UCSF
UC San Francisco Electronic Theses and Dissertations

Title

Expanding the Covalent Chemistry Toolbox: Arginine-Selective Probes for Functional Proteomics

Permalink

https://escholarship.org/uc/item/12222665

ISBN

9798293854837

Author

Ecker, Andrew Knight

Publication Date

2025-09-19

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Expanding the Covalent Chemistry Toolbox: Arginine-Selective Probes for Functional Proteomics

by
Andrew Ecker

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

Signed by:
Ian Seiple
0F55EF21BB3D419...

Ian Seiple
Chair

DocuSigned by:
Balyn Zaro
DocuSigned by:

Balyn Zaro

James Fraser
430BBB9A04D24A3...

James Fraser

Committee Members

Copyright 2025

by

Andrew Ecker

“All of the Science We Do Stands on the Shoulders of Giants. It Cost’s You Nothing to Recognize What They’ve Done to Make Your Work Possible” – Taia Wu

This thesis is dedicated to the memory of my grandparents Deborah and Hoyt Ecker as well as Arthur and Bonnie Knight.

Thank you for always pushing me to keep learning and for all the time you all dedicated to teaching me both academically and about how to appreciate the world in all of its beauty.

Acknowledgements

First, I want to thank you, Ian, my professor, prinseiple advisor, and great friend. You've always made me feel welcome and like a real genuine scientist. Even before I applied to graduate school, your enthusiasm when we first met and the exciting work you were doing inspired me to pursue a PhD. You made academia feel welcoming, even while taking on some of the most challenging and fascinating chemistry out there. You've always been there when I needed help, offering your support with whatever I was facing personal or scientific, and you were just as present to celebrate every win along the way. Even after our move to Scripps, you've continued to empower me and make me feel like I can do anything. You truly are an inspiration to me. I've also really appreciated our friendship and the way we can just chitchat, thank you for everything. I am so thankful to not be closing the chapter of doing research with you just yet, I still have so much more to learn from you and I'm very excited to see what I learn next and what we discover next.

To the members of the Seiple lab, thank you for being part of such a meaningful chapter of my life. First, to Izzy, my lab mate from day one. When we both got into UCSF, I had a feeling we were going to be friends, and I was absolutely thrilled when we ended up joining the same lab. Sharing a hood with you that first summer was such a joy, and I've loved watching you grow into the iconic chemist you are today. Our cutesy chitchat sessions have meant more to me than you know, I've learned so much from them, and from you. To Quinn, thank you for being one of my best friends in the lab. Your humor, your brilliant mind, and the kindness and joy you brought to work made such a difference. Whether I needed help troubleshooting something or a push to go on a run, you were always there. I can't wait to see what incredible things you'll do next. To Arthur and Jesus, thank you for being not just mentors, but good friends. Your attention to detail, your knowledge, and your ability to make me laugh helped shape my time in the lab for the better. And

to the rest of the Seiple lab, thank you for welcoming me so fully and for creating a positive, inspiring space to grow as a scientist. I feel lucky to have been a part of this group. To the new Seiple lab at Scripps, thank you for putting up with me as the external graduate student I am and for giving me such a great space to continue to learn and grow as a chemist and mentor, I have already learned a lot from you and am excited to continue learning with you all. Thanks for being friends.

I want to thank all of the biology/chemical biology mentors I've had the pleasure of getting to know throughout my time at UCSF. When I started at UCSF as a chemist in a chemical biology program, I felt remarkably out of place having very little understanding of biology let alone chemical biology. Proteins, and nucleic acids were more or less imaginary to me and the thought of moving them around from tube to tube to make science happen sounded magical. Jumping into the world of chemical biology, I am so thankful to Professor Kevan Shokat for giving me the opportunity to rotate in his lab and to learn from now Professor Ziyang Zhang on such a remarkable project as the KRAS G12R project. It was the perfect introduction for me to developing a covalent probe towards what became my favorite amino acid, arginine. Ziyang, thank you so much for teaching me nearly all of the chemical biology skills I now have within a short and impactful 3-month period. You made chemical biology feel accessible and showed me such great patience as I learned to manipulate DNA, culture human cells, and run a western blot making many mistakes along the way. You also inspired me to think outside of the box when looking at arginine and how to meaningfully design a covalent ligand towards it which has greatly shaped the science that I ended up working on in my thesis. Diving into true fundamental and structural biology, I am so thankful for Professor James Fraser who gave me the opportunity to learn from him for three months even though he wouldn't be able to take me on as a student. While my rotation with you

was an amazing time for me to build a love and appreciation for the work that goes into CryoEM, you have long continued to be a role model to me to keep asking questions down to the fundamental principles that drive what we are doing instead of being complacent. I greatly look forward to where our scientific paths cross next and to see what new things I can learn from you.

Of all of the chemical biology mentors I've had at UCSF, the greatest appreciation needs to go to Professor Balyn Zaro. You were my first interviewer and one of the first people from UCSF to show me kindness and a true love for what you do which really inspired me. Whenever I felt scared, challenged or overwhelmed by chemical biology, you were there to lift me up and teach me something new. I am so thankful you were willing to put some faith into a wild idea I had and as it became a major portion of my thesis you were there to give me support, resources and your time to show me how to do the experiments I was struggling with. Thank you for giving me space in your lab and treating me like one of your students even if it was not ever on record, I will always fondly think of my time working in your group. I know we are going to continue to have a great time exploring the ligandable world of proteins and I am so thankful you opened my eyes up to the world of activity-based protein profiling.

To all of the friends I have made at UCSF, thank you for being a community that challenged me to grow, celebrated my victories, and helped me through my challenges (yes, even enzyme kinetics). Gabe, thank you for being such an inspiring, kind, generous, and funny soul. I've thoroughly enjoyed all of our adventures, from casual Friday lunches to so many unique hikes. Even though we worked on opposite ends of CCB, you were one of my favorite people to talk to, and you taught me a lot about gratitude and how to appreciate life more fully. Sara and Madison, my other two hiking adventure cohort buds, thank you for your support, your patience, and your kindness. It was such a joy to share our adventures together, including our unforgettable trip to

Boston. Steph and Katie, my two best friends at UCSF. You both made everything better, no matter how hard things got. Every seminar, lunch, nature walk, and random afternoon in a park became some of the most fun I had in San Francisco because of the two of you. You'll always be the two biggest icons in CCB in my mind. You've taught me so much about exploring chemical biology through computers, working with proteins, and how to have a good laugh (or a good cry) after an experiment didn't work out. I love you both so much and can't wait for all of our adventures still to come. Robbie, thank you for spotting me that day while I was just trying to enjoy my lunch in peace and you decided I needed a queer scientific role model. You taught me so much about the value of building balance between research and enjoying one's own individual life during grad school. I loved getting to work with you on the chitinase project and that project gave me such a deeper appreciation for truly taking advantage of structural biology to answer an interesting chemistry question, the intersect of research I want to eventually study independently. Beyond the science you were always such a kind and dear friend to me who was there for the best and the worst of times but always was there to support me and lift me up.

I want to extend my deepest gratitude to my first academic mentors at UCSD. Drs. Bill and Lena Gerwick, thank you for taking a chance on me and giving me the opportunity to fall in love with research in your lab. You gave me a broader perspective and appreciation on the world we live in, and the science one can pursue, while also giving me the freedom to attempt my first total synthesis. I learned how to work hard and think deeply about my research under your guidance. Evgenia Glukhov, thank you for dedicating so much time to training me in good, reproducible science. From tissue culture to entertaining all of my wild ideas, you always made space for me to learn and grow. Aurora, thank you for connecting me with the Gerwick lab and for being my first friend in research. Working alongside you gave me so much comfort and taught me the importance

of keeping a good lab notebook. You gave me the confidence to take a moment and just dance in lab before getting back to work, and I'm so grateful for that joy. For all of my friends at UCSD from oSTEM who empowered me to believe in a world where queer scientists are celebrated and create space for the next generation of amazing diverse scientists.

To all of my family, thank you for being so patient with me as I've reclused a bit during the course of my PhD. I love you all deeply and am so grateful for the kindness and support you've extended to me throughout this journey. Mom, thank you for always encouraging me to always finish my homework, to work hard, and to follow through on commitments, even when they were tough. You've been there with me at every stage of this journey, and even when I moved the farthest, I've ever lived from you, you traveled often to see me which brought me great joy and respite from my studies. You instilled in me a deep love for the natural world and give me the freedom to be who I want to be. I wouldn't be here without your love and support. To my sister, thank you for being my number one stan from day one. You've supported me every step of the way, even when I was bad at responding to your messages. Our adventures, both in San Francisco and beyond, offered me needed breaks from school and filled me with joy. Thank you for always being there, in the big moments and the small ones.

Finally, to Jake. Thank you for being there for me every single day, no matter how late I came home or how early I left in the morning. I always knew I could count on you to listen to me talk about my day and to celebrate the little wins with such genuine enthusiasm and interest. Thank you for being willing to uproot your life to join me on the adventure we had in SF and to do it again just to return to San Diego once we fell in love with SF. Thank you for reminding me to slow down and find balance whenever I tried to speed run my life. You were there to take care of me when I felt like giving up and to remind me why I chose to pursue a PhD in the first place. Thank

you for everything, and for all the continued support you offer me as I move forward. And thank you, too, for all of our in-depth conversations about magnetic resonance and electron spin, you inspired me to ask weirder, better questions about the science right in front of me. Our sweet cats, Garfunkel Zoboomafoo and Lilith Boady, thank you so much for the many stimulating late-night conversations about science, research, and vibes in general. Thank you for always waking me up early in the morning so I wouldn't be late for lab, and for usually greeting me at the door after a long day with your love.

Contributions

The chapters in this dissertation were performed under the guidance of Dr. Ian Seiple and Dr. Balyn Zaro, with the collaboration, mentorship, and feedback of many Seiple and Zaro lab members. Dr. Jaime Fraser, on my thesis committee, and Dr. Ziyang Zhang, who previously mentored me in developing a covalent probe for arginine, both provided great intellectual support and mentorship throughout the development of the project. Dr. Minh Tran provided significant contributions toward the initial development of arginine and citrulline covalent strategies and toward the in vitro validation of ninhydrin reactivity. Andy Langen and José Montaño spent considerable time and effort greatly improving the proteomic coverage of our covalent probes towards arginine both in improving the sample preparation and in the processing of proteomic data. The automated synthesis facility at Scripps Research Institution provided HR-MS analysis for all compounds synthesized in this manuscript.

Andrew Ecker provided the experimentation and manuscript preparation for most of the following dissertation, which is a substantive contribution comparable to other dissertations in Chemistry & Chemical Biology. Ian Seiple and Balyn Zaro directed and supervised the research and provided guidance and feedback throughout.

Expanding the Covalent Chemistry Toolbox: Arginine-Selective Probes for Functional Proteomics

Andrew Ecker

Abstract

Covalent molecules have emerged as powerful tools and next-generation therapeutics, promising improved specificity and sustained target engagement. Chemical proteomic methods to screen for reactive cysteine and lysine residues on proteins amenable to covalent engagement have transformed small molecule discovery pipelines. Covalent molecules have improved our understanding of fundamental biological processes and changed the way we treat human disease, but their application remains limited to sites where reactive cysteines and lysines are present. Here we report a ninhydrin-based warhead that selectively modifies arginine residues. We developed alkyne-functionalized variants of ninhydrin to establish an arginine-specific chemical proteomics platform, enabling the classification of over 6,800 arginines. These studies uncovered potential modification sites on disease-relevant proteins, including reactive arginines within catalytic sites that is essential for function. By endowing a reversible small molecule inhibitor of cyclophilin A with our ninhydrin warhead, we achieved selective, covalent engagement, highlighting the potential for targeting arginines in therapeutic development. These findings establish ninhydrin as a warhead for studying arginine reactivity and modulating protein function, enabling new avenues for covalently engaging proteins.

In chapter **1**, I introduce the rationale for covalently targeting arginine residues and describe the early stages of ninhydrin development as a covalent warhead. It begins with foundational concepts in activity-based protein profiling (ABPP) and discusses the unique chemical challenges

posed by arginine. The chapter presents in vitro studies characterizing the kinetics and selectivity of ninhydrin reactivity toward free arginine and model proteins, establishing proof-of-concept for selective arginine modification.

In chapter **2**, I describe the development and implementation of the Reactive Arginine Profiling (RAP) platform to map reactive arginines across the proteome using alkyne-functionalized ninhydrin probes. It includes structural bioinformatic analyses to assess reactivity determinants and contextual features of modified residues. Mechanistic studies are presented on the inhibition of aconitase by ninhydrin and a structure-guided approach is used to design a covalent probe that selectively targets R55 in cyclophilin A. These efforts collectively demonstrate the utility of ninhydrin probes in functional proteomics and ligand development.

Chapter **3**, contains the extended data, detailed experimental protocols, synthetic procedures for ninhydrin probes and analogs and NMR characterization of compounds. It provides the technical foundation for the studies described in Chapters 1 and 2 and serves as a resource for reproducing or extending this work.

Table of Contents

Chapter 1. Development of Ninhydrin as a Selective Covalent Warhead for Proteinaceous Arginines 1

1.1 Activity Based Protein Profiling Introduction.....2

1.2 In-vitro characterization of arginine engagement of ninhydrin.....4

1.3 Ninhydrin Probe Development.....6

Chapter 2. Global Profiling and Mechanistic Studies of Arginine Labeling 10

2.1 Chemical proteomic profiling of arginine labeling. 11

2.2 Potential mechanism for ninhydrin inhibition of aconitase. 17

2.3 Development of a pipeline for ligand-directed arginine modification. 17

2.4 Discussion 20

Chapter 3. Experimental Methods and Supporting Information 24

3.1 Extended Data Figures 25

3.2 Methods.....29

3.3 General Synthesis Procedures 47

3.4 NMR Spectra for Compounds.....59

References 89

List of Figures

Figure 1.1 | Profiling the reactivity of ninhydrin with arginine in vitro. 6

Figure 1.2 | Comparative profiling and biochemical characterization of arginine-reactive electrophiles. 9

Figure 2.1 | Structural and functional context of arginines identified by RAP in Mino cell lysates 12

Figure 2.2 | Biophysical and functional features between reactive arginine residues. 16

Figure 2.3 | Structure-guided development of arginine-selective ligands 19

Extended Data Figure 3.1 | Ninhydrin displays selectivity for citrulline at pH 4. 25

Extended Data Figure 3.2 | Lysozyme is covalently modified by ninhydrin. 25

Extended Data Figure 3.3 | Glutathione reactivity assay for arginine reactive warheads. 26

Extended Data Figure 3.4 | Comparative analysis of intermolecular interactions surrounding reactive arginines. 26

Extended Data Figure 3.5 | Solvent accessibility and secondary structure profiles of reactive arginines. 27

Extended Data Figure 3.6 | Predicted pKa distributions of reactive residues. 27

Extended Data Figure 3.7 | Waterfall plot depicting the regioselective enrichment of peptides labeled by 4-CA-Alk and 5-CA-Alk in Mino cell lysate. 28

Extended Data Figure 3.8 | Structural comparison of cation–π interactions at hyper-reactive versus general reactive arginines. 28

List of Abbreviations

ABPP – activity-based protein profiling

ACN – acetonitrile

ACO2 – aconitase 2

AF – AlphaFold

BSA – bovine serum albumin

CuAAC – copper(I)-catalyzed azide-alkyne cycloaddition

CypA – cyclophilin A

DCM – dichloromethane

DFT – density functional theory

DMF – dimethylformamide

DMSO – dimethylsulfoxide

DTB – desthiobiotin

GSH – glutathione

HPLC – high performance liquid chromatography

HRMS – high resolution mass spectrometry

IA-Alk – iodoacetamide-alkyne

LC-MS/MS – liquid chromatography-tandem mass spectrometry

NMR – nuclear magnetic resonance

PDB – Protein Data Bank

PGO – phenyl glyoxal

PPI – protein-protein interaction

RAP– Reactive Arginine Profiling

RSA – relative solvent accessibility

SP3 – single-pot solid-phase-enhanced sample preparation

SuFEx – Sulfur(VI) Fluoride Exchange

TFA – trifluoroacetic acid

TLC – thin–layer chromatography

UV – ultraviolet

Chapter 1. Development of Ninhydrin as a Selective Covalent Warhead for Proteinaceous Arginines

1.1 Activity Based Protein Profiling Introduction

Covalent therapeutics have revolutionized drug discovery by enabling precise and irreversible engagement of protein targets. Cysteine has been the predominant residue leveraged for covalent modifications, with approved drugs such as ibrutinib (BTK Cys481), afatinib (EGFR Cys797), and adagrasib/sotorasib (KRAS G12C) exemplifying this strategy. Evaluating and exploiting proteome-wide cysteine reactivity has been transformed by the use of broad-spectrum electrophiles, such as iodoacetamide-alkyne, that enable both reactivity profiling and competitive screening to assess a molecule of interest's selectivity[1,2]. More recently, lysine residues have also emerged as attractive covalent drug targets, aided by probes that form imines or related adducts[3–5]. Beyond warheads selectively targeting a specific amino acid residue, such as cysteine and lysine, activity-based protein profiling (ABPP) strategies have begun to extend toward other nucleophilic side chains, although these remain less comprehensively mapped[6–8]. SuFEx-based probes also enable the profiling of lysines and less conventional nucleophilic residues across the proteome, including histidine and tyrosine[9,10].

Arginine plays indispensable roles in enzymatic catalysis, molecular recognition, and nucleic acid interactions. The positively charged guanidinium side-chain engages in a diverse array of noncovalent interactions, including salt-bridging and cation-π contacts, enabling arginine to stabilize transition states, mediate protein–protein interaction, and recognize aromatic functional groups within biomolecular targets[11]. However, its high pKa, resonance-stabilized charge distribution, and limited nucleophilicity complicate the development of covalent modification strategies. The reaction between vicinal dicarbonyl compounds with amidine and guanidine moieties has long been investigated from initial studies by Diels and Schleich[12] to the Maillard reaction of sugar fragments with arginine[13], yet the direct application of these reactions in

biological systems has been quite limited. Although previous approaches, such as α,β-unsaturated amides[14] and phenyl glyoxal derivatives [7,15–17], have been explored, they often suffer from poor biocompatibility and moderate selectivity. The development of a selective and robust warhead for covalent arginine modification would provide a powerful tool for studying arginine reactivity and function in complex biological contexts and facilitate the design of new therapeutic strategies. Post-translational modifications (PTMs) of arginine, including methylation, citrullination, and ADP-ribosylation, play essential roles in regulating chromatin structure, gene expression, and immune signaling[11]. Despite their biological relevance, these PTMs remain challenging to detect and study in native proteomes due to the lack of selective reagents and enrichment strategies. Citrullination, in particular, represents a non-reversible transformation of the arginine guanidinium group into a neutral urea moiety, catalyzed by peptidylarginine deiminases (PADs)[11]. This modification profoundly alters the physicochemical properties of the residue, yet tools for its global profiling remain limited.

Ninhydrin, an indanetrione, has long been utilized for detecting amino acids through its chromogenic reaction with primary amines at high temperature. Despite its established role in analytical chemistry, its potential as a selective covalent modifier of arginine has remained underexplored. Given its electrophilic properties and ability to rapidly form stabilized cyclic adducts with guanidine[18], we hypothesized that ninhydrin could be repurposed for targeted covalent modification of arginine in biological systems. Furthermore, we postulated that if ninhydrin were reactive enough across proteinaceous arginine residues, further functionalization with a bioorthogonal handle, such as an alkyne, would allow for the development of a pan-arginine chemical proteomics platform akin to the iodoacetamide-alkyne probe and the isoTOPP-ABPP workflow, as developed by Cravatt and co-workers for cysteine[1].

Our initial interest in leveraging ninhydrin for selective guanidine reactivity stemmed from exploring whether citrulline and arginine could be differentially targeted under specific conditions. In a simplified model system, we observed that at pH 4.0, Fmoc-Citrulline undergoes nearly complete conversion to a stable adduct with ninhydrin after 65 hours, whereas Fmoc-Arginine remains largely unreacted under the same conditions, as assessed by analytical HPLC (**Extended Data Fig. 3.1A**). Furthermore, in a protein context, we compared ninhydrin reactivity between unmodified bovine serum albumin (BSA) and its citrullinated counterpart using a bioorthogonal alkyne-functionalized ninhydrin probe. Following treatment, protein samples were subjected to copper-catalyzed azide–alkyne cycloaddition (CuAAC) with a TAMRA-azide reporter and analyzed by in-gel fluorescence. This assay revealed preferential labeling of citrullinated BSA over unmodified BSA after only one hour of probe treatment (**Extended Data Fig. 3.1B**). These findings suggested that ninhydrin could be tuned for citrulline selectivity under acidic conditions; however, the slower kinetics and narrow pH window prompted us to instead focus our platform on developing a more generalizable strategy for targeting proteinaceous arginine residues.

1.2 In-vitro characterization of arginine engagement of ninhydrin.

Many vicinal dicarbonyl compounds in aqueous environments exist predominantly in their hydrate state, rendering them relatively inert to nucleophilic attack. The equilibrium between this hydrate state and the reactive dehydrated state governs their ability to engage in covalent modification[19]. To investigate this equilibrium for ninhydrin versus the known arginine-reactive but promiscuous warhead phenyl glyoxal (PGO)[20], we performed density functional theory (DFT) calculations (**Fig. 1.1A**). These values predict the relative occupancy of each state, allowing us to assess their potential for selective arginine modification. Our calculations revealed that both probes predominantly exist in their hydrate state; however, ninhydrin has a calculated K_{eq} of 3.55×10^{-3},

indicating a higher fraction of the molecule resides in the reactive dehydrated state. Conversely, PGO has a significantly lower calculated K_{eq} of 1.33×10^{-6}, suggesting a much lower occupancy in the reactive form (**Fig. 1.1A**). This favorable equilibrium positioning suggests ninhydrin would have a strong potential for arginine modification.

To validate these predictions, we characterized ninhydrin reactivity toward arginine using an HPLC-based kinetics assay monitoring the consumption of Fmoc-Arginine (**Fig. 1.1B**). This approach provided direct evidence that the reaction proceeds at the terminal guanidine, while also enabling precise quantification via UV absorption. The reaction kinetics were evaluated across a pH gradient ranging from 6.0 to 10.0, with time points recorded at 2, 7, 15, 30, and 60 minutes. At pH 7.5, complete consumption of free arginine was observed within 60 minutes, whereas in more basic conditions reaction completion was achieved within 30 minutes (**Fig. 1.1B**). The relatively fast 2nd order reaction kinetics are comparable to that of iodoacetamide with cysteine and further supported our hypothesis that ninhydrin may be a warhead amenable to engaging arginines in biological systems on a reasonable timescale[21].

To probe the selectivity of ninhydrin labeling and its potential to react with proteinaceous arginines, we examined its modification stoichiometry on a model protein, bovine serum albumin (BSA), which contains 23 arginine residues (**Fig. 1.1C**). Following 90 minutes of ninhydrin treatment, intact protein mass spectrometry revealed selective engagement of up to 9 residues under physiologic conditions (**Fig. 1.1D**), with increased reactivity in a pH-dependent manner, as we observed in our kinetics assay. Similar studies were also conducted for lysozyme, revealing a comparable pH-dependent reactivity profile (**Extended Data Fig. 3.2**).

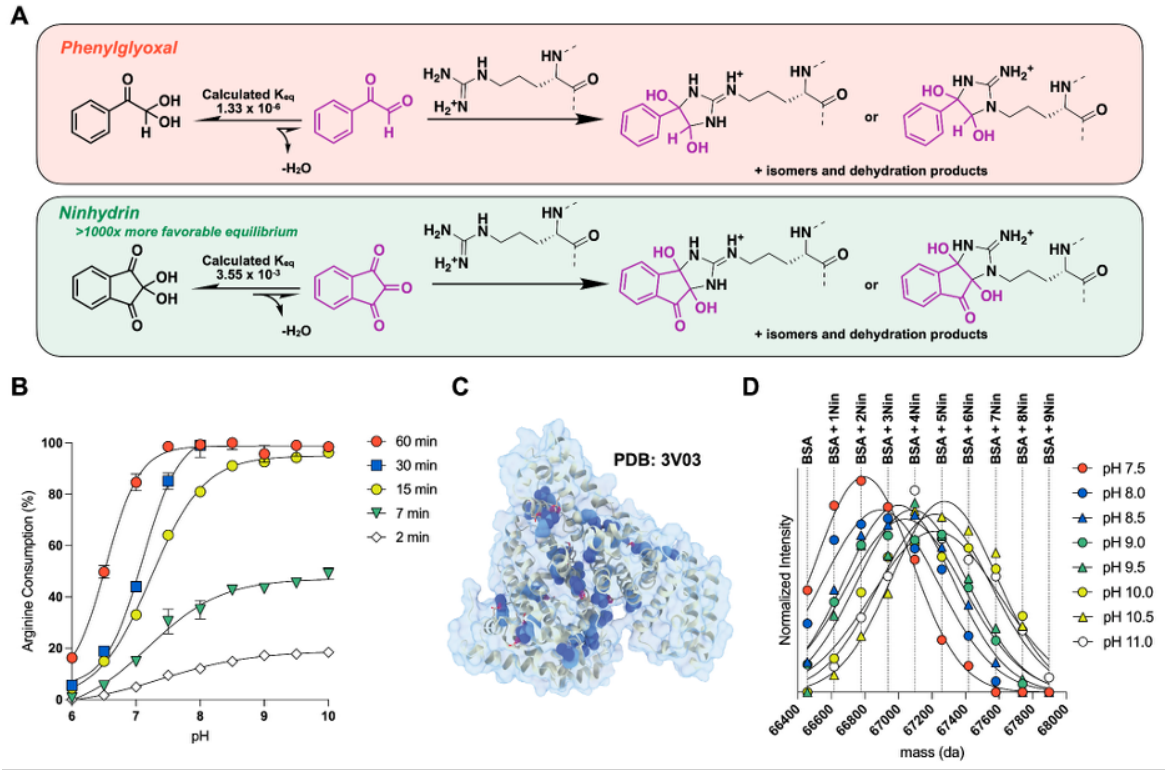


Figure 1.1 | Profiling the reactivity of ninhydrin with arginine in vitro.

A. Proposed reactions between arginine and phenylglyoxal and arginine and ninhydrin, showing formation of stable adducts. Calculated equilibrium constants (Keq) were obtained by density functional theory (DFT) at the B3LYP/6-311G** level with implicit water solvation. **B.** Kinetic analysis of pH- and time-dependent consumption of Fmoc-Arg (500 µM) by ninhydrin (15 mM, 30x) in PBS at 20 °C. Arginine consumption was quantified by LC-MS over 2 to 60 minutes. **C.** Surface rendering of bovine serum albumin (BSA, PDB: 3V03) showing the distribution of 23 arginine residues (dark blue), highlighting potential reactive sites. **D.** Intact protein mass spectrometry analysis of BSA (1.5 µM) treated with 150 µM ninhydrin at 37 °C for 90 minutes across pH 7.5 to 11. Reactions were quenched by dilution into 0.1% aqueous TFA and analyzed by LC–MS on a Xevo G2-XS system.

1.3 Ninhydrin Probe Development.

Encouraged by robust arginine modification on a purified protein, we pursued the development of a ninhydrin chemical probe for testing in complex proteomes. We synthesized two regioisomeric ninhydrin probes (**4-Nin-Alk** and **5-Nin-Alk)** and the reported arginine-reactive probe **PGO**[7] (**Fig. 1.2A**). We visualized their relative reactivity across the proteome using gel-based chemical probe techniques (**Fig. 1.2B**). Mino cell lysates were treated with the indicated probe for 1h. Following incubation, lysates were subjected to treatment with an azide-conjugated TAMRA fluorochrome under Cu(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC)[22]

conditions. Methanol was added, and precipitated proteins were resuspended and separated by SDS-PAGE. In-gel fluorescence scanning revealed dose-dependent increases in labeling across the proteome. As a benchmark, we compared proteome-wide labeling by **4-Nin-Alk** and **5-Nin-Alk** to the pan-cysteine probe iodoacetamide-alkyne (**IA-Alk**)[1]. Ninhydrin probes at 100 μM exhibited proteomic coverage comparable to **IA-Alk** at 10 μM, whereas **PGO** displayed markedly reduced labeling efficiencies, despite a known labeling promiscuity beyond arginine[7]. These results suggested that ninhydrin-based probes achieve improved proteomic labeling efficiencies compared to the best-in-class PGO probe.

We next asked if our ninhydrin probes were capable of labeling live cells in a time-dependent manner. A549 cells were treated for 0, 30, or 60 min with the indicated probe (10 μM). Labeled cells were pelleted, lysed, and subjected to conjugation with azide-TAMRA as described previously. In-gel fluorescence scanning of SDS-PAGE separated proteins revealed time-dependent incorporation of both ninhydrin probes compared to **PGO** (**Fig. 1.2C)**. Having demonstrated the ability of both ninhydrin probes to label proteins, we opted to focus our next studies on the **5-Nin-Alk**, given its superior labeling in cells.

To evaluate **5-Nin-Alk**'s potential as a pan-arginine chemical probe, we developed a workflow utilizing the isotopically distinct azide-bearing desthiobiotin (isoDTB) tags developed by Hacker and co-workers [23] (**Fig. 1.2D**). Cell lysates were treated with **5-Nin-Alk** at either 100 μM or 1 mM for 1h, and the resulting alkyne-labeled arginines were conjugated to light or heavy IsoDTB tags via copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC). Labeled proteome pairs were combined and processed by single-pot, solid-phase sample preparation (SP3), yielding enriched, desalted peptides for subsequent mass spectrometry analysis[24]. We call this pipeline Reactive Arginine Profiling (RAP).

We first performed an open search to directly assess the selectivity of our probe for arginines across the proteome. Open searches characterize the frequency of mass modifications to peptides, which allows for the identification of large mass differences between unmodified peptide sequences and experimentally observed precursors[25]. For our **5-Nin-Alk**-treated lysates, we observed one major mass modification, +701.308600 da, which corresponds to the cyclic mono-dehydration adduct of arginine and **5-Nin-Alk** clicked to the heavy isoDTB tag. A less abundant mass corresponding to the cyclic mono-dehydration adduct of arginine and **5-Nin-Alk** clicked to the light isoDTB tag was also detected (+695.310118 da, **Fig. 1.2E**), which we expected given the 10-fold lower concentration of **5-Nin-Alk** used in the light channel. pLOGO analysis of peptides with a modified arginine found no significant 2D sequence determinants of labeling with some minor enrichment for acidic residues 3-4 residues away (**Fig. 1.2F**). Finally, in further support of arginine selectivity, our RAP samples were prepared in the absence of a reducing agent, which is often required for the trapping of lysine modifications[5]. To evaluate potential free thiol engagement through orthogonal methods, we performed an in vitro glutathione (GSH) reactivity assay[26]. While the **PGO** probe displayed some GSH engagement, the **5-Nin-Alk** probe appeared unreactive, further solidifying that ninhydrin has the potential to function as a selective warhead engaging arginine in complex physiologic environments **(Extended Data Figure 3.3**).

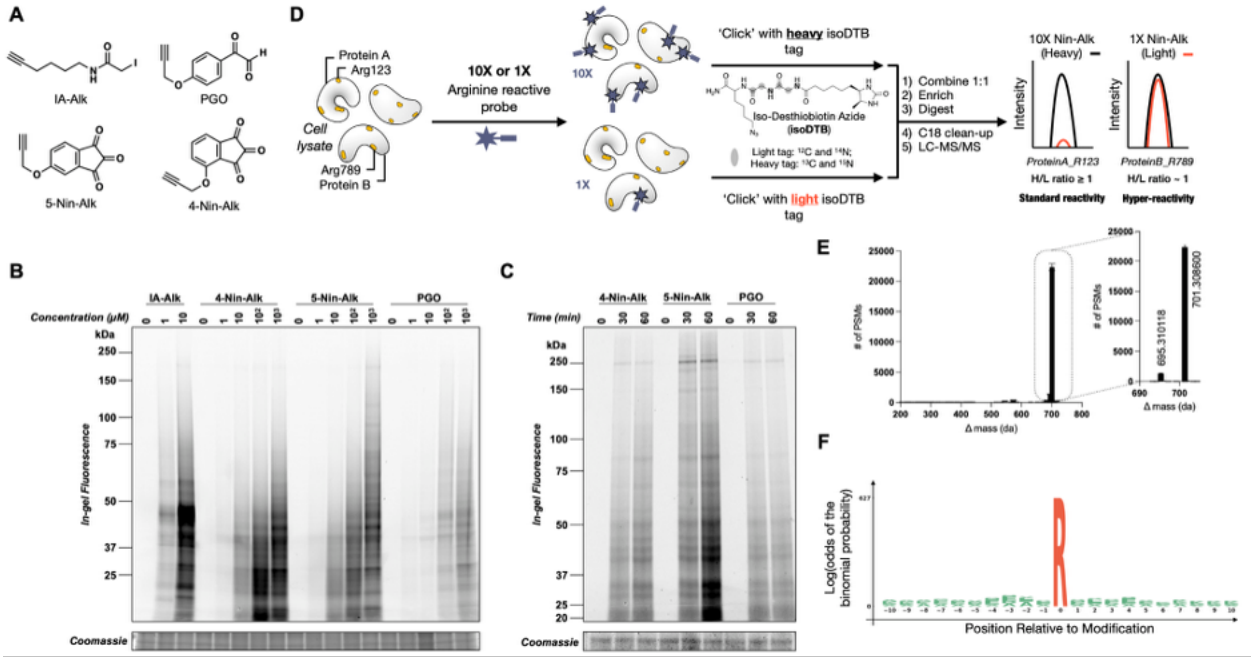


Figure 1.2 | Comparative profiling and biochemical characterization of arginine-reactive electrophiles.

