Prebiotic and Unnatural Amino Acids and Peptides

By

Dean Raymond Jarois

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN—MADISON

2026

Date of final oral examination: 01/13/2026

The dissertation is approved by the following members of the Final Oral Committee:

Samuel H. Gellman, Professor, Chemistry

Robert J. McMahon, Professor, Chemistry

David A. Baum, Professor, Botany

Zoe Todd, Assistant Professor, Chemistry and Astronomy

Acknowledgements

First and foremost, I'd like to remind you, dear reader, that science cannot be adequately conducted in a vacuum (though my friends in astrochemistry might disagree). None of this would be possible without the support of the **American taxpayers** as well as the **huge list of scientists who I cite throughout this thesis** for their hard work that enabled these discoveries. Support for this work was provided by the **National Science Foundation**, including the **Graduate Research Fellowship Program**, and the **National Institutes of Health**.

Entire UW—Madison Chemistry Department: It has been a pleasure working in such a collaborative environment. Together we show the world that building each other up is much more productive than being internally competitive. You are all awesome! I especially benefitted from wisdom from friends in the Weix Lab, Stahl Lab, and Boros Lab.

My Committee Members, whom I chose because I respect them as scientists: **Professors Samuel Gellman, Robert McMahon, David Baum, and Zoe Todd**

Chemical Instrumentation Center Scientists: Charles Fry, Heike Hofstetter, Cathy Clewett, Gabi Carosio, Martha Vestling, and Ilia Guzei

Department Staff and Administration, especially **Francisca Jofre, Pat Egan, Char Horsfall, Jeff Nielsen, Ashley Petersen, Aaron Smith, and Marc Willadsen**

Physical Plant Staff who keep our labs clean and operable

Collaborators from UCLA: Prof. Tamir Gonen, Dr. Johan Unge, Dr. Jieye Li

Sam—doubtless there were numerous times I've driven you crazy throughout these last 5 ½ years. Thanks for putting up with me—I am better off because of it. Grad school is humbling, and I believe that I am leaving your lab not only as a better scientist and a more strategic critical thinker, but also a more patient person, a more efficient communicator, and so much more. In addition to the science you instilled into me, your writing feedback taught me the concepts ~~which~~ that I probably should have learned in English class, and your advice to "be relentless" when addressing the needs of my family gave me courage that I wish I had long ago. However, some of my most treasured interactions with you were when I feared that you were going to ask me for an update on an experiment I forgot to do, but instead you wanted to talk about Moses or share a Psalm with me in Hebrew.

Past and Present Gellman Lab Members: Dr. Victor Outlaw, Dr. Zhen Yu, Dr. Kyle Brown, Dr. Ariel Kuhn, Dr. Ash Sarkar, Dr. Joseph Swain, Dr. Adam Kleman, Dr. Brian Cary, Dr. Lei Liu, Dr. Xinyu 'Jenny' Liu, Dr. Ruslan Gibadullin, Dr. Jiani Niu, Dr. Mary Katherine 'MK' Andrews, Dr. Tae Kim, Dr. John Mannone, Dr. Rylie Morris, Dr. Philip Lampkin, Dr. Ruiwen 'Irian' Xu, Yareslie Cruz Rivera, Lauren Tran, Aevi van der Stok, Theodore Gierszal, Eden Xu, Justin Suter, Yucheng Jiang, Charlie Roberts, Rui Wang, Quan Nguyen, Bret Perkins … where do I begin? I am going to miss you all. I can't lie, sometimes the drama felt like the television show *Survivor*.

Most of the time the camaraderie was wonderful. Nothing beat increasing lab morale by posting a new meme on the meme board, ordering bubble tea on a Friday afternoon, or hanging out at the Terrace. Thanks for everything you taught me, and I ask that future lab members are patient with Aevi, Eden, and Charlie who will be fixing your HPLCs (I am NOT going to miss that job!).

Speaking of which, thanks to **Agilent Technologies** especially **Carolyn Steele** for building HPLCs that are superior in every way to those of your competitor that rhymes with 'Daughters.'

Thank you, **Irian Xu**, who joined the Gellman lab at the same time as me but completed all of her requirements just a little bit ahead of me so that I could burden her for details.

Dr. Ariel Kuhn decided she was too good for us this past year, but I enjoyed all the laughs (and frequent eye rolls) sharing a desk with her.

Eden—you are smart and strong. Keep up the hard work. Feel free to keep texting me Spotify links to classical music (as long as you keep your Mozart to yourself). I hope Mendelssohn and Bach continue to grow on you, just as Mahler is growing on me.

Aevi, Yucheng, and Rui—keep up the good work on those droplets! We are onto something really cool.

Aevi and Lauren—You two are next! I am wishing you both the best of luck as you continue your scientific journey.

Justin—we'll probably all be working for you someday. Remember to always stand up straight with your shoulders back.

Lars and Charlie—who taught me how to be a mentor—all the best in your future studies.

Cathy Clewett—I have always cherished your counsel and I look up to you as such a great and humble scientific role model.

To my 'brother from another mother,' **Yngve Guttormsen**: thanks for all the science brainstorming sessions, for helping me stay sane during this insane time, for threatening to jiu-jitsu me if I didn't hurry up and finish this thesis, and for being a fine uncle to my kids.

Dad and Mom Terjesen—thank you for your support throughout this process. Mom—your positivity and optimism is contagious. Dad—your discussions with me about preaching are easily adapted in preparing to give an oral thesis defense, or any other talk for that matter. Preparation and effective delivery of important material is a such an amazing skill. Thanks for the wisdom you imparted on me.

Dad, Mom, and Mamma: You prepared me for this from a young age. Lessons taught at home stick more than those taught at school. I cannot forget when you, dad, taught 7-year-old me that steam expands 1700x and therefore how energy like heat can be used to do work. If only university physics professors could explain concepts half as well as you! Or how you, mom, taught me all about the history of x-rays starting with Roentgen. Or mamma—all the encouragement I received from you and papa—doing baking soda and vinegar experiments in your kitchen or you guys taking me to meet professors at Michigan State. I love you all and could not have done any of this without you.

I will never forget my middle-school science teacher, **Mrs. Castelli**, who helped me discover this knack that I had for science, which became an interest, and then an obsession, and now a career. It is in this way that science teachers, broadly, make *massive* contributions to scientific progress. Thank you.

Joel Raymond, Celeste Joyce, and future baby Jarois—maybe one day you'll read this. When you do, keep in mind that it took an incredible amount of perseverance to finish this program. And as hard as that was, you kids kept me focused on our future and that kept me going. Joel—I hope you remember getting to come to the lab with me and see the machines and draw lions on my fume hood. Or Celeste, when you stole the attention at my graduation when you yelled 'It's Dada!' as I walked across the front of the hall.

Juliana… this accomplishment is just as much yours as it is mine. Most of my peers remark that they could not imagine completing graduate school with children. Having a supportive spouse is what makes all the difference. We've both had to make a ton of sacrifices to get to this point… and here we are. I couldn't imagine a better wife nor could I wish for a better mother for my children. I love you and I can't wait to see what the next adventure brings. What therefore God hath joined together, let not man put asunder.

Lastly, I acknowledge **Jesus of Nazareth** as my Savior.

General Abstract

Prebiotic or unnatural amino acids and peptides can help inform our understanding of the origin of life as well as the structure and function of biomolecules. In chapter 1, the plausible prebiotic reactivity of the ammonium ion is investigated. The reaction of pyruvate and glyoxylate in the presence of ammonium resulted in numerous organic compounds containing new carbon-carbon bonds and nitrogen heteroatoms. Chapter 2 describes the aza-Cannizzaro reaction, a mechanism by which glyoxylate was converted into glycine that parallels the well-known Cannizzaro reaction. Liquid-liquid phase separation of short heterochiral peptides with hexametaphosphate is described in Chapter 3. Factors influencing coacervation included heterochirality, length, and spacing of heterochiral amino acid residues. Lastly, Chapter 4 describes the synthesis of terphenyl amino acids and their oligomers. The secondary structure of these terphenyl oligomers was investigated.

Table of Contents

Table of Figures

Table of Schemes

Table of Tables

Abbreviations

ALS	Amyotrophic lateral sclerosis
ATP	Adenosine triphosphate
BBFO	Broadband fluorine observe (probe)
Bpin	Boronic pinacol ester
C18	Octadecyl silane
CD	Circular dichroism
$CDCl_3$	Deuterated chloroform
COMU	(1-cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate
Cryo-EM	Cryogenic electron microscopy
2-CTC	2-chlorotrityl chloride
D_2O	Deuterium oxide
DCl	Detuerated hydrochloric acid
DCM	Dichloromethane
DIC	Diisopropylcarbodiimide
di-DHKG	Di-dehydroketoglutarate
DMF	*N,N*-dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DOSY	Diffusion ordered spectroscopy
eq	Equivalent(s)
ESI	Electrospray ionization
Fmoc	Fluorenylmethyloxycarbonyl protecting group
FmocOSu	Fmoc succinimide
FT-ICR	Fourier-transform ion cyclotron resonance (mass spectrometry)
FUS	Fused in Sarcoma protein
GC-MS	Gas chromatography – mass spectrometry
HATU	Hexafluorophosphate Azabenzotriazole Tetramethyl Uronium

HBTU	Hexafluorophosphate Benzotriazole Tetramethyl Uronium
HILIC	Hydrophilic interaction chromatography
HMP	Hexametaphosphate
HOD	Partially-deuterated water
HPLC	High performance liquid chromatography
hr	Hours
HRMS	High resolution mass spectrometry
IDR	Intrinsically disordered region
LC-MS	Liquid chromatography - mass spectrometry
LLPS	Liquid-liquid phase separation
m/z	Mass-to-charge ratio
MeCN	Acetonitrile
MicroED	Microcrystal electron diffraction
MLO	Membraneless organelle
MRE	Mean residue ellipticity
MS	Mass spectrometry
NBS	*N*-bromosuccinimide
NMM	N-methylmorpholine
NMR	Nuclear magnetic resonance
nOe	Nuclear Overhauser effect
OD_{600}	Optical density at 600 nm
OMPD	2-methyl-4-oxopene-2-enedioate
OoL	Origin of Life
Oxyma	Ethyl cyanohydroxyiminoacetate
polyP	Polyphosphate
ppm	Parts per million
presat	Presaturation (NMR solvent suppression technique)
Q-TOF	Quadrupole – time of flight

RNA	Ribonucleic acid
RP	Reverse phase
rTCA	Reductive tricarboxylic Acid [cycle]
TCA	Tricarboxylic acid [cylce]
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TMP	Trimetaphosphate
TMS	Tetramethylsilane
TPP	Tripolyphosphate
TRIS	tris(hydroxymethyl)aminomethane
TSP	3-(trimethylsilyl)propionic-2,2,3,3-d4 acid, sodium salt

Chapter 1

Prebiotic Reactivity with the Ammonium Ion

1.1. Abstract: Ammonium Effects on Prebiotic Ketoacid Reactivity

Living organisms require a metabolism to build and break down biological molecules, but extant metabolisms depend heavily on enzymes. Progress in prebiotic chemistry depends on finding similar reaction pathways that operate non-enzymatically. Recent research highlights how pyruvate and glyoxylate can form key metabolic intermediates. Here we highlight how the addition of ammonium ions into this pathway can result in numerous new compounds and act as a primordial pathway for nitrogen assimilation.

1.2. Introduction to Prebiotic Chemistry

> How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry? The preliminary answer . . . can be summarized as follows: The obvious inability of present-day physics and chemistry to account for such events is no reason at all for doubting that they can be accounted for by those sciences.
>
> - Erwin Schrödinger. *What is Life?* **1944**

Prebiotic chemistry is the chemists' attempt to account for the events enabling and leading up to life's origins by studying the molecules, reactions and pathways that could have occurred on the primitive Earth.[1] These molecules and reactions of interest are mainly organic and could have occurred naturally on Earth or in other planetary environments. Supposing that the universe is not endowed with a specific cosmic teleonomy that favors biological systems, reactions that are both productive and nonproductive for the Origin of Life (OoL) must be considered even though prebiotic chemistry is necessarily concerned with the OoL.[2]

1.2.1. The Goal of Prebiotic Chemistry

Life is chemistry, albeit very complex chemistry. The reactions taking place in biological systems are far from thermodynamic or kinetic favorability and only work because of the constant input of energy and the assistance of astonishingly complex biomolecules. At some point, one leaves the realm of chemistry and enters the realm of biology. Prebiotic chemistry is the attempt to bridge this chasm.[3]

It must be clearly acknowledged that the exact circumstances surrounding the OoL are unknowable; no amount of careful science will ever be able to reconstruct the events by which chemical systems became biological systems. As captured in the quote by Erwin Schrödinger at the beginning of this chapter, this inability to know the OoL with any certainty does not mean that it could not have happened naturally. However, when one approaches prebiotic chemistry with the goal of ‘knowing’ the OoL, he or she will be quickly disappointed. The value of science does not rest in its ability to prove the way something happened, but rather the ability of a theory to predict future outcomes.[4,5] Investigating molecules and reactions that could have led up to the OoL should be understood as a fundamental scientific pursuit. It will help biological scientists understand what life is and how it operates.[6] Understanding what molecules might have been useful for the OoL can also help planetary investigators direct their searches for life on other planets.[7]

1.2.2. The History and Philosophy of Prebiotic Chemistry

As a concept, prebiotic chemistry goes back to the early 20th century when Oparin and Haldane proposed that a reducing atmosphere (although the nature of Earth's early atmosphere is now heavily disputed) and a 'primordial soup' containing relevant organic compounds could have given rise to life.[8,9] This 'heterotroph hypothesis' that life depended on simple abiotic chemical components coming together provided a testable framework for future studies in prebiotic chemistry.

One such study was the famous Miller-Urey experiment that demonstrated biologically-useful amino acids could be produced from electric discharge in a mixture of water vapor, hydrogen, ammonia, and methane.[10,11] When Miller and Urey first conducted these experiments in the 1950s, molecules were detected with paper chromatography and the nature of the primitive atmosphere was mere speculation. More recent experiments, including re-analysis of Miller-Urey reaction mixtures with modern analytical techniques, variants of the Miller-Urey experiment with different atmospheres and variants meant to simulate volcanic conditions, all continued to result in robust amino acid formation.[9,12–15] Not long after the Miller-Urey experiments were originally published, Oró and Kimball demonstrated that hot hydrogen cyanide can polymerize to adenine.[16,17] This foundational work demonstrated for the first time that biomonomers could indeed be made abiotically, and countless more modern studies, like the ones in this thesis, continue to expand on this idea.

In the decades following Miller and Oró, work continued on prebiotic synthesis, but with the attention turning away from simple homogeneous mixtures. A.G. Cairns-Smith, Leslie Orgel, and Günter Wächterhäuser pioneered some of the fundamental hypotheses that challenged the idea that ‘prebiotic soups’ were uniquely responsible for the OoL and encouraged their peers to consider heterogeneous environments such as minerals, vents, and interfaces.[18–22]

While the idea that nucleic acids may have been involved in the OoL existed since the 1960s,[23,24] the idea that they may have played a catalytic role (and still do in extant biology) was not fully considered until the discovery of ribozymes, RNAs with enzyme-like activity, in 1982.[25,26] This enabled OoL researchers to consider the possibility that RNA could have served simultaneously as an information-storage polymer and a functional catalyst, supporting both replication and evolution before protein-based enzymes evolved. The new understanding that nucleic acids can have both genetic and catalytic functions led to the ‘RNA-World Hypothesis’, which remains under active consideration.[1,27–31]

Modern prebiotic chemistry challenges us to expand our focus to more broadly understand how molecules fit together, not as isolated reactions, but rather as small steps in a much larger network of reactivity. This ‘systems chemistry’ approach, in which emergent properties that derive from complex systems of reactions are studied, advances us from chemistry towards biology because ‘life’ depends not on single reactions in isolation, but on coordinated webs of reactivity where the output of one reaction becomes the input for another.[3,32] For example, even though nucleic acid monomers—adenine, guanine, cytosine, and uracil—can be made individually using prebiotically-plausible reactions, assembly into RNA is much more complicated. Powner, Gerland and Sutherland set aside traditional ‘stepwise’ reactions and addressed this problem by proposing a plausible route to activated ribonucleotides, ready to assemble into RNA.[33]

John Sutherland continued to look at systems chemistry to more recently show how a HCN-based network can give rise to amino acids, lipids, and nucleotides.[34] Jack Szostak has conducted work on fatty acid vesicles as potential protocells that can grow, divide and encapsulate genetic information.[35] These approaches are better for investigating vast chemical networks towards an integrated hypothesis of the chemical OoL. Future progress will be assessed not by looking at isolated steps, but rather by identifying pieces that fit into a 'bigger picture.'[1,6,36]

1.2.3. Prebiotic Plausibility

Prebiotic chemistry should encompass only molecules and reactions that could have occurred on the primitive Earth. This means that many of the reactions and techniques used to make molecules today are off-limits. While synthetic chemists today conduct reactions in anhydrous organic solvents, at temperatures ranging from -80°C to many degrees past the boiling point of the solvent, extreme pH ranges, in the presence of exotic or designer catalysts, and under inert atmosphere, researchers conducting prebiotic experiments only have access to water, moderate temperatures and pH values, simple catalysts, and atmospheric gases. Assembling a toolbox of reactions that operate under these constraints enables progress in prebiotic chemistry.

In this chapter, the investigations leading to and including a new prebiotically-plausible chemical reaction mechanism are described. The criteria for prebiotic relevance are briefly assessed, as well as how the new reaction might fit into the 'big picture' of prebiotic chemistry.

1.3. A Prebiotic Metabolism

The chemistry of life involves not only molecules but also the chemical processes enabling biology to function. One cannot investigate the intersection of chemical systems and prebiotic chemistry without considering how biological metabolisms may have emerged.[37]

One can begin to consider how a primordial metabolism might have worked through by tracing back the evolutionary trajectory from modern biology. However, the mechanisms that life uses to build biological molecules can be much different from the pathways proposed by prebiotic chemists. The nucleobase work of Oró provides a relevant example.[16,17] Purines today are made from amino acids; never hydrogen cyanide! A key issue is that the catalysts that carry out many metabolic reactions today—enzymes—were likely inaccessible before biology. Francis Crick introduced the 'Frozen Accident Hypothesis,' which is the idea that the genetic code arose by historical accident but was frozen in time.[23] Some OoL researchers argue that the *molecules themselves* that extant biology uses may also be types of 'Frozen Accidents.' In other words, the molecules may have already been present before the OoL, found a role in protobiology, and were carried over in extant life.[38,39]

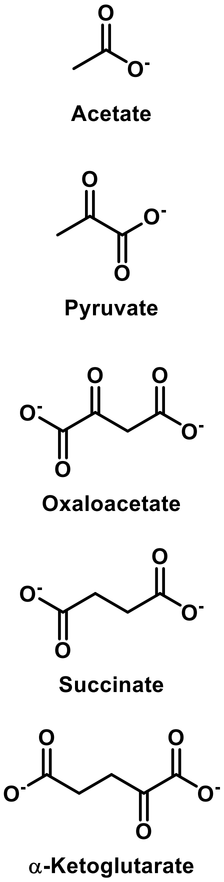


Figure 1.1 Five universal metabolic precursors.

Among the molecules that are frozen into biology are the five universal metabolic precursors (**Figure 1.1**), identified by Smith and Morowitz, which play a central role in all metabolic pathways:[37,40] acetate (precursor to lipids and terpenoids), pyruvate (precursor to sugars and amino acids), oxaloacetate (precursor to amino acids and pyrimidines), succinate (precursor to many cofactors), and α-ketoglutarate (precursor to amino acids). Due to the key roles played by the universal metabolic precursor molecules in the tricarboxylic acid cycle (TCA cycle, Krebs cycle, or citric acid cycle), as well as the central role of the TCA cycle in biochemistry, the TCA cycle and especially the autocatalytic rTCA (reductive TCA) cycle are seen as likely candidates for early prebiotic metabolisms.[21,40,41]

There are several reasons why the rTCA cycle is viewed as more prebiotically-plausible than the oxidative TCA cycle. Even though the typical oxidative TCA cycle is catabolic and has only one C-C bond-forming step, while the rTCA cycle is anabolic with three C-C bond forming steps (**Figure 1.2**), the rTCA is able to fix carbon dioxide and can operate in the absence of molecular oxygen. The rTCA cycle also has fewer ATP-dependent steps than the oxidative TCA cycle, making it better suited prior to the standardization of a biological energy currency. Some of the transformations in the rTCA cycle can also be carried out with nonenzymatic catalysts, increasing their prebiotic relevance.[42]

One problem with proposing the rTCA cycle as an early metabolic pathway is that the C-C bond-forming steps are chemically challenging. Much of the C-C bond chemistry in the rTCA cycle proceeds through enol(ate) intermediates, but the proton at the α-carbon atom on many rTCA carboxylates is not sufficiently acidic for enolization to occur outside of an enzyme. Stubbs and coworkers have overcome this limitation by demonstrating an alternative pathway by which the rTCA cycle intermediates might have been produced prebiotically. Pyruvate and glyoxylate react spontaneously in warm aqueous buffer to form a series of α-ketoacids (**Figure 1.3**), any of which may be oxidatively decarboxylated to form a corresponding carboxylate intermediate of the rTCA cycle.[43]

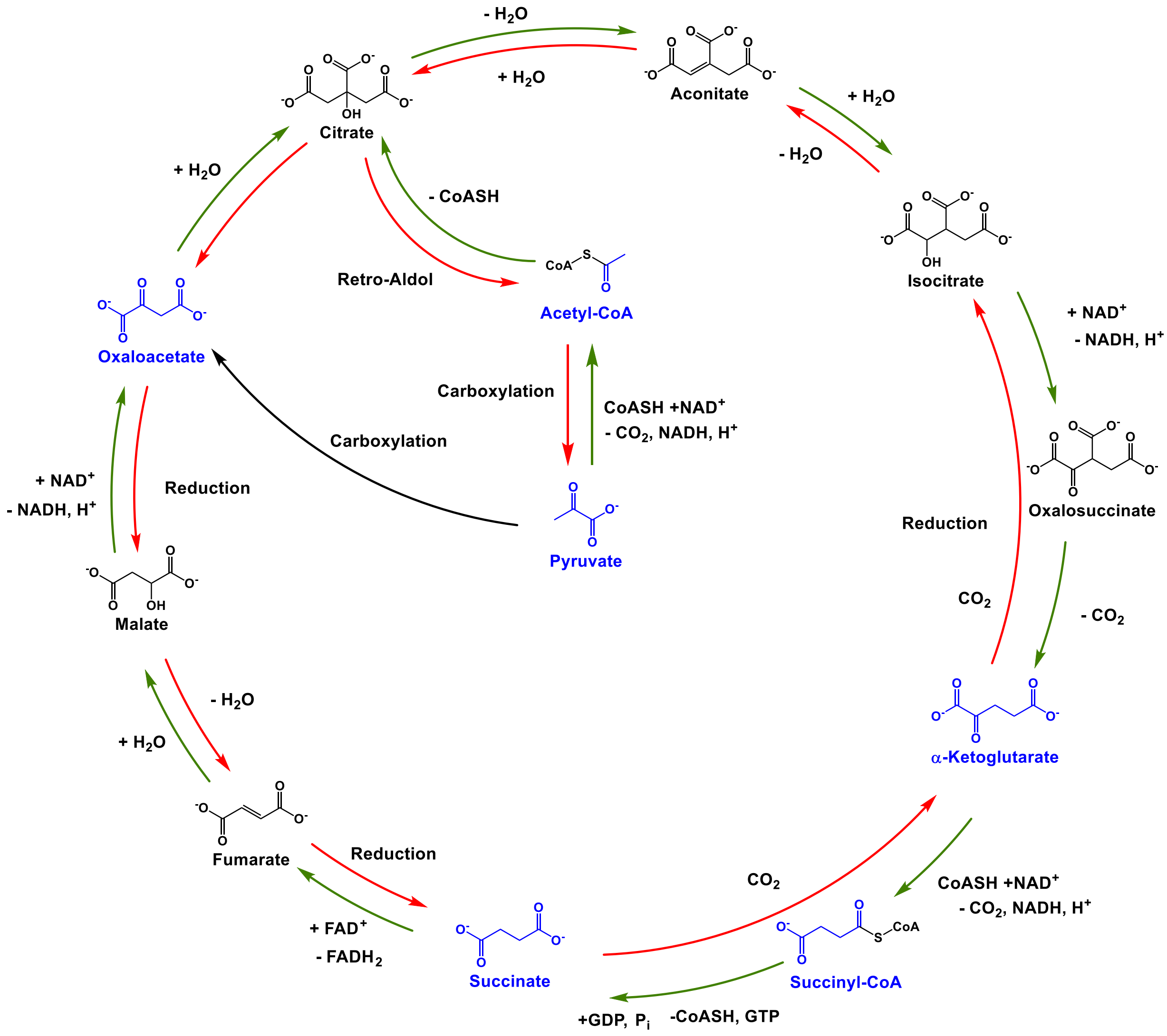


Figure 1.2 The TCA cycle (green arrows) and rTCA cycle (red arrows) with the five universal metabolic precursors depicted in blue. Note how various carboxylation reactions in the rTCA cycle are able to fix carbon dioxide into more complex carbon skeletons. Other steps include reduction, hydration, and dehydration. In modern biology, all rTCA steps are enzyme-catalyzed, but non-enzymatic prebiotic versions of some steps have been discovered.

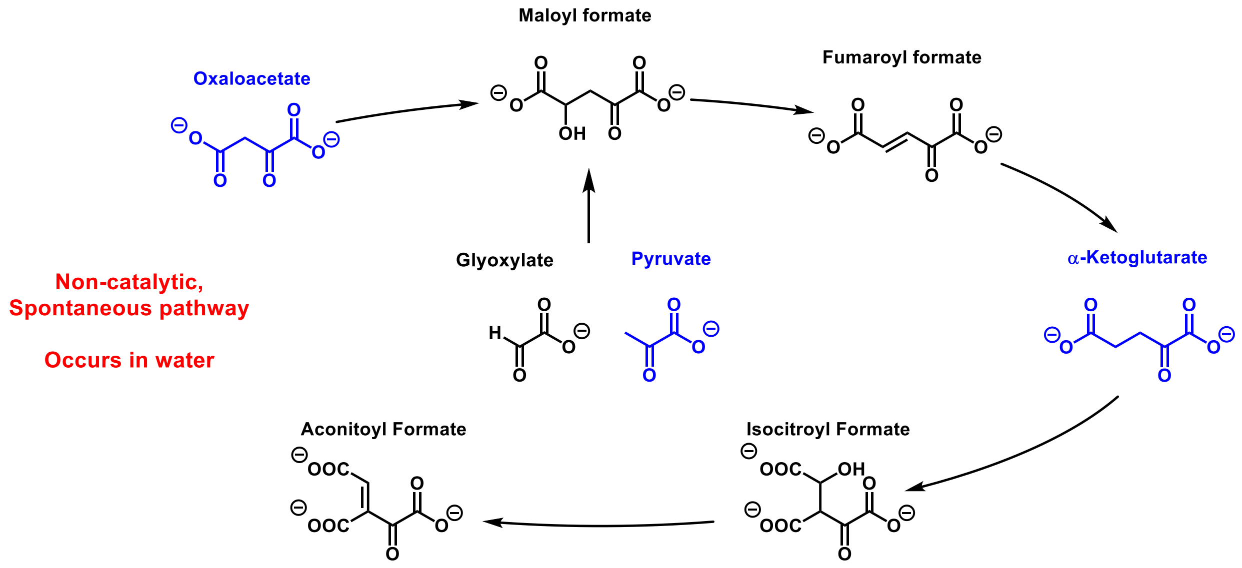


Figure 1.3 The α-ketoacid pathway demonstrated by Stubbs et al. may be initiated either by oxaloacetate or glyoxylate and pyruvate.[43] Both result in a series of metabolic ketoacid precursors, any of which can be decarboxylated to form (r)TCA cycle intermediates.

1.3.1. Iminium Ion Catalysis of Aldol Reactions

It is well-precedented that simple amines can catalyze aldol reactions via several different activated intermediates.[44,45] The most relevant strategy is 'iminium catalysis,' by which amines condense onto carbonyls to form imines, which are typically protonated at neutral conditions (pKa≈7). Because iminium ions are both more electrophilic than their aldehyde/ketone analogues and able to react in the absence of strong bases, iminium catalysis strategies have been used to form carbon-carbon bonds for over one hundred years.[46–48] In some species, imines may tautomerize to form nucleophilic enamines, allowing for a double-activation mechanism.[46] Iminium catalysis has also been employed as a strategy to catalyze Diels-Alder reactions, conjugate additions, and numerous other reactions.[44]

Drawing on our knowledge that aldol chemistry can be catalyzed by amines that activate aldehydes and ketones by forming electrophilic iminium ions,[44,45] and that these amines can also be scaffolded onto a short peptide that adopts a stable secondary structure (a 'foldamer', **Figure 1.4**),[49–52] we set out to determine whether iminium catalysis could affect the rates or product distribution of the ketoacid pathway (**Figure 1.3**) described by Stubbs et al.[43] Doing so would introduce a framework upon which more complex catalysts (enzymes) could later evolve.

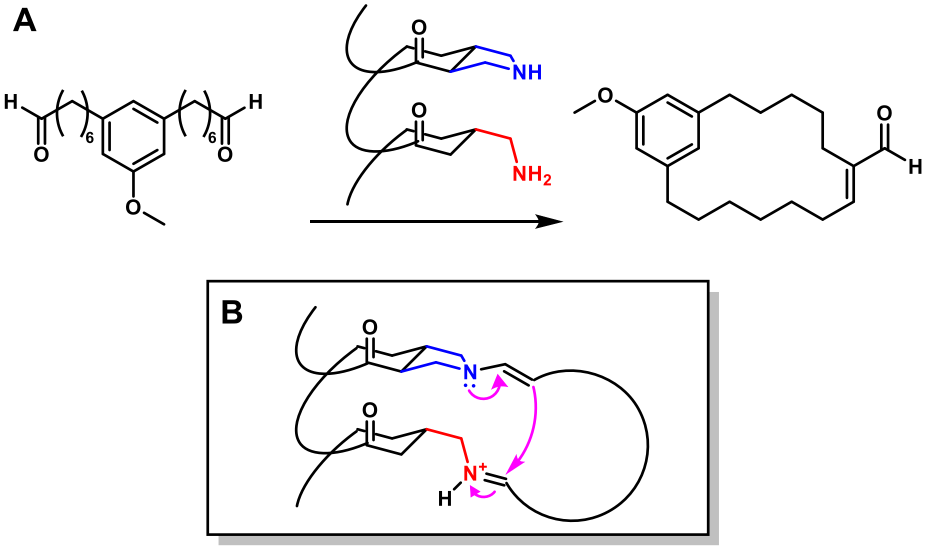


Figure 1.4 Example of an iminium/enamine catalyzed reaction on a foldamer backbone from the Gellman lab.[51] (A) A natural product precursor can undergo a challenging aldol condensation into a macrocycle. The mechanistic step allowing this transformation involves (B) the reaction between an enamine and an iminium, both formed after the starting aldehydes condense onto the foldamer scaffold. We imagine that early enzymes could have resulted from simple catalysts like these.

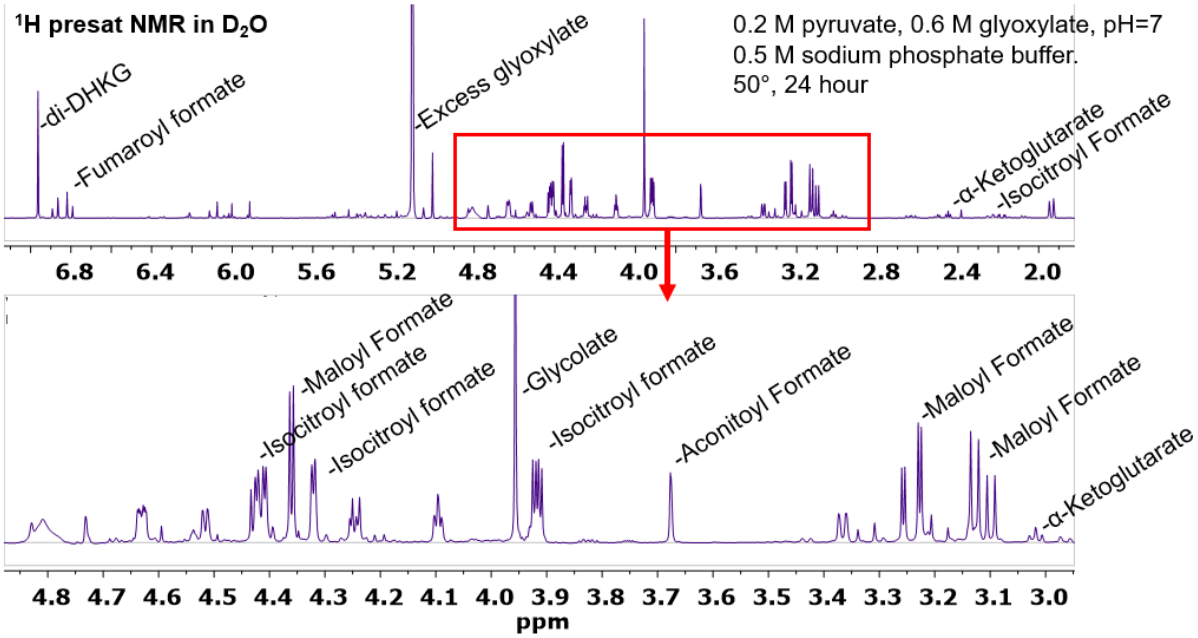


Figure 1.5 ^{1}H NMR with solvent suppression showing reaction mixture resulting from pyruvate and glyoxylate in sodium phosphate pH 7 buffer. Note all major peaks are identifiable and correspond to compounds in **Figure 1.3.**

After replicating the experiments by Stubbs and Springsteen and verifying that all resulting ketoacids are visible and identifiable by proton NMR (**Figure 1.5**), we selected 12 different amines to add to the reaction mixture at 20 mol%. These included ammonium hydroxide, hydrazine, primary aliphatic amines (methylamine, propylamine), secondary amines (diethylamine, 2-methylaziridine, morpholine, pyrrolidine), proline, imidazole, and several ‘mixed’ diamines (*N*-methylethylenediamine, *N*-methyl-1,3-propanediamine). A mixture of pyrrolidine and propylamine was also evaluated. In most cases (except for the mixture containing ammonium hydroxide) the presence of the amine did not seem to have any measurable effect on the reaction. No apparent changes in product identity or distribution were observed (**Figure 1.6)**.

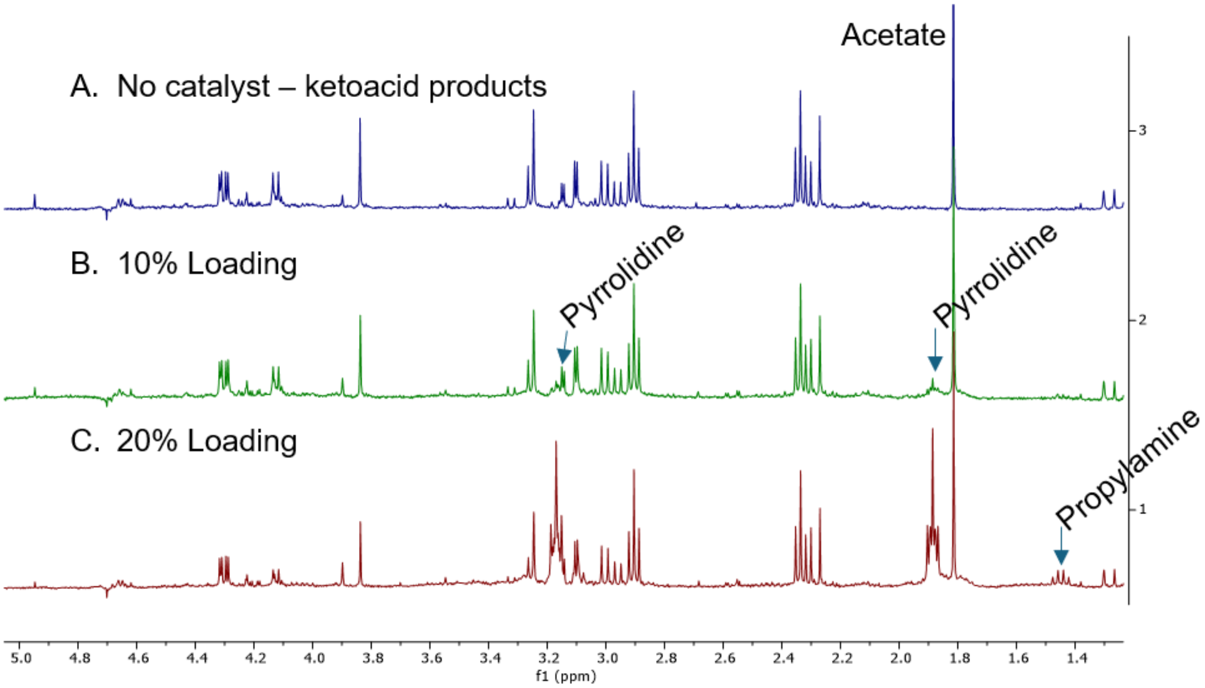


Figure 1.6 400 MHz ^{1}H presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard)spectra acquired of α-ketoacid pathway product mixtures. Each reaction originally contained 1 mmol pyruvate and 3 mmol glyoxylate in pH 7 500 mM sodium phosphate buffer and was incubated at 50°C for 72 hours. While (A) contains only the starting materials listed above, (B) and (C) contain 10% and 20% (respectively) of pyrrolidine (signals at 1.9 and 3.15 ppm) and propylamine (signals at 1.45 and (broad) 3.15 ppm; third signal upfield) for possible iminium catalysis. Note that products and distribution remain unchanged.

1.3.2. Ammonium Effects on Product Distribution

When ammonium hydroxide (35 mol%) was incubated (50°C, 24 hours) with 1 mM pyruvate and 3 mM glyoxylate in sodium phosphate buffer (500 mM, pH 7), several small new signals were observed, and the product distribution appeared slightly different from the reaction mixture in the absence of ammonium hydroxide (**Figure 1.7**).

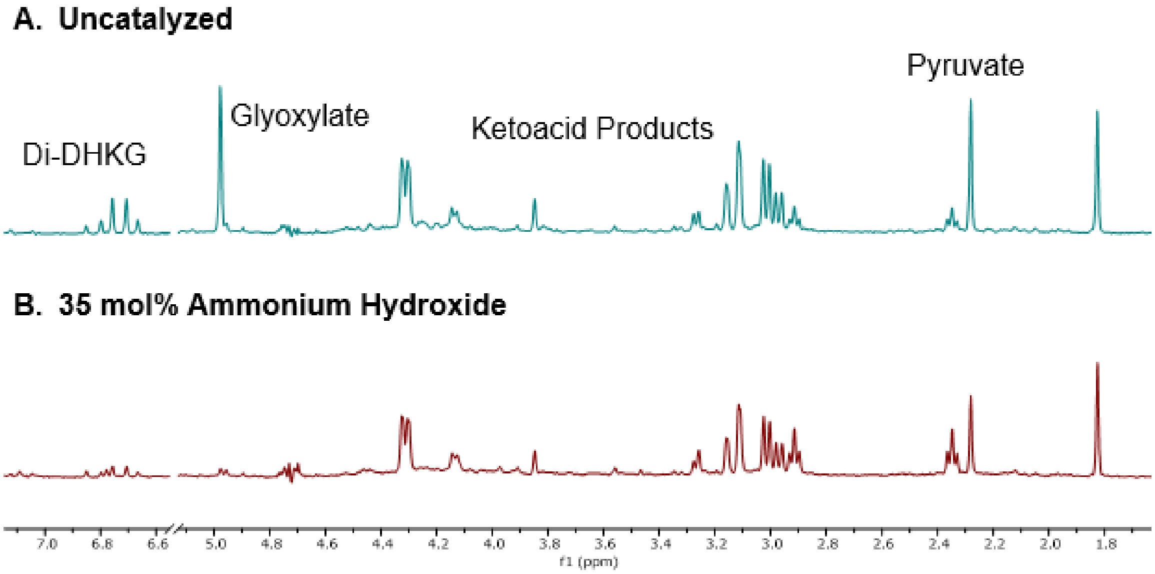


Figure 1.7 400 MHz ^1H presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard) spectra acquired of α-ketoacid pathway product mixtures. Each reaction originally contained 1 mmol pyruvate and 3 mmol glyoxylate in pH 7 500 mM sodium phosphate buffer and was incubated at 50°C for 24 hours. While (A) contains only the starting materials listed above, (B) contains 35 mol% ammonium hydroxide for possible iminium catalysis. The ammonium-catalyzed reaction does not contain di-dehydroketoglutarate (di-DHKG), the starting materials are more consumed, and the products appear to be in a different ratio than the uncatalyzed version.

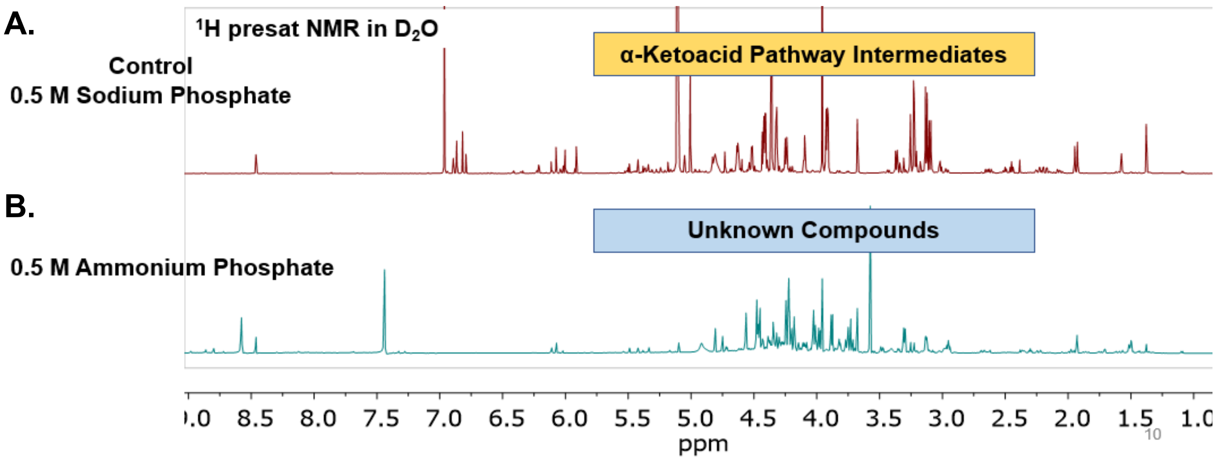


Figure 1.8 ^{1}H NMR with solvent suppression showing reaction mixture resulting from pyruvate and glyoxylate in (A) 500 mM sodium phosphate buffer and (B) 500 mM ammonium phosphate buffer at pH 7 when incubated for 24 hours at 50°C. While all compounds are identifiable in reaction (A) (see **Figure 1.5**), the origins of signals in (B) are mostly unknown.

This inspired us to raise the concentration of ammonia by switching the identity of the buffer entirely. While Springsteen and Stubbs conducted their experiments in 500 mM sodium phosphate (corrected to pH 7 with NaOH), we tried a variation with 500 mM ammonium phosphate (corrected to pH 7 with NH_4OH). The new buffer very quickly resulted in a brown reaction mixture containing numerous unidentified NMR signals (**Figure 1.8**). The behavior of this reaction was remarkably different from the sodium-buffered reaction: products were formed rapidly in the presence of ammonium hydroxide, even at room temperature (the sodium-buffered reaction requires 50°C for a reasonable reaction rate). Different products were produced in the sodium-buffered and ammonium-buffered reactions at every pH we tested from 2 to 11 (**Figures 1.9-1.10**).

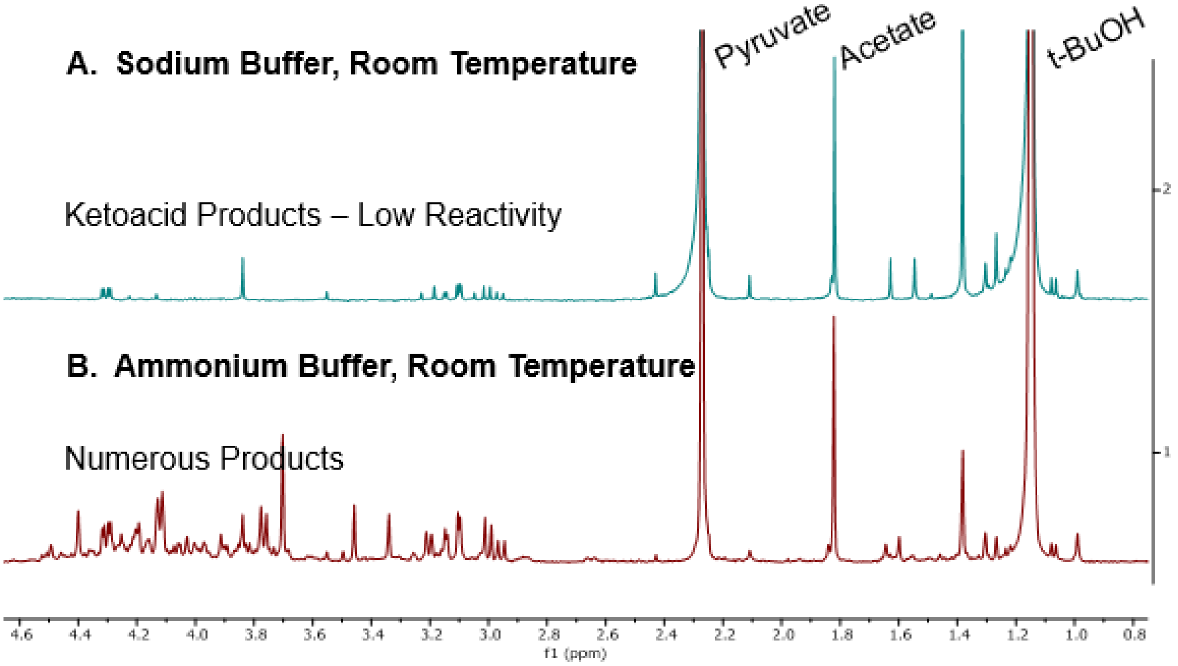


Figure 1.9 400 MHz ^{1}H presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR (500 MHz; pulse sequence=zgpr30) spectra acquired of pyruvate+glyoxylate product mixtures. Each reaction originally contained 1 mmol pyruvate and 3 mmol glyoxylate in pH 7 500 mM phosphate buffer and **kept at room temperature** for 24 hours. (A) contains sodium phosphate buffer and reactivity at room temperature is sluggish. While some ketoacid products are still identified in (B) with ammonium hydroxide, the alternative pathway resulting in unidentified products predominates. t-Butanol was added as an internal standard.

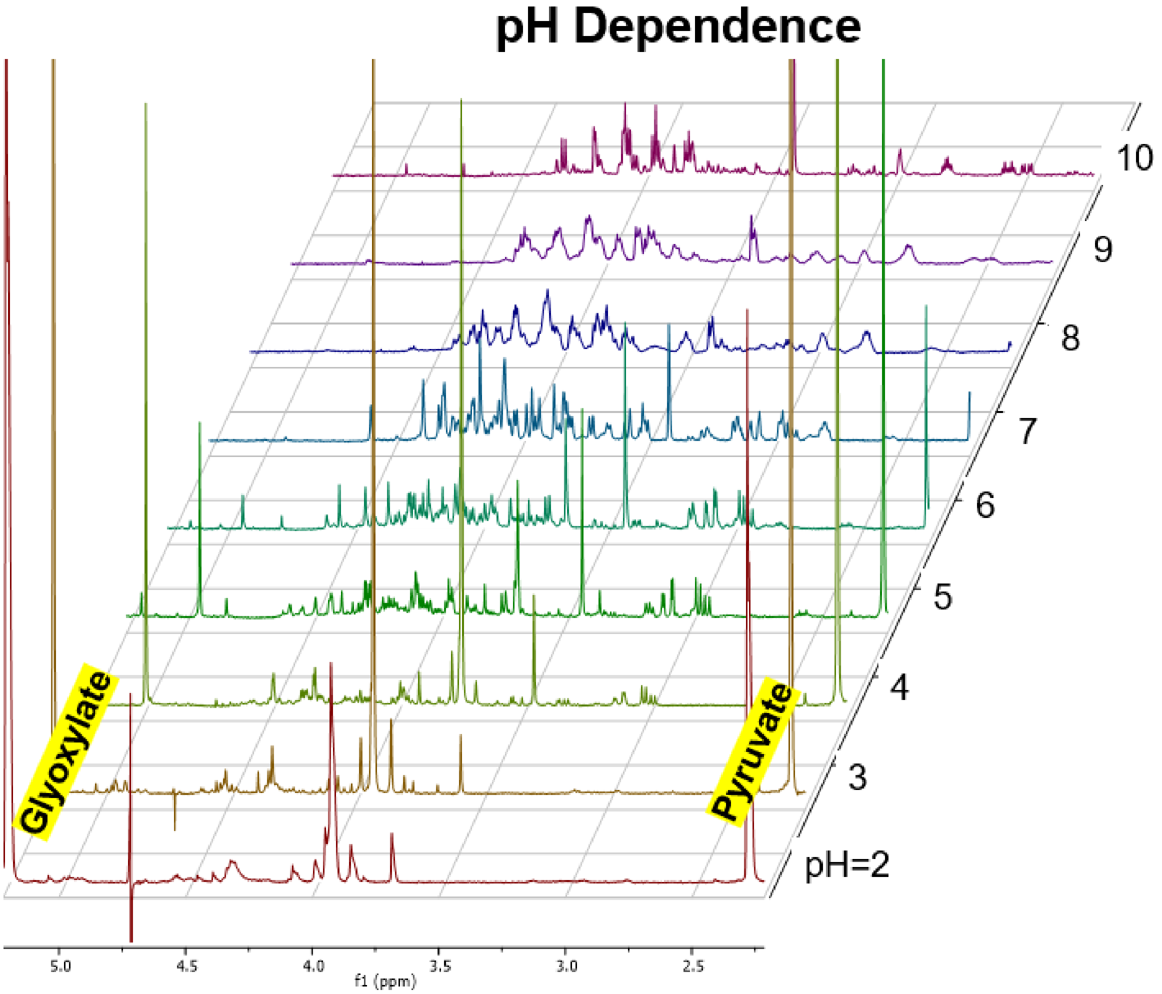


Figure 1.10 400 MHz ^{1}H presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard) spectra acquired of ammonium-containing glyoxylate+pyruvate product mixtures. Each reaction originally contained 1 mmol pyruvate and 3 mmol glyoxylate in 500 mM phosphate buffer and incubated at 50°C for 24 hours. Reactivity was observed at every pH from 2 to 10, however, number, identity, and distribution of products vary depending on pH. Starting materials were completely consumed at pH > 7.

1.3.3. Attempts to Analyze Reaction Products

We sought to identify the new products observed via proton NMR in the ammonium-buffered reaction. Prior experience in the interpretation of NMR spectra caused us to initially speculate that these molecules were similar to carbohydrates in nature based on the chemical shifts and coupling patterns. The complex coupling behavior observed via NMR strongly suggests that new carbon-carbon bonds were generated. However, other methods of analysis were problematic. GC-MS was not possible because products of this reaction were soluble in only aqueous media. The high concentration of phosphate and highly-charged products made LC-MS difficult. Attempts to divert the phosphate away from the spectrometer by discarding the first column volume of the run also led to the diversion of the compounds that were poorly retained on a C18 column. Derivatization with acetic anhydride in pyridine did not seem to yield different compounds. The inability of this mixture to undergo acetylation, in addition to ^{13}C NMR evidence (**Figure 1.11**), suggests carboxylates are abundant in the product mixture.

In an effort to gain insight on the products from the ammonium-buffered reaction, we replaced pyruvate with monofluoropyruvate (which is very toxic) or trifluoropyruvate. We hoped that ^{19}F NMR would help us identify the resulting products. When each of the pyruvate derivatives was combined with glyoxylate in ammonium phosphate buffer and incubated, proton NMR suggested that simpler reaction mixtures were obtained relative to the reaction with pyruvate itself (**1.12**). More surprisingly, no new signals were observed via ^{19}F NMR, suggesting that neither monofluoropyruvate nor trifluoropyruvate (which is nonenolizable) reacted with glyoxylate (**Figure 1.13**).

