7.19 (m, 5H), 5.54 (d, 1H, *J* = 10.2 Hz), 5.49 (s, 2H), 3.11 (t, 2H, *J* = 7.1 Hz), 2.65 (t, 2H, *J* = 7.5 Hz), 2.42 (m, 1H), 1.82 (m, 2H), 1.10-0.88 (m, 2H), 0.80-0.75 (m, 6H).



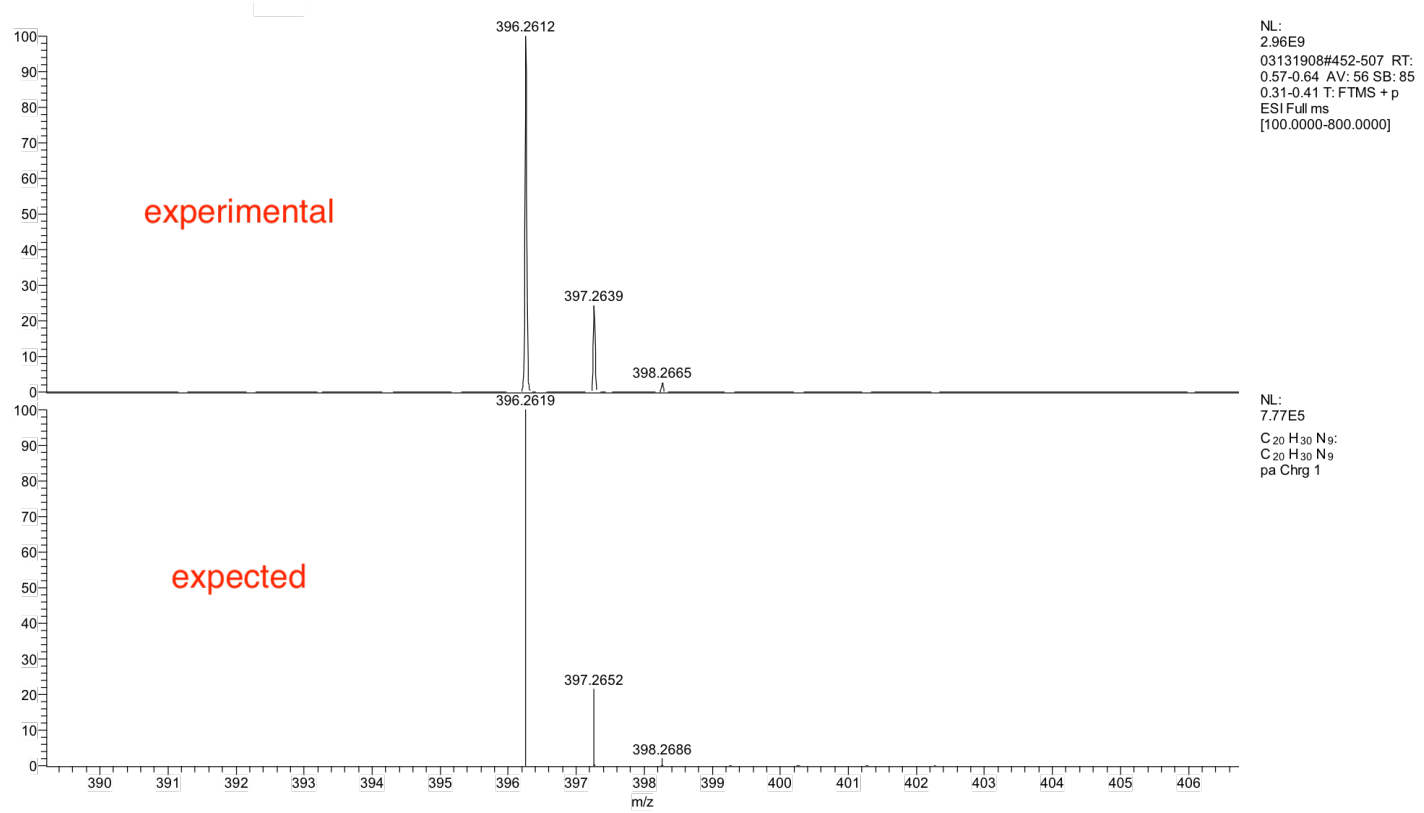
^{13}C NMR & DEPT135 (d_4-methanol, 500 MHz). δ 157.4, 146.6, 145.0, 135.2, 128.7, 128.3, 127.7, 123.8, 121.3, 62.0, 53.7, 40.3, 39.3, 28.1, 25.1, 21.7, 14.8, 9.4.



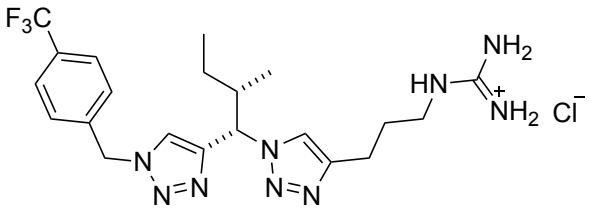
HSQC.



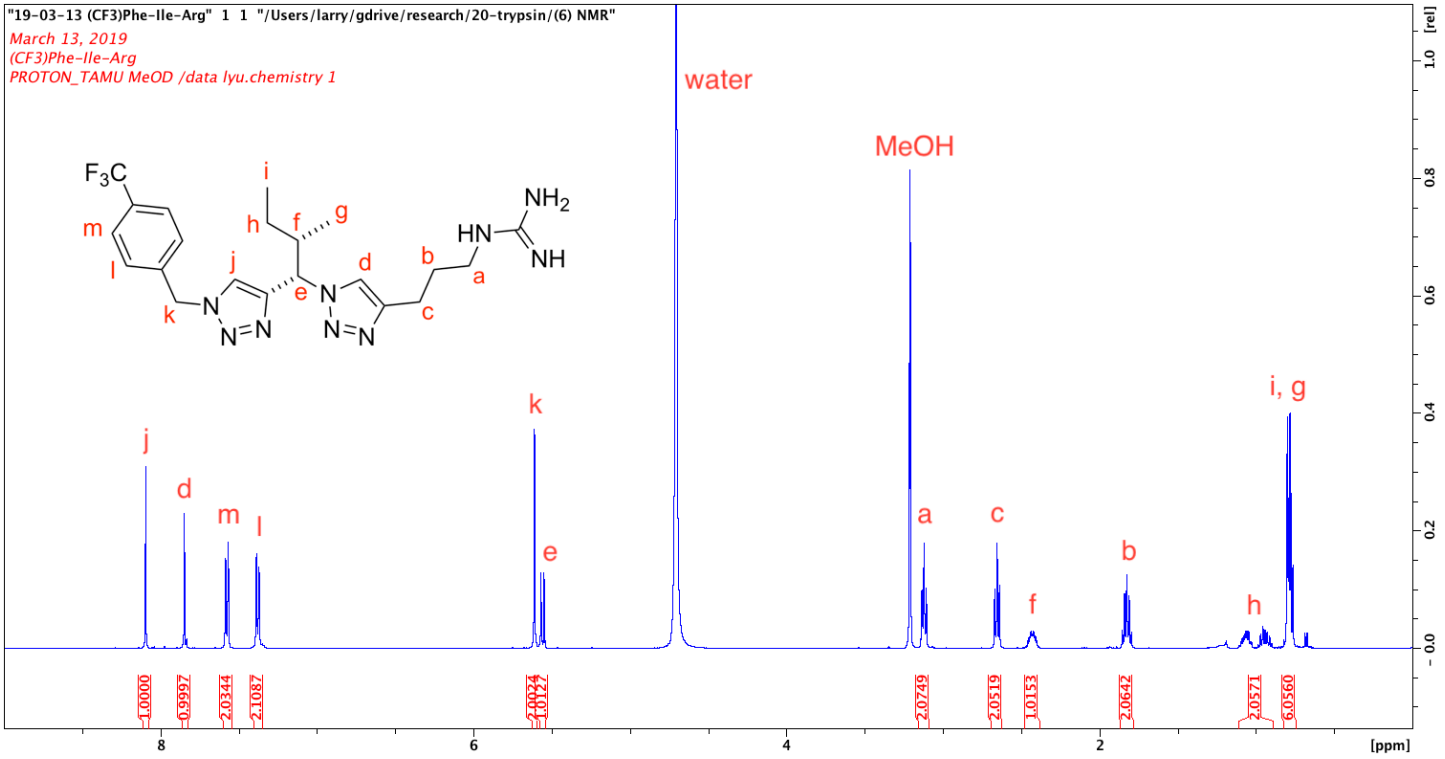
ESI-MS. Expected: 396.2619; measured: 396.2612.



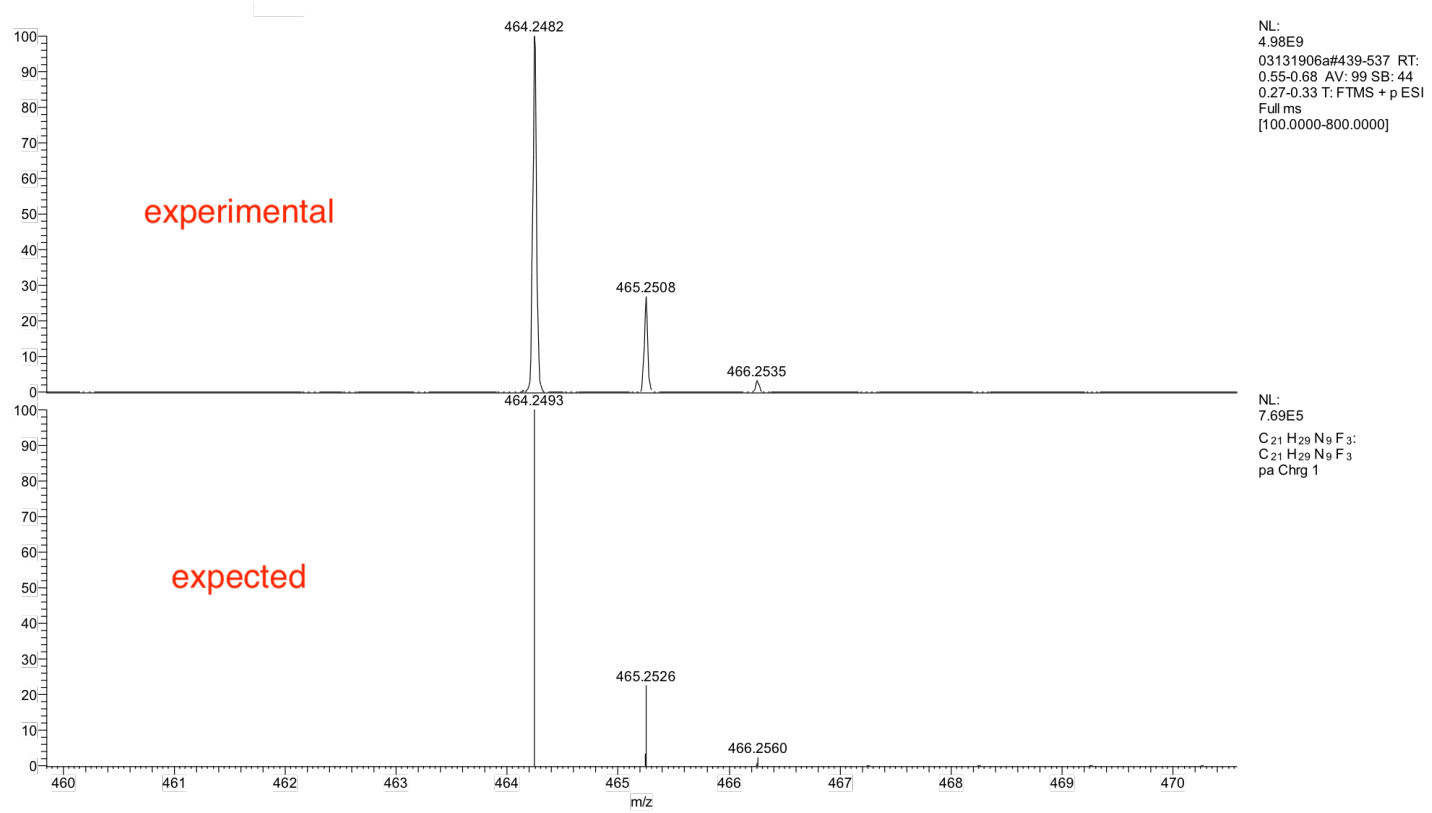
1b



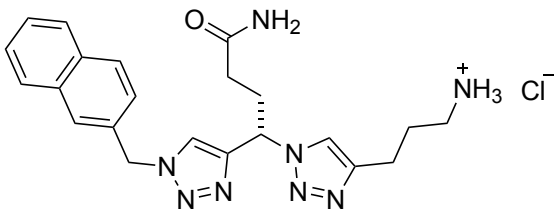
[1]H NMR (d_4-methanol, 500 MHz). δ 8.09 (s, 1H), 7.85 (s, 1H), 7.57 (d, 2H, *J* = 8.2 Hz), 7.38 (d, 2H, *J* = 8.1 Hz), 5.60 (s, 2H), 5.56 (d, 1H, *J* = 10.2 Hz), 3.12 (t, 2H, *J* = 7.2 Hz), 2.65 (t, 2H, *J* = 7.5 Hz), 2.43 (m, 1H), 1.82 (m, 2H), 1.10-0.89 (m, 2H), 0.81-0.77 (m, 6H).



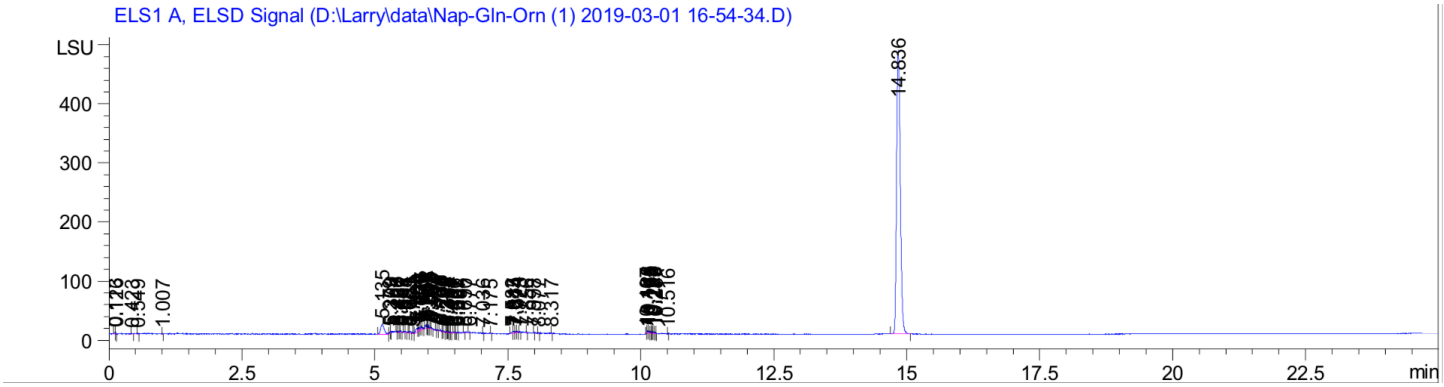
ESI-MS. Expected: 464.2493; measured: 464.2482.