A. Chemical structures of electrophilic probes used in this study, including IA-Alk (iodoacetamide alkyne, cysteine-reactive control), phenylglyoxal (PGO), and two alkyne-functionalized ninhydrin derivatives, 4-Nin-Alk and 5-Nin-Alk. **B.** Qualitative, gel-based assessment of proteome-wide reactivity of alkyne-bearing covalent probes in Mino cell lysate. Soluble proteome (1.5 μg μL^{-1}) was treated with each probe at the indicated concentration for 1 h, followed by CuAAC conjugation with TAMRA-N_3, separation by SDS–PAGE, and in-gel fluorescence scanning. **C.** Qualitative time-dependent reactivity of 4-Nin-Alk, 5-Nin-Alk, and PGO in live cells. A549 cells were treated with 10 μM of the indicated probe for 0, 30, or 60 min, followed by lysis, CuAAC conjugation with TAMRA-N_3, separation by SDS–PAGE, and in-gel fluorescence scanning. **D.** Schematic of the Reactive Arginine Profiling (RAP) workflow to assess arginine reactivity. Cell lysates were treated with an arginine-reactive electrophile at high (1 mM) or low (100 μM) concentrations and subjected to copper-catalyzed azide-alkyne cycloaddition ("click" chemistry) with isotopically light or heavy isoDTB tags. Samples were combined 1:1, enriched, digested, and analyzed by LC-MS/MS. Peptides from hyper-reactive arginine residues show similar heavy/light (H/L) intensities, whereas less reactive sites are preferentially labeled at higher probe concentrations, yielding elevated H/L ratios. **E.** Open-mass search analysis of modified peptides reveals a predominant Δmass of 701.308600 and 695.310118 Da, corresponding to covalent modification of arginine by 5-Nin-Alk and heavy or light isoDTB tag conjugation, respectively. Error bar represents SD **F.** pLogo motif analysis[27] of the ±10-residue sequence window surrounding all modified arginines (foreground: 6,913 modified sites; background: 773,893 human arginines).

Chapter 2. Global Profiling and Mechanistic Studies of Arginine Labeling

2.1 Chemical proteomic profiling of arginine labeling.

Having validated the selectivity of **5-Nin-Alk**, we next leveraged the data to identify reactive arginines embedded in proteins (**Fig. 1.2D**). Raw data collected from the experiment described above were subjected to a modified isoDTB closed-search workflow.[23] From Mino cell lysates, we captured 10,642 modified peptide ions mapping to 6,888 unique arginines across 2,076 distinct proteins (**Fig. 2.1A**). Statistical overrepresentation analysis of proteins containing reactive arginines revealed enrichment for protein classes associated with protein translation, mRNA processing, and nucleic acid binding, among others (**Fig. 2.1B**). We also cross-referenced proteins harboring reactive arginines with Pharos, a database developed by the NIH to categorize a protein's status as a drug target based upon available chemical and biological evidence[29]. This analysis revealed that the majority of proteins harboring reactive arginines are potential therapeutic targets supported by strong biological evidence but currently lack chemical tools for modulation (**Fig. 2.1C**). Complementary annotations from DrugBank further suggest an opportunity to expand covalent small-molecule engagement to a broader set of potentially druggable proteins through reactive arginine engagement (**Fig. 2.1D**).

Inspired by the unique roles arginines play in protein structure and scaffolding, we sought to characterize the broader structural and functional context of **5-Nin-Alk** enriched residues. Through secondary-structure mapping across the PDB, we found reactive arginines are enriched in helices and more flexible loops compared to β-sheets (**Fig. 2.1E**). To gauge local flexibility, we used IUPred2A[31], an energy-based predictor of intrinsic disorder, to compute per-residue scores from 0 (ordered) to 1 (disordered). Reactive arginines peaked in the moderate-disorder regime (0.3 - 0.4) yet ~ 40% reside in well-ordered contexts (scores < 0.2), demonstrating that both structural support and some degree of flexibility facilitate efficient covalent engagement (**Fig. 2.1F**).

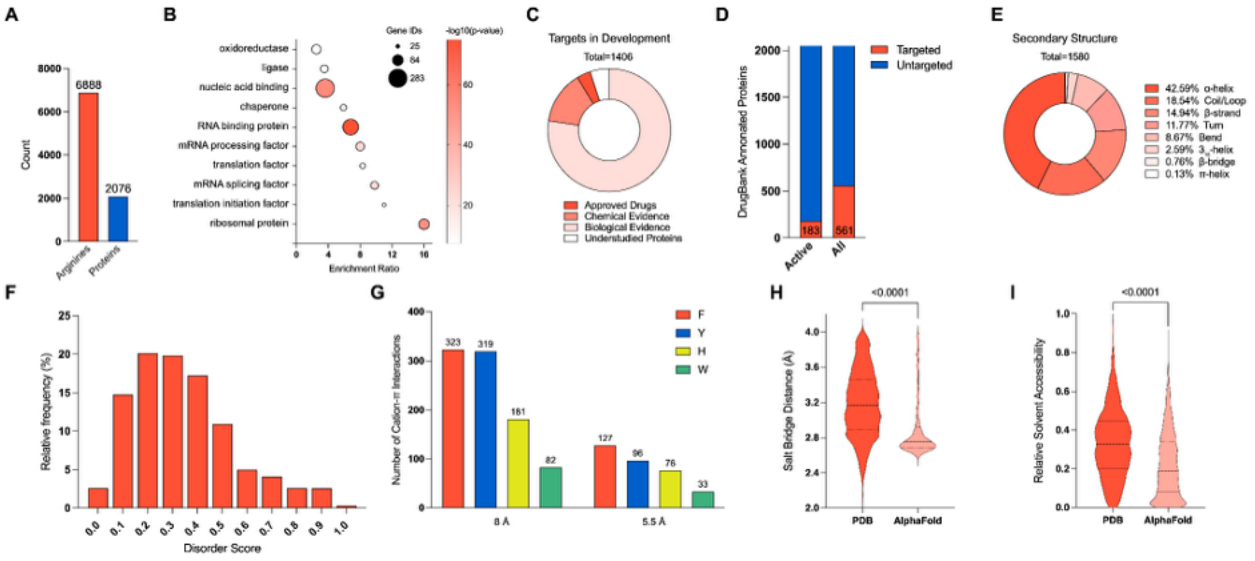


Figure 2.1 | Structural and functional context of arginines identified by RAP in Mino cell lysates. A. Total number of unique arginines and corresponding proteins identified using the RAP workflow (**see Fig. 2D**). **B.** Bubble plot of the top 10 most significantly enriched PANTHER protein classes.28 Statistical significance was calculated using Fisher's Exact test and adjusted using the Benjamini–Hochberg procedure to control the false discovery rate at $\alpha = 0.05$; only terms with adjusted p-value $< 10^{-9}$ are shown. Bubble size reflects the number of proteins in each class. **C.** Target development level of proteins containing reactive arginines, based on Target Central Resource Database classification, where Tclin = Approved Drug, Tchem = Chemical Evidence, Tbio = Biological Evidence, Tdark = Understudied Proteins29. D. Number of proteins detected in our RAP dataset that are annotated as targeted in the DrugBank. **E.** Distribution of secondary structure assignments for reactive arginines, as annotated using DSSP30 on PDB structures. **F.** Histogram of predicted intrinsic disorder scores at modified arginine sites, calculated using IUPred2A. Scores > 0.5 indicate disordered regions. (n = 3630) **G.** Quantification of cation–π interactions between reactive arginines and aromatic residues in PDB deposited structures. Interaction counts are shown at distance thresholds of 8.0 or 5.5 Å, measured from the guanidinium CZ atom to the centroid of the aromatic ring. Geometric constraints were applied for inclusion of parallel interactions (face-on) when the angle between the CZ-to-centroid vector and the normal vector of the aromatic plane was ≤ 30°, and T-shaped interactions (edge-on) when deviation from 90° was ≤ 20°. **H.** Violin plot showing the distribution of salt bridge distances between guanidinium nitrogen atoms of reactive arginines and carboxylate oxygens of acidic residues (Asp/Glu) with discrepancies between PDB (n = 1594) and AlphaFold (n = 1445) models. Welch's t-test (two-tailed, parametric) I. Predicted relative solvent accessibility (RSA) of reactive arginines, computed using the Shrake–Rupley algorithm on PDB (n = 1649) and predicted AlphaFold (n = 4309) structures, indicating an over-representation of arginine surface burial in AlphaFold models compared to PDB models, Welch's t-test (two-tailed, parametric).

Beyond these structural contexts, we turned to cation-π contacts with aromatic side chains, quantifying interactions across all available PDB models for each protein (**Fig. 2.1G**). We applied two distance thresholds, 8 Å, which is the commonly used cutoff, and a more stringent 5.5 Å cutoff for true energetic relevance[3], measured from the arginine CZ atom to the aromatic ring-centroid. To ensure only geometrically plausible face-on and edge-on interactions were included, we added angular filters: for parallel (face-on) interactions, the CZ-to-centroid vector was required to lie

within 30º of the ring normal, while for T-shaped (edge-on) interactions, deviation from 90º had to be ≤ 20º. Using these criteria, we identified 905 total cation-π contacts at 8.0 Å, while only 332 satisfied the stricter 5.5 Å threshold (**Fig. 2.1G**). Phenylalanine and tyrosine accounted for the majority of these contacts, compared to histidine and tryptophan. At the more stringent distance threshold, the distribution of aromatic partners mirrors the natural abundance of aromatic residues in the human proteome. Unlike recent efforts to selectively label tryptophans that showed enrichment for tryptophans engaged in cation-π interactions[8], we do not see an enrichment of arginines engaged in cation-π interactions (**Extended Data Fig. 3.4A**). This indicates that arginine reactivity is not biased by a specific type of π-donor or the stabilizing effect of a cation–π interaction.

We next measured salt bridges as the minimum distance between any arginine guanidinium nitrogen and Asp/Glu carboxylate oxygens for both PDB entries and AlphaFold[33] models. In AlphaFold structures, these distances cluster tightly around ~2.6 to 2.8 Å (**Extended Data Fig. 3.4B, 3.4C, and 3.4D**), whereas PDB entries span a broader 2.5 to 4.0 Å range with a median closer to 3.2 Å (**Fig. 2.1H, Extended Data Fig. 3.4E and 3.4F**). Likewise, relative solvent accessibility values derived from solvent accessible surface area (SASA)[34] analysis on PDB models are systematically higher and more variable than those from AlphaFold (median RSA for AlphaFold predicted models = 0.19, median RSA for PDB experimental models = 0.33; (**Fig. 2.1I, Extended Data Figure 3.5A, 3.5B**). Although AlphaFold models offer near proteome-wide coverage, their monomeric, in-silico nature often yields overly "idealized" geometries that may not reflect the full range of biologically relevant conformations[35,36]. This was exemplified in our secondary structure analysis on AlphaFold models, which showed a greater enrichment of alpha helices than determined in the PDB models (**Fig. 2.1E and Extended Data Fig. 3.5C**), which is a

common feature of AlphaFold structures[37]. To address this, we prioritized experimentally determined PDB structures, trading breadth for depth by focusing on a smaller set of proteins with multiple crystallographic or complexed snapshots that better approximate physiological contexts. In both cases, the experimental data capture greater conformational diversity and exposure compared to the predicted models, which we believe are critical parameters for accurately mapping covalent-probe hotspots.

Transitioning from these broader structural contexts, we sought to dissect the biophysical and functional determinants that govern differential arginine reactivity using datasets of reactive and hyper-reactive arginines (**Fig. 2.2A**). We defined hyperreactive arginines as those which display similar enrichment at high and low concentrations (comparative ratio of intensity between conditions of less than 2), whereas less reactive arginines display concentration-dependent labeling (comparative ratio of at least 2) (**Fig. 2.2A and Fig. 1.2D**). The median $\log_2$ ratio of enrichment between 1 mM and 100 μM **5-Nin-Alk** across all arginines was 2.50, reflecting dose-dependent labeling of most reactive arginines. Notably, 381 sites (~ 5.5% of detected arginines) were hyperreactive, fully engaged at 100 μM, while the remaining 6478 sites exhibited clear, concentration-dependent labeling. This percentage of arginines annotated as hyper-reactive is consistent with reports for cysteine and arginine[1,2,4]. Moreover, proteins typically harbor a single hyper-reactive arginine but can contain multiple concentration-dependent reactive arginines (**Fig. 2.2B**).

With our lists of either all reactive (any arginine engaging ninhydrin) or hyper-reactive (enrichment ratio less than 2) arginines, we set out to identify distinguishing features between the two sets. Even though arginine remains protonated under physiological conditions, cysteine and lysine reactivity have been linked to residue pKa values[38,39], so we first asked whether arginine

labeling follows a similar trend. Predicted pKa values for all reactive arginines within the PDB were, as expected, tightly clustered around 12.5. Distributions for hyper-reactive versus all reactive arginine sites showed no significant differences (**Fig. 2.2C**), indicating that, unlike lysine (**Extended Data Fig. 3.6B**), intrinsic protonation state does not drive labeling efficiency. This finding was consistent for predicted AlphaFold structures (**Extended Data Fig. 3.6A**). Given the established importance of solvent accessibility for cysteine modification by iodoacetamide[40], we next assessed the relative solvent exposure of reactive arginines. SASA analysis revealed comparable relative solvent exposure distributions for hyper-reactive and all reactive sites, with both populations exhibiting partial rather than full surface exposure (median RSA for hyper-reactive = 0.31, median RSA for all reactive = 0.33, mean RSA for both = 0.33; **Fig. 2.2D**). This pKa and solvent accessibility independence of arginine labeling highlights a prime advantage for probe development: arginine-directed covalent targeting can reliably engage targets across diverse structural and environmental contexts, enabling the design of broadly applicable and highly selective chemical probes.

To examine whether salt-bridge geometry contributes to hyper-reactivity, we compared the salt bridge distances for hyperreactive and all reactive arginines in PDB models (**Fig. 2.2E**). Both populations exhibit broad, overlapping distributions (median distances = 3.2 Å for both hyperreactive and all reactive sites), indicating that salt-bridge proximity alone does not distinguish the hyper-reactive subset. Lastly, we explored functional constraints by mapping mean pathogenicity scores (AlphaMissense) onto each reactive arginine (**Fig. 2.2F**). A high pathogenicity score indicates that mutations of a given residue are associated with significant functional impairment or disease-related phenotypes[41]. Both hyperreactive and all reactive residues are enriched at the high end of the pathogenicity scale, with 51% of enriched arginines

falling into the likely pathogenic range, indicating a labeling bias towards arginines that are mutationally constrained, likely reflecting their critical structural or functional roles in maintaining cellular integrity. Since these arginines play critical roles in cellular function, they are less likely to undergo genetic mutations that could evade covalent engagement. Thus, covalent engagement at these essential positions likely mimics genetic perturbation, directly impacting residue functionality and consequently cellular processes, making them attractive and robust targets for drug and tool compound development.

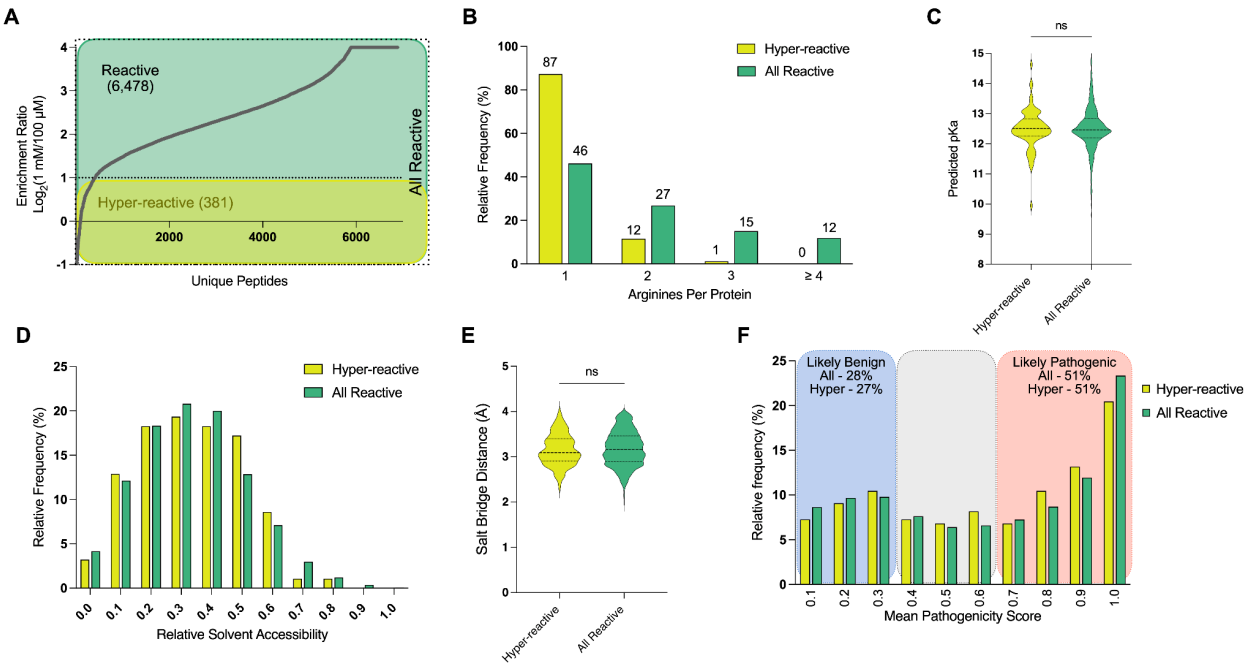


Figure 2.2 | Biophysical and functional features between reactive arginine residues. A. Waterfall plot showing the distribution of median RAP enrichment scores (log_2 heavy-to-light [H/L] ratios) for all quantified arginines across 4 biological replicates. Arginine residues with log_2(H/L) ≤1 and detected in ≥2 unique peptides were classified as hyper-reactive (n = 381), representing a subset of the total reactive arginines identified (n = 6,859). See **Fig. 2D** for a schematic of the RAP workflow. **B.** Distribution of reactive and hyper-reactive arginines per unique protein. **C.** Distribution of predicted pKa values for hyper-reactive (n = 85) and all reactive arginine residues (n = 1570), calculated using PROPKA3 and the PDB.42,43 Welch's t-test (two-tailed, parametric) **D.** Predicted relative solvent accessibility (RSA) of hyper-reactive (n = 93) and reactive arginines (n = 1649), computed using the Shrake–Rupley algorithm34 on PDB structures, indicating a shared preference for surface exposure. **E.** Violin plot showing the distribution of salt bridge distances between guanidinium nitrogen atoms of hyper-reactive (n = 101) or all reactive arginines (n = 1445) and carboxylate oxygens of acidic residues (Asp/Glu) in PDB models. Welch's t-test (two-tailed, parametric) F. Mean pathogenicity score of mutations to each hyper-reactive (n = 220) and reactive arginine (n = 4865), predicted using AlphaMissense.

2.2 Potential mechanism for ninhydrin inhibition of aconitase.

Ninhydrin has been reported to inhibit the catalytic activity of mitochondrial aconitase at doses comparable to those used in our assays[44], but the mechanism of this inhibition has not been elucidated. Mitochondrial aconitase plays a central role in the tricarboxylic acid cycle, linking cellular energy production to redox homeostasis and iron-sulfur cluster maintenance[45]. Dysregulation of its activity has been implicated in metabolic disorders, neurodegeneration, and tumor progression[46]. While not widely targeted clinically, mitochondrial aconitase has garnered interest as a metabolic vulnerability in cancer and as a redox-sensitive node in neurodegenerative diseases, with small molecules and metabolic poisons such as fluoroacetate demonstrating proof-of-concept for pharmacological modulation[47]. Our **5-Nin-Alk** hyper-reactivity RAP dataset revealed a single modified reactive arginine on aconitase, R479, despite the presence of multiple arginine residues in the active site. Structural analysis of an aconitase crystal structure (PDB 1B0J)[48] identified R479 as a key residue in the positioning of aconitase's native substrate (**Fig. 2.3A**). Covalent docking simulations to this crystal structure further demonstrated that ninhydrin modification at this site would sterically occlude the active site, likely preventing substrate binding (**Fig. 2.3B**). These findings provide mechanistic insight into how covalent modification of R479 influences aconitase function and suggest that ninhydrin-mediated modification of at this arginine contributes to the observed inhibitory effect.

2.3 Development of a pipeline for ligand-directed arginine modification.

To systematically identify which reactive arginines in our RAP dataset would be amenable to targeting with ninhydrin-endowed ligands, we developed an in-house computational structural proteomics pipeline (**Fig. 2.3C**). The workflow compiles a list of UniProt IDs[49] alongside their respective modified arginine residues and cross-references them with available high-resolution

structures in the Protein Data Bank (PDB).[42] The analysis parsed through 9,348 unique PDB files, identifying 33,402 occurrences of arginines from our dataset. For each structurally resolved arginine, the pipeline identifies nearby ligands within 3 Å of the guanidinium group. From these occurrences, 292 arginines were found to engage in 884 unique contacts with 461 distinct ligands (data not shown). To prioritize functionally relevant ligands, we exclude primary human metabolites, crystallographic ions, and buffer components.

Applying this strategy, we identified R55 in cyclophilin A (CypA) as a reactive arginine proximal to recently reported tri-vector ligands[50] that provide synthetically tractable starting points for medicinal chemistry (**Fig. 2.3D** and **2.3E**). Structure-guided ligand optimization led to the development of **4-CA-Alk** (**Fig. 2.3D**), where the aniline NH_2 was replaced with an isosteric trifluoromethyl group, the urea was substituted with an amide, and the ethyl ester was replaced with a ninhydrin warhead. This design positioned the ninhydrin warhead appropriately for covalent engagement with R55 in our docking models to the crystal structure of cyclophilin A in complex with a tri-vector ligand (PDB 6GJI)[50] (**Fig. 2.3F**). A regioisomer, featuring the amide substitution at the 5 position of ninhydrin, was designed to spatially disfavor R55 modification, **5-CA-Alk**.

To quantitatively assess the proteome-wide selectivity of **4-CA-Alk**, we performed a probe-versus-no-probe RAP experiment. At 100 μM, **4-CA-Alk** enriched 287 modified peptides, indicating that 4-CA-Alk selectively modifies a limited subset of reactive arginines under physiologic conditions (data not shown). At 10 μM, only 7 modified peptides were detected, demonstrating a marked reduction in off-target engagement at lower concentrations (**Fig. 2.3G**). Notably, R55 of cyclophilin A was among these 7 peptides. The ligand scaffold likely plays a key role in directing this selectivity by engaging the binding pocket containing R55 and restricting the modification of other reactive arginines throughout the proteome. A comparative RAP analysis at

1 mM between **4-CA-Alk** and its regioisomer **5-CA-Alk**, which mispositions the ninhydrin warhead relative to R55, revealed both a significantly reduced overall reactivity for **5-CA-Alk** at CypA R55 and a median $\log_2$ selectivity between the probes ratio of 2.49 in favor of **4-CA-Alk** (**Fig. 2.3H**, **Extended Data Fig. 3.7**). These findings underscore the importance of precise warhead placement within the ligand scaffold to achieve selective and site-specific arginine labeling.

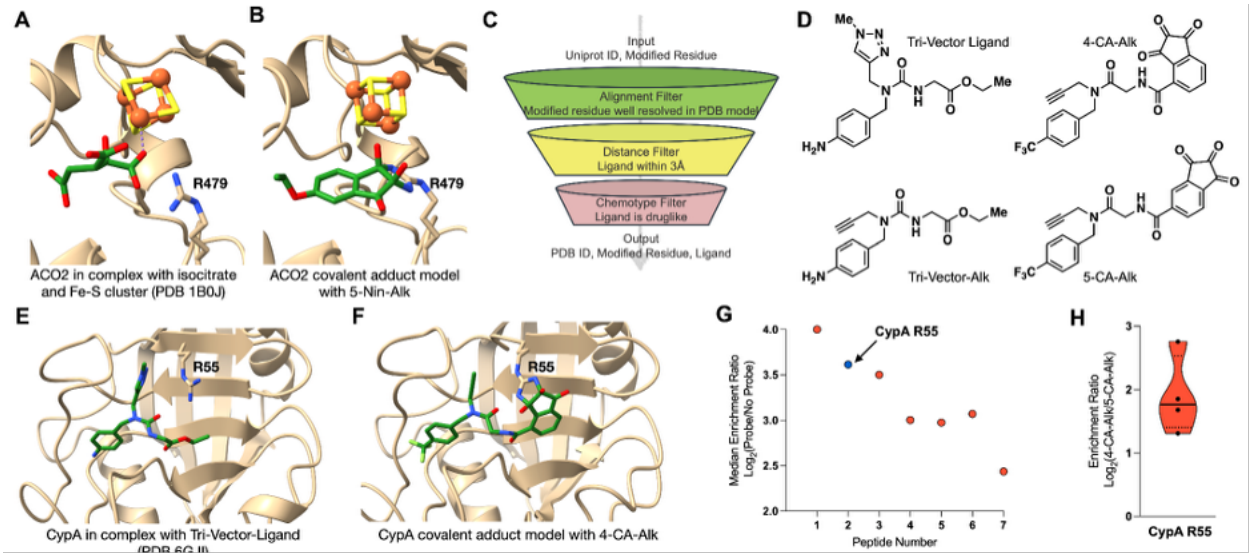


Figure 2.3 | Structure-guided development of arginine-selective ligands. A. Crystal structure of mitochondrial aconitase (ACO2) bound to isocitrate and a [4Fe-4S] cluster, highlighting the proximity of R479 to the active site (PDB: 1B0J). **B.** Covalent docking model of **5-Nin-Alk** to R479 of ACO2, showing spatial alignment with the native isocitrate ligand. **C.** Schematic of the in-house structure-based filtering pipeline used to identify candidate ligands in proximity to modified arginines. Input UniProt IDs and modified residue positions were filtered by PDB alignment, ligand proximity (≤3 Å), and drug-likeness of the ligand to prioritize hits for follow-up. **D.** Chemical structures of the tri-vector ligand and tri-vector alkyne, a clickable analog previously shown to bind cyclophilin A (CypA). The tri-vector alkyne contains an alkyne moiety that enables CuAAC-based enrichment and inspired the synthesis of **5-CA-Alk** and **4-CA-Alk**. **E.** Published co-crystal structure of CypA bound to the tri-vector ligand (PDB: 6GJI),[50] showing R55 positioned at the periphery of the ligand binding site. **F.** Covalent docking model of **4-CA-Alk** to R55 of CypA, showing similar binding orientation as the tri-vector ligand. **G.** Chemoproteomic enrichment of peptides labeled by **4-CA-Alk** (10 µM, 1h) in a RAP experiment. Each data point represents a distinct modified peptide; covalent labeling of R55 in cyclophilin A (CypA) is highlighted in purple. **H.** Relative enrichment of CypA R55 between **4-CA-Alk** and **5-CA-Alk** as determined by RAP. n = 4 replicates.

2.4 Discussion

Covalent probes have transformed chemical biology and drug discovery, but broadly expanding covalent strategies to amino acids beyond cysteine and lysine has been challenging. Our study establishes ninhydrin-based probes as selective and robust tools for covalent arginine labeling, significantly expanding the chemical proteomics toolkit and enabling new approaches to study arginine-mediated processes. Our ninhydrin-derived probes leverage the equilibrium between the hydrate and reactive dehydrate states of vicinal dicarbonyls. This strategy optimizes probe reactivity while minimizing non-specific interactions and positions arginine as a promising target for broadly applicable covalent strategies. Unlike cysteine and lysine, whose reactivities heavily depend on intrinsic residue pKa and high solvent accessibility, arginine residues demonstrate consistent reactivity independent of pKa variations or extensive solvent exposure. This independence is particularly advantageous for therapeutic targeting as it allows for covalent engagement in structurally diverse and otherwise challenging environments. Both in vitro and cell-based studies revealed that modified arginines are often partially buried rather than fully solvent-exposed, emphasizing that determinants of arginine reactivity remain poorly defined, unlike other amino acids, where factors such as pKa or solvent exposure more clearly predict reactivity (**Figs. 1.1C, 2.1I and 2.2D**).

Arginines are frequently localized at protein-protein interaction (PPI) sites, where their covalent modification is poised to directly modulate or disrupt critical interaction surfaces. Both cation-π interactions and salt bridges are more prevalent in proteins from thermophilic organisms, enhancing protein stability. We observe that reactive arginines frequently engage in these stabilizing interactions, although these features were not enriched in hyper-reactive arginines compared to generally reactive ones. This trend is consistent with the absence of significant

differences in predicted pKa or solvent accessibility between the two groups. In our RAP dataset, we identify 290 unique arginines stabilizing PPIs within the PDB, forming 519 unique PPI pairs. Among these, 145 arginines participated in 193 cation-π interactions, and 212 arginines were involved in 326 unique salt bridge pairs. Despite the prevalence of these specific interactions, our analysis found that fully solvated arginines are underrepresented based on their physicochemical properties. This rarity may partly reflect the inherent dynamic nature of fully solvated arginines, as their positions might only be reliably resolved when stabilized at interaction interfaces, resulting in relatively lower RSA values being captured in our bioinformatic analyses. Importantly, these interactions provide valuable insights into structural contexts that could guide rational optimization of arginine-targeted covalent probes.

Integration of experimental PDB structures significantly enhanced our structural analyses compared to predictions based on computational AlphaFold models. Although AlphaFold offers extensive near-proteome-wide coverage, these models frequently depict idealized conformations, lacking critical contextual interactions present in biological environments. By contrast, the use of PDB structures, though computationally more intensive, owing to the need for extensive sequence alignments, the handling of large cryo-EM complexes, and the presence of multiple structural entries for individual proteins, provided more realistic and biologically representative data. Specifically, experimental structures captured greater diversity in salt bridge geometries, solvent accessibility, and interaction distances (**Extended Data Fig. 3.8A and 3.8B**), all crucial for accurately predicting arginine reactivity (**Fig. 2.1H, 2.1I, and Extended Data Fig. 3.4A – 3.4E**). This approach underscores the value of incorporating experimentally determined structures when studying residue-level reactivity, particularly for interpreting complex biophysical features that may be underrepresented in predicted models.

Recent work studying iodoacetamide-cysteine reactivity found that simple features such as 2D sequence, pKa or even data derived from AlphaFold models were not sufficient to predict a cysteine's relative reactivity. More accurate prediction required experimentally determined structures to start to model residue-level reactivity[51]. While AlphaFold offers transformative accessibility and continues to enable exciting applications such as predictive covalent docking, the current generation of models often struggles with accurate representation of sidechain packing and interaction geometries. As we expand our platform to profile reactive arginines across more cell types, the growing bulk of arginine reactivity data may make AlphaFold-derived structures increasingly useful at scale, particularly given the computational demands of mining the PDB. Future iterations of Alphafold, which are showing improvements in sidechain accuracy, may further enhance its utility for predicting covalent probe engagement.

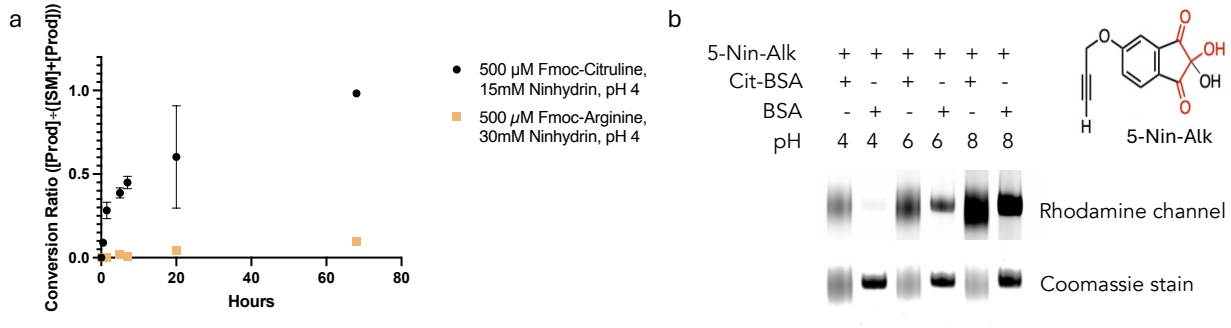
Selective modification of functionally relevant arginines is exemplified by ninhydrin-mediated labeling of aconitase at R479. Previous studies suggested that ninhydrin inhibits aconitase catalytic activity, and the RAP data extends this hypothesis by providing a putative molecular mechanism of inhibition(**Fig. 2.2A and 2.2B**). Structural analysis confirmed that R479 is critical for substrate positioning, and covalent modeling indicated that modification at this site sterically occludes substrate binding, contributing to inhibition. These findings highlight the potential of selective arginine modification as a powerful approach for probing enzyme function and developing covalent inhibitors targeting active-site arginines. Beyond aconitase, the selective modification of cyclophilin A at R55 further underscores the utility of ninhydrin-based probes in ligand-directed covalent targeting. Our compound 4-CA-Alk exhibited dose-dependent selectivity, engaging a limited subset of peptides, including cyclophilin A R55 at 10 μM (**Fig. 2.3**). Structural comparisons with a regioisomeric probe (5-CA-Alk), which mispositioned the ninhydrin warhead

relative to R55, emphasized the importance of precise warhead placement in achieving selective, site-specific arginine engagement. These results demonstrate how integrated computational modeling, proteomics, and structure-guided ligand design systematically guide the development of selective covalent ligands.