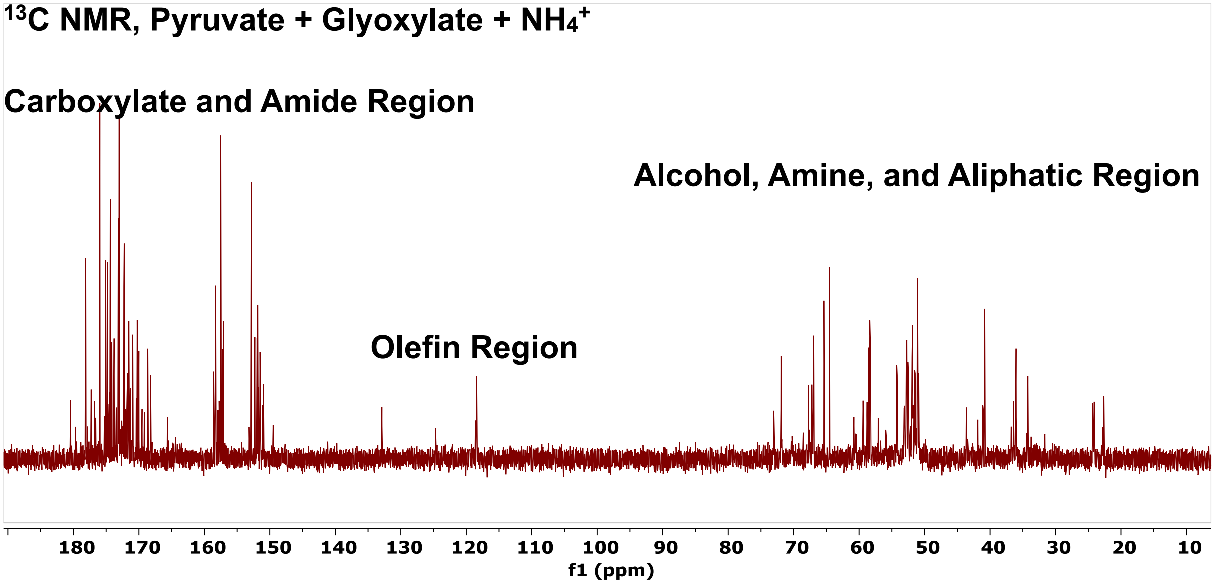


Figure 1.11 400 MHz ^{13}C NMR spectrum acquired of ammonium-containing glyoxylate+pyruvate product mixture. Reaction originally contained 1 mmol pyruvate and 3 mmol glyoxylate in 500 mM pH 7 ammonium phosphate buffer and incubated at 50°C for 24 hours. Note the abundance of signals in the carboxyl/amide region of the spectrum.

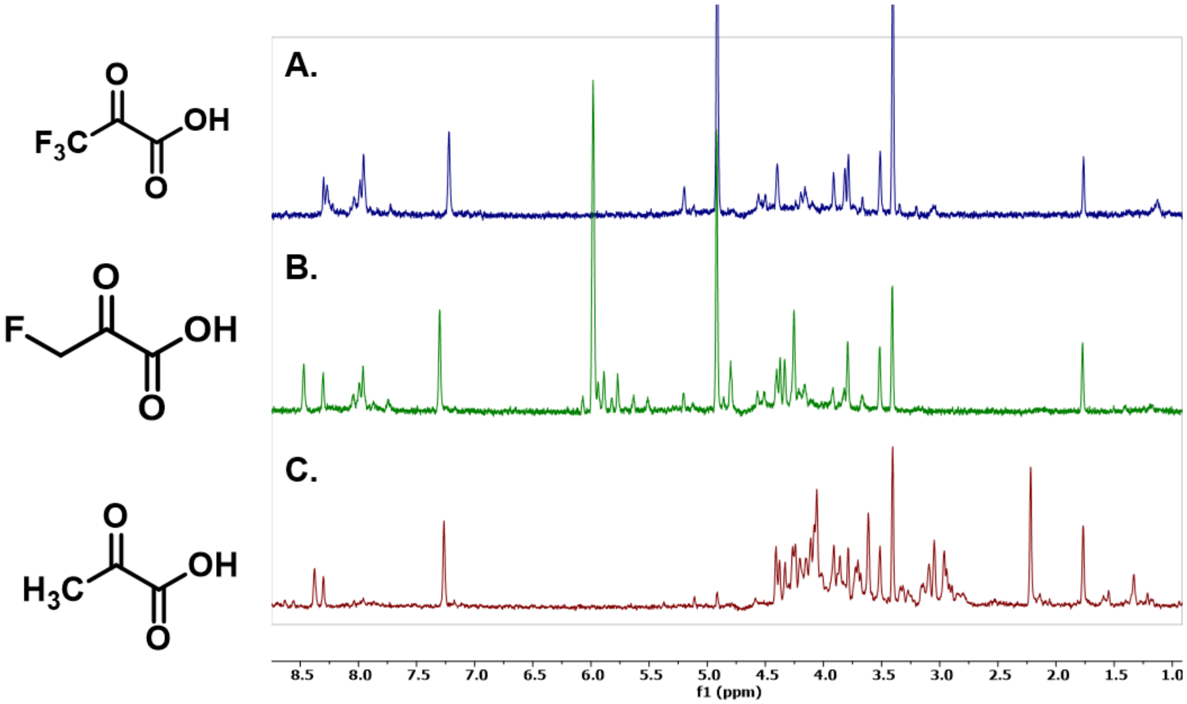


Figure 1.12 400 MHz ^{1}H presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard) spectra of 3 eq. glyoxylate reacted with 1 mM (A) trifluoropyruvate, (B) monofluoropyruvate, and (C) the pyruvate control in ammonium phosphate buffer (pH 7, 500 mM) for 24 hours at 50°C. Note how the control reaction with pyruvate is more complicated than the fluorine-containing derivatives. Mono- and trifluoro pyruvate seem to result in similar products, even though trifluoropyruvate is not enolizable. This suggests that reactivity is possible even without a pyruvate enolate/enamine.

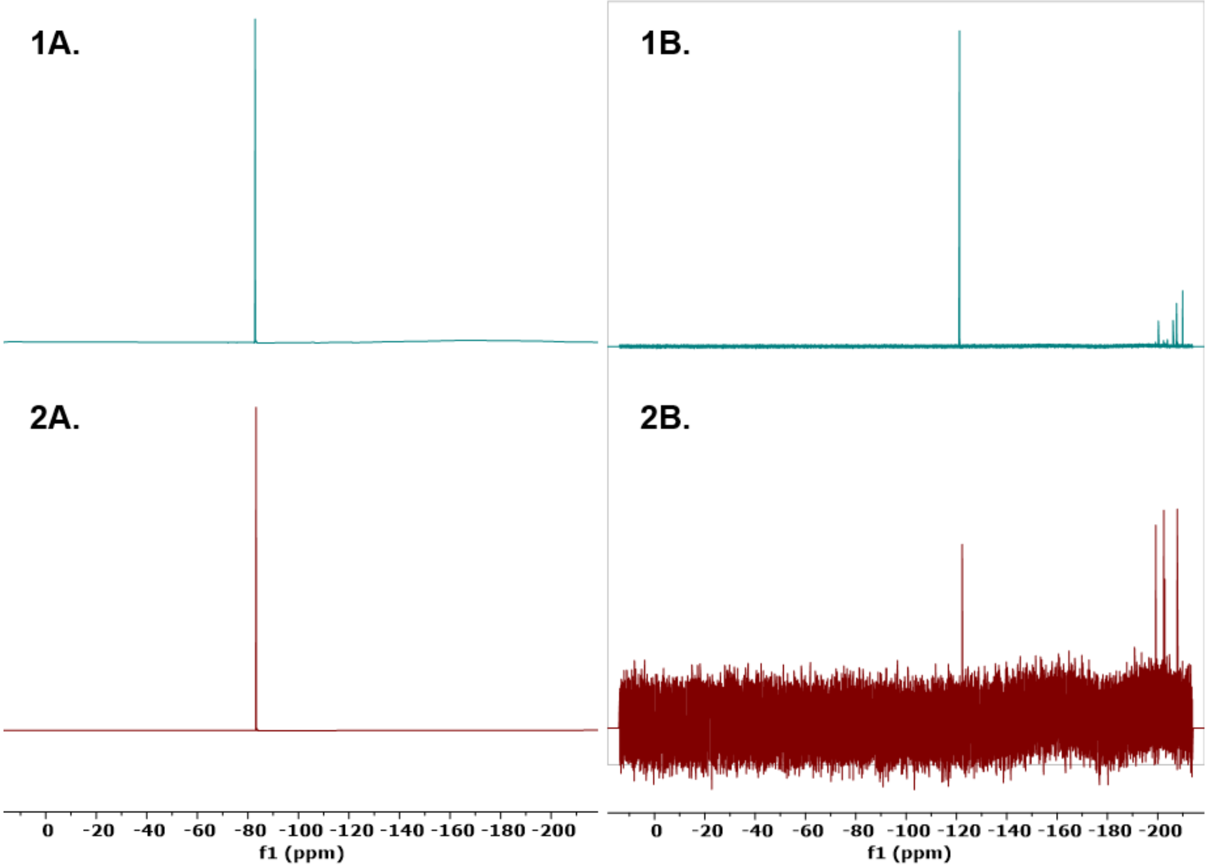
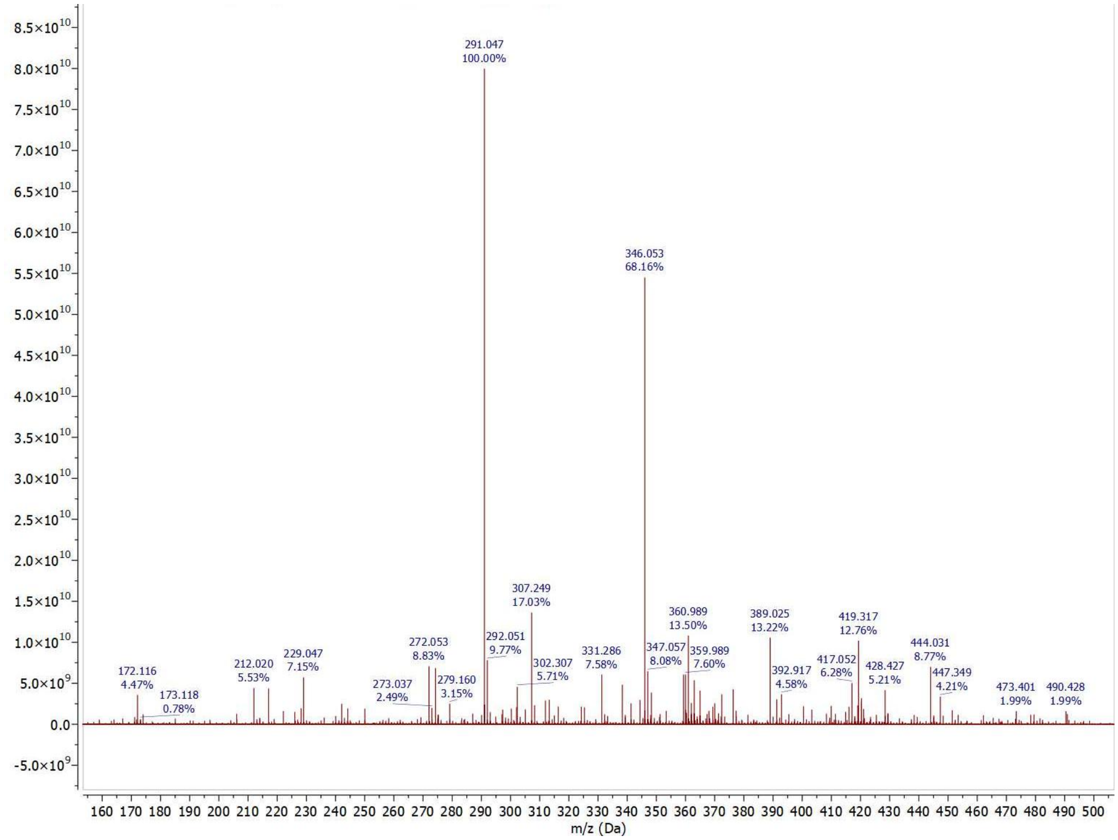
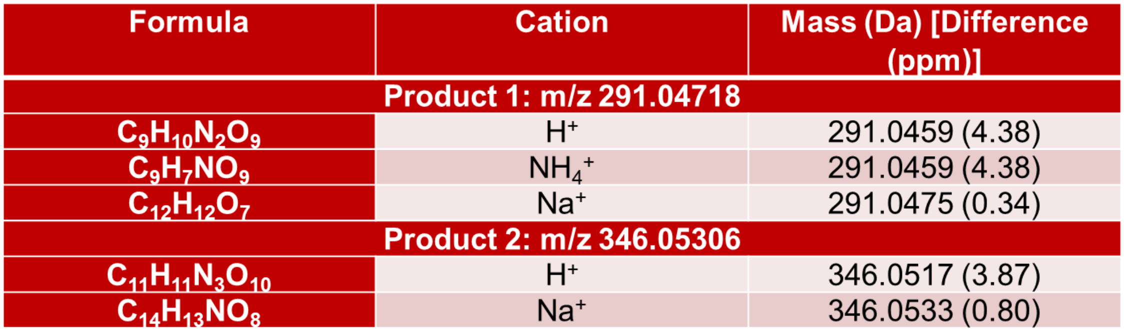


Figure 1.13 375 MHz ^{19}F NMR spectra (1) before and (2) after the reaction of 3 eq. glyoxylate with 1 mM (A) trifluoropyruvate or (B) monofluoropyruvate in ammonium phosphate buffer (pH 7, 500 mM) for 24 hours at 50°C. Note that spectrum 2B has a high baseline due to lower NMR sample concentration.

The abundance of carboxylates and the lack of a true chromophore forced us to adopt an alternative strategy toward the analysis of these compounds. A 1.5 meter-long column (3 cm diameter) was packed with silica and conditioned with water to perform hydrophilic interaction chromatography (HILIC) to separate the components of this reaction mixture. Eluting first with 10% water in acetonitrile helped desalt the mixture while most organics were retained. The organics were then eluted with increasing aqueous solvent (10-40%). Evaporation of the solvent yielded one band containing considerable solute. Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry was used to acquire the high-resolution masses; two highly abundant masses (**Figure 1.14**) were observed at *m*/*z* 291.04718 and *m*/*z* 346.05306 (positive mode ionization; negative mode was unsuccessful). Following an additional clean-up with preparative HILIC HPLC (40% water/MeCN, 0.1% TFA) the NMR spectrum was significantly simpler than the original mixture (**Figure 1.15 A**). Diffusion ordered spectroscopy (DOSY) was used to further separate the NMR signals into those that are associated with the smaller component and those associated with the larger component (**Figure 1.15 B**). Even with this information, 1) the high number of heteroatoms suggested by the formula make NMR interpretation difficult, and 2) it is not clear to what degree these two unknown compounds represent significant components of the reaction mixture. It is, however, evident that 1) nitrogen from the buffer is being incorporated into these molecules (nitrogen fixation, not simply iminium catalysis) and 2) new carbon-carbon bonds are being formed by an unknown mechanism.





Formula	Cation	Mass (Da) [Difference (ppm)]
Product 1: m/z 291.04718		
$C_9H_{10}N_2O_9$	H^+	291.0459 (4.38)
$C_9H_7NO_9$	NH_4^+	291.0459 (4.38)
$C_{12}H_{12}O_7$	Na^+	291.0475 (0.34)
Product 2: m/z 346.05306		
$C_{11}H_{11}N_3O_{10}$	H^+	346.0517 (3.87)
$C_{14}H_{13}NO_8$	Na^+	346.0533 (0.80)

Figure 1.14. High resolution mass spectrum acquired with Bruker FT-ICR MS (credit to Dr. Ying Ge for allowing access to this instrument, and to Dr. Kyle Brown for his assistance). Positive mode direct injection of pyruvate+glyoxylate+ammonium compounds after HILIC desalting and separation. Two compounds are high abundance. Possible molecular formulae are listed above. It is not clear to what extent these compounds represent the overall mixture.

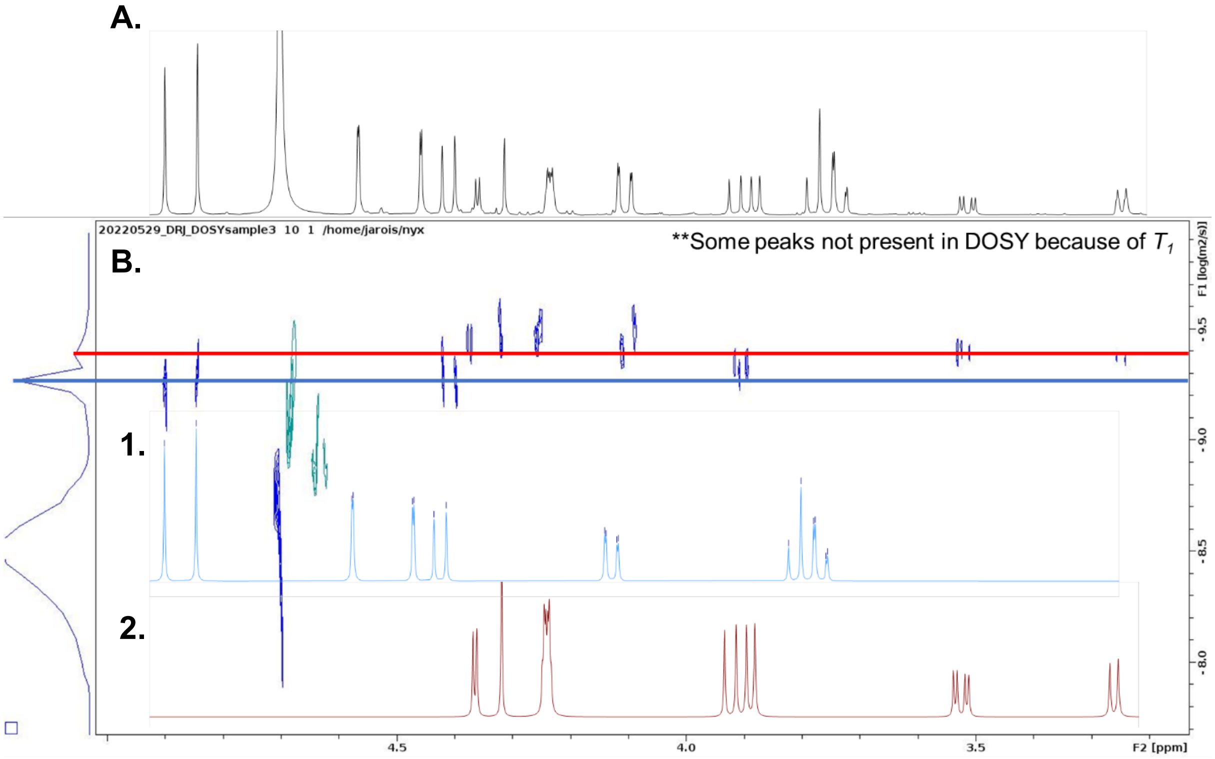


Figure 1.15 500 MHz (A) ^{1}H NMR spectrum of pyruvate + glyoxylate + ammonium compounds after HILIC desalting and separation. It is not clear to what extent these extracted compounds represent the overall mixture. (B) DOSY pseudo-2D spectrum helps distinguish which peaks are associated with each molecule. The digitally-separated NMR spectra are shown as B1 and 2. (B1) represents the compound with the smaller diffusion coefficient and is likely larger in mass than (B2). Based on the HRMS data in Figure 1.14., it is likely that B1 has an *m/z* of 346 while B2 has a *m/z* of 291.

Because of the difficulty involved with analyzing complex reaction mixtures in salty aqueous buffer, we decided to try to simplify the challenge by focusing on reactivity of glyoxylate alone rather than the glyoxylate + pyruvate mixture. This new approach led us to discover a previously unknown reaction mechanism, which will be the topic of the following chapter.

1.4. Conclusion

Here we highlight how ammonium ions can drastically change the reactivity of relevant prebiotic reactions, allowing not only a form of catalysis but also new reactions by which nitrogen may be incorporated into organic molecules. Although many products in the reaction of pyruvate and glyoxylate in ammonium phosphate buffer could not be identified, we have sufficient evidence to demonstrate the formation of new carbon-carbon bonds as well as the likely incorporation of nitrogen into these organic molecules. We therefore urge OoL researchers to consider what effect ammonium might have on proposed prebiotic reactions or systems.

1.5. Experimental

Water was purified to 18.2 MΩ with a Thermo Barnstead Nanopure system. Reagents were purchased from the following manufacturers: pyruvic acid (ACS, Sigma Aldrich), monofluoropyruvic acid (CombiBlocks), trifluoropyruvic acid (CombiBlocks), glyoxylic acid (50% in water, AlfaAesar), glyoxylic acid monohydrate (98%, Thermo), ammonium phosphate dibasic (reagent grade, Ward’s), ammonium phosphate monobasic (lab grade, Ward’s), ammonium hydroxide (30%, ACS grade, Sigma Aldrich), deuterium oxide (99.8%, Acros Organics), sodium phosphate monobasic (BioXtra, Sigma Aldrich), and sodium phosphate dibasic (ACS, Sigma Aldrich).

NMR spectra were collected on a Bruker Neo 500 MHz spectrometer with a Prodigy cryoprobe. High throughput NMR studies were collected on a Bruker Avance 400 MHz spectrometer with a BBFO probe. NMR samples were prepared by mixing equal volumes of the reaction mixture and D_2O. 3-(Trimethylsilylpropionic-2,2,3,3-d4 acid sodium salt ("TSP," 98%, Sigma Aldrich) was used as a chemical shift reference. In quantitative studies, TSP was added prior to the reaction and used as an integration standard to account for possible evaporation or dilution when adding ammonium hydroxide. pH measurements were acquired on a Mettler AB15 instrument. FT-ICR HRMS was conducted with a Bruker 12T SolariX FTMS with electrospray ionization. Waters LCMS was performed with an Acquity ARC and QDa ESI mass spectrometer. Most reactions were carried out in a New Brunswick Scientific temperature-controlled shaker.

Conversion of pyruvate and glyoxylate into α-Ketoacids.

Pyruvic acid (0.085 mL, 1 mmol, 200 mM) was dissolved in 4 mL of sodium phosphate buffer (pH 7, 500 mM). Glyoxylic acid (50% in water, 0.33 mL, 3 eq) was added. Concentrated NaOH solution was added until the pH was 7 (unless otherwise noted). Volume was corrected to 5 mL. The reaction was incubated at 40°C to 50°C for 24 hours and checked with ^{1}H presat NMR to find α-ketoacid products. Any deviation from this general procedure is clearly described.

Conversion of pyruvate and glyoxylate into unknown products.

Pyruvic acid (0.085 mL, 1 mmol, 200 mM) was dissolved in 4 mL of ammonium phosphate buffer (pH 7, 500 mM). Glyoxylic acid (50% in water, 0.33 mL, 3 eq) was added. Concentrated NH4OH solution was added until the pH was 7 (unless otherwise noted) and volume made to 5 mL. The reaction was incubated at 40°C to 50°C for 24 hours and checked with ^{1}H presat NMR to find products. Any deviation from this general procedure is clearly described.

1.6. References

(1) Leslie E., O. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39* (2), 99–123. https://doi.org/10.1080/10409230490460765.

(2) Cleaves, H. J. Prebiotic Chemistry: What We Know, What We Don't. *Evol. Educ. Outreach* **2012**, *5* (3), 342–360. https://doi.org/10.1007/s12052-012-0443-9.

(3) Pross, A. *What Is Life? How Chemistry Becomes Biology*; Oxford university press: Oxford, 2012.

(4) Kuhn, T. S.; Hacking, I. *The Structure of Scientific Revolutions*, 4th ed.; University of Chicago press: Chicago, 2012.

(5) Eschenmoser, A. Etiology of Potentially Primordial Biomolecular Structures: From Vitamin B_{12} to the Nucleic Acids and an Inquiry into the Chemistry of Life's Origin: A Retrospective. *Angew. Chem. Int. Ed.* **2011**, *50* (52), 12412–12472. https://doi.org/10.1002/anie.201103672.

(6) Ruiz-Mirazo, K.; Briones, C.; De La Escosura, A. Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. *Chem. Rev.* **2014**, *114* (1), 285–366. https://doi.org/10.1021/cr2004844.

(7) Pace, N. R. The Universal Nature of Biochemistry. *Proc. Natl. Acad. Sci.* **2001**, *98* (3), 805–808. https://doi.org/10.1073/pnas.98.3.805.

(8) Lazcano, A. Alexandr I. Oparin and the Origin of Life: A Historical Reassessment of the Heterotrophic Theory. *J. Mol. Evol.* **2016**, *83* (5–6), 214–222. https://doi.org/10.1007/s00239-016-9773-5.

(9) Cleaves, H. J.; Chalmers, J. H.; Lazcano, A.; Miller, S. L.; Bada, J. L. A Reassessment of Prebiotic Organic Synthesis in Neutral Planetary Atmospheres. *Orig. Life Evol. Biospheres* **2008**, *38* (2), 105–115. https://doi.org/10.1007/s11084-007-9120-3.

(10) Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **1953**, *117* (3046), 528–529. https://doi.org/10.1126/science.117.3046.528.

(11) Miller, S. L. Production of Some Organic Compounds under Possible Primitive Earth Conditions [1]. *J. Am. Chem. Soc.* **1955**, *77* (9), 2351–2361. https://doi.org/10.1021/ja01614a001.

(12) Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. The Miller Volcanic Spark Discharge Experiment. *Science* **2008**, *322* (5900), 404–404. https://doi.org/10.1126/science.1161527.

(13) Bada, J. L. New Insights into Prebiotic Chemistry from Stanley Miller's Spark Discharge Experiments. *Chem. Soc. Rev.* **2013**, *42* (5), 2186. https://doi.org/10.1039/c3cs35433d.

(14) Parker, E. T.; Cleaves, J. H.; Burton, A. S.; Glavin, D. P.; Dworkin, J. P.; Zhou, M.; Bada, J. L.; Fernández, F. M. Conducting Miller-Urey Experiments. *J. Vis. Exp.* **2014**, No. 83, 51039. https://doi.org/10.3791/51039.

(15) Parker, E. T.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Callahan, M.; Aubrey, A.; Lazcano, A.; Bada, J. L. Primordial Synthesis of Amines and Amino Acids in a 1958 Miller H_2 S-Rich Spark Discharge Experiment. *Proc. Natl. Acad. Sci.* **2011**, *108* (14), 5526–5531. https://doi.org/10.1073/pnas.1019191108.

(16) Oró, J. Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* **1961**, *191* (4794), 1193–1194. https://doi.org/10.1038/1911193a0.

(17) Oró, J.; Kimball, A. P. Synthesis of Purines under Possible Primitive Earth Conditions. I. Adenine from Hydrogen Cyanide. *Arch. Biochem. Biophys.* **1961**, *94* (2), 217–227. https://doi.org/10.1016/0003-9861(61)90033-9.

(18) Cairns-Smith, A. G. *Genetic Takeover and the Mineral Origins of Life*; Cambridge university press: Cambridge New York New Rochelle [etc.], 1987.

(19) Schwartz, A. W.; Orgel, L. E. Template-Directed Polynucleotide Synthesis on Mineral Surfaces. *J. Mol. Evol.* **1985**, *21* (3), 299–300. https://doi.org/10.1007/BF02102362.

(20) Wächtershäuser, G. Groundworks for an Evolutionary Biochemistry: The Iron-Sulphur World. *Prog. Biophys. Mol. Biol.* **1992**, *58* (2), 85–201. https://doi.org/10.1016/0079-6107(92)90022-X.

(21) Wächtershäuser, G. Evolution of the First Metabolic Cycles. *Proc. Natl. Acad. Sci.* **1990**, *87* (1), 200–204. https://doi.org/10.1073/pnas.87.1.200.

(22) Wächtershäuser, G. Before Enzymes and Templates: Theory of Surface Metabolism. *Microbiol. Rev.* **1988**, *52* (4), 452–484. https://doi.org/10.1128/mr.52.4.452-484.1988.

(23) Crick, F. H. C. The Origin of the Genetic Code. *J. Mol. Biol.* **1968**, *38* (3), 367–379. https://doi.org/10.1016/0022-2836(68)90392-6.

(24) Orgel, L. E. Evolution of the Genetic Apparatus. *J. Mol. Biol.* **1968**, *38* (3), 381–393. https://doi.org/10.1016/0022-2836(68)90393-8.

(25) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **1982**, *31* (1), 147–157. https://doi.org/10.1016/0092-8674(82)90414-7.

(26) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell* **1983**, *35* (3), 849–857. https://doi.org/10.1016/0092-8674(83)90117-4.

(27) Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319* (6055), 618–618. https://doi.org/10.1038/319618a0.

(28) Robertson, M. P.; Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (5), a003608–a003608. https://doi.org/10.1101/cshperspect.a003608.

(29) Sharp, P. A. On the Origin of RNA Splicing and Introns. *Cell* **1985**, *42* (2), 397–400. https://doi.org/10.1016/0092-8674(85)90092-3.

(30) Pace, N. R.; Marsh, T. L. Rna Catalysis and the Origin of Life. *Orig. Life Evol. Biosph.* **1985**, *16* (2), 97–116. https://doi.org/10.1007/BF01809465.

(31) Lewin, R. RNA Catalysis Gives Fresh Perspective on the Origin of Life: The Old Chicken-and-Egg Problem of the Origin of Life Is Illuminated in Unexpected Ways by Recent Results on the Splicing of RNA Precursors. *Science* **1986**, *231* (4738), 545–546. https://doi.org/10.1126/science.231.4738.545.

(32) Ashkenasy, G.; Hermans, T. M.; Otto, S.; Taylor, A. F. Systems Chemistry. *Chem. Soc. Rev.* **2017**, *46* (9), 2543–2554. https://doi.org/10.1039/C7CS00117G.

(33) Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature* **2009**, *459* (7244), 239–242. https://doi.org/10.1038/nature08013.

(34) Sutherland, J. D. The Origin of Life—Out of the Blue. *Angew. Chem. Int. Ed.* **2016**, *55* (1), 104–121. https://doi.org/10.1002/anie.201506585.

(35) Schrum, J. P.; Zhu, T. F.; Szostak, J. W. The Origins of Cellular Life. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (9), a002212–a002212. https://doi.org/10.1101/cshperspect.a002212.

(36) Sutherland, J. D. Opinion: Studies on the Origin of Life — the End of the Beginning. *Nat. Rev. Chem.* **2017**, *1* (2), 0012. https://doi.org/10.1038/s41570-016-0012.

(37) Muchowska, K. B.; Varma, S. J.; Moran, J. Nonenzymatic Metabolic Reactions and Life's Origins. *Chem. Rev.* **2020**, *120* (15), 7708–7744. https://doi.org/10.1021/acs.chemrev.0c00191.

(38) Koonin, E. Frozen Accident Pushing 50: Stereochemistry, Expansion, and Chance in the Evolution of the Genetic Code. *Life* **2017**, *7* (2), 22. https://doi.org/10.3390/life7020022.

(39) Leister, D. Thawing out Frozen Metabolic Accidents. *BMC Biol.* **2019**, *17* (1), 8. https://doi.org/10.1186/s12915-018-0621-5.

(40) Smith, E.; Morowitz, H. J. Universality in Intermediary Metabolism. *Proc. Natl. Acad. Sci.* **2004**, *101* (36), 13168–13173. https://doi.org/10.1073/pnas.0404922101.

(41) Morowitz, H. J.; Kostelnik, J. D.; Yang, J.; Cody, G. D. The Origin of Intermediary Metabolism. *Proc. Natl. Acad. Sci.* **2000**, *97* (14), 7704–7708. https://doi.org/10.1073/pnas.110153997.

(42) Zubarev, D. Y.; Rappoport, D.; Aspuru-Guzik, A. Uncertainty of Prebiotic Scenarios: The Case of the Non-Enzymatic Reverse Tricarboxylic Acid Cycle. *Sci. Rep.* **2015**, *5* (1), 8009. https://doi.org/10.1038/srep08009.

(43) Stubbs, R. T.; Yadav, M.; Krishnamurthy, R.; Springsteen, G. A Plausible Metal-Free Ancestral Analogue of the Krebs Cycle Composed Entirely of α-Ketoacids. *Nat. Chem.* **2020**, *12* (11), 1016–1022. https://doi.org/10.1038/s41557-020-00560-7.

(44) Erkkilä, A.; Majander, I.; Pihko, P. M. Iminium Catalysis. *Chem. Rev.* **2007**, *107* (12), 5416–5470. https://doi.org/10.1021/cr068388p.

(45) Erkkilä, A.; Pihko, P. M. Rapid Organocatalytic Aldehyde-Aldehyde Condensation Reactions. *Eur. J. Org. Chem.* **2007**, *2007* (25), 4205–4216. https://doi.org/10.1002/ejoc.200700292.

(46) Knoevenagel, E. Ueber Eine Darstellungsweise Der Glutarsäure. *Berichte Dtsch. Chem. Ges.* **1894**, *27* (2), 2345–2346. https://doi.org/10.1002/cber.189402702229.

(47) Knoevenagel, E. Condensation von Malonsäure Mit Aromatischen Aldehyden Durch Ammoniak Und Amine. *Berichte Dtsch. Chem. Ges.* **1898**, *31* (3), 2596–2619. https://doi.org/10.1002/cber.18980310308.

(48) Crowell, T. I.; Peck, D. W. Kinetic Evidence for a Schiff Base Intermediate in the Knoevenagel Condensation^1. *J. Am. Chem. Soc.* **1953**, *75* (5), 1075–1077. https://doi.org/10.1021/ja01101a018.

(49) Gellman, S. H. Foldamers: A Manifesto. *Acc. Chem. Res.* **1998**, *31* (4), 173–180. https://doi.org/10.1021/ar960298r.

(50) Girvin, Z. C.; Gellman, S. H. Exploration of Diverse Reactive Diad Geometries for Bifunctional Catalysis via Foldamer Backbone Variation. *J. Am. Chem. Soc.* **2018**, *140* (39), 12476–12483. https://doi.org/10.1021/jacs.8b05869.

(51) Girvin, Z. C.; Andrews, M. K.; Liu, X.; Gellman, S. H. Foldamer-Templated Catalysis of Macrocycle Formation. *Science* **2019**, *366* (6472), 1528–1531. https://doi.org/10.1126/science.aax7344.

(52) Andrews, M. K.; Liu, X.; Gellman, S. H. Tailoring Reaction Selectivity by Modulating a Catalytic Diad on a Foldamer Scaffold. *J. Am. Chem. Soc.* **2022**, *144* (5), 2225–2232. https://doi.org/10.1021/jacs.1c11542.

Chapter 2

The Aza-Cannizzaro Reaction

Portions of this chapter are summarized in the following article:

Jarois, D.R.; Schimmelpfennig, L.E.; Gellman, S.H. A New Mechanism for Formation of Glycine from Glyoxylic acid: the Aza-Cannizzaro Reaction. *Chem. Eur. J.* **2024**, *30,* e202403202. doi: 10.1002/chem.202403202

2.1. Abstract: The Aza-Cannizzaro Reaction

Glyoxylic acid and glycine are considered to be important building blocks relevant to the origin of life. Numerous reactions and mechanisms have been previously reported for the interconversion of glyoxylic acid and glycine. Here a new reaction is described in which glyoxylic acid in ammonium-rich aqueous buffers forms a hemiaminal/imine mixture that disproportionates into oxamate and glycine. Quantitative NMR studies, isotope labelling, and kinetics studies all provide evidence to suggest that a hydride transfer from the glyoxylate hemiaminal to the iminium ion underlies the mechanism. This parallels the well-known Cannizzaro reaction. This ‘aza-Cannizzaro’ discovery offers a new perspective on how nitrogen may have been incorporated into more complex prebiotic molecules and systems.

2.2. Pyruvate and Glyoxylate Revisited

In Chapter 1, the reaction of glyoxylate and pyruvate in aqueous ammonium phosphate buffer was investigated, and numerous new products, many of which remain unidentified, were observed. In order to better understand what reactions might occur between pyruvate and glyoxylate in ammonium phosphate buffer, we decided to simplify the system and run control experiments with each starting material individually to understand if there is any self-reactivity of either pyruvate or glyoxylate under these conditions.

The self-condensation of pyruvate into parapyruvate, zymonate, and 2-methyl-4-oxopene-2-enedioate (OMPD) has been recently investigated.[1] Our own experimental evidence are consistent with the reported reaction pathway in both sodium and ammonium-rich environments (**Figure 2.1**).

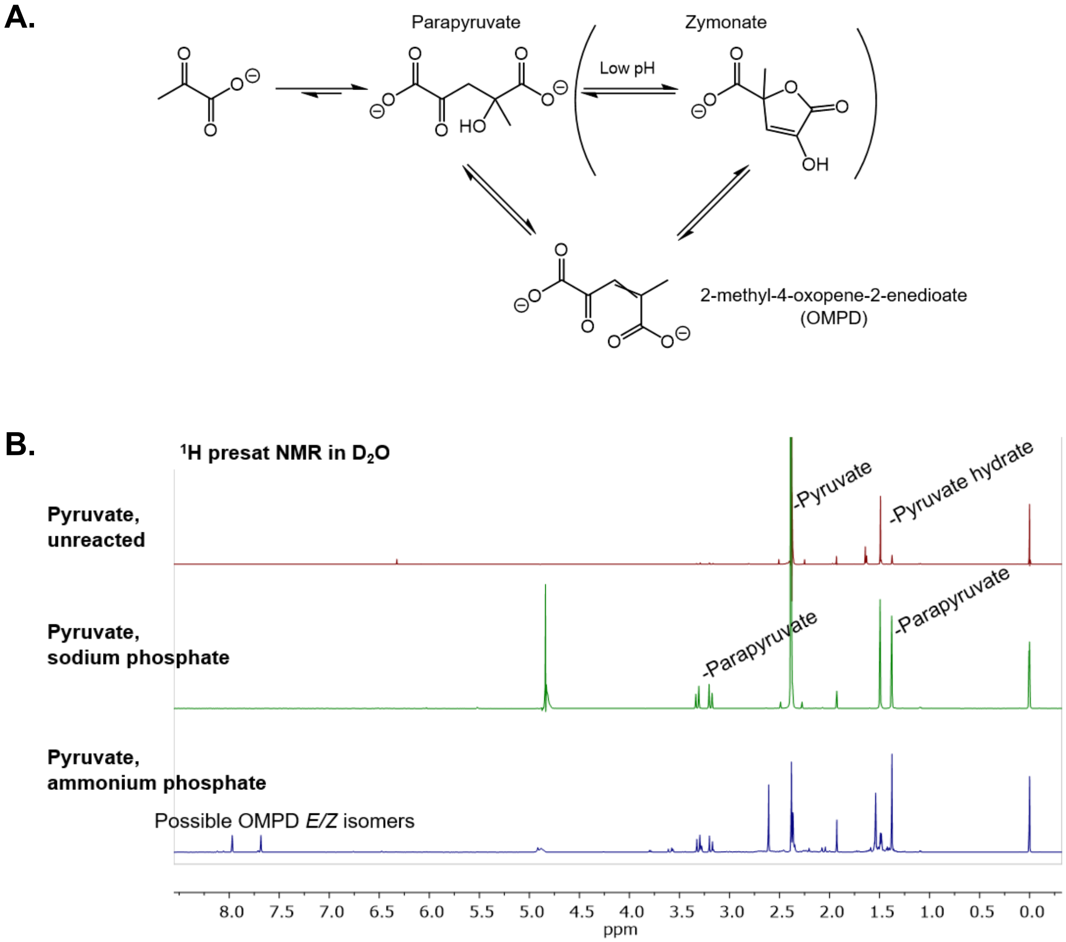


Figure 2.1 Pyruvate, when heated at 50°C for 24 hours in aqueous buffer (here, 500 mM sodium or ammonium phosphate, pH 7) forms parapyruvate, zymonate, and OMPD. (A) shows the overall reaction, while (B) shows NMR characterization of the reaction mixtures in aqueous buffers with 500 MHz ^{1}H presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d$_4$ acid, sodium salt) standard).

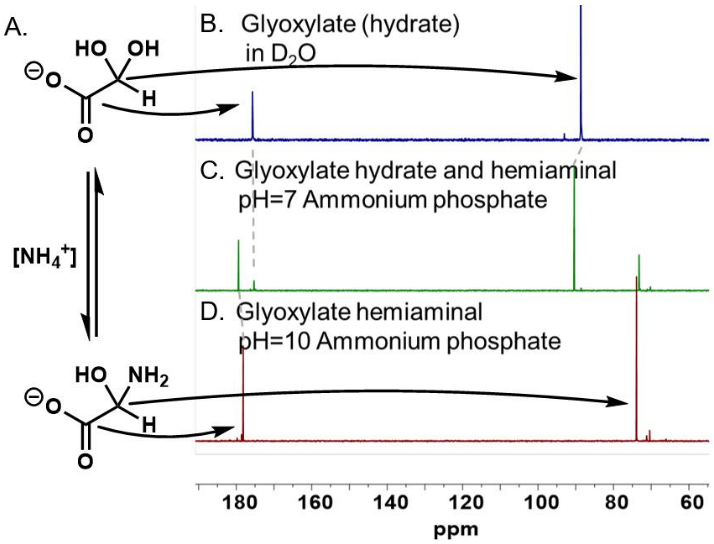


Figure 2.2 Glyoxylate forms the hemiaminal in the presence of ammonium phosphate. (A) Equilibrium between the hydrate and hemiaminal forms of glyoxylate. (B) In unbuffered D_2O, ^{13}C NMR (125 MHz) indicates that a 500 mM solution of glyoxylate contains mostly the hydrate form (δ175 and 88 ppm). (C) In ammonium phosphate buffer, pH 7, the hemiaminal (δ178 and 73 ppm) is observed in addition to the hydrate. (D) In ammonium phosphate buffer, pH 10, only the hemiaminal is observed.

Most studies of the reaction of ammonium with glyoxylate are dated.[2,3] In the 1990s, the reactivity between glyoxylate and ammonia was revisited, and it was concluded using solid state NMR experiments that ammonia reacts with glyoxylic acid to generate the corresponding hemiaminal.[4] Computational studies by Mayer and Moran support the conclusion that the iminium/hemiaminal pair is lower in energy than the glyoxylate and the hydrate.[5] Our own NMR studies also support the complete conversion of glyoxylate into the iminium/hemiaminal mixture as ammonia is added to the solution (**Figure 2.2**).

To investigate whether the glyoxylate iminium/hemiaminal pair could undergo further reaction, we incubated glyoxylate with 500 mM ammonium phosphate buffer (pH set to 7 with ammonium hydroxide). Upon incubation for 48 hours at 50°C, numerous products were observed by ^{1}H NMR including one that gave rise to a large singlet at 3.57 ppm (**Figure 2.3**). Other prebiotic chemistry studies indicate that glycine has a proton resonance with this chemical shift.[6] We observed that adding glycine into our reaction mixture led the singlet at 3.57 ppm to become more intense (**Figure 2.4**). Parallel observations were made via ^{13}C NMR. These NMR results strengthened our conclusion that glycine was being formed under these conditions.

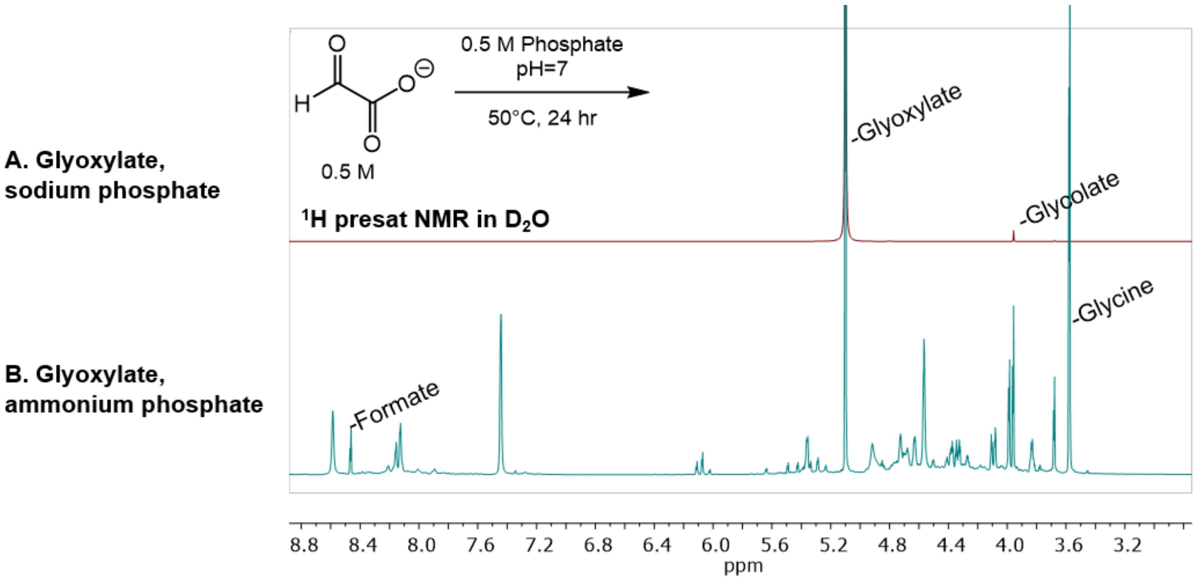


Figure 2.3 ^{1}H-presat (selective pulse at 4.7 ppm to attenuate HOD signal) (500 MHz) with full-size inset to show relative abundance of products obtained from 0.5 M glyoxylate in 0.5 M (A) sodium phosphate or (B) ammonium phosphate neutralized to pH 7 and heated 24 hr at 50 °C. Aliquot was mixed with an equal volume of D_2O prior to analysis. Identifiable resonances are labelled.

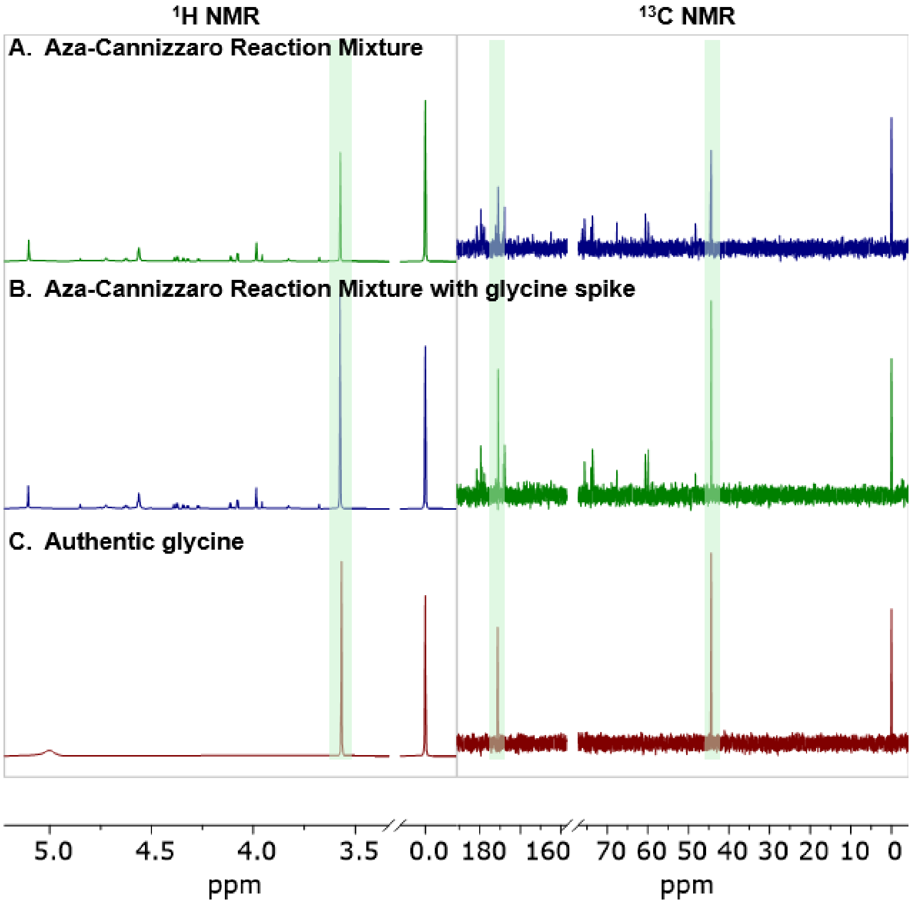


Figure 2.4 Identification of glycine in the aza-Cannizzaro reaction mixture. The presence of glycine in the aza-Cannizzaro reaction mixture was established by NMR, with 3.8 mM TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) as an internal standard. (A) Selected portions of ^{1}H NMR (selective pulse at 4.7 ppm to attenuate HOD resonance) (500 MHz; pulse sequence=zgpr30; relaxation delay=15s) and ^{13}C NMR spectra (125 MHz, zgpg30; relaxation delay=5s) for glyoxylate (500 mM) in pH 7 ammonium phosphate (500 mM) after 48 hr incubation at 50°C; the reaction solution was mixed with an equal volume of D_2O before analysis. (B) Selected portions of ^{1}H and ^{13}C NMR spectra of the same sample after addition of 23 mM glycine. (C) Selected portions of ^{1}H and ^{13}C NMR spectra of 23 mM glycine in 250 mM ammonium phosphate in 1:1 D_2O:H_2O.

2.3. Glyoxylate and Glycine

Glyoxylate is often thought to have been a precursor for other prebiotic molecules.[7–9] Eschenmoser famously postulated a ‘glyoxylate scenario’ by which glyoxylate was argued to serve as a starting point for both sugars and amino acids through catalytic and autocatalytic networks with other compounds.[7,10] Glyoxylate was detected in recent versions of the Miller-Urey experiment and is thought to occur in interstellar ices.[11,12] Glycine is also an important prebiotic feedstock molecule. Although glycine has not been confirmed in the interstellar medium, it has been found on comets.[13–15] Prebiotic processes that generated glycine may have been involved in more complex reaction cycles because glycine undergoes transamination with α-ketoacids to form other amino acids.[16–18]

Numerous publications focusing on prebiotic chemistry have reported the transformation of glyoxylate into glycine previously (**Figure 2.5**). Glyoxylic acid and ammonia can form glycine with iron(II) catalysis through single electron transfer from the Fe(II) ion to the imine.[19] A similar transformation is reported to occur in sodium silicate.[20] Others note that glycine imine can be prepared through photochemical denitrogenation of 2-azidoacetic acid, and glyoxylate can reduce the imine to glycine.[21] Glyoxylic acid can undergo an acid-catalyzed reaction with ammonium sulfate to form *N*-oxalylglycine, which can undergo subsequent hydrolysis to form glycine.[22–24] A comparable transformation can occur in tetrahydrofuran with ethyl glyoxylate to generate the diethyl ester of *N*-oxalylglycine.[25] Most recently, the reaction of glyoxylic acid and diamidophosphate in the presence of cyanide and carbonate buffer was reported to form glycine through a Bucherer-Bergs-type mechanism.[26,27]

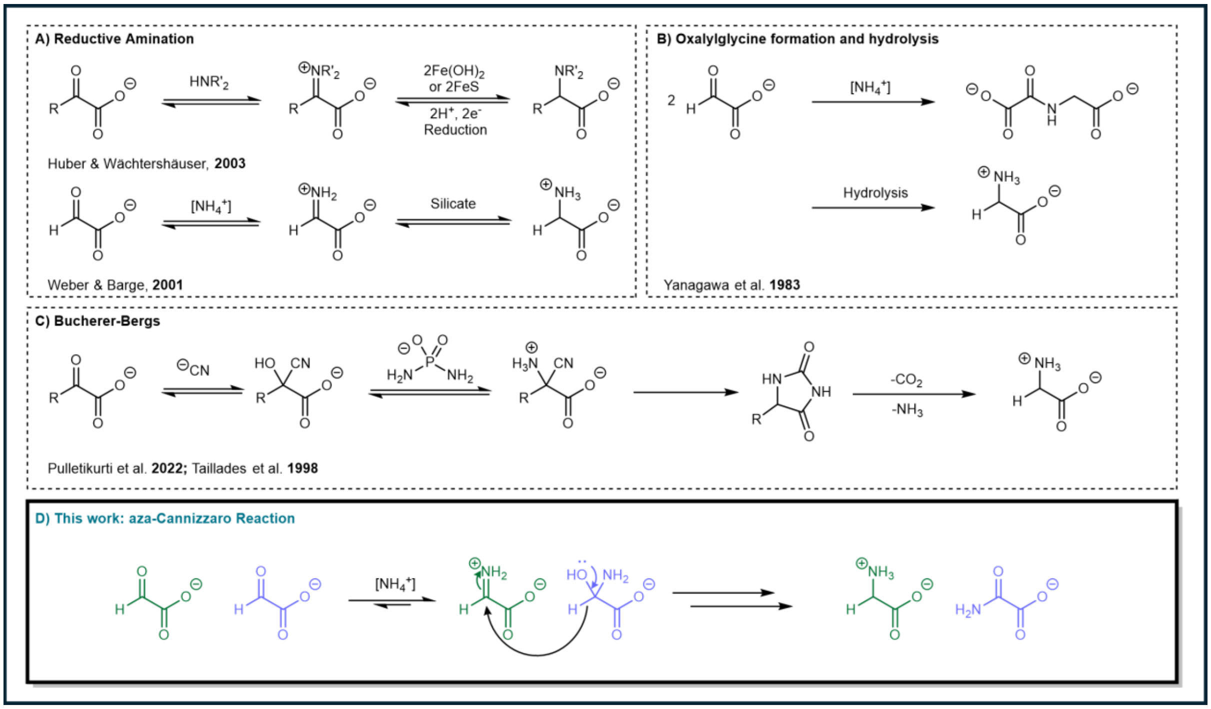


Figure 2.5 Several previously reported transformations of 2-oxoacids to α-amino acids proceed through: A) reductive amination; B) transamination-like mechanisms resulting in oxalylglycine and subsequent hydrolysis; or C) a Bucherer-Bergs reaction. Our work (D) highlights a previously unknown aza-Cannizzaro reaction.