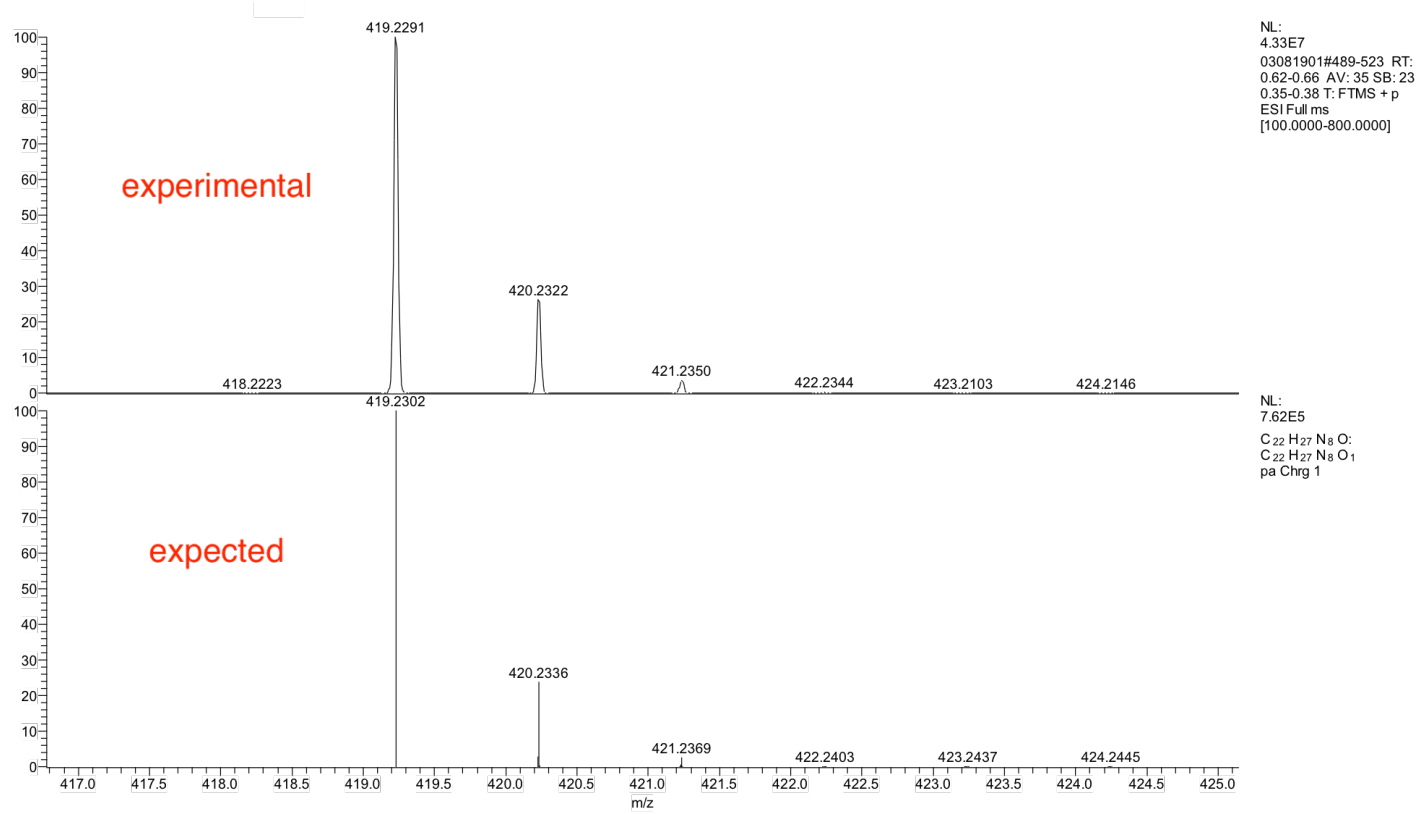
1c



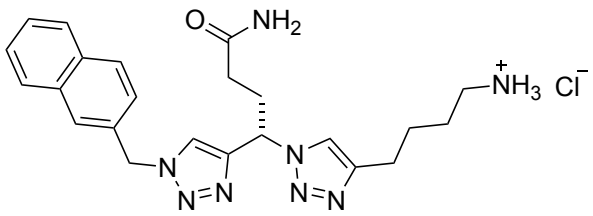
HPLC.



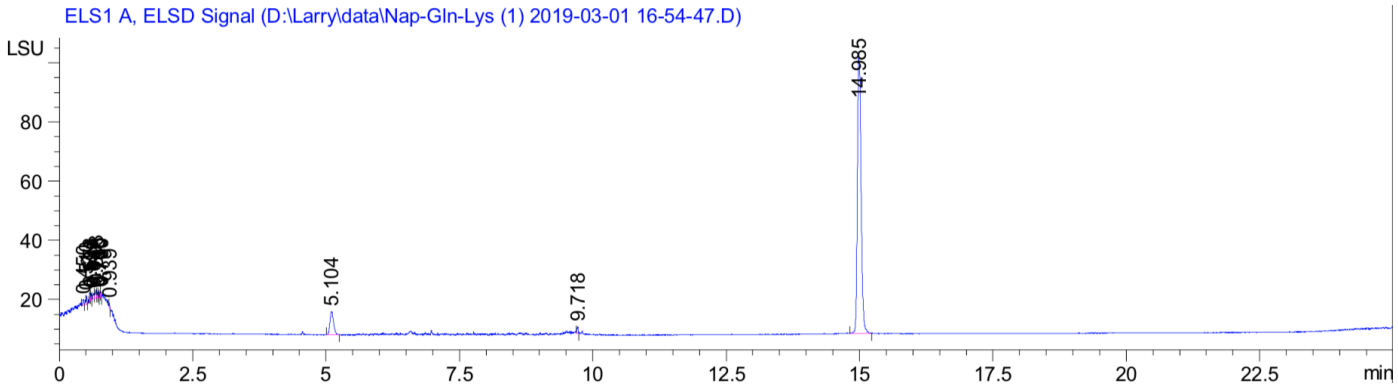
ESI-MS. Expected: 419.2302; measured: 419.2291.



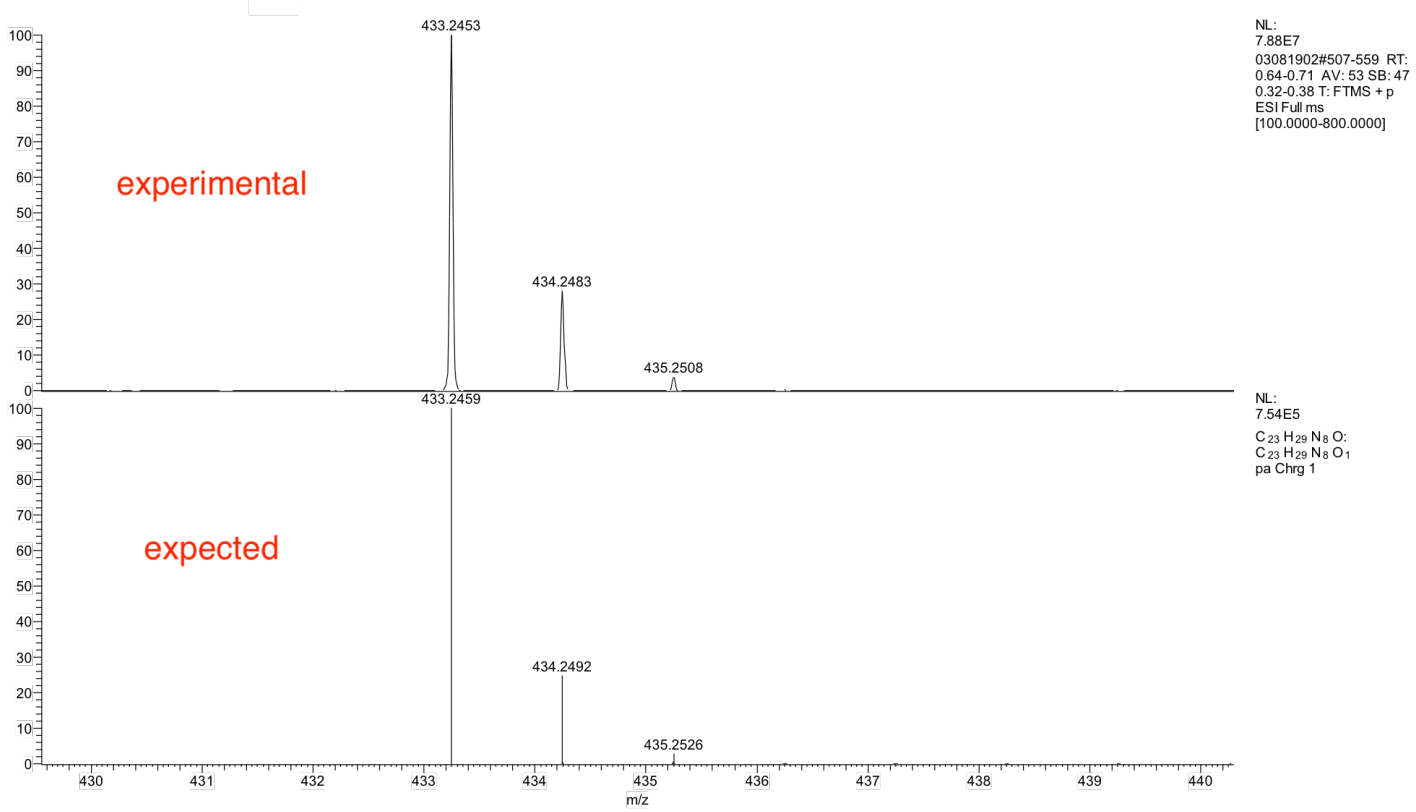
1d



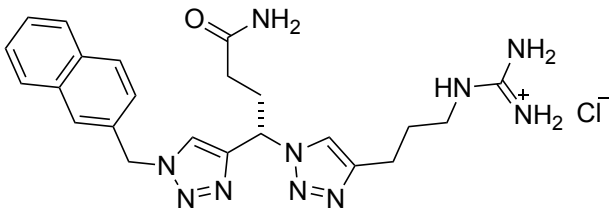
HPLC.



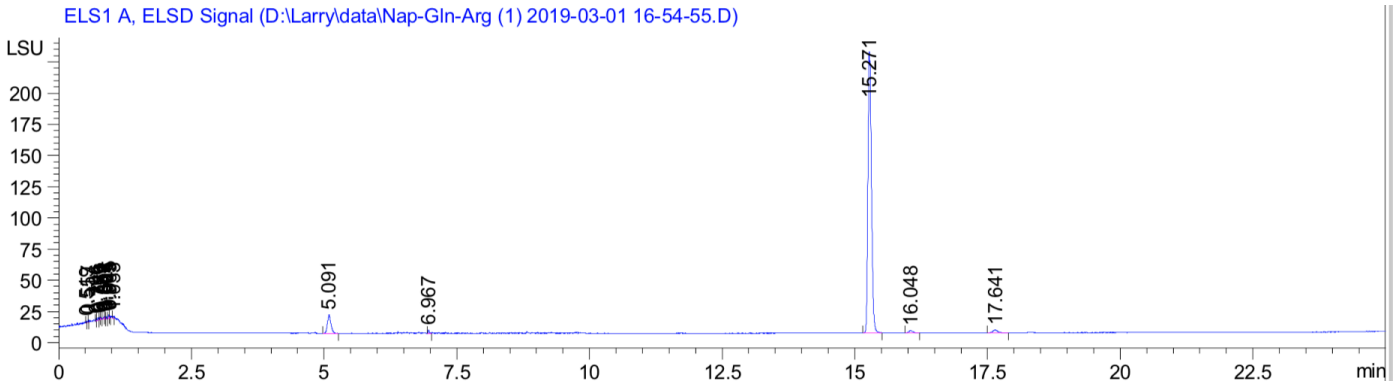
ESI-MS. Expected: 433.2459; measured: 433.2453.



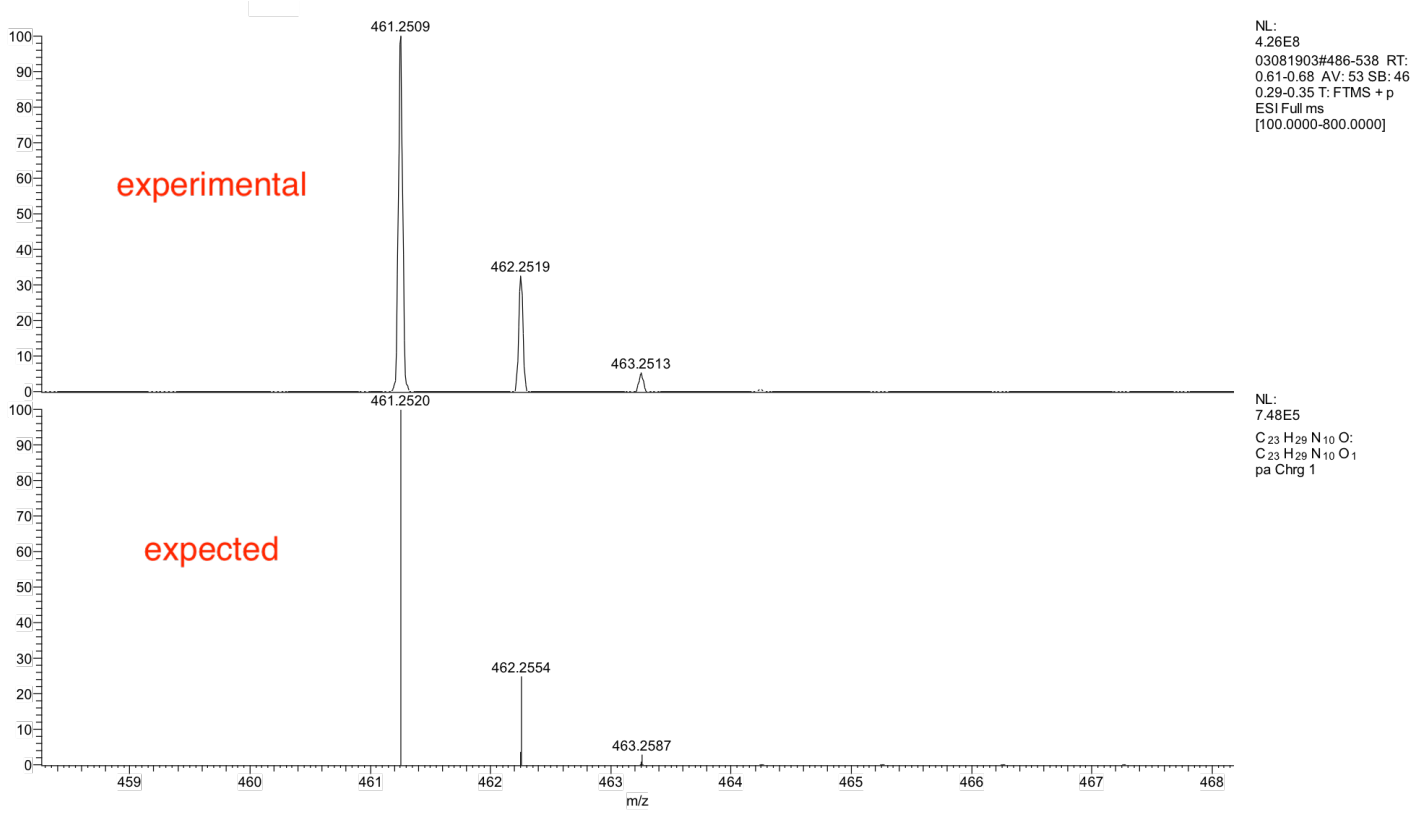
1e



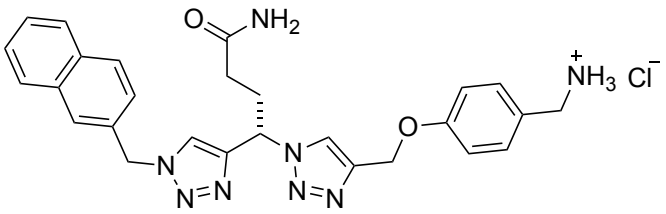
HPLC.



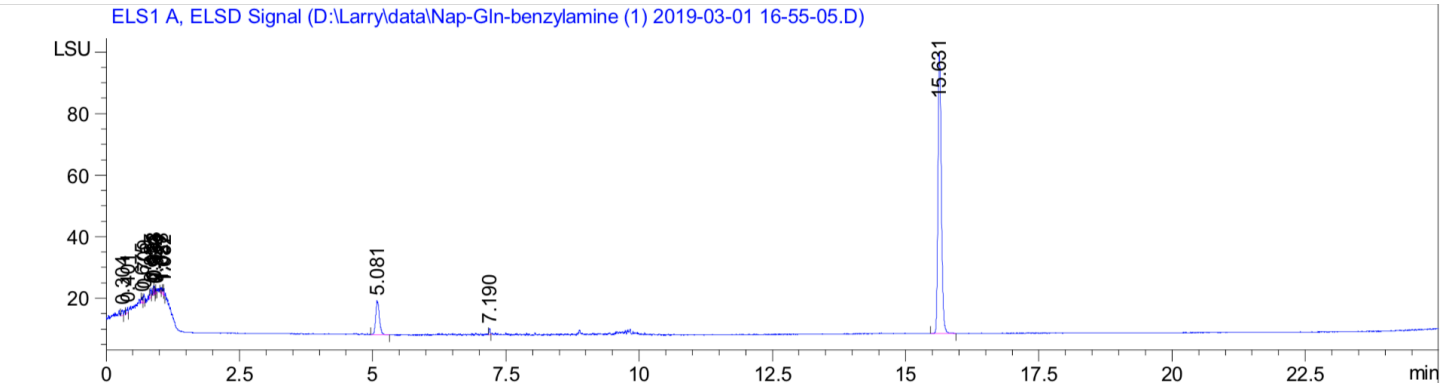
ESI-MS. Expected: 461.2520; measured: 461.2509.