This study establishes a robust framework for selective covalent arginine modification using ninhydrin-derived probes. We introduce RAP as a chemical proteomic approach to classify arginine reactivity, facilitating rational probe development and the identification of key arginines for therapeutic targeting. Future research will focus on refining probe specificity, elucidating additional structural determinants of arginine reactivity, expanding the application of ninhydrin-based probes to a broader range of therapeutically relevant targets, and leveraging the RAP platform as a competitive screening tool for the discovery of arginine-reactive small molecules. Importantly, the ability to perform competitive screening without requiring enrichment handles will simplify the evaluation of candidate molecules, enhancing throughput and reducing synthetic complexity, thus accelerating the development of novel arginine-targeted therapeutics. By advancing strategies for arginine-selective covalent modification, this work paves the way for novel chemical biology tools and next-generation covalent therapeutics.

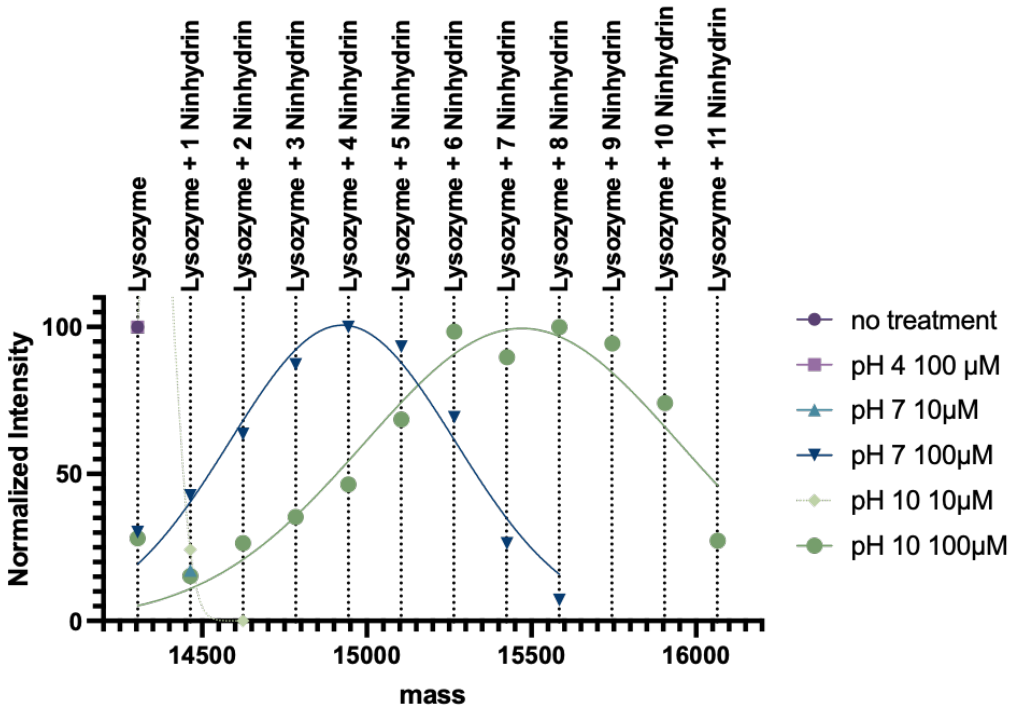
Chapter 3. Experimental Methods and Supporting Information

3.1 Extended Data Figures



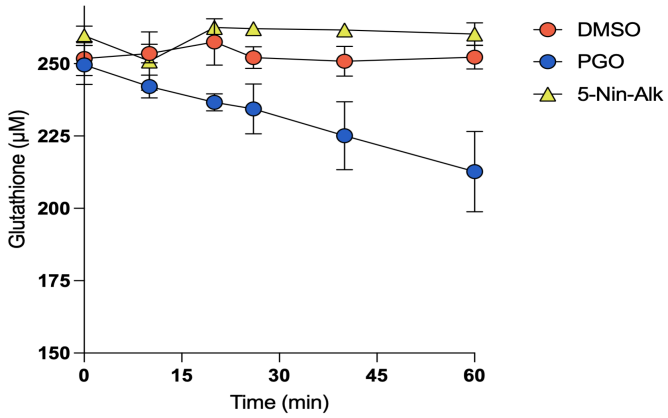
Extended Data Figure 3.1 | Ninhydrin displays selectivity for citrulline at pH 4.

A. Kinetic analysis of pH- and time-dependent consumption of Fmoc-Arg (500 µM) or Fmoc-Cit by ninhydrin (15 mM, 30x) in pH 4.0 PBS at 20 °C. Consumption was quantified by UV absorption by HPLC over 2 minutes to 65 hours **B.** Qualitative, gel-based assessment of proteome-wide reactivity of alkyne-bearing ninhydrin (5-Nin-Alk) with either citrullinated BSA(Cit-BSA) or commercially available BSA (1.5 µg µL^{-1}). Each were treated with ethe probe at the indicated concentration for 1 h, followed by CuAAC conjugation with TAMRA-N_3, separation by SDS–PAGE, and in-gel fluorescence scanning.



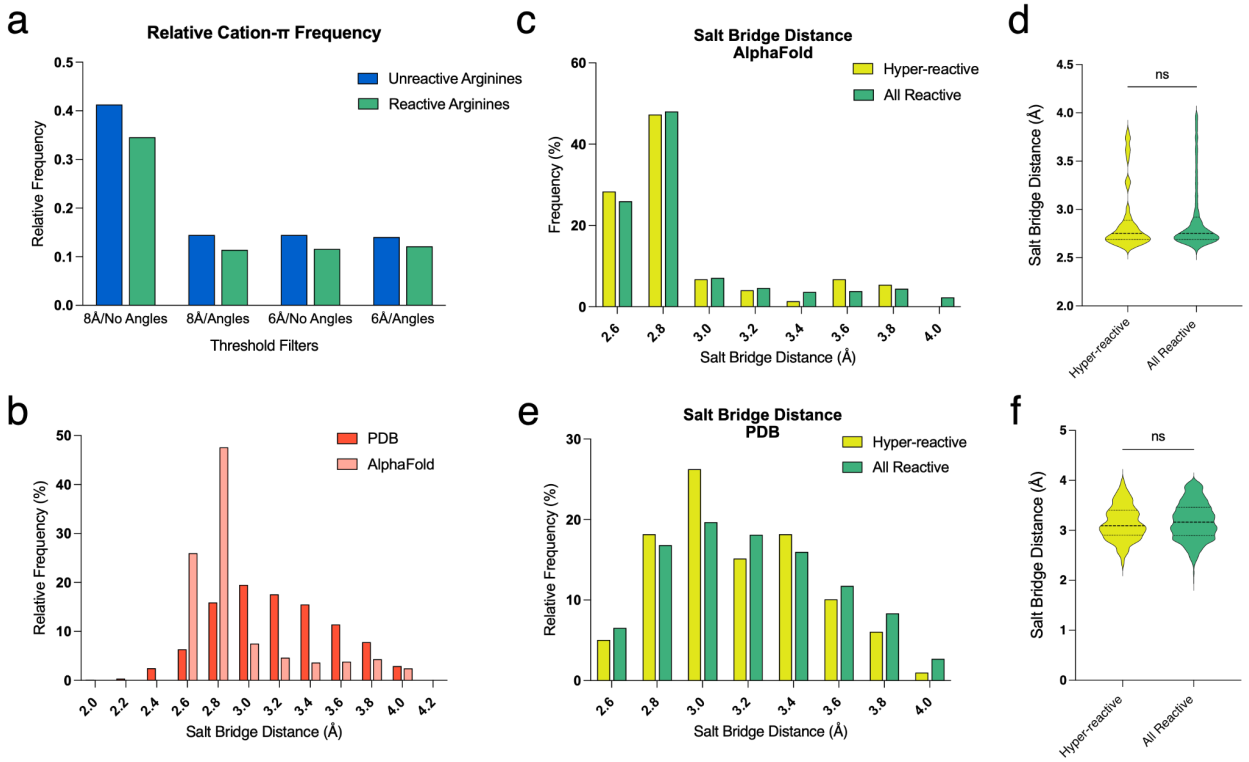
Extended Data Figure 3.2 | Lysozyme is covalently modified by ninhydrin.

Intact protein mass spectrometry analysis of Lysozyme (1.5 µM) treated with 10 or 100 µM ninhydrin at 37 °C for 90 minutes across pH 4.0, 7.0 and 10.0. Reactions were quenched by dilution into 0.1% aqueous TFA and analyzed by LC–MS on a Xevo G2-XS system.



Extended Data Figure 3.3 | Glutathione reactivity assay for arginine reactive warheads.

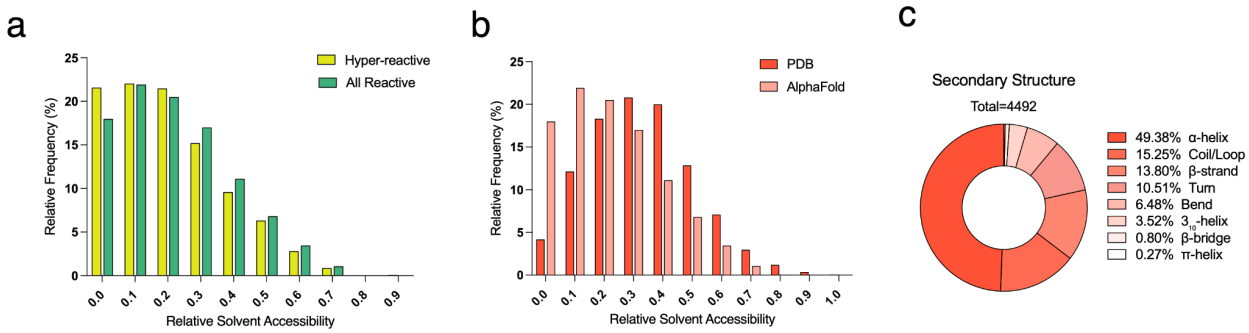
Compounds were incubated with reduced glutathione (250 μM) in assay buffer (100 mM Tris pH 9.0, 50% MeOH) at a final concentration of 500 μM. At the indicated time points (0–60 min), Ellman's reagent was added and residual free thiols were quantified via absorbance at 412 nm. Values reflect means of triplicate measurements with individual replicates shown. DMSO control (orange circles), phenyl glyoxal (blue circles), 5-Nin-Alk (yellow triangles).



Extended Data Figure 3.4 | Comparative analysis of intermolecular interactions surrounding reactive arginines.

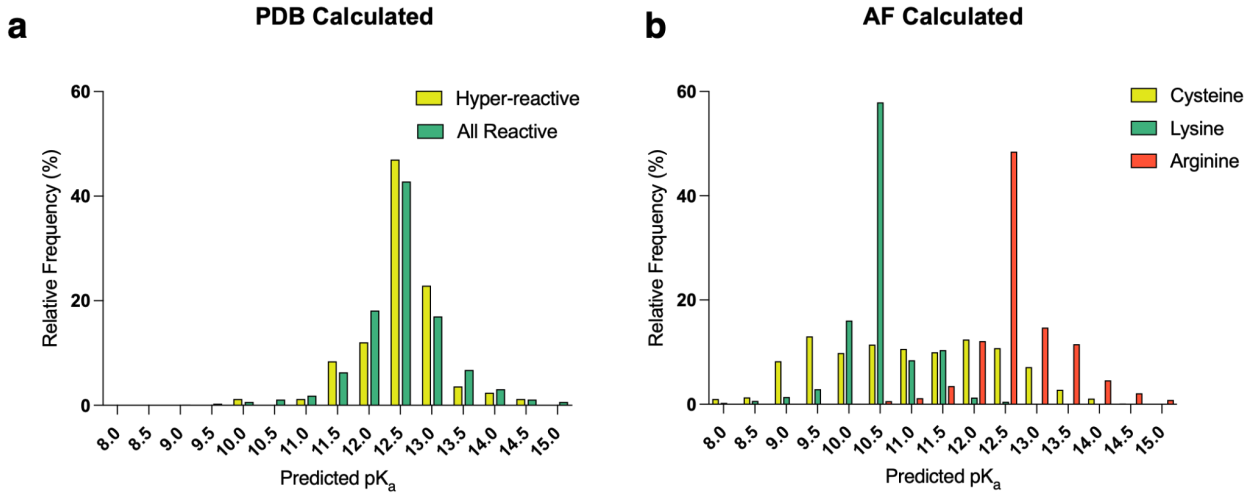
A. Frequency of arginines engaged in cation–π interactions across reactive (green) and unreactive (blue) datasets, based on AlphaFold-predicted models and varying distance (6 Å or 8 Å) and angle filters. No significant enrichment or depletion of cation–π interactions was observed among reactive residues. **B.** Comparison of salt bridge distances between reactive arginines and acidic residues in AlphaFold (light coral, n = 1594) versus experimentally determined PDB structures (terra cotta, n = 1445), showing generally shorter predicted distances in AlphaFold models. **C.** Histogram of salt bridge distances from AlphaFold models comparing (Figure caption continued on the next page)

(Figure caption continued from the previous page) hyper-reactive (yellow, n = 74) and all reactive (green, n = 1594) arginines. **D.** Violin plot showing the distribution of salt bridge distances from AlphaFold models for hyper-reactive and all reactive arginines; no statistically significant difference observed (unpaired t-test). **E.** Histogram of salt bridge distances derived from experimental PDB structures comparing hyper-reactive (yellow, n =101) and all reactive arginines (green, n = 1445). **F.** Violin plot of salt bridge distances in PDB models; consistent with AlphaFold data, no statistically significant difference was observed between hyper-reactive and all reactive arginines.



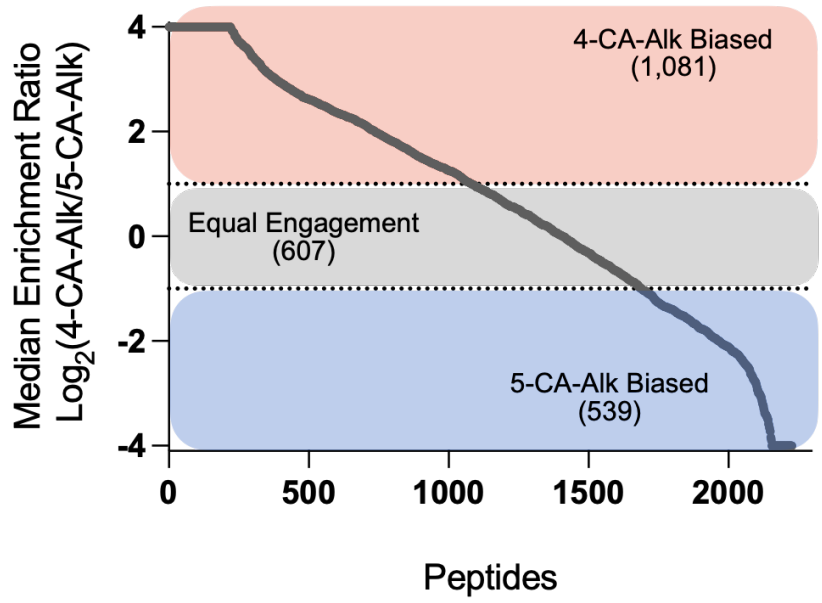
Extended Data Figure 3.5 | Solvent accessibility and secondary structure profiles of reactive arginines.

A. Distribution of relative solvent accessibility (RSA) for hyper-reactive (yellow, n = 1960) versus all reactive (green, n = 4309) arginines in AlphaFold-predicted structures. Both sets show a similar solvent exposure profile. **B.** Comparison of RSA values for all reactive arginines across AlphaFold (light coral, n = 4309) and PDB (terra cotta, n = 1649) models reveals a modest shift toward higher accessibility in AlphaFold predictions. **C.** Secondary structure distribution of all reactive arginines based on DSSP annotation of AlphaFold models (n = 4492 residues). The majority localize within α-helices (49.38%), followed by coil/loop regions and β-strands.



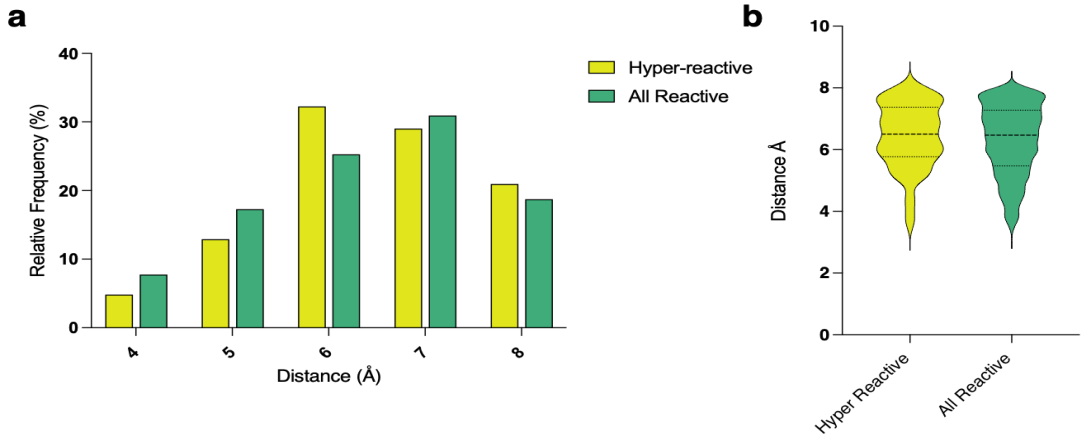
Extended Data Figure 3.6 | Predicted pKa distributions of reactive residues.

A. Distribution of PROPKA-predicted pKa values for reactive arginines calculated using experimental PDB structures, comparing hyper-reactive (yellow, n = 85) and all reactive (green, n = 1570) residues. **B.** Predicted pKa distributions for reactive arginines(orange, n = 4334), cysteines(yellow, n = 3988), and lysines (green, n = 3863) using AlphaFold structures. Arginine predictions are from the current dataset, cysteine data was based on matched cell analysis with Mino cell lysate and iodoacetamide alkyne as the electrophile and lysine data are derived from previously reported reactivity profiles from Hacker et al. (2017)4. No significant pKa shift distinguishes hyper-reactive from reactive residues across any residue class.



Extended Data Figure 3.7 | Waterfall plot depicting the regioselective enrichment of peptides labeled by 4-CA-Alk and 5-CA-Alk in Mino cell lysate.

Peptides identified by the RAP workflow were classified based on relative enrichment at 1 mM probe concentration in biological triplicates. Each point represents the $\log_2$ ratio of 4-CA-Alk to 5-CA-Alk enrichment for an individual peptide. Peptides were grouped into 4-CA-Alk biased (red, n = 1,081), 5-CA-Alk biased (blue, n = 539), and equally engaged (gray, n = 607) categories



Extended Data Figure 3.8 | Structural comparison of cation–π interactions at hyper-reactive versus general reactive arginines.

A. Histogram depicting the frequency distribution of cation–π interactions among hyper-reactive arginines (yellow, n = 62) compared to all reactive arginines (green, n = 905) based on structural analysis of the experimental models from the PDB. **B**. Violin plot showing the same comparison as in (**A**), highlighting the broader distribution of cation–π interactions across the two arginine populations.

3.2 Methods

Density Functional Theory (DFT) Calculations Initial molecular models were prepared using Avogadro[52]. All calculations were performed using GAMESS[53] on ChemCompute.org[53]. Geometry optimization was performed with the PM3 basis set, restricted Hartree-Fock (RHF) molecular orbital method, and B3LYP DFT functional. Solvent effects were modeled using the water PCM solvent model on a single processor core. The optimized structures were subjected to a single-point energy calculation with the 6-311G** basis set. All output models were reviewed in Avogadro for realistic geometry, and the delta G free energy in solvent between the hydrated and dehydrated states (with an additional water molecule) for both phenyl glyoxal and ninhydrin was calculated to determine the relative equilibrium constants.

Arginine Kinetics Assay In triplicate, 190 µL of PBS buffer (pH 6.0 – 10.0, in 0.5 pH unit increments) was allotted into 27 Eppendorf tubes. A 5 µL aliquot of a 20 mM stock solution of Fmoc-Arg-OH in DMSO was added to each tube, mixed thoroughly, and briefly centrifuged to ensure complete pooling of the solution. A 600 mM Ninhydrin solution (1:1 DMSO/DI water) was prepared alongside a 2 M HCl quenching solution. All tubes were pre-opened, and a timer, vortex mixer, waste container, and fresh pipette tips were arranged for efficient handling. Reactions were initiated by adding 5 µL of Ninhydrin solution to the first tube while starting the timer, followed by immediate capping and vortex mixing. This process was repeated every 10 seconds for the subsequent tubes, ensuring all reactions were initiated within 4.5 minutes. At designated time points, the reactions were quenched by sequentially adding 10 µL of 2 M HCl to each tube, following the same 10-second interval sequence as for reaction initiation. Samples were then centrifuged, and 200 µL of the supernatant was carefully transferred using gel-loading tips into pre-labeled LC-MS vials with conical inserts, ensuring minimal air bubble formation. Samples

were analyzed using a Waters Acquity UPLC SQD2 detector LC-MS system on a polar gradient method (20–55% acetonitrile/water + 0.1% formic acid). The UV detection window was set to 300 nm for Fmoc absorbance, and free arginine (SM) and adduct formation were identified based on the extracted ion chromatograms. Peak integrations at 300 nm were used to calculate the conversion ratio of reactants to products.

Intact Protein Mass Spectrometry Reactions were performed with 1.5 μM BSA in PBS, incubated with 150 μM ninhydrin at 37 ºC for 90 minutes. Reactions were quenched by a 100% dilution with 0.1% TFA (aq) and analyzed by intact protein LC/MS using a Waters Xevo G2-XS system equipped with an Acquity UPLC BEH C4 1.7 μm column. The mobile phase was a linear gradient of 5–95% acetonitrile/water + 0.05% formic acid. The spectra were processed using QuantLynx to determine the extent of modification.

Glutathione reactivity assay Reduced glutathione (GSH) was prepared at a concentration of 250 μM in assay buffer (100 mM Tris pH 9.0, 50% MeOH as co-solvent). In triplicate, 100 μL of GSH (250 μM) in assay buffer was added to each well of a clear, flat-bottom 96-well plate (Greiner Bio-one). To each well, 1 μL of 100x inhibitor stocks in DMSO (final concentration 500 μM) or DMSO vehicle was added, and the plate was incubated at room temperature for the indicated time (0–60 min). Following incubation, Ellman's reagent (Thermo) (10 μL, 50 mM stock in DMSO, final concentration 5 mM) was added, and the reaction was incubated at room temperature for 10 minutes. Absorbance was measured at 412 nm using a plate reader (Cytation 5, BioTek). The remaining concentration of free thiols was quantified using a glutathione standard curve.[26]

Cell Culture Mino and A549 cells were cultured under standard conditions (37°C, 5% CO_2) in T175 flasks. Both cell lines were revived from frozen stocks stored in liquid nitrogen. Cryovials containing the cells were rapidly thawed in a 37°C water bath and immediately transferred to pre-

warmed complete culture media. To remove residual DMSO, cells were centrifuged at 300 × g for 5 min, resuspended in fresh culture media, and plated in appropriate flasks. Cells were allowed to recover for 24–48 hours before the first media change and were subsequently maintained under routine culture conditions.

Mino cells were maintained in non-treated T175 flasks in Cytiva RPMI-1640 medium (no HEPES, stored at 4ºC) supplemented with 10% fetal bovine serum (HyClone), L-glutamine (100 μg mL^{-1}), 0.05 mM β-mercaptoethanol, and penicillin-streptomycin (100 μg mL^{-1}). Cells were routinely passaged every 2–3 days when they reached a density of ~8×10^5 cells per mL to maintain exponential growth. Passaging was performed by centrifuging the culture at 300 × g for 5 min, discarding the supernatant, and resuspending the cell pellet in fresh pre-warmed medium at a 1:3 dilution.

A549 cells were cultured in tissue culture-treated T175 flasks using Cytiva Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS, GlutaMAX, and Pen/Strep. Cells were passaged at ~80–90% confluency using enzymatic detachment. The culture media was aspirated, and cells were washed once with phosphate-buffered saline (PBS) before treatment with 0.05% Trypsin-EDTA at 37°C for 2–5 min. Detachment was confirmed under a microscope, and trypsinization was quenched by adding complete medium. The cells were then collected, pelleted at 300 × g for 5 min, resuspended in fresh media, and replated at a 1:3 dilution.

Both cell lines were harvested after three passages to ensure robust growth and adaptation to culture conditions. Mino cells were collected from suspension culture, pelleted at 300 × g for 5 min, and washed once with ice-cold PBS. A549 cells were harvested by aspirating the culture medium, washing twice with PBS, and detaching with Trypsin-EDTA as described above. Detached A549 cells were collected in fresh media, pelleted at 300 × g for 5 min, and washed once

with ice-cold PBS. Washed cell pellets were resuspended in ice-cold Dulbecco's phosphate-buffered saline (DPBS) and sonicated on ice (2 × 20 pulses, 20% amplitude, 1 s pulse rate) using a tip sonicator. The lysed suspension was centrifuged at 21,000 × g for 20 min at 4°C, and the supernatant was transferred to a lobind tube. Protein concentrations were determined using the bicinchoninic acid (BCA) assay (Thermo Scientific), and the resulting lysate was used fresh for subsequent experiments.

Gel-Based Probe Reactivity Profiling 50 µL of freshly prepared cell lysates (2 mg mL^{-1} total protein) were treated as indicated in PBS buffer (pH 7.4) at 37°C. Lysates were dosed with a 100X probe in DMSO or an equal volume of DMSO as a control. Following probe treatment, CuAAC click chemistry was performed by sequentially adding 2 µL TAMRA-azide (5 mM, stock in DMSO), 2 µL tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (50 mM, freshly prepared in water), 6 µL tris[(1-benzyl-1-H-1,2,3-triazol-4-yl)methyl]amine (TBTA) (1.7 mM, 3:1 tert-butanol:DMSO), and 2 µL CuSO4•5H2O (50 mM in water) with vortexing between additions. The reaction was incubated at room temperature for 1 h with gentle mixing. Reactions were quenched by adding 250 µL of ice-cold methanol, followed by incubation at -20°C for at least 2 hours. Precipitated proteins were pelleted by centrifugation (10,000 × g, 10 min, 4°C), and the supernatant was discarded. Tubes were inverted and allowed to dry for 30 minutes to ensure complete evaporation of methanol. Protein pellets were resuspended in 75 µL of 4% SDS in PBS (pH 7.5) and bath sonicated (35 kHz, 10 min) to achieve complete solubilization. Following resuspension, 25 µL of 4× Laemmli loading buffer (Bio-Rad, containing 10% β-mercaptoethanol) was added, and samples were vortexed gently before boiling at 95°C for 5 minutes. After cooling to room temperature, 20 µg of protein per lane was loaded onto a 4–15% Tris-Glycine SDS-PAGE gel (Bio-Rad, Criterion TGX) for separation. Fluorescently labeled proteins were visualized using a Bio-

Rad ChemiDoc fluorescence scanner, and total protein staining with Coomassie was performed for normalization.

RAP Workflow 250µL freshly prepared cell lysate at 2 µg mL^{-1} protein concentration was treated in lobind tubes with either 100µM probe or 1.0mM probe for 1 hr at 37 ºC in pairs. Following probe treatment, CuAAC click chemistry was performed by sequentially adding 25 µL isoDTB tags (light tags to the 100µM treated, heavy tags to the 1.0mM treated, (Vector Labs, CCT-1565-L and CCT-1565-H, 10 mM, stock in DMSO), 25 µL tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (50 mM, freshly prepared in water), 75 µL tris[(1-benzyl-1-H-1,2,3-triazol-4-yl)methyl]amine (TBTA) (1.7 mM, 3:1 tert-butanol:DMSO), and 25 µL CuSO4•5H2O (50 mM in water) with mixing between additions. The reaction was incubated at room temperature for 1 hour with gentle mixing. Protein samples were solubilized by adding 200 µL of 10% SDS in PBS, gently mixed and individually sonicated (35 kHz, 2 min) to dissolve any precipitate formed during the click chemistry reaction. Equal amounts of samples were combined in a 2 mL lobind Eppendorf tube and treated with 1 µL of benzonase, followed by incubation at 37°C for 30 minutes at 800 rpm.

For protein precipitation and cleanup, SP3 bead purification was employed. 200 µL Sera-Mag SpeedBeads (GE45152105050250, GE65152105050250), were combined in a 1:1 ratio, washed three times with 1mL water, and resuspended in 250 µL of water before being added to the sample. After incubation at room temperature with shaking at 1000 rpm for 5 minutes, 1 mL of absolute ethanol was added, and the sample was incubated under the same conditions for another 5 minutes. The beads were collected on a magnetic rack, and the supernatant was removed. Beads were washed three times with 1 mL of freshly prepared 80% ethanol.

To prepare proteins for digestion, the beads were resuspended in 500 µL of fresh 2 M urea in 0.5% SDS in PBS. Proteins were reduced with 10 mM DTT at 65°C for 15 minutes and then alkylated with 20 mM iodoacetamide at 37°C for 30 minutes at 500 rpm. The proteins were re-precipitated with 1 mL ethanol and subjected to the same washing steps as before.

Proteins were resuspended in 500 µL of fresh 2 M urea in PBS, and trypsin was added at a 1:100 enzyme-to-protein (w/w) ratio for overnight digestion at 37°C with shaking at 500 rpm. Following digestion, peptides were transferred to a 15mL falcon tube containing 10mL acetonitrile to achieve >95% final volume ACN. Samples were incubated at room temperature with shaking at 1000 rpm for 10 minutes before peptides were collected by centrifugation at 1600 rcf for 5 minutes. The pellet was washed three times with 1 mL ACN, then transferred to a 1.5 mL LoBind Eppendorf tube. The samples were placed on a magnetic rack and washed once more in 1mL fresh ACN. After magnetic separation, ACN was removed and peptides were eluted twice using 100 µL of 2% DMSO in water, incubated at 37°C for 30 minutes at 1000 rpm.

For enrichment, 250 µL of high-capacity NeutrAvidin resin (Thermo Fisher, 29204) was washed three times with 10 mL of IAP buffer (50 mM MOPS pH 7.4, 10 mM sodium phosphate, and 50 mM NaCl), then resuspended in a buffer volume adjusted to achieve a final peptide concentration of 0.2 µg/µL. Peptides were incubated with the resin for 2 hours at room temperature on a rotator, then washed twice with PBS and twice with water, with centrifugation at 1600 rcf between each wash. The resin was transferred to a bio-spin column (Bio-Rad, 7326204), washed twice with 500 µL of water, and centrifuged at 800 rcf for 30 seconds.

Enriched peptides were eluted in 60 µL of 80% acetonitrile in water containing 0.1% formic acid, incubated for 10 minutes at room temperature, then centrifuged at 1600 rcf for 1 minute. This elution was repeated at 72°C for 10 minutes at 300 rpm, and a final elution was

performed at room temperature, followed by centrifugation at 2600 rcf for 2 minutes. The combined eluates were dried using a SpeedVac.

Crude peptides were then resuspended in 22 µL ACN and sonicated (35 kHz, 10 min). 425 µL water and 2.3 µL trifluoroacetic acid were then added followed by another sonication. C18 resin (Thermo Fisher, 89870) was activated twice with 200 µL of 50% methanol in water, followed by three washes with 5% ACN in water containing 0.1% TFA. Resin was then placed into 2mL lobind tubes for sample loading. Samples were then added onto the resin over three 150 µL additions. After complete loading, 200 µL of flowthrough was re-added to the resin twice to ensure complete binding. Samples were then washed three times as mentioned before. All centrifugation steps were performed at 300 rcf for 15 seconds. Samples were then eluted twice with 20 µL 70% ACN in water containing 0.1% formic acid after a 1 minute incubation (1600 rcf, 1 min). A third elution was performed similarly, but at 2600 rcf for 2 minutes. Eluted, desalted peptides were then dried using a SpeedVac. Samples were resuspended in 20 µL LC-LOAD (PreOmics) followed by vortexing and sonication (35 kHz, 10 min). Samples were centrifuged (21000 rcf, 10 min) and 15 µL were transferred into MS vials (Thermo Fisher, 200 046) for analysis by mass spectrometry.

Mass Spectrometry Data Acquisition A nanoElute2 was attached in line to a timsTOF Pro2 equipped with a CaptiveSpray Source (Bruker, Hamburg, Germany). Chromatography was conducted at 40°C through a 25 cm reversed-phase C18 column (PepSep, 1893476) at a constant flow rate of 0.5 µL min^{-1}. Mobile phase A was 98/2/0.1% water/MeCN/formic acid (v/v/v, Thermo Fisher, LS118, LS120) and phase B was MeCN with 0.1% formic acid (v/v). During a 120 min method, peptides were separated by a 3-step linear gradient (4% to 40% B over 95 min, 40% to 60% B over 10 min, 60% to 90% B over 10 min) followed by a 10 min isocratic flush at 95% before washing and a return to low organic conditions. Experiments were run as data-dependent

acquisitions with ion mobility activated in PASEF mode. MS and MS/MS spectra were collected with m/z 100 to 1700 and ions with z = +1 were excluded.