It was apparent to us that none of the previously reported transformations could explain our observations. Only glyoxylate, ammonium phosphate, and water were present without any strong acids or bases, metal ions, cyanide, or other reducing agents.

2.3.1. Glycine from Glyoxylate

After observing glycine formation with only glyoxylic acid and aqueous ammonium phosphate, we decided to focus our efforts on understanding this reaction. In addition to glycine, other products were observed in this reaction mixture. Glycolic acid, oxalylglycine, ammonium formate, oxalate, and bicarbonate were some of the signals identifiable based on ^{1}H and/or ^{13}C NMR, each confirmed with spiking authentic samples (**Figure 2.6**). Several other signals, mostly lower in intensity, were also observed but could not be identified. Buffering anions such as sulfate, phosphate, or even acetate were necessary for the reaction to proceed (**Figure 2.7**). The absence of an oligo-oxo counterion diminished glycine formation significantly, indicating that the oligo-oxo anions sulfate, phosphate and acetate may serve a catalytic function (**Figure 2.8**). Glycine formation was also diminished below pH 7 (**Figure 2.9**). The reaction proceeded most quickly at elevated temperatures, and the exclusion of atmospheric oxygen did not change glycine yield (**Figure 2.10**). The overall transformation, including identifiable products and buffer effects, is summarized in **Figure 2.11.**

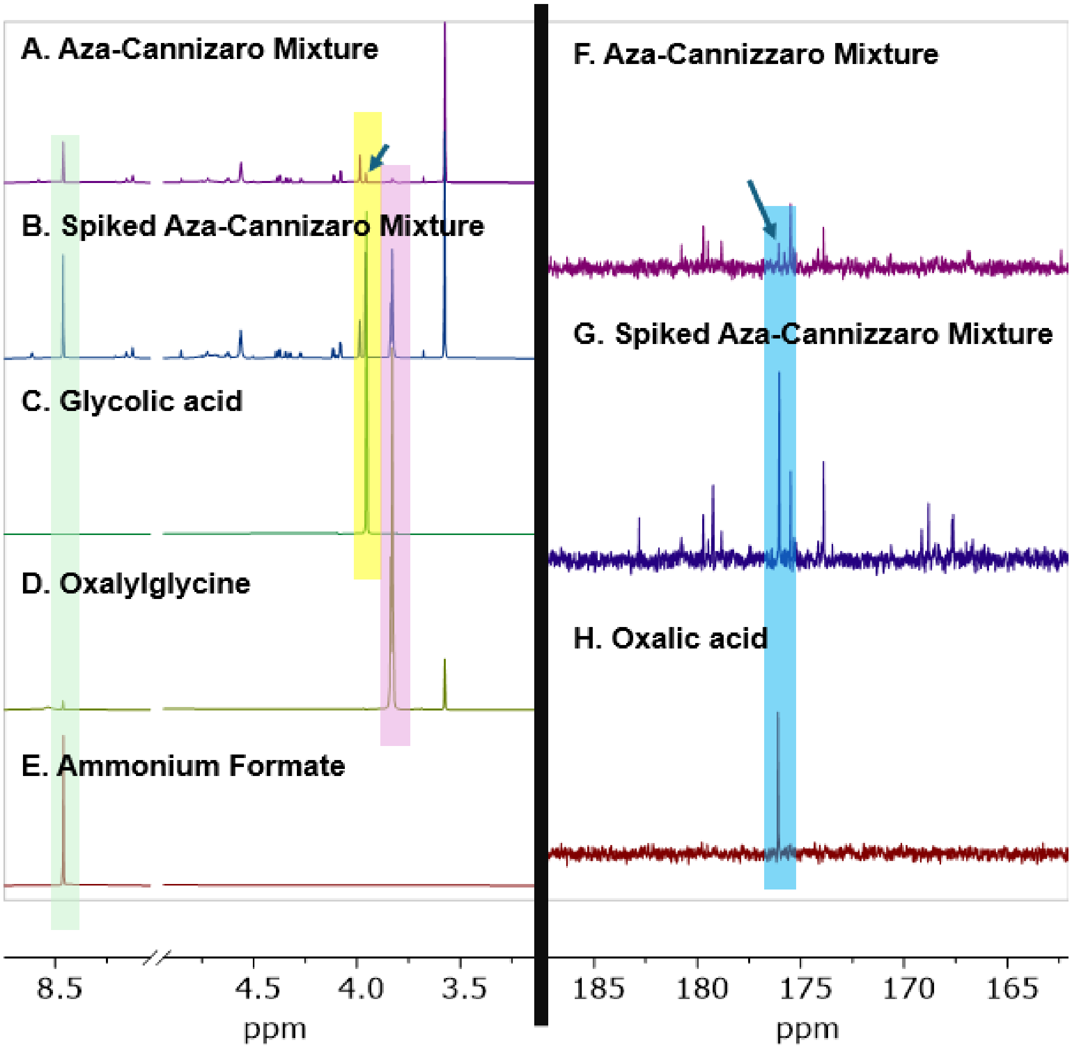


Figure 2.6 Identification of side products in the reaction mixture. The presence of glycolic acid (yellow shading), oxalylglycine (pink shading), formate (green shading), and oxalate (blue shading) in the aza-Cannizzaro reaction mixture was established by NMR with 3.8 mM TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) as an internal standard. (A) and (F) represent ^{1}H-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) (500 MHz; pulse sequence=zgpr30; relaxation delay=15s) and ^{13}C NMR spectra (125 MHz, zgpg30; relaxation delay=5s) for glyoxylate (500 mM) in pH=7 ammonium phosphate (500 mM) after 48 hr incubation at 50°C; the reaction solution was mixed with an equal volume of D_2O. (B) and (G) show the same sample spiked with 23 mM each of glycolic acid, oxalylglycine, ammonium formate, and oxalic acid. Note the highlighted signals. (C-E) and (H) are pure 23 mM samples of the nominal compound in 250 mM ammonium phosphate in 1:1 D_2O:H_2O. Glycolate and oxalate are present in very small amounts in the original sample and the corresponding resonances in (A) and (F) (respectively) are indicated with arrows.

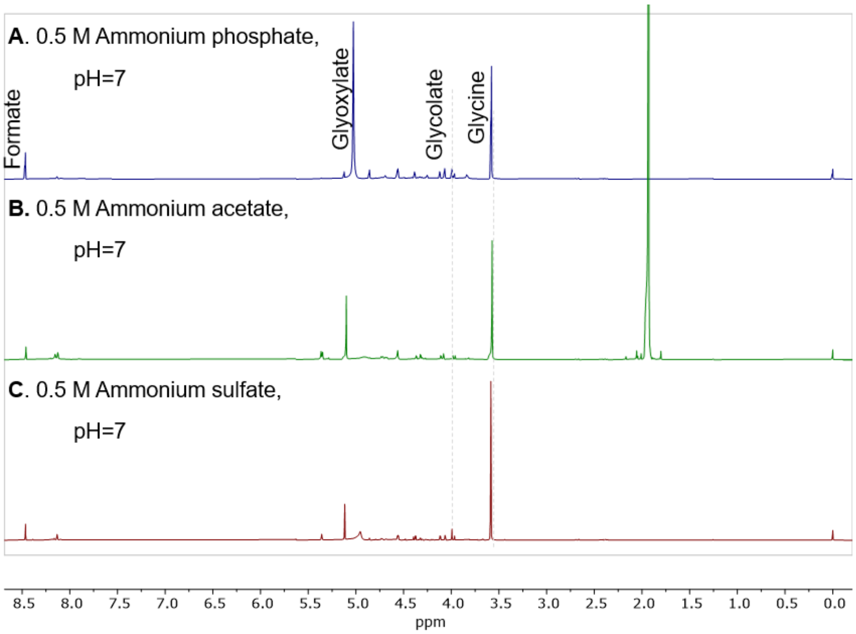


Figure 2.7 The aza-Cannizzaro reaction occurs in the presence of multiple buffer anions. ^{1}H-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) (500 MHz; pulse sequence=zgpr30, with TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard) NMR spectra of glyoxylate (500 mM) in pH=7 ammonium solution (500 mM) after 48 hr incubation at 50°C; aliquot of reaction was mixed with an equal volume of D_2O before analysis. Note that glycine formation occurs in (A) ammonium phosphate, (B), ammonium acetate (acetate signal at 1.8 ppm), and (C) ammonium sulfate.

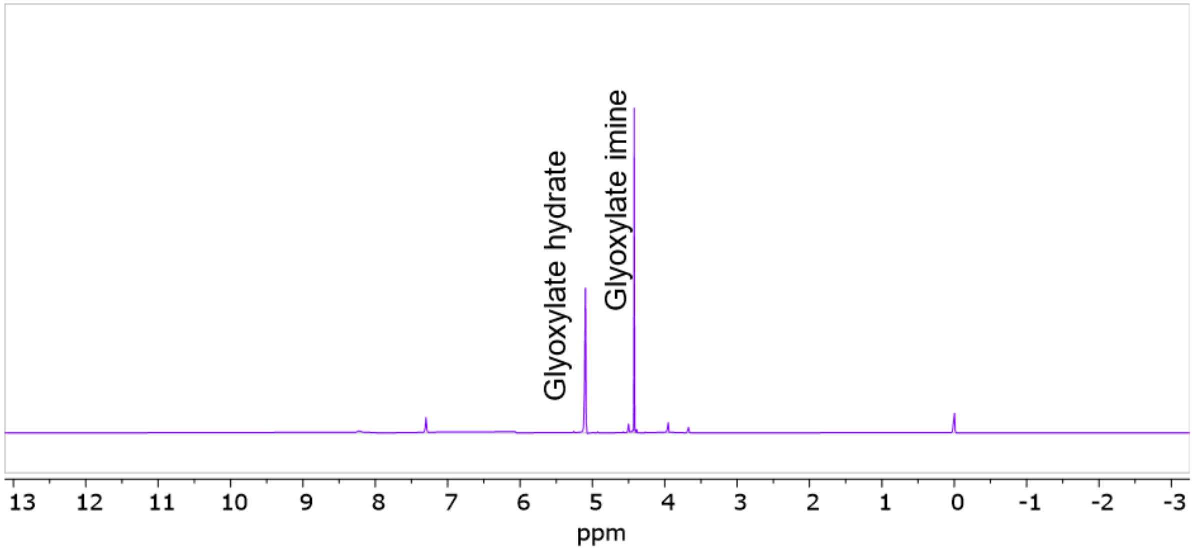


Figure 2.8 The aza-Cannizzaro reaction is impaired in the absence of phosphate, sulfate, or acetate. Glyoxylate (500 mM) was neutralized with ammonium hydroxide to pH 7 and incubated at 50°C for 48 hr. Reaction aliquot mixed with an equal volume of D_2O prior to analysis. Only small amounts of Cannizzaro and aza-Cannizzaro products were detected via ^1H NMR (selective pulse at 4.7 ppm to attenuate HOD resonance) (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard).

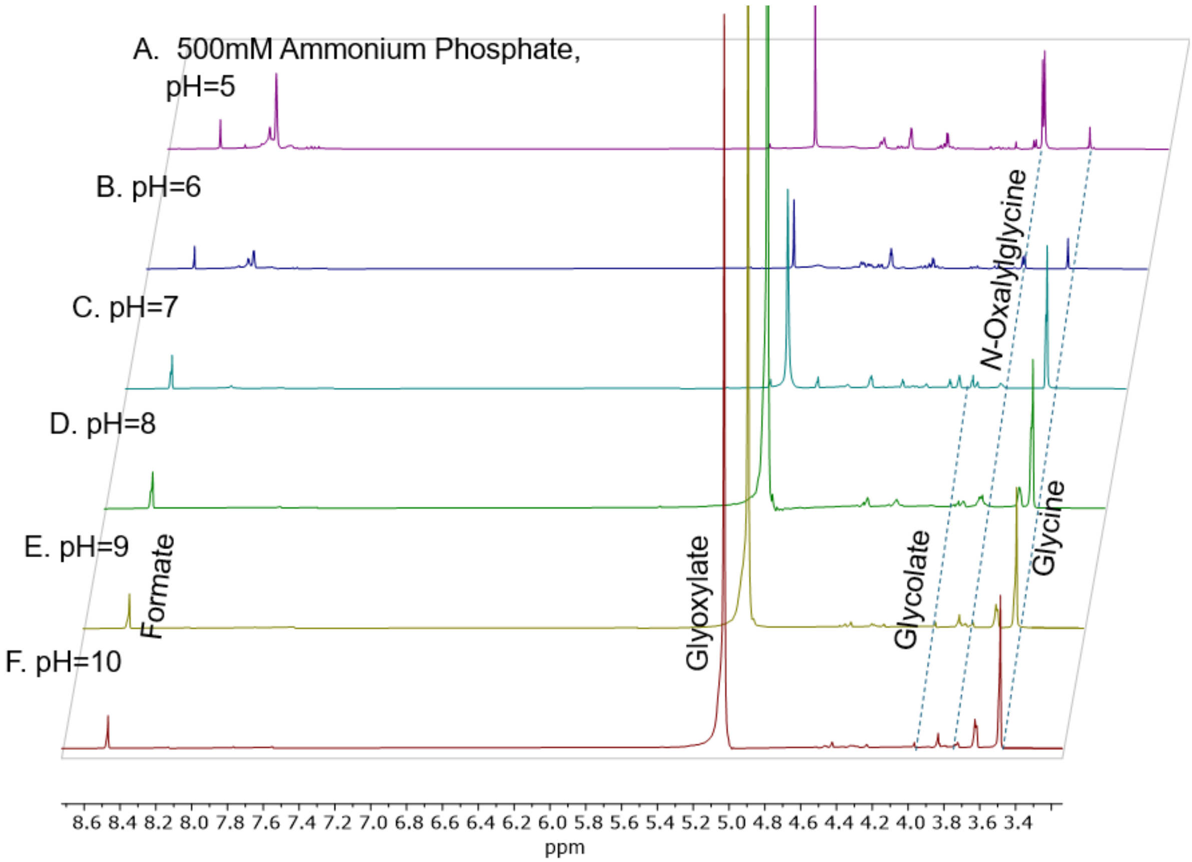


Figure 2.9 Reactivity was observed over wide pH range. Glyoxylate (500 mM) in ammonium phosphate (500 mM, pH 5-10) was incubated at 50°C for 48 hr. Reaction aliquots were combined with an equal volume of D_2O prior to analysis. Glycine formation was observed at each pH via ^1H NMR-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard). Glycine formation is most favorable above pH 6 (C-F). Other products are observed. Oxalylglycine formation is most prominent at pH 5 (A).

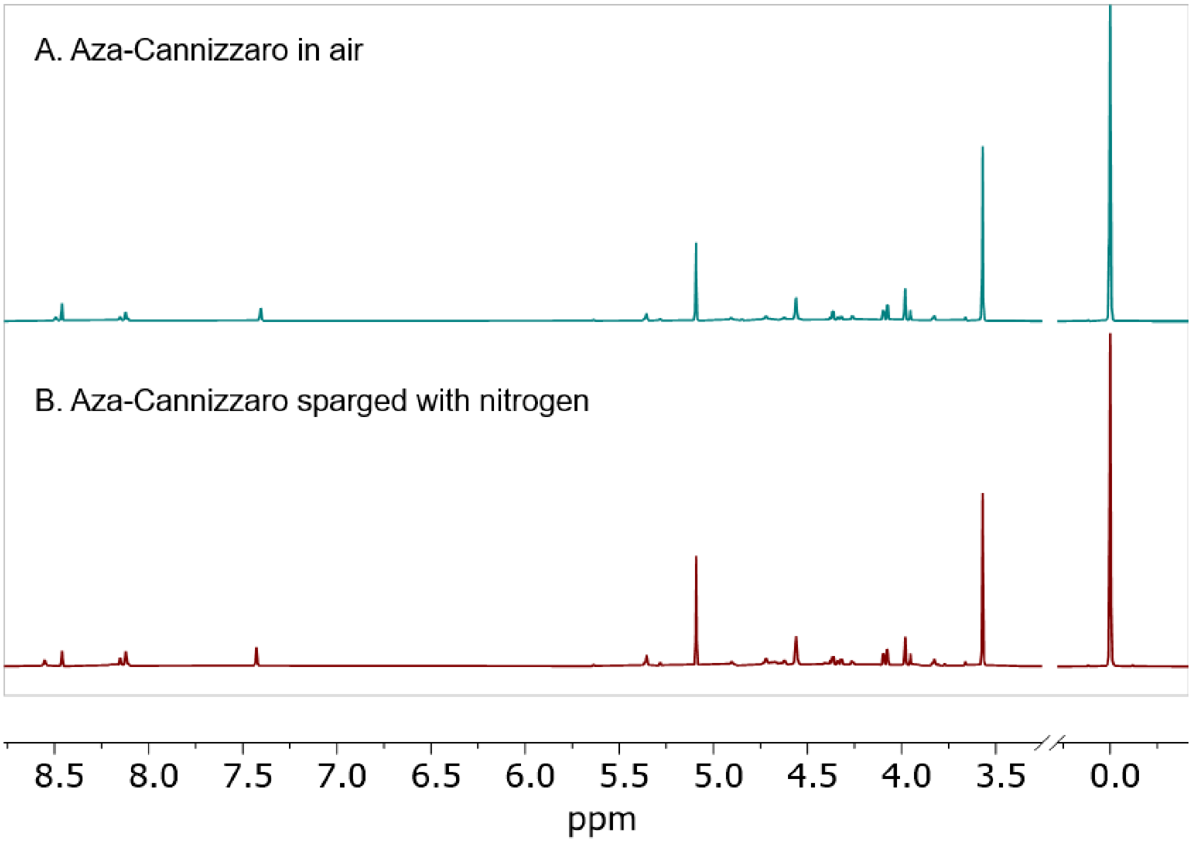
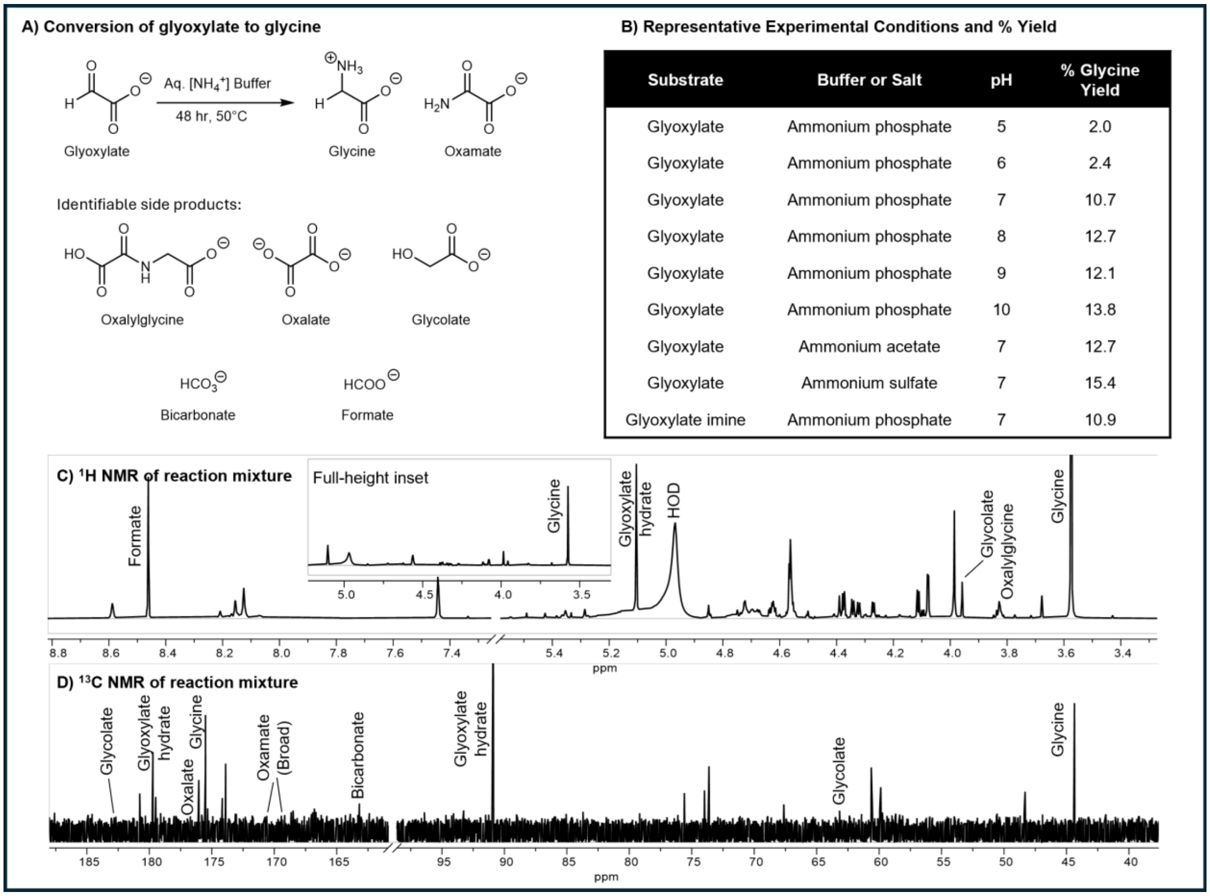


Figure 2.10 Atmospheric oxygen has no effect on the reaction. A solution containing glyoxylate (500 mM) in ammonium phosphate (500 mM, pH 7) was either (A) shaken to introduce oxygen or (B) sparged with nitrogen and incubated at 50°C for 48 hr. Reaction aliquot combined with an equal volume of D_2O prior to analysis. Glycine formation was observed in both samples via ^1H NMR-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard) with negligible observable differences between the two conditions.



Substrate	Buffer or Salt	pH	% Glycine Yield
Glyoxylate	Ammonium phosphate	5	2.0
Glyoxylate	Ammonium phosphate	6	2.4
Glyoxylate	Ammonium phosphate	7	10.7
Glyoxylate	Ammonium phosphate	8	12.7
Glyoxylate	Ammonium phosphate	9	12.1
Glyoxylate	Ammonium phosphate	10	13.8
Glyoxylate	Ammonium acetate	7	12.7
Glyoxylate	Ammonium sulfate	7	15.4
Glyoxylate imine	Ammonium phosphate	7	10.9

Figure 2.11 Reaction Summary: Glyoxylate is transformed to glycine when heated to 50°C for 48 hr in ammonium-rich buffers (0.5 M) at varying pH. A) Summary of reaction outcome. B) Glycine yield under differing experimental conditions based on integration of NMR signals relative to a TSP(3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard (note that maximum glycine yield is 50%; see text). Representative NMR data: C) ^{1}H-presat (selective pulse at 4.7 ppm to attenuate HOD signal) (500 MHz) with full-size inset to show relative abundance of glycine; and D) ^{13}C NMR (125 MHz, with 5 s relaxation delay to visualize all signals). Both spectra obtained from 0.5 M glyoxylate and 0.5 M ammonium phosphate that was neutralized to pH 7 and heated 48 hr at 50 °C. Aliquot was mixed with an equal volume of D_2O prior to analysis (see experimental information). Identifiable resonances are labelled.

2.3.2. The Aza-Cannizzaro Hypothesis

Because of the absence of other reducing agents in the glyoxylate and ammonium reaction mixture, it seemed most likely that the glyoxylate imine/hemiaminal pair served a dual purpose as the oxidizing agent and the reducing agent in the glycine-forming system. This disproportionation parallels the well-known Cannizzaro reaction, by which one equivalent of an aldehyde is oxidized while a second equivalent is reduced.

Cannizzaro reactions typically occur at high pH, requiring the nucleophilic attack of hydroxide on an aldehyde to form an alkoxide intermediate that acts as a hydride donor to a second equivalent of aldehyde. The alkoxide is oxidized to the carboxylic acid; the aldehyde is reduced to an alcohol.[28] This 'classic' reaction is often taught with benzaldehyde, the historic example,[29] but other aldehydes undergo this reaction as well, including glyoxylate (**Figure 2.12 A**).[16] When glyoxylate is incubated in pH 9 sodium phosphate buffer for 21 hours at 40°C, oxalate and glycolate are both observed via ^{13}C NMR, consistent with the Cannizzaro reaction products (**Figure 2.12 B**). This traditional Cannizzaro reaction in the absence of ammonium does not occur at pH 7. We propose that similar reactivity might occur in ammonium buffers between the hemiaminal and iminium of glyoxylate (**Figure 2.12 A**) to form the observed products, including glycine (**Figure 2.12 C**).

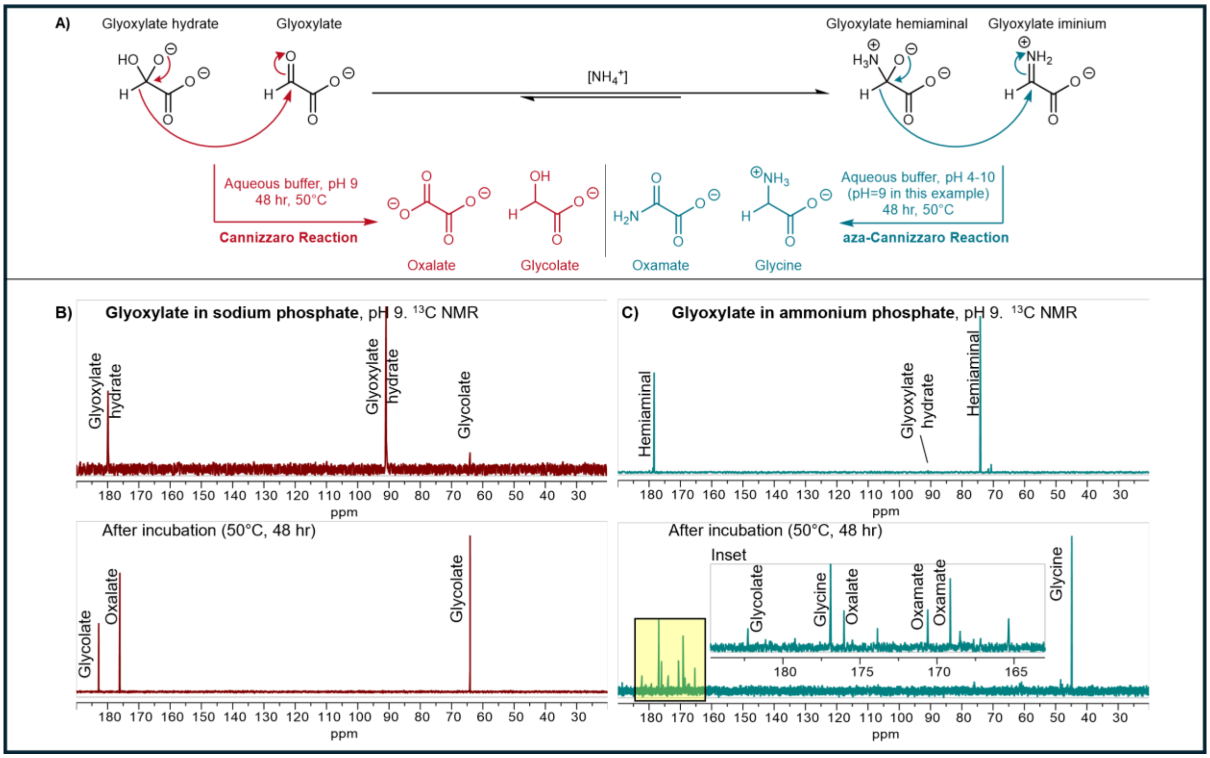


Figure 2.12 A) Glyoxylate undergoes the well-known Cannizzaro reaction at elevated pH. In the presence of ammonium, glyoxylate exists in equilibrium with the corresponding hemiaminal and iminium, which are proposed to disproportionate into glycine and oxamate via an "aza-Cannizzaro" mechanism. B,C) The Cannizzaro transformation of glyoxylate into glycolate and oxalate and the aza-Cannizzaro transformation into glycine and oxamate can be monitored by ^{13}C NMR (125 MHz). Glyoxylate (0.5 M) was neutralized and incubated at 50°C for 48 hr in pH 9 sodium phosphate buffer (0.5 M) or pH 9 ammonium phosphate (0.5 M) buffer to form the products indicated.

To test whether this glycine-forming reaction proceeds through a Cannizzaro-like mechanism, we first sought spectroscopic evidence for the predicted by-product. If one equivalent of glyoxylate iminium is reduced to glycine, another equivalent should be oxidized to oxamic acid. Oxamic acid does not have any non-exchangeable protons and is therefore not observed via ^{1}H NMR in D_2O. Observation via^{13}C NMR spectroscopy is challenging because the molecule has broad signals and a long relaxation time. When a pulse sequence that does not rely on nOe enhancement via ^{1}H presaturation is selected (viz. zgig30 on Bruker instruments) and the NMR relaxation delay is elongated to 20 seconds, oxamate can be detected (confirmed with spike, **Figure 2.13**) and the signals can be integrated to indicate an approximate 1:1 ratio with glycine (**Figure 2.14**).

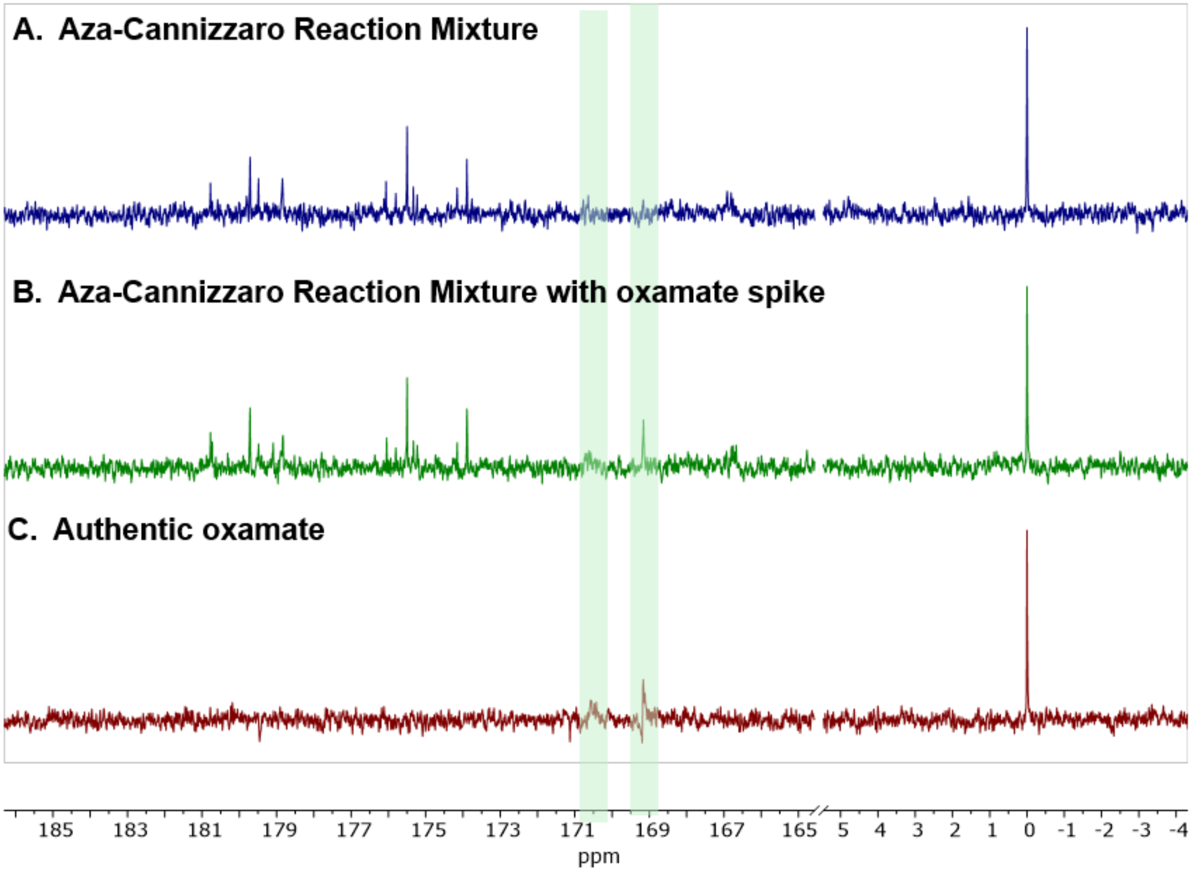


Figure 2.13 Identification of oxamic acid in the aza-Cannizzaro reaction mixture. The presence of oxamate in the aza-Cannizzaro reaction mixture was established by ^{13}C NMR, with 3.8 mM TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) as an internal standard. (A) Selected portions of the ^{13}C NMR spectrum (125 MHz, zgpg30; relaxation delay=5s) for glyoxylate (500 mM) in pH 7 ammonium phosphate (500 mM) after 48 hr incubation at 50°C; the reaction solution was mixed with an equal volume of D_2O before analysis . (B) Selected portions of the ^{13}C NMR spectrum of the same sample after addition of 23 mM oxamate. (C) Selected portions of ^{3}C NMR spectrum of 23 mM oxamate in 250 mM ammonium phosphate in 1:1 D_2O:H_2O. Note that oxamate signals are significantly broadened at pH 7 and require long relaxation times.

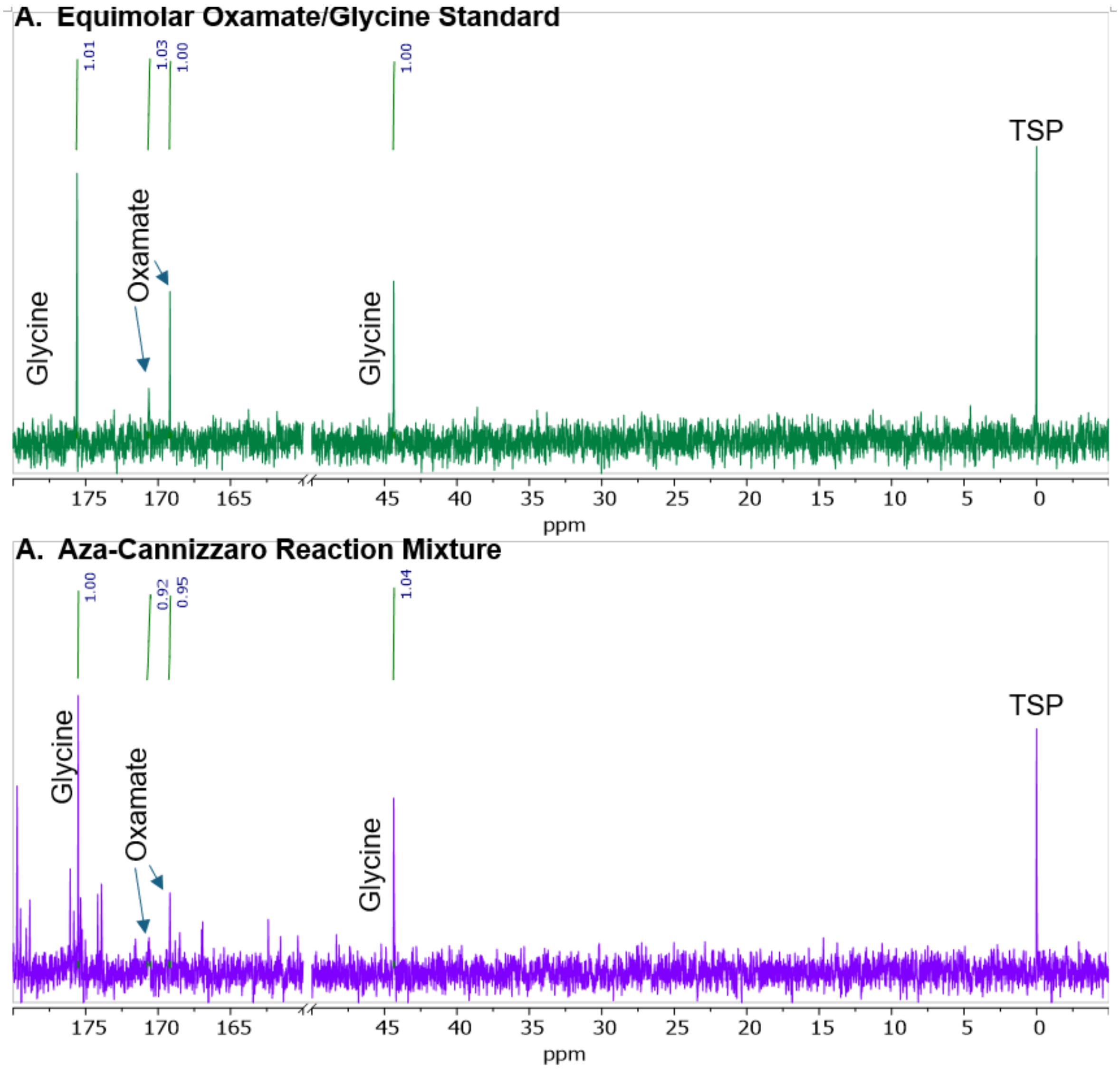


Figure 2.14 Glycine and oxamic acid are formed in comparable quantities from glyoxylate in ammonium phosphate buffer. Oxamate can be difficult to detect via ^{13}C NMR because (1) neither carbon is bonded to hydrogen, (2) there is significant resonance broadening at pH 7, and (3) the carbon signals have long relaxation times. A zgig30 pulse sequence was used with 20 s relaxation delay and without nOe enhancement to probe whether oxamic acid and glycine were formed in comparable quantities, as predicted by the aza-Cannizzaro mechanistic hypothesis. (A) ^{13}C NMR data for a solution containing equimolar amounts of commercial oxamic acid and glycine in ammonium phosphate buffer, pH 7, mixed with an equal volume D_2O. The ratio of integrals was found to be approximately 1:1. (B) ^{13}C NMR data for a reaction mixture containing 500 mM each of glyoxylate and ammonium phosphate, pH 7, incubated at 50°C for 48 hr. The ratio of integrals was found to be approximately 1:0.9.

While a 1:1 ratio of glycine with oxamate is consistent with the proposed aza-Cannizzaro mechanism, we set out to collect further evidence by deuterium-labelling the glyoxylate starting material. Since the aldehyde hydrogen is predicated to migrate via hydride transfer to the aldehyde carbon (**Figure 2.12 A**), a similar reaction with deuterated glyoxylate should result in a doubly-deuterated glycine (glycine-d_2), if the mechanistic hypothesis is valid.

Labelling glyoxylate, a hygroscopic liquid at room temperature, is difficult. Several possible synthetic strategies were envisioned (**Figure 2.15**) including the partial reduction of oxalate acid in deuterium oxide, [30,31,32] H/D exchange on a 1,3-dithiane, and $NaIO_4$ cleavage of labelled tartaric acid. Tartaric acid cleavage was immediately ruled out because of the prohibitive cost of fumarate-d_2. Synthesis of deuterated fumarate would be laborious. We therefore first tried the partial reduction of oxalic acid. Oxalic acid was co-evaporated several times with deuterium oxide to ensure complete exchange, and the product was dissolved in cold D_2O with dry nitrogen bubbling through it. Partial reduction

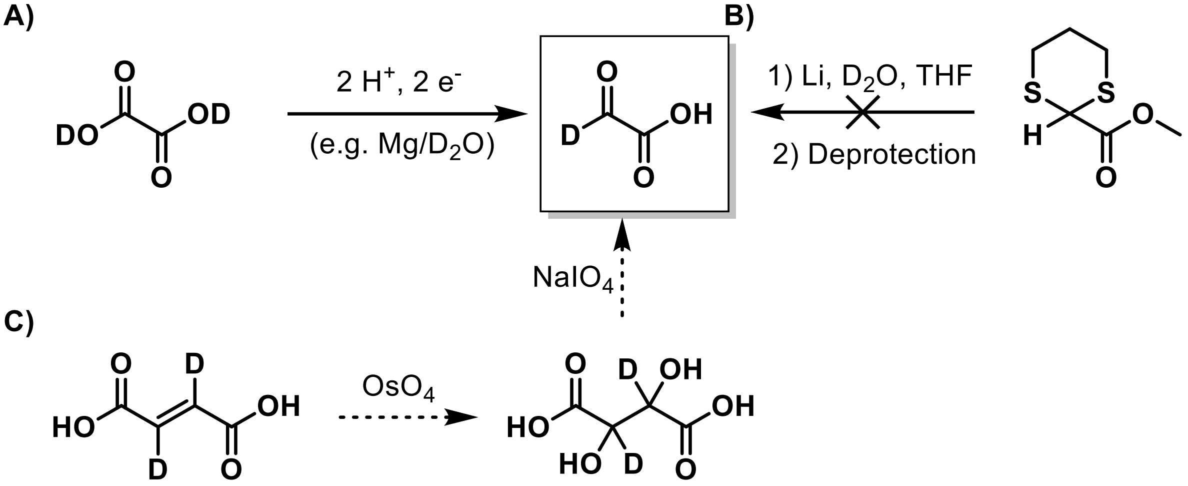


Figure 2.15 Strategies to prepare deuterium-labelled glyoxylate: (A) partial reduction of deuterated oxalic acid either through electrochemistry, sodium amalgam, or magnesium in deuterium oxide. These strategies were successful at producing the desired product with varying amounts of glycolate-d_2 as a byproduct. (B) Umpolung strategy of first replacing a proton on a dithiane with deuterium and simultaneous ester deprotection. While the first step was successful, the dithiane proved difficult to remove. (C) A route starting with the Upjohn dihydroxylation of deuterium-labelled fumarate followed by cleavage with periodate. This route was avoided due to the high cost of the starting material and the hazards of working with osmium tetroxide.

was attempted with sodium amalgam, but the reaction was inefficient: just a slight increase in temperature resulted in over-reduction into deuterated glycolate. Electrochemical reduction resulted in the same problems—small temperature fluctuations tend ruin the experiment. Furthermore, electrochemical reduction releases heavy metal ions from the electrodes into the solution, which can catalyze the traditional Cannizzaro reaction. The dithiane route showed promise initially; H/D exchange and deprotection were accomplished simultaneously and cleanly with lithium metal in THF/D_2O followed by neutralization with DCl. The dithiane-protected glyoxylate-d was extracted with DCM. Removal of the dithiane, however, proved challenging. Neither mercury-assisted cleavage (HgO, $HgCl_2$, HgTFA/polystyrene), oxidation (NBS, chlorine, H_2O_2/I_2) nor alkylation (MeI, which was kept *far away* from the mercury salts!) resulted in any detectable product formation.[33–36] Luckily, an alternative, mercury-free route was found using magnesium as the partial reductant.[37] While this method resulted in low concentrations of glyoxylate-d, it was good enough to provide clean product and was safer (with adequate ventilation to remove deuterium gas) than the alternative routes for large scale reactions (**Figure 2.16**).

After deuterated glyoxylate was produced, the solution was mixed with ammonium phosphate, the pH was corrected to 7 with ammonium hydroxide, and the mixture was heated for 48 hours at 50°C. Deuteride transfers are very slow due to a large kinetic isotope effect. After lyophilization, the resulting glycine-d2 was derivatized with Fmoc-OSu in pyridine, and the product was isolated using RP-HPLC. When compared to an authentic sample of Fmoc-Gly-OH, the Fmoc-Gly-d2-OH was missing the ^{1}H NMR signal

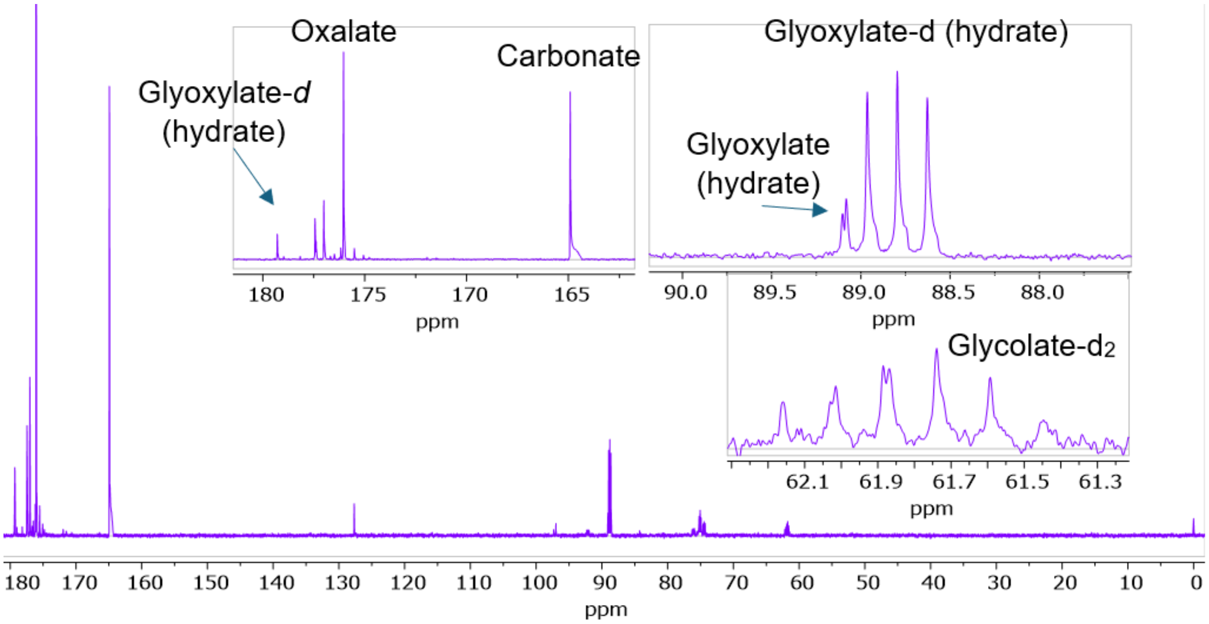


Figure 2.16 ^{13}C NMR (200 MHz) analysis of glyoxylate-d prepared via magnesium reduction of oxalic acid-d$_2$. Insets for selected regions are provided. These data show that the sample contains oxalate (unreacted starting material) and glycolate-d$_2$ (by-product from over-reduction).

at 3.5 ppm. This signal could, however, be observed with ^{2}H NMR. The evidence from ^{13}C NMR was even more striking; the α-carbon resonance almost entirely disappeared due to the lack of nOe enhancement and split into a quintet because deuterium has a nuclear spin of one (**Figure 2.17**). HRMS data were consistent with glycine-d_2 (**Figure 2.18**).

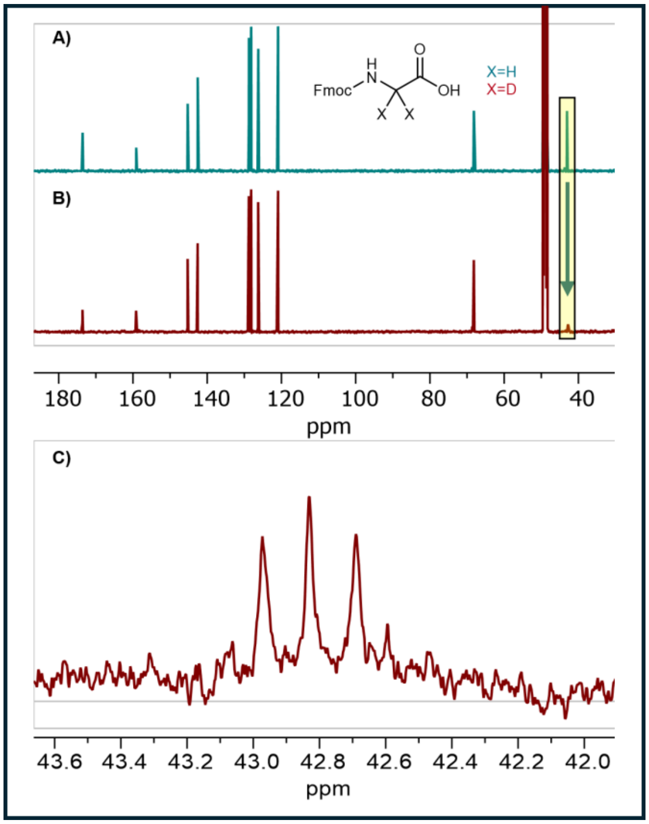
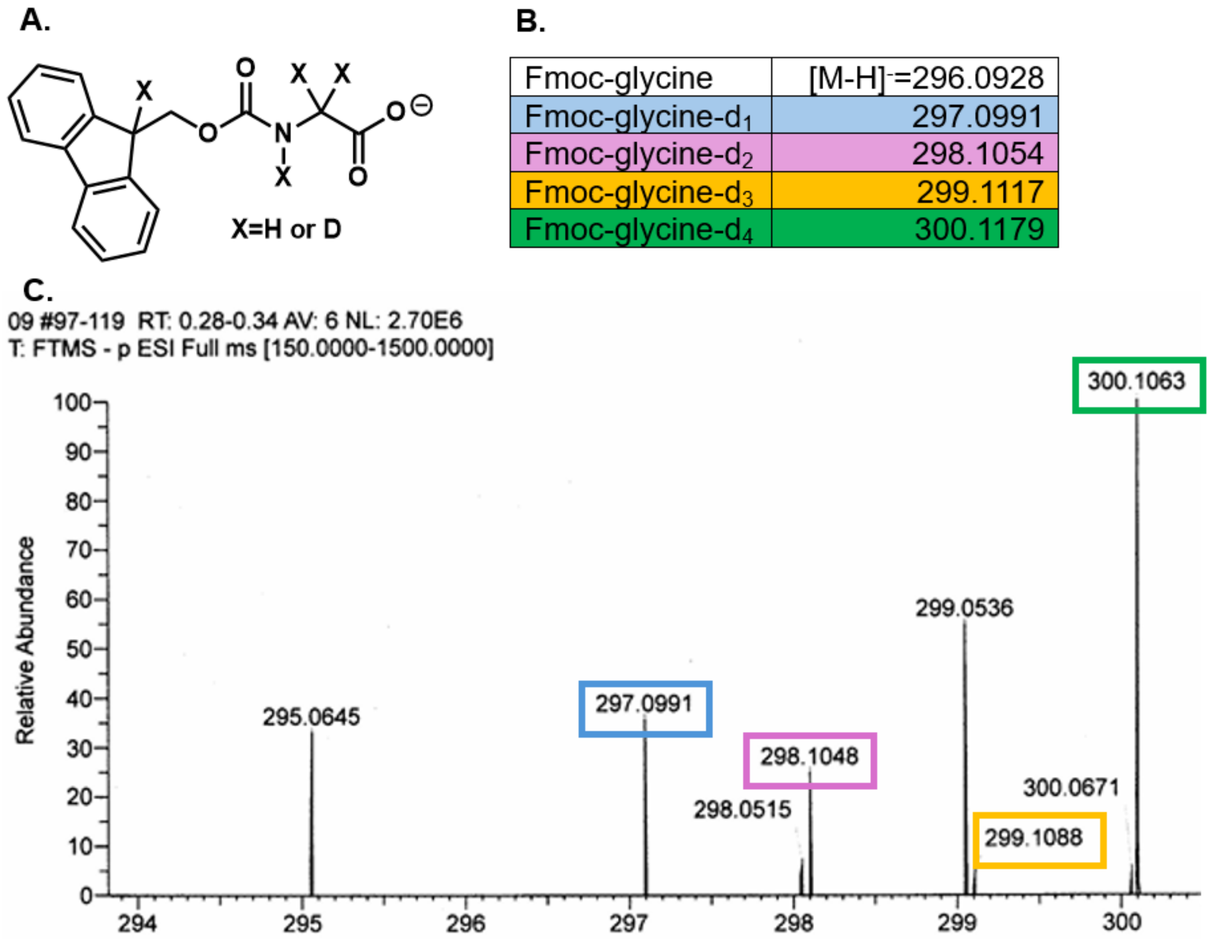


Figure 2.17. Representative ^{13}C NMR (200 MHz) spectra. A) commercial Fmoc-Gly-OH in methanol-d_4. B) Fmoc-Gly-OH isolated from a reaction mixture that was generated from glyoxylic acid-d. Note attenuation of the α-carbon signal in B relative to A (highlighted in yellow), which is consistent with the predicted lack of nOe enhancement. C) Expansion of the ^{13}C NMR signal near 42.8 ppm for Fmoc-Gly-OH formed from glyoxylic acid-d. This signal appears to be split into a quintet, as expected if there are two D atoms attached.



Fmoc-glycine	$[M-H]^-$=296.0928
Fmoc-glycine-d_1	297.0991
Fmoc-glycine-d_2	298.1054
Fmoc-glycine-d_3	299.1117
Fmoc-glycine-d_4	300.1179

Figure 2.18 Mass spectrum of Fmoc-glycine-d_x. High resolution ESI M- mass spectrum showing multiple species including Fmoc-glycine-d_x from the aza-Cannizzaro reaction with glyoxylic acid-d as described in the main text. (A) Chemical structure of Fmoc-glycine anion showing four possible positions of deuteration (X). (B) Exact masses calculated for Fmoc-glycine anion containing one, two, three or four deuterium atoms. (C) Fmoc-glycine containing one, two, three or four deuterium atoms detected via mass spectrometry. The species containing zero deuterium atoms was not detected.