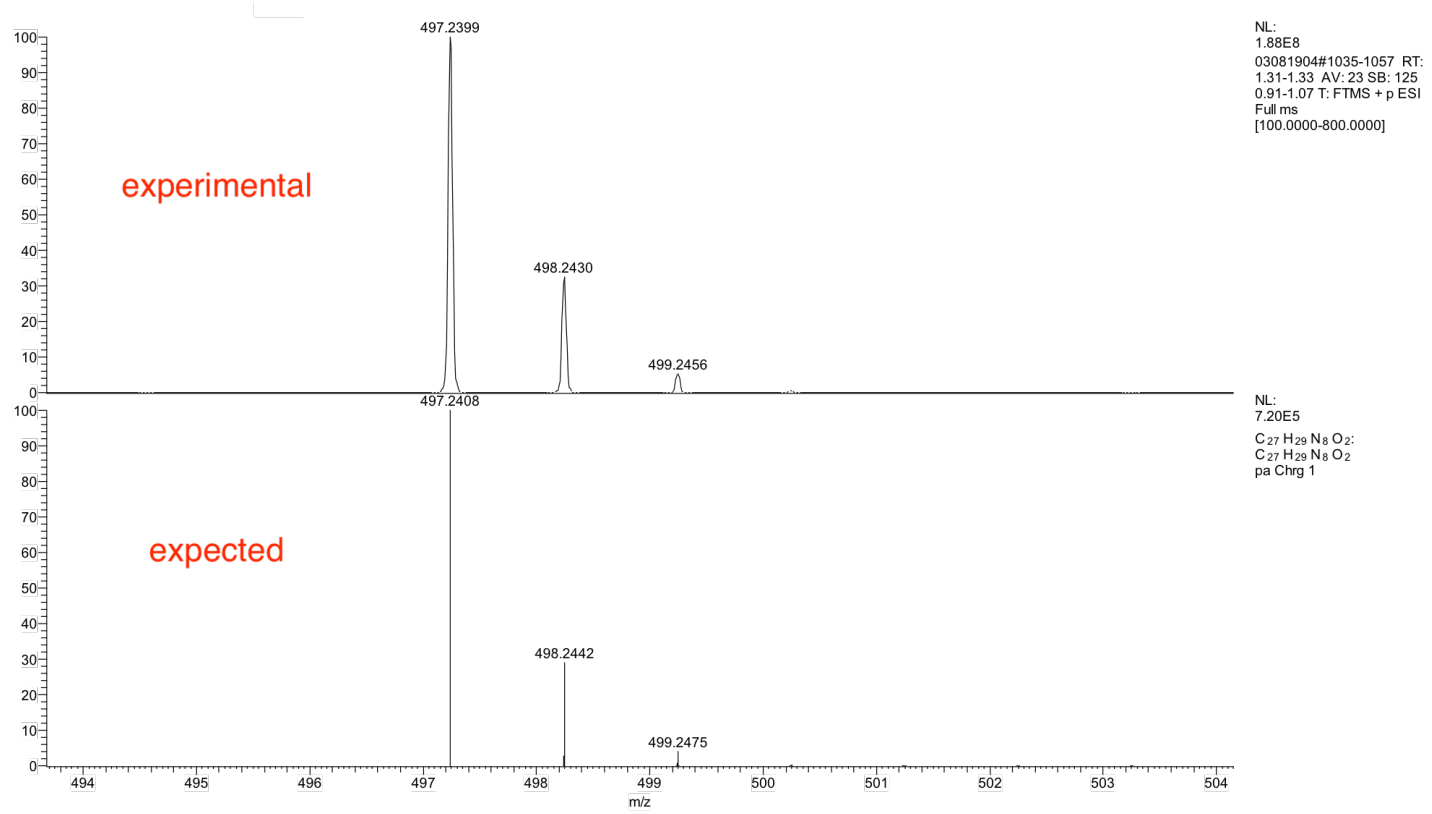
1f



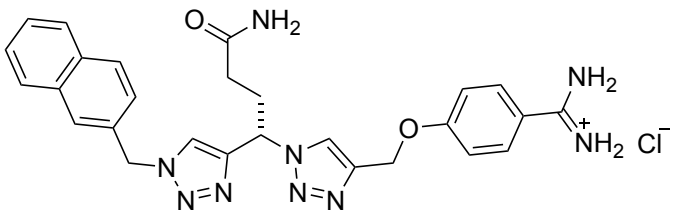
HPLC.



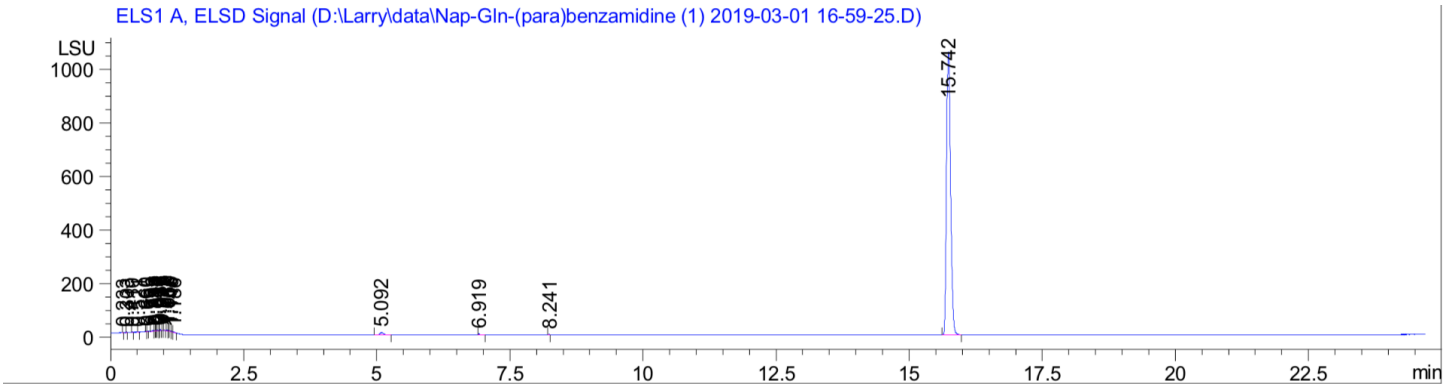
ESI-MS. Expected: 497.2408; measured: 497.2399.



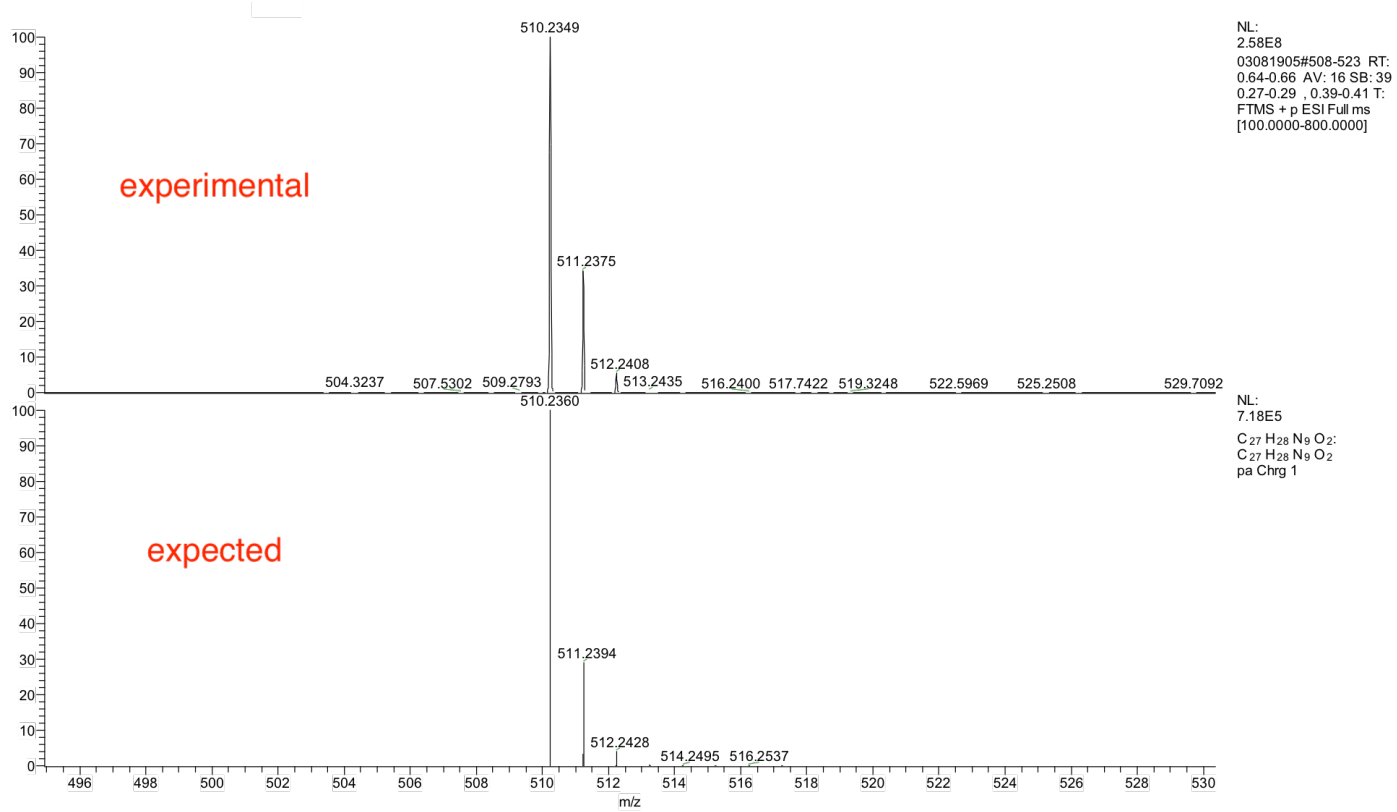
1g



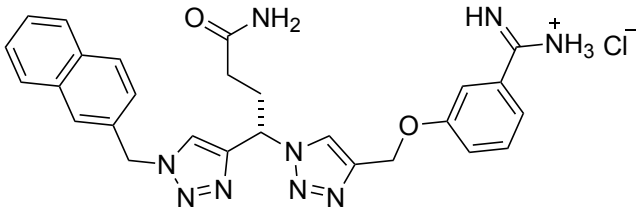
HPLC.



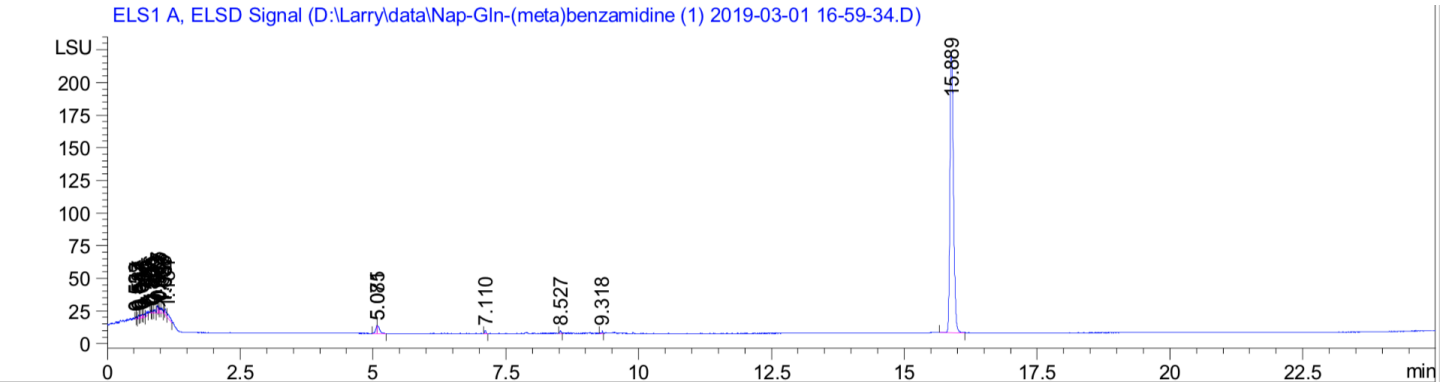
ESI-MS. Expected: 510.2360; measured: 510.2349.



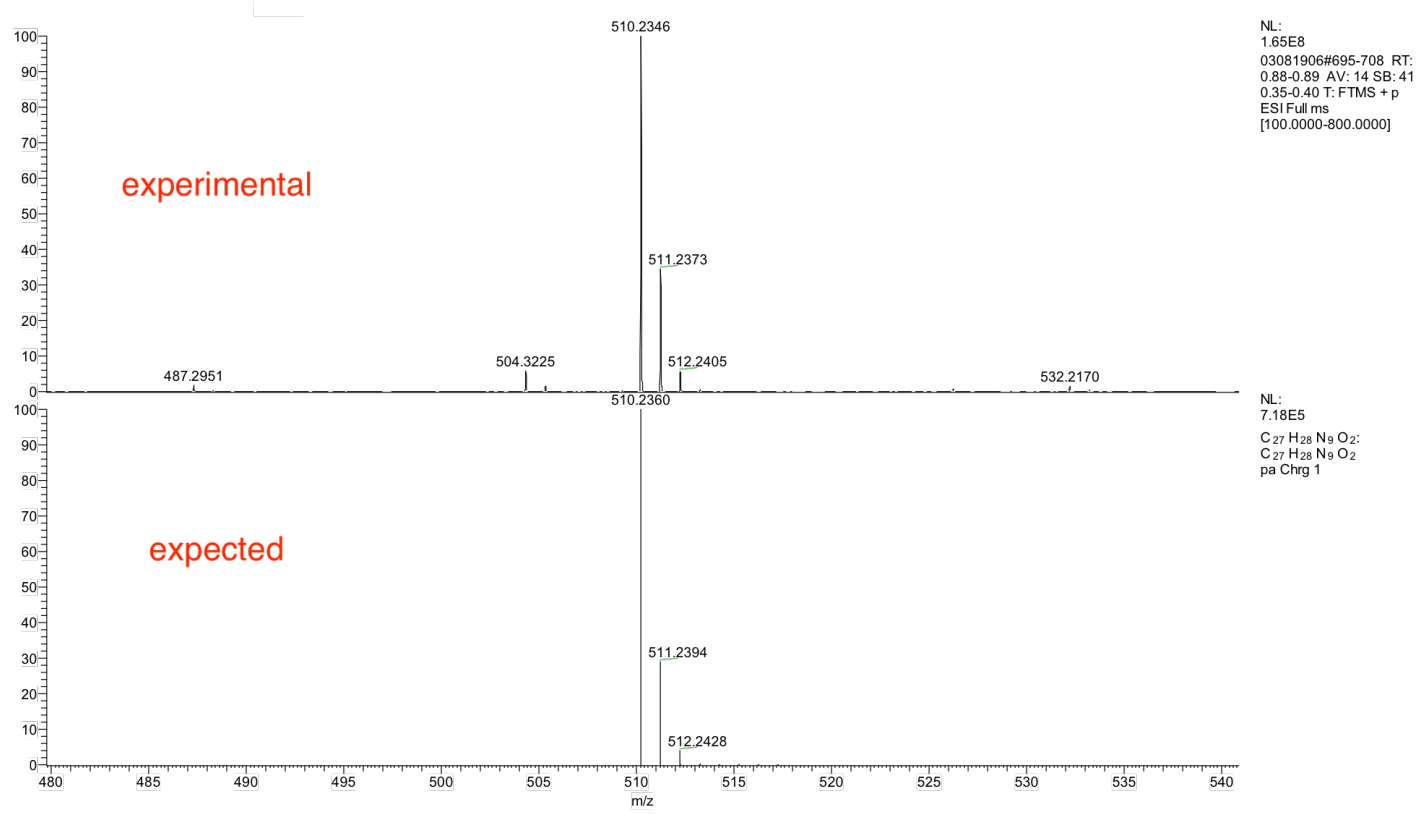
1h



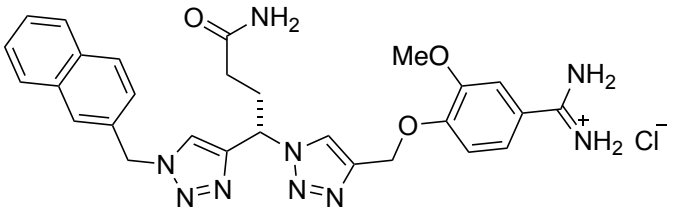
HPLC.



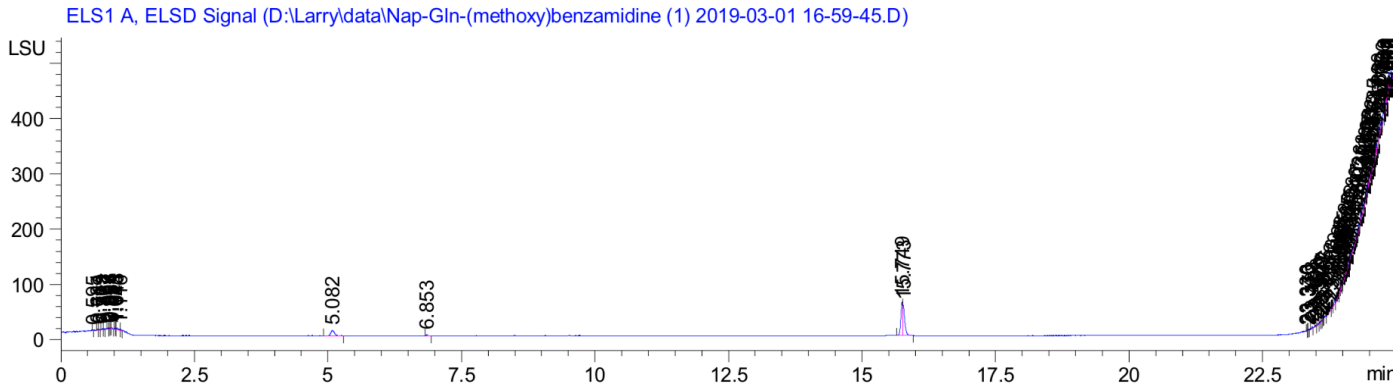
ESI-MS. Expected: 510.2360; measured: 510.2346.