FragPipe Open Search To comprehensively survey mass shifts present in peptides across the dataset, an Open Search was performed using MSFragger within FragPipe v22.0. The search was configured with the following parameters: precursor mass range −10 to +850 Da, fragment mass tolerance 40 amu, isotope error set to 0, and enzyme set to "trypsin" with cleavage after "KR" and not before "P" (C-terminal specificity). Enzymatic digestion was assumed, allowing up to 2 missed cleavages. Minimum and maximum peptide lengths were set to 7 and 50 residues, respectively, and peptide mass range from 500 to 5000 Da. N-terminal methionine clipping was enabled, and a minimum of 15 peaks and a 0.01 intensity ratio were required per spectrum. Fragment ions considered included *b* and *y* series, with a maximum fragment charge of 2+. Calibration and parameter optimization were disabled, and deisotoping and de-neutralloss filtering were enabled. No fixed modifications were applied, and variable modifications were included but not enforced in the initial search (used only for localization and PTM profiling downstream).PeptideProphet was enabled with the following command-line options: --nonparam --expectscore --decoyprobs --masswidth 1000.0 --clevel -2, and results from multiple replicates were combined. PTMProphet was disabled. ProteinProphet was run with the setting --maxppmdiff 2000000. PTM-Shepherd was enabled for post-processing and mass shift profiling. Peak picking parameters were set to a smoothing factor of 2 bins, prominence ratio of 0.3, and peak picking width of 0.002 Da. Precursor mass tolerance was 0.01 Da. Fragment ion types considered for modification annotation were *b* and *y*, with a maximum fragment charge of 2+. PTM-Shepherd was run with default UniMod annotations disabled to allow unsupervised delta mass discovery, and glycan annotation was enabled but no glycan database was used. Glycan delta masses were removed before peak picking.

Normalization of mass shifts was performed across PSMs (not scans). Report generation was enabled with the following options: --sequential --mapmods --prot 0.01. MS1-based quantification, TMT-Integrator, and Spectral Library generation were disabled. For downstream analysis, the global.modsummary.tsv file from PTM-Shepherd was used to extract the number of PSMs associated with each observed mass shift. These counts were plotted against the theoretical mass shift to visualize the landscape of modifications in the range of +200 to +750 Da.

FragPipe Closed Search Mass spectrometry data were analyzed using FragPipe v22.0. Closed Search was performed using MSFragger with the following parameters: precursor mass tolerance −50 to +50 ppm, fragment mass tolerance 50 amu, isotope error set to 0/1/2, enzyme name "stricttrypsin", cleavage at "KR" but not before "P" (C-terminal cleavage), enzymatic cleavage with up to 2 missed cleavages, peptide length between 7 and 50 residues, and peptide mass range from 500 to 5000 Da. Fragment ion types included *b* and *y* ions, with a maximum fragment charge of 2+. N-terminal methionine clipping was enabled. The number of variable modifications per peptide was limited to 4, and the number of allowed variable modification combinations was capped at 5000. Spectra were required to have a minimum of 15 peaks and a minimum peak intensity ratio of 0.01. Mass calibration was set to "Mass Calibration and Parameter Optimization," and deisotoping and de-neutralloss processing were enabled.

Variable modifications included oxidation of methionine (+15.9949 Da), N-terminal acetylation (+42.0106 Da), and a pair of arginine labels (+695.2942 Da and +701.3014 Da) each with a maximum of 1 occurrence. Cysteine was either unmodified or labeled with a variable modification of +57.02146 Da (up to 3 occurrences). Fixed modifications were not applied except in specific runs evaluating cysteine, where carbamidomethylation (+57.02146 Da) was fixed on C.

Crystal-C, PTMProphet, and TMT-Integrator were disabled. PeptideProphet was also disabled. ProteinProphet was enabled and run with the following setting: --maxppmdiff 2000000. Percolator was used for peptide-spectrum match (PSM) validation with the following command-line options: --only-psms --no-terminate --post-processing-tdc, and a minimum probability threshold of 0.5.

MS1-based quantification was enabled using IonQuant with the following parameters: m/z tolerance of 10 ppm, retention time (RT) tolerance of 0.4 min, ion mobility tolerance 0.05 (although ion mobility was not used), labeling mode enabled with two arginine labels (+695.2942 Da and +701.3014 Da), re-quantification enabled, minimum frequency of 0.5, a minimum of 2 isotopes and 1 scan per feature, top 3 ions used for quantification, and no normalization applied. Match-between-runs was disabled. Label-free quantification was not used.

MSBooster, PTMShepherd, SAINTexpress, Skyline, and Spectral Library generation were disabled. Report generation was enabled with the following options: --sequential --razor --mapmods --prot 0.01, with decoys and known contaminants removed. Summary tables were generated at both peptide and protein levels.

SASA Analysis Pipeline Output modified peptides from Fragpipe were processed using a custom Python-based pipeline to identify modified residues based on their sequence position in UniProt. Protein identifiers were initially provided as UniProt names, which were converted to their corresponding UniProt accession numbers using the UniProt API (https://rest.uniprot.org/). Structural models were then downloaded from the European Bioinformatics Institute (EBI) AlphaFold Database (https://alphafold.ebi.ac.uk/). Each protein structure was parsed using Bio.PDB (Biopython) and the Shrake-Rupley algorithm[34] was applied to compute solvent-accessible surface area (SASA) at the atomic level. Residue positions were adjusted by shifting input values down by one to ensure alignment with AlphaFold numbering. To ensure the reliability

of structural predictions, residues with a predicted local distance difference test (pLDDT) score below 70 were excluded from the analysis. The relative solvent accessibility (RSA) was then calculated by normalizing SASA values against a reference maximum exposure of 274 $Å^2$. Computational storage was optimized by removing downloaded PDB files immediately following analysis, and results were iteratively written to an output file. This automated workflow enabled high-throughput extraction of solvent accessibility data from AlphaFold models, facilitating the structural interpretation of specific residues within predicted protein conformations.

For experimental structures, PDB accessions were retrieved from UniProt cross-references and corresponding mmCIF files were downloaded using Biopython's MMCIFParser. Structures with ribosomal content or more than 10 chains were excluded. Each PDB structure was aligned to the UniProt sequence to correctly map the modified residue to the structural coordinates. SASA values were computed using the same Shrake-Rupley algorithm, and RSA values were calculated using the same normalization scheme. In cases where multiple structures existed for a given protein, all qualifying entries were processed independently, and the mean and standard deviation of RSA values for each residue were calculated and used in downstream analysis.

PDB Proximal Ligand Search Pipeline To assess the structural context of modified arginine residues in proteins, an in-house computational pipeline was developed to map modified arginines from the Fragpipe output onto full-length UniProt sequences, align them to resolved PDB structures, and evaluate their spatial proximity to bound ligands. The input dataset consisted of a CSV file containing UniProt accession numbers and peptide sequences with modified arginine residues denoted as R[701.3014]. Other modifications annotated in brackets, such as oxidation or carbamidomethylation, were ignored. The modified residue's position within the peptide was extracted, and all bracketed modifications were removed to generate a cleaned peptide sequence

for further processing. To determine the position of the modified arginine in the full-length protein sequence, the cleaned peptide was aligned to the UniProt sequence using Biopython's PairwiseAligner with the BLOSUM62 substitution matrix, and affine gap penalties (-10 for opening, -0.5 for extension). The modified residue's position in the peptide was mapped to its corresponding position in the aligned UniProt sequence, yielding the residue number in the full-length protein. The script then queried the UniProt cross-references to retrieve all associated PDB accession numbers for each protein. Structures were downloaded in the PDBx/mmCIF format using MMCIFParser from Biopython. To improve computational efficiency and ensure relevance to the analysis, structures containing ribosomal components were excluded, and entries with more than 10 chains were filtered out. For structural mapping, the UniProt sequence was aligned to the sequences of all polypeptide chains in each PDB structure, allowing for potential discrepancies in residue numbering between sequence databases and structural models. The modified residue's position in the PDB structure was identified based on this alignment, ensuring that subsequent analyses were performed on the correct structural residue. To assess the molecular environment of modified arginines, the spatial relationship between the CZ atom of the modified residue and bound ligands was analyzed. NeighborSearch from Biopython[54] was used to determine whether any ligand atoms were within 6.0 Å of the arginine CZ atom. Water molecules (residue name 'HOH') were excluded to focus the analysis on small molecules and cofactors. If a ligand met the distance threshold, its residue name and chain position were recorded. To ensure robustness in handling large datasets, results were written incrementally to a CSV file containing the UniProt ID, PDB ID, chain identifier, mapped residue position in the PDB, and the nearest ligand within the specified distance.

AlphaMissense Pathogenicity Annotation To evaluate the potential functional relevance of modified arginines, AlphaMissense pathogenicity scores were incorporated into the dataset. UniProt accession numbers and residue positions for all modified arginines were used to generate a queryable index. A custom Python script was developed to query the AlphaMissense API () for each protein and residue pair using concurrent requests with up to 10 worker threads to accelerate the process. Isoform-specific identifiers (e.g., P12345-2) were excluded to ensure consistency with canonical AlphaMissense annotations. For each query, the API returned a predicted mean pathogenicity score and a classification (e.g., 'likely pathogenic', 'likely benign'). If a classification was not explicitly returned, a custom threshold-based scheme was applied: scores ≤ 0.34 were labeled 'likely benign', scores ≥ 0.564 were labeled 'likely pathogenic', and intermediate values were designated 'ambiguous'. All results were written to a CSV file including the protein ID, residue position, score, and classification.

Secondary Structure Assignment via DSSP To assess local structural context of modified residues, DSSP-based secondary structure annotations were computed for both AlphaFold-predicted and experimentally resolved PDB structures. For AlphaFold models, UniProt IDs were first mapped to primary accessions using the UniProt API. Structural models were downloaded from the AlphaFold Protein Structure Database. For each protein, the corresponding PDB file was parsed using Biopython's PDBParser, and DSSP secondary structure assignments were computed using Bio.PDB.DSSP. Modified residue positions were matched to AlphaFold residue indices (adjusted for 0-based numbering), and residues with a predicted local distance difference test (pLDDT) score below 75 were excluded. For residues passing the pLDDT filter, a ±4 residue sliding window was used to collect local secondary structure context. Both the DSSP code for the modified residue and its surrounding window were recorded.

For experimental structures, UniProt cross-references were used to identify relevant PDB accessions. Structures were downloaded in mmCIF format and parsed using Biopython's MMCIFParser. Resolution and structural complexity filters were applied: only structures with ≤10 chains and ≤3.0 Å resolution were retained. The best-aligned chain was selected by global pairwise sequence alignment using Biopython's PairwiseAligner, and modified residues were mapped from UniProt positions to structure-specific indices. DSSP was run on each selected PDB model to obtain secondary structure codes. For each modified residue, the DSSP code and ±3 residue sequence context were recorded. Results were aggregated across all qualifying structures for each protein, and the most frequently observed DSSP code (majority vote) was assigned as the representative secondary structure for that residue. Confidence scores were calculated as the frequency of the major assignment relative to total observations.

Motif Enrichment Analysis via pLogo To investigate sequence preferences surrounding modified arginines, motif enrichment analysis was performed using pLogo (https://plogo.uconn.edu). Foreground sequences were generated from a list of experimentally identified arginine modification sites, with UniProt accession numbers and residue positions extracted from FragPipe output. A custom Python script queried the UniProt API to retrieve the full protein sequences and extracted a 21-residue window (±10 amino acids) centered on each modified arginine. If a full window could not be extracted due to proximity to sequence termini, the missing positions were padded with 'X' characters to maintain fixed length. Background sequences were generated separately from the full UniProt human proteome FASTA (including isoforms). For each protein sequence, a sliding window approach was used to extract 21-residue sequences centered on every amino acid position, again padding with 'X' as needed. This resulted in an unfiltered and comprehensive background set. Due to file size limitations on the pLogo web

interface, the background set was downsampled to 1,000,000 sequences using Python's random.sample() method. Foreground and background sets were submitted to the pLogo web interface, specifying arginine (R) as the central anchor residue. A binomial statistical model with Bonferroni correction was used to evaluate motif significance. Enrichment scores and residue logos were downloaded and replotted using custom scripts to standardize styling across figures.

Salt Bridge Identification in AlphaFold Structures To identify potential salt bridge interactions between modified arginines and acidic residues, AlphaFold-predicted structural models were analyzed using a custom Python pipeline. UniProt IDs and modified residue positions were extracted from a CSV input file. Isoform identifiers were excluded. For each canonical UniProt ID, the AlphaFold API was queried to retrieve the corresponding structure metadata and download the associated PDB file. Models were stored locally and parsed using Biopython's PDBParser. Each arginine residue was scanned for proximity to nearby aspartate (ASP) or glutamate (GLU) side chain atoms (OD1/OD2 or OE1/OE2, respectively). Only models with pLDDT scores ≥ 70 for both arginine and acidic residues were considered. Salt bridges were defined by the presence of at least one NE, NH1, or NH2 atom from the modified arginine within 4.0 Å of the target carboxylate atom. When multiple atoms satisfied the distance criterion, the closest atom pair was selected. Electrostatic interaction energies were estimated using a simplified Coulombic potential: assuming unit charges, a dielectric constant of 20.0, and interatomic distance in Å. For each interaction, the chain ID, residue names and numbers, interacting atoms, interatomic distance, pLDDT scores, and computed Coulombic energy were recorded. Following analysis, downloaded PDB files were removed to conserve disk space. Final results were compiled into a CSV file for downstream analysis and visualization.

Cation–π Interaction Identification in AlphaFold Structures To identify potential cation–π interactions involving modified arginines, AlphaFold-predicted structures were evaluated using a geometry-based Python workflow. For each UniProt ID and arginine position, the corresponding AlphaFold model was downloaded from the EBI database. PDB files were parsed with Biopython's PDBParser, and residues were scanned for interactions between arginine CZ atoms and aromatic residues (PHE, TYR, TRP, HIS). For each aromatic residue, centroid coordinates were computed from annotated ring atoms (e.g., CZ, CE1, CE2, etc.), and the Euclidean distance to the arginine CZ atom was calculated. Interactions within 6.0 Å were considered candidates. If angular filtering was enabled, the orientation between the CZ–centroid vector and the ring plane normal was used to classify geometry as parallel (face-to-face) or T-shaped (edge-to-face), with thresholds of ≤30° and 90°±20°, respectively. Models with pLDDT scores below 70 for either the arginine or aromatic residue were excluded. For each identified interaction, the protein ID, chain IDs, residue positions and types, distance, angle, geometry classification, and pLDDT scores were recorded. Only one interaction was logged per arginine to avoid multiple counting. Results were written to two CSV outputs: a detailed interaction table and a simplified summary indicating whether each modified arginine participates in a cation–π interaction.

Salt Bridge and Cation–π Interaction Analysis in Experimental PDB Structures To validate and expand upon AlphaFold-based interaction predictions, salt bridges and cation–π interactions were also identified using experimentally resolved PDB structures. For each UniProt ID and modified residue position, all associated PDB structures were retrieved via UniProt cross-references. Structures were excluded if they represented ribosomal proteins, large macromolecular complexes (taxonomy count > 15), NMR structures, or models containing more than 10 chains. PDBx/mmCIF files were downloaded and parsed using Biopython's MMCIFParser. For each

candidate arginine, neighboring residues were identified using NeighborSearch, and interactions were classified as salt bridges or cation–π based on geometric and electrostatic criteria. Salt bridges were defined as any acidic residue (ASP, GLU) within 4.0 Å of NE, NH1, or NH2 atoms of the arginine. Cation–π interactions were defined as cases where the arginine CZ atom was within 8.0 Å of the centroid of an aromatic side chain (PHE, TYR, TRP, HIS). When enabled, angular filtering was used to categorize interaction geometries as face-to-face (≤30°) or T-shaped (90° ± 20°). All interactions were annotated with residue identity, chain location, distance, geometry (if applicable), whether the partner residue was part of the same polypeptide chain, and optionally, an estimated electrostatic energy. Final results were compiled into a single CSV table containing all identified intermolecular interactions from experimental PDB models.

pKa Estimation for Arginines Using Experimental PDB Structures To estimate side-chain protonation states of modified arginines, PROPKA 3.0 was used to calculate predicted pKa values from experimental structures in the Protein Data Bank (PDB). Modified arginines were identified by UniProt accession and residue position. Isoforms and ribosomal proteins were excluded. Associated PDB structures were retrieved via UniProt cross-references and filtered to exclude NMR structures, entries with >10 polymer entities, or structures with >10 chains. Structures were downloaded in mmCIF format and parsed using Biopython's MMCIFParser. Each structure was aligned to the canonical UniProt sequence using PairwiseAligner to identify the best-matching chain and map UniProt residue indices to structural coordinates. Only arginines with a resolved CZ atom and a B-factor ≥ 30.0 were considered. Structures were converted to PDB format using Biopython's PDBIO to ensure compatibility with PROPKA. The PROPKA executable was run via subprocess call, and predicted pKa values were parsed from the resulting .pka files. For each modified arginine, the output included UniProt ID, PDB ID, chain ID, UniProt and author residue

numbers, and pKa value. Structures and temporary output files were deleted after use to conserve disk space.

pKa Estimation for Arginine, Lysine or Cysteine from AlphaFold Models To complement structure-based pKa estimation, PROPKA was also applied to AlphaFold-predicted structures. AlphaFold models were downloaded in PDB format for each UniProt accession. Only residues with pLDDT confidence scores >70 were considered for analysis. For each AlphaFold structure, PROPKA 3.0 was executed using the PDB file as input. Predicted pKa values were extracted using a custom parser that accommodated multiple possible output formats. Residue-level pKa values were reported for each modified arginine, cysteine or lysine, alongside the corresponding pLDDT score. Proteins for which PROPKA failed to produce valid output were excluded. Temporary AlphaFold PDB and PROPKA output files were deleted after processing.

Intrinsic Disorder Prediction via IUPred2A To evaluate the intrinsic disorder propensity of modified arginine residues, disorder scores were retrieved using the IUPred2A web API (https://iupred2a.elte.hu). A CSV file containing UniProt IDs and modified residue positions served as input. For each protein, the full-length disorder profile was fetched in JSON format and parsed to extract per-residue IUPred2 scores. Modified arginine positions were indexed using 1-based UniProt residue numbering. Prior to extraction, each index was converted to 0-based to match Python list conventions. If a position fell within the length of the retrieved disorder profile, the corresponding IUPred2 score was recorded. Otherwise, 'NA' was assigned. Each entry in the output file included the UniProt ID, residue position, and corresponding disorder score. Results were written iteratively to a CSV file for downstream integration with structural and reactivity data.

Molecular Docking and Covalent Adduct Modeling Molecular docking was performed using the Molecular Operating Environment (MOE2024)[55]. The crystal structures of aconitase 2 (PDB: 1B0J) and cyclophilin A (PDB: 6GJI) were obtained, with all crystallographic ions removed. Each structure was solvated in a 10 Å water sphere to mimic physiological conditions before preparation using MOE's QuickPrep feature, which optimized hydrogen placement, assigned protonation states, and performed an initial energy minimization. For aconitase, the docking site was defined based on the known binding site of isocitrate, and the dehydrate of ninhydrin was docked using MOE's docking protocol. For cyclophilin A, the ligand 4'CA-Alk was docked using MOE's template-based docking to ensure proper positioning. In both cases, the top-scoring pose was selected and loaded into the MOE Window, where a covalent adduct was modeled. Energy minimization was performed using the Amber10:EHT force field with gradient-based optimization until the root mean square gradient reached 0.05 kcal/(mol·Å). Visualization and figure preparation were performed using ChimeraX 1.6.1[56].

3.3 General Synthesis Procedures

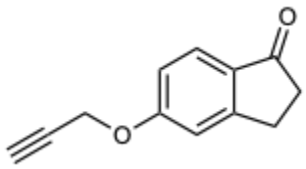
General Experimental Procedures: All reactions were performed in flame- or oven-dried glassware fitted with rubber septa under a positive pressure of nitrogen or argon, unless otherwise noted. All reaction mixtures were stirred throughout the course of each procedure using Teflon-coated magnetic stir bars. Air- and moisture-sensitive liquids were transferred via syringe or stainless steel cannula. Solutions were concentrated by rotary evaporation below 35 °C. Analytical thin-layer chromatography (TLC) was performed using glass plates pre-coated with silica gel (0.25-mm, 60-Å pore size, 230−400 mesh, SILICYCLE INC) impregnated with a fluorescent indicator (254 nm). TLC plates were visualized by exposure to ultraviolet light (UV), and then

were stained by submersion in a basic aqueous solution of potassium permanganate ($KMnO_4$) or with an acidic ethanolic solution of ninhydrin, followed by brief heating.

Materials: Dichloromethane (DCM), dimethylformamide (DMF), tetrahydrofuran (THF), ethyl ether, and acetonitrile to be used in anhydrous reaction mixtures were dried by passage through activated alumina columns immediately prior to use. Hexanes used were ≥85% n-hexane. Other commercial solvents and reagents were used as received, unless otherwise noted.

Instrumentation: Unless otherwise noted, proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on a JEOL JNM-ECZ400R 400 MHz NMR equipped with 5 mm H/F/X Royal Probe or a Bruker AVIII HD 600 MHz NMR equipped with 5 mm CPDCH CryoProbe at 23 °C. Proton chemical shifts are expressed in parts per million (ppm, δ scale) and are referenced to residual protium in the NMR solvent ($CDCl_3$: δ 7.26, CD_3CN: δ 1.94). Carbon chemical shifts are expressed in parts per million (ppm, δ scale) and are referenced to the carbon resonance of the NMR solvent ($CDCl_3$: δ 77.16, CD_3CN: δ 118.26). Data are represented as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, dd = doublet of doublets, dt = doublet of triplets, sxt = sextet, m = multiplet, br = broad, app = apparent), integration, and coupling constant (J) in hertz (Hz). Mass measurements for high-resolution mass spec (HRMS) were performed on a Waters Xevo G2-XS TOF calibrated against sodium formate clusters and using a LeuEnk lockmass. Expected monoisotopic masses were calculated using MassLynx 4.1 and the *m/z* values for calibrant and lockmass were MassLynx-default values.

Compound SI-1



DMF (13.5 mL) was added to a round-bottom flask containing K_2CO_3 (1.87 g, 13.50 mmol, 2.00 equiv) and 5-hydroxy-2,3-dihydro-1H-inden-1-one (1.00 g, 6.75 mmol, 1 equivl). Propargyl bromide (80 wt% in toluene, 829 µL, 1.10 g, 7.42 mmol, 1.10 equiv) was added dropwise over 1 minute. The reaction vessel was heated in a 60 °C oil bath for 16 h. After cooling to ambient temperature, the mixture was diluted with DCM (30 mL) and washed with DI water (3 × 30 mL) followed by brine (2 × 30 mL). The organic layer was dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure to afford **SI-1** as a brown solid, which was used in subsequent steps without further purification (1.22 g, 97% yield).

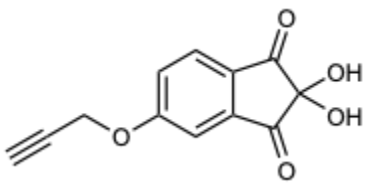
TLC (hexanes:ethyl acetate = 8:2): R_f = 0.25 (UV, $KMnO_4$)

^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.71 (d, *J* = 8.4 Hz, 1H), 7.00 (s, 1H), 6.97 (d, *J* = 8.7 Hz, 1H), 4.77 (t, *J* = 2.1 Hz, 2H), 3.11 (t, *J* = 5.4 Hz, 2H), 2.69 (t, *J* = 5.9 Hz, 2H), 2.57 (s, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ: 205.41, 163.06, 158.06, 131.22, 125.50, 115.88, 111.00, 77.74, 76.35, 56.05, 36.53, 25.99.

HRMS (ESI$^+$): *m/z* calcd for $C_{12}H_{11}O_2^+$ ($[M+H]^+$): 187.0759; found: 187.0756.

5-Nin-Alk



Selenium dioxide (745 mg, 6.71 mmol, 2.5 equiv) was added to a solution of **SI-1** (500 mg, 2.69 mmol, 1 equiv) in 1,4-dioxane (3.25 mL) and DI water (325 µL). The reaction mixture was heated to reflux in a 96 ºC oil bath for 4 hours. After cooling to ambient

temperature, the mixture was diluted with ethyl acetate (15 mL) and washed successively with DI water (2 × 10 mL), 1 M aqueous HCl (2 × 10 mL). The organic layer was washed with brine (1 × 10 mL), dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure. The crude residue was purified by flash chromatography (30–50% acetone in hexanes) to afford a red solid which was still contaminated with what may be some selenium byproducts. Final purification was performed by reverse-phase HPLC (2–25% MeCN with 0.1% TFA in milliQ water with 0.1% TFA), and lyophilization of the collected fractions yielded the product as a white powder (561 mg, 90% yield).

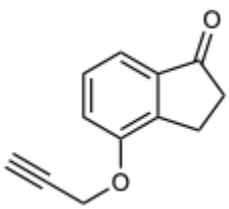
TLC (hexanes:acetone = 1:1): R_f = 0.34 (UV, $KMnO_4$)

^{1}H NMR (400 MHz, 10:1 CD_3CN + D_2O) δ: 7.91 (d, J = 8.6 Hz, 1H), 7.47 (dd, J = 8.6, 2.5 Hz, 1H), 7.40 (d, J = 2.4 Hz, 1H), 4.88 (d, J = 2.4 Hz, 2H), 2.95 (t, J = 2.4 Hz, 1H)

^{13}C NMR (101 MHz, 10:1 CD_3CN + D_2O) δ: 211.86, 196.70, 195.22, 164.89, 141.85, 133.24, 126.78, 126.31, 107.41, 77.81, 57.05, 28.79

HRMS (ESI$^+$): *m/z* calcd for $C_{12}H_9O_5^+$ ($[M+H]^+$): 233.0450; found: 233.0448.

Compound SI-2



To a round-bottom flask containing K_2CO_3 (1.87 g, 13.50 mmol, 2.00 equiv) and 4-hydroxy-2,3-dihydro-1H-inden-1-one (1.00 g, 6.75 mmol, 1 equiv) was added anhydrous DMF (13.5 mL). Propargyl bromide (80 wt% in toluene, 829 µL, 1.10 g, 7.42 mmol, 1.10 equiv) was added dropwise over the course of 1 minute. The vessel was heated in a 60 °C oil bath for 16 h. After cooling to ambient temperature, the mixture was diluted with DCM (30 mL) and washed with DI water (3 × 30 mL) followed by brine (2 × 30 mL). The organic layer was dried over Na_2SO_4, filtered, and the filtrate

was concentrated under reduced pressure to afford **SI-2** as a tan solid, which was used in subsequent steps without further purification (1.18 g, 94% yield).

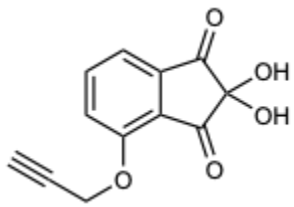
TLC (hexanes:ethyl acetate = 8:2): R_f = 0.40 (UV, $KMnO_4$)

^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.40 (d, *J* = 7.6 Hz, 1H), 7.36 (t, *J* = 7.6 Hz, 1H), 7.17 (d, *J* = 7.7 Hz, 1H), 4.80 (s, 2H), 3.07 (t, *J* = 6.0 Hz, 2H), 2.69 (t, *J* = 6.0 Hz, 2H), 2.55 (s, 1H)

^{13}C NMR (101 MHz, $CDCl_3$) δ: 207.23, 155.10, 144.57, 139.04, 128.86, 116.50, 116.45, 78.18, 76.22, 56.04, 36.26, 22.70.

HRMS (ESI$^+$): *m/z* calcd for $C_{12}H_{11}O_2^+$ ($[M+H]^+$): 187.0759; found: 187.0757

4-Nin-Alk



To a solution of **SI-2** (500 mg, 2.69 mmol, 1 equiv) in 1,4-dioxane (3.25 mL) and DI water (325 µL) was added selenium dioxide (745 mg, 6.71 mmol, 2.50 equiv). The reaction mixture was heated to reflux in a 96 ºC oil bath for 4 hours. After cooling to ambient temperature, the mixture was diluted with ethyl acetate (15 mL) and washed successively with DI water (2 × 10 mL) and 1 M aqueous HCl (2 × 10 mL). The organic layer was washed with brine (1 × 10 mL), dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure. The crude residue was purified by flash chromatography (30–50% acetone in hexanes) to yield a red solid. Final purification was performed by reverse-phase HPLC (2–25% MeCN with 0.1% TFA in milliQ water with 0.1% TFA) to remove selenic acid byproducts yielding a white powder after lyophilization (532 mg, 2.29 mmol, 85% yield).

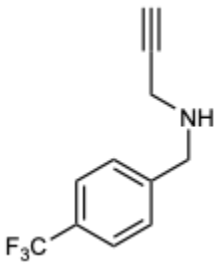
TLC (hexanes:acetone = 1:1): R_f = 0.36 (UV, $KMnO_4$)

^{1}H NMR (400 MHz, 10:1 $CD_3CN + D_2O$) δ: 7.93 (dd, J = 8.3, 7.6 Hz, 1H), 7.58 (d, J = 7.0 Hz, 1H), 7.55 (d, J = 7.9 Hz, 1H), 4.96 (d, J = 2.4 Hz, 2H), 2.97 (t, J = 2.4 Hz, 1H)

^{13}C NMR (101 MHz, 10:1 $CD_3CN + D_2O$) δ: 196.78, 194.41, 156.65, 141.17, 139.49, 127.15, 121.36, 116.92, 87.68, 78.00, 77.95, 57.26

HRMS (ESI$^+$): *m/z* calcd for $C_{12}H_9O_5^+$ ($[M+H]^+$): 233.0450; found: 233.0446

Compound SI-3



A solution of 4-(trifluoromethyl)benzaldehyde (1.74 g, 10.0 mmol, 1 equiv), propargylamine (826 mg, 15.0 mmol, 2.50 equiv), sodium triacetoxyborohydride (5.30 g, 25.0 mmol, 2.50 equiv), and glacial acetic acid (600 mg, 10.0 mmol, 1 equiv) in DCM (95 mL) was stirred at ambient temperature for 2.5 h. The reaction mixture was diluted with DCM (40 mL) and saturated aqueous $NaHCO_3$ (40 mL) was added, and the layers were mixed vigorously and then separated. The aqueous layer was extracted with DCM (2 × 20 mL), and the combined organic layers were concentrated under reduced pressure. The residue was dissolved in ether (50 mL) and extracted with 1.0 M aqueous HCl (3 × 30 mL). The combined aqueous extracts were basified to pH ~11–12 (as judged by pH strips) with 1.0 M aqueous NaOH and extracted with ether (4 × 20 mL). The combined organic layers were dried over Na_2SO_4, filtered, and the filtrate was concentrated. The crude residue was purified by flash chromatography (silica gel, gradient from 10-25% ethyl acetate in hexanes) afforded N-(4-(trifluoromethyl)benzyl)prop-2-yn-1-amine (2.05 g, 96% yield) as a colorless oil.

TLC (hexanes:ethyl acetate = 4:1): R_f = 0.32 (UV, ninhydrin)

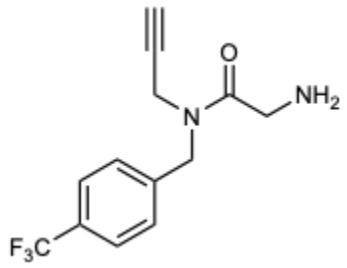
^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.59 (d, J = 7.9 Hz, 2H), 7.48 (d, J = 7.9 Hz, 2H), 3.95 (s, 2H), 3.43 (s, 2H), 2.27 (s, 1H)

^{13}C NMR (101 MHz, $CDCl_3$) δ: 143.58(d, J = 32.57), 129.73, 128.67, 125.42(q, J= 3.8 Hz), 124.23 (q, J = 271 Hz) 81.78, 71.94, 51.68, 37.41

^{19}F NMR (376 MHz, $CDCl_3$) δ: -62.31

HRMS (ESI$^+$): *m/z* calcd for $C_{11}H_{11}F_3N^+$ ([M+H]$^+$): 214.0884; found: 214.0842

Compound SI-4



To a solution of **SI-3** (365 mg, 1.71 mmol, 1 equiv) in DMF (8.56 mL) was added DIPEA (664 mg, 895 µL, 5.14 mmol, 3.00 equiv), followed by (tert-butoxycarbonyl)glycine (330 mg, 1.88 mmol, 1.10 equiv) and PyAOP (976 mg, 2.57 mmol, 1.50 equiv). The mixture was briefly sonicated and stirred at ambient temperature for 1 hour. The reaction was diluted with ethyl acetate (50 mL) and sequentially washed with 1.0 M aqueous HCl (2 × 15 mL), DI water (2 × 15 mL), saturated aqueous $NaHCO_3$ (2 × 15 mL), and brine (2 × 15 mL). The organic layer was dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure to afford the Boc-protected intermediate as a white solid. This material was then dissolved in 4 M aqueous HCl in dioxane and stirred for 3 hours at ambient temperature to yield the HCl salt of the product as a white solid after diluting in toluene and concentrating under reduced pressure. The HCl salt was freebased with 2 mL aqueous saturated $NaHCO_3$ and extracted with 20 mL ethyl acetate which was subsequently dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure to yield pure **SI-4** (392.7 mg, 85% yield) as a white foam which was used in subsequent steps without further purification.