Because the glyoxylate imine/hemiaminal species are involved in the rate-limiting hydride transfer step of this reaction, we predicted the reaction would be second order in glyoxylate concentration. We previously noted that glycine formation proceeded slowly at low concentrations of glyoxylate. The data in **Figure 2.19** show the effect of starting glyoxylate concentration on the reaction outcome after 48 hr. When the starting glyoxylate concentration was 50 mM, the ^{1}H NMR resonance for glyoxylate hydrate was stronger than the resonance for glycine after 48 hr. When the starting glyoxylate concentration was

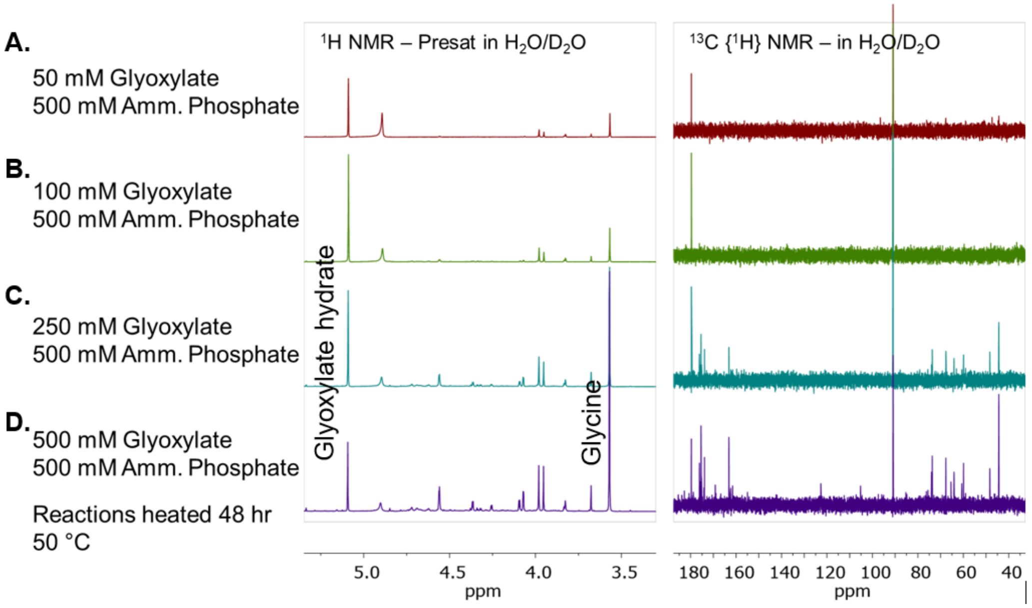


Figure 2.19 The aza-Cannizzaro reaction is sensitive to the starting concentration of glyoxylate. Glyoxylate at (A) 50 mM, (B), 100 mM, (C), 250 mM, or (D) 500 mM in ammonium phosphate (500 mM, pH 7) was incubated at 50°C for 48 hr. Reaction aliquot combined with an equal volume of D_2O prior to analysis. Glycine formation was observed in all samples via ^{1}H NMR-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard) and ^{13}C NMR spectroscopy (125 MHz, zgpg30; relaxation delay=2s). Starting material was also observed in each sample.

500 mM, however, the glycine resonance was dominant after 48 hr. This difference demonstrates that the reaction rate is heavily dependent on glyoxylate concentration, which is what one would expect for a second-order reaction. Measurement of initial rates of glyoxylate consumption via ^{1}H NMR provided further evidence of a second-order dependence on glyoxylate concentration, which is consistent with our mechanistic hypothesis (**Figure 2.20**). However, it should be noted that multiple reactions occur concurrently, and this complexity may influence the results of this initial rate investigation.

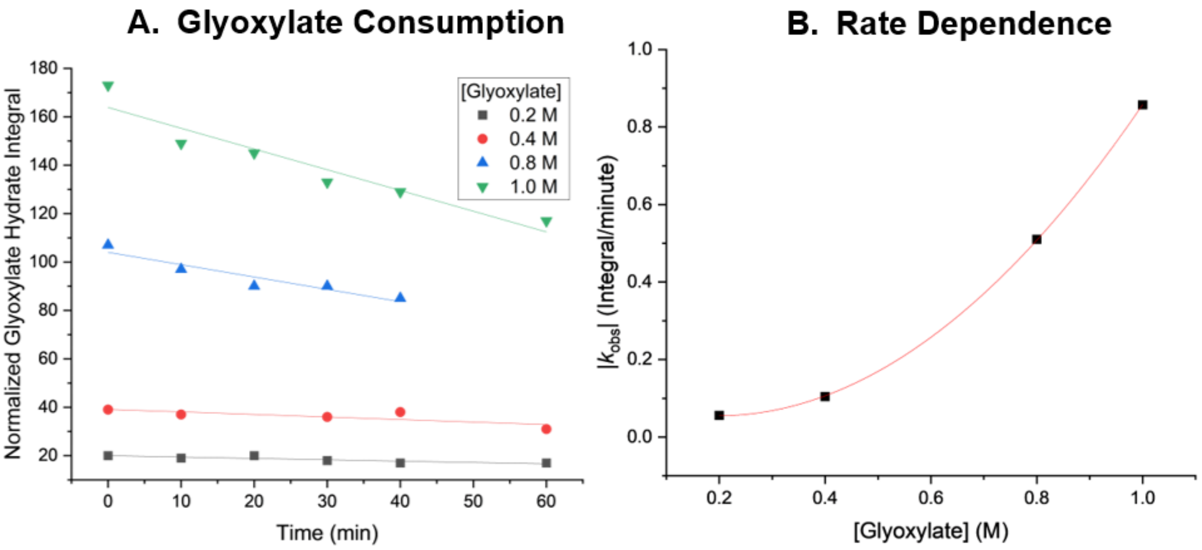


Figure 2.20 Second-order glyoxylate consumption under aza-Cannizzaro reaction conditions. Dependence of initial rate of glyoxylate consumption on [glyoxylate] in the aza-Cannizzaro reaction at pH 7 and 50°C. Aliquots were acquired throughout the first hour and spiked with a constant concentration of 3-(trimethylsilyl)propionic acid-d_4 as an NMR standard. A) The normalized glyoxylate hydrate integral (δ5.10 ppm) plotted as a function of time. B) The rate of glyoxylate consumption exhibits a second order dependence on concentration (2^{nd} order polynomial fit; $y=0.10258-0.48801x+1.24354x^2$) $R^2=0.99998$).

Interestingly, at elevated pH, where both the traditional and aza-Cannizzaro routes are plausible options, aza-Cannizzaro products predominate while only trace amounts of oxalate and glycolate are observed. This observation suggests that the aza-Cannizzaro reaction outcompetes the traditional Cannizzaro reaction (**Figure 2.12 C**). One likely explanation for this preference is that at high ammonium concentrations, the iminium/hemiaminal are predominant, which would mean that there is little of the hydrate/aldehyde pair required for a conventional Cannizzaro reaction.[5]

The trace amounts of *N*-oxalylglycine we observed via NMR indicate that the process proposed by Yanagawa may occur simultaneously with the aza-Cannizzaro reaction, albeit to a limited extent. However, formation of *N*-oxalylglycine does not generate glycine under our conditions, because hydrolysis of *N*-oxalylglycine requires strong acid. We determined that *N*-oxalylglycine is stable to hydrolysis in ammonium phosphate buffer at 50°C for over 5 days (**Figure 2.21**). Furthermore, Yanagawa reported that the highest *N*-oxalylglycine yields occurred at low pH, while the aza-Cannizzaro reaction described here is most favorable at pH>7.

Sud et al. used the term “aza-Cannizzaro” to describe a reaction of ethyl glyoxylate in 2017.[25] We believe that this designation was inappropriate. The mechanism proposed by Sud et al. featured an intramolecular 1,3-sigmatropic hydride shift, which is forbidden according to orbital symmetry.[28] Their own computational results suggested that this 1,3 shift is less favorable than a 1,2 shift followed by proton transfer. Neither of these mechanistic pathways corresponds to an analogue of the accepted Cannizzaro reaction mechanism. Thus, we assert that our experiments have identified the first example of an authentic aza-Cannizzaro reaction.

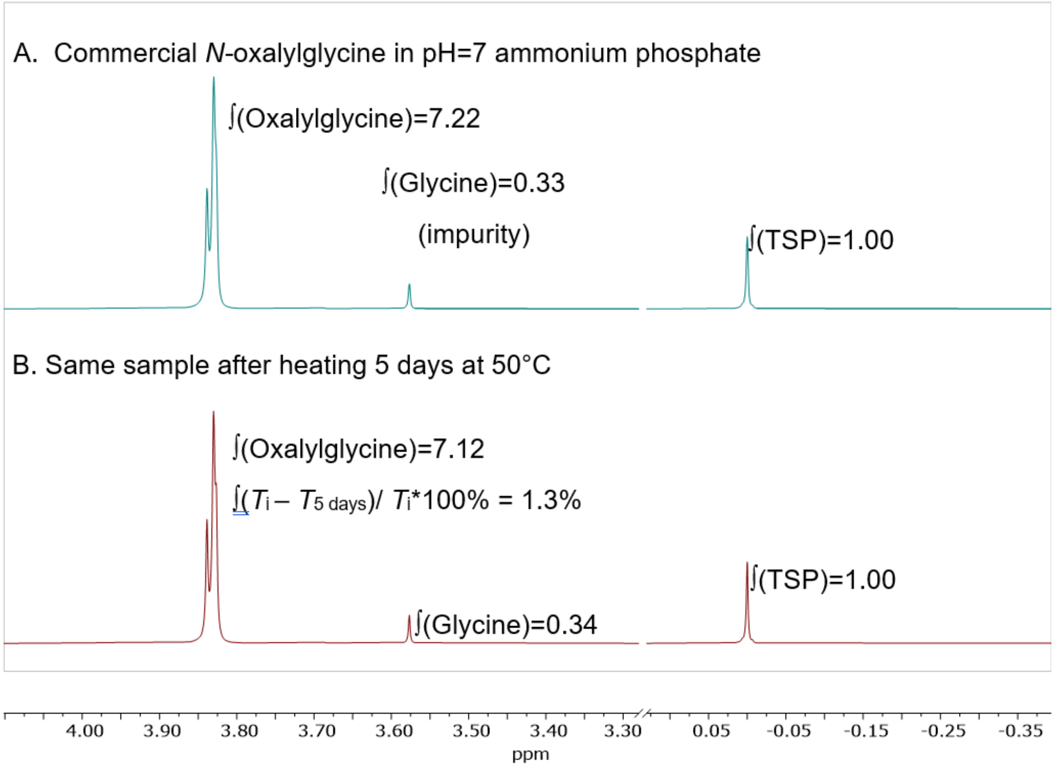


Figure 2.21 Oxalylglycine is stable to hydrolysis in pH 7 ammonium phosphate buffer at 50°C. (A) A commercial sample of oxalylglycine was dissolved in 500 mM ammonium phosphate buffer, pH 7. An aliquot was mixed with an equal volume of D_2O containing TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) as a standard, and an ^1H-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR spectrum (500 MHz) was acquired. (B) The oxalylglycine solution was incubated at 50°C for 5 days, and then another aliquot was analyzed after mixing with an equal volume of D_2O containing TSP. Only 1.3% of the oxalylglycine underwent hydrolysis under these conditions, based on the normalized integral. These findings support our hypothesis that glycine formed upon heating glyoxylic acid in ammonium phosphate buffer arises via an aza-Cannizzaro mechanism, rather than an alternative mechanistic path involving oxalylglycine formation followed by hydrolysis.

Numerous other research groups published reactions that were related to the aza-Cannizzaro reaction, but did not achieve this reactivity.[4,21–23,25,27] One notable example is the recent description by Krishnamurthy's group of a Bucherer-Bergs mechanism for glycine formation starting with glyoxylate and cyanide.[27] When we added potassium cyanide to glyoxylate and pH 7 ammonium phosphate, the aza-Cannizzaro reaction was suppressed; glycine formation was not observed (**Figure 2.22**). We hypothesize that cyanohydrin formation is both more favorable and more rapid than hemiaminal formation. This hypothesis is consistent with the observation that cyanide inhibits the cyanoborohydride reduction of 2-oxoacids.[5]

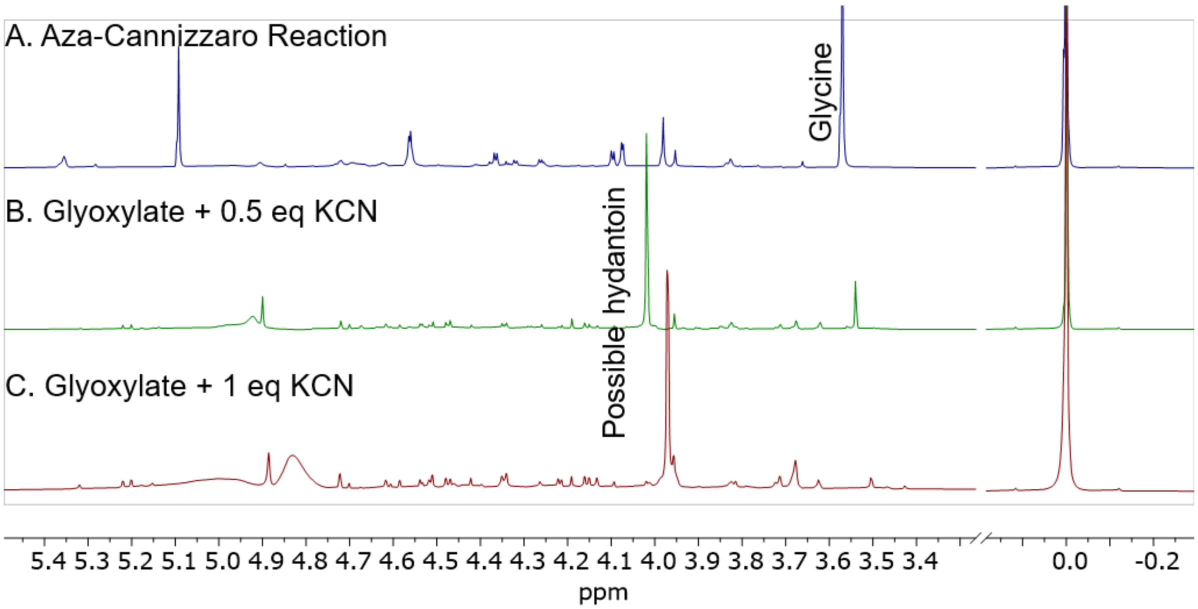


Figure 2.22 Cyanide inhibits the aza-Cannizzaro reaction. Glyoxylate (500 mM) in ammonium phosphate (500 mM, pH 7) was heated to 50°C for 48 hr with: (A) no additive, (B) 0.5 eq KCN, or (C) 1 eq KCN. Reaction aliquot combined with an equal volume of D_2O prior to analysis. ^{1}H-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR data (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt) standard) show that glycine formation was suppressed in the cyanide-containing solutions.

2.3.3. Reaction with Alkyl Amines

We wondered from a chemical standpoint if conducting the aza-Cannizzaro experiment in alkylammonium phosphates rather than ammonium phosphate would result in the formation of *N*-alkylglycine. We tested this hypothesis first by preparing a 0.5 M solution of ethylammonium phosphate by bubbling ethylamine gas through a solution of 0.5 M phosphoric acid and 0.5 M glyoxylic acid until neutralized. This solution was incubated at 50°C overnight and analyzed via ^{1}H NMR. Although the ethylamine signals complicate analysis, *N*-ethylglycine can be clearly observed (**Figure 2.23**).

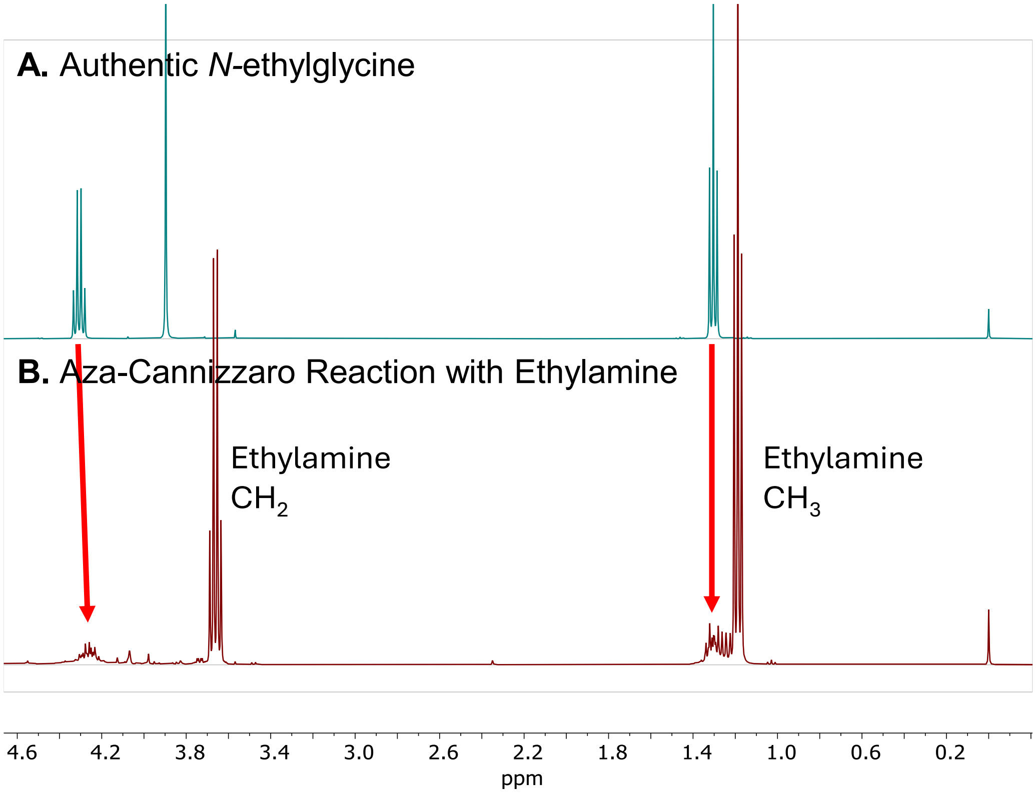


Figure 2.23 *N*-ethylglycine can be formed by conducting the aza-Cannizzaro reaction in ethylammonium phosphate. A) Commercial sample of *N*-ethylglycine. B) ^{1}H NMR of reaction mixture resulting from glyoxylate (500 mM) in ethylammonium phosphate (500 mM, pH 7) that was heated to 50°C for 24 hr. Reaction aliquot combined with an equal volume of D_2O prior to analysis. ^{1}H-presat (selective pulse at 4.7 ppm to attenuate HOD resonance) NMR data (500 MHz; pulse sequence=zgpr30, TSP (3-(trimethylsilyl)propionic-2,2,3,3-d$_4$ acid. While the largest signals are attributed to ethylammonium, *N*-ethylglycine can still be observed.

2.3.4. Implications of the Aza-Cannizzaro Mechanism

While Cannizzaro reactions have been considered in prebiotic systems (especially because of their competition with the formose reaction),[38,39] a nitrogenous analogue (viz. the aza-Cannizzaro reaction) has not been previously reported to our knowledge. Glycine is the simplest proteinogenic amino acid and can serve as a prebiotic feedstock molecule, reacting with α-ketoacids to form more complex α-amino acids via transamination.[16–18] Because transamination reactions with glycine regenerate one equivalent of glyoxylate, one can imagine that the aza-Cannizzaro reaction in conjunction with other reactions like transamination could have allowed prebiotic assimilation of nitrogen from ammonia into the building blocks of life.[40] While some highlight the unknown availability of ammonia on early earth, ammonium-containing minerals are fairly abundant and can result in high local concentrations, which might make aza-Cannizzaro reactivity plausible as a prebiotic process.[41]

2.4. Conclusion

In an attempt to understand the reactivity of pyruvate and glyoxylate in ammonium buffers, we chanced across an unprecedented reaction mechanism that is a nitrogenous analogue of the Cannizzaro reaction. In this aza-Cannizzaro reaction, glyoxylic acid in ammonium-rich aqueous solutions spontaneously forms glycine; other products form under these conditions as well, presumably via other mechanisms. Because glycine can undergo transamination to form other α-amino acids and produce glyoxylate in the process,[16–18] one can envision linked cycles of reactivity allowing nitrogen to be nonenzymatically incorporated into more complex molecules.[40]

2.5. Experimental

Water was purified to 18.2 MΩ with a Thermo Barnstead Nanopure system. Reagents and their origins: glyoxylic acid (50% in water, AlfaAesar), glyoxylic acid monohydrate (98%, Thermo), ammonium phosphate dibasic (reagent grade, Ward's), ammonium phosphate monobasic (lab grade, Ward's), ammonium hydroxide (30%, ACS grade, Sigma Aldrich), *N*-oxalylglycine (>98%, Santa Cruz Biotechnologies), glycine (electrophoresis grade, IBI), glycolic acid (ACS, Sigma Aldrich), oxalic acid anhydrous (98%, Thermo), oxamic acid (>95%, TCI), FmocOSu (ChemImpex), deuterium oxide (99.8%, Acros Organics), magnesium (ACS, Sigma Aldrich), sodium phosphate monobasic (BioXtra, Sigma Aldrich), sodium phosphate dibasic (ACS, Sigma Aldrich), ammonium acetate (molecular biology grade, Sigma Aldrich), ammonium sulfate (molecular biology grade, Sigma Aldrich), potassium cyanide (ACS, Sigma Aldrich), pyridine (ACS, Sigma Aldrich), methanol, (HPLC, Fisher), N-ethyglycine (TCI), ethylamine (Sigma Aldrich) and phosphoric acid (85%, Sigma Aldrich).

NMR spectra were collected on a Bruker Neo 500 MHz spectrometer with a Prodigy cryoprobe, or, for measurements with deuterium-labelled samples, on a Bruker Avance 600 MHz spectrometer with a TCI-F cryoprobe. NMR samples were prepared by mixing equal volumes of the reaction mixture and D_2O. 3-(Trimethylsilylpropionic-2,2,3,3-d_4 acid sodium salt ("TSP," 98%, Sigma Aldrich) was used as a chemical shift reference. In quantitative studies, TSP was added prior to the reaction and used as an integration standard to account for possible evaporation or dilution when adding ammonium hydroxide. Preparative HPLC was performed on an Agilent 1260 Infinity II instrument with a Waters X-Select C18 CSH column (19x250 mm) with MeCN/water+0.1%TFA mobile phase. pH measurements were acquired on a Mettler AB15 instrument. Lyophilization was carried out on a Labconco FreeZone instrument. Most reactions were carried out in a New Brunswick Scientific temperature-controlled shaker.

Conversion of glyoxylate to glycine and oxamic acid (the aza-Cannizarro Reaction). Glyoxylic acid monohydrate (230 mg, 2.5 mmol, 0.5 M, 1 eq) was dissolved in 5 mL aqueous ammonium buffer (pH=7, 0.5 M ammonium phosphate, unless otherwise specified). Ammonium hydroxide (30% in water) was added dropwise until the pH returned to 7.0. The reaction vessel was capped and incubated for 48 hr at 50°C to produce glycine (10% conversion), oxamate (10%), and other products.

The Cannizzaro Reaction of glyoxylate. Glyoxylic acid monohydrate (230 mg, 2.5 mmol, 0.5 M, 1 eq) was dissolved in 5 mL aqueous sodium phosphate buffer (pH 9, 0.5 M sodium phosphate). Sodium hydroxide (10 M in water) was added dropwise until the pH returned to 9.0. The reaction vessel was capped and incubated for 48 hr at 50°C to produce glycolate and oxalate.

Preparation of glyoxylic acid-d. Oxalic acid anhydrous (4.50 g, 0.050 mol) was mixed with 10 mL deuterium oxide, and the solution was concentrated *in vacuo*. This procedure was repeated three times to provide oxalic acid-d_2. Meanwhile magnesium metal (5.00 g, 206 mmol, 4.1 eq) was added to 20 mL deuterium oxide. The mixture was cooled in an ice bath. A steady stream of dry nitrogen gas was directed into the flask. The oxalic acid-d_2 was dissolved in 50 mL D_2O, and this solution was added dropwise to the stirring magnesium slurry. The reaction vessel remained open to air to allow D_2 gas to escape. After the bubbling ceased, 6.7 mL phosphoric acid (85%) was added, and the solution was filtered to remove insoluble magnesium salts. The volume was corrected to 100 mL with nanopure water. If conversion were complete, this procedure would generate a 0.5 M glyoxylic acid-d solution. The actual concentration was estimated by ^{13}C NMR and typically found to be less than 0.2 M. The solution was used without further purification.

Conversion of glyoxylic acid-d to Fmoc-glycine-OH-d_2. 100 mL of the glyoxylic acid-d solution was mixed with 5.72 g ammonium phosphate monobasic (0.048 mol), and ammonium hydroxide (30%) was added dropwise to reach pH 7.0. This solution was incubated for 48 hr at 50°C; the solution was then frozen and lyophilized. The powder was mixed with 100 mL pyridine, 15.3 g FmocOSu (0.050 mol) was added, and the mixture was heated for 24 hr at 50°C. After cooling, the solution was filtered, and the filtrate was concentrated *in vacuo.* The dried residue was combined with 10 mL methanol, and this mixture was concentrated to 2 mL, which was passed through a 0.22 μm filter. Fmoc-Gly-d_2 was isolated via reverse-phase HPLC. The isolated fractions were concentrated *in vacuo* and analyzed via NMR in DMSO-d_6.

The aza-Cannizzaro Reaction of glyoxylate in ethylammonium phosphate. Into a solution of glyoxylic acid monohydrate (0.5 M) in aqueous phosphoric acid (0.5 M) was bubbled ethylamine gas. The pH was intermittently checked with pH strips until pH 7, at which point the ethylamine gas was ceased and the mixture was incubated for 24 hr at 50°C to produce *N*-ethylglycine.

2.6. References

(1) Rios, A. C.; Bera, P. P.; Moreno, J. A.; Cooper, G. Pyruvate Aldol Condensation Product: A Metabolite That Escaped Synthetic Preparation for Over a Century. *ACS Omega* **2020**, *5* (25), 15063–15068. https://doi.org/10.1021/acsomega.0c00877.

(2) Perkin, W. H. IX.—On Dibromacetic and Glyoxylic Acids. *J Chem Soc* **1877**, *32* (0), 90–103. https://doi.org/10.1039/JS8773200090.

(3) Debus, H. CXLI.—Contributions to the History of Glyoxylic Acid. *J Chem Soc Trans* **1904**, *85* (0), 1382–1403. https://doi.org/10.1039/CT9048501382.

(4) Hoefnagel, A. J.; Van Bekkum, H.; Peters, J. A. The Reaction of Glyoxylic Acid with Ammonia Revisited. *J. Org. Chem.* **1992**, *57* (14), 3916–3921. https://doi.org/10.1021/jo00040a035.

(5) Mayer, R. J.; Moran, J. Quantifying Reductive Amination in Nonenzymatic Amino Acid Synthesis. *Angew. Chem. Int. Ed.* **2022**, *61* (48), e202212237. https://doi.org/10.1002/anie.202212237.

(6) Clay, A. P.; Cooke, R. E.; Kumar, R.; Yadav, M.; Krishnamurthy, R.; Springsteen, G. A Plausible Prebiotic One-Pot Synthesis of Orotate and Pyruvate Suggestive of Common Protometabolic Pathways. *Angew. Chem.* **2022**, *134* (11), e202112572. https://doi.org/10.1002/ange.202112572.

(7) Eschenmoser, A. The Search for the Chemistry of Life’s Origin. *Tetrahedron* **2007**, *63* (52), 12821–12844. https://doi.org/10.1016/j.tet.2007.10.012.

(8) Marín-Yaseli, M. R.; González-Toril, E.; Mompeán, C.; Ruiz-Bermejo, M. The Role of Aqueous Aerosols in the “Glyoxylate Scenario”: An Experimental Approach.

Chem. – Eur. J. **2016**, *22* (36), 12785–12799. https://doi.org/10.1002/chem.201602195.

(9) Krishnamurthy, R.; Liotta, C. L. The Potential of Glyoxylate as a Prebiotic Source Molecule and a Reactant in Protometabolic Pathways—The Glyoxylose Reaction. *Chem* **2023**, *9* (4), 784–797. https://doi.org/10.1016/j.chempr.2023.03.007.

(10) Andersen, J. L.; Flamm, C.; Merkle, D.; Stadler, P. F. *In Silico* Support for Eschenmoser's Glyoxylate Scenario. *Isr. J. Chem.* **2015**, *55* (8), 919–933. https://doi.org/10.1002/ijch.201400187.

(11) Cleaves, H. J.; Chalmers, J. H.; Lazcano, A.; Miller, S. L.; Bada, J. L. A Reassessment of Prebiotic Organic Synthesis in Neutral Planetary Atmospheres. *Orig. Life Evol. Biospheres* **2008**, *38* (2), 105–115. https://doi.org/10.1007/s11084-007-9120-3.

(12) Eckhardt, A. K.; Bergantini, A.; Singh, S. K.; Schreiner, P. R.; Kaiser, R. I. Formation of Glyoxylic Acid in Interstellar Ices: A Key Entry Point for Prebiotic Chemistry. *Angew. Chem. Int. Ed.* **2019**, *58* (17), 5663–5667. https://doi.org/10.1002/anie.201901059.

(13) Snyder, L. E.; Lovas, F. J.; Hollis, J. M.; Friedel, D. N.; Jewell, P. R.; Remijan, A.; Ilyushin, V. V.; Alekseev, E. A.; Dyubko, S. F. A Rigorous Attempt to Verify Interstellar Glycine. *Astrophys. J.* **2005**, *619* (2), 914–930. https://doi.org/10.1086/426677.

(14) Lattelais, M.; Pauzat, F.; Pilmé, J.; Ellinger, Y.; Ceccarelli, C. About the Detectability of Glycine in the Interstellar Medium. *Astron. Astrophys.* **2011**, *532*, A39. https://doi.org/10.1051/0004-6361/201016039.

(15) Rivilla, V. M.; Sanz-Novo, M.; Jiménez-Serra, I.; Martín-Pintado, J.; Colzi, L.; Zeng, S.; Megías, A.; López-Gallifa, Á.; Martínez-Henares, A.; Massalkhi, S.; Tercero, B.; De Vicente, P.; Martín, S.; Andrés, D. S.; Requena-Torres, M. A.; Alonso, J. L. First Glycine Isomer Detected in the Interstellar Medium: Glycolamide ($NH_2C(O)CH_2OH$). *Astrophys. J. Lett.* **2023**, *953* (2), L20. https://doi.org/10.3847/2041-8213/ace977.

(16) Stubbs, R. T.; Yadav, M.; Krishnamurthy, R.; Springsteen, G. A Plausible Metal-Free Ancestral Analogue of the Krebs Cycle Composed Entirely of α-Ketoacids. *Nat. Chem.* **2020**, *12* (11), 1016–1022. https://doi.org/10.1038/s41557-020-00560-7.

(17) Herbst, R. M.; Engel, L. L. A REACTION BETWEEN α-KETONIC ACIDS AND α-AMINO ACIDS. *J. Biol. Chem.* **1934**, *107* (2), 505–512. https://doi.org/10.1016/S0021-9258(18)75369-6.

(18) Huber, C.; Wächtershäuser, G. Primordial Reductive Amination Revisited. *Tetrahedron Lett.* **2003**, *44* (8), 1695–1697. https://doi.org/10.1016/S0040-4039(02)02863-0.

(19) Plater, M. J.; Vassiliev, K. The Detection of Glycine from the Treatment of Glyoxylic Acid with Iron(II) Sulfate and Ammonia in Water. *J. Chem. Res.* **2011**, *35* (3), 129–132. https://doi.org/10.3184/174751911X12964930076809.

(20) Weber, J. M.; Barge, L. M. Iron-Silicate Chemical Garden Morphology and Silicate Reactivity with Alpha-Keto Acids. *ChemSystemsChem* **2021**, *3* (3), e2000058. https://doi.org/10.1002/syst.202000058.

(21) Paczelt, V.; Wende, R. C.; Schreiner, P. R.; Eckhardt, A. K. Glycine Imine—The Elusive *α* -Imino Acid Intermediate in the Reductive Amination of Glyoxylic Acid.

Angew. Chem. Int. Ed. **2023**, *62* (11), e202218548. https://doi.org/10.1002/anie.202218548.

(22) Yanagawa, H.; Makino, Y.; Sato, K.; Nishizawa, M.; Egami, F. A Novel Way for the Formation of α-Amino Acids and Their Derivatives in an Aqueous Medium. *Adv. Space Res.* **1983**, *3* (9), 69–74. https://doi.org/10.1016/0273-1177(83)90043-1.

(23) Yanagawa, H.; Makino, Y.; Sato, K.; Nishizawa, M.; Egami, F. Novel Formation of α-Amino Acids and Their Derivatives from Oxo Acids and Ammonia in an Aqueous Medium. *J. Biochem. (Tokyo)* **1982**, *91* (6), 2087–2090. https://doi.org/10.1093/oxfordjournals.jbchem.a133902.

(24) Koga, T.; Naraoka, H. Synthesis of Amino Acids from Aldehydes and Ammonia: Implications for Organic Reactions in Carbonaceous Chondrite Parent Bodies. *ACS Earth Space Chem.* **2022**, *6* (5), 1311–1320. https://doi.org/10.1021/acsearthspacechem.2c00008.

(25) Sud, A.; Chaudhari, P. S.; Agarwal, I.; Mohammad, A. B.; Dahanukar, V. H.; Bandichhor, R. Discovery of Redox System Enabling C–N–C Bonds Formation: Unprecedented Aza-Cannizzaro Reaction. *Tetrahedron Lett.* **2017**, *58* (19), 1891–1894. https://doi.org/10.1016/j.tetlet.2017.04.010.

(26) Taillades, J.; Beuzelin, I.; Garrel, L.; Tabacik, V.; Bied, C.; Commeyras, A. N-Carbamoyl-α-Amino Acids Rather than Free α-Amino Acids Formation in the Primitive Hydrosphere: A Novel Proposal for the Emergence of Prebiotic Peptides. *Orig. Life Evol. Biosph.* **1998**, *28* (1), 61–77. https://doi.org/10.1023/A:1006566810636.

(27) Pulletikurti, S.; Yadav, M.; Springsteen, G.; Krishnamurthy, R. Prebiotic Synthesis of α-Amino Acids and Orotate from α-Ketoacids Potentiates Transition to Extant Metabolic Pathways. *Nat. Chem.* **2022**, *14* (10), 1142–1150. https://doi.org/10.1038/s41557-022-00999-w.

(28) Swain, C. G.; Powell, A. L.; Sheppard, W. A.; Morgan, C. R. Mechanism of the Cannizzaro Reaction. *J. Am. Chem. Soc.* **1979**, *101* (13), 3576–3583. https://doi.org/10.1021/ja00507a023.

(29) Cannizzaro, S. Ueber Den Der Benzoësäure Entsprechenden Alkohol. *Justus Liebigs Ann. Chem.* **1853**, *88* (1), 129–130. https://doi.org/10.1002/jlac.18530880114.

(30) Pickett, D. J.; Yap, K. S. A Study of the Production of Glyoxylic Acid by the Electrochemical Reduction of Oxalic Acid Solutions. *J. Appl. Electrochem.* **1974**, *4* (1), 17–23. https://doi.org/10.1007/BF00615902.

(31) Ok, H.; Caldwell, C.; Schroeder, D. R.; Singh, A. K.; Nakanishi, K. Synthesis of Optically Active 3-Diazoacetylretinals with Triisopropylphenylsulfonylhydrazone. *Tetrahedron Lett.* **1988**, *29* (19), 2275–2278. https://doi.org/10.1016/S0040-4039(00)86036-0.

(32) Hopkins, F. G.; Cole, S. W. On the Proteid Reaction of Adamkiewicz, with Contributions to the Chemistry of Glyoxylic Acid. *Proc. R. Soc. Lond.* **1901**, *68* (442–450), 21–33. https://doi.org/10.1098/rspl.1901.0008.

(33) Wuts, P. G. M. *Greene’s Protective Groups in Organic Synthesis*, 1st ed.; Wiley, 2025. https://doi.org/10.1002/9781394233199.

(34) Liu, J.; Wong, C.-H. An Efficient Method for the Cleavage of P-Methoxybenzylidene (PMP), Tetrahydropyranyl (THP) and 1,3-Dithiane Protecting Groups by Selectfluor^{TM}. *Tetrahedron Lett.* **2002**, *43* (22), 4037–4039. https://doi.org/10.1016/S0040-4039(02)00740-2.

(35) Ganguly, N.; Barik, S. A Facile Mild Deprotection Protocol for 1,3-Dithianes and 1,3-Dithiolanes with 30% Hydrogen Peroxide and Iodine Catalyst in Aqueous Micellar System. *Synthesis* **2009**, *2009* (08), 1393–1399. https://doi.org/10.1055/s-0028-1088023.

(36) Yus, M.; Nájera, C.; Foubelo, F. The Role of 1,3-Dithianes in Natural Product Synthesis. *Tetrahedron* **2003**, *59* (33), 6147–6212. https://doi.org/10.1016/S0040-4020(03)00955-4.

(37) Benedict, S. R. A Note on the Preparation of Glyoxylic Acid as a Reagent. *J. Biol. Chem.* **1909**, *6* (1), 51–52. https://doi.org/10.1016/S0021-9258(18)91634-0.

(38) Omran, A.; Gonzalez, A.; Menor-Salvan, C.; Gaylor, M.; Wang, J.; Leszczynski, J.; Feng, T. Serpentinization-Associated Mineral Catalysis of the Protometabolic Formose System. *Life* **2023**, *13* (6), 1297. https://doi.org/10.3390/life13061297.

(39) Kim, H.-J.; Ricardo, A.; Illangkoon, H. I.; Kim, M. J.; Carrigan, M. A.; Frye, F.; Benner, S. A. Synthesis of Carbohydrates in Mineral-Guided Prebiotic Cycles. *J. Am. Chem. Soc.* **2011**, *133* (24), 9457–9468. https://doi.org/10.1021/ja201769f.

(40) Frenkel-Pinter, M.; Samanta, M.; Ashkenasy, G.; Leman, L. J. Prebiotic Peptides: Molecular Hubs in the Origin of Life. *Chem. Rev.* **2020**, *120* (11), 4707–4765. https://doi.org/10.1021/acs.chemrev.9b00664.

(41) Johnson, B.; Goldblatt, C. The Nitrogen Budget of Earth. *Earth-Sci. Rev.* **2015**, *148*, 150–173. https://doi.org/10.1016/j.earscirev.2015.05.006.

Chapter 3

Stereochemical Differences Modulate Phase Separation Propensities of Homochiral and Heterochiral Peptides

3.1. Abstract

Liquid-liquid phase separation (LLPS) is responsible for the formation of membraneless organelles in cells and has been suggested as a mechanism for prebiotic compartmentalization. While LLPS has been widely studied with homochiral peptides and proteins, the effect of peptide heterochirality on phase separation is received little attention. Here we investigate how the length, configuration, and sequence of cationic peptides can affect phase separation propensity with hexametaphosphate. Phase behavior was determined with microscopy and optical density measurements. We found that replacing L residues with D isomers in peptide 13-mers caused differences in LLPS propensity. Different spatial arrangement of D residues led to different behaviors. We propose more experiments that can determine whether residue identity, spacing, backbone disruptions, or a combination of these factors govern this effect.

3.2. Liquid-Liquid Phase Separation

Liquid-liquid phase separation (LLPS) is a phenomenon by which biopolymers such as peptides, proteins, and nucleic acids form membraneless compartments with dynamic environments.[1–3] While the solvent of the ‘condensed phase’ is aqueous, as is also true of the surrounding ‘dilute phase’, the increased local concentration of biomolecules within the condensed phase significantly changes molecular dynamics, chemical reactivity, and biomolecular folding propensities within these compartments.[4–7]

These compartments formed via LLPS are often referred to as droplets, coacervates, or even granules.[8] In this chapter, these terms will be used interchangeably with a preference for droplets and coacervates.

Within many cells, LLPS causes the formation of membraneless organelles (MLOs) such as the nucleolus, stress granules, germ cell granules, and P-bodies.[8,9] Many of these organelles were thought to be solid at the time of their discoveries.[9] Only in the last 20 years was it determined that these MLOs are liquid. In 2009, Brangwynne et al. demonstrated that P-granules from *C. elegans* were spherical, could fuse, allowed molecular exchange with the surrounding environment, and could dissolve and reform in response to changes in temperature or concentration.[10] This evidence that P-granules were liquid-like inspired similar experiments on other membraneless organelles and established the importance of LLPS for cellular function.[9,11–13]

Many cellular functions are facilitated by membraneless organelles that can recruit and concentrate specific biological molecules. RNA splicing and transcription, as well as ribosome assembly, are examples of functions mediated by membraneless organelles.[13–16] Interestingly, LLPS is also central to the pathogenesis of numerous illnesses, including

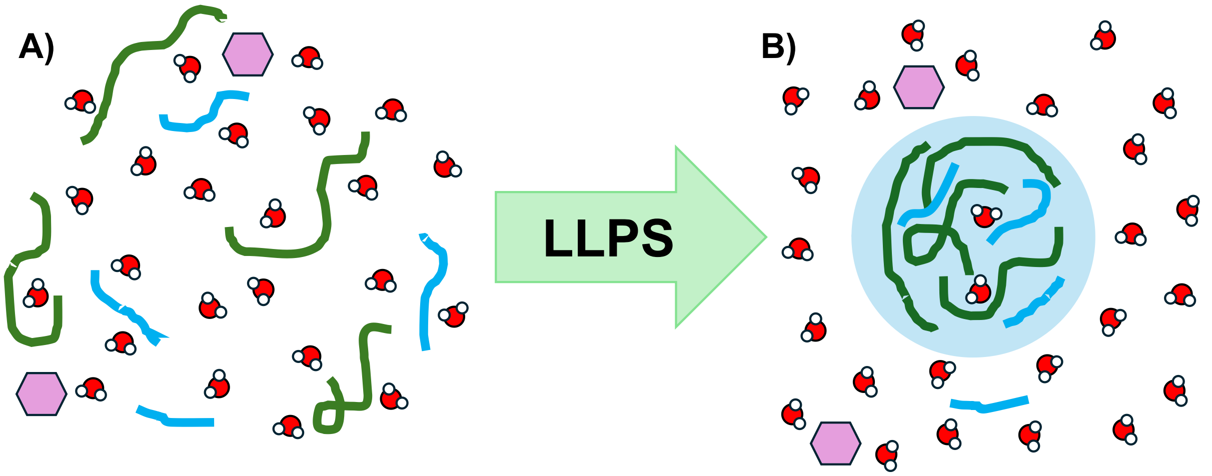


Figure 3.1 Cartoon representation of LLPS. A) Before LLPS demixing, all components are heterogeneous and solvated. B) LLPS causes molecules to become concentrated in the condensed phase. Note that that some molecules (such as the biopolymer represented by the blue line) may partition in and out of the condensed phase. Some LLPS systems have a selection bias preventing certain molecules (such as the one represented by the pink hexagons) from colocalizing into the droplet.

Amyotrophic lateral sclerosis (ALS), Alzheimer's Disease, cancer, viral infections, and autoimmune diseases.[11,15–22] While the mechanistic connection between condensate formation and disease varies, all demonstrate that LLPS is vital to cellular function, and when droplets fail to form, are disturbed or mature into a gel-like or solid form, that function is lost.[23]

3.2.1. Molecules in the Condensed Phase

LLPS occurs when a polymer-rich phase spontaneously separates from a more dilute polymer solution (**Figure 3.1**). Phase separation becomes possible when enthalpically favorable interactions between polymer molecules overcome the entropic cost of de-mixing.[24] Some biomolecules are capable of forming condensed phases by themselves in a process known as simple coacervation.[25,26] The RNA-binding Fused in Sarcoma (FUS) protein, one of the most well-studied phase-separating proteins,[27–31] displays this behavior. Simple coacervation occurs via intermolecular interactions

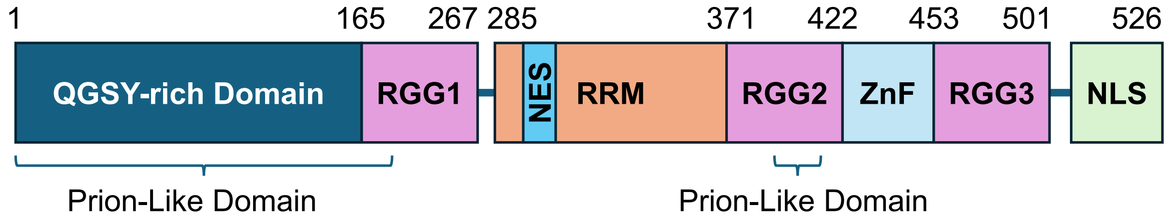


Figure 3.2 Schematic of FUS protein (1-526) showing an N-terminal region that is rich in glutamine, glycine, serine, and tyrosine (QGSY-rich domain), several arginine-glycine-glycine repetitive regions (RGG1-3), an RNA recognition motif (RMM) containing a nuclear export signal (NES), a zinc finger (ZnF), and a nonclassical nuclear positioning signal (NLS) at the C-terminus. The RGG-ZnF-RGG is thought to be responsible for RNA binding. Two prion-like domains are likely responsible for LLPS and aggregation.[28, 31]

between the tyrosine-rich prion-like domain near the FUS N-terminus and the low-complexity RGG repeats toward the C-terminus (**Figure 3.2**).These favorable contacts have been hypothesized to be cation-π interactions,[29,30] although recent evidence suggests that hydrogen bonding might be a more important factor.[32]

Complex coacervation involves two interacting molecules, often oppositely charged.[33] RNA and positively charged proteins can form electrostatic interactions that result in LLPS.[33,34] Histones and nucleic acids form nuclear condensates that assist in chromatin organization.[35] Synthetic cationic polymers and polyanions can be used as model systems for biopolymer phase separation.[25,36,37]

Formation of coacervates requires a delicate balance of molecular interactions.[13] Numerous strong noncovalent interactions will cause molecules to precipitate as a solid, whereas insufficient noncovalent interactions will prevent phase separation. Non-covalent forces governing LLPS can include charge-charge interactions, cation-π, dipole interactions, aromatic interactions, and the hydrophobic effect.[30,38] Biomolecule concentration, salt concentration, and charge ratios influence condensate formation.[39,40] For complex coacervates, if the concentration of either component is too low, phase separation will not occur. If concentrations are too high, one might observe reentrant behavior in which a condensate returns to a homogenous solution due to saturation of binding sites or charge imbalance.[41,42] Alternatively, some molecules precipitate when the concentration becomes too high.

Because most phase-separating peptides or proteins are either intrinsically disordered or contain one or more intrinsically disordered regions (IDRs), it is thought that intrinsic disorder plays a role in LLPS.[1,38,43,44] IDRs are a feature of RNA-binding proteins that unite with RNA to form membraneless organelles such as stress granules, P-bodies, the nucleolus, and germ granules.[8,13,38]IDRs do not adopt a specific conformation and often have sequences containing repeats of several amino acids, such as RGG repeats. Because of their inherent flexibility, IDRs can engage in transient interactions and can support multiple interactions simultaneously.[1,27] A sticker-and-spacer model is sometimes used to help recognize or design IDRs: regions of disorder contain 'sticky' residues prone to non-covalent attractions and flexible 'spacer' residues between them.[45–47] The RGG domain of the FUS protein is an excellent example: cationic arginine residues are dispersed amongst numerous flexible glycine residues.[30]

Molecules in a condensed phase behave differently from those in the dilute phase condensates provide a unique microenvironment.[45,48,49] While molecular dynamics might be slowed down to some extent within a condensed phase, molecular motions must remain facile if the condensate is to remain liquid. Molecules can readily diffuse into, out from, and within condensed liquid phases, although dynamics may be slow compared to bulk solution.[6,49,50] Solute diffusion is much slower from gel-like droplets or solid aggregates. Because of the crowded microenvironment within a droplet, some biomolecules exhibit different folding propensities and thus secondary structures relative to dilute solution.[51–53] The pH or pK_a of a molecule might be different in droplets relative to a dilute phase.[6,54] Reaction equilibria and enzyme activity may be altered. Selective partitioning and high local concentrations can enhance the rate of chemical reactions occurring in the condensed phase.[6,55–57]

3.2.2. Prebiotic Chemistry and LLPS

Origin of Life (OoL) researchers have long considered the possible role of LLPS in the origin of life. In 1924, Alexander Oparin proposed that life did not begin with cells bound by lipid membranes, but rather with coacervate droplets.[25,58,59] This was the first conceptualization of protocells: primordial compartments that mimic cellular behavior without 'being alive' in a conventional sense. Protocells are thought to be important for the origin of life because they provide organization, which counters intrinsic molecular tendencies toward dilution and dispersion.[60,61] Without organization, reactive molecules might never meet, and reactive networks would rarely link together.[61] Additionally, merging and dividing droplets generate populations needed for natural selection to occur. It is proposed that condensate formation leads to molecular concentration, and high local concentrations accelerate reactions.[6]

The complicated biosynthesis of lipids, complex folding proteins, and genetic encoding all pose serious challenges in terms of understanding the origin of life at a molecular level. LLPS is a physical process; it does not require lipid membranes, folded proteins, or genetic encoding. Additionally, LLPS complements the RNA world hypothesis that states life began with self-replicating RNA that was used for both information storage and catalysis. It does this by proposing routes by which RNA can be protected from hydrolysis, how RNA can be compartmentalized and still show enhanced catalytic activity relative to bulk solution, and how selection of RNA sequences and catalytic functions can occur.[62]

Perhaps one of the most attractive features of protocells formed via LLPS in the context of the origin of life is the selection bias that such condensates can display toward solute molecules. Certain solutes can be selectively recruited into condensed phases based on intermolecular interactions.[60] Selective partitioning establishes a microenvironment that is distinct from the dilute phase, reminiscent of cellular life in which living cells are out-of-equilibrium systems. A recruitment bias maintains high concentrations of specific molecules while excluding others, thereby selecting specific species or reactions. As a result, preferentially-recruited molecules may undergo chemical reactions not only with faster rates, but fewer side products due to reduced chemical competition.[6] Additionally, the molecules that drive the phase separation may participate in some selection bias. This bias is described further in section 3.2.3. and is the theme that underlies this entire chapter.

3.2.3. Peptide Chirality and LLPS

Biopolymers are typically constructed from subunits with a single absolute configuration. The ribose and deoxyribose forming the backbones of RNA and DNA are exclusively D-. In ribosomally-synthesized peptides and proteins, amino acids bearing a side chain are exclusively L-. The origin of homochirality, here referring to the one-handedness of individual monomers in large biomolecules, proves to be yet another mystery in OoL research. Presumably, ancient reactions that produced amino acids yielded racemic mixtures. Heterochiral peptides (mixed L- and D- residues) were likely common products of amino acid condensation reactions. Some processes must have favored formation or concentration homochiral peptides for living organisms to arise. LLPS is one possible mechanism by which homochiral enrichment could have occurred. Heterochiral peptides may have been selectively enriched into the condensed phase while homochiral peptides were recruited to a lesser extent, or vice versa, to allow homochiral peptides to evolve into biological systems (**Figure 3.3**). For this mechanism to be plausible, however, homochiral peptides and heterochiral peptides would need to exhibit different phase separation propensities.

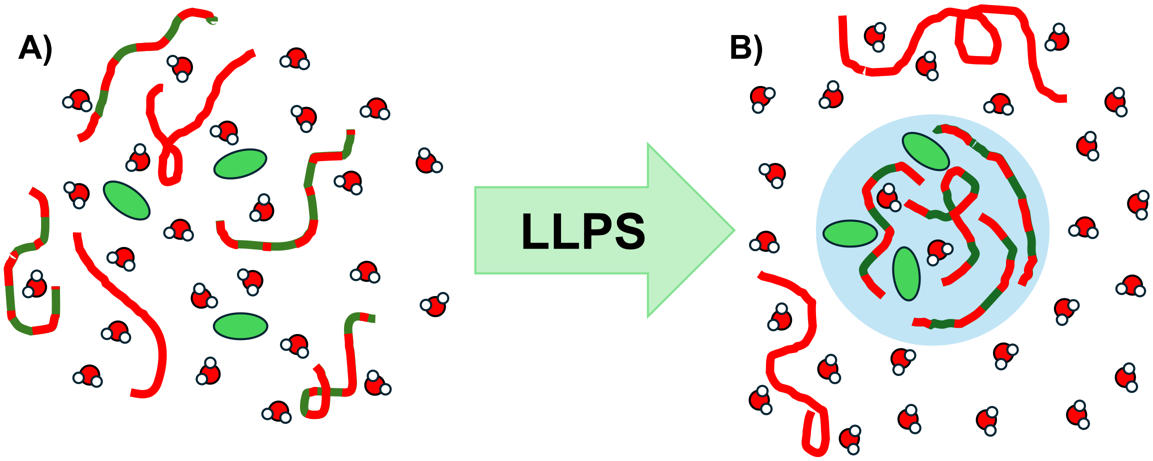


Figure 3.3 Cartoon representation of hypothetical selective LLPS. A) Both homochiral (red) and heterochiral (red/dark green) biopolymers are dispersed in solution with complex coacervation partner (green oval). B) Heterochiral biopolymers may be concentrated in the condensed phase while homochiral biopolymers are less recruited. This is a potential prebiotic method of selecting homochiral peptides.