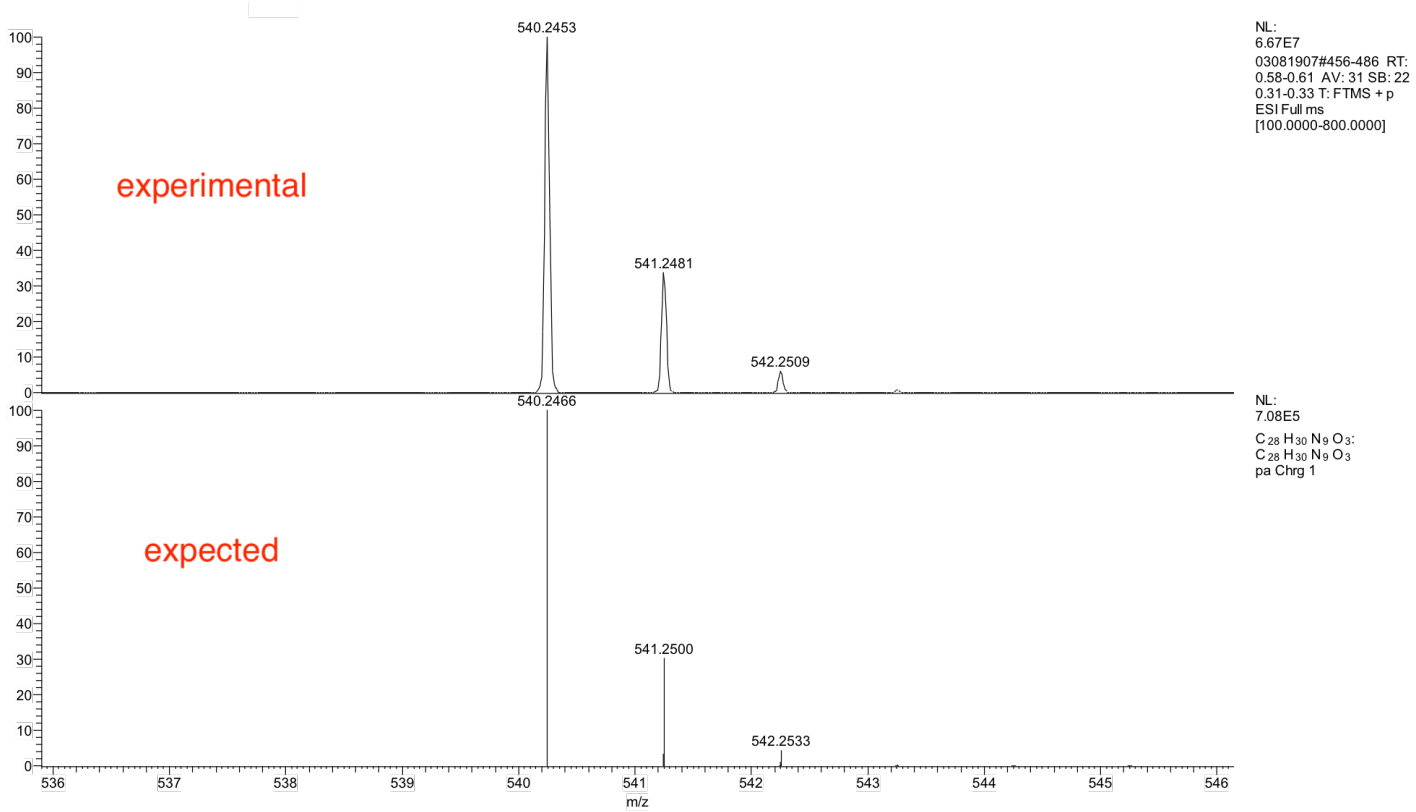
1i



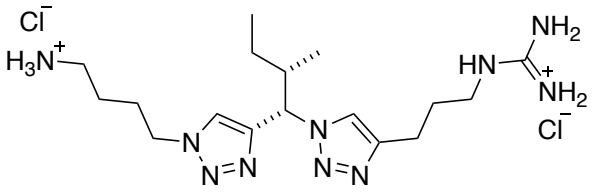
HPLC.



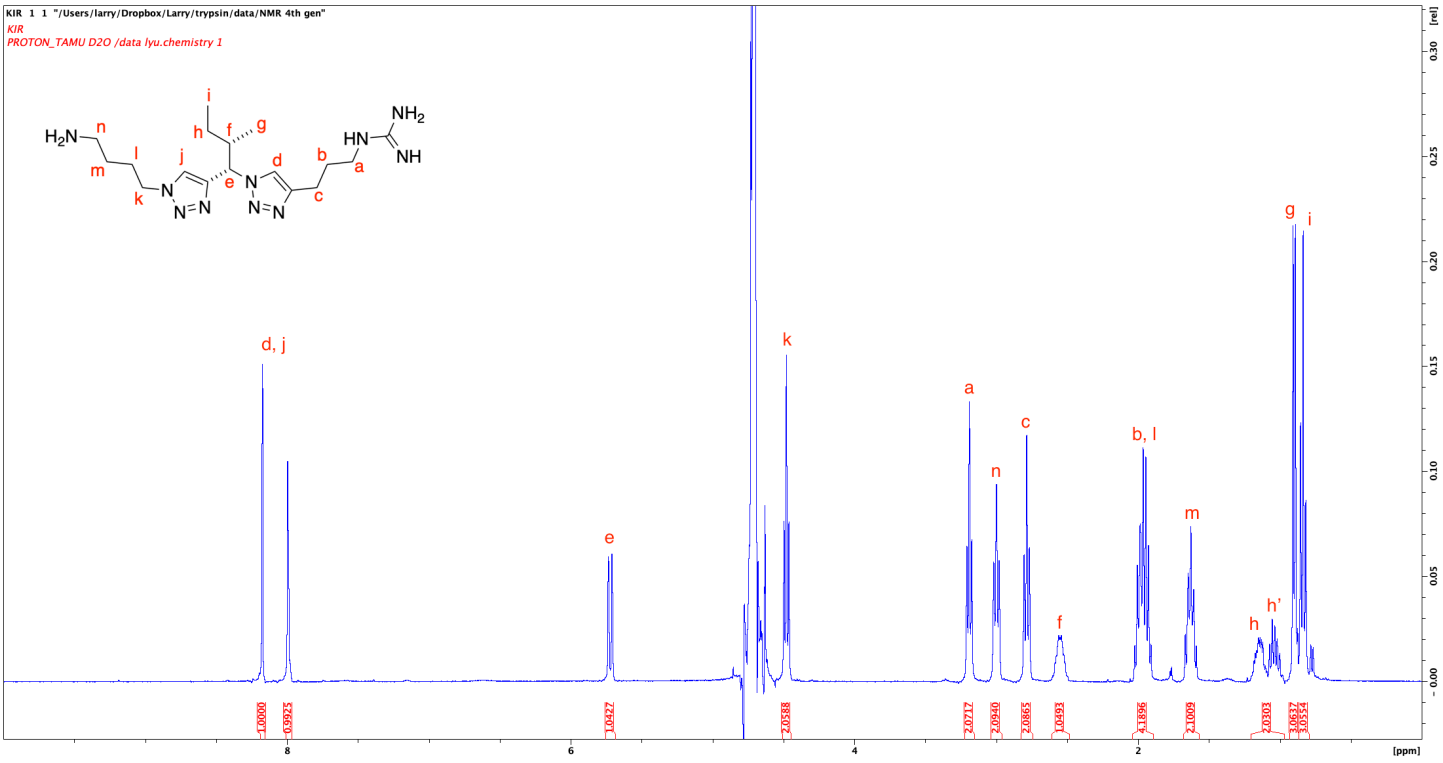
ESI-MS. Expected: 540.2466; measured: 540.2453.



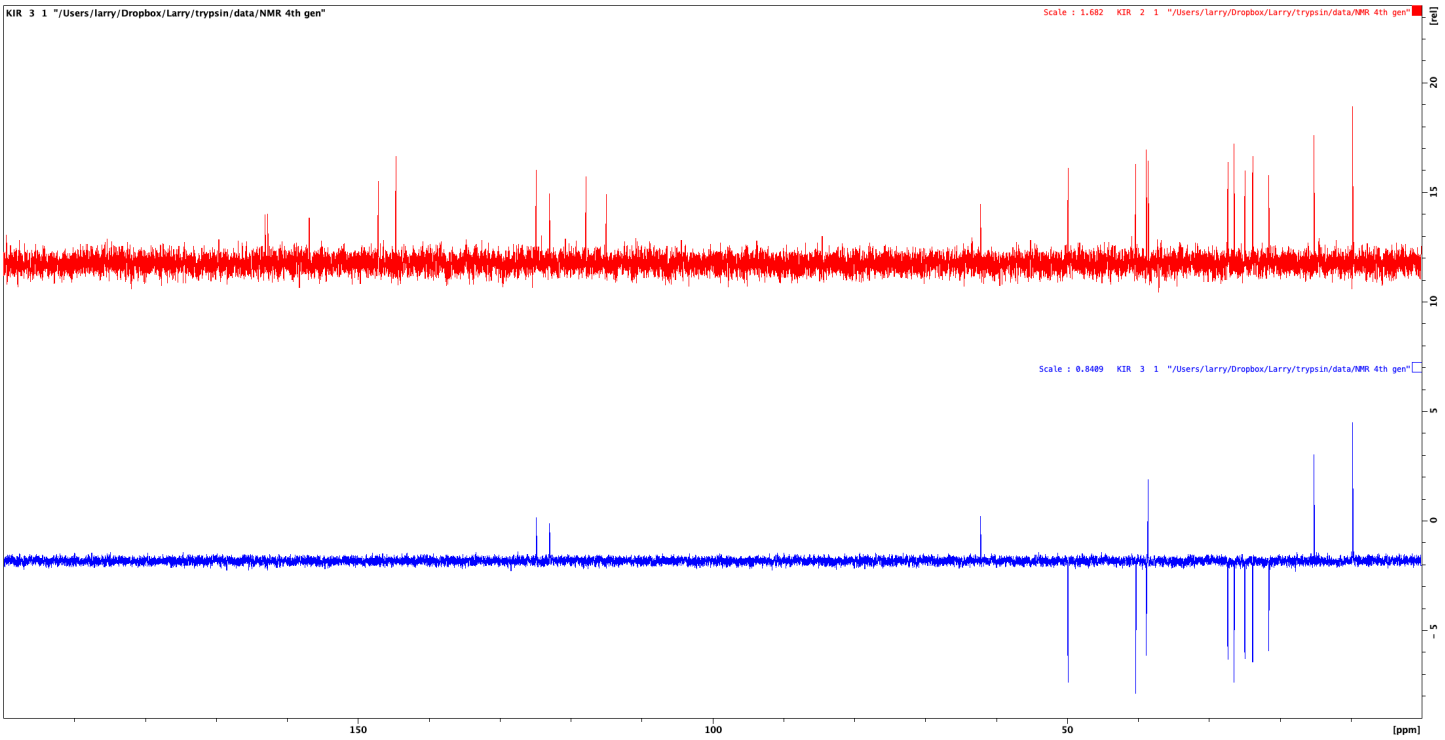
1j



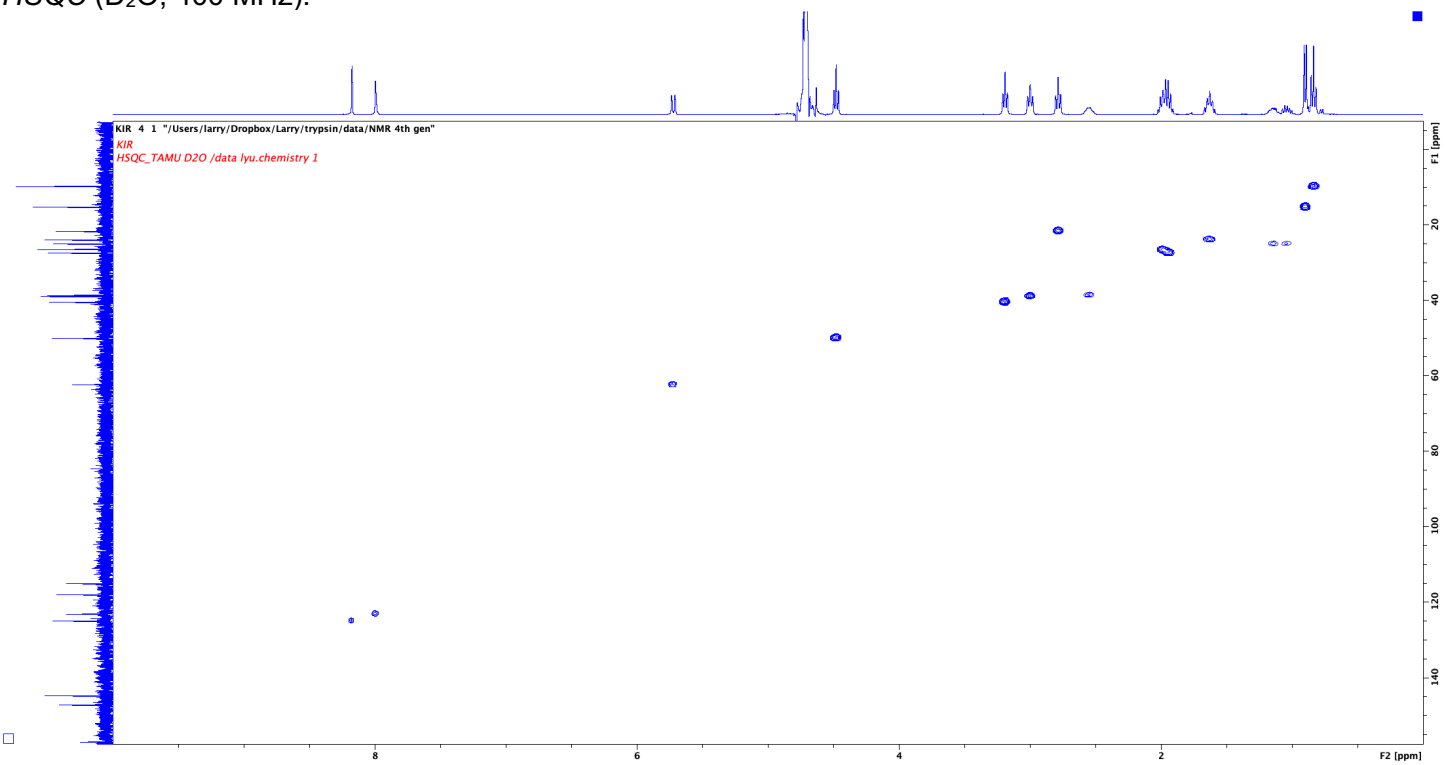
[1]H NMR (D_2O, 400 MHz). δ 8.17 (s, 1H), 8.00 (s, 1H), 5.72 (d, 1H, *J* = 10.0 Hz), 4.48 (t, 2H, *J* = 6.9 Hz), 3.19 (t, 2H, *J* = 6.8 Hz), 3.00 (t, 2H, *J* = 7.6 Hz), 2.78 (t, 2H, *J* = 7.5 Hz), 2.55 (m, 1H), 1.96 (m, 4H), 1.63 (m, 2H), 1.20-0.98 (m, 2H), 0.90 (d, 3H, *J* = 6.6 Hz), 0.84 (t, 3H, *J* = 7.4 Hz).