TLC (hexanes:ethyl acetate = 1:1): R_f = 0.35 (UV, ninhydrin)

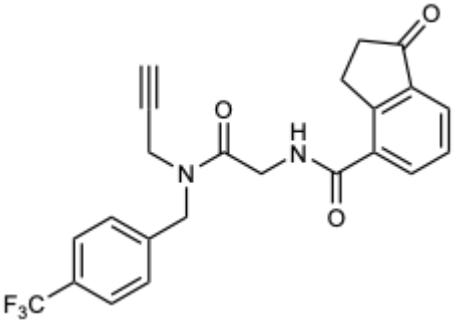
^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.54 (d, *J* = 8.0 Hz, 1H), 7.49 (d, *J* = 8.0 Hz, 1H), 7.42 (d, *J* = 8.0 Hz, 1H), 7.33 (d, *J* = 8.0 Hz, 1H), 4.60 (s, 2H, Ha, major rotamer), 4.62 (s, 2H, Ha′, minor rotamer), 4.29 (s, 2H, Hb, major rotamer), 4.35 (s, 2H, Hb′, minor rotamer), 4.00 (s, 2H, Hc, major rotamer), 3.96 (s, 2H, Hc′, minor rotamer), 2.26 (s, 1H, Hd, major rotamer), 2.14 (s, 1H, Hd′, minor rotamer), 1.91 (b, 3H, He, both rotamers overlapping). Signals marked with Ha/Ha′–Hd/Hd′ correspond to rotameric pairs arising from restricted rotation about the amide bond. Rotamers are observed in a ~55:45 ratio (major:minor).

^{13}C NMR (101 MHz, $CDCl_3$) δ173.25, 140.71, 139.54, 130.31, 129.97, 128.65, 127.07, 126.16(q, J = 3.8 Hz), 125.76 (q, J = 3.8 Hz), 124.14(q, J = 271 Hz),78.23, 77.35, 73.54, 72.82, 48.85, 43.51, 35.78, 35.11

^{19}F NMR (376 MHz, $CDCl_3$) δ: -62.58Hz (d, J = 36.0 Hz, 3F)

HRMS (ESI$^+$): *m/z* calcd for $C_{13}H_{14}F_3N_2O_3^+$ ($[M+H]^+$): 271.1058; found: 271.1050

Compound SI-5



To a solution of **SI-4** (100 mg, 370 µmol, 1 equiv) in DMF was added DIPEA (191 mg, 270 µL, 1.48 mmol, 4.00 equiv), followed by 1-oxo-2,3-dihydro-1H-indene-4-carboxylic acid (72 mg, 407 µmol, 1.10 equiv) and HATU (211 mg, 555 µmol, 1.50 equiv). The reaction was stirred for 14 hours at ambient temperature. The reaction was diluted with ethyl acetate (50 mL) and washed with 1.0 M aqueous HCl (2 × 15 mL), DI water (2 × 15 mL), saturated aqueous $NaHCO_3$ (2 × 15 mL), and brine (2 × 15 mL). The organic layer was dried over Na_2SO_4, filtered, and the filtrate was

concentrated under reduced pressure. Purification by flash chromatography (silica gel, gradient 40–60% ethyl acetate in hexanes) afforded **SI-5** as a white foam (137.7 mg, 87% yield) as a complex mixture of rotamers.

TLC (hexanes:ethyl acetate = 2:3): R_f = 0.31 (UV, $KMnO_4$)

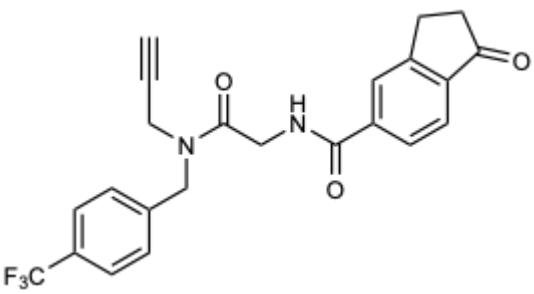
^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.95 (t, J = 8.5 Hz, 1H), 7.88 (d, J = 7.7 Hz, 1H), 7.65 (d, J = 8.0 Hz, 1H), 7.60 (d, J = 7.9 Hz, 1H), 7.46 (t, J = 7.8 Hz, 1H), 7.42 – 7.34 (m, 2H), 7.24 (br, 1H), 4.79 (s, 2H, major rotamer), 4.76 (s, 2H, minor rotamer) 4.47 (d, J = 3.4 Hz, 2H major rotamer), 4.36 (d, J = 4.0 Hz, 1H), 4.29 (s, 1H), 4.02 (s, 1H), 3.53 – 3.42 (m, 2H), 2.75 (q, J = 5.4 Hz, 2H), 2.34 (d, J = 34.4 Hz, 1H)

^{13}C NMR (101 MHz, $CDCl_3$) δ: 206.50, 168.54, 168.34, 166.77, 154.18, 139.99, 138.42, 133.20, 132.53, 132.43, 128.64, 127.86, 127.20, 126.75, 126.34 (q, J = 3.7 Hz), 125.88 (q, J = 3.8 Hz), 77.55, 76.76, 74.32, 73.33, 49.15, 48.99, 42.01, 41.95, 36.26, 36.18, 26.35

^{19}F NMR (376 MHz, $CDCl_3$) δ: -62.50, -62.55

HRMS (ESI$^+$): *m/z* calcd for $C_{23}H_{20}F_3N_2O_3^+$ ($[M+H]^+$): 429.1426; found: 429.1437

Compound SI-6



To a solution of **SI-4** (60 mg, 222 µmol, 1 equiv) in DMF was added DIPEA (57 mg, 77 µL, 444 µmol, 4.00 equiv), followed by 1-oxo-2,3-dihydro-1H-indene-5-carboxylic acid (43 mg, 244 µmol, 1.10 equiv) and HATU (127 mg, 333 µmol, 1.50 equiv). The reaction mixture was stirred at ambient temperature for 14 hours. The reaction mixture was

diluted with ethyl acetate (50 mL) and sequentially washed with 1.0 M aqueous HCl (2 × 15 mL), DI water (2 × 15 mL), saturated aqueous $NaHCO_3$ (2 × 15 mL), and brine (2 × 15 mL). The organic layer was dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure. The crude product was purified by flash column chromatography (silica gel, gradient 40–60% ethyl acetate in hexanes) to afford the product as a white foam (87.6 mg, 92% yield) as a complex mixture of rotamers.

TLC (hexanes:ethyl acetate = 2:3): R_f = 0.29 (UV, $KMnO_4$)

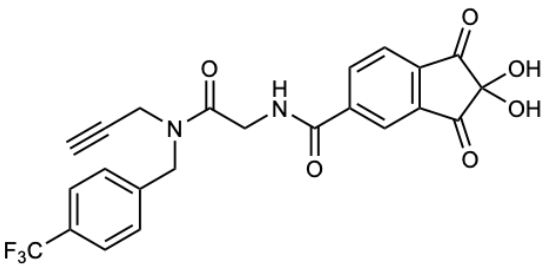
^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.12 (m, 1H), 7.85 (m, 1H), 7.59 (m, 3H), 7.41 (m, 3H), 4.78 (s, 2H, major rotamer), 4.76 (s, 2H, minor rotamer), 4.45 (t, J = 4.3 Hz, 2H, major rotamer), 4.33 (t, J = 4.1 Hz, 2H, minor rotamer), 4.26 (t, J = 2.6 Hz, 2H, minor rotamer), 4.02 (t, J = 3.1 Hz, 1H, major rotamer), 3.22 – 3.13 (m, 2H), 2.77 – 2.67(m, 2H), 2.40 – 2.33 (m, 1H, minor rotamer), 2.31 – 2.24 (m, 1H, major rotamer)

^{13}C NMR (101 MHz, $CDCl_3$) δ: 206.17, 168.61, 168.55, 168.43, 166.63 (m), 166.35, 158.52, 155.37, 140.11, 139.52, 139.39, 139.29, 137.36, 133.78, 133.74, 133.30, 133.18, 128.65, 127.21, 126.30 (q, *J* = 3.7 Hz), 125.83 (q, *J* = 3.2 Hz), 124.05, 122.15, 77.68, 76.89, 74.31, 74.22, 73.29, 73.23, 49.12, 49.03, 42.11, 42.04, 36.60, 36.56, 36.19, 35.13, 26.02, 25.89

^{19}F NMR (376 MHz, $CDCl_3$) δ: -62.48(major rotamer), -62.54 (minor rotamer)

HRMS (ESI$^+$): *m/z* calcd for $C_{23}H_{20}F_3N_2O_3^+$ ($[M+H]^+$): 429.1426; found: 429.1438

5-CA-Alk



Selenium dioxide (79 mg, 716 µmol, 2.1 equiv) was added to a solution of **SI-6** (146 mg, 341 µmol, 1 equiv) in 1,4-dioxane (1.5 mL) and DI water (150 µL). The reaction mixture was heated to reflux in a 96 ºC oil bath for 4 hours. After cooling to ambient temperature, the mixture was diluted with ethyl acetate (15 mL) and washed successively with DI water (2 × 10 mL), 1 M aqueous HCl (2 × 10 mL). The organic layer was washed with brine (1 × 10 mL), dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure. The crude mixture was purified by reverse-phase HPLC (25-50% MeCN with 0.1% TFA in milliQ water with 0.1% TFA), and lyophilization of the collected fractions yielded the product as a white powder (18 mg, 11% yield) as a complex mixture of rotamers.

TLC (hexanes:acetone = 1:2): R_f = 0.39 (UV, $KMnO_4$)

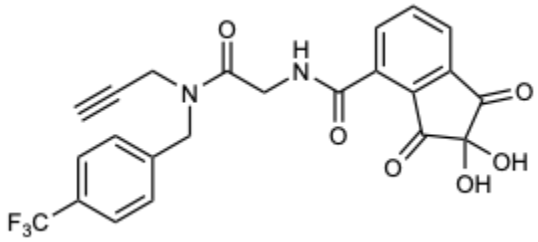
^{1}H NMR (600 MHz, 10:1 CD_3CN + D_2O) δ 7.98 (2H, m), 7.67 (3H, m), 7.49 (3H, m), 4.75 (2H, m), 4.29 (2H, m), 4.17 (2H, m), 2.60 (1H, s)

^{13}C NMR (151 MHz, 10:1 CD_3CN + D_2O) δ: 169.78, 169.64, 169.19, 168.83, 166.68, 153.87, 142.62, 141.88, 136.91, 136.70, 133.61, 131.39, 130.95, 130.53, 129.58, 129.21, 128.41, 126.43, 126.13, 124.30, 123.56, 79.43, 78.86, 74.53, 73.36, 50.20, 49.94, 42.32, 42.16, 37.46, 35.89

^{19}F NMR (376 MHz, CD_3CN + 10% D_2O) δ: -63.68(major rotamer), -63.71 (minor rotamer), -76.55 (reference TFA)

HRMS (ESI$^+$): *m/z* calcd for $C_{23}H_{18}F_3N_2O_6^+$ ($[M+H]^+$): 475.1117; found: 475.1113.

4-CA-Alk



Selenium dioxide (140 mg, 1.3 mmol, 2.2 equiv) was added to a solution of **SI-5** (245 mg, 590 µmol, 1 equiv) in 1,4-dioxane (2.6 mL) and DI water (260 µL). The reaction mixture was heated to reflux in an oil bath at 96 ºC for 4 hours. After cooling to ambient temperature, the mixture was diluted with ethyl acetate (15 mL) and washed successively with DI water (2 × 10 mL), 1 M aqueous HCl (2 × 10 mL). The organic layer was washed with brine (1 × 10 mL), dried over Na_2SO_4, filtered, and the filtrate was concentrated under reduced pressure. The crude mixture was purified by reverse-phase HPLC (25-50% MeCN with 0.1% TFA in milliQ water with 0.1% TFA), and lyophilization of the collected fractions yielded the product as a white powder (26 mg, 10 % yield) as a complex mixture of rotamers.

TLC (hexanes:acetone = 1:2): R_f = 0.42 (UV, $KMnO_4$)

^{1}H NMR (600 MHz, 10:1 CD_3CN + D_2O) δ: δ 8.05 (2H, m), 7.67 (4H, m), 7.51 (2H, m), 4.78 (2H, m), 4.40 (2H, m), 4.19 (2H, m), 2.68 (1H, m), 2.55 (2H, s)

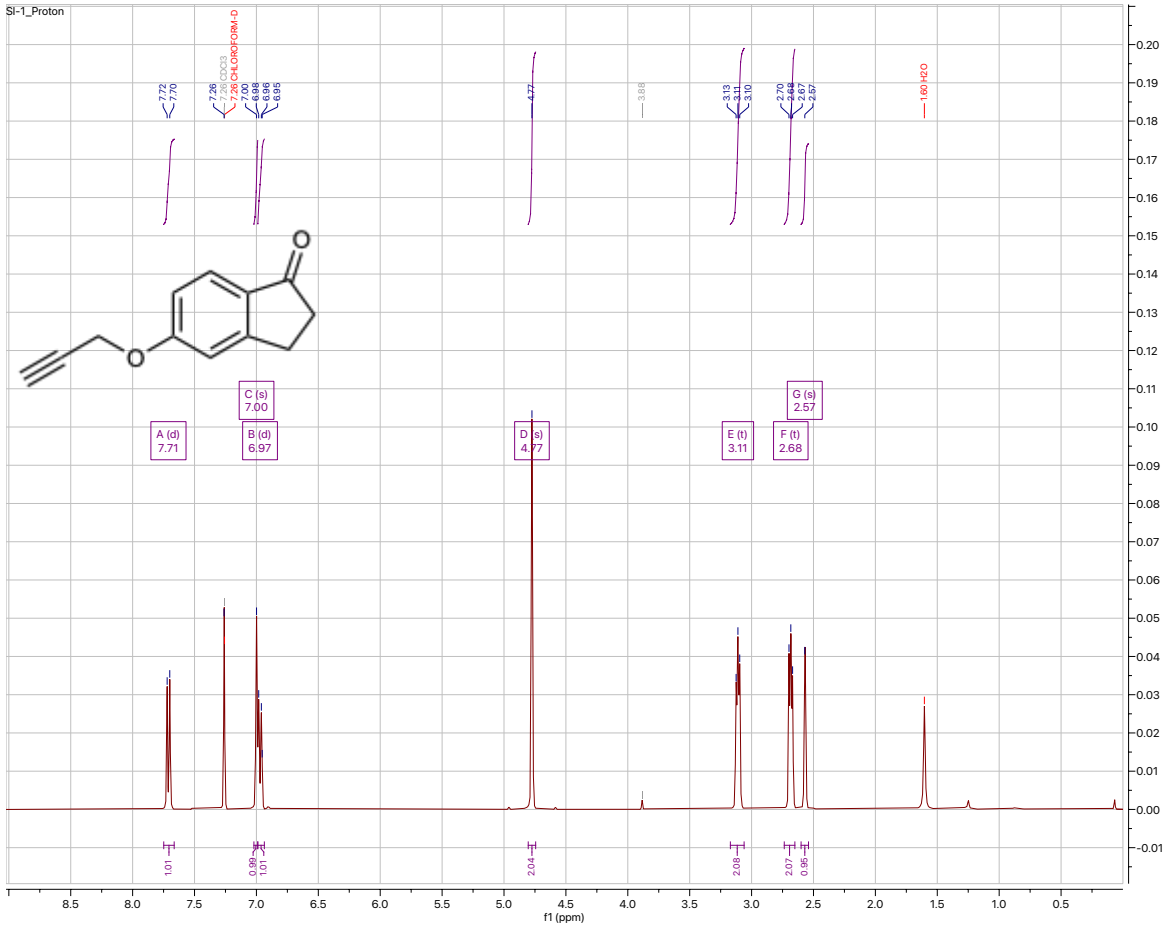
^{13}C NMR (151 MHz, 10:1 CD_3CN + D_2O_3) δ: δ 169.38, 169.27, 166.87, 165.13, 162.89, 149.24, 146.14, 140.70, 138.49, 137.80, 135.73, 135.59, 134.43, 131.75, 130.21, 129.79, 129.72, 129.50, 128.98, 128.21, 126.85, 126.22, 125.92, 124.10, 79.12, 74.43, 73.23, 50.12, 49.89, 49.74, 42.65, 42.52, 40.34, 37.31, 35.76

^{19}F NMR (376 MHz, CD_3CN + 10% D_2O) δ: -63.71(major rotamer), -63.74 (minor rotamer), -76.55 (reference TFA)

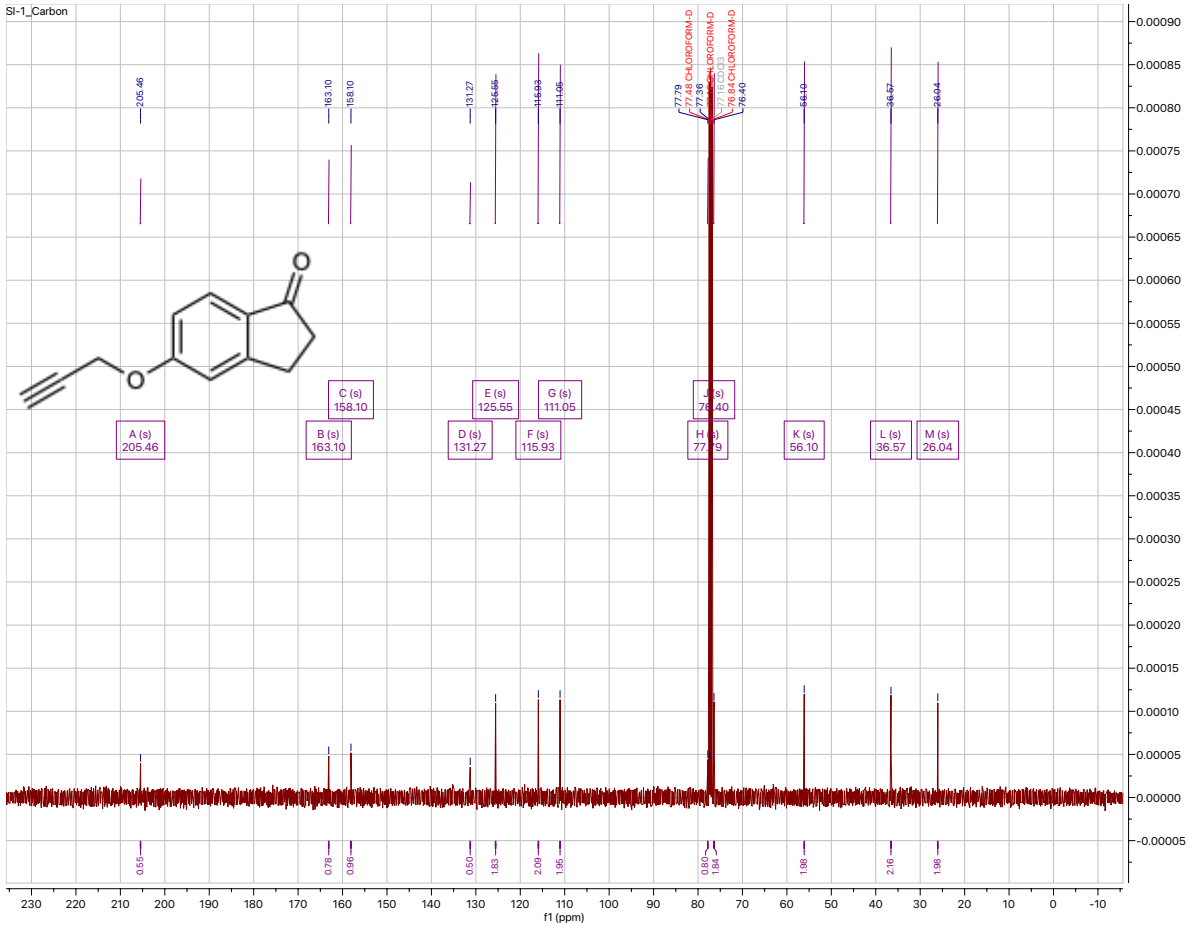
HRMS (ESI$^+$): *m/z* calcd for $C_{23}H_{18}F_3N_2O_6^+$ ($[M+H]^+$): 475.1117; found: 475.1122

3.4 NMR Spectra for Compounds

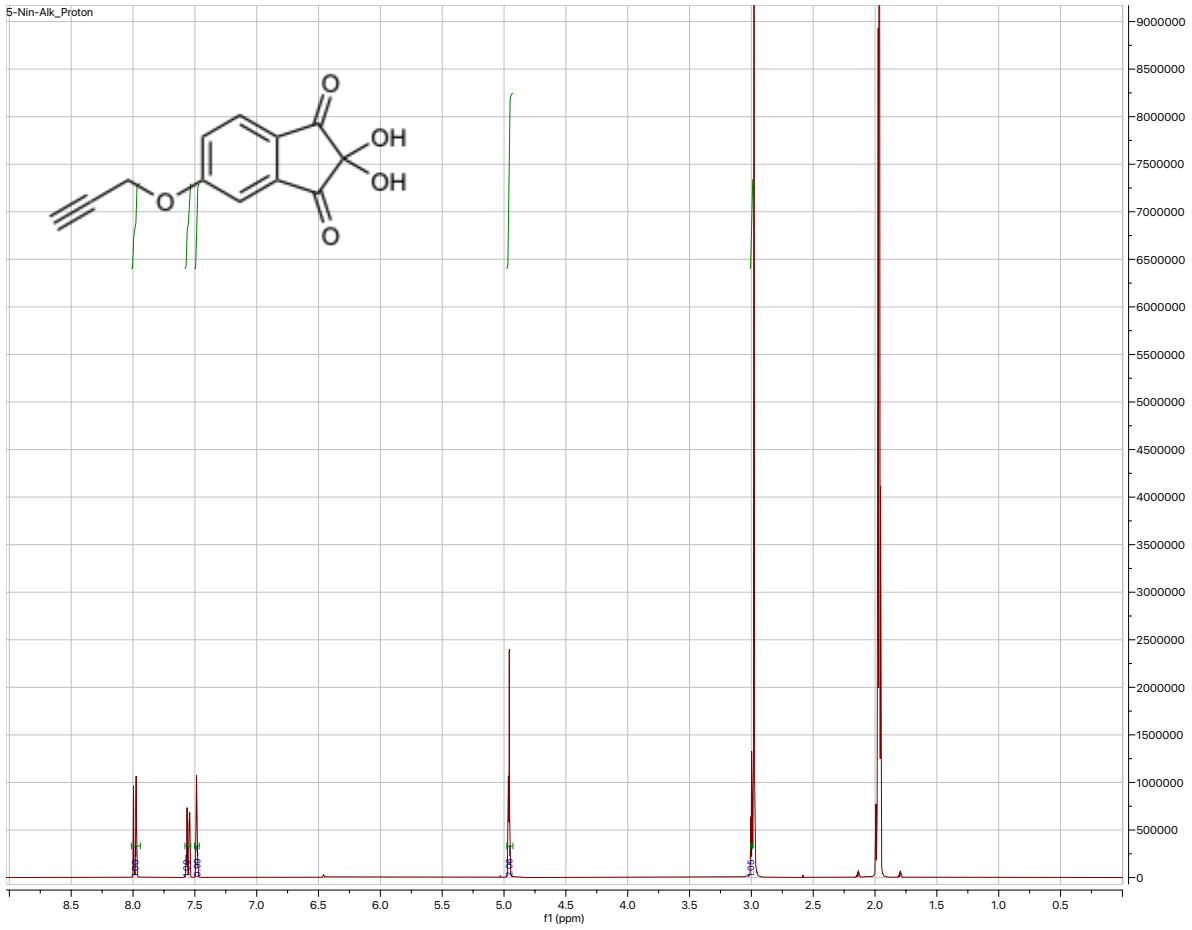
SI-1 1H



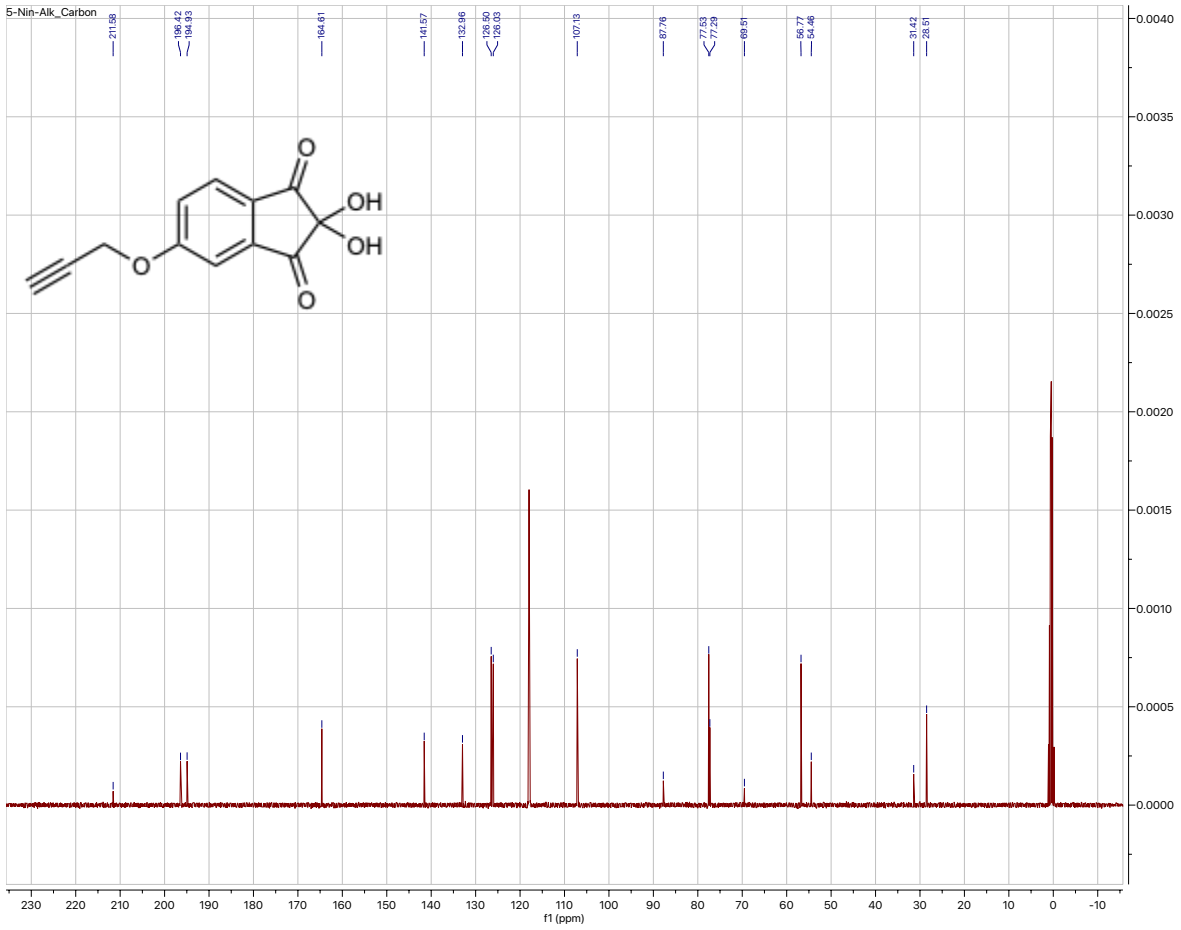
SI-1 13C


SI-1_Carbon
A (s) 205.46
B (s) 163.10
C (s) 158.10
D (s) 131.27
E (s) 125.55
F (s) 115.93
G (s) 111.05
H (s) 77.79
J (s) 76.40
K (s) 56.10
L (s) 36.57
M (s) 26.04
f1 (ppm)

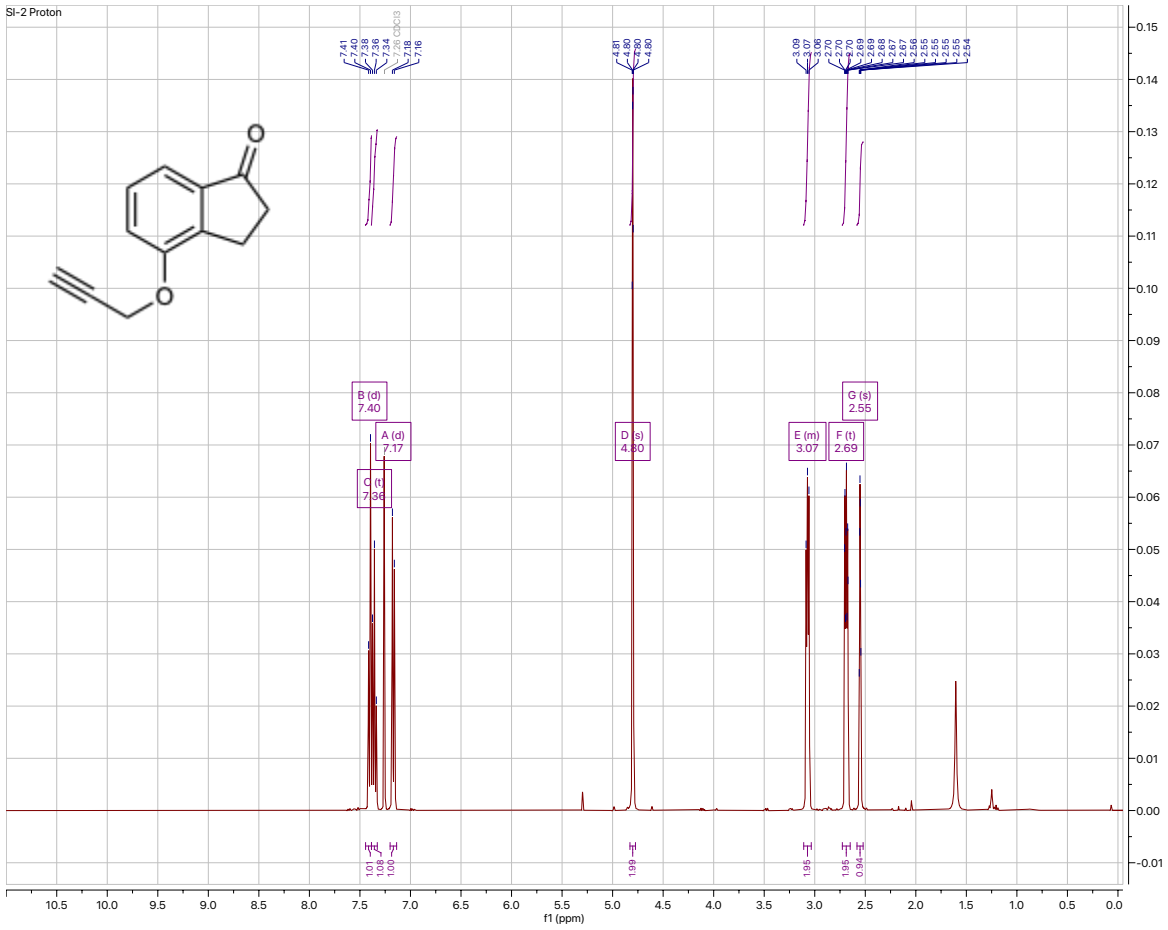
5-Nin-Alk 1H



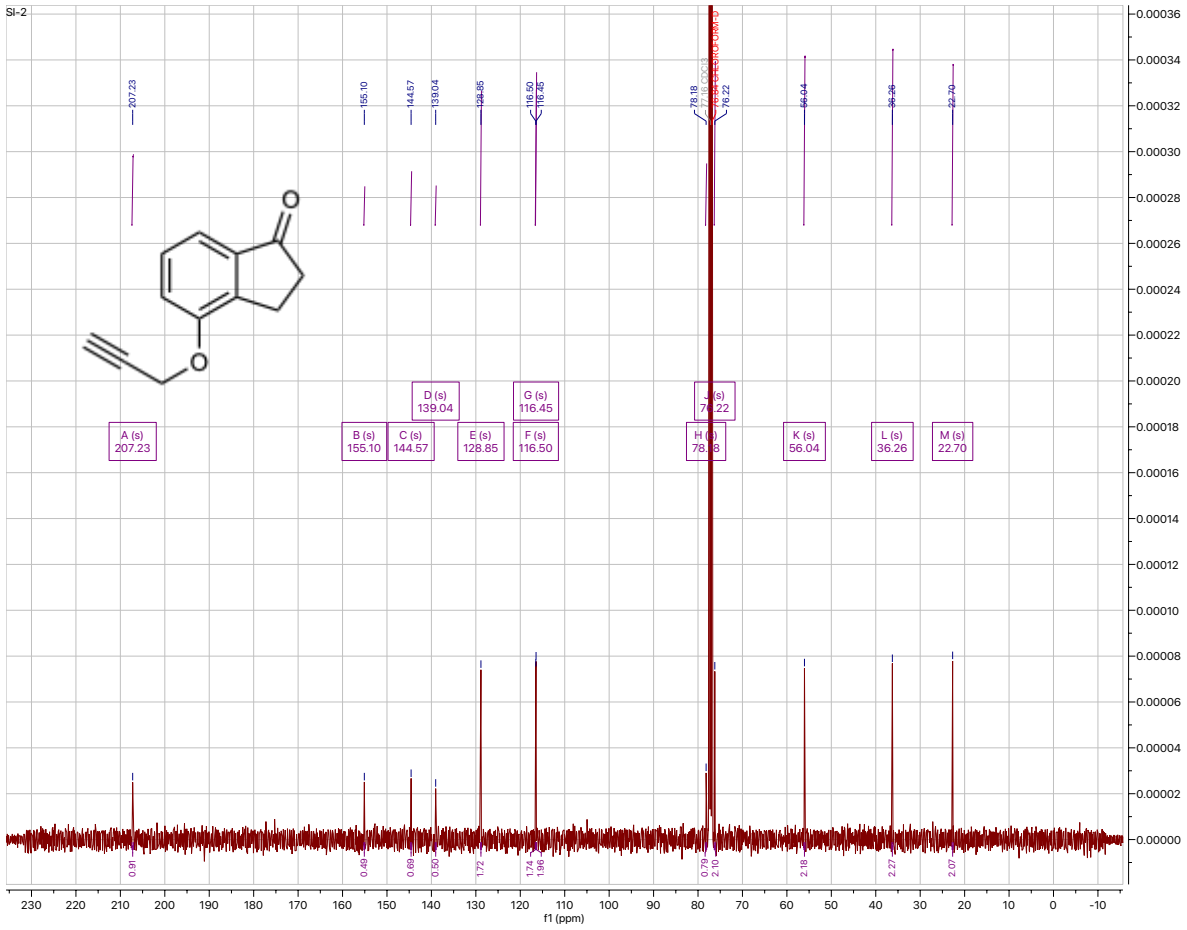
5-Nin-Alk 13C


5-Nin-Alk_Carbon
211.58
196.42
194.93
164.61
141.57
132.96
126.50
126.03
107.13
87.76
77.53
77.29
69.61
56.77
54.46
31.42
28.51
OH
OH
O
O
O
f1 (ppm)