Previous studies have found evidence that the chirality of peptides can play some role in selection and propensity in LLPS, but these studies were somewhat limited in scope, and the mechanisms by which the observed selection occurs remain unclear. Pokhrel et al. determined that poly(lysine) preferentially forms droplets with poly(G-quadruplex) nucleic acids when the chirality was matched between the two components: D- nucleic acids with poly-D-lysine.[63] Mirlohi et al. investigated the presence and morphology of droplets between homochiral and heterochiral poly(glutamate) and poly(lysine). These authors concluded that homochiral polypeptides tended to form solid aggregates while heterochiral peptide mixtures favored LLPS.[64] Similar results were described by Perry et al., who also conducted work on poly(lysine) and poly(glutamate) demonstrating that liquid coacervates are observed during the complexation of D/L polymers.[65] Seal et al. showed that homochiral and syndiotactic (alternating D/L) peptides show different reentrant behaviors with RNA.[66] Although these studies are relevant to our understanding of LLPS, the systems used were often polydisperse, and both components of complex coacervation are chiral. Few mechanistic insights have been gained from these experiments. More investigations are necessary to propose an explanation for how homo- and heterochirality translate into selective phase behavior.

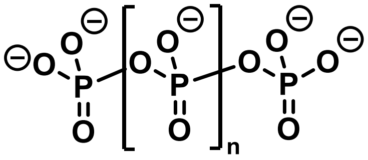
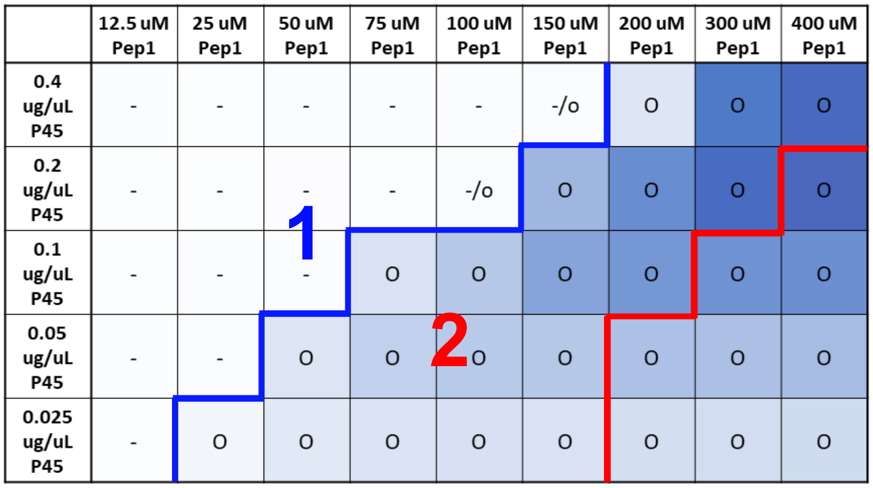


Figure 3.4 Structure of polyphosphate.

Unpublished data by van der Stok[67] revealed significant differences between homo- and heterochiral cationic peptide diastereomers when mixed with polyphosphate (polyP). In these studies, the polyP had an average length of 45 subunits. van der Stok designed an all L- peptide that adopts an α-helical conformation in solution and forms robust droplets with polyP.[67] To test the effect of intrinsic disorder on LLPS, van der Stok made several heterochiral variants in which L-alanine residues were replaced with D-alanine.[67] The helical peptide (homochiral; **1**) and one of the heterochiral diastereomers (**2**) are shown below.

`Ac-QAARAQAQRRAAEREAARRQAQRRAAEY-NH2`; “**1**”

`Ac-QAARaQAQRRAaEREAaRRQaQRRAaEY-NH2`; “**2**”



	12.5 uM Pep1	25 uM Pep1	50 uM Pep1	75 uM Pep1	100 uM Pep1	150 uM Pep1	200 uM Pep1	300 uM Pep1	400 uM Pep1
0.4 ug/uL P45	-	-	-	-	-	-/o	o	o	o
0.2 ug/uL P45	-	-	-	-	-/o	o	o	o	o
0.1 ug/uL P45	-	-	-	o	o	o	o	o	o
0.05 ug/uL P45	-	-	o	o	o	o	o	o	o
0.025 ug/uL P45	-	o	o	o	o	o	o	o	o

Figure 3.5 Phase diagram of the helical peptide AV-RRAA in 50 mM TRIS, 5 mM NaCl buffer with the phase boundary marked by a blue line. Shading represents turbidity (OD_{600}) and coacervate-forming conditions (assessed with optical microscopy) marked with "o." The phase boundary of heterochiral diastereomer AV-5DA-S is overlayed in red (OD_{600} shading of peptide 2 not shown). Peptide concentrations between 12.5 and 400 μM were tested with polyP concentrations between 0.025 and 0.4 μg/μL. Data and figure courtesy of Aevi van der Stok.

Replacing L-Ala with D-Ala destabilizes a α-helix by ~1 kcal/mol.[68–70] Thus, the heterochiral peptide **2** might be expected to display more conformational disorder than homochiral peptide **1**. As noted above, natural proteins that support condensate formation often contain large disordered domains. If polypeptide disorder is conducive to LLPS, then one would predict that heterochiral peptide **2** would have a higher propensity to support condensate formation with polyP relative to homochiral peptide **1**. It is therefore surprising that van der Stok discovered that heterochiral peptide **2** exhibited lower LLPS propensity upon mixing with polyP relative to homochiral peptide **1. Figure 3.5** helps visualize the differences in phase behavior between peptides **1** and **2**. Each column represents a different concentration of peptide **1**, while rows represent varying concentration of polyP. Darker shades of blue represent greater optical density, and cells with “o” inside were confirmed to contain coacervates using optical microscopy. A phase boundary is the line of demarcation between the conditions that result in coacervation and those that remain homogeneous. The blue line represents the phase boundary for peptide **1.** Coacervates formed under conditions on the right side of the blue line. The red line represents the phase boundary for peptide **2.** Peptide **2** therefore requires higher peptide concentrations than peptide **1** to form droplets at any tested polyP concentration. The observation that the heterochiral peptide has a lower LLPS propensity raises two possibilities: either 1) intrinsic disorder is not a defining predictor of LLPS propensity, or 2) heterochiral peptides can sometimes display greater conformational order than homochiral peptides. Both possibilities challenge our understanding of what features enable peptides to undergo LLPS by suggesting peptide stereochemistry modulates LLPS in ways not represented by current disorder-based models. Systematic studies are required to gain mechanistic insight into the effect of heterochiral residues on coacervation.

In this chapter, we will describe a series of synthetic peptides with gradually increased complexity through sequence expansion and heterochiral residues to systematically determine how heterochirality can influence LLPS propensities.

3.3. Design of Model LLPS System

Many of the current studies on heterochiral peptide LLPS utilize polydisperse polymers of one amino acid, such as poly(lysine) or poly(glutamic acid).[63,64] We chose instead to use relatively short synthetic peptides that could be prepared and studied in pure form. This approach was intended to help us elucidate relationships between peptide structure and phase separation propensity. Peptide **1** of van der Stok was studied as a pure compound, but it was designed to be α-helical and is therefore relatively long (28 amino acid residues) and contains five different types of amino acid residue.[67]

To systematically test multiple peptide parameters including length, the number of heterochiral substitutions, and sequence, as well as to maximize prebiotic relevance by minimizing complexity, we started our LLPS experiments with L-arginyl-L-arginine. Our own studies show that L-arginyl-L-arginine undergoes phase separation with polyP, and the simplicity of its sequence renders it a suitable starting point for systematic heterochiral LLPS studies. Simple peptides can be built up from that starting point using a simple sticker-and-spacer model: charged arginine residues with the ability to form electrostatic contacts (“stickers”) can be interspersed amongst alanine residues (“spacers”). RRAA repeats are an important feature in the phase-separating peptides developed by van der Stok.[67] In some of the sequences, we added tyrosine at the C-terminus as a chromophore tag to assist in the purification of these peptides and in concentration determination. The peptides described in this chapter were synthesized from Fmoc-protected amino acids using solid-phase methods with either DIC/oxyma or HBTU/NMM as coupling agents.[70]

In keeping with a minimalist design strategy, we chose the simple cyclic polyphosphate sodium hexametaphosphate (HMP) as our polyanion for complex coacervate formation. HMP is achiral, prebiotically relevant,[71,72] and is the smallest polyP with which we observed LLPS upon mixing with Arg-rich peptides. Trimetaphosphate (TMP) and tripolyphosphate (TPP) (**Figure 3.6**) each yielded homogeneous solutions when combined with various cationic species including poly-L-lysine and poly-L-arginine,

Ca^{2+} and short peptides such as Arg-Arg, Arg-Glu-Arg and Arg-Gly-Gly-Arg. HMP appears to be stable to hydrolysis: ^{31}P NMR shows that aqueous HMP solutions in pH 7 TRIS buffer are bench-stable for at least a week (**Figure 3.7**).

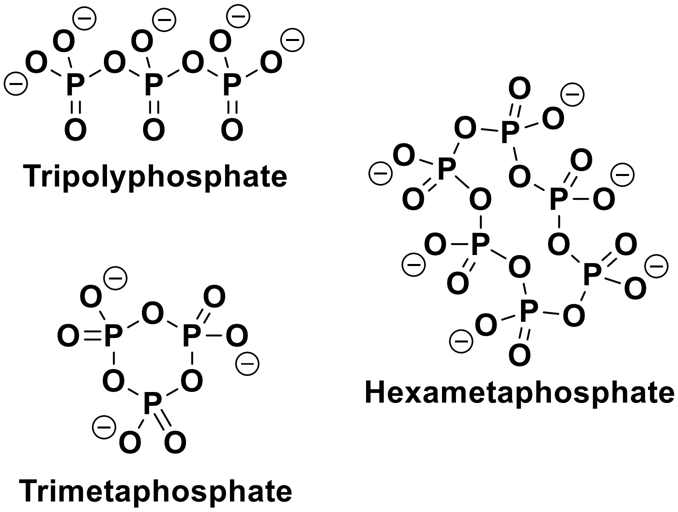


Figure 3.6 Structures of tripolyphosphate, trimetaphosphate, and hexametaphosphate.

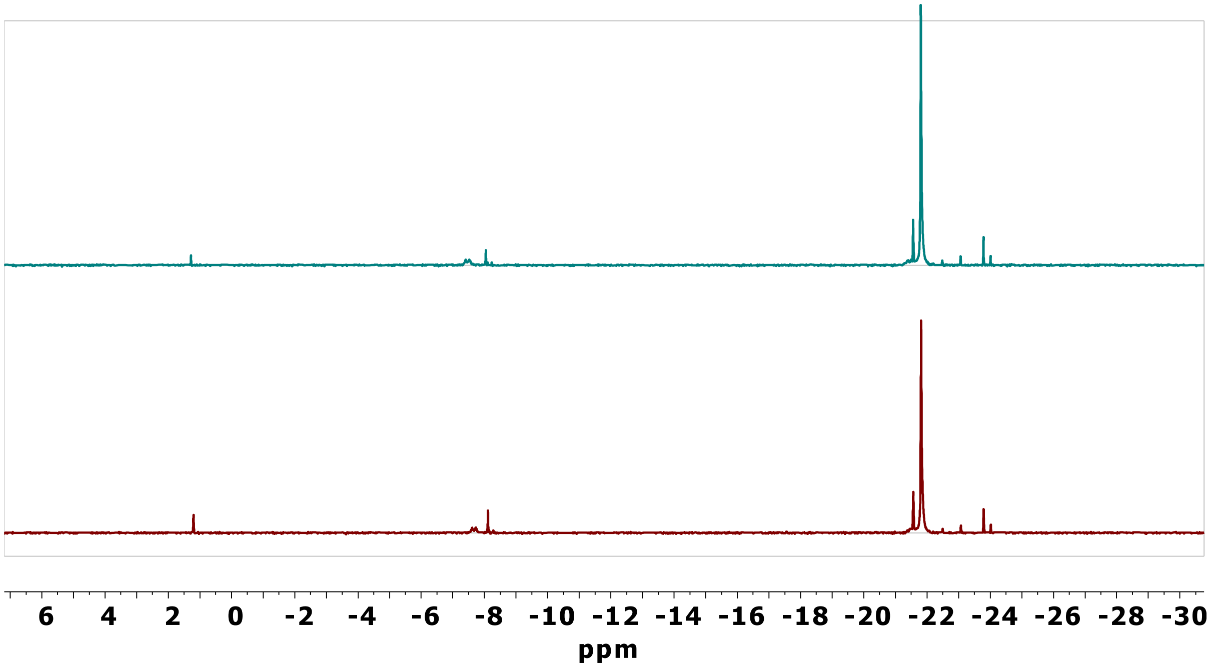


Figure 3.7 ^{31}P NMR (162 MHz) spectrum of 25 mM sodium hexametaphosphate in room temperature pH 7 TRIS buffer for A) 30 minutes and B) one week. NMR sample prepared by mixing an equal volume of HMP solution with D_2O.

3.4. Assessment of LLPS with Microscopy and Optical Density

Condensate formation induced by mixing a peptide and HMP was assessed in clear polystyrene well plates so that 96 different conditions could be screened simultaneously. Typically, we varied the concentration of HMP along one axis of the plate and the concentration of peptide along the other, seeking to identify the phase boundary, which in practice is the distinction between conditions that support LLPS and conditions that result in a homogeneous solution. For complex coacervates, there may be two phase boundaries as one of the component concentrations is varied. A lower concentration will separate the condensate region from conditions that are too dilute for LLPS, and a higher concentration will separate the condensate region from conditions that lead to reentrant behavior. To obtain reproducible results, peptide solutions must be transferred with high-quality, “low-retention” pipette tips from a reputable manufacturer (Eppendorf, Corning, Sartorius), and the number of times a solution is transferred or mixed must be consistent across experiments. Serial dilution—a methodical process in which a solution is diluted sequentially with a multichannel pipette—produces the most consistent results, and concentrations are usually varied as factors of two. Inconsistency in liquid handling technique can cause inconsistency in experimental results. These inconsistencies may be due, at least in part, to peptide adherence to polypropylene pipette tip surfaces.[73]

The presence of droplets was assessed using two different methods: microscopy and optical density measurement (**Figure 3.8**). The droplets formed by the short peptides listed in the previous section with HMP were generally large, most being approximately 5-20 μm in diameter. Droplets resulting from the LLPS of other systems are usually smaller; arginine-rich peptides and RNA produce droplets <5 μm in diameter,[29] and most

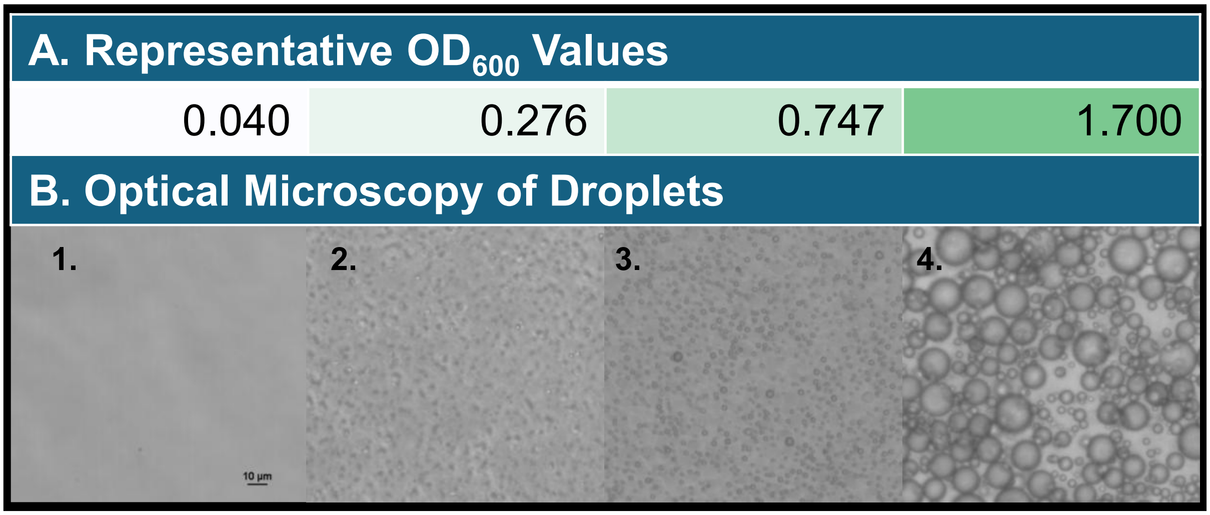


Figure 3.8 Comparison between A) optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader and B) optical microscopy on a Nikon TS2. Scale bar in the first panel is 10 µm. Data are representative from mixing HMP with the all-L peptide sequence H_2N-RRAARRAARRAAY-OH in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the following proportions: Panel 1- no droplets, 0.2 mM HMP, 0.94 mM peptide; Panel 2- Small droplets, 0.2 mM HMP, 30 mM peptide; Panel 3- Small and medium droplets, 0.4 mM HMP, 7.5 mM peptide; Panel 4- Large droplets, 12.5 mM HMP, 7.5 mM peptide.

published micrographs of FUS coacervates are even smaller.[33] Peptide/HMP droplets were easily visualized with a Nikon TS2 inverted microscope with phase contrast using either 200x or 400x magnification (**Figure 3.8 B**). Droplets could also be visualized in three dimensions using a Leica M205 stereomicroscope with up to 160x magnification. Microscopy is especially useful to confirm the liquidity of droplets by assessing whether they are capable of fusing together. Liquid droplets will fuse upon collision; gels will not.

Although droplet visualization with microscopy provides excellent insight, there are two major drawbacks to microscopy: 1) acquiring and processing images of every well in a plate can be time-consuming, especially for large experiments. This format is tedious, and the time variation between evaluation of the first and last wells can influence the data. Condensates in the last well to be photographed had more time to settle and mature than condensates in the first well. 2) Microscopy data are often qualitative. Even if micrographs are properly processed, and droplet areas are accurately measured and averaged, droplet volumes are unknown. Moreover, droplets that have not settled to the bottom of the well can be missed.

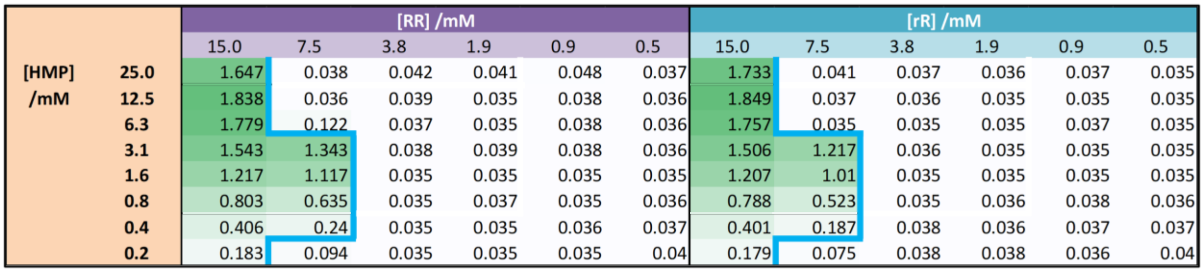
Optical density measurements at 600 nm (OD_{600}, **Figure 3.8 A**) offer the ability to acquire quantitative data for an entire plate in seconds. OD_{600} in this context is a turbidity measurement. Lower OD_{600} values indicate that light is transmitted through a sample with limited scattering, while higher values are associated with a larger extent of light scattering. 600 nm light is commonly used because few biomolecules absorb light at this wavelength.

OD_{600} measurements must be used with prudence given the limitations. Quantification based on OD_{600} values is problematic because the extent to which light is scattered is influenced in a non-linear way by the size of the objects that cause the scattering (in this case, the condensates).[74] Linearity is especially skewed when the size of the turbidity-causing particles approaches 600 nm. OD_{600} measurements must be treated as comparative rather than absolute measurements of droplet size and number. Solid aggregates can skew OD_{600} measurements by scattering light more efficiently than droplets,[74] and OD_{600} might not necessarily distinguish between large droplets and multiple small ones. Even with these limitations, OD_{600} is a useful screening tool that can save time and offer quantitative measurements in coacervation experiments that probe LLPS systems with which the experimentalist is familiar. OD_{600} increases, albeit nonlinearly, with particle size, particle number, and refractive index.[75,76] Thorough LLPS studies should have evidence from both microscopy and OD_{600}.

3.5. Length Dependence of Homo-/Heterochiral LLPS Propensity Differences

The following studies were conducted during a time at which in-house analytical abilities were limited due to the fouling of our HPLC columns. The results herein should be interpreted as provisional until replicated with peptides of established purity.

Starting with the peptide dimer Arg-Arg, we set out to determine the minimum length of peptides that support LLPS when combined with HMP for which there are differences among peptide diastereomers. We compared L-Arg-L-Arg (RR) with D-Arg-L-Arg (rR; note use of lower-case letter to indicate residues with D configuration) in mixtures with HMP (**Figure 3.9**). At 15 mM [peptide], both diastereomers exhibited robust LLPS with [HMP] between 0.4 and 25 mM. At 7.5 mM [peptide], LLPS was observed between only 0.4 and 3.1 mM HMP. No LLPS was observed at peptide concentrations <7.5 mM. We therefore failed to detect a difference between the phase boundaries of RR and rR.

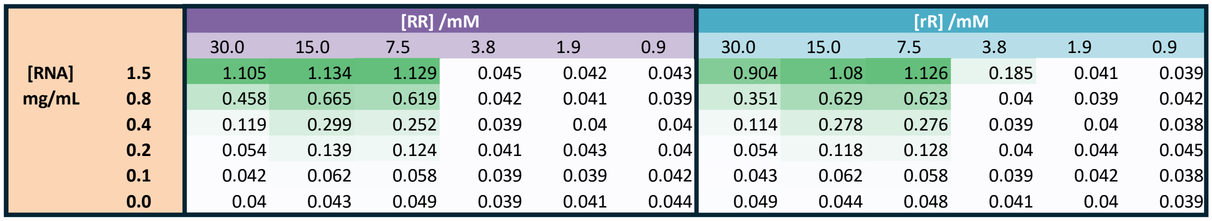


		[RR] /mM						[rR] /mM					
		15.0	7.5	3.8	1.9	0.9	0.5	15.0	7.5	3.8	1.9	0.9	0.5
[HMP] /mM	**25.0**	1.647	0.038	0.042	0.041	0.048	0.037	1.733	0.041	0.037	0.036	0.037	0.035
	12.5	1.838	0.036	0.039	0.035	0.038	0.036	1.849	0.037	0.036	0.035	0.035	0.035
	6.3	1.779	0.122	0.037	0.035	0.038	0.036	1.757	0.035	0.035	0.035	0.037	0.035
	3.1	1.543	1.343	0.038	0.039	0.038	0.036	1.506	1.217	0.036	0.035	0.035	0.035
	1.6	1.217	1.117	0.035	0.035	0.035	0.035	1.207	1.01	0.035	0.035	0.035	0.035
	0.8	0.803	0.635	0.035	0.037	0.035	0.036	0.788	0.523	0.035	0.036	0.038	0.036
	0.4	0.406	0.24	0.035	0.035	0.036	0.037	0.401	0.187	0.038	0.036	0.037	0.037
	0.2	0.183	0.094	0.035	0.035	0.035	0.04	0.179	0.075	0.038	0.038	0.036	0.04

Figure 3.9 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing HMP with the peptide sequence H_2N-RR-OH and H_2N-rR-OH in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. Presence of droplets was established with optical microscopy (not shown) prior to measurement. Blue line represents phase boundary (OD_{600} cutoff for LLPS boundary is 0.15).

Because HMP is achiral, we wondered whether rR and RR would exhibit different behavior with a chiral polyanion. We chose total yeast RNA and varied the RNA concentration between 1.5 and 0.05 mg mL^{-1}. RR and rR concentrations were varied between 30 mM and 0.9 mM. Once again, no droplets were observed when [peptide] <7.5

mM. Phase boundaries were nearly identical between the two diastereomers (**Figure 3.10**). OD_{600} varied as a function of RNA concentration.



		[RR] /mM						[rR] /mM					
		30.0	15.0	7.5	3.8	1.9	0.9	30.0	15.0	7.5	3.8	1.9	0.9
[RNA]	**1.5**	1.105	1.134	1.129	0.045	0.042	0.043	0.904	1.08	1.126	0.185	0.041	0.039
mg/mL	**0.8**	0.458	0.665	0.619	0.042	0.041	0.039	0.351	0.629	0.623	0.04	0.039	0.042
	0.4	0.119	0.299	0.252	0.039	0.04	0.04	0.114	0.278	0.276	0.039	0.04	0.038
	0.2	0.054	0.139	0.124	0.041	0.043	0.04	0.054	0.118	0.128	0.04	0.044	0.045
	0.1	0.042	0.062	0.058	0.039	0.039	0.042	0.043	0.062	0.058	0.039	0.042	0.038
	0.0	0.04	0.043	0.049	0.039	0.041	0.044	0.049	0.044	0.048	0.041	0.04	0.039

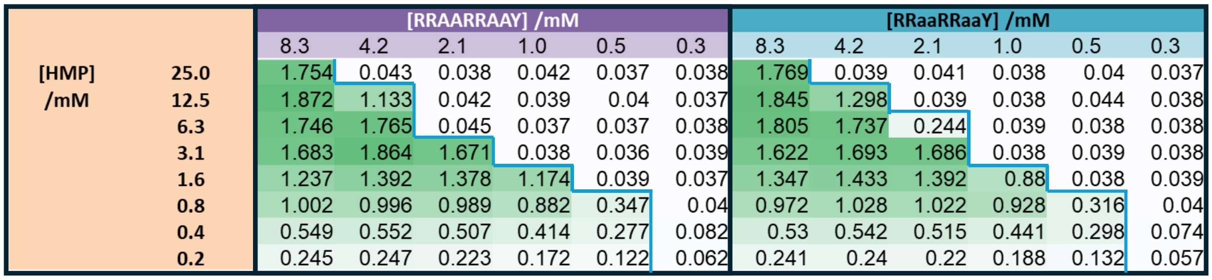
Figure 3.10 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing total yeast RNA with the peptide sequence H_2N-RR-OH and H_2N-rR-OH in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above.

After determining that diastereomers RR and rR have similar LLPS propensities when mixed with HMP or total yeast RNA, we turned to longer Arg-rich peptides. The following 9-mer peptides were synthesized:

`H2N- R R A A R R A A Y -OH;` “RRAA-9”

`H2N- R R a a R R a a Y -OH;` “RRaa-9”

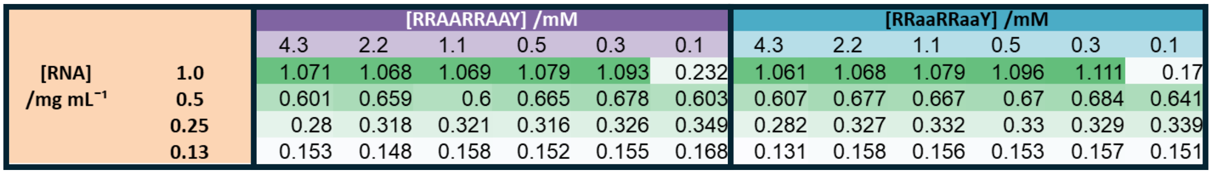
These peptides follow the typical sticker-and-spacer model and the RRAA motif is present in the heterochiral phase separating construct developed by van der Stok.[67] The terminal tyrosine is not intended to act as a ‘sticker,’ but rather as a chromophore, although we cannot rule out the possibility that the Tyr residue may play some role in LLPS. The stereochemistry of Y remained L here and in all future experiments.



		[RRAARRAAY] /mM						[RRaaRRaaY] /mM					
		8.3	4.2	2.1	1.0	0.5	0.3	8.3	4.2	2.1	1.0	0.5	0.3
[HMP] /mM	25.0	1.754	0.043	0.038	0.042	0.037	0.038	1.769	0.039	0.041	0.038	0.04	0.037
	12.5	1.872	1.133	0.042	0.039	0.04	0.037	1.845	1.298	0.039	0.038	0.044	0.038
	6.3	1.746	1.765	0.045	0.037	0.037	0.038	1.805	1.737	0.244	0.039	0.038	0.038
	3.1	1.683	1.864	1.671	0.038	0.036	0.039	1.622	1.693	1.686	0.038	0.039	0.038
	1.6	1.237	1.392	1.378	1.174	0.039	0.037	1.347	1.433	1.392	0.88	0.038	0.039
	0.8	1.002	0.996	0.989	0.882	0.347	0.04	0.972	1.028	1.022	0.928	0.316	0.04
	0.4	0.549	0.552	0.507	0.414	0.277	0.082	0.53	0.542	0.515	0.441	0.298	0.074
	0.2	0.245	0.247	0.223	0.172	0.122	0.062	0.241	0.24	0.22	0.188	0.132	0.057

Figure 3.11 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing HMP with the peptides RRAA-9 and RRaa-9 het in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. Presence of droplets was established with optical microscopy (not shown) prior to measurement. Blue line represents phase boundary (OD_{600} cutoff for LLPS boundary is 0.1).

With [HMP] between 0.2 and 25 mM and [9-mer peptide] between 8.3 and 0.3 mM, we determined that RRAA-9 and RRaa-9-het exhibit very similar LLPS phase boundaries (**Figure 3.11**). However, the 9-mer peptides formed coacervates at much lower concentrations relative to the dimers. The 9-mer peptides displayed minimum concentrations for LLPS of 0.5 mM when [HMP] < 0.8 mM. We evaluated phase separation also with total yeast RNA (**Figure 3.12**). No difference in the phase boundaries was observed between RRAA-9 and RRaa-9-het. We observed that OD_{600} remains



		[RRAARRAAY] /mM						[RRaaRRaaY] /mM					
		4.3	2.2	1.1	0.5	0.3	0.1	4.3	2.2	1.1	0.5	0.3	0.1
[RNA] /mg mL^{-1}	1.0	1.071	1.068	1.069	1.079	1.093	0.232	1.061	1.068	1.079	1.096	1.111	0.17
	0.5	0.601	0.659	0.6	0.665	0.678	0.603	0.607	0.677	0.667	0.67	0.684	0.641
	0.25	0.28	0.318	0.321	0.316	0.326	0.349	0.282	0.327	0.332	0.33	0.329	0.339
	0.13	0.153	0.148	0.158	0.152	0.155	0.168	0.131	0.158	0.156	0.153	0.157	0.151

Figure 3.12 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing total yeast RNA with the peptides RRAA-9 and RRaa-9 het in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above.

consistent at a given RNA concentration, showing little to no variation relative to peptide concentration.

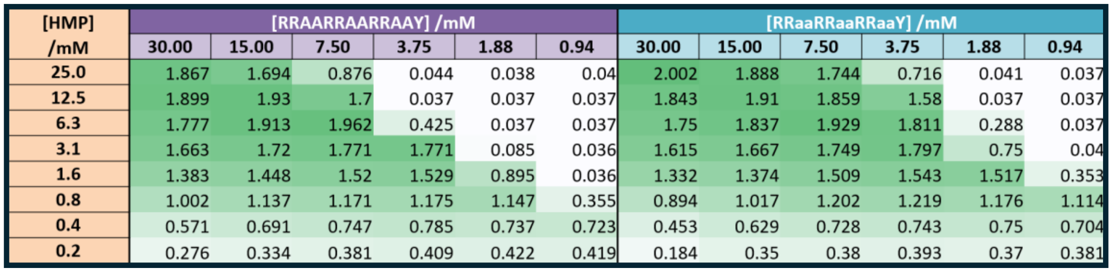
Continuing to systematically increase the peptide length, we synthesized a 13-mer RRAA-based peptide and a heterochiral diastereomer:

`H2N- R R A A R R A A R R A A Y -OH;` “RRAA-13”

`H2N- R R a a R R a a R R a a Y -OH;` “RRaa-13-het”

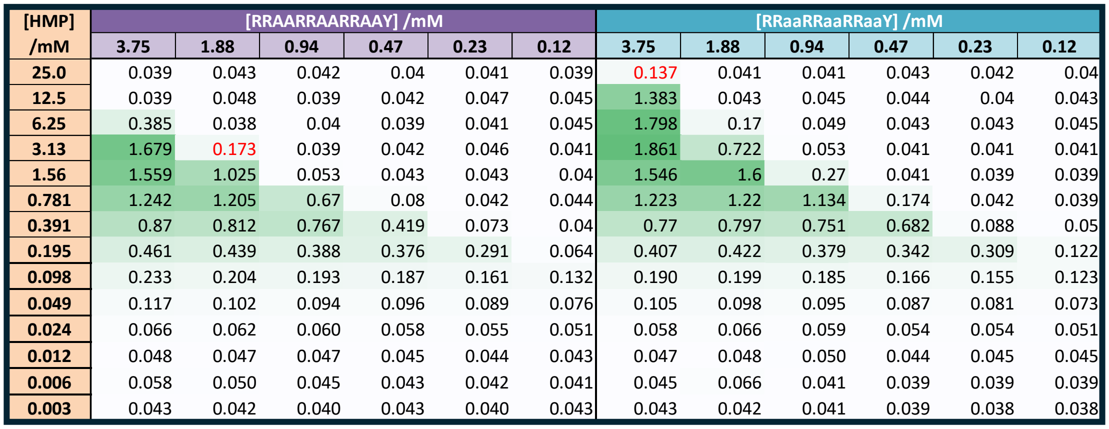
Several trials measured the phase diagrams of these 13-mer peptides from 30 mM to 0.12 mM [13-mer peptide] and [HMP] from 25 mM to 0.003 mM (**Figure 3.13**). At 25 mM hexametaphosphate, we observed droplets forming when [RRAA-13] was >7.5 mM; at the same concentration of HMP, heterochiral RRaa-13 het exhibited LLPS propensity at >3.75 mM—half the concentration as the homochiral variant. At HMP concentrations ≥0.781 mM, coacervation of RRaa-13-het occurred at approximately one half the peptide concentration relative to RRAA-13. At very low concentrations of hexametaphosphate, LLPS propensities between RRAA-13 and RRaa-13 het were more similar; both peptides had similar propensities when [HMP] <0.781 mM. This observation demonstrates that the differences in LLPS behavior relate more to reentrant behavior (the top of the phase boundary) rather than the lower concentration threshold for coacervation to occur. Optical microscopy data (**Figure 3.14**) supported the assignment of the phase boundary. These observations show that homochiral and heterochiral peptides can have different LLPS propensities upon mixing with HMP when the peptide reaches a certain length; for these RRAA peptides, that length is between 9 and 13 residues.

A. High Peptide Concentration



[HMP] /mM	[RRAARRAARRAAY] /mM						[RRaaRRaaRRaaY] /mM					
	30.00	15.00	7.50	3.75	1.88	0.94	30.00	15.00	7.50	3.75	1.88	0.94
25.0	1.867	1.694	0.876	0.044	0.038	0.04	2.002	1.888	1.744	0.716	0.041	0.037
12.5	1.899	1.93	1.7	0.037	0.037	0.037	1.843	1.91	1.859	1.58	0.037	0.037
6.3	1.777	1.913	1.962	0.425	0.037	0.037	1.75	1.837	1.929	1.811	0.288	0.037
3.1	1.663	1.72	1.771	1.771	0.085	0.036	1.615	1.667	1.749	1.797	0.75	0.04
1.6	1.383	1.448	1.52	1.529	0.895	0.036	1.332	1.374	1.509	1.543	1.517	0.353
0.8	1.002	1.137	1.171	1.175	1.147	0.355	0.894	1.017	1.202	1.219	1.176	1.114
0.4	0.571	0.691	0.747	0.785	0.737	0.723	0.453	0.629	0.728	0.743	0.75	0.704
0.2	0.276	0.334	0.381	0.409	0.422	0.419	0.184	0.35	0.38	0.393	0.37	0.381

B. Low Peptide Concentration



[HMP] /mM	[RRAARRAARRAAY] /mM						[RRaaRRaaRRaaY] /mM					
	3.75	1.88	0.94	0.47	0.23	0.12	3.75	1.88	0.94	0.47	0.23	0.12
25.0	0.039	0.043	0.042	0.04	0.041	0.039	0.137	0.041	0.041	0.043	0.042	0.04
12.5	0.039	0.048	0.039	0.042	0.047	0.045	1.383	0.043	0.045	0.044	0.04	0.043
6.25	0.385	0.038	0.04	0.039	0.041	0.045	1.798	0.17	0.049	0.043	0.043	0.045
3.13	1.679	0.173	0.039	0.042	0.046	0.041	1.861	0.722	0.053	0.041	0.041	0.041
1.56	1.559	1.025	0.053	0.043	0.043	0.04	1.546	1.6	0.27	0.041	0.039	0.039
0.781	1.242	1.205	0.67	0.08	0.042	0.044	1.223	1.22	1.134	0.174	0.042	0.039
0.391	0.87	0.812	0.767	0.419	0.073	0.04	0.77	0.797	0.751	0.682	0.088	0.05
0.195	0.461	0.439	0.388	0.376	0.291	0.064	0.407	0.422	0.379	0.342	0.309	0.122
0.098	0.233	0.204	0.193	0.187	0.161	0.132	0.190	0.199	0.185	0.166	0.155	0.123
0.049	0.117	0.102	0.094	0.096	0.089	0.076	0.105	0.098	0.095	0.087	0.081	0.073
0.024	0.066	0.062	0.060	0.058	0.055	0.051	0.058	0.066	0.059	0.054	0.054	0.051
0.012	0.048	0.047	0.047	0.045	0.044	0.043	0.047	0.048	0.050	0.044	0.045	0.045
0.006	0.058	0.050	0.045	0.043	0.042	0.041	0.045	0.066	0.041	0.039	0.039	0.039
0.003	0.043	0.042	0.040	0.043	0.040	0.043	0.043	0.042	0.041	0.039	0.038	0.038

Figure 3.13 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing HMP with the peptides RRAA-13 and RRaa-13 het in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above—(A) a high concentration measurement and (B) a low concentration measurement were both performed. Discrepancies between these trials are marked in red.

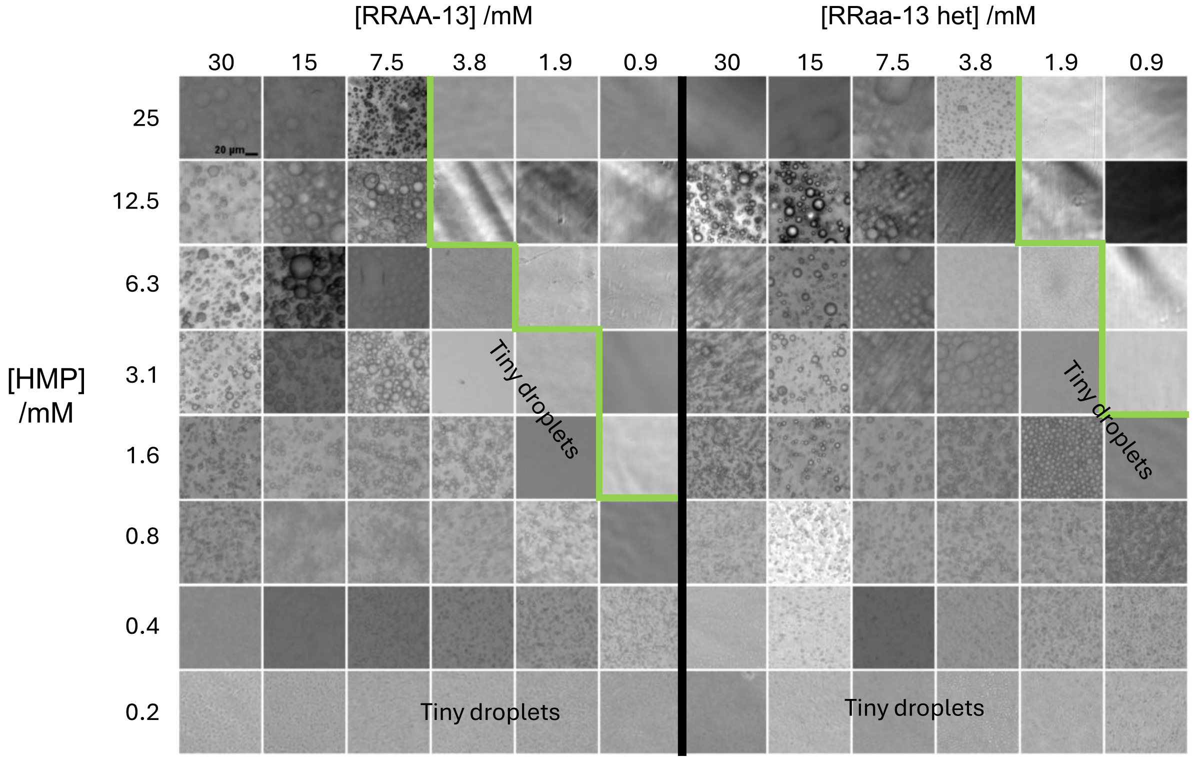


Figure 3.14 Optical microscopy images acquired on a Nikon TS2 after mixing HMP with the peptides RRAA-13 and RRaa-13 het in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. Green line represents phase boundary, supporting the OD_{600} values in **Figure 3.13**. Scale bar in first panel represents 10 µm.

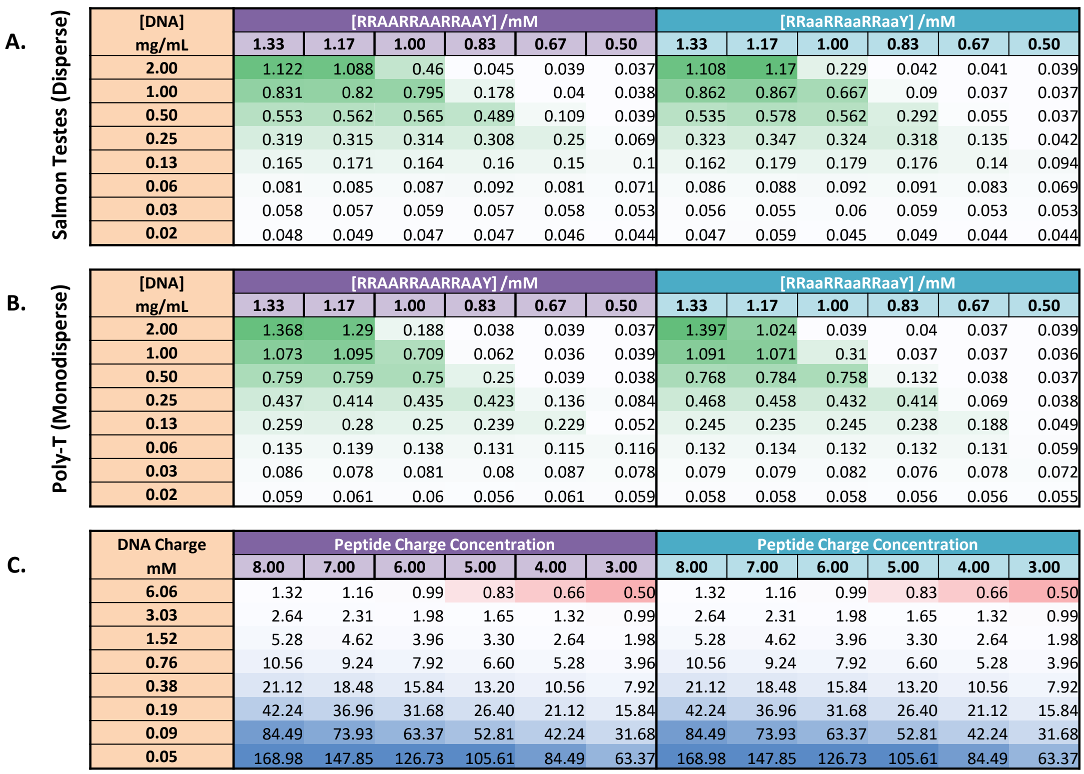
To determine whether a chiral polyanion would influence the differences in LLPS propensities between RRAA-13 and RRaa-13-het, we ran a trial using single-stranded DNA (ss-DNA) from salmon testes (as opposed to our trials with shorter peptides, in which we used total yeast RNA) as seen in **Figure 3.15 A**. We focused on the peptide concentration where we expected to see a phase boundary, here between 1.3 and 0.5 mM. When [DNA] was varied between 0.02 and 2.0 mg mL^{-1}, we did not observe any notable differences in phase boundaries but once again noticed that the OD_{600} remained constant at most peptide concentrations at a constant DNA concentration, right up to the edge of the phase boundary.

We considered multiple hypotheses to explain why OD_{600} changes as a factor of nucleic acid concentration in all of our peptide/nucleic acid LLPS experiments. One possibility we considered is that the nucleic acid is polydisperse while the length of the peptide is constant. To test this hypothesis, we ordered a defined sequence of ss-DNA—poly-T with a length of 60. Not only did we find that the trend of OD_{600} being dependent on [nucleic acid] continued with monodisperse poly-T, but we also observed that the phase behavior was very similar to the behavior in the trial containing polydisperse ss-DNA (**Figure 3.15 B**). We therefore determined that polydispersity of nucleic acids does not cause this OD_{600} effect.

Another possibility was that a large nucleic acid molecule acts as a sort of template, drawing a very specific ratio of smaller peptides to it. When there are too few peptides, coacervation does not occur. If there are too many peptides, what is not templated onto the nucleic acid remains in solution. In this situation, the size and volume of the coacervates are limited by the nucleic acid, the ‘limiting reagent,’ in a sense. This would explain why the turbidity remains nearly stable while peptide concentration is increased.

We predicted that in coacervates that are likely stabilized by charge-charge interactions (such as those between arginine side chains and phosphate), charge ratios could be useful in testing our templating hypothesis. Overall, charge ratios will always balance due to the counterions in solution as well as the buffer. In this analysis we are referring specifically to the ratios of charge concentrations (molarity of the species multiplied by the overall charge) of the biomolecules responsible for coacervation. In a situation where large nucleic acids are templating a certain number of positively-charged peptides, we would hypothesize that excess positively-charged peptides would remain in the dilute phase. Under conditions with negative charge ratios there may not be sufficient peptide for droplets to form.

When looking at relative charge ratios (**Figure 3.15 C**) between the nucleic acids and peptides, we noticed that excess negative charge prevented coacervates from forming, while positive charge is tolerated at higher DNA concentrations This trend in charge was even more obvious with a 9-mer peptide that did not phase separate at charge ratios < 0.33 (**Figure 3.14**). This observation makes sense in the context of the working hypothesis that excess positively-charged peptide species will remain in solution.



A.

Salmon Testes (Disperse)

[DNA] mg/mL	[RRAARRAARRAAY] /mM						[RRaaRRaaRRaaY] /mM					
	1.33	1.17	1.00	0.83	0.67	0.50	1.33	1.17	1.00	0.83	0.67	0.50
2.00	1.122	1.088	0.46	0.045	0.039	0.037	1.108	1.17	0.229	0.042	0.041	0.039
1.00	0.831	0.82	0.795	0.178	0.04	0.038	0.862	0.867	0.667	0.09	0.037	0.037
0.50	0.553	0.562	0.565	0.489	0.109	0.039	0.535	0.578	0.562	0.292	0.055	0.037
0.25	0.319	0.315	0.314	0.308	0.25	0.069	0.323	0.347	0.324	0.318	0.135	0.042
0.13	0.165	0.171	0.164	0.16	0.15	0.1	0.162	0.179	0.179	0.176	0.14	0.094
0.06	0.081	0.085	0.087	0.092	0.081	0.071	0.086	0.088	0.092	0.091	0.083	0.069
0.03	0.058	0.057	0.059	0.057	0.058	0.053	0.056	0.055	0.06	0.059	0.053	0.053
0.02	0.048	0.049	0.047	0.047	0.046	0.044	0.047	0.059	0.045	0.049	0.044	0.044

B.

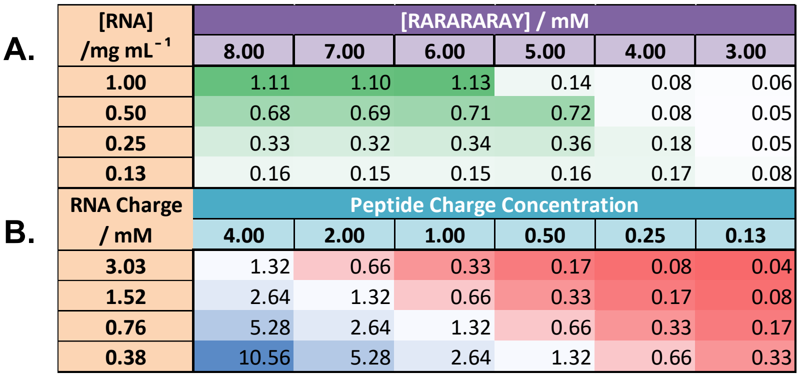
Poly-T (Monodisperse)

[DNA] mg/mL	[RRAARRAARRAAY] /mM						[RRaaRRaaRRaaY] /mM					
	1.33	1.17	1.00	0.83	0.67	0.50	1.33	1.17	1.00	0.83	0.67	0.50
2.00	1.368	1.29	0.188	0.038	0.039	0.037	1.397	1.024	0.039	0.04	0.037	0.039
1.00	1.073	1.095	0.709	0.062	0.036	0.039	1.091	1.071	0.31	0.037	0.037	0.036
0.50	0.759	0.759	0.75	0.25	0.039	0.038	0.768	0.784	0.758	0.132	0.038	0.037
0.25	0.437	0.414	0.435	0.423	0.136	0.084	0.468	0.458	0.432	0.414	0.069	0.038
0.13	0.259	0.28	0.25	0.239	0.229	0.052	0.245	0.235	0.245	0.238	0.188	0.049
0.06	0.135	0.139	0.138	0.131	0.115	0.116	0.132	0.134	0.132	0.132	0.131	0.059
0.03	0.086	0.078	0.081	0.08	0.087	0.078	0.079	0.079	0.082	0.076	0.078	0.072
0.02	0.059	0.061	0.06	0.056	0.061	0.059	0.058	0.058	0.058	0.056	0.056	0.055

C.

DNA Charge mM	Peptide Charge Concentration						Peptide Charge Concentration					
	8.00	7.00	6.00	5.00	4.00	3.00	8.00	7.00	6.00	5.00	4.00	3.00
6.06	1.32	1.16	0.99	0.83	0.66	0.50	1.32	1.16	0.99	0.83	0.66	0.50
3.03	2.64	2.31	1.98	1.65	1.32	0.99	2.64	2.31	1.98	1.65	1.32	0.99
1.52	5.28	4.62	3.96	3.30	2.64	1.98	5.28	4.62	3.96	3.30	2.64	1.98
0.76	10.56	9.24	7.92	6.60	5.28	3.96	10.56	9.24	7.92	6.60	5.28	3.96
0.38	21.12	18.48	15.84	13.20	10.56	7.92	21.12	18.48	15.84	13.20	10.56	7.92
0.19	42.24	36.96	31.68	26.40	21.12	15.84	42.24	36.96	31.68	26.40	21.12	15.84
0.09	84.49	73.93	63.37	52.81	42.24	31.68	84.49	73.93	63.37	52.81	42.24	31.68
0.05	168.98	147.85	126.73	105.61	84.49	63.37	168.98	147.85	126.73	105.61	84.49	63.37

Figure 3.15 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing A) ss-DNA from salmon testes or B) monodisperse poly-T with the peptides RRAA-13 and RRaa-13 het in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. C) Charge ratio (positive charge/negative charge) is shown for each condition. Red represents excess negative charge, while blue represents excess positive charge.



A.

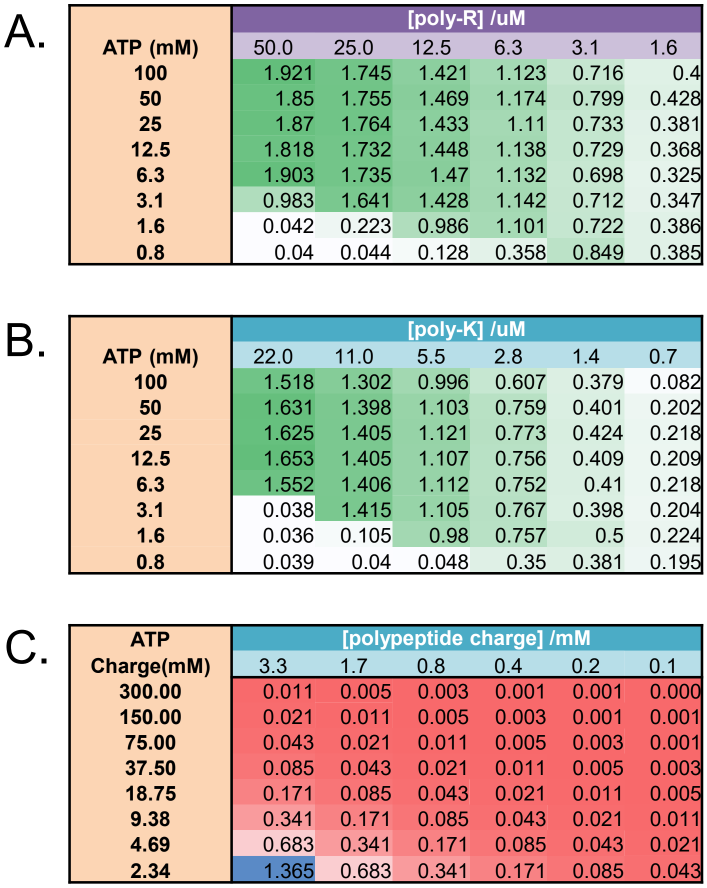
[RNA] /mg mL^{-1}	[RARARARAY] / mM					
	8.00	7.00	6.00	5.00	4.00	3.00
1.00	1.11	1.10	1.13	0.14	0.08	0.06
0.50	0.68	0.69	0.71	0.72	0.08	0.05
0.25	0.33	0.32	0.34	0.36	0.18	0.05
0.13	0.16	0.15	0.15	0.16	0.17	0.08

B.