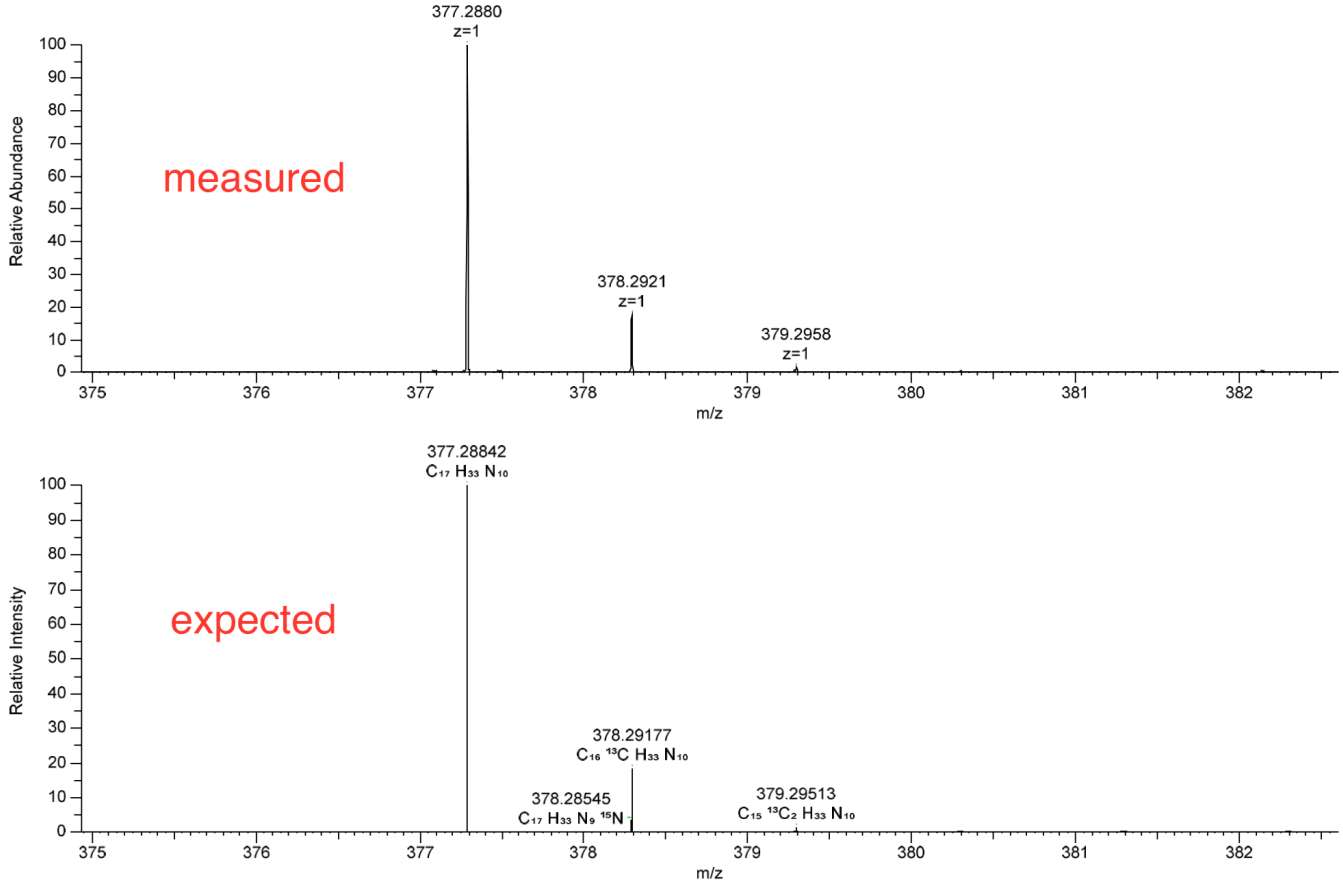
^{13}C NMR & DEPT135 (D_2O, 400 MHz). δ 163.1, 162.7, 156.9, 147.1, 144.6, 124.8, 123.0, 117.8, 114.9, 62.2, 49.9, 40.3, 38.8, 38.6, 27.3, 26.4, 24.9, 23.8, 21.6, 15.2, 9.7.



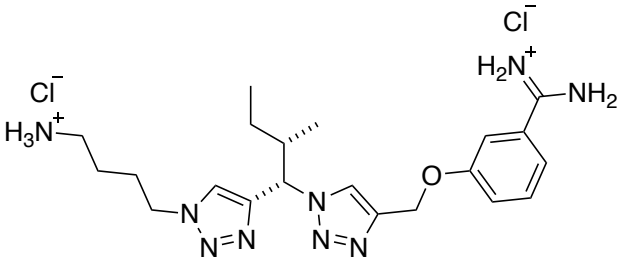
HSQC (D_2O, 400 MHz).



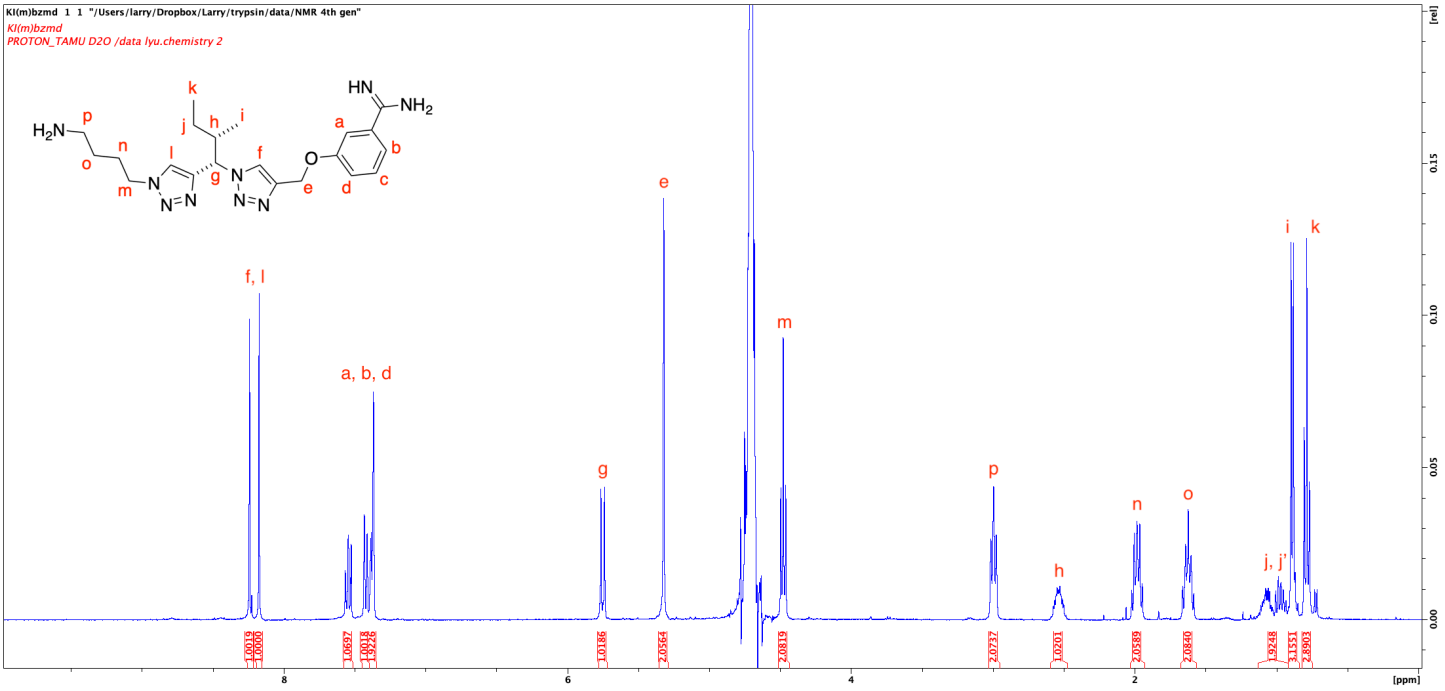
ESI-MS. Expected: 377.2884; found: 377.2880



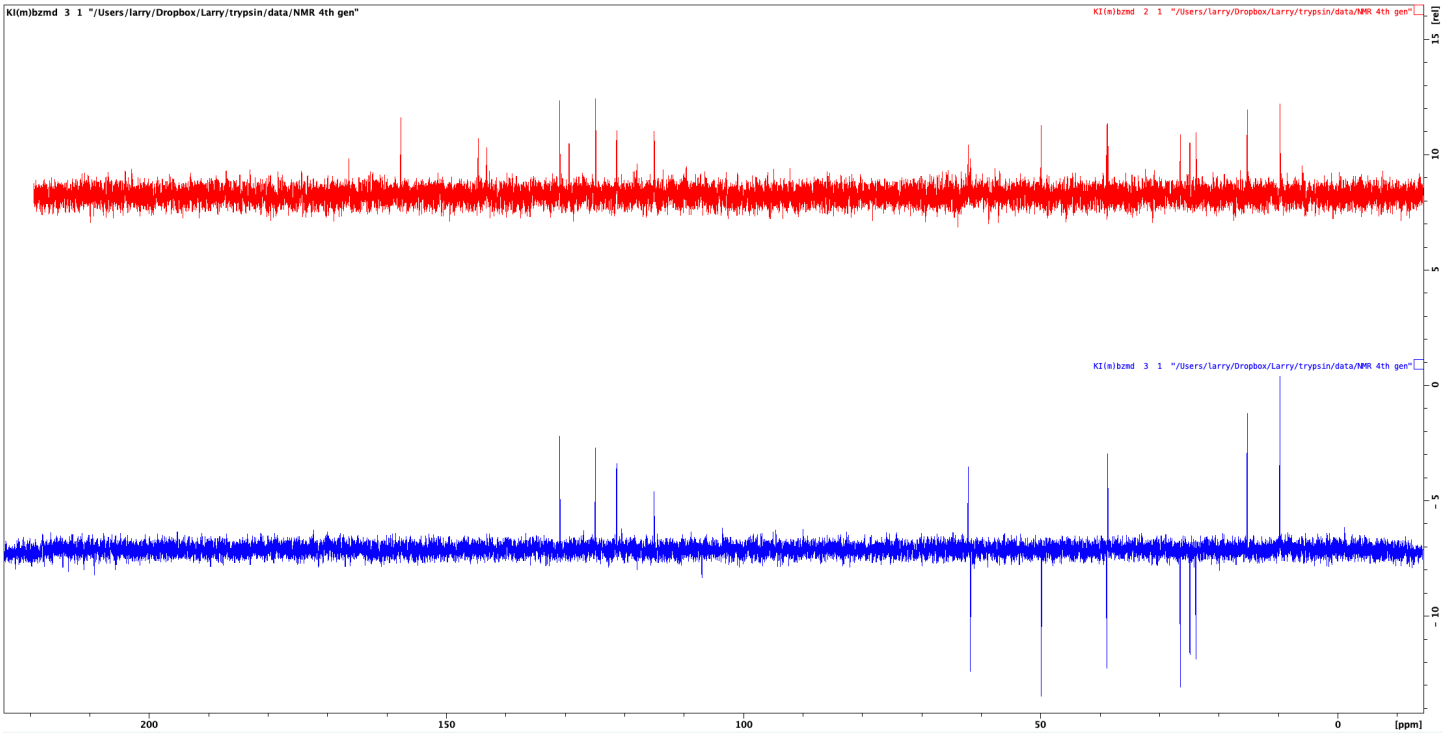
1k



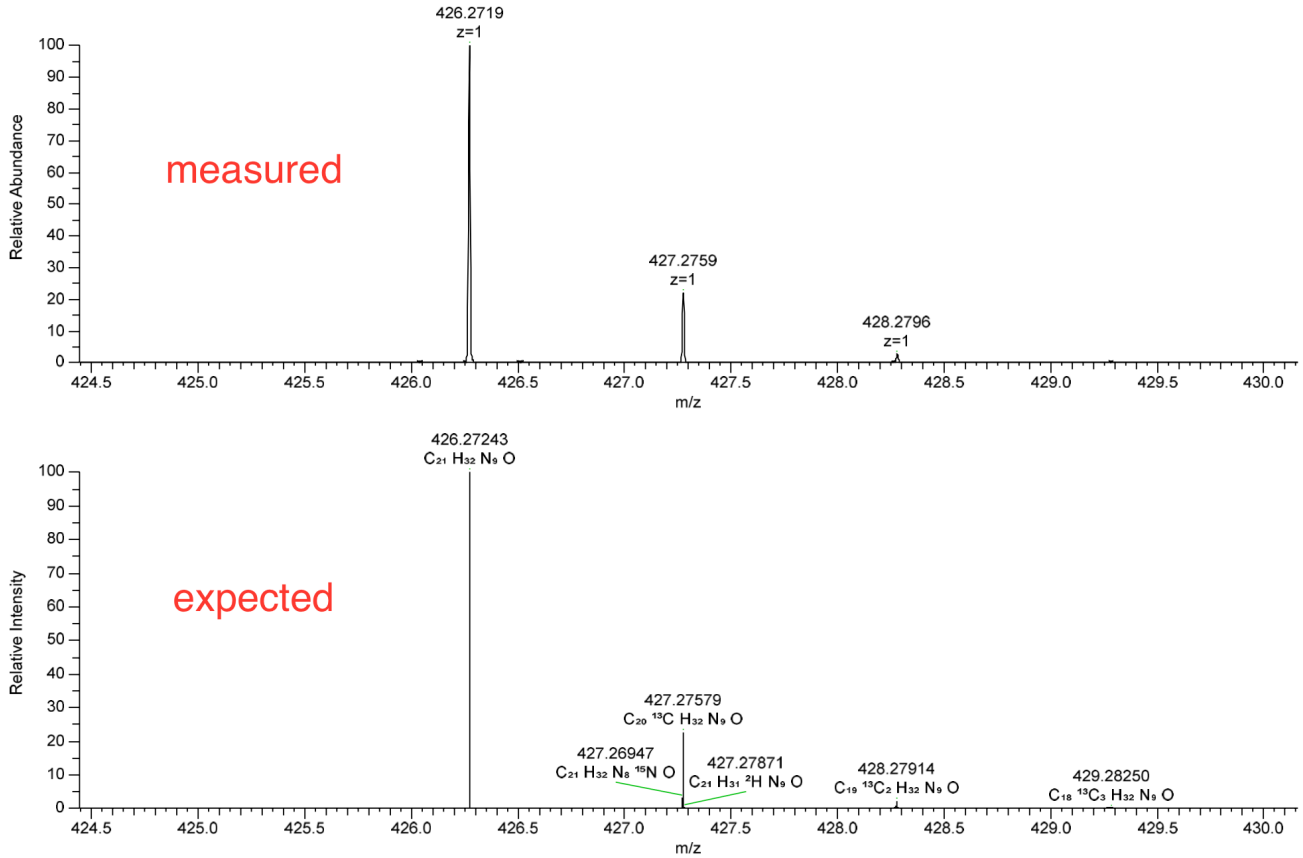
[1]H NMR (D_2O, 400 MHz). δ 8.24 (s, 1H), 8.17 (s, 1H), 7.54 (t, 1H, *J* = 3.4 Hz), 7.42 (d, 1H, *J* = 7.9 Hz), 7.38 (d, 1H, *J* = 3.9 Hz), 5.75 (d, 1H, *J* = 10.1 Hz), 5.31 (s, 2H), 4.48 (t, 2H, *J* = 6.9 Hz), 2.99 (t, 2H, *J* = 7.6 Hz), 2.54 (m, 1H), 1.99 (m, 2H), 1.62 (m, 2H), 1.12-0.92 (m, 2H), 0.89 (d, 3H, *J* = 6.7 Hz), 0.79 (t, 3H, *J* = 7.4 Hz).



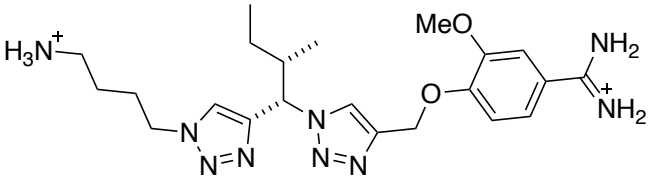
^{13}C NMR & DEPT135 (D_2O, 400 MHz). δ 157.6, 144.6, 143.2, 130.9, 129.3, 124.9, 121.3, 115.0, 62.1, 49.9, 38.8, 38.7, 26.4, 24.8, 23.8, 15.2, 9.7.



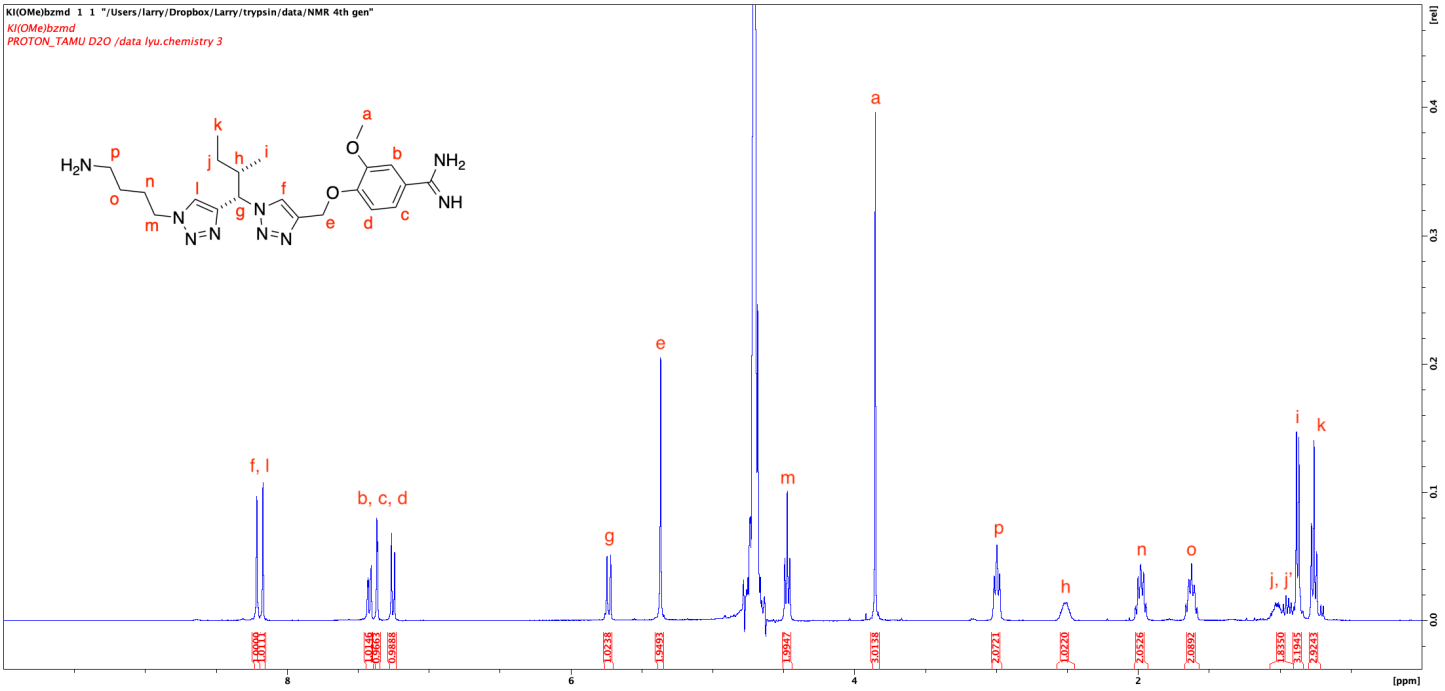
ESI-MS. Expected: 426.2724; found: 426.2719.