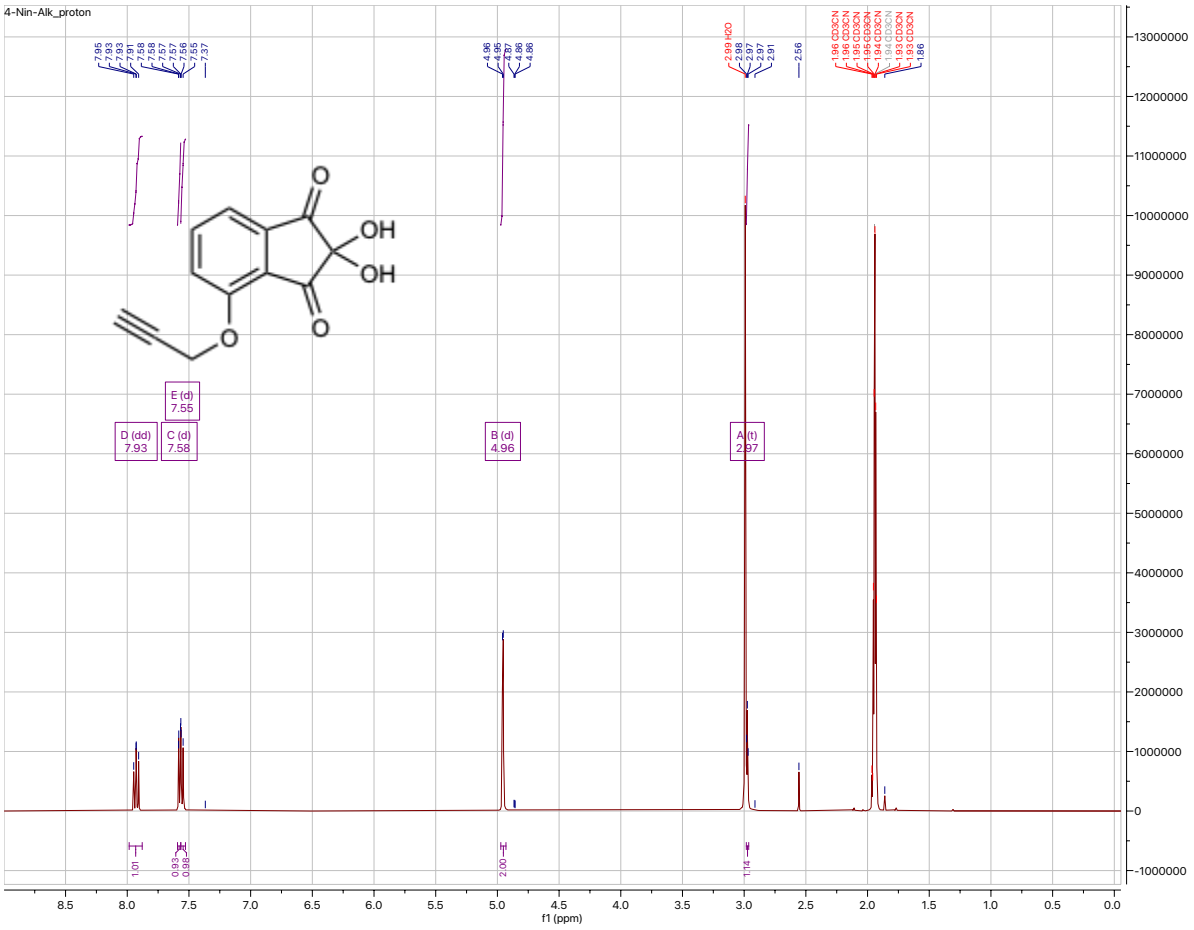
SI-2 1H



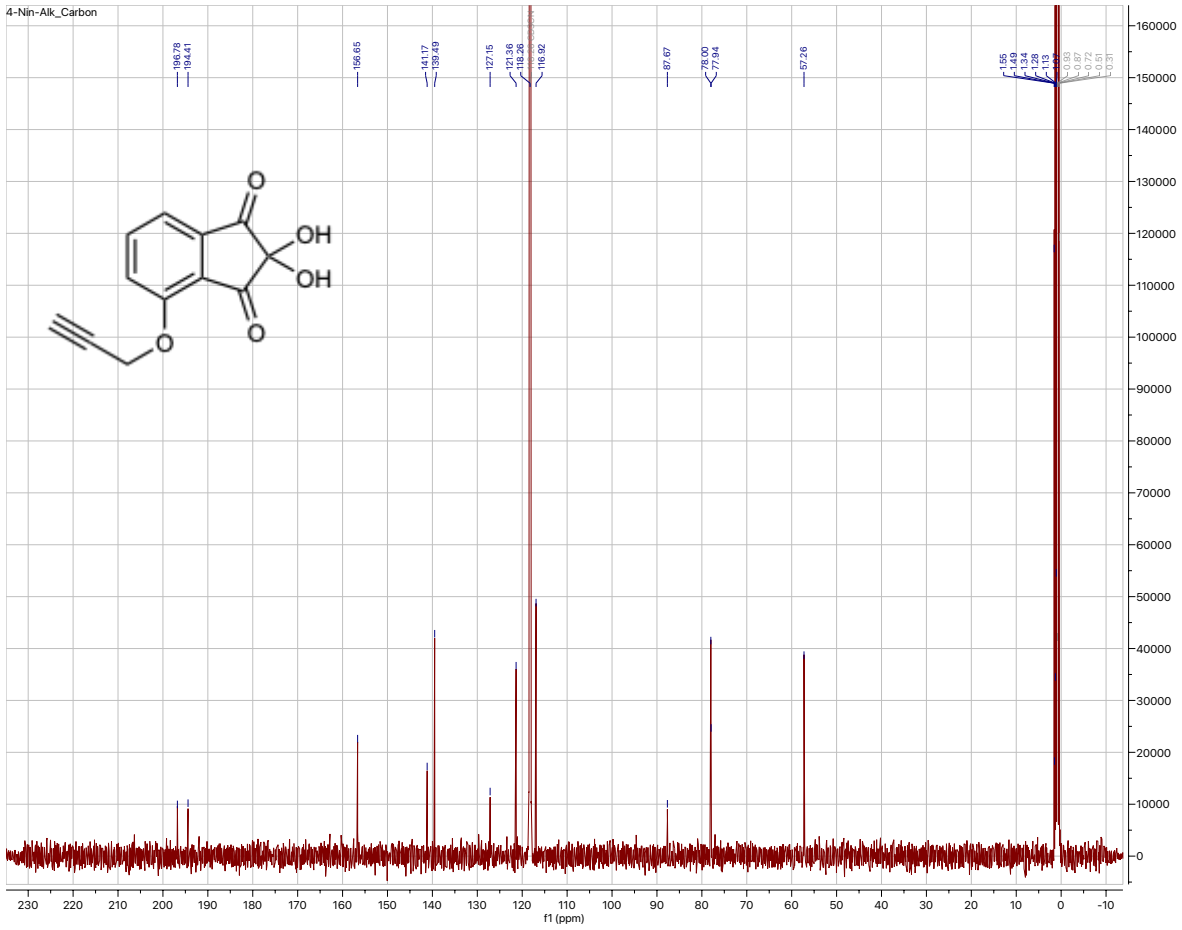
SI-2 13C


SI-2
207.23
155.10
144.57
139.04
128.85
116.50
116.45
78.18
76.22
56.04
36.26
22.70
A (s) 207.23
B (s) 155.10
C (s) 144.57
D (s) 139.04
E (s) 128.85
F (s) 116.50
G (s) 116.45
H (s) 78.18
K (s) 56.04
L (s) 36.26
M (s) 22.70
f1 (ppm)

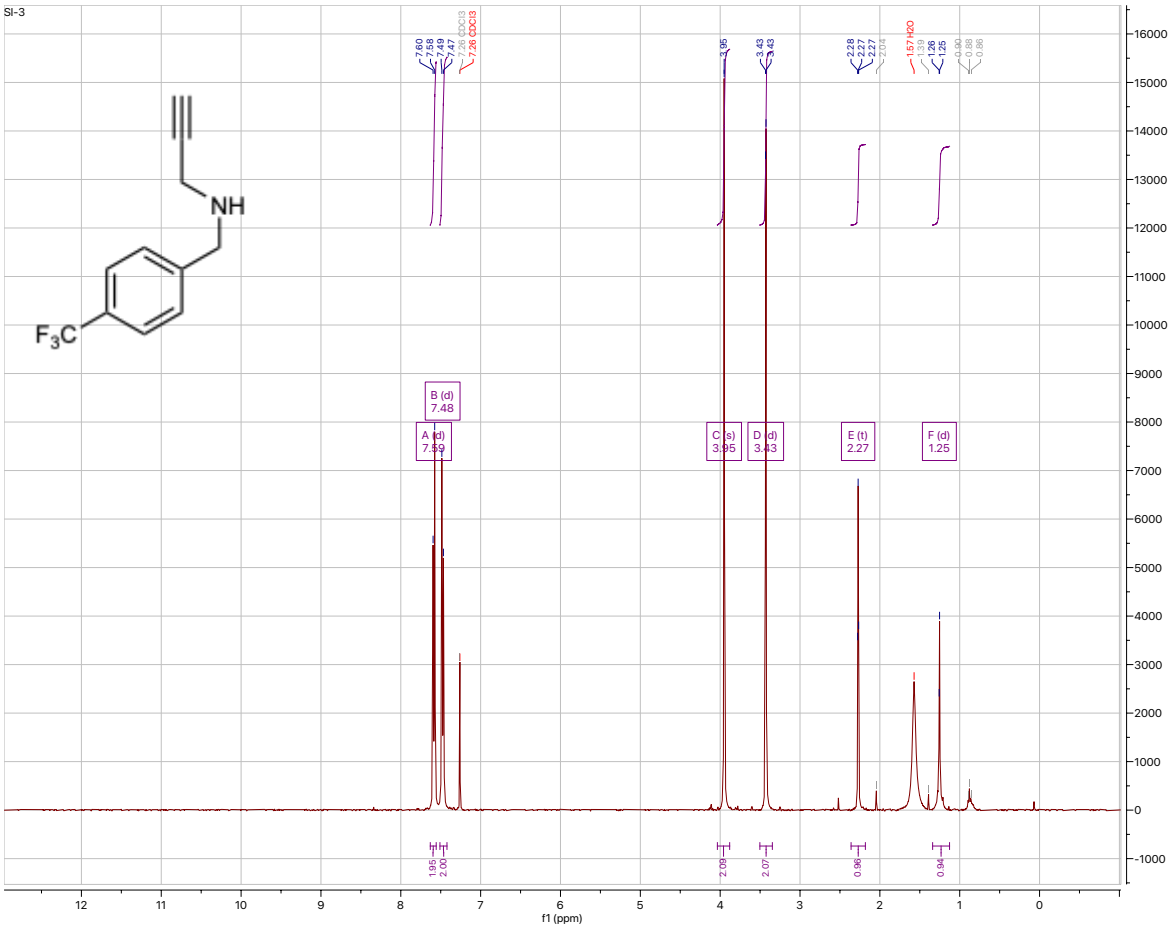
4-Nin-Alk 1H


4-Nin-Alk_proton
OH
OH
O
O
O
E (d)
7.55
D (dd)
7.93
C (d)
7.58
B (d)
4.96
A (t)
2.97
f1 (ppm)

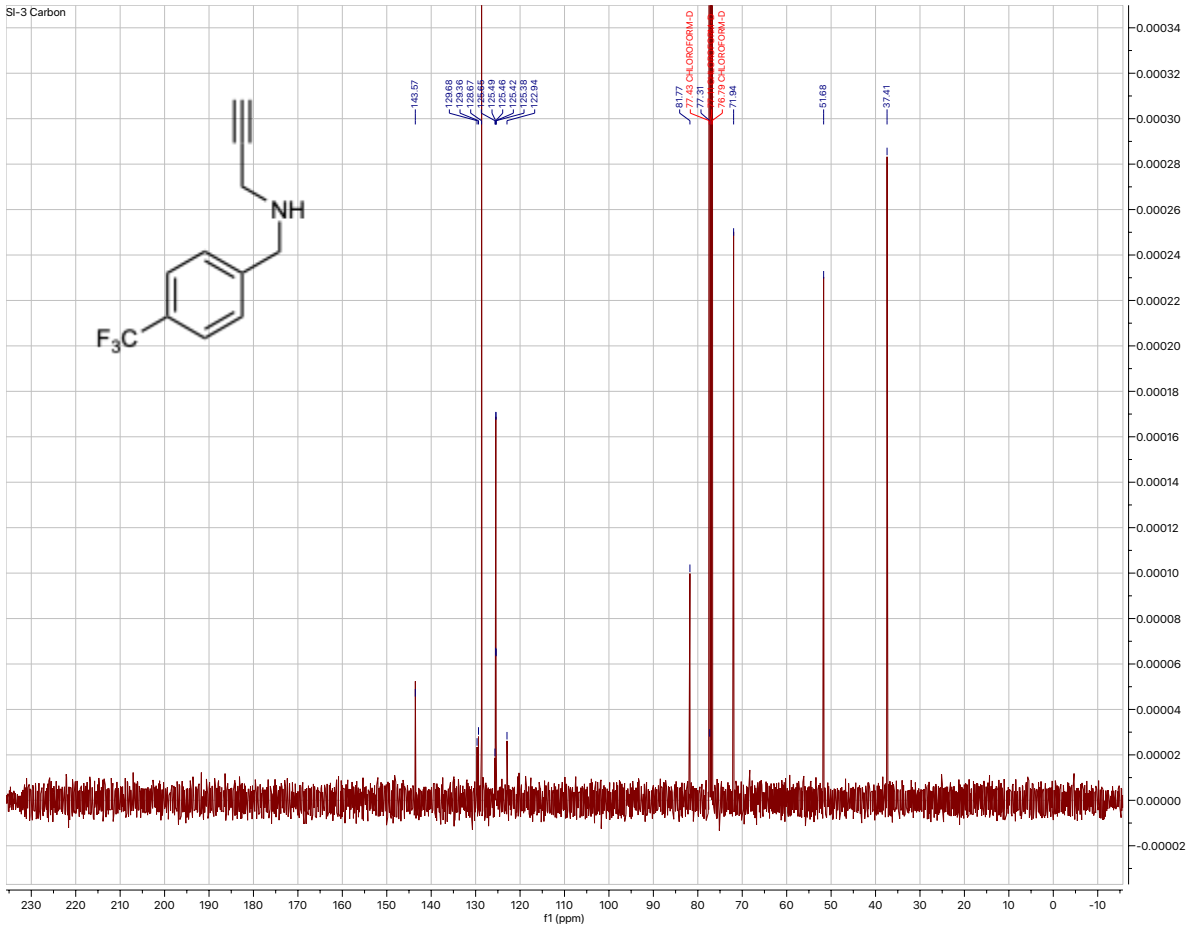
4-Nin-Alk 13C


4-Nin-Alk_Carbon
196.78
194.41
166.65
141.17
139.49
127.15
121.36
118.26
116.92
87.67
78.00
77.94
57.26
1.55
1.49
1.34
1.28
1.13
0.93
0.87
0.72
0.51
0.31
OH
OH
O
O
O
f1 (ppm)

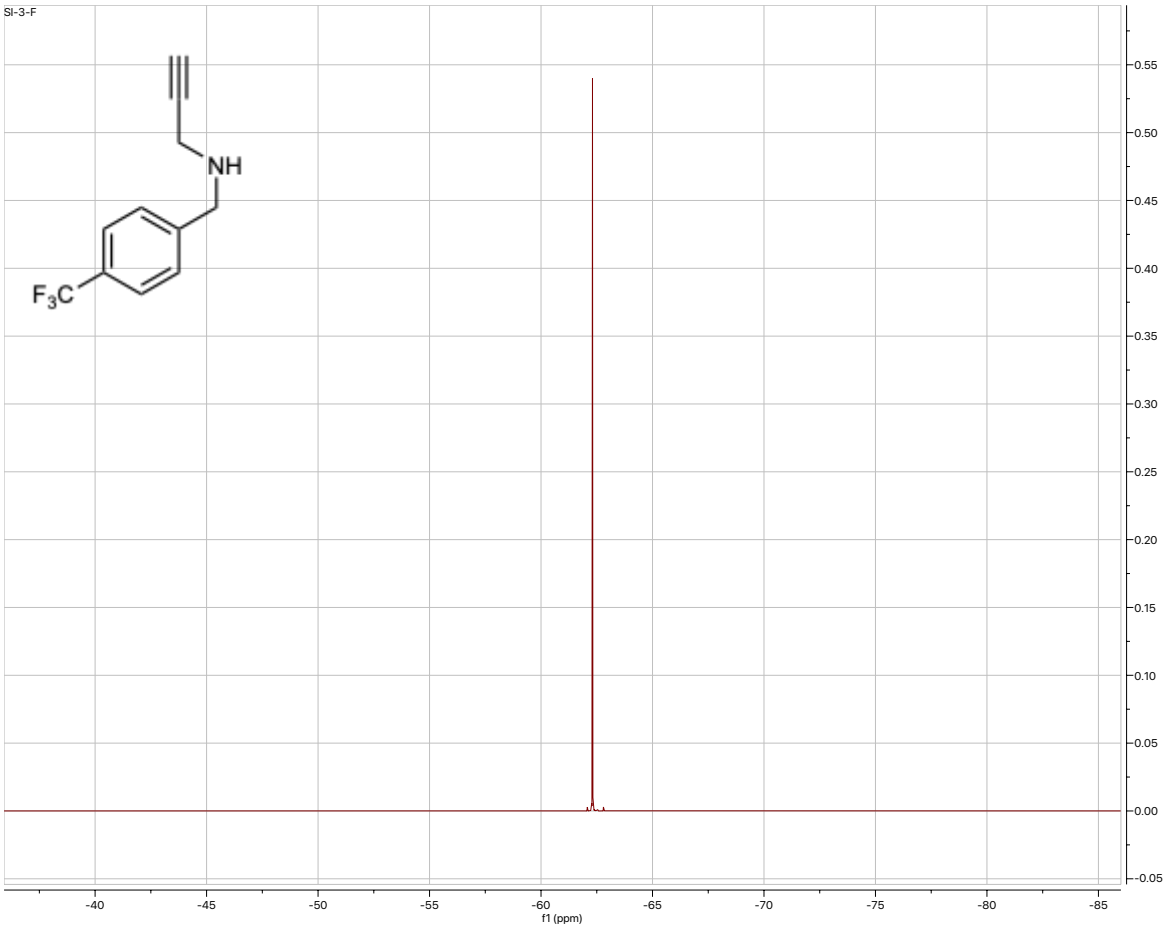
SI-3 1H



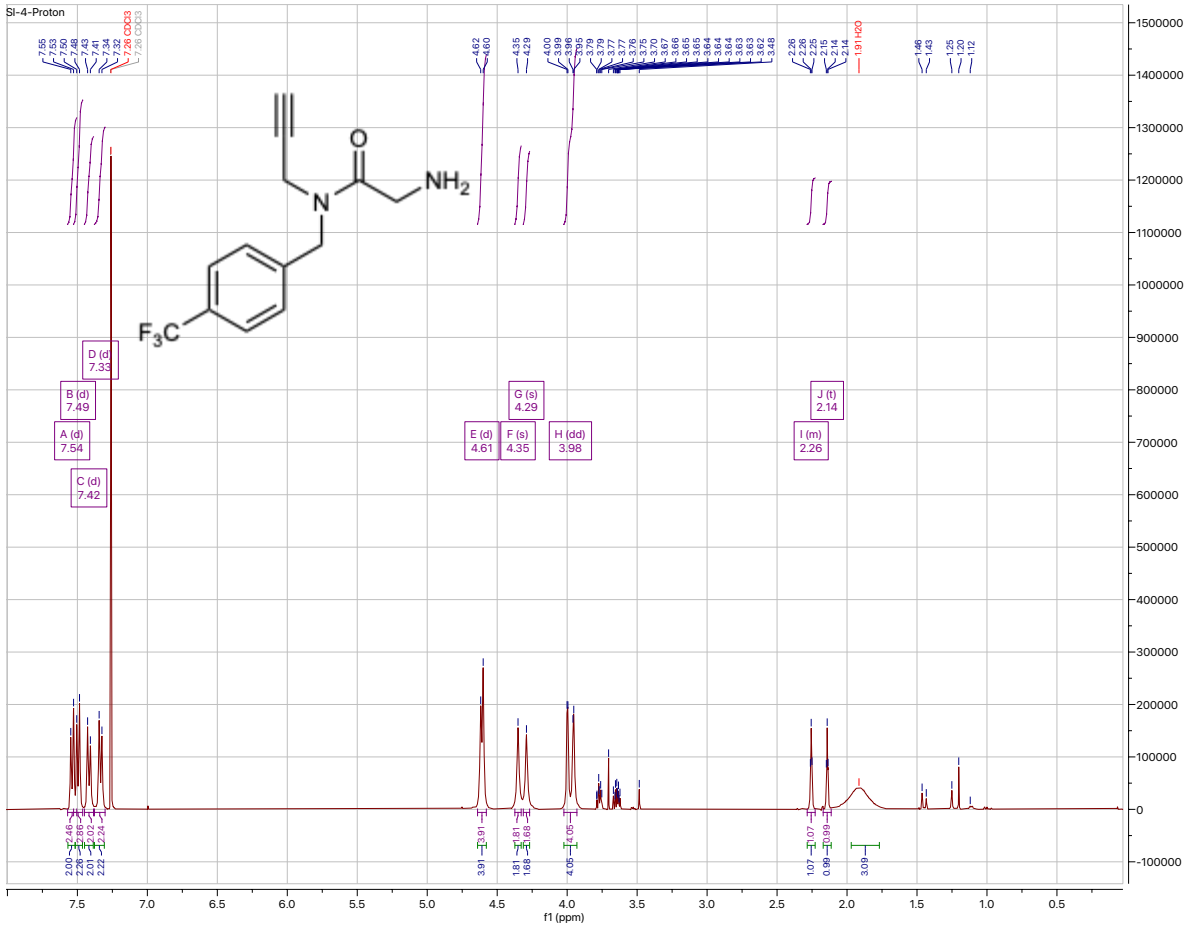
SI-3 13C


SI-3 Carbon
143.57
129.68
129.36
128.67
125.46
125.42
125.38
122.94
81.77
77.48 CHLOROFORM-D
76.79 CHLOROFORM-D
71.94
51.68
37.41
NH
F3C
f1 (ppm)

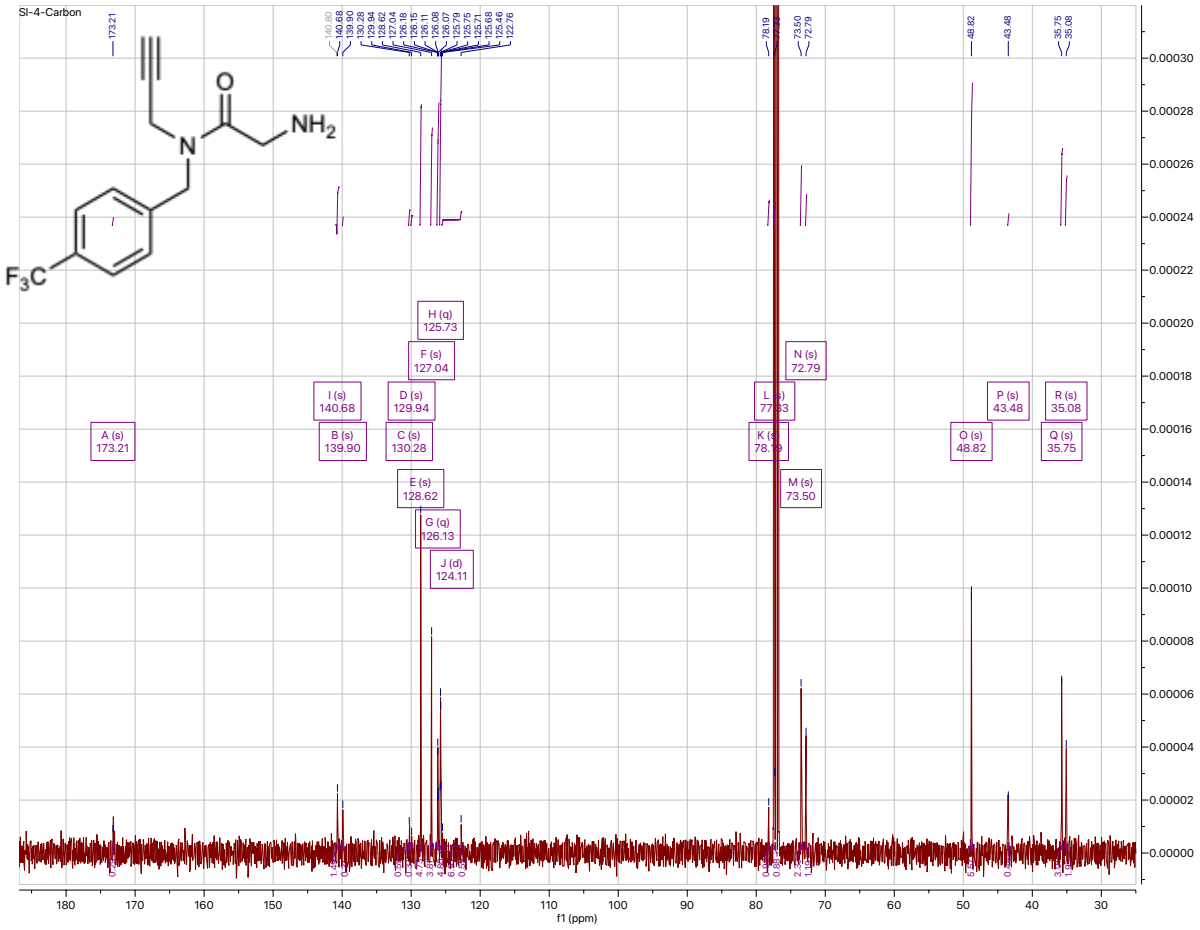
SI-3 19F


SI-3-F
NH
F3C
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-40
-45
-50
-55
-60
-65
-70
-75
-80
-85
f1 (ppm)

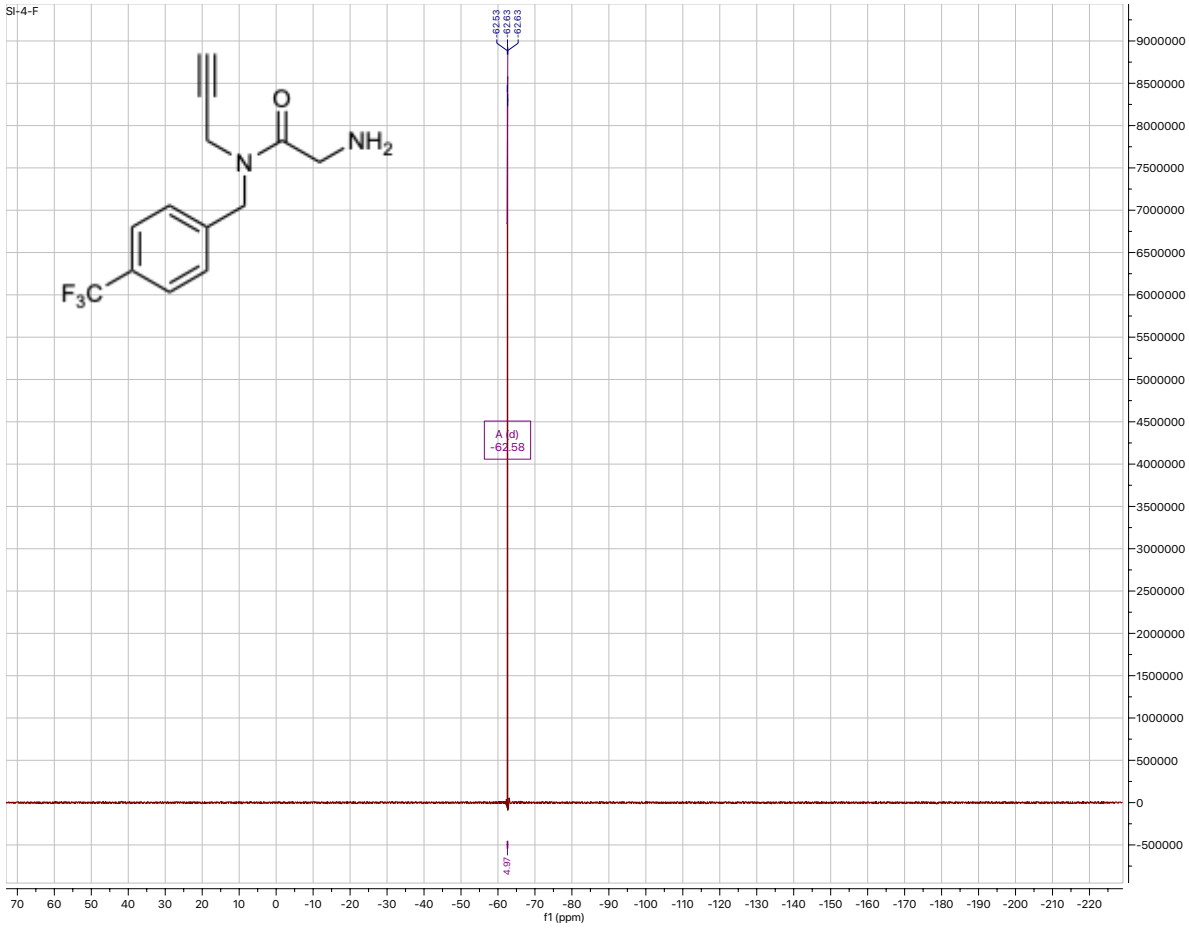
SI-4 1H


SI-4-Proton
O
NH2
N
F3C
D (d) 7.33
B (d) 7.49
A (d) 7.54
C (d) 7.42
G (s) 4.29
E (d) 4.61
F (s) 4.35
H (dd) 3.98
J (t) 2.14
I (m) 2.26
f1 (ppm)

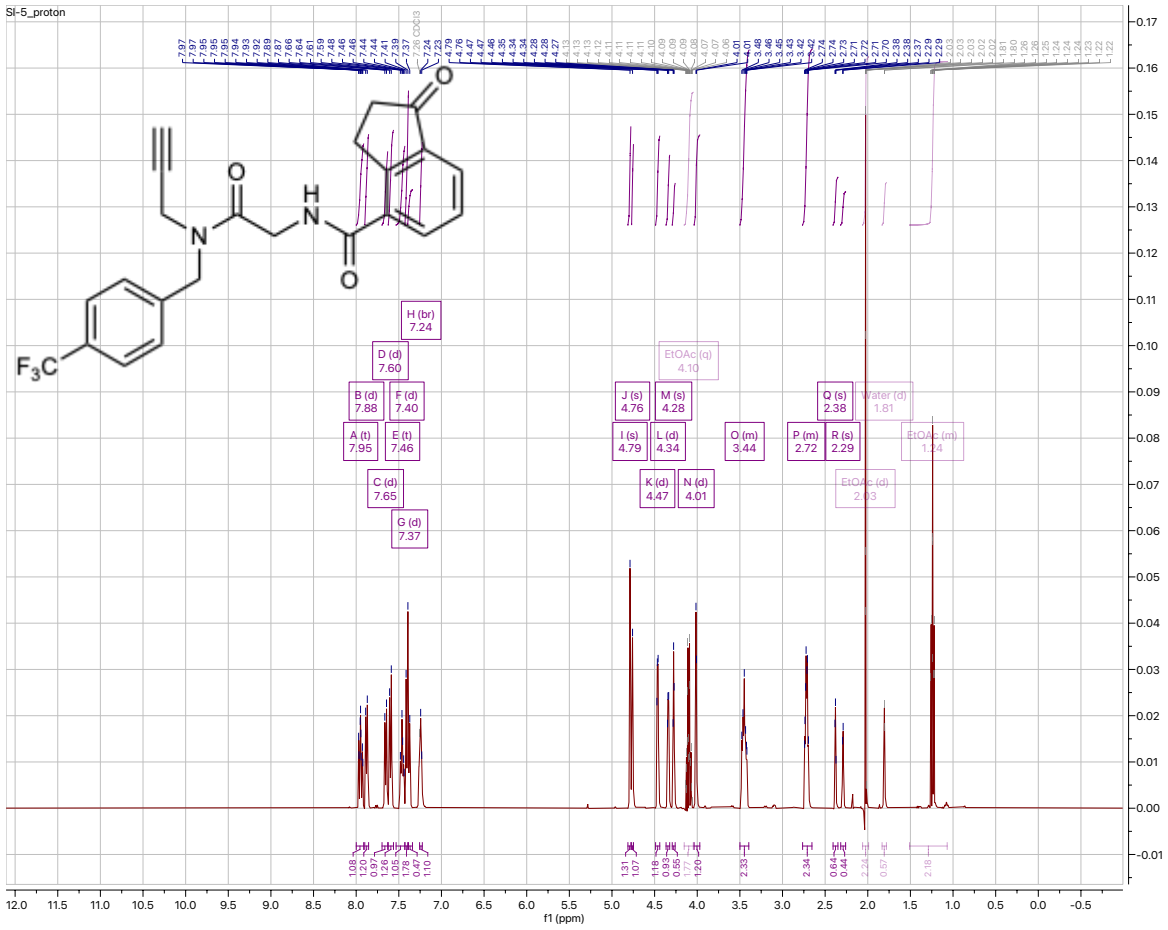
SI-4 13C


SI-4-Carbon
O
NH2
N
F3C
A (s) 173.21
B (s) 139.90
C (s) 130.28
D (s) 129.94
E (s) 128.62
F (s) 127.04
G (q) 126.13
H (q) 125.73
I (s) 140.68
J (d) 124.11
K (s) 78.19
L (s) 77.33
M (s) 73.50
N (s) 72.79
O (s) 48.82
P (s) 43.48
Q (s) 35.75
R (s) 35.08
f1 (ppm)