RNA Charge / mM	Peptide Charge Concentration					
	4.00	2.00	1.00	0.50	0.25	0.13
3.03	1.32	0.66	0.33	0.17	0.08	0.04
1.52	2.64	1.32	0.66	0.33	0.17	0.08
0.76	5.28	2.64	1.32	0.66	0.33	0.17
0.38	10.56	5.28	2.64	1.32	0.66	0.33

Figure 3.16 A) Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing total yeast RNA with the peptide H_2N-RARARARAY-OH in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. B) Charge ratio (positive charge/negative charge) is shown corresponding to each condition. Red represents excess negative charge, while blue represents excess positive charge.

To further test the templating hypothesis, we decided to switch the nature of the positively- and negatively-charged species and pair a large polymeric polycation with a small polyanion (**Figure 3.17**). Poly-L-lysine and poly-L-arginine both form coacervates with ATP, and in this case we determined that the resulting OD_{600} is dependent on the polycation. It seems likely that in this 'flipped' scenario, the polycationic peptide is recruiting a certain number of ATP molecules while excess negative charge remains in solution, which supports our hypothesis that large molecules act as 'templates' that can recruit a specific number of smaller equivalents into the coacervates.



A.

ATP (mM)	[poly-R] /uM					
	50.0	25.0	12.5	6.3	3.1	1.6
100	1.921	1.745	1.421	1.123	0.716	0.4
50	1.85	1.755	1.469	1.174	0.799	0.428
25	1.87	1.764	1.433	1.11	0.733	0.381
12.5	1.818	1.732	1.448	1.138	0.729	0.368
6.3	1.903	1.735	1.47	1.132	0.698	0.325
3.1	0.983	1.641	1.428	1.142	0.712	0.347
1.6	0.042	0.223	0.986	1.101	0.722	0.386
0.8	0.04	0.044	0.128	0.358	0.849	0.385

B.

ATP (mM)	[poly-K] /uM					
	22.0	11.0	5.5	2.8	1.4	0.7
100	1.518	1.302	0.996	0.607	0.379	0.082
50	1.631	1.398	1.103	0.759	0.401	0.202
25	1.625	1.405	1.121	0.773	0.424	0.218
12.5	1.653	1.405	1.107	0.756	0.409	0.209
6.3	1.552	1.406	1.112	0.752	0.41	0.218
3.1	0.038	1.415	1.105	0.767	0.398	0.204
1.6	0.036	0.105	0.98	0.757	0.5	0.224
0.8	0.039	0.04	0.048	0.35	0.381	0.195

C.

ATP Charge(mM)	[polypeptide charge] /mM					
	3.3	1.7	0.8	0.4	0.2	0.1
300.00	0.011	0.005	0.003	0.001	0.001	0.000
150.00	0.021	0.011	0.005	0.003	0.001	0.001
75.00	0.043	0.021	0.011	0.005	0.003	0.001
37.50	0.085	0.043	0.021	0.011	0.005	0.003
18.75	0.171	0.085	0.043	0.021	0.011	0.005
9.38	0.341	0.171	0.085	0.043	0.021	0.011
4.69	0.683	0.341	0.171	0.085	0.043	0.021
2.34	1.365	0.683	0.341	0.171	0.085	0.043

Figure 3.17 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing ATP with the polypeptides A) poly-L-arginine or B) poly-L-lysine in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. C) Charge ratio (positive charge/negative charge) is shown corresponding to each condition.

3.6. Heterochiral Character Dependence on LLPS Propensity Differences

The following studies were conducted during a time at which in-house analytical abilities were limited due to the fouling of our HPLC columns. The results herein should be interpreted as provisional until replicated with peptides of established purity.

After determining that two RRAA 13-mer diastereomers exhibit different LLPS propensities while two diastereomers of 9-mers did not, we wanted to investigate the amount of “heterochiral character” needed to cause a shift in propensity. Here we define heterochiral character as the fraction of amino acid residues that differ in chirality from the dominant chirality in the backbone. For example, a 10-mer peptide with 8 L-amino acids and 2 D-amino acids will be described to have a heterochiral character of 20%; the enantiomer of that same peptide will have the same heterochiral character.

We therefore synthesized the following 13-mer peptides, with varying heterochiral character (note that L-tyrosine was not included in the heterochiral character calculations).

`0%  H2N- R R A A R R A A R R A A Y -OH;` “RRAA-13”

`16% H2N- R R A a R R A A r R A A Y -OH;` “RRAA-13 het-16”

`25% H2N- R R a A R r A A r R A A Y -OH;` “RRAA-13 het-25”

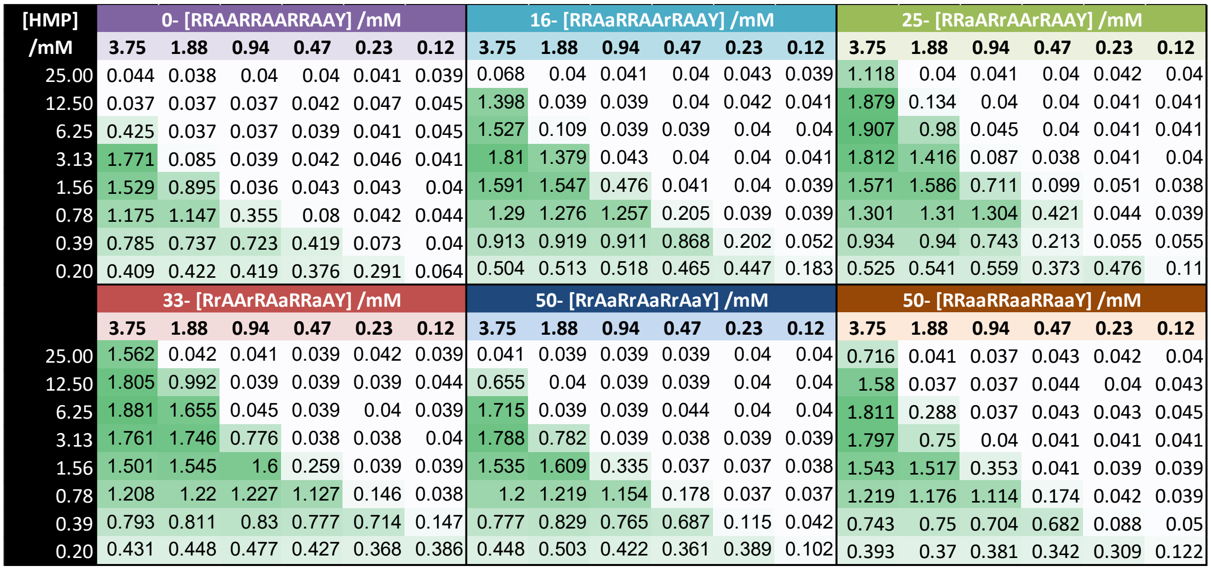
`33% H2N- R r A A r R A a R R a A Y -OH;` “RRAA-13 het-33”

`50% H2N- R r A a R r A a R r A a Y -OH;` “RRAA-13 het-50”

Note that two sequences have so far been described with 50% heterochiral character: RRaa-13-het, which was used in the length-dependence experiments described in section 3.5., and RRAA-13 het-50, which is new in this section. The RRaa-13-het has all D-alanine residues in short clusters, while D- residues are evenly spaced in the RRAA-13 het-n% sequences. We did not synthesis peptides with >50% heterochiral character, since such sequences would be very similar from a stereochemical perspective (except for tyrosine) to those listed above.

The phase boundary was determined for each new peptide (**Figure 3.16**). We tested 3.75 mM – 0.12 mM peptide with 25 mM – 0.19 mM HMP. Turbidity differences between the same concentrations for different peptides were immediately observed, and OD_{600} measurements were made to gain further insight. In this dataset, it is best to compare one condition or set of conditions at a time across all diastereomers. For example, at a peptide concentration of 3.75 mM, the peptides with 25% and 33% heterochiral character underwent coacervation at all tested HMP concentrations; those with 16% and 50% (RRAA-13 het-50) underwent coacervation at [HMP] ≤ 12.5 mM. The homochiral diastereomer underwent coacervation at [HMP] ≤ 6.25 mM. Looking at the concentrations of both HMP and peptide at the phase boundary indicate that all heterochiral RRAA-13 diastereomers show a greater LLPS propensity than the homochiral RRAA-13. The best-performing peptide is RRAA-13 het-33, which seems to coacervate at ~4x lower peptide concentrations compared to homochiral RRAA-13 at each HMP concentration. For each RRAA-13 het-% diastereomer, the summation of OD600 values across all peptide and HMP concentrations was determined. This sum can be thought of as a comparative readout for the relative LLPS propensities (**Figure 3.17**).

One way to interpret these data is to suggest that introducing a D residue every third residue of an otherwise L peptide backbone increases the ability for that peptide to undergo LLPS. Each residue with inverted chirality introduces an unconformity (a 'kink') to the backbone,[77] and collectively these kinks might be expected to increase disorder in the peptide relative to an all-L peptide. This interpretation is speculative. The RRAA 13-mers are likely quite disordered regardless of heterochiral character.



[HMP]	0- [RRAARRAARRAAY] /mM						16- [RRAaRRAArRAAY] /mM						25- [RRaARrAArRAAY] /mM					
/mM	**3.75**	**1.88**	**0.94**	**0.47**	**0.23**	**0.12**	**3.75**	**1.88**	**0.94**	**0.47**	**0.23**	**0.12**	**3.75**	**1.88**	**0.94**	**0.47**	**0.23**	**0.12**
25.00	0.044	0.038	0.04	0.04	0.041	0.039	0.068	0.04	0.041	0.04	0.043	0.039	1.118	0.04	0.041	0.04	0.042	0.04
12.50	0.037	0.037	0.037	0.042	0.047	0.045	1.398	0.039	0.039	0.04	0.042	0.041	1.879	0.134	0.04	0.04	0.041	0.041
6.25	0.425	0.037	0.037	0.039	0.041	0.045	1.527	0.109	0.039	0.039	0.04	0.04	1.907	0.98	0.045	0.04	0.041	0.041
3.13	1.771	0.085	0.039	0.042	0.046	0.041	1.81	1.379	0.043	0.04	0.04	0.041	1.812	1.416	0.087	0.038	0.041	0.04
1.56	1.529	0.895	0.036	0.043	0.043	0.04	1.591	1.547	0.476	0.041	0.04	0.039	1.571	1.586	0.711	0.099	0.051	0.038
0.78	1.175	1.147	0.355	0.08	0.042	0.044	1.29	1.276	1.257	0.205	0.039	0.039	1.301	1.31	1.304	0.421	0.044	0.039
0.39	0.785	0.737	0.723	0.419	0.073	0.04	0.913	0.919	0.911	0.868	0.202	0.052	0.934	0.94	0.743	0.213	0.055	0.055
0.20	0.409	0.422	0.419	0.376	0.291	0.064	0.504	0.513	0.518	0.465	0.447	0.183	0.525	0.541	0.559	0.373	0.476	0.11

	33- [RrAArRAaRRaAY] /mM						50- [RrAaRrAaRrAaY] /mM						50- [RRaaRRaaRRaaY] /mM					
	3.75	**1.88**	**0.94**	**0.47**	**0.23**	**0.12**	**3.75**	**1.88**	**0.94**	**0.47**	**0.23**	**0.12**	**3.75**	**1.88**	**0.94**	**0.47**	**0.23**	**0.12**
25.00	1.562	0.042	0.041	0.039	0.042	0.039	0.041	0.039	0.039	0.039	0.04	0.04	0.716	0.041	0.037	0.043	0.042	0.04
12.50	1.805	0.992	0.039	0.039	0.039	0.044	0.655	0.04	0.039	0.039	0.04	0.04	1.58	0.037	0.037	0.044	0.04	0.043
6.25	1.881	1.655	0.045	0.039	0.04	0.039	1.715	0.039	0.039	0.044	0.04	0.04	1.811	0.288	0.037	0.043	0.043	0.045
3.13	1.761	1.746	0.776	0.038	0.038	0.04	1.788	0.782	0.039	0.038	0.039	0.039	1.797	0.75	0.04	0.041	0.041	0.041
1.56	1.501	1.545	1.6	0.259	0.039	0.039	1.535	1.609	0.335	0.037	0.037	0.038	1.543	1.517	0.353	0.041	0.039	0.039
0.78	1.208	1.22	1.227	1.127	0.146	0.038	1.2	1.219	1.154	0.178	0.037	0.037	1.219	1.176	1.114	0.174	0.042	0.039
0.39	0.793	0.811	0.83	0.777	0.714	0.147	0.777	0.829	0.765	0.687	0.115	0.042	0.743	0.75	0.704	0.682	0.088	0.05
0.20	0.431	0.448	0.477	0.427	0.368	0.386	0.448	0.503	0.422	0.361	0.389	0.102	0.393	0.37	0.381	0.342	0.309	0.122

Figure 3.16 Optical densities as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing HMP with the RRAA-13 series peptides in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above.

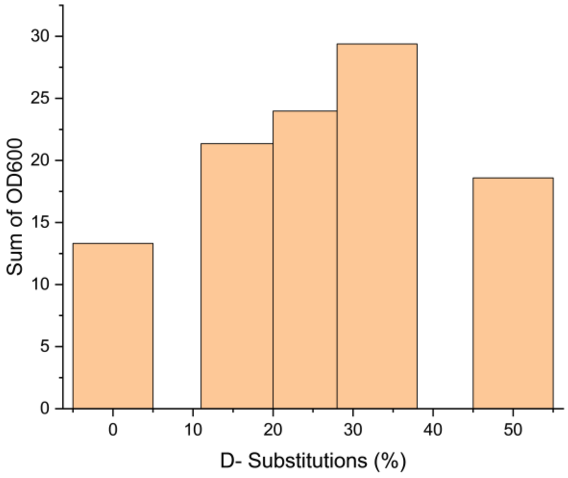


Figure 3.17 Summation of all OD_{600} values acquired on a Biotek Synergy 2 plate reader for each RRAA-13 het-% series peptide from the phase diagrams in **Figure 3.16**.

As seen in the two lower-right panels in **Figure 3.16,** the two RRAA 13-mers with 50% heterochiral characters—RRAA-13 het-50 and RRaa-13 het—behave differently. RRaa-13 het (bottom right), with clustered D-alanine residues forms coacervates at slightly lower peptide concentrations than RRAA-13 het-50% (bottom middle) with alternating L- and D- amino acids. RRaa-13 het requires half the peptide concentration to undergo LLPS at HMP concentrations of 25 and 6.25 mM relative to RRAA-13 het-50%. This observation very clearly demonstrates that net heterochiral character is not the only factor to consider, but also perhaps the identity or spacing of the opposite-chirality residues in the sequence.

We considered why the RRAA-13 het-50 with alternating L- and D- residues has a lower LLPS propensity than RRaa-13 het with clustered D-alanine residues. One possibility is the number or frequency of kinks in the backbone. The RRaa-13 het has 6 kinks, i.e., sites at which L and D residues are adjacent. The RRAA-13 het-50 has twice as many such sites. In the sequences below, each of these sites is represented with an asterisk:

`H2N- R R*a a*R R*a a*R R*a a*Y -OH;` “RRaa-13-het”

`H2N- R*r*A*a*R*r*A*a*R*r*A*a*Y -OH;` “RRAA-13 het-50”

Interestingly, the RRAA-13 het-25 and RRAA-13 het-33, which exhibited the highest LLPS propensity, have 6 and 8 kinks in their sequences, respectively:

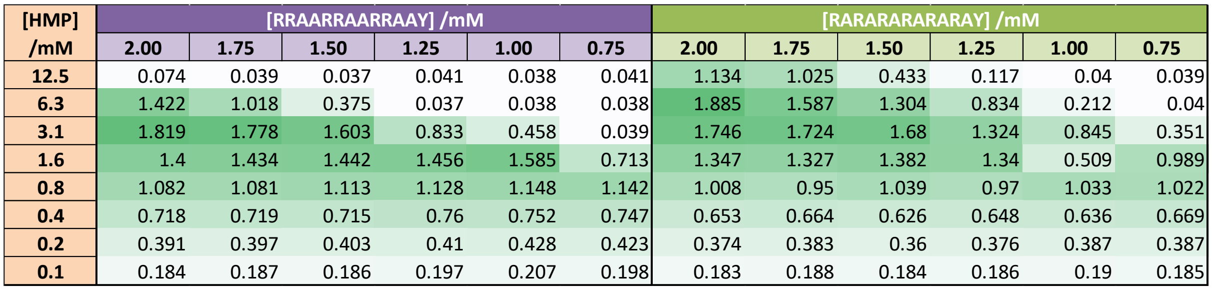
`25% H2N- R R*a*A R*r*A A*r*R A A Y -OH;` “RRAA-13 het-25”

`33% H2N- R*r*A A*r*R A*a*R R*a*A Y -OH;` “RRAA-13 het-33”

This led us to consider another peptide sequence:

H2N- R A R A R A R A R A R A Y -OH; “RARA-13”

H2N- R a R a R a R a R a R a Y -OH; “RaRa-13 het-50”



[HMP] /mM	[RRAARRAARRAAY] /mM						[RARARARARARAY] /mM					
	2.00	1.75	1.50	1.25	1.00	0.75	2.00	1.75	1.50	1.25	1.00	0.75
12.5	0.074	0.039	0.037	0.041	0.038	0.041	1.134	1.025	0.433	0.117	0.04	0.039
6.3	1.422	1.018	0.375	0.037	0.038	0.038	1.885	1.587	1.304	0.834	0.212	0.04
3.1	1.819	1.778	1.603	0.833	0.458	0.039	1.746	1.724	1.68	1.324	0.845	0.351
1.6	1.4	1.434	1.442	1.456	1.585	0.713	1.347	1.327	1.382	1.34	0.509	0.989
0.8	1.082	1.081	1.113	1.128	1.148	1.142	1.008	0.95	1.039	0.97	1.033	1.022
0.4	0.718	0.719	0.715	0.76	0.752	0.747	0.653	0.664	0.626	0.648	0.636	0.669
0.2	0.391	0.397	0.403	0.41	0.428	0.423	0.374	0.383	0.36	0.376	0.387	0.387
0.1	0.184	0.187	0.186	0.197	0.207	0.198	0.183	0.188	0.184	0.186	0.19	0.185

Figure 3.18 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing HMP with the peptides RRAA-13 and RARA-13 in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. Presence of droplets was established with optical microscopy (not shown) prior to measurement.

The sequence of RARA-13 is different from that of the RRAA-13 sequence, but RaRa-13 het-50 has the same number of kinks as RRAA-13 het-50, so we hoped to gain insight from the phase separation behavior of these peptides. Although we estimate by comparing turbidity measurements that RARA-13 has a slightly better phase separation propensity than RRAA-13 (**Figure 3.18**), it is so far not clear whether RARA-13 and RaRa-13 het-50 exhibit much of a difference (**Figure 3.19**). More research must be conducted on similar peptides to better understand how peptide sequence and stereochemistry are related to the ability to form complex coacervates.

[HMP] /mM	[RARARARARARAY] /mM						[RaRaRaRaRaRaY] /mM					
	2.00	**1.75**	**1.50**	**1.25**	**1.00**	**0.75**	**2.00**	**1.75**	**1.50**	**1.25**	**1.00**	**0.75**
12.5	1.134	1.025	0.433	0.117	0.04	0.039	1.299	1.176	0.755	0.148	0.061	0.037
6.3	1.885	1.587	1.304	0.834	0.212	0.04	1.721	1.681	1.397	0.77	0.325	0.037
3.1	1.746	1.724	1.68	1.324	0.845	0.351	1.691	1.69	1.509	1.549	0.988	0.247
1.6	1.347	1.327	1.382	1.34	0.509	0.989	1.391	1.388	1.447	1.418	1.427	1.056
0.8	1.008	0.95	1.039	0.97	1.033	1.022	1.097	1.085	1.106	1.068	1.109	1.056
0.4	0.653	0.664	0.626	0.648	0.636	0.669	0.715	0.74	0.733	0.719	0.711	0.641
0.2	0.374	0.383	0.36	0.376	0.387	0.387	0.403	0.411	0.416	0.41	0.417	0.37
0.1	0.183	0.188	0.184	0.186	0.19	0.185	0.189	0.192	0.192	0.191	0.176	0.176

Figure 3.19 Optical density as OD_{600} values acquired on a Biotek Synergy 2 plate reader after mixing HMP with the peptides RARA-13 and RaRa-13 het-50 in low-ion TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$ at pH 8) in the proportions shown above. Presence of droplets was established with optical microscopy (not shown) prior to measurement.

3.7. Future Work

Several key differences have been established by the experiments in this chapter between both homochiral and heterochiral 13-mers and between heterochiral 13-mers with different sequences or heterochiral character. In all cases, heterochiral RRAA-13 diastereomers seem to undergo LLPS at lower concentrations than homochiral RRAA-13. The RRAA-13 het-% sequence with four evenly-distributed D-residues exhibits superior LLPS propensity to the diastereomers with two, three, or six distributed D-residues. The RRaa-13 het sequence with clustered D-alanine substitutions forms coacervates at lower concentrations than the RRAA-13 het-50% variant with distributed D-residues, even though they have the same amount of heterochiral character. In order to better understand the physical origin of these differences, several other peptides should be synthesized and tested:

`H2N- R A r a R A r a R A r a Y -OH;` RAra-13 het

Results from RAra-13 het would be insightful. This peptide has the same overall sequence as the RARA-13 class described above, with two key differences. First, this sequence has 6 kinks in the backbone, similar to the RRaa-13 het from section 3.5. Second, and unlike RRaa-13 het, this sequence contains both L- and D-alanine and L- and D-arginine. If RAra-13-het exhibits a better LLPS propensity than the RAra-13 or RAra-13 het-50 diastereomers, it is likely that the number of kinks in the backbone play a key role in LLPS.

Other RARA-13 diastereomers can be synthesized with 25% and 33% heterochiral character. Like the RRAA-13 het-25 and RRAA-13 het-33, these isomers will have 6 or 8 kinks in the backbone while only containing 2 or 4 D- residues:

25% H2N- R A r A R a R A r A R A Y -OH; RARA-13 het-25

33% H2N- R a R A r A R a R A r A Y -OH; RARA-13 het-33

If these two peptides, like the analogous peptides in the RRAA-13 series, show better LLPS propensity than the homochiral RARA-13, the hypothesis that exchanging every third residue for one of the opposite chirality has a positive effect on LLPS is further supported.

One must also wonder if the identity of amino acids were exchanged for the opposite chirality might play any role in LLPS. This hypothesis is more subtle and difficult to test. RRaa-13 het contains all L-arginine and D-alanine. Although the overall peptide is heterochiral, there is homochirality among the individual residues. RRAA-13 het-50, on the other hand, contains both stereoisomers of alanine and arginine. It does not seem likely that homo- and heterochirality among the individual residues would play a significant role. After all, RRAA-13 het-33 contains both L- and D- forms of both alanine and arginine and has a better phase separation propensity than the homochiral RRAA-13. Regardless, one could still test this hypothesis with alternative versions of the equally-spaced RRAA-

13 het series peptides and determining if their phase separation propensities are similar to the previous versions:

0% H2N- R R A A R R A A R R A A Y -OH; “RRAA-13”

16% H2N- R R A a R R A A r R A A Y -OH; “RRAA-13 het-16”

16% H2N- R R A A r R A A r R A A Y -OH; “RRAA-13 het-16_alt”

25% H2N- R R a A R r A A r R A A Y -OH; “RRAA-13 het-25”

25% H2N- R R a A R R a A R R a A Y -OH; “RRAA-13 het-25_alt”

33% H2N- R r A A r R A a R R a A Y -OH; “RRAA-13 het-33”

33% H2N- R r A A r R A A r R A a Y -OH; “RRAA-13 het-33_alt”

If these ‘alternate’ peptides, which have the same heterochiral character but contain alternate chiral substitutions that are primarily one amino acid, have similar phase separation propensities as the original variants, then we can rule out the possibility that homo- or heterochirality among residues plays a significant role in LLPS propensity.

To conclude this investigation, the two-dimensional phase diagrams shared above must be replicated using fresh peptide. Peptide purity should be assessed on a clean ultra-high pressure liquid chromatography (UHPLC) column. Wells should be screened for droplets with both OD_{600} and optical microscopy to ensure droplet formation and continue to rule out formation of solid aggregates (so far, aggregates of these RRAA-13 peptides have *not* been observed). Additional peptides from this section should be synthesized and tested to gain insight into how heterochirality might affect LLPS.

3.8. Conclusion

Systematic modifications of short, cationic peptide sequences were used in this chapter in attempt to isolate the effects of stereochemistry on phase separating systems with hexametaphosphate. This approach was aimed to help us understand how peptide length, chiral composition, and spacing of chiral residues affect phase separation.

The results so far indicate that heterochirality in simple 13-mer RRAA-repeat peptides promotes phase separation. Specifically, it seems that the influence of heterochirality on LLPS is governed not only by peptide length but also by the composition and spacing of chiral residues. Differences in phase separation propensity were observed in neither Arg-Arg diastereomers nor RRAA-repeating 9-mer peptides. However, 13-mer peptides demonstrated significant stereochemistry-dependent behaviors with respect to LLPS. This observation establishes a length threshold with the RRAA peptide construct. Differences in LLPS propensity were also observed in RRAA peptides with varying numbers of heterochiral residues, with all heterochiral peptides undergoing LLPS at lower concentrations than the homochiral diastereomer.. Importantly, peptides with the identical number of heterochiral residues but differing in distribution can exhibit different LLPS propensities. These findings suggest that subtle features can modulate intermolecular interactions that contribute to LLPS. Further study is necessary to probe the precise mechanism.

In a prebiotic context, this work supports the hypothesis that LLPS could have contributed to the selection of homo- or heterochiral peptides. Differences in phase behavior could have contributed to the enrichment of homochiral compounds.

These results are provisional until replicated with peptides of established purity.

3.9. Experimental Methods and Spectra

Instrumentation: Peptides were synthesized on either a CEM Liberty Blue microwave-assisted peptide synthesizer with a high-throughput resin transfer module or a CEM MultiPep parallel synthesizer. Peptides were purified on an Agilent 1260/1290 Infinity II instrument with an MSD detector using a Waters XSelect C18 CSH 19x250 mm column. Peptide Concentration was measured on a Thermo NanoDrop One. Nanopure water for preparative HPLC was from a Barnsted instrument. Microscopy was conducted wither on a Nikon TS2 inverted microscope with phase contrast objectives or a Leica M205 stereo-zoom microscope. Turbidity measurements were conducted on a BioTek Synergy 2 plate reader. High resolution mass spectra with optional sequencing were acquired with a Bruker Impact II Q-TOF that was a gift from the Bender family to the Department of Chemistry. ^{31}P NMR was acquired on a Bruker Avance Neo 500 funded by NSF award CHE-2017891.

Consumables: Manual peptide synthesis vessels were purchased from Torviq. 96-well plates and pipette tips used to handle peptides and conduct assays were manufactured by Corning. Other plasticware where retention is not an issue was purchased from DOT scientific.

Chemicals: Peptide synthesis resins, including 2-CTC resin and preloaded Fmoc-Tyr(tBu)-Wang resin, were purchased from Aapptec. Unless otherwise noted, peptide synthesis reagents including Fmoc-protected amino acids and coupling reagents were purchased from Chem-Impex International. Reagents from other manufacturers include: Fmoc-Arg(Pbf)-OH (Ambeed), trifluoroacetic acid (Honeywell), HPLC grade acetonitrile (Sigma Aldrich), *N,N*-dimethylformamide (Sigma Aldrich), HMP (Sigma Aldrich), TRIS

(Sigma Aldrich), potassium chloride (Sigma Aldrich), magnesium chloride (Sigma Aldrich), dichloromethane (Honeywell), diethyl ether (Sigma Aldrich), methanol (Sigma Aldrich).

Synthesis of Peptide Dimers

Fmoc-Arg(Pbf)-OH was loaded onto 2-chlorotritylchloride resin as follows. 2-CTC substitution was stated on the bottle. 1.0 eq of Fmoc-Arg(Pfb)-OH was added to dry DCM (10 mL per gram resin) followed by 2.5 eq Hünig's base. The protected arginine only fully dissolved after the base was added. The mixture was stirred overnight in a closed bottle. The next morning, remaining 2-CTC groups were capped by adding methanol (1 mL per gram resin) and filtered. The resin was dried under nitrogen, and arginine substitution was determined gravimetrically by comparing the mass of the dried resin to the amount of resin originally used.

The resin was deprotected by adding 20% piperidine or 4-methylpiperidine in DMF and stirring at room temperature for 40 minutes. The solution was drained out of the reaction vessel and the resin was rinsed.

The next amino acid residue was coupled by preactivating the protected amino acid (2 eq) with HATU (2 eq) and Hünig's base (4 eq) in DMF and adding it to the resin. This was stirred for 2 hours before being drained and washed. A double coupling may be performed for better yield.

Synthesis of Peptide 9- and 13-mers

SPPS of 9- and 13-mer peptides was carried out in a Liberty Blue automated peptide synthesizer at 0.1 mmol scale. 1 M DIC and Oxyma were used for coupling and 0.2 M amino acid solutions were prepared. Preloaded tyrosine Wang resin was pre-swelled in DCM/DMF. 20% 4-methyl piperidine was used for deprotection. Double couplings were carried out on all amino acids. Regular coupling cycles were conducted at 90° for two minutes, and arginine residues were installed using a 10-minute extended coupling.

Alternatively, a non-microwave assisted parallel synthesizer (MultiPep, CEM) may be used with HBTU and NMM.

Arginine-Rich Peptide Cleavage

After N-terminal deprotection with 20% piperidine or 4-methylpiperidine in DMF, the resin was rinsed with methanol and dried. Peptide was cleaved off the resin using 95% TFA in water. After 4 hours of cleavage (at which point the solution should be clear, not cloudy), the TFA was drained from the reaction vessel and reduced in volume under a stream of nitrogen. Cold diethyl ether was used to precipitate the crude peptide, which was pelletized under centrifugation and washed twice with more ether.

Purification of Arginine-Rich Peptides

A general HPLC method has been developed that results in good separation of all 9-mer and 13-mer arginine-rich peptides described thus far. Water and acetonitrile were run at 20 mL/min starting with a 3 minute isocratic hold at 5% organic. Acetonitrile was then increased to 15% over the next 12 minutes. This is followed by a wash cycle to remove other impurities from the column and reequilibration. Using this method, 9- and 13-mer peptides typically elute between 8 and 14 minutes and show an intense signal at 205 and 280 nm.

Concentration Determination

Peptides from HPLC fractions are isolated and lyophilized. After redissolving in minimum buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$), aliquots were extracted and diluted 10 or 100 times in buffer to fall within the dynamic range of the NanoDrop One. Extinction coefficients were predicted at 205 and 280.[78] Concentration was determined via Beer's Law at both wavelengths and averaged.

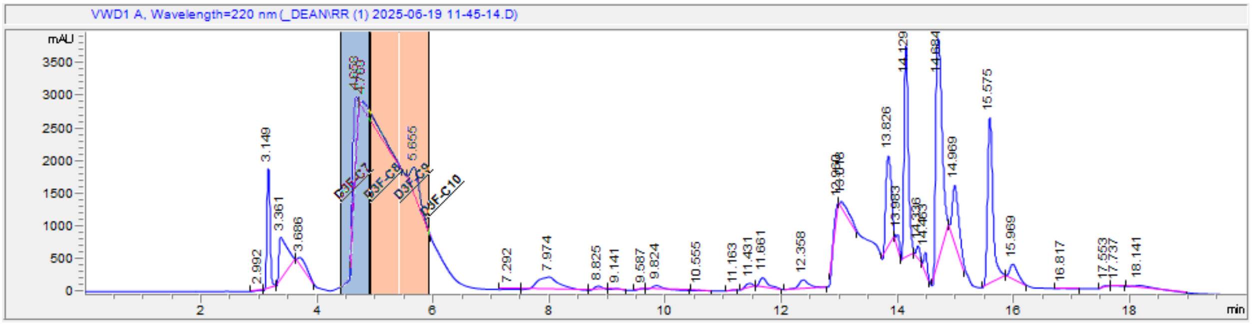
LLPS Assay Setup

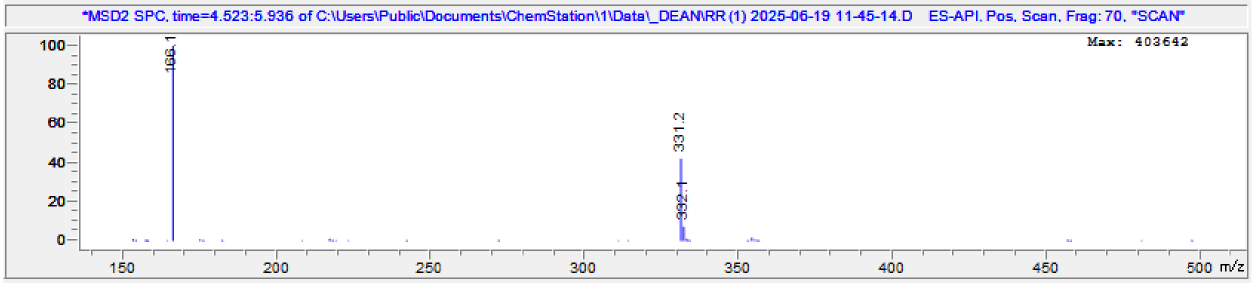
To determine the phase boundary of a given peptide, an assay was conducted as follows: With a multichannel pipette, peptide concentration was varied by factors of two across the columns of a 96-well plate using serial dilution. The final volume of the peptide solution in each well should not exceed 50% of the working volume of the plate. For each dilution, the plunger on the pipette was pushed three times for thorough mixing but limited retention. In a separate plate, polyphosphate (HMP) solution was prepared via an orthogonal serial dilution, where each row is made with half the HMP concentration as the one prior. The rows of HMP solution are then pipetted into the rows of peptide solution, mixing three times with the pipette for each row. The plate was measured at once via OD_{600} or optical microscopy.

Chromatograms and Mass Spectra from Prep LC-MS:

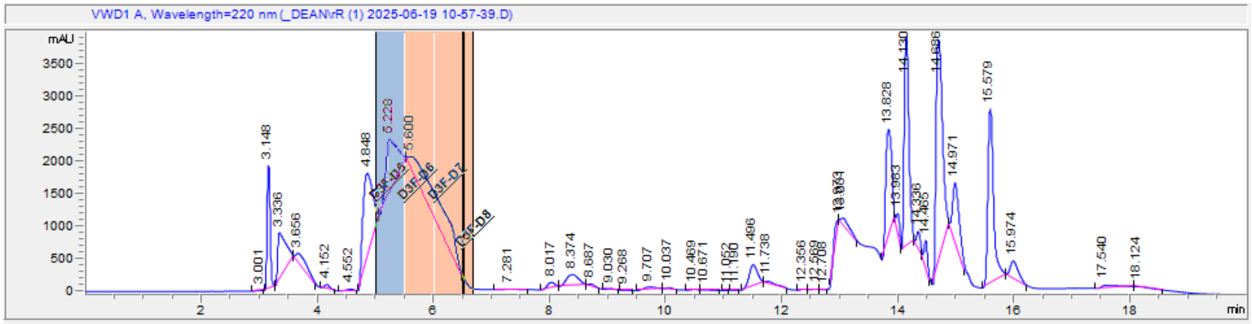
*Note: some spectra exhibit odd peak shape. This is likely due to detector saturation.

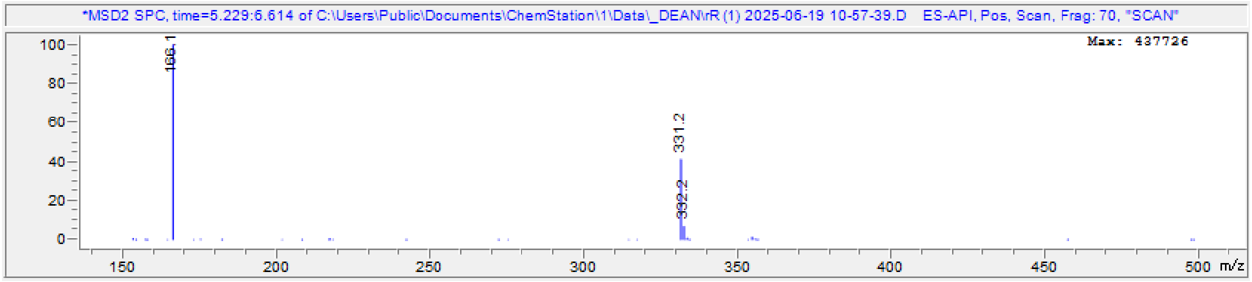
`H2N- R R -OH`



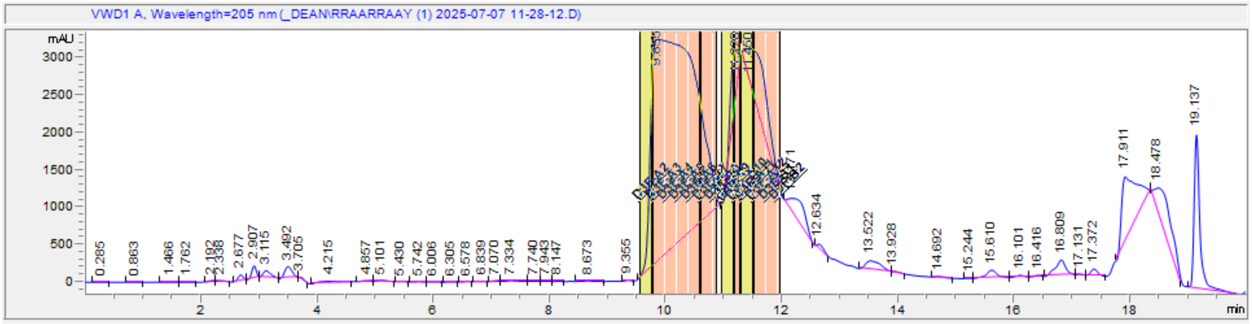


`H2N- r R -OH`

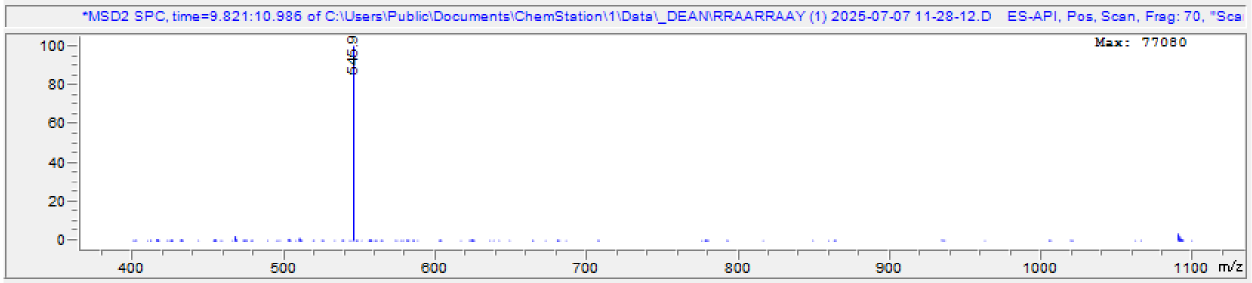




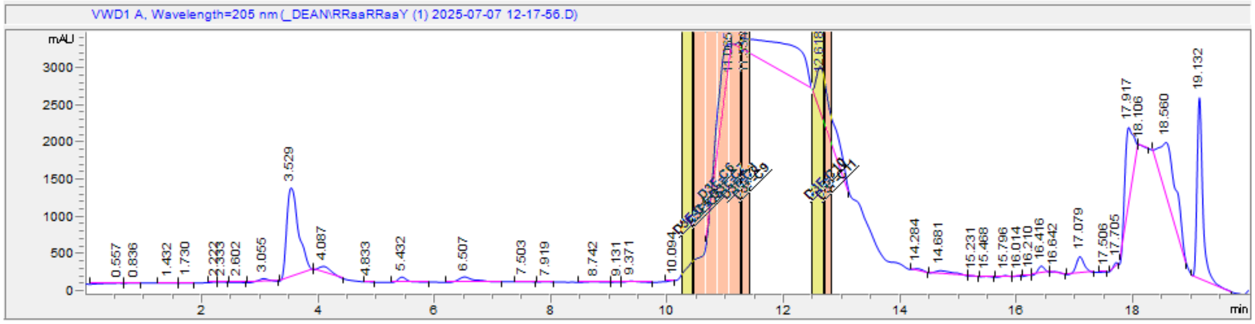
H2N- R R A A R R A A Y -OH; “RRAA-9”

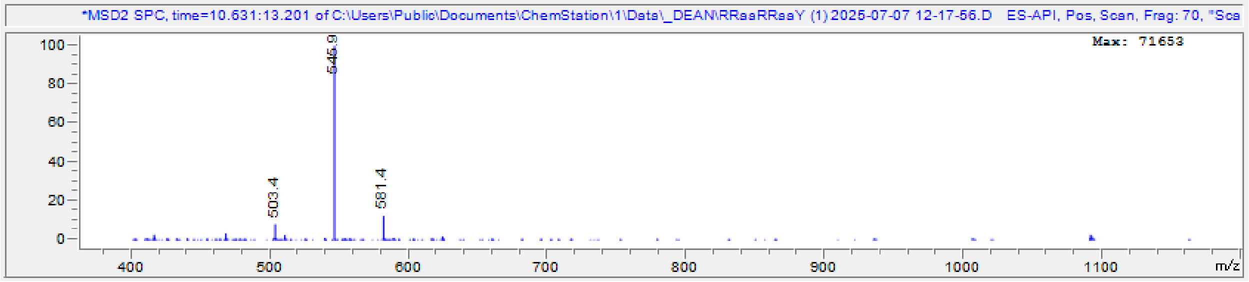


(Leftmost peak contains peptide of interest)

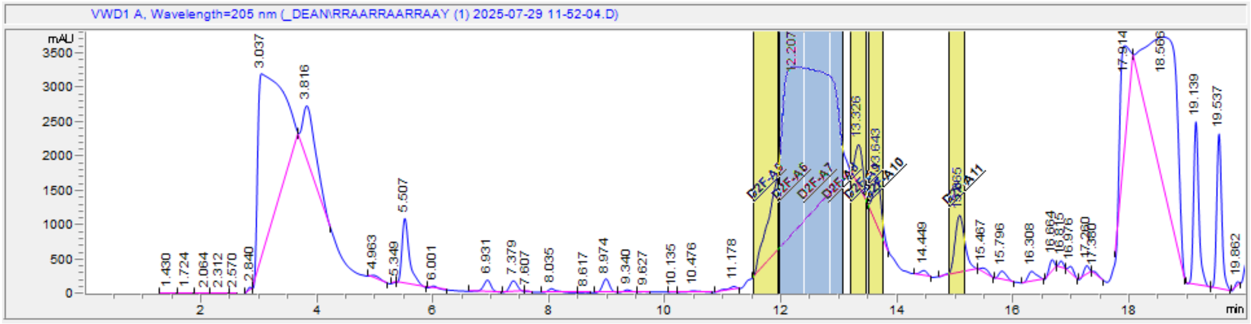


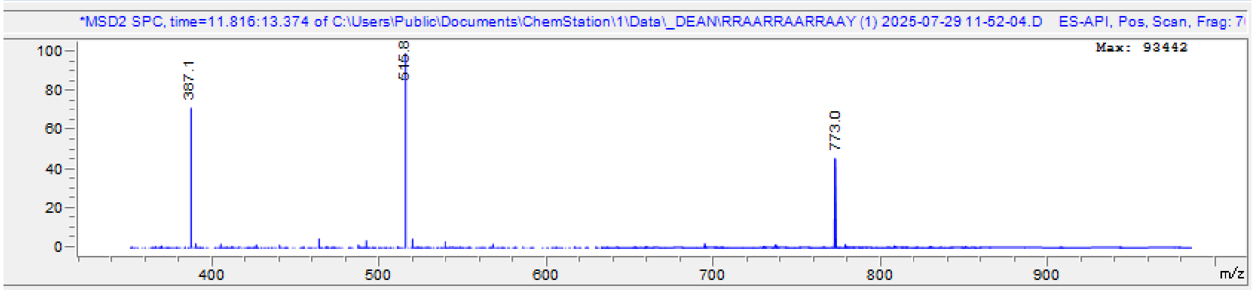
H2N- R R a a R R a a Y -OH; “RRaa-9”



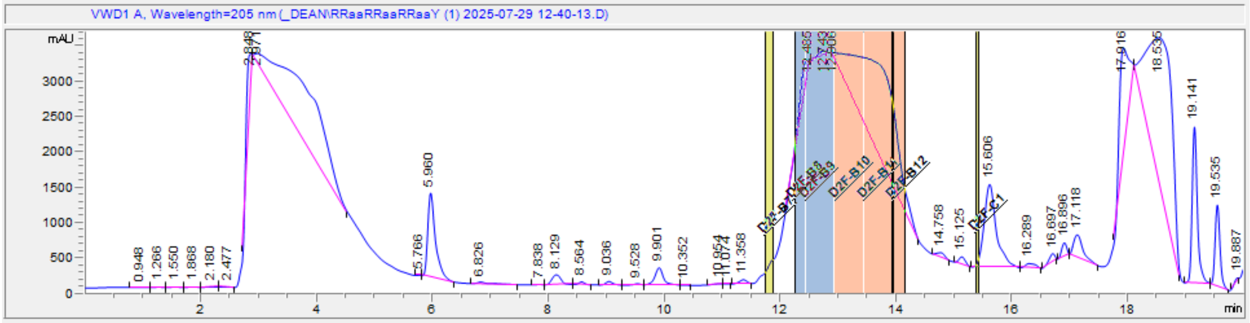


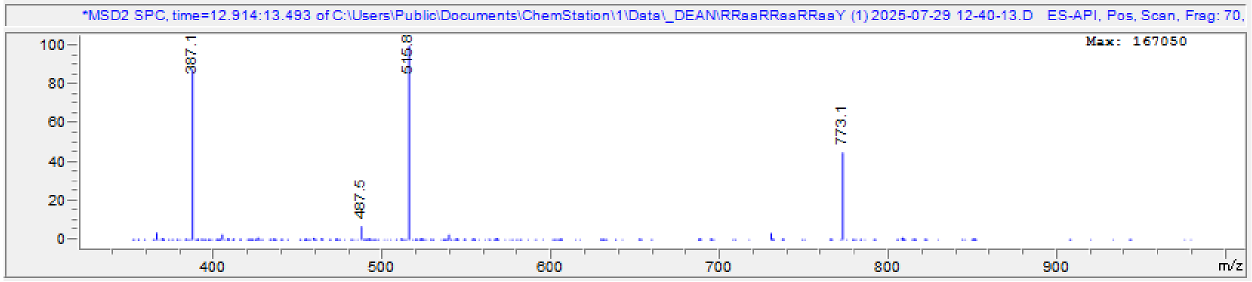
H2N- R R A A R R A A R R A A Y -OH; “RRAA-13”





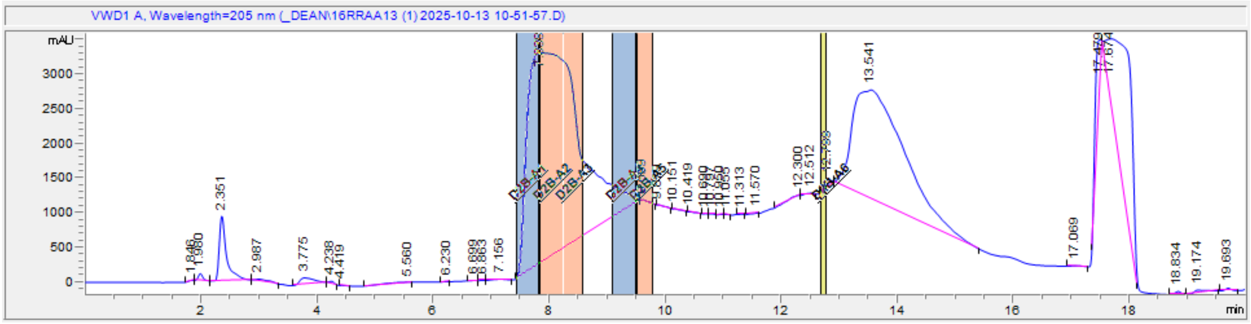
H2N- R R a a R R a a R R a a Y -OH; “RRaa-13-het”

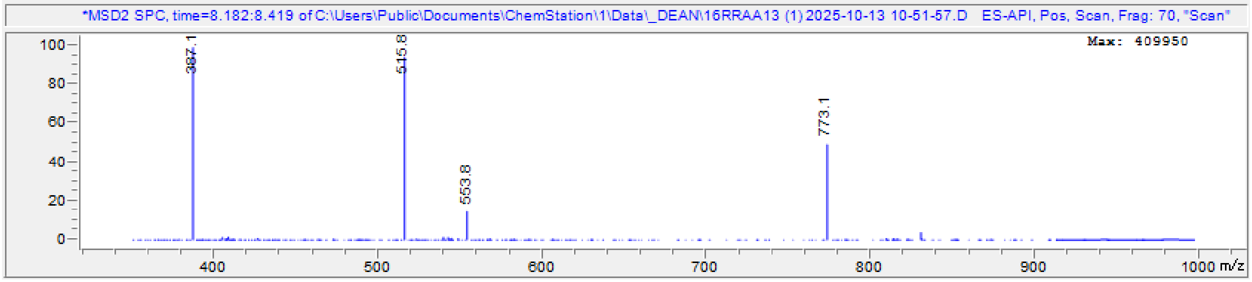




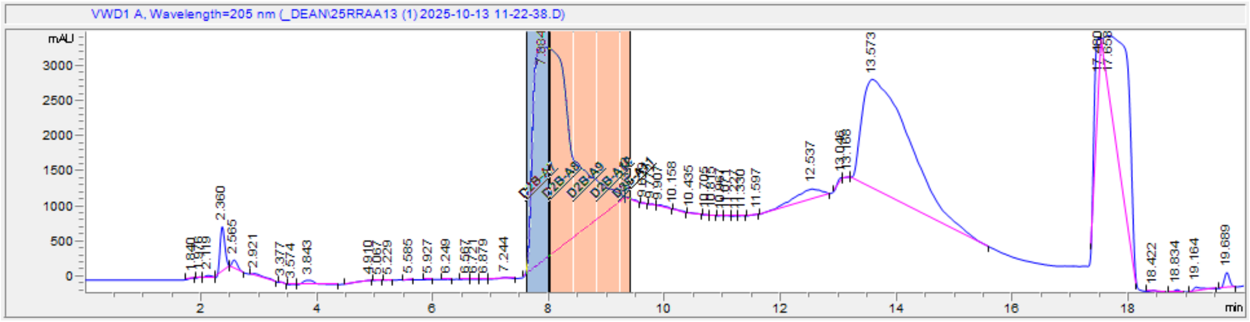
*Peak shape in the following peptides result of column fouling.