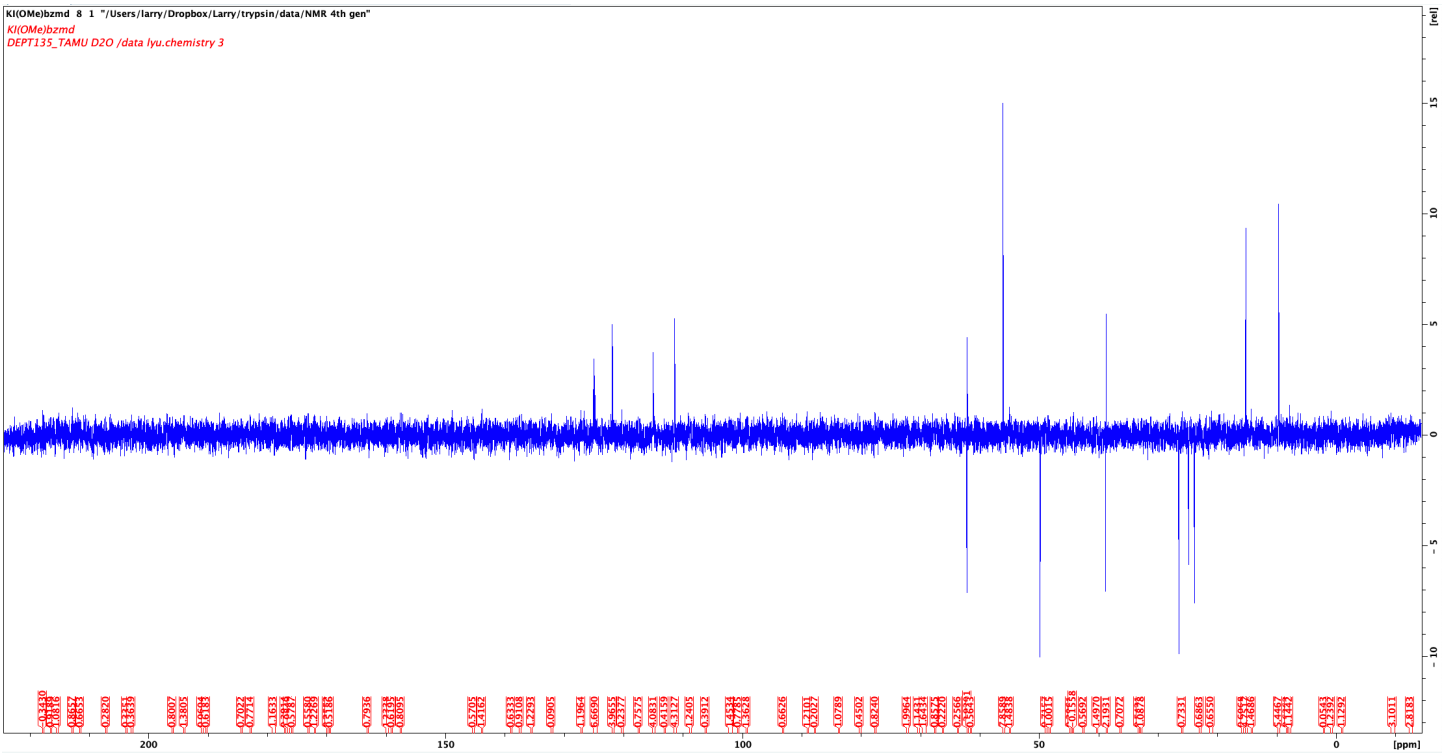
1l



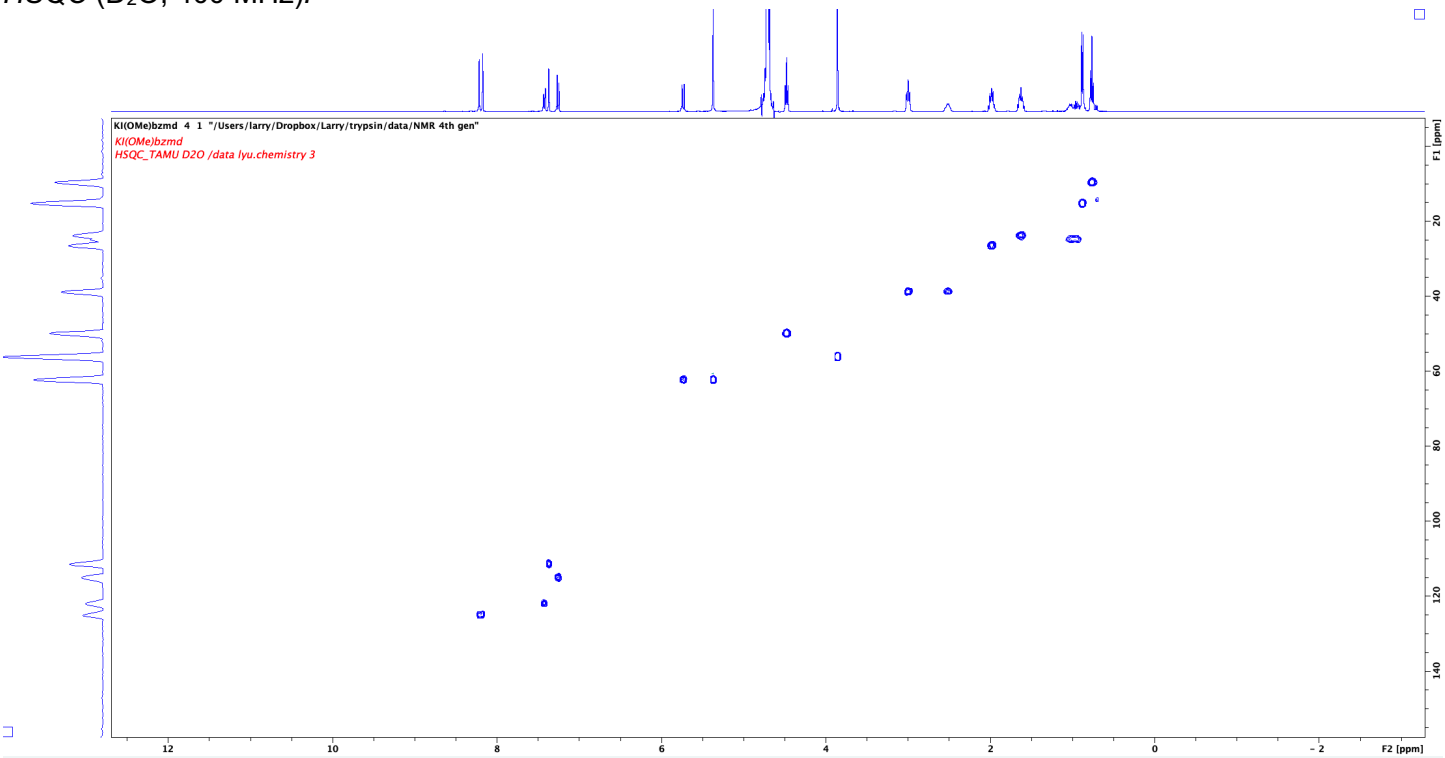
[1]H NMR (D_2O, 400 MHz). δ 8.21 (s, 1H), 8.17 (s, 1H), 7.42 (dd, 1H, *J* = 8.5, 2.1 Hz), 7.36 (d, 1H, *J* = 2.1 Hz), 7.25 (d, 1H, *J* = 8.5 Hz), 5.73 (d, 1H, *J* = 10.1 Hz), 5.37 (s, 2H), 4.47 (t, 2H, *J* = 6.9 Hz), 3.85 (s, 3H), 2.99 (t, 2H, *J* = 7.6 Hz), 2.51 (m, 1H), 1.98 (m, 2H), 1.62 (m, 2H), 1.07-0.90 (m, 2H), 0.87 (d, 2H, *J* = 6.7 Hz), 0.76 (t, 3H, *J* = 7.4 Hz).



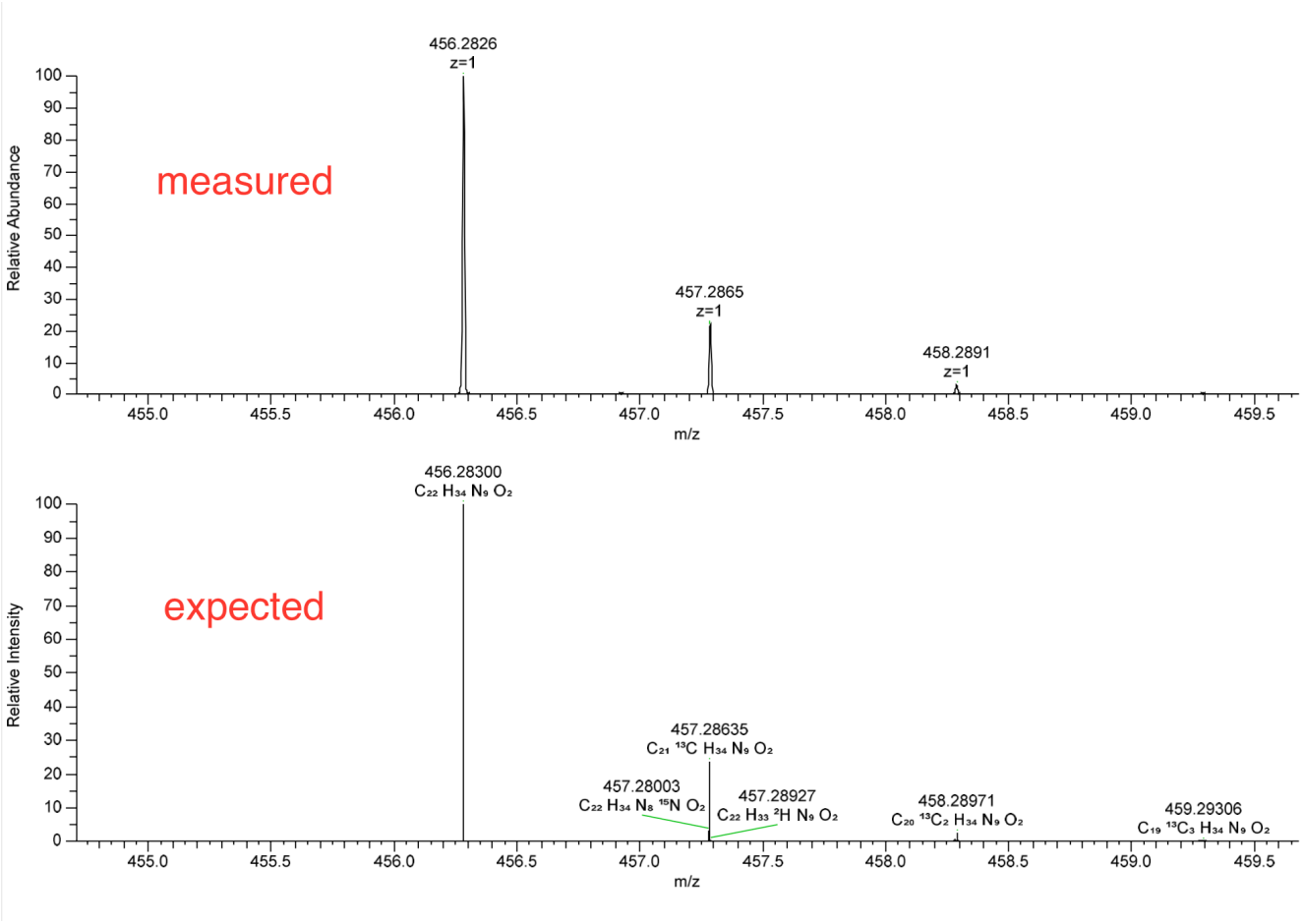
DEPT135 (D_2O, 400 MHz). δ 125.0, 124.9, 121.9, 115.0, 111.4, 62.2, 62.1, 56.1, 49.9, 38.8, 38.7, 26.4, 24.8, 23.8, 15.2, 9.7.



HSQC (D_2O, 400 MHz).

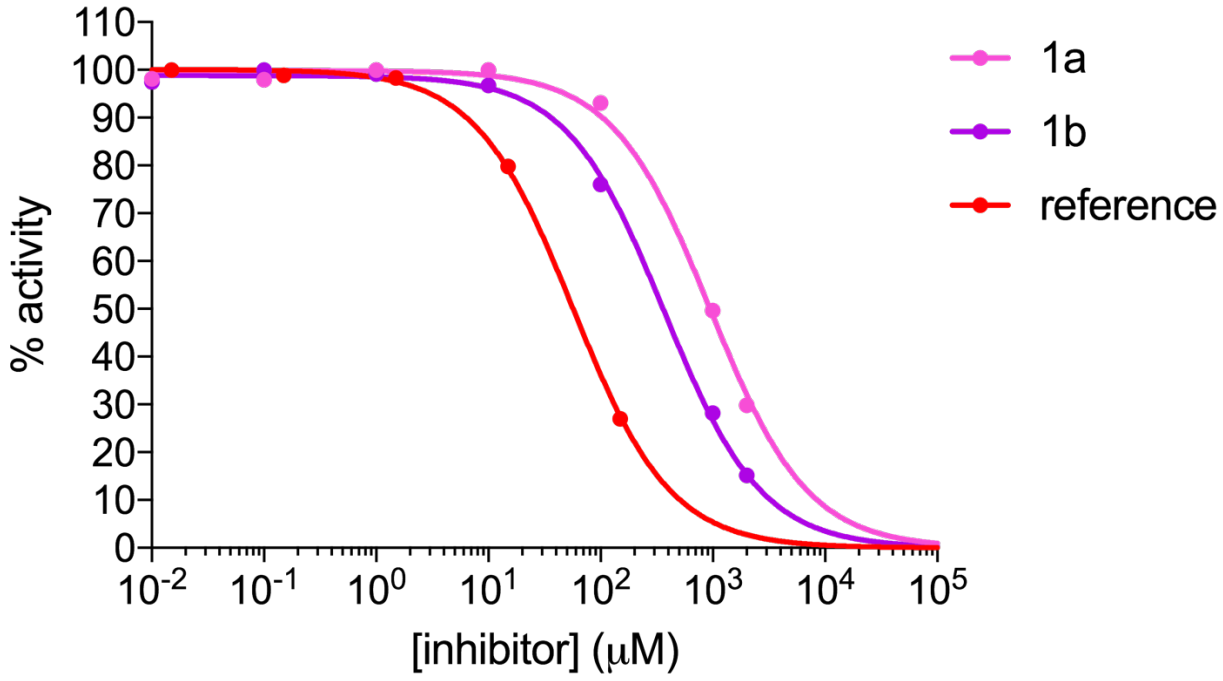


ESI-MS. Expected: 456.2830; found: 456.2826.