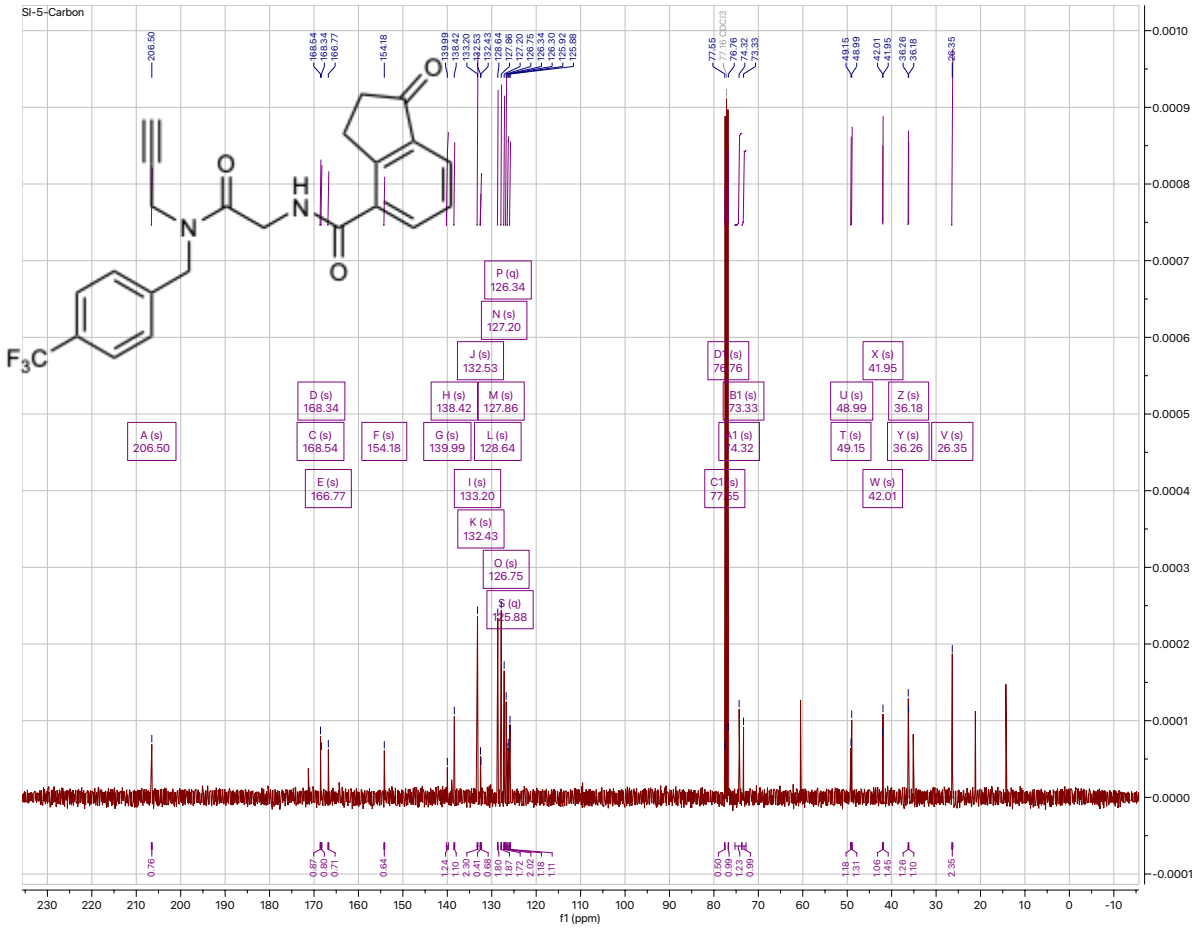
SI-4 19F


SI-4-F
-62.63
-62.63
-62.63
O
N
NH2
F3C
A (d)
-62.58
4.97
9000000
8500000
8000000
7500000
7000000
6500000
6000000
5500000
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
-500000
70 60 50 40 30 20 10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190 -200 -210 -220
f1 (ppm)

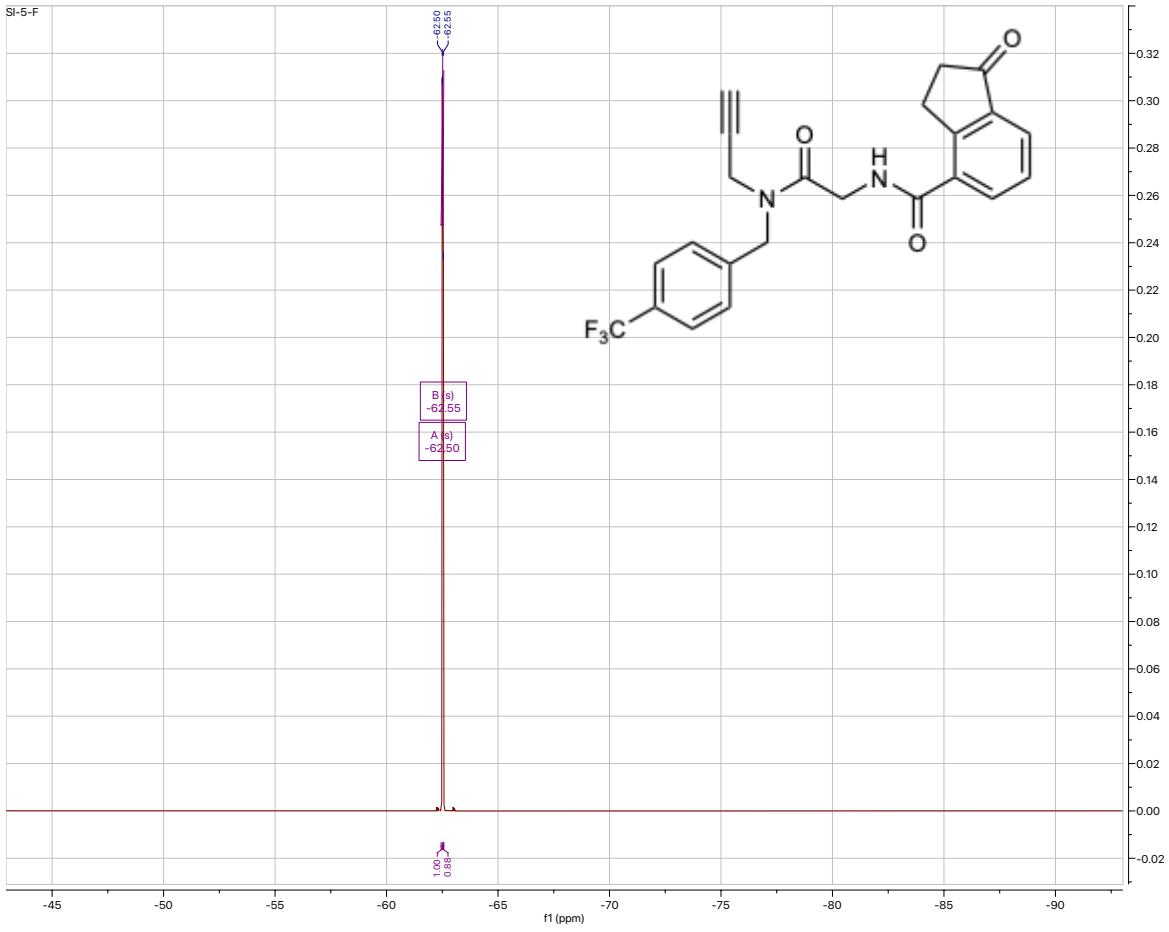
SI-5 1H


SI-5_proton
H (br) 7.24
D (d) 7.60
B (d) 7.88
F (d) 7.40
A (t) 7.95
E (t) 7.46
C (d) 7.65
G (d) 7.37
EtOAc (q) 4.10
J (s) 4.76
M (s) 4.28
I (s) 4.79
L (d) 4.34
K (d) 4.47
N (d) 4.01
O (m) 3.44
Q (s) 2.38
Water (d) 1.81
P (m) 2.72
R (s) 2.29
EtOAc (m) 1.24
EtOAc (d) 2.03
f1 (ppm)

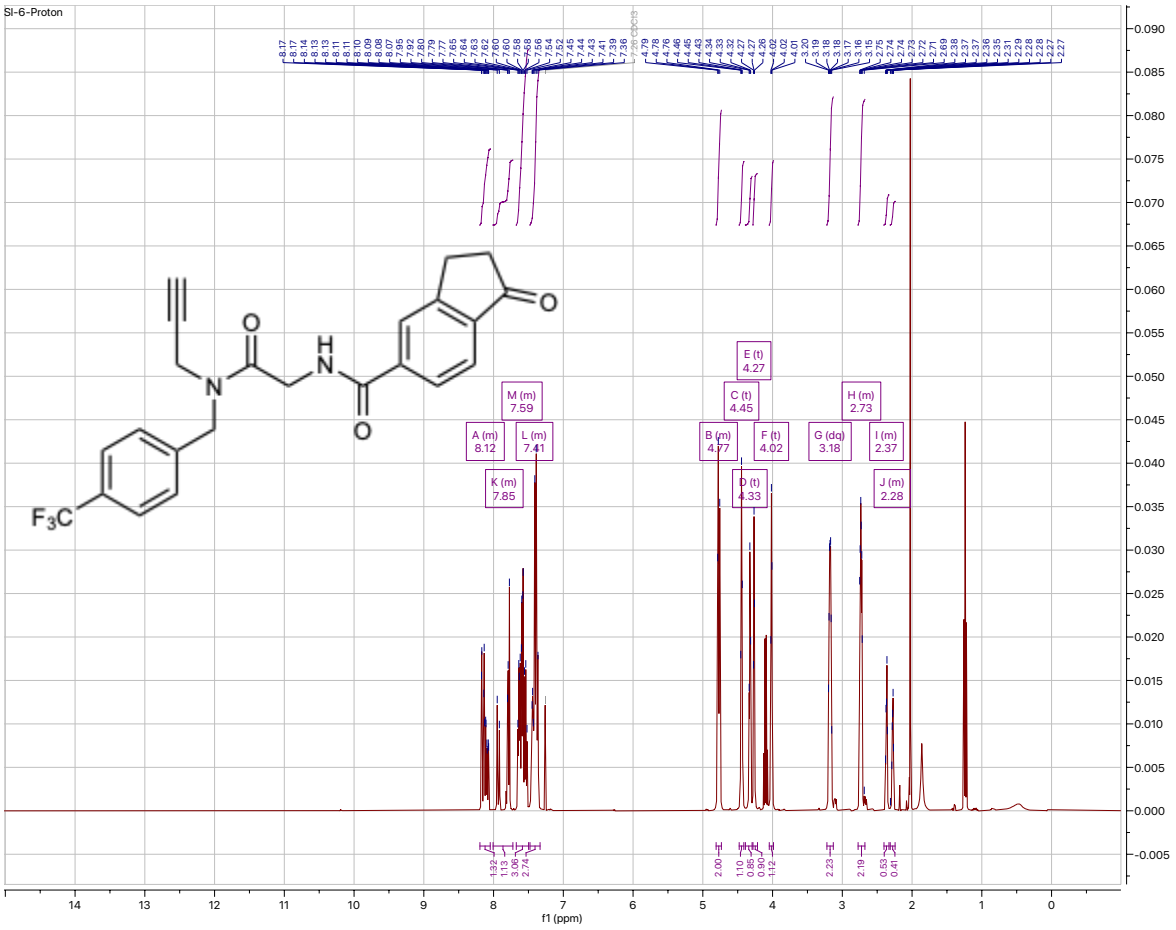
SI-5 13C



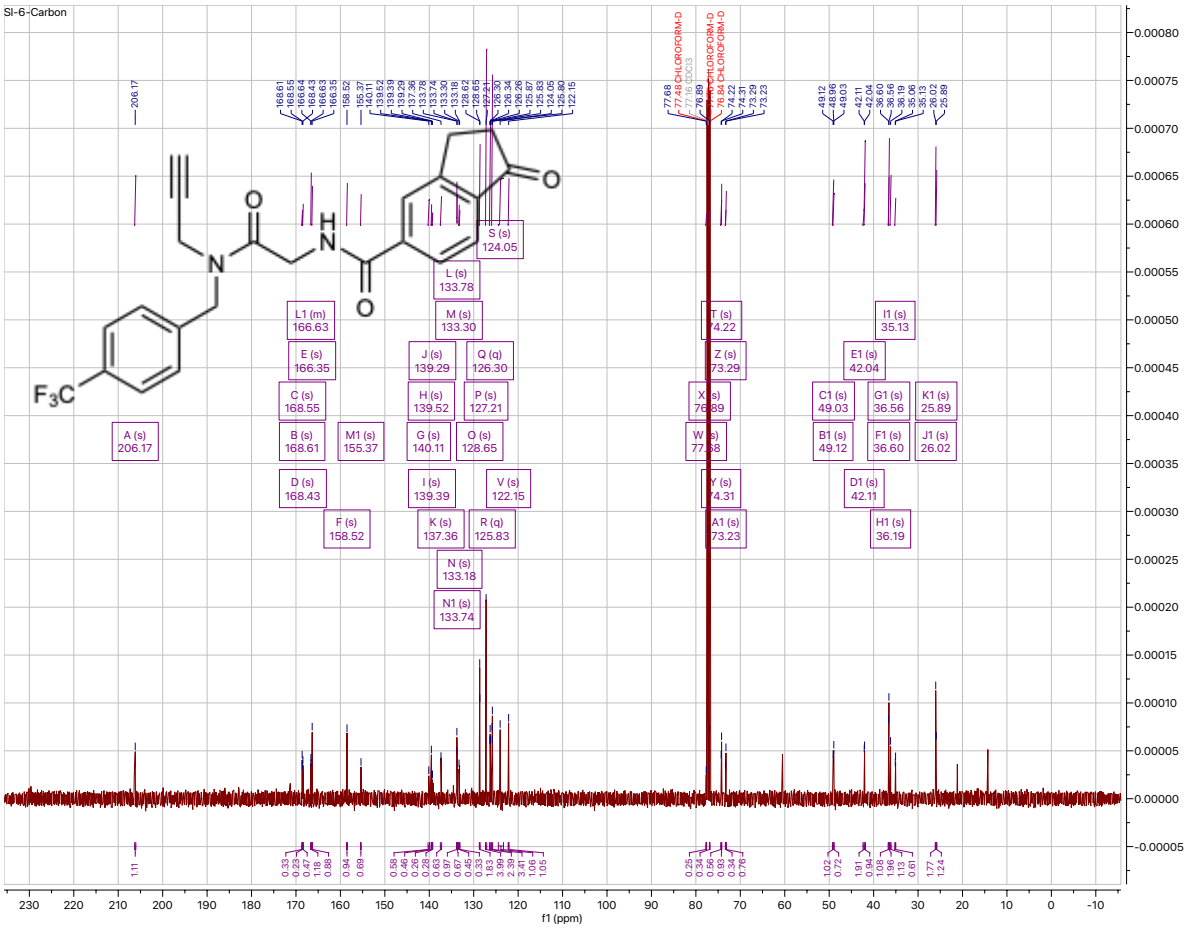
SI-5 19F



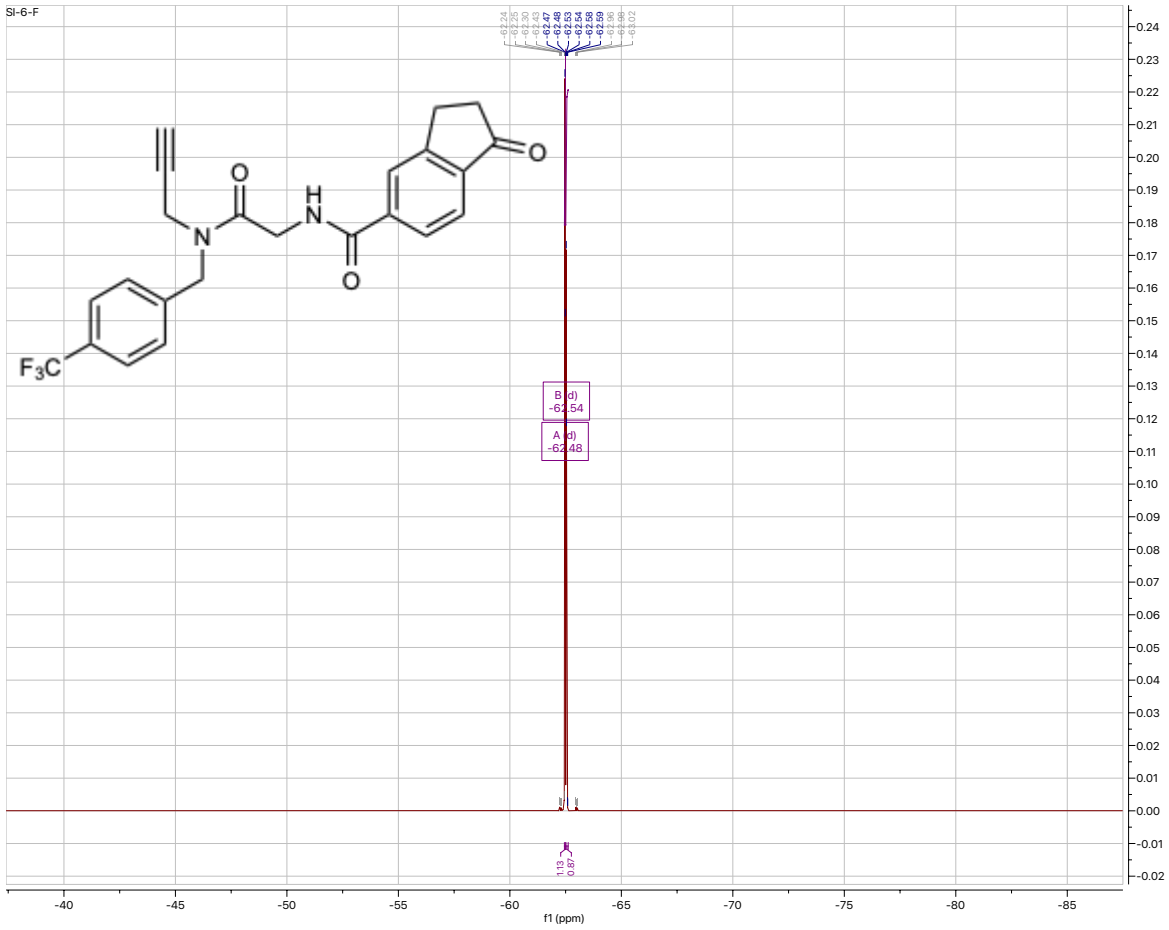
SI-6 1H


SI-6-Proton
A (m) 8.12
K (m) 7.85
L (m) 7.41
M (m) 7.59
B (m) 4.77
C (t) 4.45
D (t) 4.33
E (t) 4.27
F (t) 4.02
G (dq) 3.18
H (m) 2.73
I (m) 2.37
J (m) 2.28
f1 (ppm)

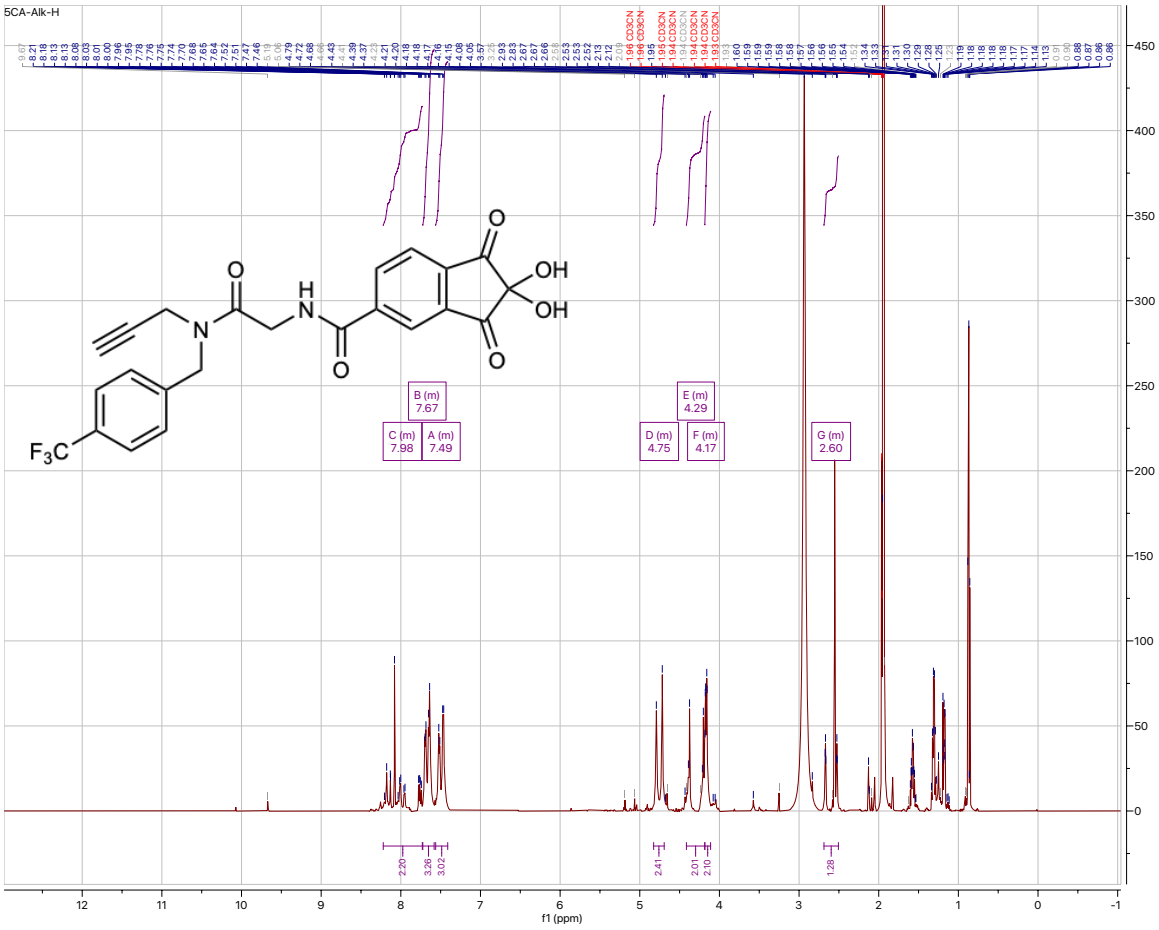
SI-6 13C



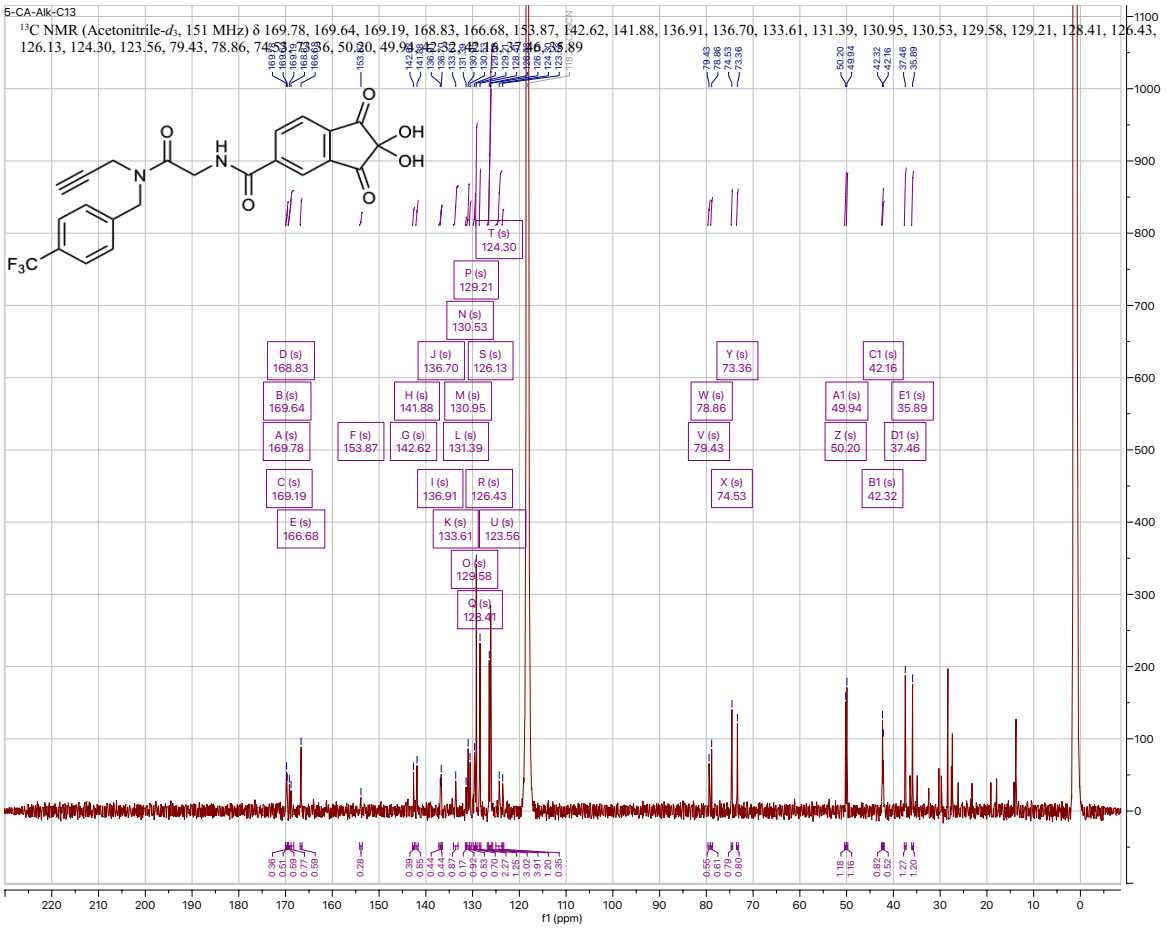
SI-6 19F



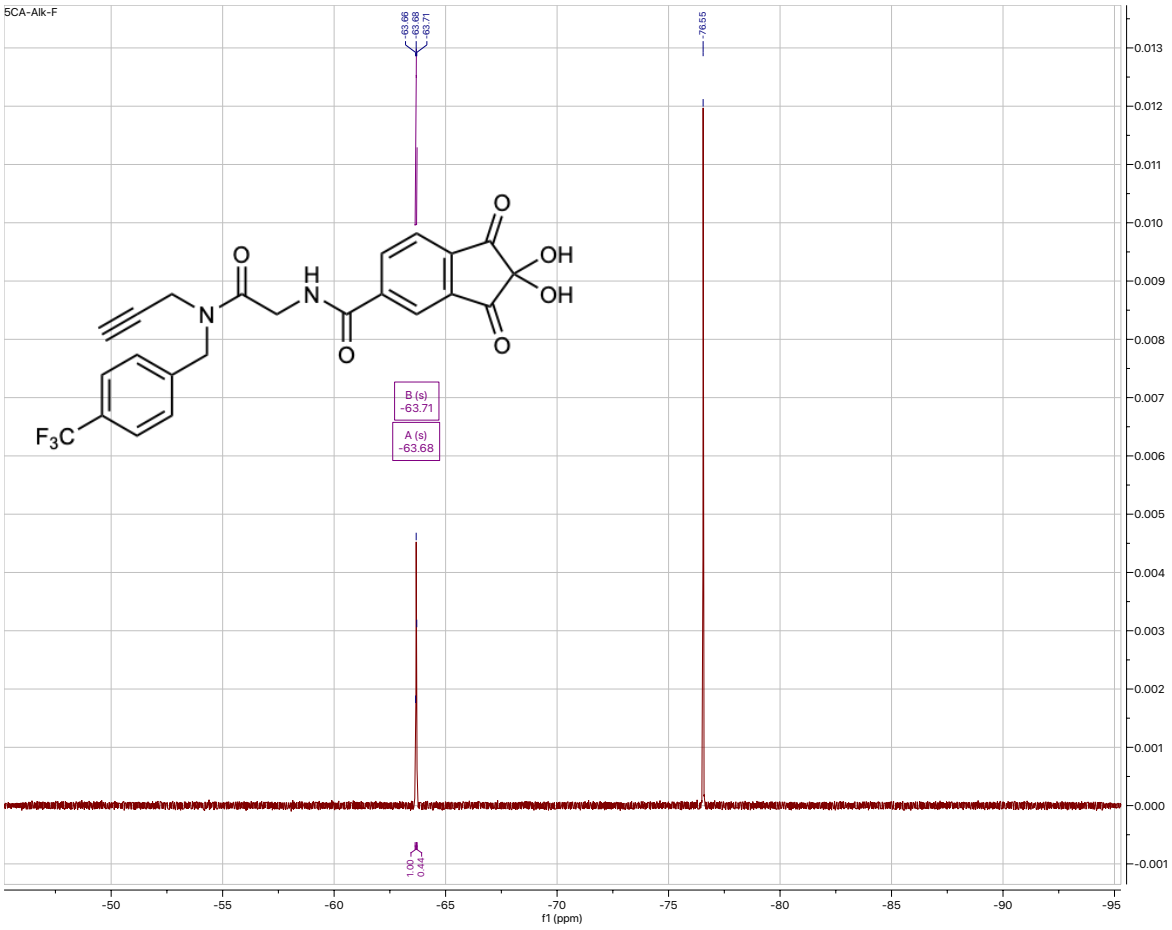
5-CA-Alk 1H



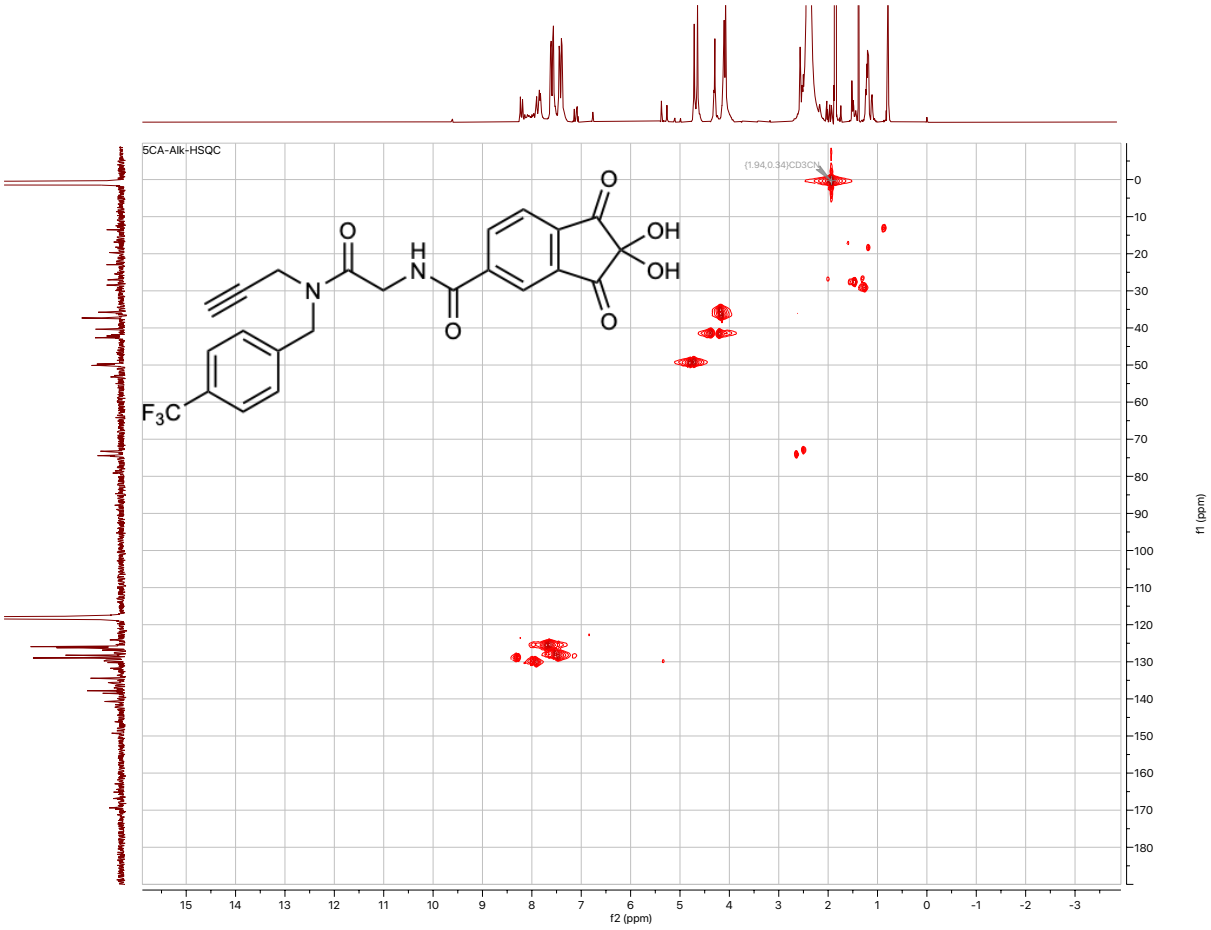
5-CA-Alk 13C


5-CA-Alk-C13
13C NMR (Acetonitrile-d3, 151 MHz) δ 169.78, 169.64, 169.19, 168.83, 166.68, 153.87, 142.62, 141.88, 136.91, 136.70, 133.61, 131.39, 130.95, 130.53, 129.58, 129.21, 128.41, 126.43, 126.13, 124.30, 123.56, 79.43, 78.86, 74.53, 73.36, 50.20, 49.94, 42.32, 42.16, 37.46, 35.89
T (s) 124.30
P (s) 129.21
N (s) 130.53
D (s) 168.83
J (s) 136.70
S (s) 126.13
Y (s) 73.36
C1 (s) 42.16
B (s) 169.64
H (s) 141.88
M (s) 130.95
W (s) 78.86
A1 (s) 49.94
E1 (s) 35.89
A (s) 169.78
F (s) 153.87
G (s) 142.62
L (s) 131.39
V (s) 79.43
Z (s) 50.20
D1 (s) 37.46
C (s) 169.19
I (s) 136.91
R (s) 126.43
X (s) 74.53
B1 (s) 42.32
E (s) 166.68
K (s) 133.61
U (s) 123.56
O (s) 129.58
Q (s) 128.41
f1 (ppm)