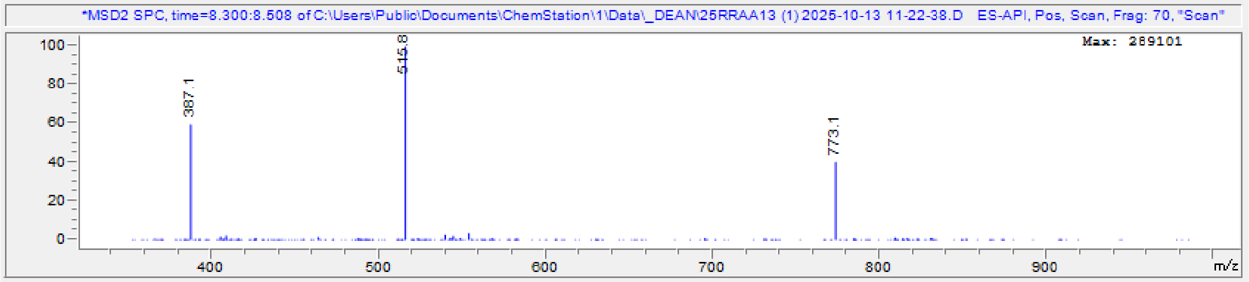
16% H2N- R R A a R R A A r R A A Y -OH; “RRAA-13 het-16”



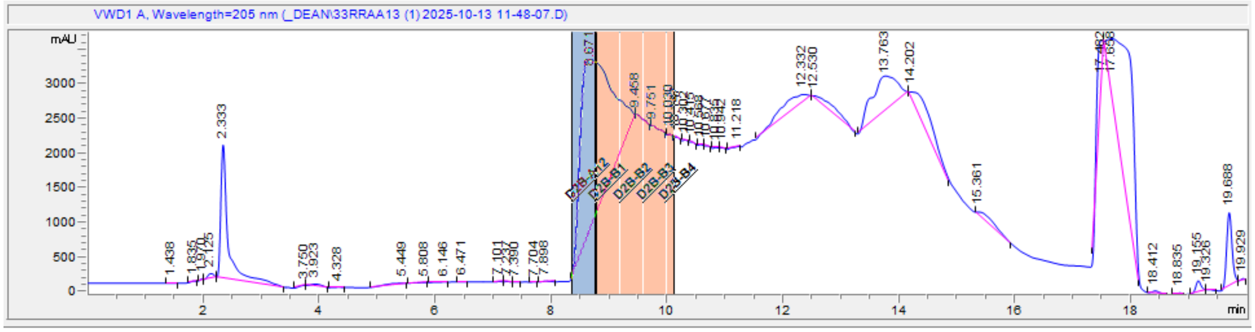


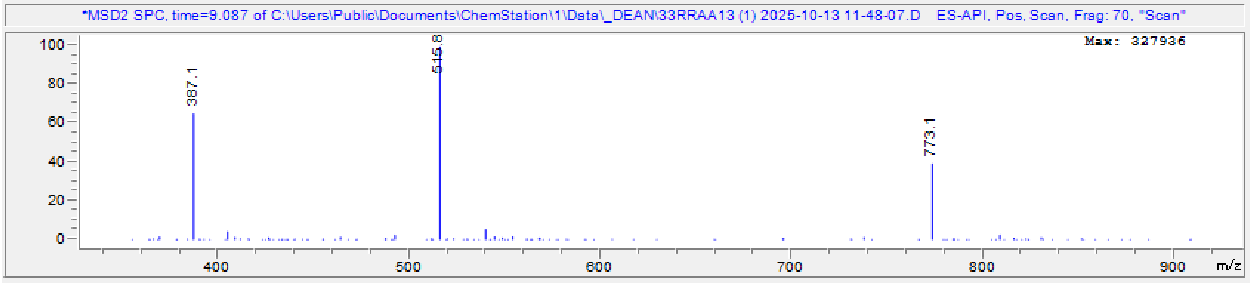
25% H2N- R R a A R r A A r R A A Y -OH; “RRAA-13 het-25”



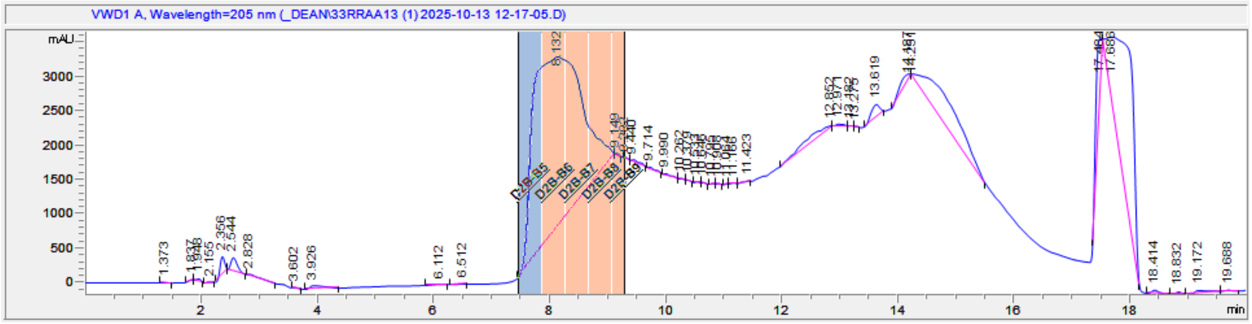


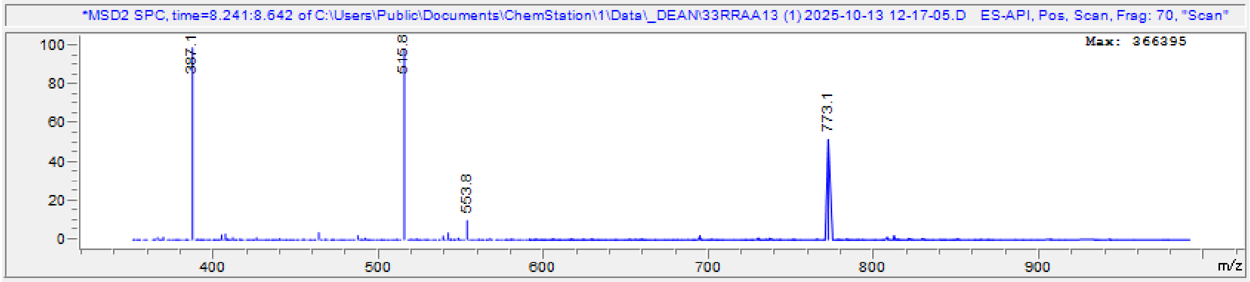
33% H2N- R r A A r R A a R R a A Y -OH; “RRAA-13 het-33”



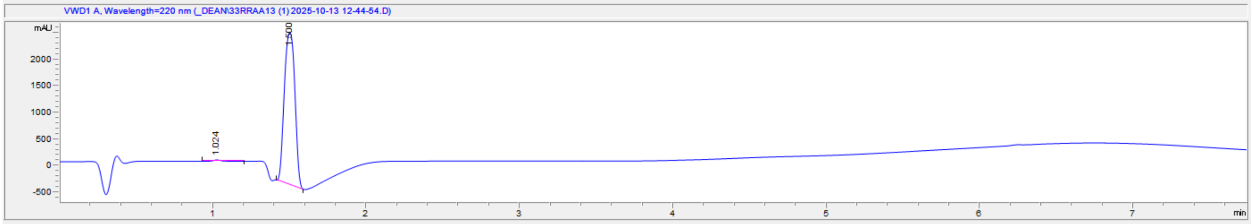


50% H2N- R r A a R r A a R r A a Y -OH; “RRAA-13 het-50”

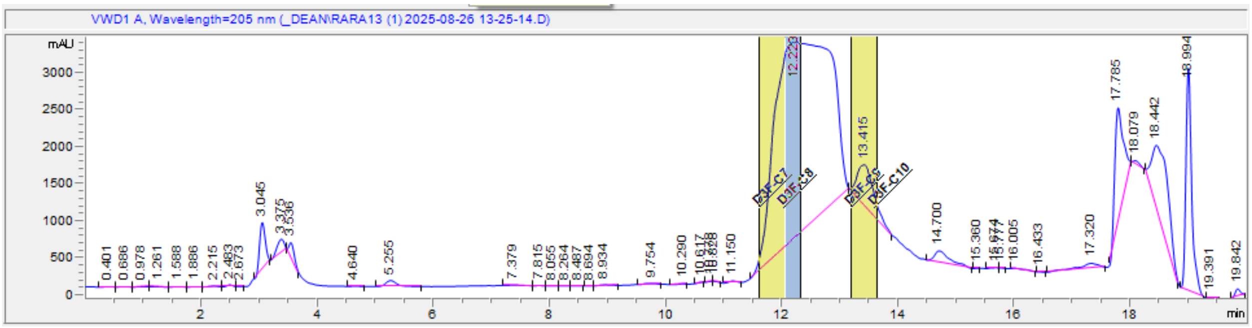


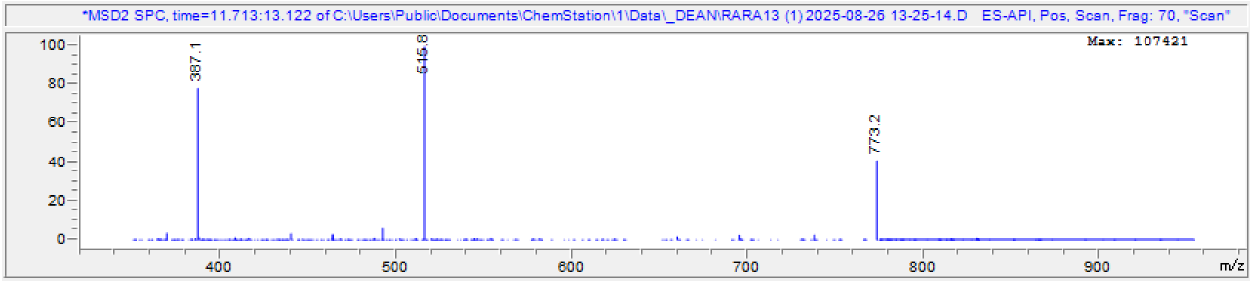


Despite weird peak shape, analytical chromatography shows purified sample to be 99.3% pure, like the RRAA-13 het-50% shown below. Most of these samples, however, were not tested on analytical mode due to issues with column fouling at the time these experiments were being performed. Future experiments must quantitatively establish the purity of each peptide.

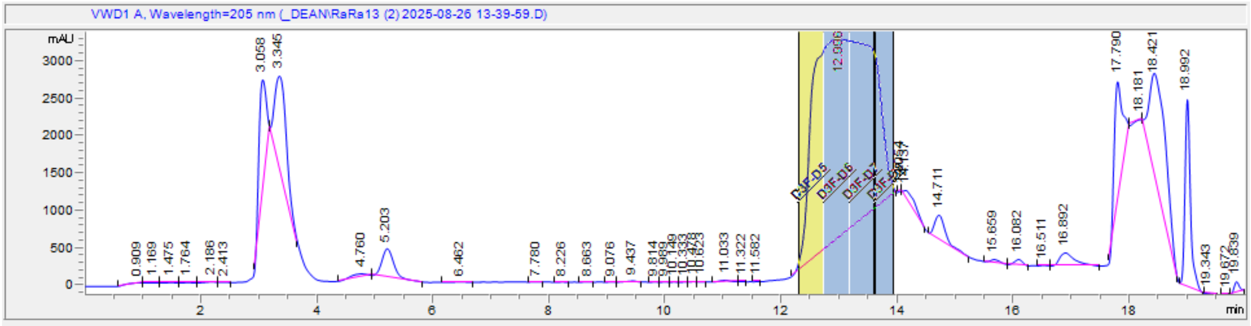


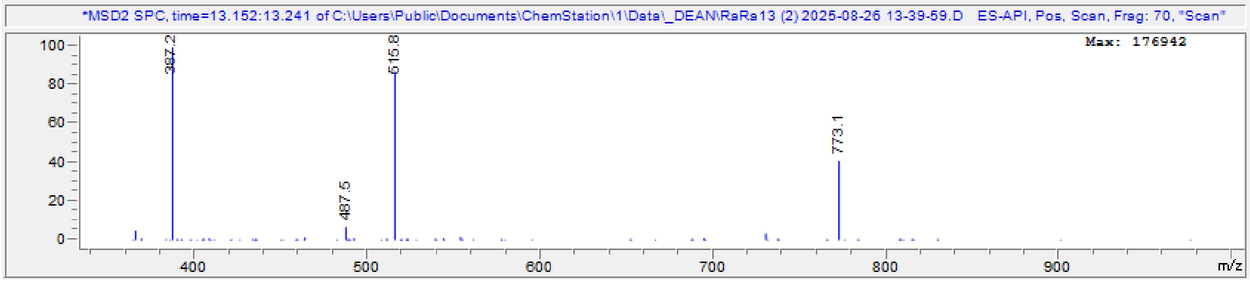
H2N- R A R A R A R A R A R A Y -OH; “RARA-13”





H2N- R a R a R a R a R a R a Y -OH; “RaRa-13 het-50”





Mass Spectrometry Sequencing Data

Mass spectrometry cannot distinguish the chirality of residues. Information gained from the following study is limited to peptide sequence.

H2N- R R a a R R a a R R a a Y -OH; “RRaa-13-het”

Calculated [M+H]$^+$ 1544.9105

Found 1544.9105 (<0.1 ppm)

Calculated fragment ions (**Bold** ions have been identified in MS/MS data):

Seq	#	B	Y	# (+1)
R	1	157.10843	1544.91056	13
R	**2**	**313.20954**	1388.80945	12
A	**3**	**384.24666**	1232.70834	11
A	**4**	**455.28377**	1161.67122	10
R	**5**	**611.38488**	**1090.63411**	9
R	**6**	**767.48599**	**934.53300**	8
A	7	838.52310	**778.43189**	7
A	8	909.56022	**707.39478**	6
R	9	1065.66133	**636.35766**	5
R	10	1221.76244	**480.25655**	4
A	11	1292.79955	**324.15544**	3
A	12	1363.83667	**253.11833**	2
Y	13	1526.89999	182.08121	1

3.10. References

(1) Uversky, V. N. Biological Liquid–Liquid Phase Separation, Biomolecular Condensates, and Membraneless Organelles: Now You See Me, Now You Don't. *Int. J. Mol. Sci.* **2023**, *24* (17), 13150. https://doi.org/10.3390/ijms241713150.

(2) Perry, S. L. Phase Separation: Bridging Polymer Physics and Biology. *Curr. Opin. Colloid Interface Sci.* **2019**, *39*, 86–97. https://doi.org/10.1016/j.cocis.2019.01.007.

(3) Singh, A. N.; Yethiraj, A. Liquid–Liquid Phase Separation As the Second Step of Complex Coacervation. *J. Phys. Chem. B* **2021**, *125* (12), 3023–3031. https://doi.org/10.1021/acs.jpcb.0c07349.

(4) Keating, C. D.; Pappu, R. V. Liquid–Liquid Phase Separation: A Widespread and Versatile Way to Organize Aqueous Solutions. *J. Phys. Chem. Lett.* **2021**, *12* (45), 10994–10995. https://doi.org/10.1021/acs.jpclett.1c03352.

(5) Jan-Stefan, V. Coacervation-Enhanced Peptide Catalysis. *Nat. Catal.* **2024**, *7* (11), 1144–1144. https://doi.org/10.1038/s41929-024-01267-5.

(6) Smokers, I. B. A.; Visser, B. S.; Slootbeek, A. D.; Huck, W. T. S.; Spruijt, E. How Droplets Can Accelerate Reactions─Coacervate Protocells as Catalytic Microcompartments. *Acc. Chem. Res.* **2024**, *57* (14), 1885–1895. https://doi.org/10.1021/acs.accounts.4c00114.

(7) Oh, K.-I.; You, X.; Flanagan, J. C.; Baiz, C. R. Liquid–Liquid Phase Separation Produces Fast H-Bond Dynamics in DMSO–Water Mixtures. *J. Phys. Chem. Lett.* **2020**, *11* (5), 1903–1908. https://doi.org/10.1021/acs.jpclett.0c00378.

(8) Courchaine, E. M.; Lu, A.; Neugebauer, K. M. Droplet Organelles? *EMBO J.* **2016**, *35* (15), 1603–1612. https://doi.org/10.15252/embj.201593517.

(9) Gomes, E.; Shorter, J. The Molecular Language of Membraneless Organelles. *J. Biol. Chem.* **2019**, *294* (18), 7115–7127. https://doi.org/10.1074/jbc.TM118.001192.

(10) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **2009**, *324* (5935), 1729–1732. https://doi.org/10.1126/science.1172046.

(11) Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **2015**, *163* (1), 123–133. https://doi.org/10.1016/j.cell.2015.09.015.

(12) Lin, Y.; Protter, D. S. W.; Rosen, M. K.; Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **2015**, *60* (2), 208–219. https://doi.org/10.1016/j.molcel.2015.08.018.

(13) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular Condensates: Organizers of Cellular Biochemistry. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (5), 285–298. https://doi.org/10.1038/nrm.2017.7.

(14) Li, Y.; Liu, Y.; Yu, X.-Y.; Xu, Y.; Pan, X.; Sun, Y.; Wang, Y.; Song, Y.-H.; Shen, Z. Membraneless Organelles in Health and Disease: Exploring the Molecular Basis, Physiological Roles and Pathological Implications. *Signal Transduct. Target. Ther.* **2024**, *9* (1), 305. https://doi.org/10.1038/s41392-024-02013-w.

(15) Thakur, D. K.; Padole, S.; Sarkar, T.; Arumugam, S.; Chattopadhyay, S. Liquid–Liquid Phase Separation: Mechanisms, Roles, and Implications in Cellular Function

and Disease. *FASEB BioAdvances* **2025**, *7* (11), e70054. https://doi.org/10.1096/fba.2025-00140.

(16) Yang, S.; Shen, W.; Hu, J.; Cai, S.; Zhang, C.; Jin, S.; Guan, X.; Wu, J.; Wu, Y.; Cui, J. Molecular Mechanisms and Cellular Functions of Liquid-Liquid Phase Separation during Antiviral Immune Responses. *Front. Immunol.* **2023**, *14*, 1162211. https://doi.org/10.3389/fimmu.2023.1162211.

(17) Choi, H.-J.; Lee, J. Y.; Kim, K. Glutathionylation on RNA-Binding Proteins: A Regulator of Liquid–liquid Phase Separation in the Pathogenesis of Amyotrophic Lateral Sclerosis. *Exp. Mol. Med.* **2023**, *55* (4), 735–744. https://doi.org/10.1038/s12276-023-00978-2.

(18) Zheng, L.-W.; Liu, C.-C.; Yu, K.-D. Phase Separations in Oncogenesis, Tumor Progressions and Metastasis: A Glance from Hallmarks of Cancer. *J. Hematol. Oncol.J Hematol Oncol* **2023**, *16* (1), 123. https://doi.org/10.1186/s13045-023-01522-5.

(19) Wei, W.; Bai, L.; Yan, B.; Meng, W.; Wang, H.; Zhai, J.; Si, F.; Zheng, C. When Liquid-Liquid Phase Separation Meets Viral Infections. *Front. Immunol.* **2022**, *13*, 985622. https://doi.org/10.3389/fimmu.2022.985622.

(20) Fu, Q.; Zhang, B.; Chen, X.; Chu, L. Liquid–Liquid Phase Separation in Alzheimer’s Disease. *J. Mol. Med.* **2024**, *102* (2), 167–181. https://doi.org/10.1007/s00109-023-02407-3.

(21) Zhang, L.; Wang, S.; Wang, W.; Shi, J.; Stovall, D. B.; Li, D.; Sui, G. Phase-Separated Subcellular Compartmentation and Related Human Diseases. *Int. J. Mol. Sci.* **2022**, *23* (10), 5491. https://doi.org/10.3390/ijms23105491.

(22) Galloux, M.; Longhi, S. Unraveling Liquid–Liquid Phase Separation (LLPS) in Viral Infections to Understand and Treat Viral Diseases. *Int. J. Mol. Sci.* **2024**, *25* (13), 6981. https://doi.org/10.3390/ijms25136981.

(23) Wang, B.; Zhang, L.; Dai, T.; Qin, Z.; Lu, H.; Zhang, L.; Zhou, F. Liquid–Liquid Phase Separation in Human Health and Diseases. *Signal Transduct. Target. Ther.* **2021**, *6* (1), 290. https://doi.org/10.1038/s41392-021-00678-1.

(24) Alemasova, E. E.; Lavrik, O. I. A sePARate Phase? Poly(ADP-Ribose) versus RNA in the Organization of Biomolecular Condensates. *Nucleic Acids Res.* **2022**, *50* (19), 10817–10838. https://doi.org/10.1093/nar/gkac866.

(25) Abbas, M.; Lipiński, W. P.; Wang, J.; Spruijt, E. Peptide-Based Coacervates as Biomimetic Protocells. *Chem. Soc. Rev.* **2021**, *50* (6), 3690–3705. https://doi.org/10.1039/D0CS00307G.

(26) Castelletto, V.; Seitsonen, J.; Pollitt, A.; Hamley, I. W. Minimal Peptide Sequences That Undergo Liquid–Liquid Phase Separation via Self-Coacervation or Complex Coacervation with ATP. *Biomacromolecules* **2024**, *25* (8), 5321–5331. https://doi.org/10.1021/acs.biomac.4c00738.

(27) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular Interactions Underlying Liquid−liquid Phase Separation of the FUS Low-Complexity Domain. *Nat. Struct. Mol. Biol.* **2019**, *26* (7), 637–648. https://doi.org/10.1038/s41594-019-0250-x.

(28) Chen, C.; Ding, X.; Akram, N.; Xue, S.; Luo, S.-Z. Fused in Sarcoma: Properties, Self-Assembly and Correlation with Neurodegenerative Diseases. *Molecules* **2019**, *24* (8), 1622. https://doi.org/10.3390/molecules24081622.

(29) Qamar, S.; Wang, G.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Ströhl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.; Tartaglia, G. G.; Farrer, L. A.; Kaminski Schierle, G. S.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **2018**, *173* (3), 720-734.e15. https://doi.org/10.1016/j.cell.2018.03.056.

(30) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174* (3), 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006.

(31) Sun, Z.; Diaz, Z.; Fang, X.; Hart, M. P.; Chesi, A.; Shorter, J.; Gitler, A. D. Molecular Determinants and Genetic Modifiers of Aggregation and Toxicity for the ALS Disease Protein FUS/TLS. *PLoS Biol.* **2011**, *9* (4), e1000614. https://doi.org/10.1371/journal.pbio.1000614.

(32) Xu, R. *Interrogation of Intermolecular Interactions in Liquid–Liquid Phase Separation*; The University of Wisconsin - Madison, 2025.

(33) Sanchez-Burgos, I.; Espinosa, J. R.; Joseph, J. A.; Collepardo-Guevara, R. RNA Length Has a Non-Trivial Effect in the Stability of Biomolecular Condensates Formed by RNA-Binding Proteins. *PLOS Comput. Biol.* **2022**, *18* (2), e1009810. https://doi.org/10.1371/journal.pcbi.1009810.

(34) Niu, J.; Qiu, C.; Abbott, N. L.; Gellman, S. H. Formation of versus Recruitment to RNA-Rich Condensates: Controlling Effects Exerted by Peptide Side Chain Identity. *J. Am. Chem. Soc.* **2022**, *144* (23), 10386–10395. https://doi.org/10.1021/jacs.2c02222.

(35) Hansen, J. C.; Maeshima, K.; Hendzel, M. J. The Solid and Liquid States of Chromatin. *Epigenetics Chromatin* **2021**, *14* (1), 50. https://doi.org/10.1186/s13072-021-00424-5.

(36) Yuan, C.; Li, Q.; Xing, R.; Li, J.; Yan, X. Peptide Self-Assembly through Liquid-Liquid Phase Separation. *Chem* **2023**, *9* (9), 2425–2445. https://doi.org/10.1016/j.chempr.2023.05.009.

(37) Dai, S.; Xie, Z.; Wang, B.; Ye, R.; Ou, X.; Wang, C.; Yu, N.; Huang, C.; Zhao, J.; Cai, C.; Zhang, F.; Buratto, D.; Khan, T.; Qiao, Y.; Hua, Y.; Zhou, R.; Tian, B. An Inorganic Mineral-Based Protocell with Prebiotic Radiation Fitness. *Nat. Commun.* **2023**, *14* (1), 7699. https://doi.org/10.1038/s41467-023-43272-5.

(38) Patel, A.; Malinovska, L.; Saha, S.; Wang, J.; Alberti, S.; Krishnan, Y.; Hyman, A. A. ATP as a Biological Hydrotrope. *Science* **2017**, *356* (6339), 753–756. https://doi.org/10.1126/science.aaf6846.

(39) Yao, Y.; Jin, Z.; Ling, B.; Malounda, D.; Shapiro, M. G. Self-Assembly of Protein Superstructures by Physical Interactions under Cytoplasm-like Conditions. *Biophys. J.* **2021**, *120* (13), 2701–2709. https://doi.org/10.1016/j.bpj.2021.05.007.

(40) Wei, M.-T.; Elbaum-Garfinkle, S.; Holehouse, A. S.; Chen, C. C.-H.; Feric, M.; Arnold, C. B.; Priestley, R. D.; Pappu, R. V.; Brangwynne, C. P. Phase Behaviour of Disordered Proteins Underlying Low Density and High Permeability of Liquid

Organelles. *Nat. Chem.* **2017**, *9* (11), 1118–1125. https://doi.org/10.1038/nchem.2803.

(41) Hong, Y.; Najafi, S.; Casey, T.; Shea, J.-E.; Han, S.-I.; Hwang, D. S. Hydrophobicity of Arginine Leads to Reentrant Liquid-Liquid Phase Separation Behaviors of Arginine-Rich Proteins. *Nat. Commun.* **2022**, *13* (1), 7326. https://doi.org/10.1038/s41467-022-35001-1.

(42) Banerjee, P. R.; Milin, A. N.; Moosa, M. M.; Onuchic, P. L.; Deniz, A. A. Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. *Angew. Chem. Int. Ed.* **2017**, *56* (38), 11354–11359. https://doi.org/10.1002/anie.201703191.

(43) Yewdall, N. A.; André, A. A. M.; Lu, T.; Spruijt, E. Coacervates as Models of Membraneless Organelles. *Curr. Opin. Colloid Interface Sci.* **2021**, *52*, 101416. https://doi.org/10.1016/j.cocis.2020.101416.

(44) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **2020**, *49* (1), 107–133. https://doi.org/10.1146/annurev-biophys-121219-081629.

(45) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains. *Science* **2020**, *367* (6478), 694–699. https://doi.org/10.1126/science.aaw8653.

(46) Holehouse, A. S.; Pappu, R. V. Functional Implications of Intracellular Phase Transitions. *Biochemistry* **2018**, *57* (17), 2415–2423. https://doi.org/10.1021/acs.biochem.7b01136.

(47) Ginell, G. M.; Holehouse, A. S. An Introduction to the Stickers-and-Spacers Framework as Applied to Biomolecular Condensates. In *Phase-Separated Biomolecular Condensates*; Zhou, H.-X., Spille, J.-H., Banerjee, P. R., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2023; Vol. 2563, pp 95–116. https://doi.org/10.1007/978-1-0716-2663-4_4.

(48) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176* (3), 419–434. https://doi.org/10.1016/j.cell.2018.12.035.

(49) Lorenz-Ochoa, K. A.; Lee, E.; Baiz, C. R. Beyond Electrostatic Interactions: H-Bond Dynamics in Arginine- versus Lysine-Rich Biocondensates. *J. Phys. Chem. Lett.* **2025**, *16* (36), 9288–9292. https://doi.org/10.1021/acs.jpclett.5c01743.

(50) Lorenz-Ochoa, K. A.; Baiz, C. R. Ultrafast Spectroscopy Reveals Slow Water Dynamics in Biocondensates. *J. Am. Chem. Soc.* **2023**, *145* (50), 27800–27809. https://doi.org/10.1021/jacs.3c10862.

(51) Abyzov, A.; Blackledge, M.; Zweckstetter, M. Conformational Dynamics of Intrinsically Disordered Proteins Regulate Biomolecular Condensate Chemistry. *Chem. Rev.* **2022**, *122* (6), 6719–6748. https://doi.org/10.1021/acs.chemrev.1c00774.

(52) Guseva, S.; Schnapka, V.; Adamski, W.; Maurin, D.; Ruigrok, R. W. H.; Salvi, N.; Blackledge, M. Liquid–Liquid Phase Separation Modifies the Dynamic Properties of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2023**, *145* (19), 10548–10563. https://doi.org/10.1021/jacs.2c13647.

(53) Lorenz-Ochoa, K. A.; Cho, M.; Parekh, S. H.; Baiz, C. R. Interaction-Dependent Secondary Structure of Peptides in Biomolecular Condensates. *J. Am. Chem. Soc.* **2024**, *146* (49), 33616–33625. https://doi.org/10.1021/jacs.4c11226.

(54) Choi, S.; Knoerdel, A. R.; Sing, C. E.; Keating, C. D. Effect of Polypeptide Complex Coacervate Microenvironment on Protonation of a Guest Molecule. *J. Phys. Chem. B* **2023**, *127* (26), 5978–5991. https://doi.org/10.1021/acs.jpcb.3c02098.

(55) Lim, S.; Clark, D. S. Phase-Separated Biomolecular Condensates for Biocatalysis. *Trends Biotechnol.* **2024**, *42* (4), 496–509. https://doi.org/10.1016/j.tibtech.2023.10.003.

(56) Harris, R.; Berman, N.; Lampel, A. Coacervates as Enzymatic Microreactors. *Chem. Soc. Rev.* **2025**, *54* (9), 4183–4199. https://doi.org/10.1039/D4CS01203H.

(57) Zhang, Y.; Narlikar, G. J.; Kutateladze, T. G. Enzymatic Reactions inside Biological Condensates. *J. Mol. Biol.* **2021**, *433* (12), 166624. https://doi.org/10.1016/j.jmb.2020.08.009.

(58) Lazcano, A. Alexandr I. Oparin and the Origin of Life: A Historical Reassessment of the Heterotrophic Theory. *J. Mol. Evol.* **2016**, *83* (5–6), 214–222. https://doi.org/10.1007/s00239-016-9773-5.

(59) Oparin, A. *Происхождение Жизни*; Moscow, 1924.

(60) Prosdocimi, F.; Farias, S. T. D. Coacervates Meet the RNP-World: Liquid-Liquid Phase Separation and the Emergence of Biological Compartmentalization. *BioSystems* **2025**, *252*, 105480. https://doi.org/10.1016/j.biosystems.2025.105480.

(61) Hansma, H. G. Liquid–Liquid Phase Separation at the Origins of Life. In *Droplets of Life*; Elsevier, 2023; pp 251–268. https://doi.org/10.1016/B978-0-12-823967-4.00006-3.

(62) Poudyal, R. R.; Pir Cakmak, F.; Keating, C. D.; Bevilacqua, P. C. Physical Principles and Extant Biology Reveal Roles for RNA-Containing Membraneless Compartments in Origins of Life Chemistry. *Biochemistry* **2018**, *57* (17), 2509–2519. https://doi.org/10.1021/acs.biochem.8b00081.

(63) Pokhrel, P.; Zhang, Z.; Ji, J.; Shakya, S.; Jenyk, J.; Lower, A.; Janssen, M.; Mao, H. Chiral Matching between Nucleic Acids and Polypeptides Facilitates Liquid–Liquid Phase Separation. *J. Phys. Chem. Lett.* **2025**, *16* (36), 9545–9552. https://doi.org/10.1021/acs.jpclett.5c02218.

(64) Mirlohi, K.; Thapa, A.; Zhang, D.; Blocher McTigue, W. C. The Effects of Chirality and Salt Addition on Phase Separation and Complexation Morphology in Mixtures of Polypeptides and Polypeptoids. *Biomacromolecules* **2025**, *26* (11), 7646–7655. https://doi.org/10.1021/acs.biomac.5c01165.

(65) Perry, S. L.; Leon, L.; Hoffmann, K. Q.; Kade, M. J.; Priftis, D.; Black, K. A.; Wong, D.; Klein, R. A.; Pierce, C. F.; Margossian, K. O.; Whitmer, J. K.; Qin, J.; De Pablo, J. J.; Tirrell, M. Chirality-Selected Phase Behaviour in Ionic Polypeptide Complexes. *Nat. Commun.* **2015**, *6* (1), 6052. https://doi.org/10.1038/ncomms7052.

(66) Seal, M.; Edelstein, I.; Scolnik, Y.; Weil-Ktorza, O.; Metanis, N.; Levy, Y.; Longo, L. M.; Goldfarb, D. RNA Binding and Coacervation Promote Preservation of Peptide

Form and Function across the Heterochiral–Homochiral Divide. *Protein Sci.* **2025**, *34* (9), e70273. https://doi.org/10.1002/pro.70273.

(67) van der Stok, A.; Gellman, S. H. Preference for Homochiral Cationic Peptides in Liquid-Liquid Phase Separation with Polyanions. *Unpubl. Data - Univ. Wis. - Madison* **2026**.

(68) DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. Protein Design, a Minimalist Approach. *Science* **1989**, *243* (4891), 622–628. https://doi.org/10.1126/science.2464850.

(69) Fisher, B. F.; Hong, S. H.; Gellman, S. H. Helix Propensities of Amino Acid Residues via Thioester Exchange. *J. Am. Chem. Soc.* **2017**, *139* (38), 13292–13295. https://doi.org/10.1021/jacs.7b07930.

(70) Xu, R.; Huang, J.; Kuhn, A. J.; Gellman, S. H. Effects of D-Amino Acid Replacements on the Conformational Stability of Miniproteins. *ChemBioChem* **2025**, *26* (7), e202500085. https://doi.org/10.1002/cbic.202500085.

(71) Glonek, T. Did Cyclic Metaphosphates Have a Role in the Origin of Life? *Orig. Life Evol. Biospheres* **2021**, *51* (1), 1–60. https://doi.org/10.1007/s11084-021-09604-5.

(72) Yamagata, Y.; Watanabe, H.; Saitoh, M.; Namba, T. Volcanic Production of Polyphosphates and Its Relevance to Prebiotic Evolution. *Nature* **1991**, *352* (6335), 516–519. https://doi.org/10.1038/352516a0.

(73) Goebel-Stengel, M.; Stengel, A.; Taché, Y.; Reeve, J. R. The Importance of Using the Optimal Plasticware and Glassware in Studies Involving Peptides. *Anal. Biochem.* **2011**, *414* (1), 38–46. https://doi.org/10.1016/j.ab.2011.02.009.

(74) Huang, Y.; Bai, Y.; Jin, W.; Shen, D.; Lyu, H.; Zeng, L.; Wang, M.; Liu, Y. Common Pitfalls and Recommendations for Using a Turbidity Assay to Study Protein Phase Separation. *Biochemistry* **2021**, *60* (32), 2447–2456. https://doi.org/10.1021/acs.biochem.1c00386.

(75) Van Lindt, J.; Bratek-Skicki, A.; Nguyen, P. N.; Pakravan, D.; Durán-Armenta, L. F.; Tantos, A.; Pancsa, R.; Van Den Bosch, L.; Maes, D.; Tompa, P. A Generic Approach to Study the Kinetics of Liquid–Liquid Phase Separation under near-Native Conditions. *Commun. Biol.* **2021**, *4* (1), 77. https://doi.org/10.1038/s42003-020-01596-8.

(76) Raymond-Smiedy, P.; Bucknor, B.; Yang, Y.; Zheng, T.; Castañeda, C. A. A Spectrophotometric Turbidity Assay to Study Liquid-Liquid Phase Separation of UBQLN2 In Vitro. In *Protein Aggregation*; Cieplak, A. S., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2023; Vol. 2551, pp 515–541. https://doi.org/10.1007/978-1-0716-2597-2_32.

(77) Zheng, Y.; Mao, K.; Chen, S.; Zhu, H. Chirality Effects in Peptide Assembly Structures. *Front. Bioeng. Biotechnol.* **2021**, *9*, 703004. https://doi.org/10.3389/fbioe.2021.703004.

(78) Anthis, N. J.; Clore, G. M. Sequence-specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Sci.* **2013**, *22* (6), 851–858. https://doi.org/10.1002/pro.2253.

Chapter 4

Regiochemical Modification of Terphenyl Amino Acids and their Oligomers

In collaboration with Leah Garman, Ilia Guzei, Lars Schimmelpfennig and Charles Lenz (UW-Madison), and Jieye Lin, Johan Unge, and Tamir Gonen (UCLA)

4.1. Abstract: Regiochemical Modification of Terphenyl Amino Acids and their Oligomers

Here we describe series of oligomers derived from terphenyl amino acids. The structure of these terphenyl oligomers might be stabilized by hydrogen bonding, aromatic interactions, or both. Several different regiochemical isomers were synthesized and their folding and self-associating properties were investigated with spectroscopic and crystallographic techniques. Although high-resolution structural characterization remains a challenge for long oligomers, x-ray crystallography and microcrystal electron diffraction have been successful at determining structures of terphenyl monomers and dimers, respectively.

4.2. Folding in Biomolecules

Numerous peptides, proteins, and nucleic acids adopt specific and intricate three-dimensional structures. Stable 3D structures endow many biomolecules with their specific biological function.[1–3] Enzymes, for example, are only functional when the active site has the correct size, geometry and contacts to catalyze the reaction of the substrate(s).[4,5] Certain biomolecules must have a significant thermodynamic benefit to reliably fold lest they are tagged for proteolysis, or in some cases, cause disease.[2,6]

The lowest energy conformation of a protein or nucleic acid is stabilized by non-covalent interactions such as hydrogen bonding, cation-anion interactions, cation-π interactions, aromatic interactions and hydrophobic effects.[3,7–9] In cells, other effects such as crowding, ion pairing or chaperones may also play a role.[10–13] Heat or the disruption of stabilizing interactions can cause denaturation of a biomolecule.[9,14–16]

4.2.1. Foldamers: Structure and Function

Nature has endowed numerous peptides, proteins, and nucleic acids secondary and tertiary structure that these molecules to engage in specific interactions. Organic chemists may draw inspiration from these classes of biological molecules to design synthetic molecules that can fold into compact conformations, possibly enabling biological or catalytic functions.[17–19] The term 'foldamer' is used to describe any synthetic polymer that adopts a specific compact conformation.[17]

Numerous synthetic foldamers have been characterized, many of which contain peptide backbones.[20] Natural peptides, made from L- α-amino acids, adopt a relatively small number of regular secondary structures, such as α-helices or β-sheets, from which we may gain inspiration in a quest for synthetic foldamers. Helices are among the most readily accessible structures for synthetic foldamers.[21] Peptidic helices (**Figure 4.1.A**) are stabilized by hydrogen bonding interactions between amide carbonyls and nearby amide

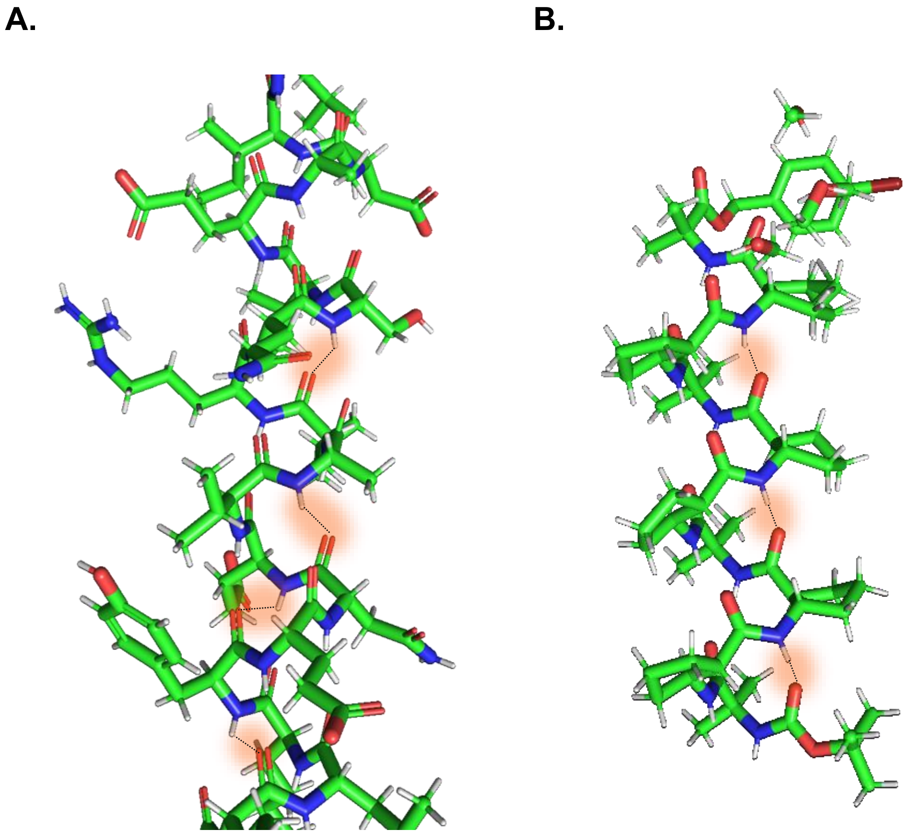


Figure 4.1 Two helical amino acid sequences: A) portion of α-helical keratin-1 structure (PDB: 6EC0),[23,24] a coiled-coil human intermediate filament protein and B) a synthetic foldamer (CSD: PUCDEX) with 1:2 α/β residues.[30] While both structures are helical and stabilized by iterturn hydrogen bonding, α-helical keratin has 3.5 amino acid residues per turn while the amide bonds of the foldamer, like the residues, are designed to be nearly perfectly aligned.[30] Several backbone interturn hydrogen bonds are represented in each structure by dotted lines highlighted in orange.

N-H hydrogen atoms. In α-helices with ~3.5 amino acid residues per turn, the carbonyl and the N-H are always four residues away from each other (*i,i*+4 amide hydrogen bonding).[22–24]

The incorporation of non-canonical amino acids into synthetic peptides enables the formation of new secondary structures.[20,25] For example, the 1:1 alteration of α and γ-amino acids into certain peptides can favor alternating 12/10 helices, where there are either 12 or 10 atoms in a ‘ring’ formed by hydrogen bonds, with alternating H-bond orientation.[26] Other foldamers with a 1:2 α:β residue ratio and a αββ backbone repeat pattern form helices with ~3 residues per turn, in which residues with i,i+3 spacing are aligned along one side of the helix (**Figure 4.1.B**).[27,28] This ‘stacked’ geometry has been used for catalysis by incorporating reactive side chains into two aligned residues.[29,30] Iminium/enamine-catalyzed aldol reactions, achieved by incorporating a primary and a secondary amine into an αββ foldamer, have displayed a 150-fold rate enhancement relative to the same aldol reaction catalyzed by a pair of simple primary and secondary amines.[30]

While many amino acids that contain *sp*^{3} carbon atoms can form foldamer helices with interturn H-bonding (**Figure 4.1**),[17,18] others have synthesized foldamers from amino acids containing exclusively *sp*^{2} carbon atoms featuring *intraturn* H-bonding.[31–33] Most examples of *sp*^{2} foldamers are composed of subunits that are largely planar, and the helical structure is stabilized by the aromatic stacking reactions between residues in adjacent turns (**Figure 4.2**).[31–34]

Foldamers are relevant to a variety of applications. The catalytic potential of foldamers has already been highlighted in the context of the aldol reaction, but foldamers have the ability to catalyze other reactions as well.[19] Other foldamers have been designed to bind small molecules or even ions.[35–38] Foldamers have even been developed to inhibit fusion of SARS-CoV-2 and other viruses with host cells.[39,40] Development of new foldamers and understanding how they fold might enable new generations of foldamers with diverse biological or catalytic functions.

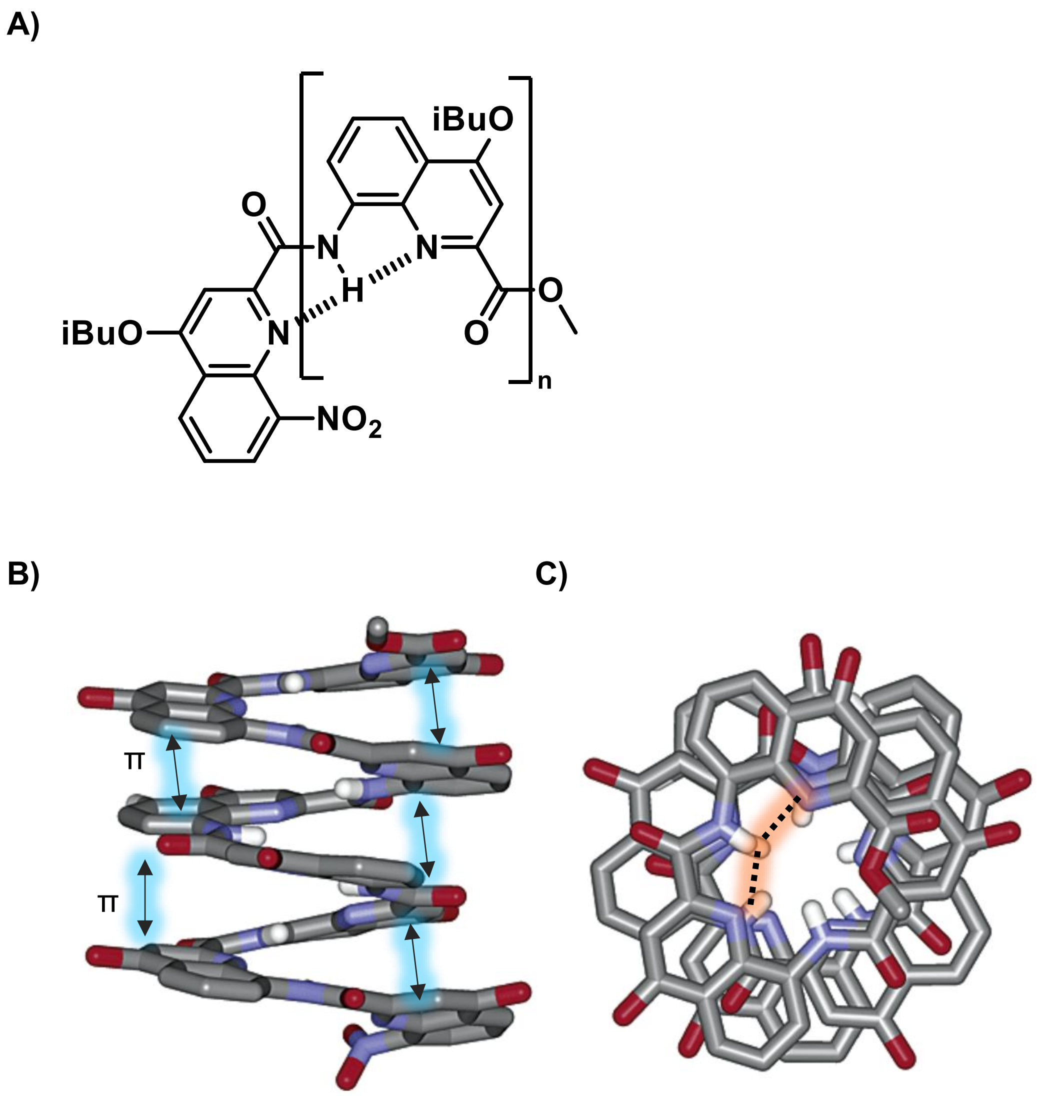


Figure 4.2 Synthetic *sp*^{2} foldamer from Jiang et al.[34] A) Structure of a foldamer subunit. B) Side view showing aromatic stacking (blue) and C) showing intraturn H-bonds (orange). Figure adapted from Jiang et al.[34]

4.2.2. Terphenyl as a Basis for New Amino Acids

While the foldamers described in the previous section can be stabilized by either interturn H-bonding (*sp*^{3}) or aromatic interactions (*sp*^{2}), a potential foldamer first designed by Adam Kleman and Samuel Gellman based on an *o*-terphenyl amino acid has the potential to be stabilized by both interactions (**Figure 4.3**).[41] A single *sp*^{3} carbon atom in the backbone provides extra flexibility in the backbone, which might allow hydrogen bonding to occur between adjacent turns. The aromatic rings of the terphenyl might align in such a way that aromatic interactions also stabilize adjacent turns. Hydrogen atoms at the *ortho*- positions on the terphenyl prevent the structure from being planar (**Figure 4.4**), causing some inherent twisting to a polymer containing *o*-terphenyl amino acids.[42] The methoxy- group on the central aryl ring was determined to be necessary to increase the solubility of these molecules.[43]

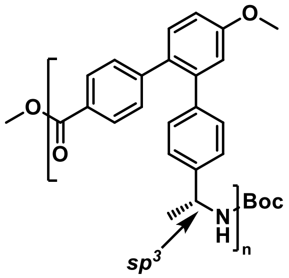


Figure 4.3 Design of *o*-terphenyl construct by Kleman et al. featuring a single *sp*^{3} carbon in the backbone.[41]

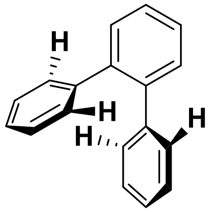


Figure 4.4 *o*-Terphenyl structure cannot be planar due to hydrogen atoms.

We later designated the terphenyl oligomer design by Klemen et al.[41] the *pop-R*-Tph-*n* series. '*pop'* refers to the regiochemical relationships among aromatic substituents: from the N-terminus, there is a *para*- relationship with the next aryl group; the N- and C-terminal phenylene rings are *ortho*- to each other; the C-terminal phenylene is *para*-substituted. The *R* refers to the stereochemistry at the *sp*^{3} carbon, and the *n* refers to the length of the oligomer (1 for a monomer, 2 for a dimer, etc.).

Kleman et al. conducted a significant amount of work on these amino acids and their oligomers, synthesizing several generations of compounds as well as hybrid peptides containing both α-amino acids and terphenyl amino acids.[41,43] Chiral *o*-terphenyl amino acids were particularly useful since conformational behavior could be investigated by circular dichroism (CD) spectroscopy.[44,45]

Kleman's *o*-terphenyl amino acid oligomers showed evidence of both folding behavior and self-association. CD spectra of longer oligomers result in more intense mean residue ellipticity (MRE, normalized according to concentration and oligomer length), suggesting there is length-dependent cooperativity.[46] ^{1}H NMR spectra of *o*-terphenyl oligomers in $CDCl_3$ contained mostly broad multiplets, which is consistent with the formation of folded (or aggregated) forms. In contrast, NMR spectra in DMSO-*d*$_6$ displayed sharper resonances. Kleman et al. concluded that folding or aggregation in $CDCl_3$ was driven at least in part by intermolecular H-bonding, and that the strong H-bonding propensity of DMSO prevented this mode of self-assembly (**Figure 4.5**).[41]

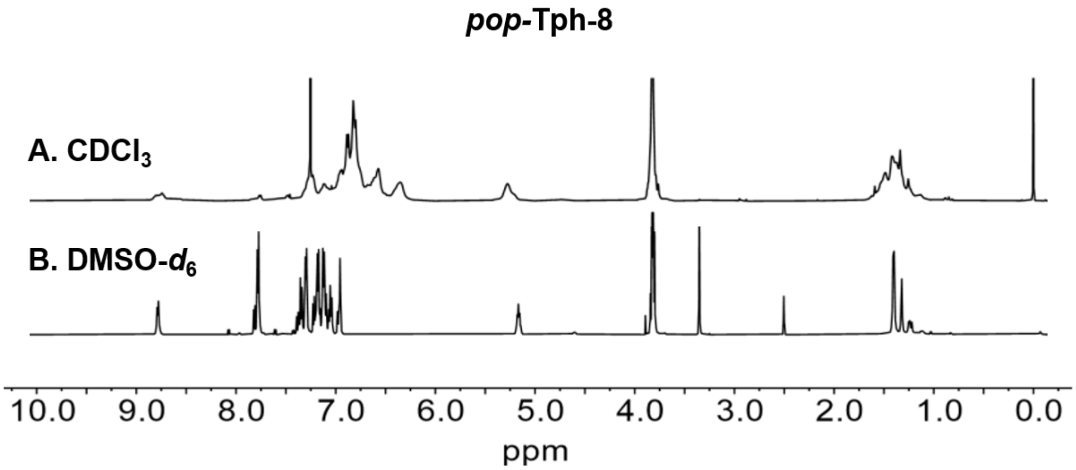


Figure 4.5 500 MHz ^{1}H NMR spectra of 10 mM *pop*-Tph-8 made and acquired by Kleman et al.[41] acquired in A) $CDCl_3$ and B) DMSO-*d*$_6$.

Evidence for the self-association of the original chiral *o*-terphenyl oligomers was found with both CD spectroscopy and diffusion-ordered NMR spectroscopy (DOSY). Both folding and self-association may cause a CD signature to change; these two possible causes can be distinguished by conducting variable-concentration measurements. A change caused by the onset of folding, for example, in a series of oligomers with increasing length, will be identified if the CD signatures do not change with variation in oligomer concentration. If concentration-dependent variations are observed, then one can conclude that changes in association state occur. When CD spectra of the *pop*-terphenyl oligomers were acquired at varying concentrations and plotted in terms of MRE, it was observed that higher concentrations resulted in a greater intensity for minima and maxima (**Figure 4.6**). These observations were interpreted to suggest that these *o*-terphenyl oligomers undergo self-association and folding simultaneously.[41,47] Further evidence was collected with DOSY.[48–50] The diffusion constant, which has an inverse relationship with molecular cross section (small molecules diffuse more quickly than large molecules), became smaller as *o*-terphenyl oligomer concentration was made greater (**Table 4.1**).[41] This observation suggests that larger assemblies form at higher concentrations.