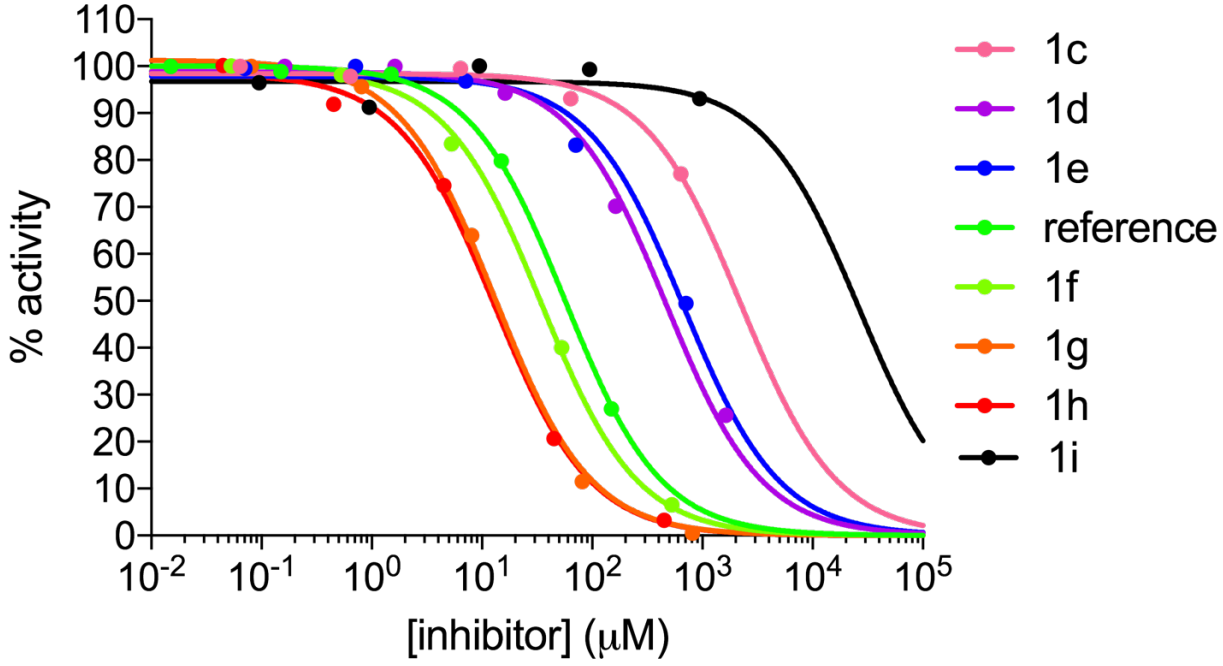


H. Assay Data

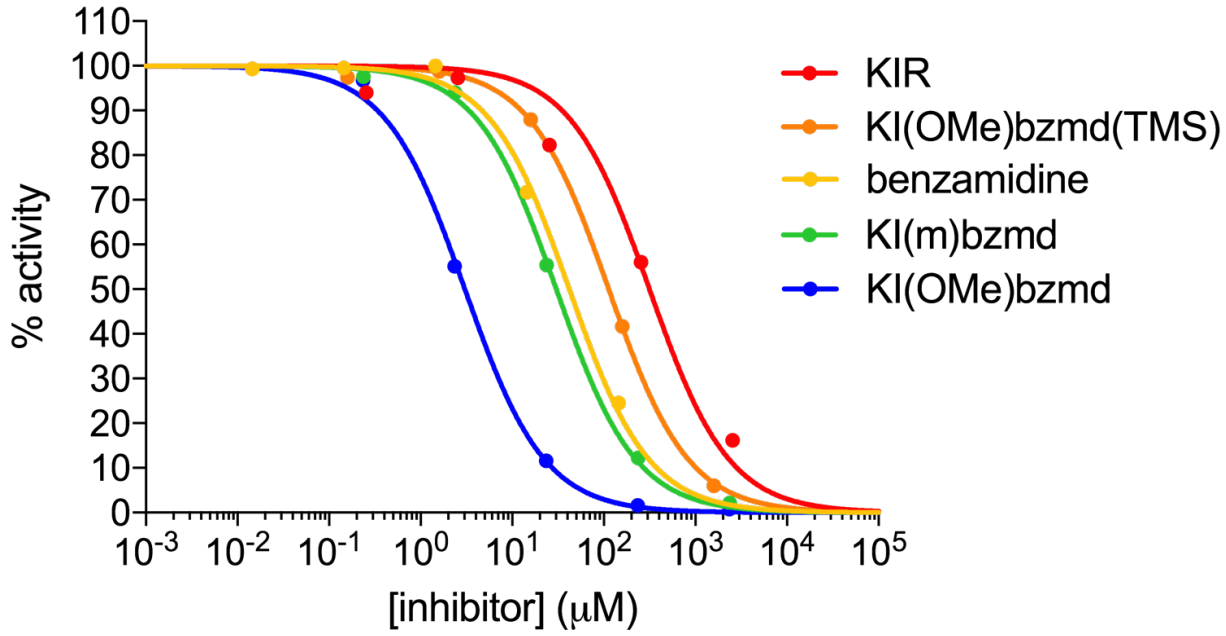
Dose Response Experiments



1st generation compounds (reference = benzamidine)



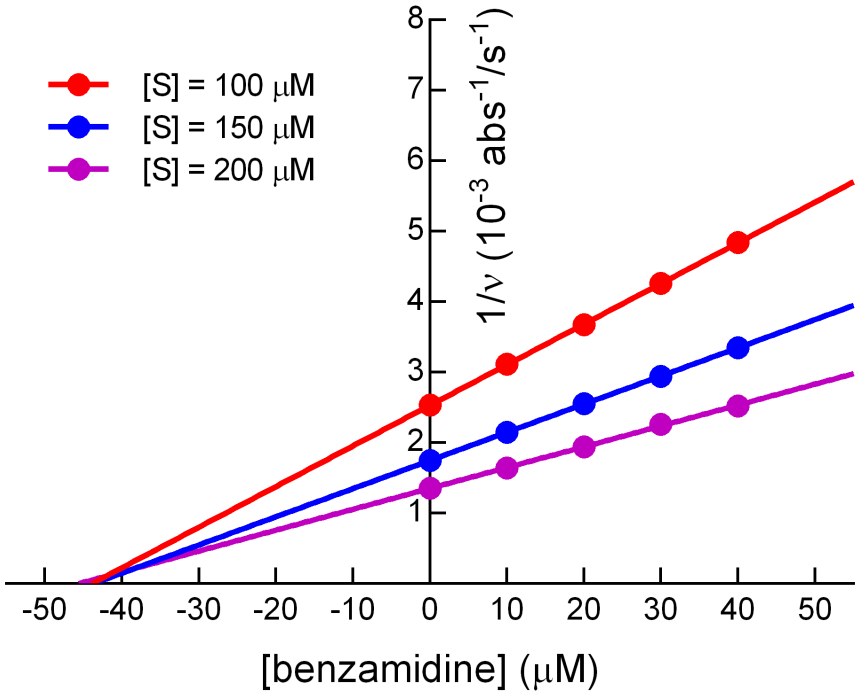
2nd generation compounds



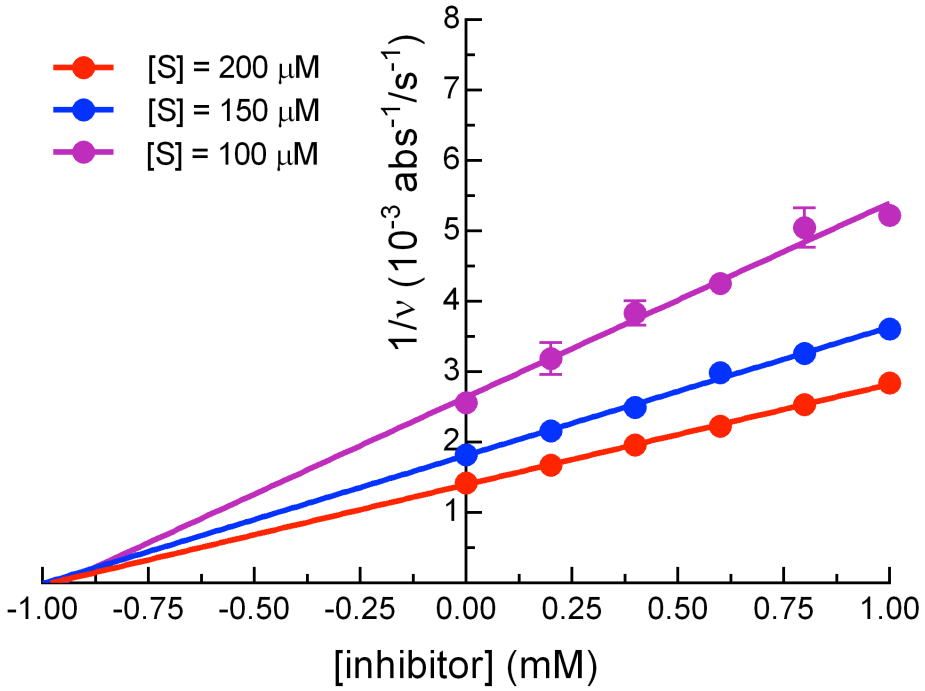
3rd generation compounds: KIR = **1j**, KI(m)bzmd = **1k**, KI(OMe)bzmd = **1l**; KI(OMe)bzmd(TMS) is a side product separated during synthesis)

K_d Measurements

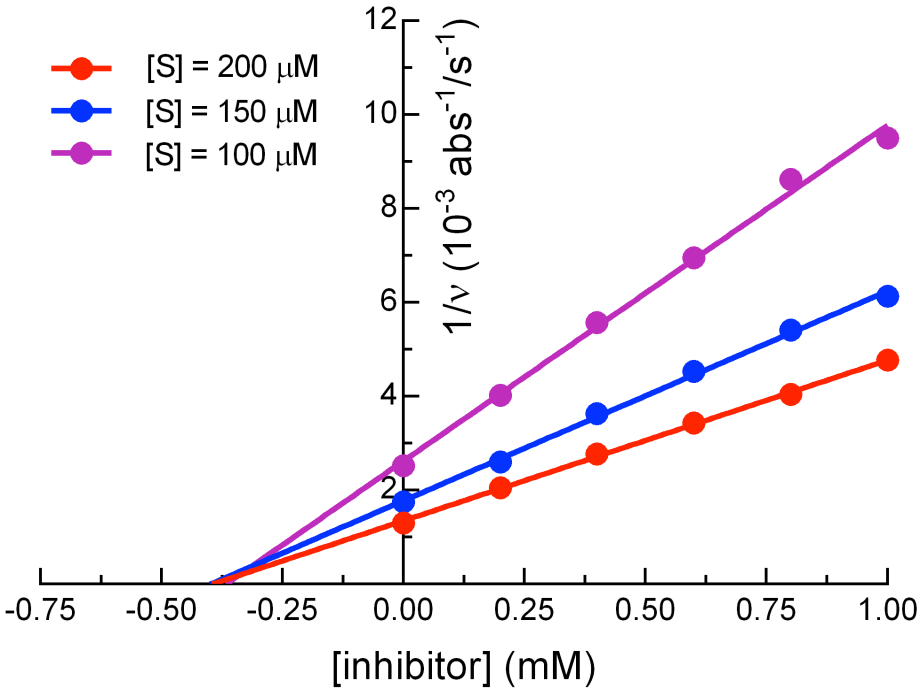
First Assay Featuring Bovine Trypsin
Based on Zimmerman *et al* ref[30] in text.



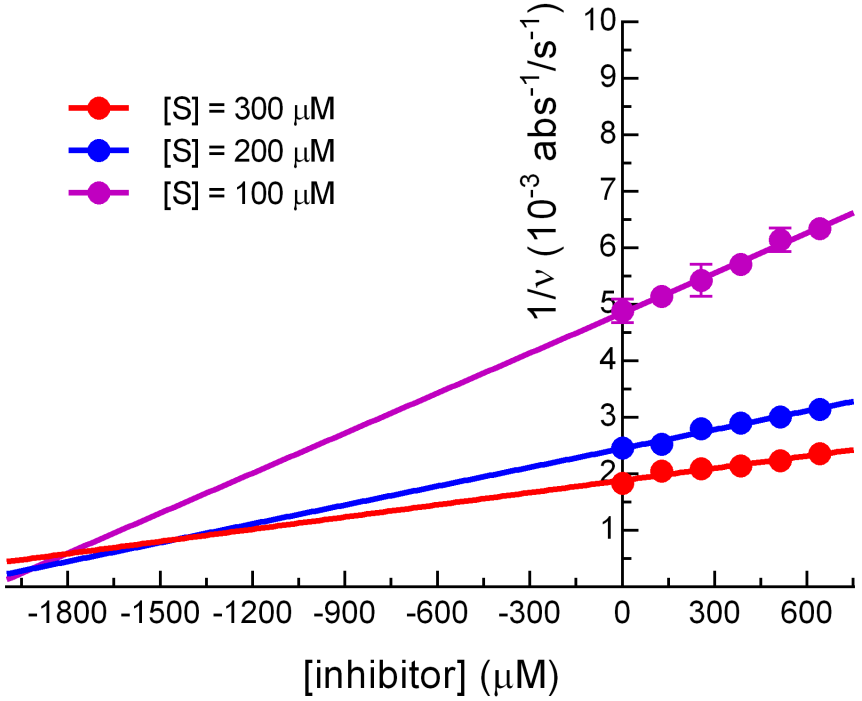
benzamidine (reference compound) binding to trypsin; error bars may be too small to see