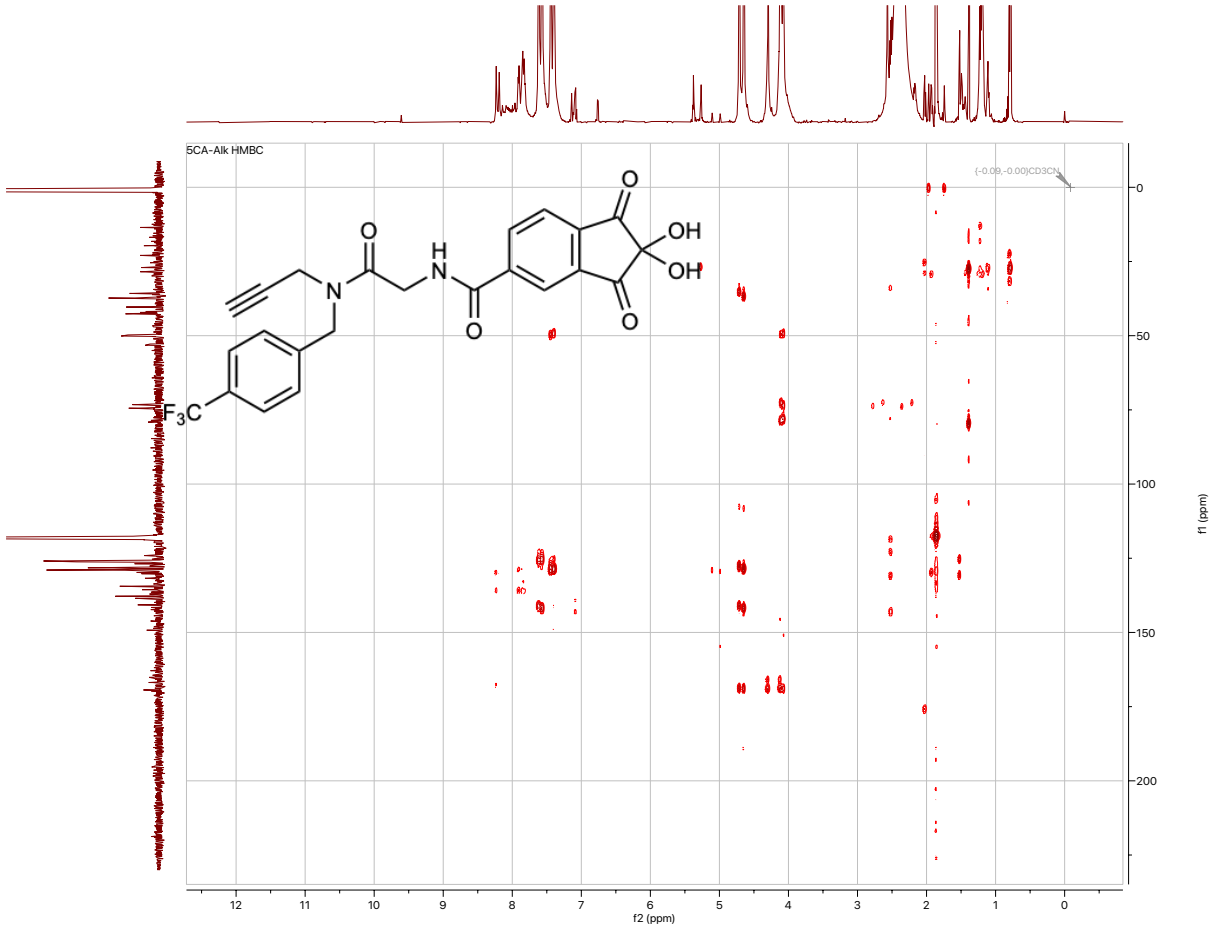
5-CA-Alk 19F



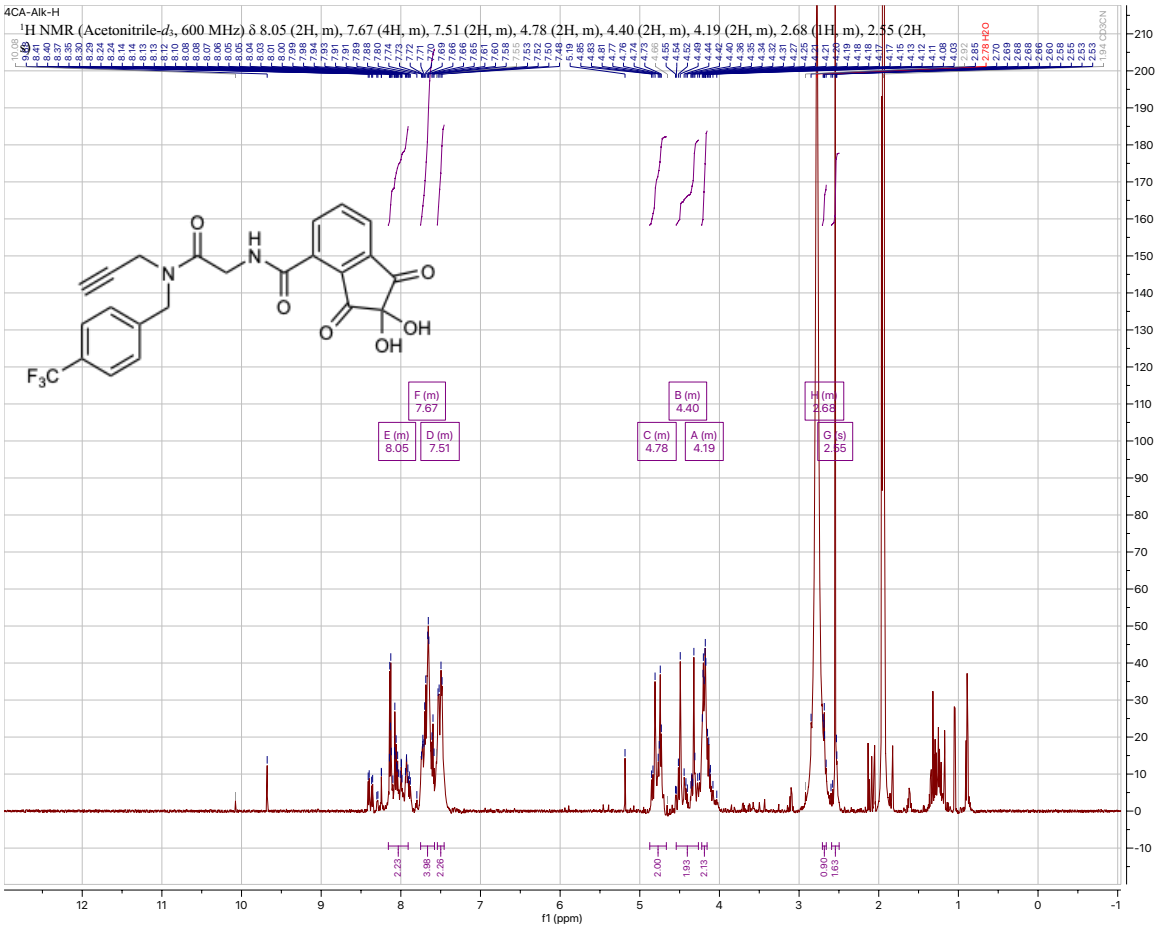
5-CA-Alk HSQC



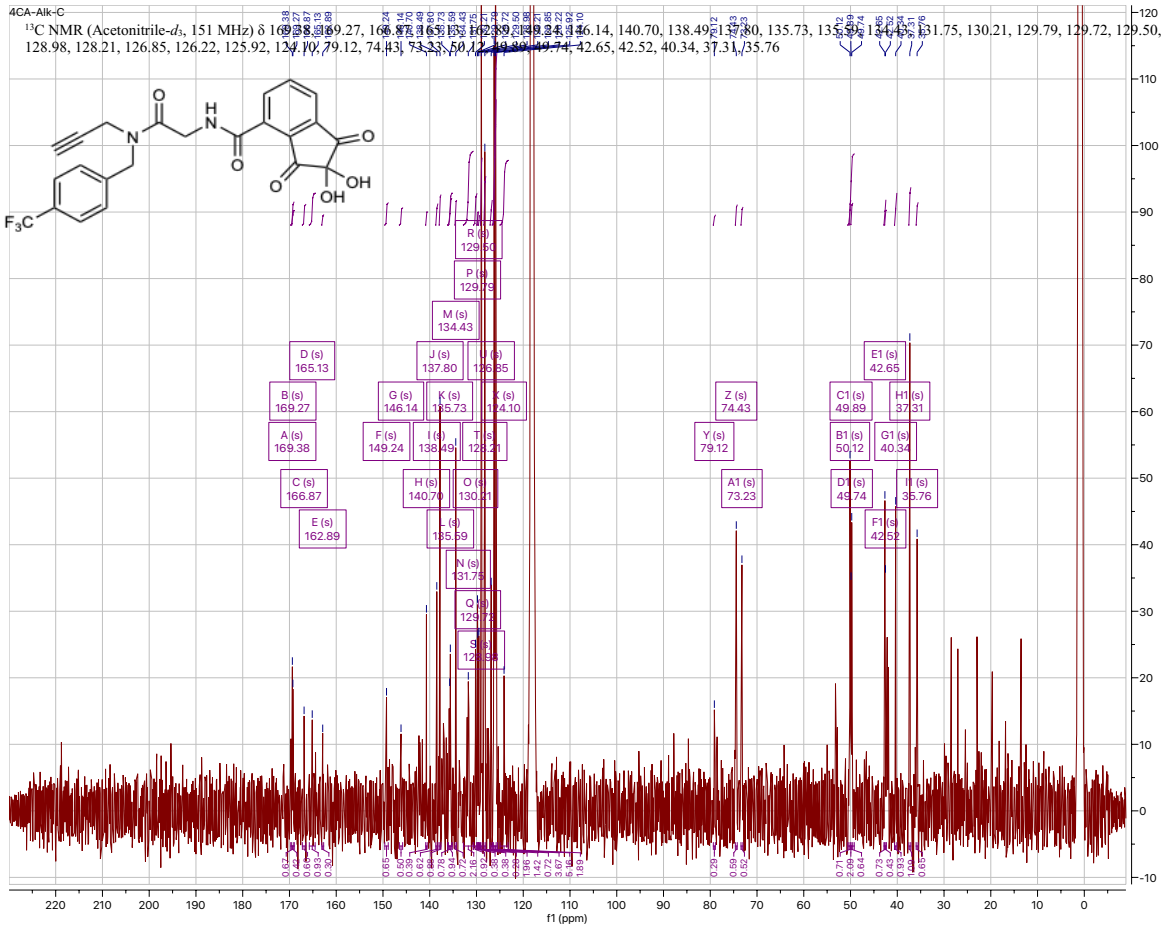
5-CA-Alk HMBC



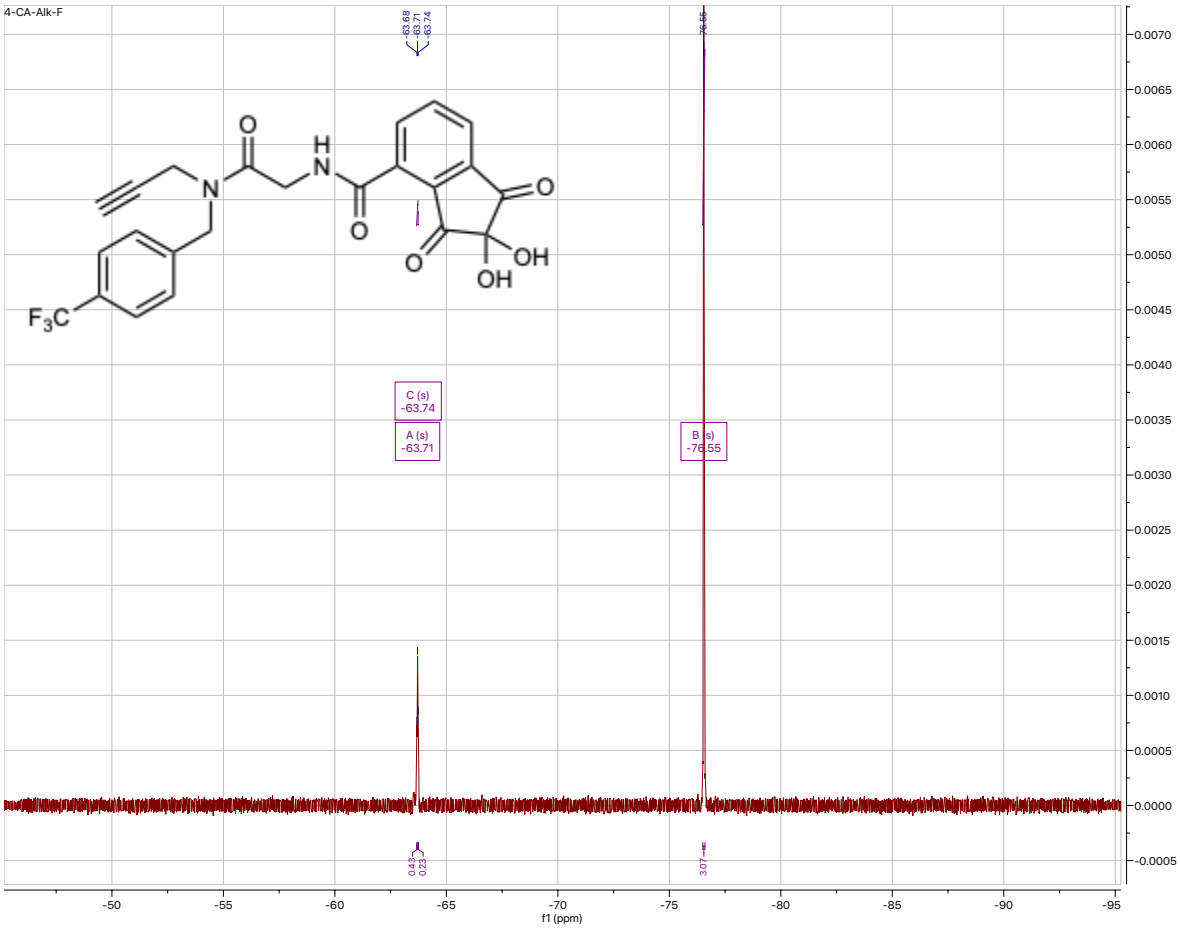
4-CA-Alk 1H


4CA-Alk-H
¹H NMR (Acetonitrile-d_3, 600 MHz) δ 8.05 (2H, m), 7.67 (4H, m), 7.51 (2H, m), 4.78 (2H, m), 4.40 (2H, m), 4.19 (2H, m), 2.68 (1H, m), 2.55 (2H,
F (m) 7.67
E (m) 8.05
D (m) 7.51
B (m) 4.40
C (m) 4.78
A (m) 4.19
G (s) 2.55
f1 (ppm)

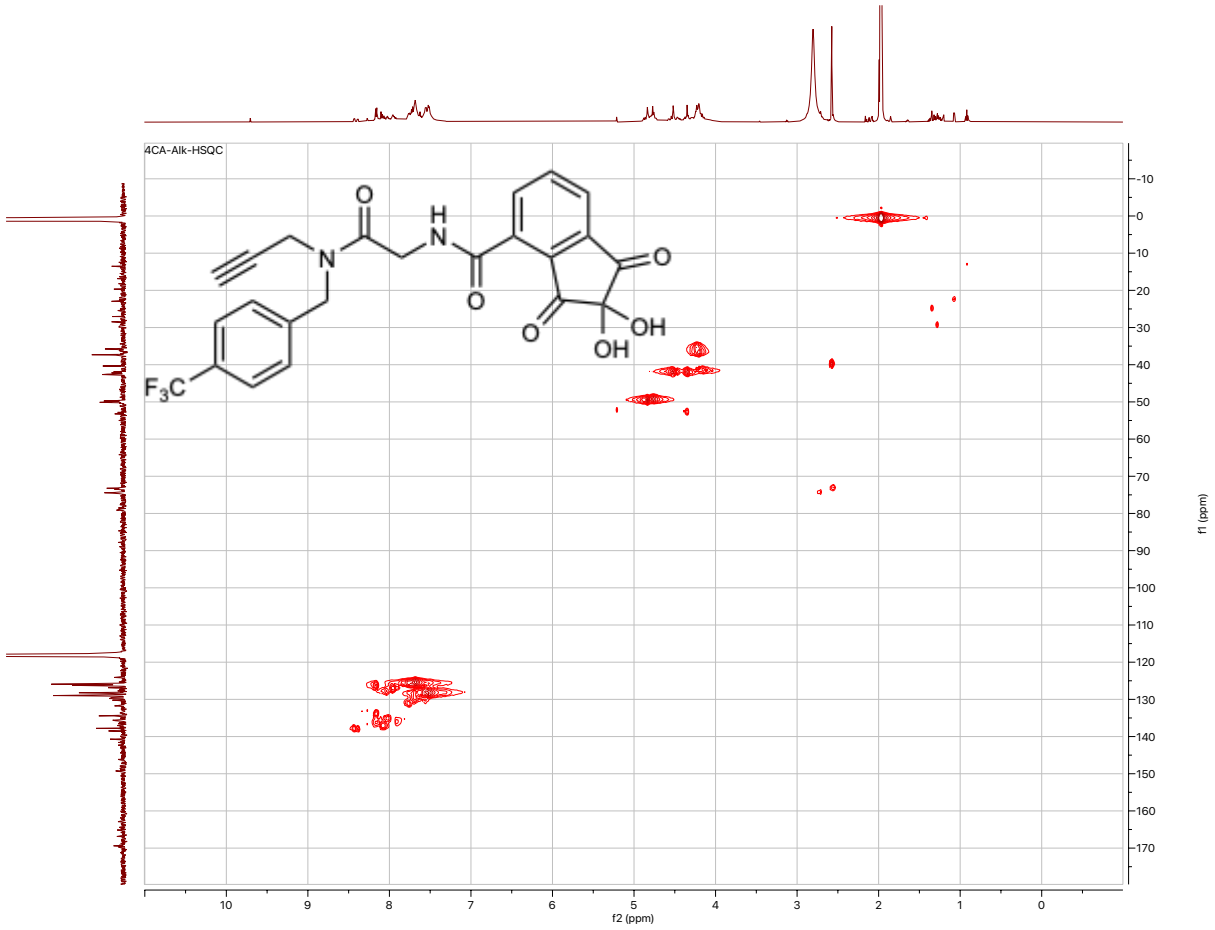
4-CA-Alk 13C


4CA-Alk-C
^{13}C NMR (Acetonitrile-d_3, 151 MHz) δ 169.38, 169.27, 166.87, 165.13, 162.89, 149.24, 146.14, 140.70, 138.49, 137.80, 135.73, 134.43, 131.75, 130.21, 129.79, 129.72, 129.50, 128.98, 128.21, 126.85, 126.22, 125.92, 124.00, 79.12, 74.43, 73.23, 50.12, 49.89, 49.74, 42.65, 42.52, 40.34, 37.31, 35.76
R (s) 129.50
P (s) 129.79
M (s) 134.43
D (s) 165.13
J (s) 137.80
O (s) 126.85
E1 (s) 42.65
B (s) 169.27
G (s) 146.14
K (s) 135.73
X (s) 124.10
Z (s) 74.43
C1 (s) 49.89
H1 (s) 37.31
A (s) 169.38
F (s) 149.24
I (s) 138.49
T (s) 128.21
Y (s) 79.12
B1 (s) 50.12
G1 (s) 40.34
C (s) 166.87
H (s) 140.70
O (s) 130.21
A1 (s) 73.23
D1 (s) 49.74
I1 (s) 35.76
E (s) 162.89
L (s) 135.59
F1 (s) 42.52
N (s) 131.75
Q (s) 129.72
f1 (ppm)

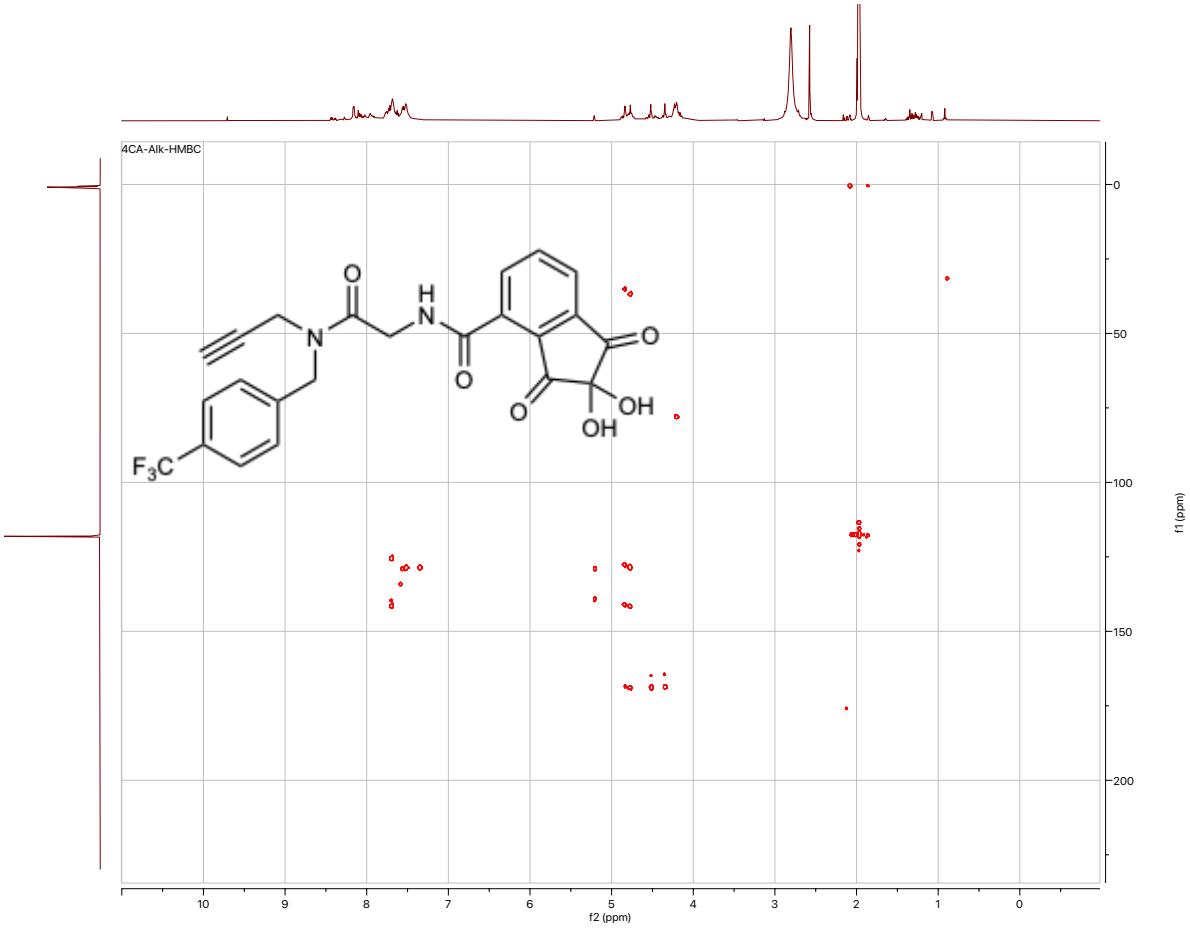
4-CA-Alk 19F


4-CA-Alk-F
-63.68
-63.71
-63.74
-76.55
C (s)
-63.74
A (s)
-63.71
B (s)
-76.55
0.43
0.23
3.07
f1 (ppm)

4-CA-Alk HSQC



4-CA-Alk HMBC


4CA-Alk-HMBC
O
H
N
N
O
O
O
OH
OH
F3C
0
50
100
150
200
f1 (ppm)
10
9
8
7
6
5
4
3
2
1
0
f2 (ppm)

References

1. Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).

2. Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570–574 (2016).

3. Abbasov, M. E. *et al.* A proteome-wide atlas of lysine-reactive chemistry. *Nat. Chem.* **13**, 1081–1092 (2021).

4. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).

5. Yang, T. *et al.* Reversible lysine-targeted probes reveal residence time-based kinase selectivity. *Nat. Chem. Biol.* **18**, 934–941 (2022).

6. Adibekian, A. *et al.* Click-generated triazole ureas as ultrapotent in vivo–active serine hydrolase inhibitors. *Nat. Chem. Biol.* **7**, 469–478 (2011).

7. Zanon, P. R. A. *et al.* Profiling the proteome-wide selectivity of diverse electrophiles. Preprint at https://doi.org/10.26434/chemrxiv-2021-w7rss-v2 (2021).

8. Xie, X. *et al.* Oxidative cyclization reagents reveal tryptophan cation–π interactions. *Nature* **627**, 680–687 (2024).

9. Hahm, H. S. *et al.* Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. *Nat. Chem. Biol.* **16**, 150–159 (2020).

10. Wang, N. *et al.* Genetically Encoding Fluorosulfate- l -tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins *in Vivo*. *J. Am. Chem. Soc.* **140**, 4995–4999 (2018).

11. Gupta, M. N. & Uversky, V. N. Biological importance of arginine: A comprehensive review of the roles in structure, disorder, and functionality of peptides and proteins. *Int. J. Biol. Macromol.* **257**, 128646 (2024).

12. Diels, O. & Schleich, K. Über Bildung und Eigenschaften der aus 1.2-Diketonen entstehenden Verbindungen. I. Diacetyl und Benzamidin. *Berichte Dtsch. Chem. Ges.* **49**, 1711–1721 (1916).

13. Hellwig, M. & Henle, T. Baking, Ageing, Diabetes: A Short History of the Maillard Reaction. *Angew. Chem. Int. Ed.* **53**, 10316–10329 (2014).

14. Zhang, Z., Morstein, J., Ecker, A. K., Guiley, K. Z. & Shokat, K. M. Chemoselective Covalent Modification of K-Ras(G12R) with a Small Molecule Electrophile. *J. Am. Chem. Soc.* **144**, 15916–15921 (2022).

15. Dovgan, I. *et al.* Arginine-selective bioconjugation with 4-azidophenyl glyoxal: application to the single and dual functionalisation of native antibodies. *Org. Biomol. Chem.* **16**, 1305–1311 (2018).

16. Jones, A. X. *et al.* Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers. *Nat. Commun.* **10**, 3911 (2019).

17. Chen, P. *et al.* Cell-Active, Arginine-Targeting Irreversible Covalent Inhibitors for Non-Kinases and Kinases. *Angew. Chem. Int. Ed.* e202422372 (2025) doi:10.1002/anie.202422372.

18. Conn, R. B. & Davis, R. B. Green Fluorescence of Guanidinium Compounds with Ninhydrin. *Nature* **183**, 1053–1055 (1959).

19. Jong, J. A. W. *et al.* Reactivity of (Vicinal) Carbonyl Compounds with Urea. *ACS Omega* **4**, 11928–11937 (2019).

20. Thompson, D. A., Ng, R. & Dawson, P. E. Arginine selective reagents for ligation to peptides and proteins. *J. Pept. Sci.* **22**, 311–319 (2016).

21. Nelson, K. J., Day, A. E., Zeng, B.-B., King, S. B. & Poole, L. B. Isotope-coded, iodoacetamide-based reagent to determine individual cysteine pKa values by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **375**, 187–195 (2008).

22. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **67**, 3057–3064 (2002).

23. Zanon, P. R. A., Lewald, L. & Hacker, S. M. Isotopically Labeled Desthiobiotin Azide (isoDTB) Tags Enable Global Profiling of the Bacterial Cysteinome. *Angew. Chem. Int. Ed.* **59**, 2829–2836 (2020).

24. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).

25. Yu, F. *et al.* Identification of modified peptides using localization-aware open search. *Nat. Commun.* **11**, 4065 (2020).

26. Böhme, A., Thaens, D., Paschke, A. & Schüürmann, G. Kinetic Glutathione Chemoassay To Quantify Thiol Reactivity of Organic Electrophiles—Application to α,β-Unsaturated Ketones, Acrylates, and Propiolates. *Chem. Res. Toxicol.* **22**, 742–750 (2009).

27. O'Shea, J. P. *et al.* pLogo: a probabilistic approach to visualizing sequence motifs. *Nat. Methods* **10**, 1211–1212 (2013).

28. Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).

29. Kelleher, K. J. *et al.* Pharos 2023: an integrated resource for the understudied human proteome. *Nucleic Acids Res.* **51**, D1405–D1416 (2023).

30. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

31. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).

32. Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci.* **96**, 9459–9464 (1999).

33. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

34. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371 (1973).

35. He, X. *et al.* AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors. *Acta Pharmacol. Sin.* **44**, 1–7 (2023).

36. He, X., Li, J., Shen, S. & Xu, H. E. AlphaFold3 versus experimental structures: assessment of the accuracy in ligand-bound G protein-coupled receptors. *Acta Pharmacol. Sin.* **46**, 1111–1122 (2025).

37. Stevens, A. O. & He, Y. Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. *Biomolecules* **12**, 985 (2022).

38. Liu, R., Yue, Z., Tsai, C.-C. & Shen, J. Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors. *J. Am. Chem. Soc.* **141**, 6553–6560 (2019).

39. Awoonor-Williams, E., Golosov, A. A. & Hornak, V. Benchmarking *In Silico* Tools for Cysteine p K_a Prediction. *J. Chem. Inf. Model.* **63**, 2170–2180 (2023).

40. Whitehurst, C. B. *et al.* Location and Role of Free Cysteinyl Residues in the Sindbis Virus E1 and E2 Glycoproteins. *J. Virol.* **81**, 6231–6240 (2007).

41. Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).

42. Burley, S. K. *et al.* RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451 (2021).

43. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).

44. Lenzen, S. & Mirzaie-Petri, M. Inhibition of aconitase by alloxan and the differential modes of protection of glucose, 3-O-methylglucose, and mannoheptulose. *Naunyn. Schmiedebergs Arch. Pharmacol.* **346**, (1992).

45. Castro, L., Tórtora, V., Mansilla, S. & Radi, R. Aconitases: Non-redox Iron–Sulfur Proteins Sensitive to Reactive Species. *Acc. Chem. Res.* **52**, 2609–2619 (2019).

46. Bak, D. W. & Weerapana, E. Proteomic strategies to interrogate the Fe-S proteome. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1871**, 119791 (2024).

47. Kim, E. *et al.* Mitochondrial aconitase suppresses immunity by modulating oxaloacetate and the mitochondrial unfolded protein response. *Nat. Commun.* **14**, 3716 (2023).

48. Lloyd, S. J., Lauble, H., Prasad, G. S. & Stout, C. D. The mechanism of aconitase: 1.8 Å resolution crystal structure of the S642A:citrate complex. *Protein Sci.* **8**, 2655–2662 (1999).

49. The UniProt Consortium *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

50. De Simone, A. *et al.* A computationally designed binding mode flip leads to a novel class of potent tri-vector cyclophilin inhibitors. *Chem. Sci.* **10**, 542–547 (2019).

51. Boatner, L. *et al.* CIAA: Integrated Proteomics and Structural Modeling for Understanding Cysteine Reactivity with Iodoacetamide Alkyne. Preprint at https://doi.org/10.26434/chemrxiv-2025-tm8ch (2025).

52. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **4**, 17 (2012).

53. Perri, M. J. & Weber, S. H. Web-Based Job Submission Interface for the GAMESS Computational Chemistry Program. *J. Chem. Educ.* **91**, 2206–2208 (2014).

54. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

55. Chemical Computing Group ULC. Molecular Operating Environment (MOE), 2024.06. Chemical Computing Group ULC (2024).

56. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Andrew Ecker
D28581F12236454...
Author Signature

6/24/2025
Date