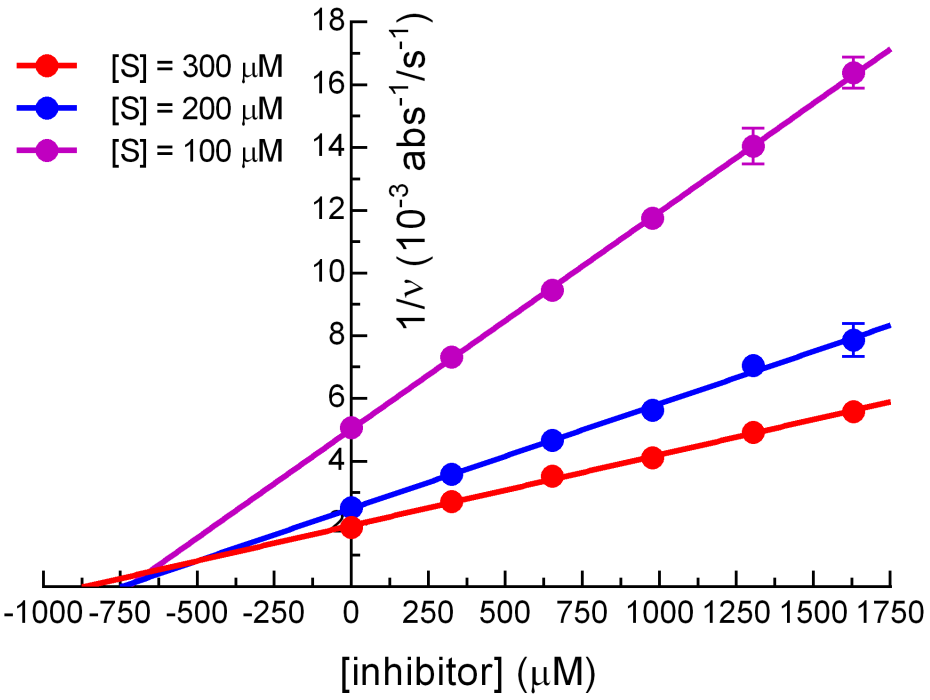
compound **1a**



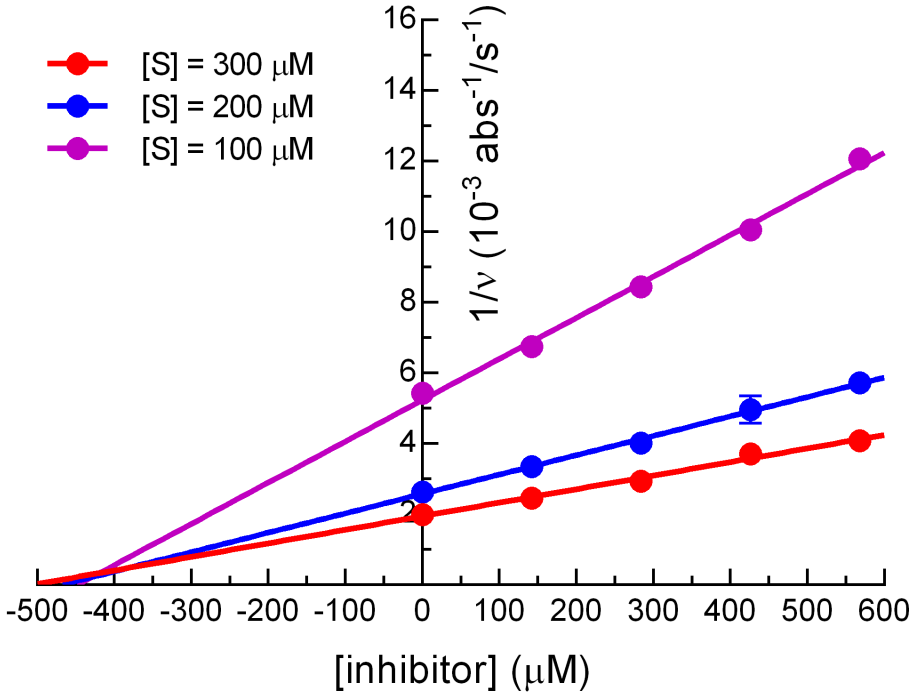
compound **1b**



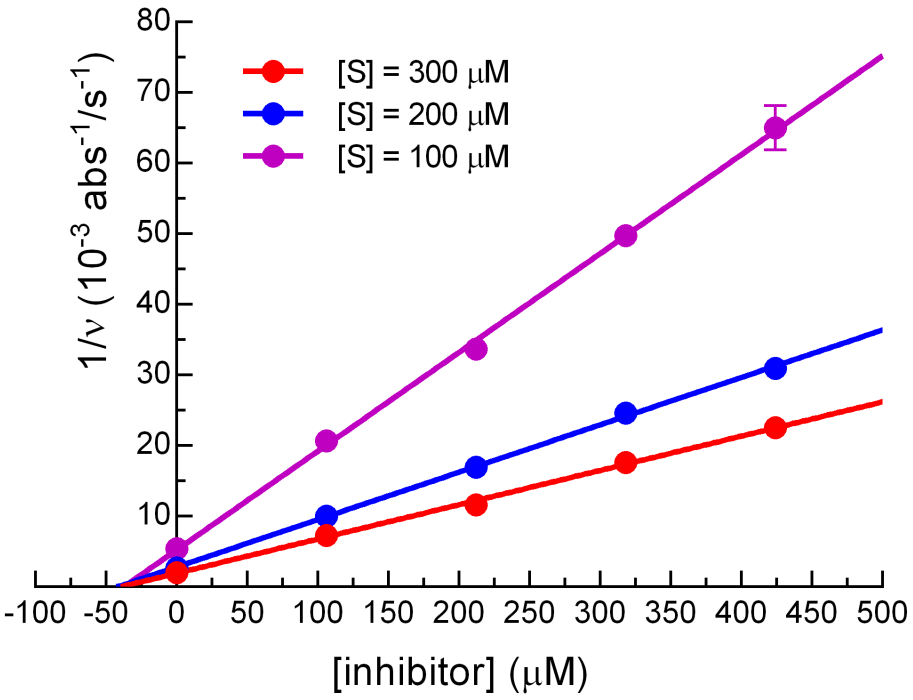
compound **1d**



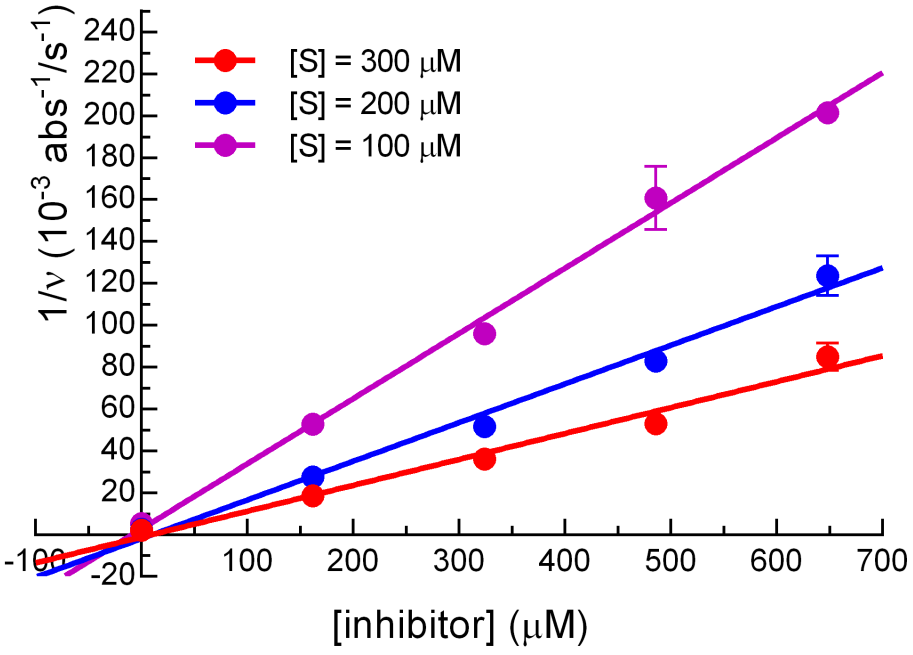
compound **1e**



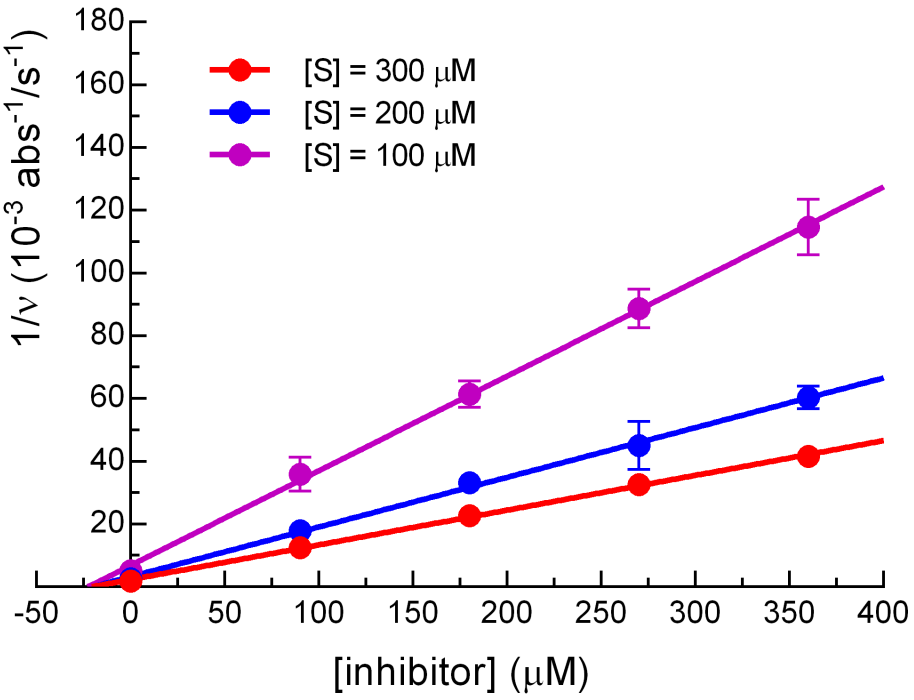
compound **1f**



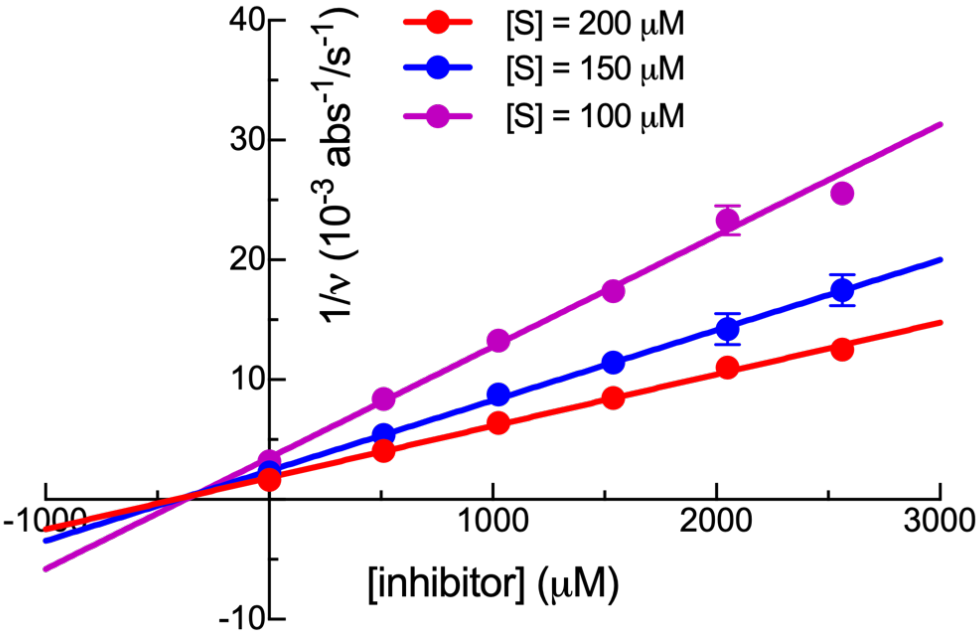
compound **1g**



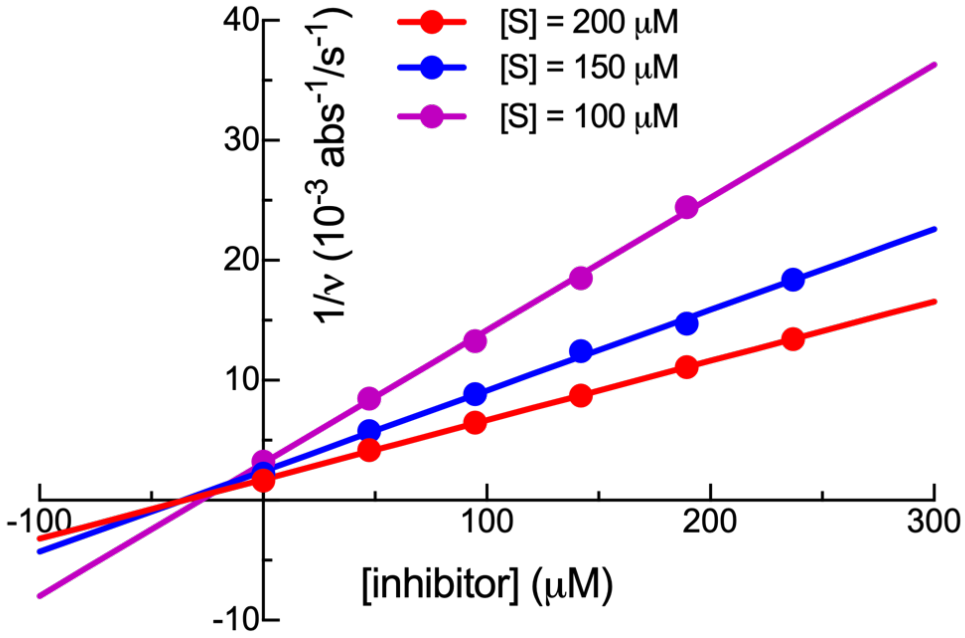
compound **1h**



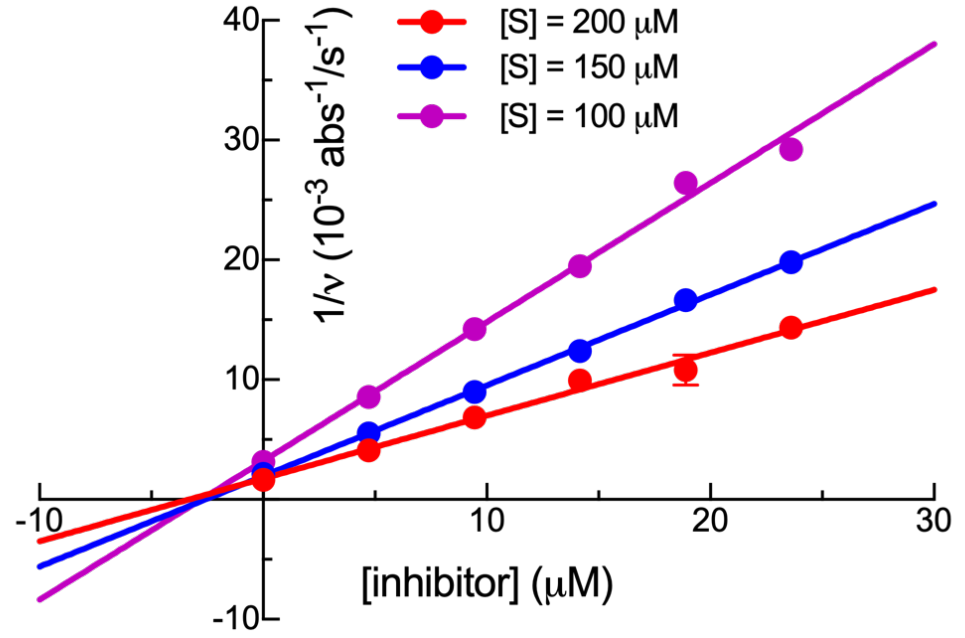
compound **1i**



compound **1j**



compound **1k**



compound **1l**

Second Assay Featuring Human Trypsin

Kd's of several selected compounds against *human* trypsin (in contrast to assays described in the previous section, which used *bovine* trypsin) were measured using an optically pure, tripeptide probe, H-Glu-Gly-Arg-pNA acetate (CAS # 67175-71-8).

Firstly, the *uninhibited* reaction rates of H-Glu-Gly-Arg-pNA hydrolysis by trypsin at various substrate concentrations were measured. According to Michaelis-Menton equation:

$$v = \frac{v_{\max}[\mathrm{S}]}{K_\mathrm{M} + [\mathrm{S}]}$$

$$\frac{1}{v} = \frac{K_\mathrm{M}}{v_{\max}} \frac{1}{[\mathrm{S}]} + \frac{1}{v_{\max}}$$

$$k = \frac{K_\mathrm{M}}{v_{\max}}$$

Plot 1/*v* – 1/[S] and calculate the slope *k*.

Next, select a constant substrate concentration, and measure the same reaction rate under various inhibitor concentrations. According to Michaelis-Menton equation:

$$v = \frac{v_{\max}[\mathrm{S}]}{K_\mathrm{M}\left(1 + \dfrac{[\mathrm{I}]}{K_\mathrm{d}}\right) + [\mathrm{S}]}$$

$$\frac{1}{v} = \frac{1}{K_\mathrm{d}} \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]}[\mathrm{I}] + \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]} + \frac{1}{v_{\max}}$$

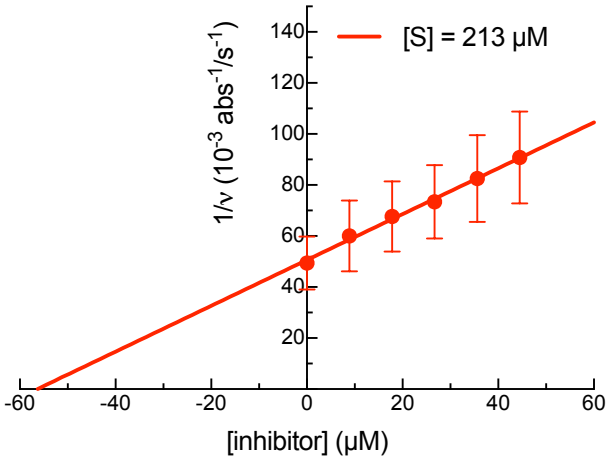
$$k' = \frac{1}{K_\mathrm{d}} \frac{K_\mathrm{M}}{v_{\max}[\mathrm{S}]} = \frac{1}{K_\mathrm{d}} \frac{k}{[\mathrm{S}]}$$

Plot 1/v – [I] and calculate the slope k’.

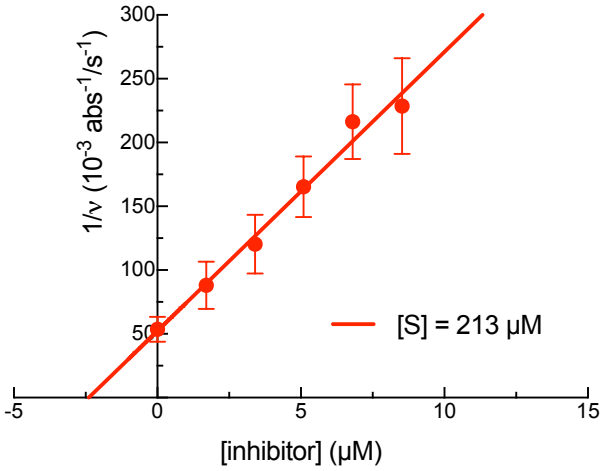
Combine the equations of k and k’, we arrive at

$$K_\mathrm{d} = \frac{k}{k'[\mathrm{S}]}$$

Buffer, enzyme, substrate, and inhibitor stock solutions were prepared in the same way as described in the previous section.



compound **1k**



compound **1l**

I. References

1. Ko, E.; Raghuraman, A.; Perez, L. M.; Ioerger, T. R.; Burgess, K., Exploring Key Orientations at Protein-Protein Interfaces with Small Molecule Probes. *J. Am. Chem. Soc.* **2013,** *135*, 167-173.
2. Xin, D.; Ko, E.; Perez, L. M.; Ioerger, T. R.; Burgess, K., Evaluating Minimalist Mimics by Exploring Key Orientations on Secondary Structures (EKOS). *Org. Biomol. Chem.* **2013,** *11*, 7789-7801.
3. Fukuyama, T.; Lin, S. C.; Li, L., Facile reduction of ethyl thiol esters to aldehydes: application to a total synthesis of (+)-neothramycin A methyl ether. *J. Am. Chem. Soc.* **1990,** *112*, 7050-1.
4. Muller, S.; Liepold, B.; Roth, G. J.; Bestmann, H. J., An Improved One-pot Procedure for the Synthesis of Alkynes from Aldehydes. *Synlett* **1996**, 521-522.
5. Roth, G. J.; Liepold, B.; Muller, S. G.; Bestmann, H. J., Further Improvements of the Synthesis of Alkynes from Aldehydes. *Synthesis* **2004**, 59-62.
6. Zimmerman, M.; Ashe, B.; Yurewicz, E. C.; Patel, G., Sensitive assays for trypsin, elastase, and chymotrypsin using new fluorogenic substrates. *Analytical Biochemistry* **1977,** *78*, 47-51.
7. Markwardt, F.; Landmann, H.; Walsmann, P., Comparative Studies on the Inhibition of Trypsin, Plasmin, and Thrombin by Derivatives of Benzylamine and Benzamidine. *European Journal of Biochemistry* **1968,** *6*, 502-506.
8. Kabsch, W., Xds. *Acta crystallographica. Section D, Biological crystallography* **2010,** *66*, 125-32.
9. Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968,** *33*, 491-7.
10. McCoy, A. J., Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica, Section D: Biological Crystallography* **2007,** *63*, 32-41.
11. Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica, Section D: Structural Biology* **2019,** *75*, 861-877.
12. Vagin, A. A.; Steiner, R. A.; Lebedev, A. A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G. N., REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica, Section D: Biological Crystallography* **2004,** *D60*, 2184-2195.
13. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallographica, Section D: Biological Crystallography* **2011,** *67*, 235-242.
14. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica, Section D: Biological Crystallography* **2004,** *D60*, 2126-2132.
15. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **1993,** *26*, 283-91.
16. Peters, B.; Moad, C.; Youn, E.; Buffington, K.; Heiland, R.; Mooney, S., Identification of similar regions of protein structures using integrated sequence and structure analysis tools. *BMC Structural Biology* **2006,** *6*, 4.