Only limited conclusions could be drawn from the *pop*-terphenyl system. High resolution data could not be obtained because these oligomers did not form high-quality crystals.[41,43,51] In addition, the tendency of these molecules to self-associate in non-polar solvents hampered computational modeling efforts.[41]

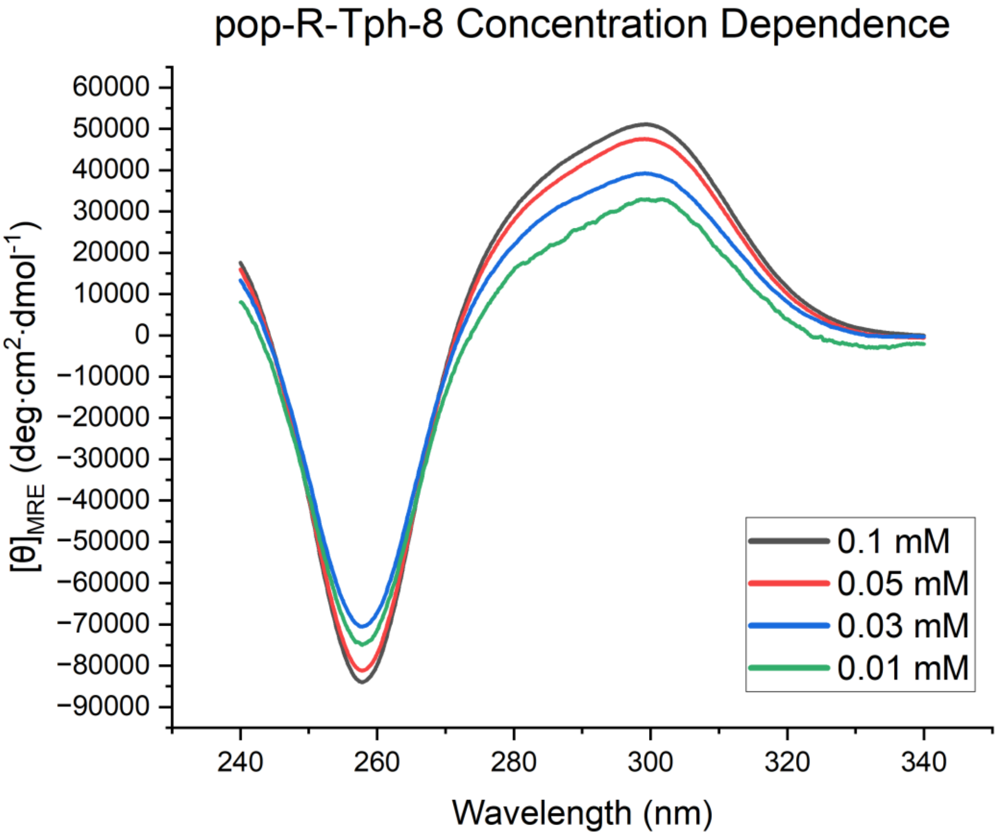


Figure 4.6 Normalized CD spectra of *o*-terphenyl octamer (*pop*-*R*-Tph-8) in chloroform at room temperature as a function of octamer concentration.[41] At the ellipticity maximum (300 nm), increasing oligomer concentrations is accompanied by a higher MRE. Since MRE is normalized to concentration, a second-order effect, such as self-association, is likely responsible for the observed variations in ellipticity, at least in part. Data generated by Kleman, A.

Table 4.1 Diffusion coefficients of *pop*-Tph-*n* in $CDCl_3$ determined by DOSY NMR by Kleman et al.[41] and normalized to diffusion of tetramethylsilane.

	Relative Diffusion Coefficient		
[Oligomer] /mM	pop-Tph-2	pop-Tph-4	pop-Tph-8
0.03			0.114
0.1	0.252	0.1949	
0.3			0.1165
1	0.2594	0.1811	0.1135
5			0.1147
10	0.2594	0.1759	
50	0.2504	0.1469	

To continue the work pioneered by Kleman et al., we investigated several regioisomers of the *o*-terphenyl amino acid he used. The original regiochemistry was characterized by the N-terminal methylbenzylamine and the C-terminal methyl benzoate connected to the middle *o*-phenylene with a *pop-* substitution pattern. We rationalized that greater structural flexibility might be achieved by changing any of these substitutional relationships to *meta-*, which might allow the structure to more easily adopt an arrangement conducive to crystallization. We therefore set out to synthesize *mop-*, *pom-*, and *pmp-* terphenyl oligomers (**Figure 4.7**) to determine whether any might adopt a stable secondary structure in the monomeric form.

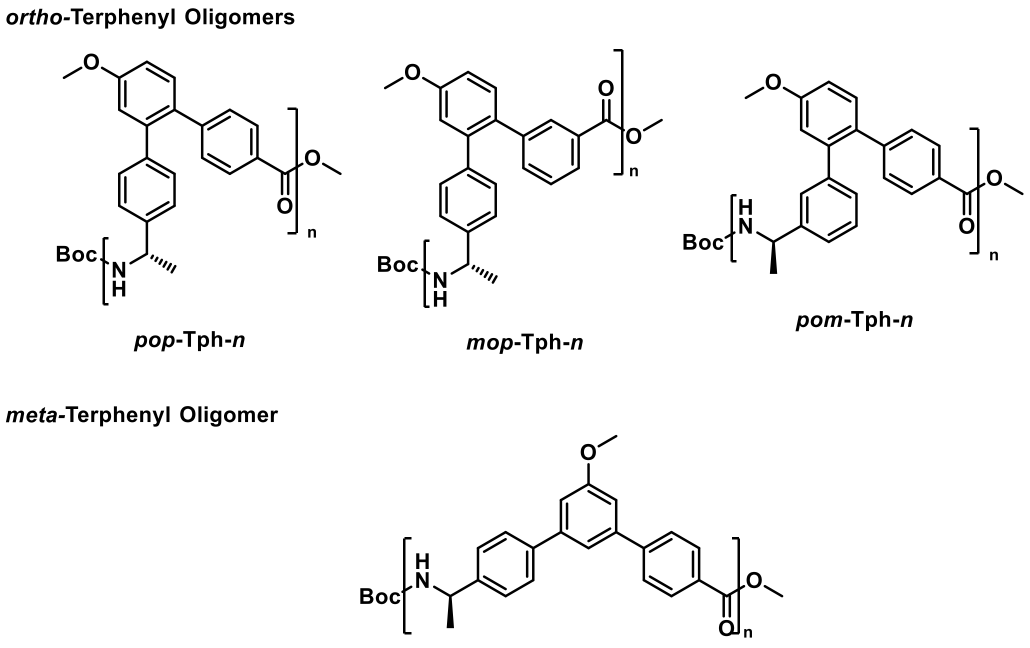
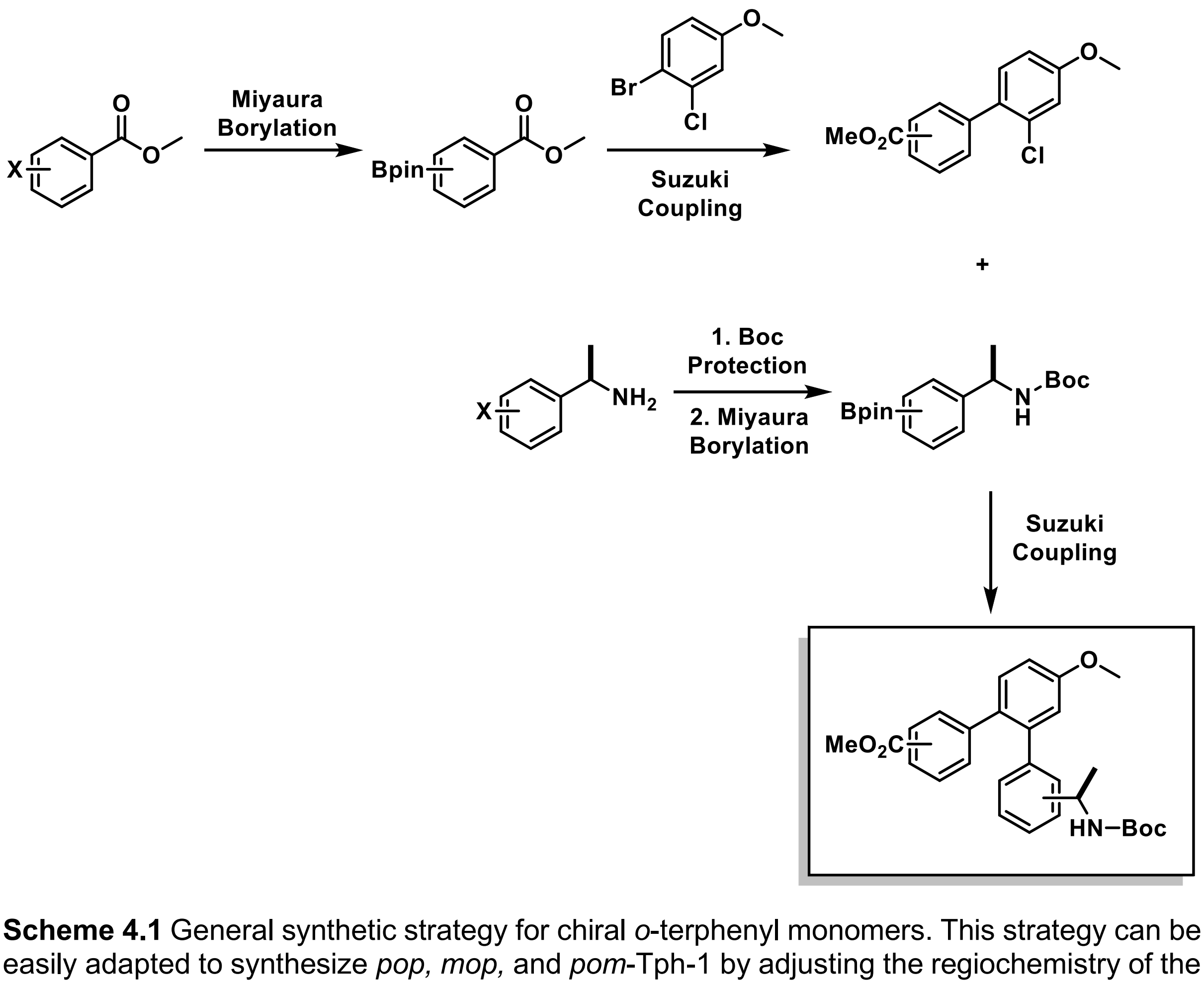


Figure 4.7 Structures of *pop*-Tph-*n* and the regioisomers described in this chapter.

4.3. Synthesis and Characterization of New Terphenyl Amino Acids

Following the general approach laid out by Kleman et al.,[41] terphenyl amino acids were synthesized by consecutive Miyaura borylations and Suzuki couplings as shown in **Scheme 4.1** A methyl benzoate bearing a boronic pinacol ester (Bpin) was either purchased or synthesized from the aryl halide. A mild, selective Suzuki coupling was then employed to couple the boronic ester to the aryl bromide of a bromo-chloro-anisole, resulting in a biaryl chloride.[52] Meanwhile, a halo(α-methylbenzylamine) was Boc-protected and converted to a boronic pinacol ester via Miyaura borylation. A final Suzuki coupling between the Bpin(α-methylbenzyl-t-butyl carbamate) and the biaryl chloride resulted in the terphenyl amino acid with a C-terminal methyl ester and an N-terminal NHBoc.



Scheme 4.1 General synthetic strategy for chiral *o*-terphenyl monomers. This strategy can be easily adapted to synthesize *pop, mop,* and *pom*-Tph-1 by adjusting the regiochemistry of the starting materials. *m*-terphenyl monomers can be synthesized as well by using 1-bromo-3-chloro-5-methoxybenzene as the anisole.

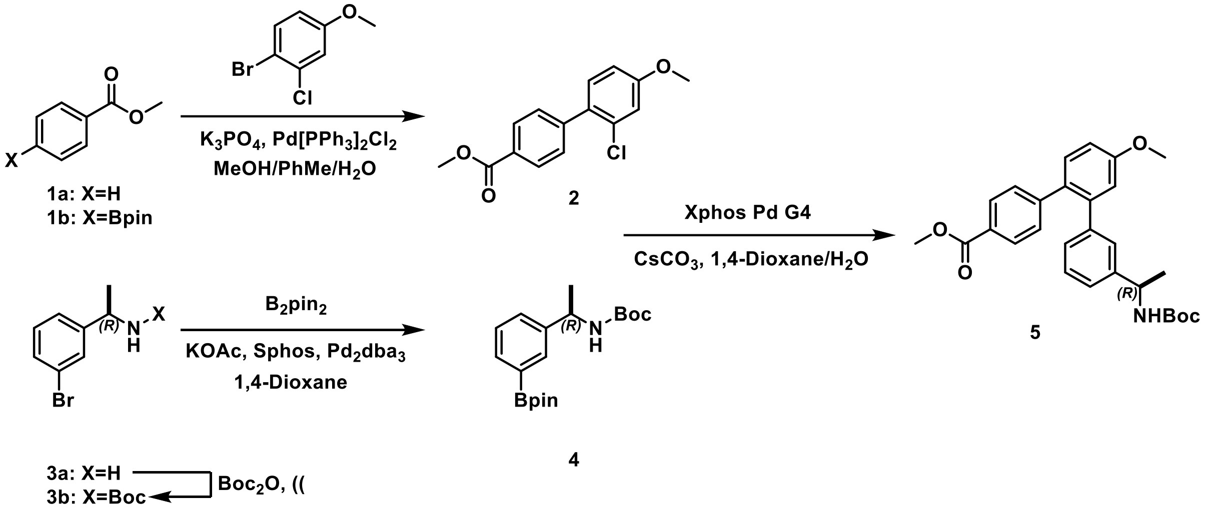
4.3.1. *mop*-Terphenyl Amino Acids

The synthesis of *mop*-terphenyl monomer **5**, outlined in **Scheme 4.2**, started with a selective Suzuki coupling of commercially available methyl 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzoate **1b** and 3-bromo-2-chloroanisole using potassium phosphate tribasic and catalytic bis(triphenylphosphine)palladium chloride.[52] The choice of solvent was critical for this reaction; high proportions of methanol or dioxane resulted in coupling with both the aryl bromide as well as the aryl chloride.[53–55] Using only toluene resulted in poor yields and overnight reaction time. Dissolving the organic reactants in 85:15 toluene/methanol and adding the base as a 2 M aqueous solution resulted in consistent 5 hour reaction times and >50% yield. The resulting biaryl **2** was recrystallized in methanol.

Original work by Kleman et al. built the N-terminal fragment starting with the Boc protection (R)-(+)-1-(4-chlorophenyl)ethylamine.[41] We needed *meta* stereochemistry on this fragment to synthesize the *mop*-terphenyl monomer. We found it more economical to use (R)-(+)-1-(3-**bromo**phenyl)ethylamine **3a**, which is less expensive than the chloride and improves the rate and yield of the following Miyaura borylation.[56] Boc protection of (R)-(+)-1-(3-bromophenyl)ethylamine was achieved quickly by reacting it neat with 1.1 eq di-*t*-butyl dicarbonate under sonication until carbon dioxide production ceased (15 min).[57] Workup and azeotropic removal of *t*-butanol afforded clean Boc-protected aryl bromide **3b** for the following Miyaura borylation, which was conducted via the protocol of Kleman et al.[41] to yield **4**.

The final Suzuki coupling between the biaryl chloride **2** and 1-(3-Bpin-phenyl)ethylamine **4** was the most challenging step in this synthesis. Originally, we tried to use the procedure adapted from Kleman et al., which employed palladium(II) acetate and Sphos ligand in water/THF.[41] Commercial palladium(II) acetate is known to be of inconsistent quality,[58,59] so we instead chose to test Xphos Pd G4 as a catalyst.[60] At the same time, we changed the solvent from 3:1 THF:water to 4:1 1,4-dioxane:water, which has a higher boiling point and would allow us to reflux this reaction mixture at a higher temperature, and we used Cs_2CO_3 as a base. These changes allowed us to prepare the desired *mop*-terphenyl monomer **5**, "*mop*-Tph-1," in 84% yield on a multigram scale

.

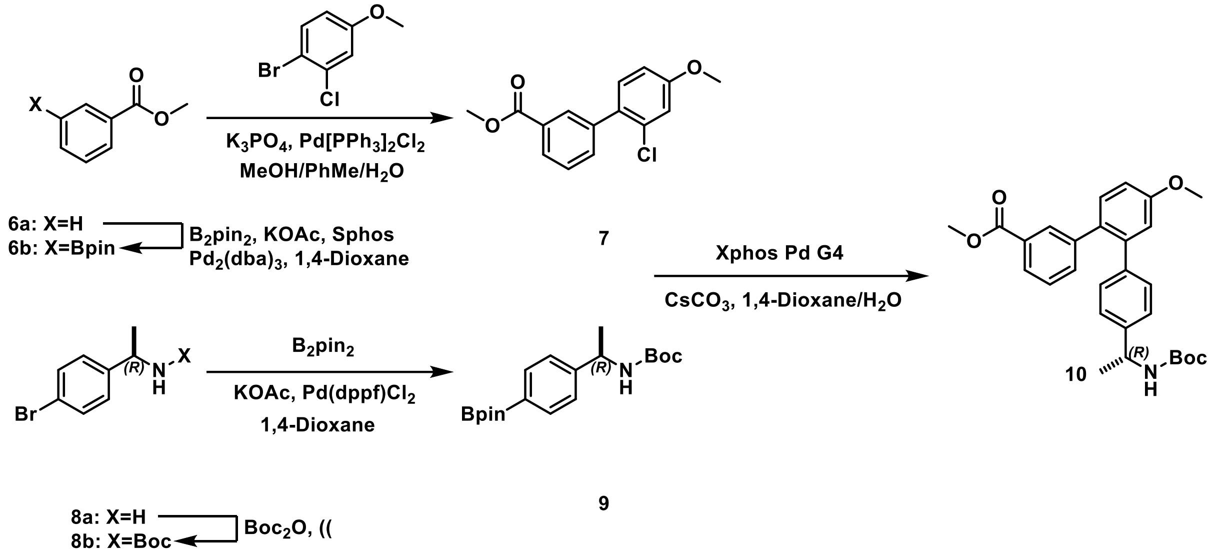


Scheme 4.2 Synthesis of *mop*-Tph-1 **5** using optimized procedures.

4.3.2. *pom*-Terphenyl Amino Acids

Taking into account the lessons learned from the synthesis of *mop*-terphenyl monomer **5**, we were able to execute the synthesis of *pom*-terphenyl monomer **10** in straightforward manner (**Scheme 4.3**). The C-terminal fragment was synthesized starting with a Miyaura borylation of methyl 3-chlorobenzoate **6a** to form **6b**.[61] This compound was easily coupled to 3-bromo-2-chloroanisole with the conditions in section 4.3.1. to form **7.** The N-terminus started with the Boc protection of (R)-(+)-1-(4-bromophenyl)ethylamine

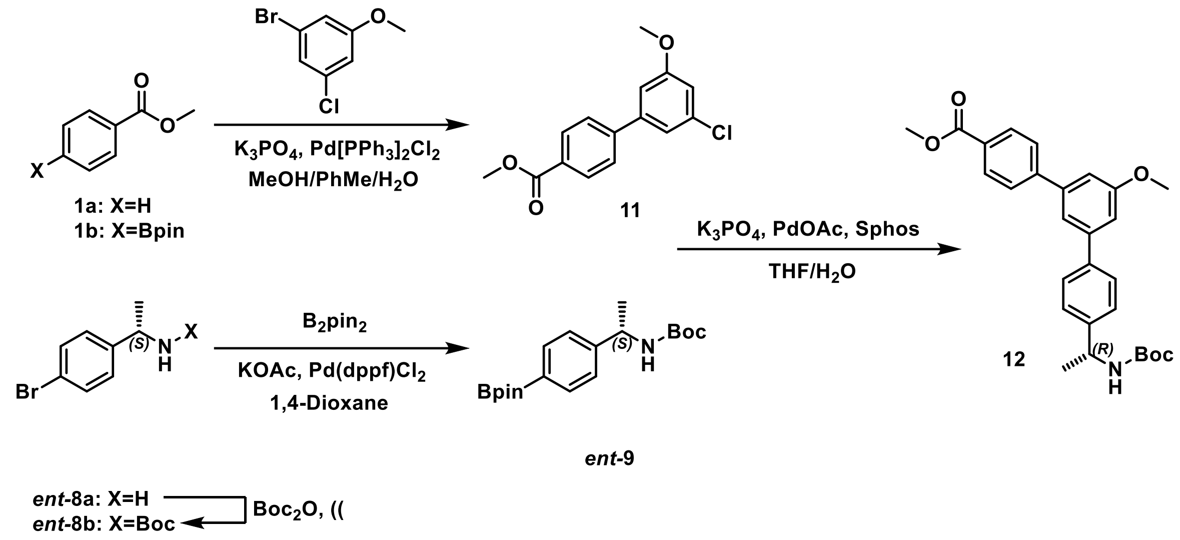
8a under neat conditions. Borylation was achieved to generate **9** using a procedure adapted from Kleman et al.[41] The final Suzuki coupling to make the *pom*-terphenyl monomer *pom*-Tph-1 **10** was conducted using the procedure that was used to make the *pom* monomer **5**.



Scheme 4.3 Synthesis of *pom*-Tph-1 **10** using optimized procedures.

4.3.3. *pmp*-Terphenyl Amino Acids

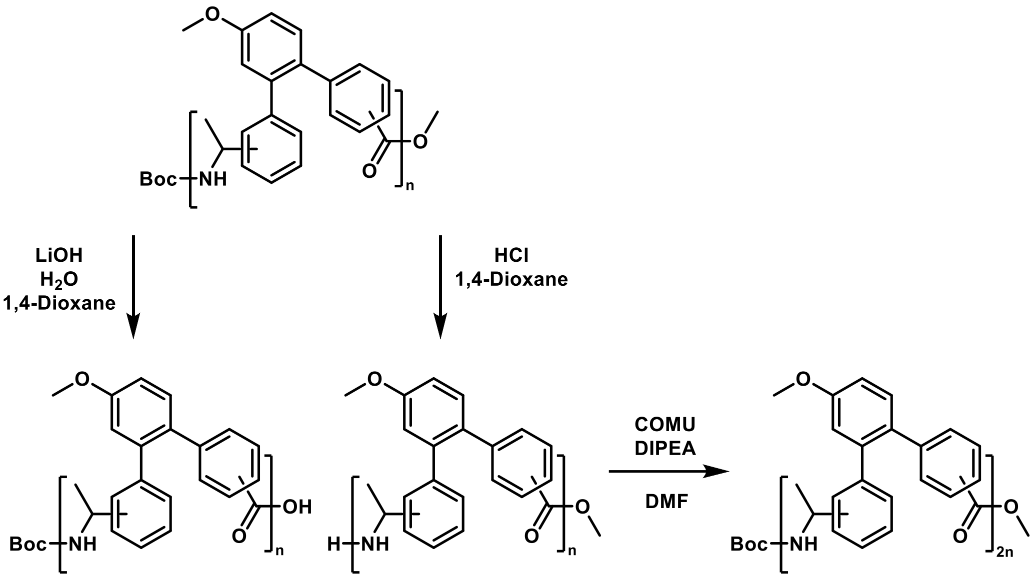
Synthesis of the *pmp*-terphenyl monomer **12** (**Scheme 4.4**) progressed in a similar fashion to the *mop*- and *pom*- isomers in sections 4.3.1. and 4.3.2. The synthesis started with the Miyaura borylation of methyl 4-chlorobenzoate **1a.**[61] The boronic ester was coupled to 1-bromo-3-chloro-5-methoxybenzene to yield the biaryl chloride **11**. The N-terminus **9** was Boc protected and borylated according to standard reactions and coupled with the biaryl chloride **11** using bis(triphenylphosphine)palladium chloride to yield the *pmp*-terphenyl monomer *pmp*-Tph-1 **12.**[41]



Scheme 4.4 Synthesis of *pmp*-Tph-1 **12** according to optimized procedures.

4.4. Synthesis and Characterization of Terphenyl Oligomers

Homo-oligomers in the *mop*, *pom* and *pmp* series with length *n* = 2, 4, and 8 were synthesized via standard solution-phase peptide coupling reactions (**Scheme 4.5**). The protected terphenyl amino acid monomer was first dissolved and separated into two equal portions. To one portion, HCl was added to remove the Boc group. To the other portion, LiOH was added to cleave the methyl ester. After workup, both the carboxylic acid and the amine-HCl salt were obtained and carried forward without further purification. These were mixed with Hünig's base and (1-cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate (COMU) for coupling overnight, followed by workup and purification to result in the dimer.[62] This procedure was repeated with the dimer or tetramer to provide the tetramer or octamer, respectively.



Scheme 4.5 General scheme for the oligomerization of terphenyl compounds.

We were prepared to face challenges regarding the solubility of these terphenyl oligomers. Kleman et al. noted that the solubility of *pop*-Tph-*n* in chloroform decreased dramatically as *n* increased.[41,43] While the *mop*-terphenyl series (*n* = 2, 4, 8) showed excellent solubility, solubility of *pom*-Tph-8 was poor and prevented us from collecting useful NMR or CD data. The *pmp*-terphenyl oligomers showed remarkably low solubility, even the dimer. No physical characterization was conducted with the *pmp* oligomers series.

4.4.1. Evidence of Folding and Self-Association

CD spectra were acquired in chloroform for tetramers and octamers in the *mop* and *pop* series (**Figure 4.8**). CD is a low resolution technique from the perspective of determining molecular conformation. Although CD cannot provide an atomic-level molecular structure, in contrast to crystallography, CD allows facilitates comparative studies that help us understand propensities of molecules to adopt secondary structure, determine what perturbs that structure, and evaluate effects of concentration.[44,45] In all *o*-

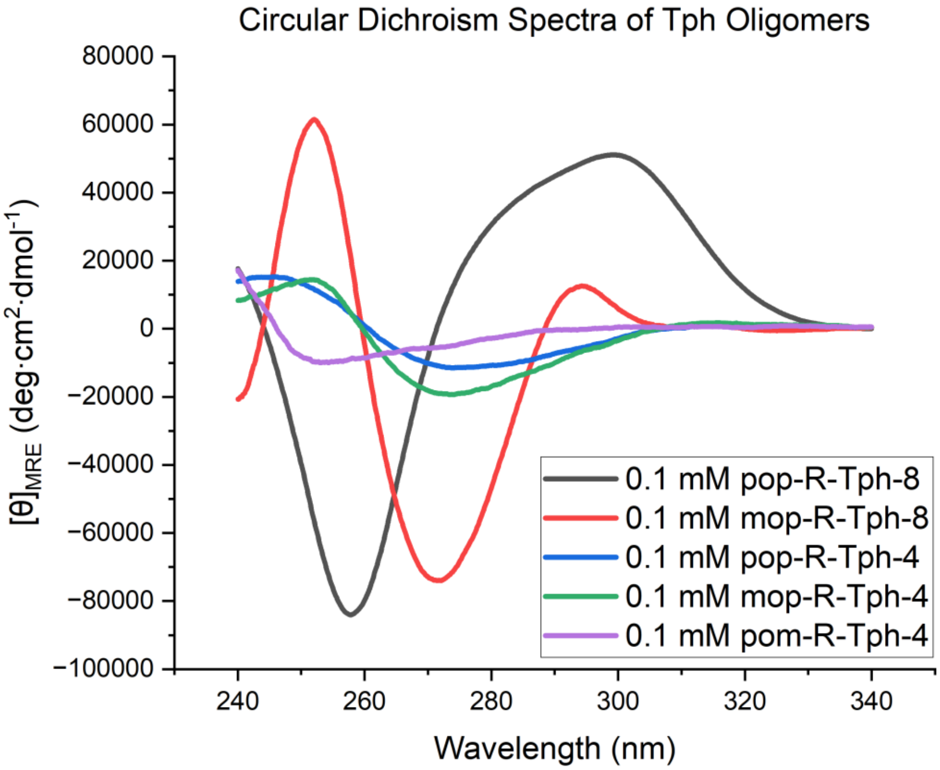


Figure 4.8 CD spectra in room temperature chloroform showing MRE of all synthesized tetramers and octamers. Terphenyl octamers exhibit more intense MRE values than the respective tetramers, demonstrating length-dependent cooperativity. *pom-R*-Tph-8 is not included in this series on account of its poor solubility.

terphenyl systems—*pop, mop,* and *pom*—the MRE signal becomes more intense as a function of oligomer length. Because MRE is normalized to length, the stronger signal at longer oligomer lengths suggests length-dependent folding cooperativity.[46] The *mop*-Tph-8 MRE displayed maxima and minima with similar intensities to those of the *pop*-Tph-8, although the CD signatures of these two octamers were quite different from one another. This observation encouraged us to focus on the *mop* series for further study.

Longer oligomers were sensitive to structural perturbation or dissociation by addition of methanol, a hydrogen bond donor, or increasing temperature. In variable temperature experiments (**Figure 4.9**), 0.1 mM *pop* and *mop* tetramers underwent a ~50%

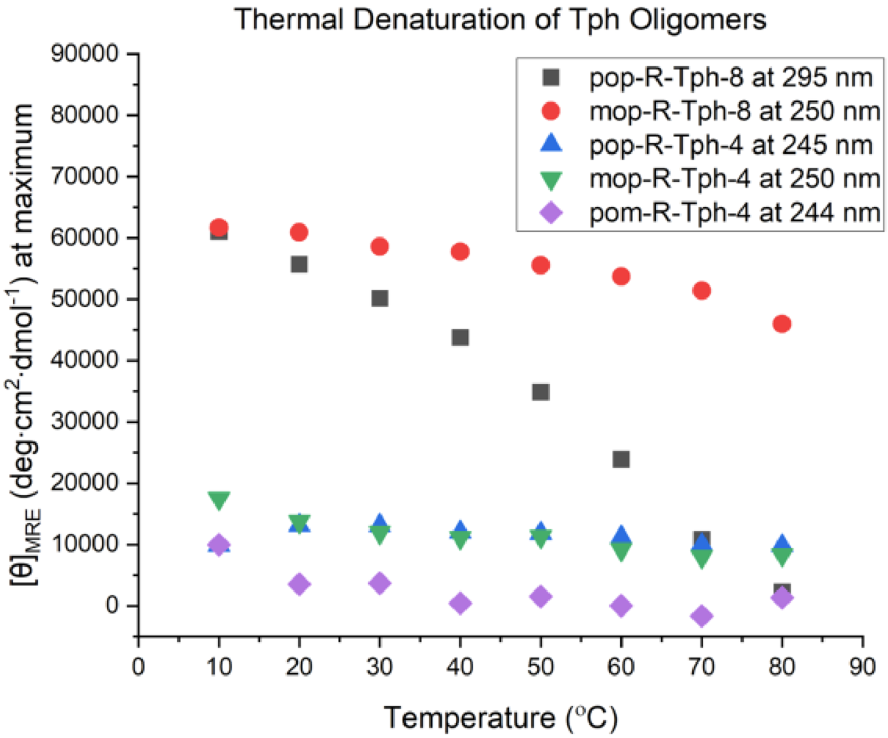


Figure 4.9 Variable temperature CD data for terphenyl tetramers and octamers (0.1 mM in chloroform). MRE of tetramers declines steadily from 10°C to 80°C. While *pop*-Tph-8 undergoes complete denaturation as the temperature approaches 80°C, *mop*-Tph-8 retains most of its ellipticity, indicating structural elements are likely still present.

decrease in MRE signal from 10°C to 80°C in chloroform; at 70°C the MRE of *pom*-Tph-4 reached zero at 244 nm. The octamers of *pop* and *mop* behaved differently across the 10°C to 80°C temperature range; while the MRE of *pop*-Tph-8 at 295 nm approached zero by 80°C, the *mop*-Tph-8 lost only ~20% of the original MRE.

Similar results were observed while varying methanol concentration with 0.1 mM terphenyl oligomers at room temperature (**Figure 4.10**). While *pop*-Tph-8 MRE reduced by >50% with only 10% methanol and no MRE was observed at 25% methanol, *mop*-Tph-8 suffered a <20% loss in signal even at 50% methanol. These results suggest that

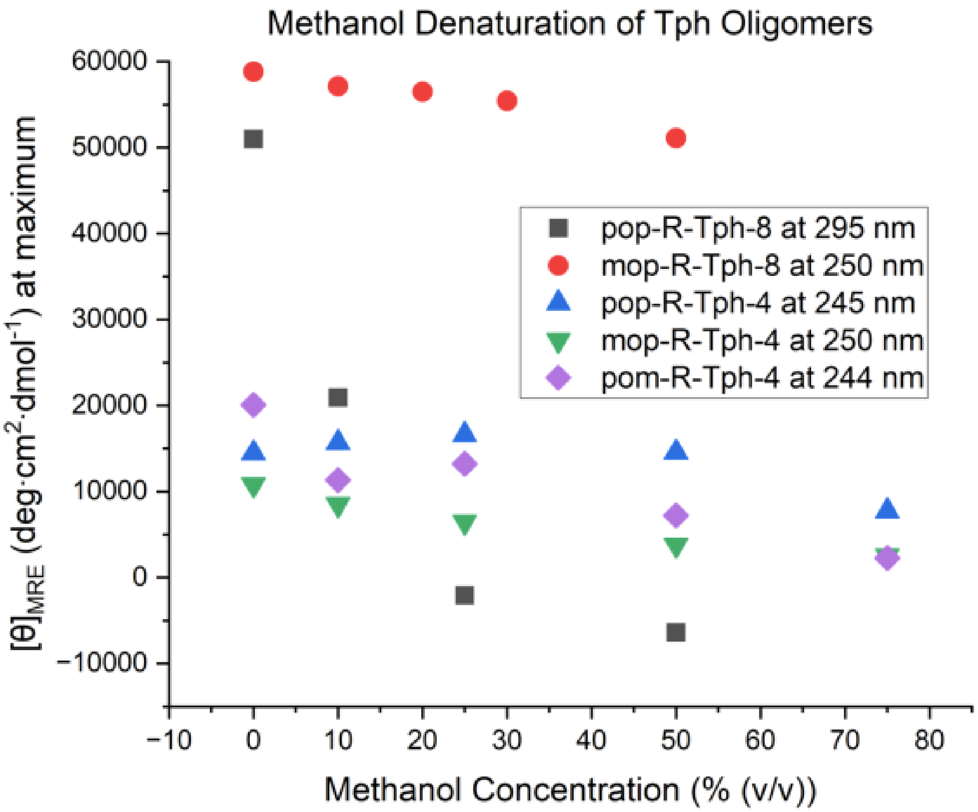


Figure 4.10 Methanol denaturation of 0.1 mM terphenyl oligomers dissolved in room temperature chloroform causes a loss in CD signal. While *pop*-Tph-8 undergoes the most drastic change, *mop*-Tph-8 retains most of its ellipticity even at 50% methanol.

the *mop*-terphenyl structure is more robust than the *pop*- counterpart, and possibly less dependent on hydrogen bonding.

Both CD and DOSY experiments seem to suggest that folding and self-association occur in concert among the *pop* terphenyl oligomers.[41] We wanted to determine whether folding and self-association were linked in the *mop*- series as well. When the concentration of *pop*-Tph-8 was increased in chloroform, the MRE intensified (**Figure 4.5**), suggesting self-association occurs at higher concentrations of *pop*-Tph-8. When the

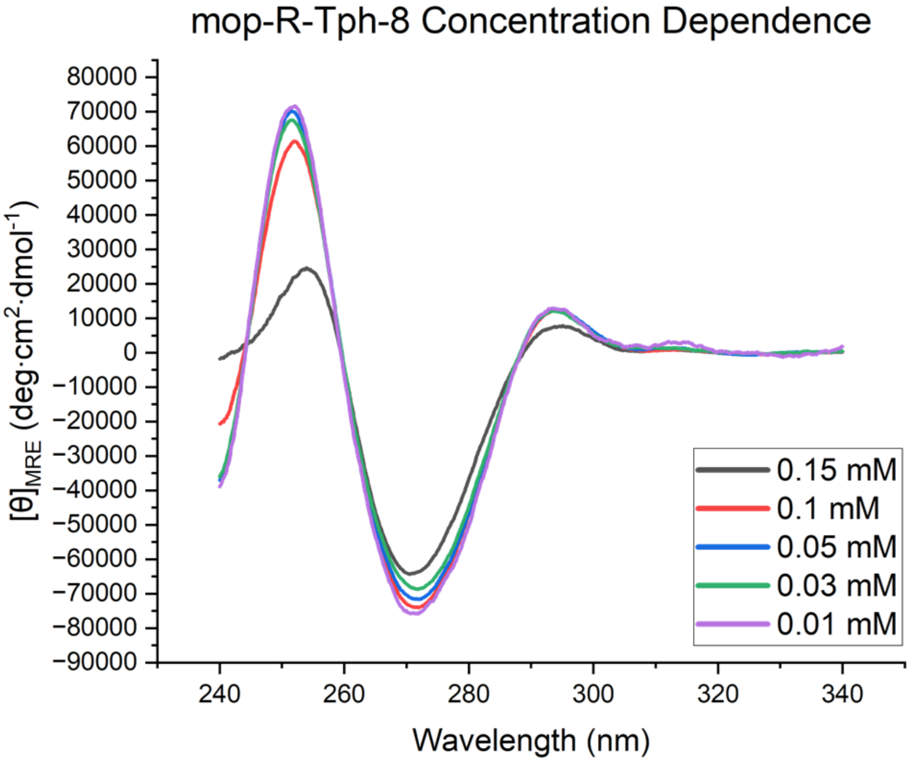


Figure 4.11 CD spectra of *mop-R*-Tph-8 in chloroform acquired at room temperature. Unlike the trend we observe with self-associating *pop*-Tph-8, at the *mop*-Tph-8 ellipticity minimum (270 nm) increasing oligomer concentrations is accompanied by a similar, or perhaps slightly less intense, MRE.

analogous experiment was conducted with *mop*-Tph-8, the intensity showed very little change, perhaps slightly decreasing from 0.01 mM to 0.15 mM in chloroform (**Figure 4.11**). These observations suggest that *mop*-Tph-8 is largely or entirely monomeric at these low concentrations. 600 MHz DOSY NMR using the Bruker dstebpgp3s pulse sequence, however, revealed that self-association of the *mop* series tetramer and octamer occurs above 1 mM (**Table 4.2.**). The diffusion coefficient remain3e constant from 0.1 to 50 mM for *mop*-Tph-2, consistent with a completely monomeric state throughout this range. Similar behavior was observed for the *pop* dimer. For *mop*-Tph-4, the diffusion coefficient became smaller as the concentration approached 50 mM. Similar

Table 4.2. Diffusion coefficients measured for *pop*-Tph-*n* in $CDCl_3$ by DOSY NMR by Kleman et al.[41] compared to diffusion coefficients of *mop*-Tph-*n*, both normalized to the experimental diffusion coefficient of tetramethylsilane.

	Relative Diffusion Coefficient					
	pop Series			*mop Series*		
[Oligomer] /mM	**pop-Tph-2**	**pop-Tph-4**	**pop-Tph-8**	**mop-Tph-2**	**mop-Tph-4**	**mop-Tph-8**
0.03			0.114			
0.1	0.252	0.1949		0.2907	0.1925	0.1387
0.3			0.1165			
1	0.2594	0.1811	0.1135	0.2914	0.1907	0.1372
5			0.1147			
10	0.2594	0.1759		0.2909	0.1575	0.0997
50	0.2504	0.1469		0.2926	0.1414	0.0696

behavior was observed for the *pop* tetramer. The Stokes-Einstein equation shows that the average volume of the *mop* tetramer species at 50 mM is approximately 2.5x the volume of the species at 0.1 mM, which is consistent with the conclusion that self-association of this tetramer occurs at 50 mM.[50] In contrast to the similarities between *mop* and *pop* dimers and tetramers, there was a clear difference between the octamers. The *pop*-Tph-8 diffusion coefficient was steady from 0.03 to 5 mM. Higher concentrations of *pop*-Tph-8 could not be measured due to solubility limitations. These data along with light scattering data were previously interpreted to indicate that the *pop* octamer self-associates even at 0.03 mM in chloroform. In contrast, we found that the diffusion coefficient for *mop*-Tph-8 diffusion coefficient is steady between 0.1 and 1.0 mM but then decreases at 10 mM and further decreases at 50 mM. We therefore conclude that the *mop* octamer is monomeric at or below 1.0 mM in chloroform, but that self-association begins between 1.0 and 10 mM.

Comparing the NMR spectra of the dimer, tetramer, and octamer at 1 mM in $CDCl_3$ (a concentration at which self-association is minimal), provides further evidence of folding.

The aromatic signal envelope of the octamer shifts upfield to ~6 ppm, while the tetramer and dimer aromatic signals only reach 6.5 and 7 ppm, respectively (**Figure 4.12**). The upfieshift is indicative of higher shielding, which is what one would expect from a more compact structure. The acquisition of proton NMR spectra in DMSO significantly separates and sharpens the signals, which may indicate hydrogen bonding contributes to the folding (**Figure 4.13**).

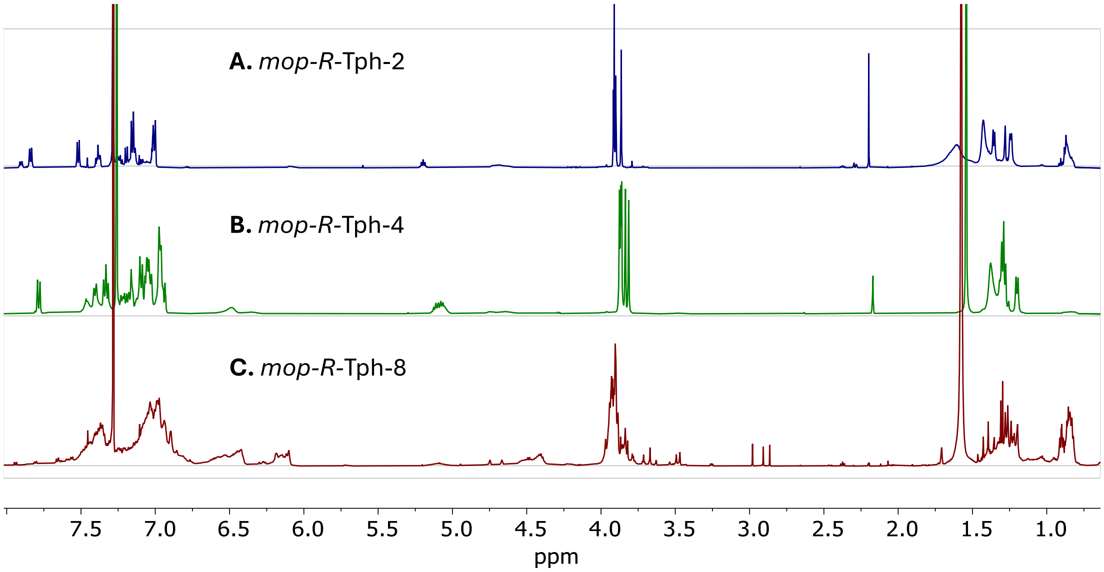


Figure 4.12 500 MHz ^{1}H NMR of *mop*-Tph A) dimer, B) tetramer, and C) octamer, each 1 mM in $CDCl_3$.

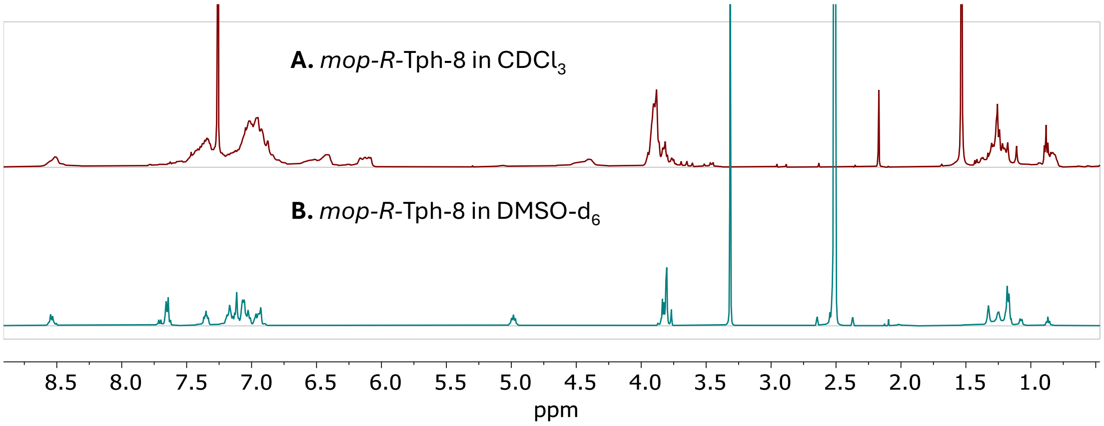


Figure 4.13 500 MHz ^{1}H NMR of *mop*-Tph-8 in A) $CDCl_3$ and B) DMSO-d_6.

4.5. Attempted Crystallization of Terphenyl Oligomers

As noted previously, terphenyl compounds have an extreme aversion to crystallization.[51] Almost all crystallization attempts conducted both here and in the previous work by Kleman et al. resulted in formation of an amorphous glass.[43]

While amorphous glass and microcrystals can be sometimes hard to distinguish with the naked eye, polarization microscopy can easily differentiate between these two types of material. A crystal that is illuminated in a cross-polarized field extinguishes polarized light when rotated. Particles of a glass will appear dark or grey at all times under these same conditions. Polarization microscopy was heavily relied upon here to assess crystallization attempts.

4.5.1. Preliminary Crystallization Attempts

Five main techniques for crystal growth were tested.

1. Slow evaporation of solvent
2. Vapor diffusion (a solution of the molecule is placed into a larger container containing the vapor of a solvent in which the solute is insoluble)
3. Layer diffusion (a solution of the molecule is ‘layered’ with another solvent in an NMR tube)
4. Slow cooling of a concentrated solution prepared by heating
5. Sublimation (the sample is heated under vacuum, and volatile material is allowed to condense on a cold finger)

Different crystallization attempts were conducted with various terphenyl oligomers and various solvents depending mostly on availability. In general, oligomers were dissolved that seemed to provide high solubility, such as toluene, ethyl acetate, methylene chloride, and chloroform. In crystallization systems with two solvents, pentane, hexanes, or methanol was used to try to initiate crystallization from a concentrated solution in a good solvent.

Slow evaporation trials in chloroform, methylene chloride, or ethyl acetate failed to produce crystalline material and often resulted in a glass. However, *mop*-Tph-1 crystals grew upon slow evaporation of a chloroform solution submitted for mass spectrometry analysis. The structure was determined via X-ray diffraction. Two molecules in slightly different conformations occur in the unit cell (**Figures 4.14-4.15**). In each conformation, the dihedral angles between the central aryl ring and the C-terminal phenylene

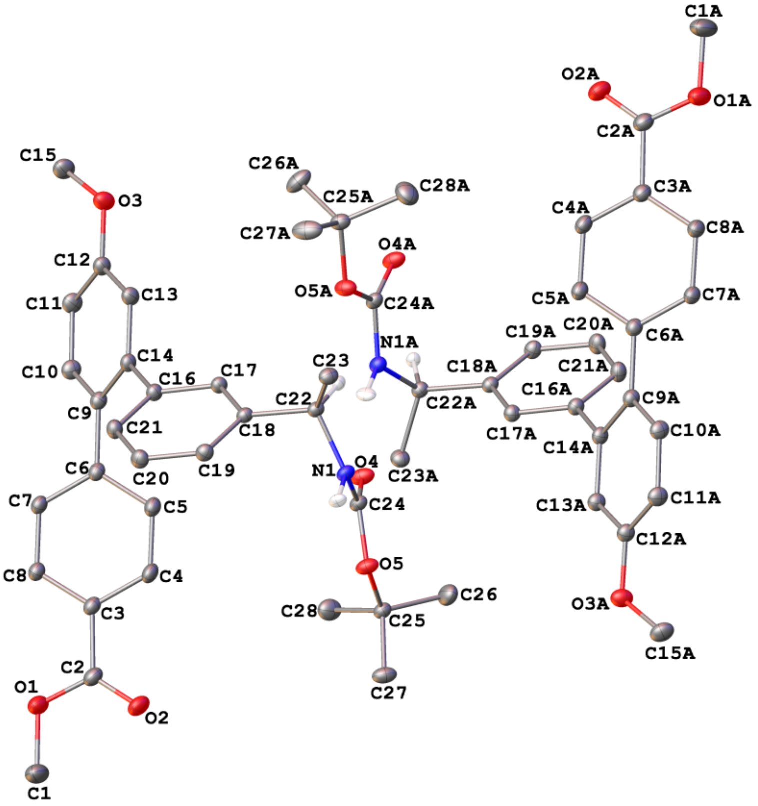


Figure 4.14 Unit cell of *mop*-Tph-1 containing two symmetry independent molecules with 50% probability ellipsoids. Resolution = 0.79 Å. Dihedral angles: C7-C6-C9-C10=46.2°; C9-C14-C16-C21=55.1°; C7A-C6A-C9A-C10A=49.6°; C9A-C14A-C16A-C21A=-56.6°

substituents are 46.2° and 49.6°. The dihedral angles between the central aryl and the N-terminal phenylene are 55.1° and -56.6°. Thus, similar angles are observed in both symmetry independent structures of *mop*-Tph-1 (**Figure 4.15**).

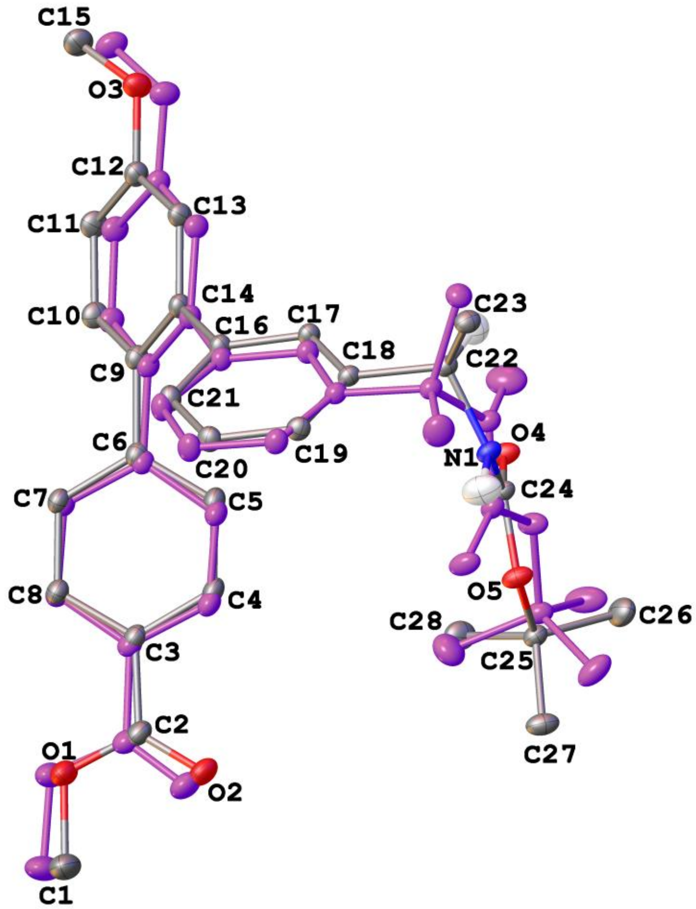
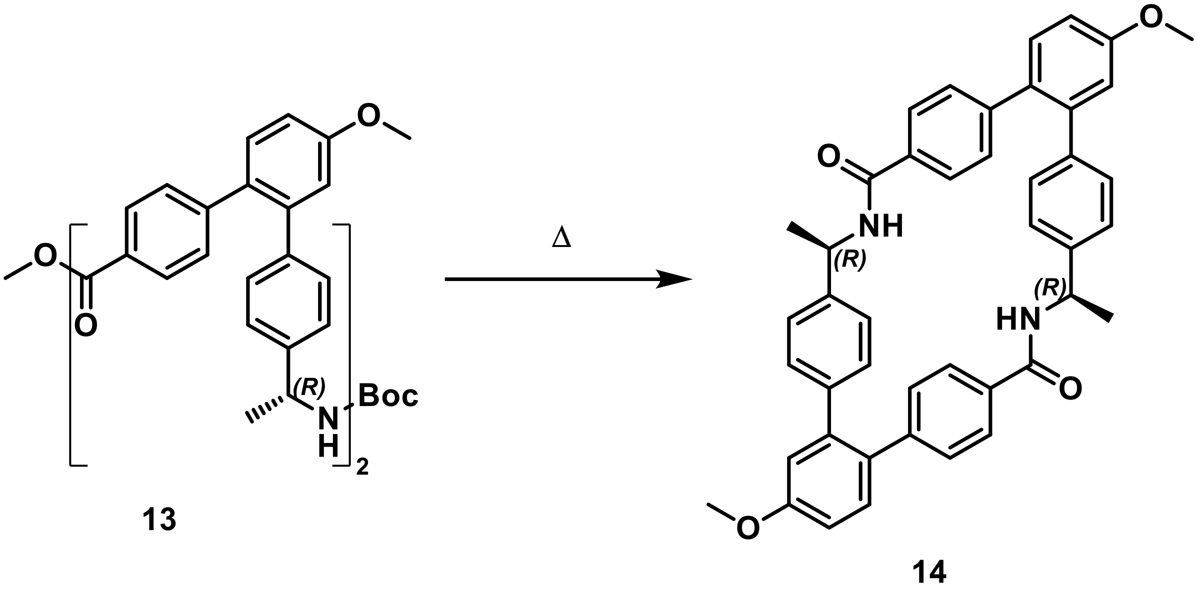


Figure 4.15 Inverted and superimposed symmetry independent conformations of *mop*-Tph-1 within the unit cell. 50% probability ellipsoids. Resolution = 0.79 Å.

Vapor diffusion trials were conducted on *mop*-Tph-*n* (*n*=2, 4, and 8) by placing either an HPLC vial with a conical bottom or a cut NMR tube (both with pierced caps) containing 0.1 mg/µL compound dissolved in chloroform, DCM, or ethyl acetate into 20 mL glass scintillation vials containing the secondary solvent. Care was taken to ensure that the inner vessel did not touch the sides of the scintillation vial! With the NMR tubes, this goal was accomplished most frequently by adding a second NMR cap to the bottom of the tube so that the tube containing the sample stood up freely. Vapor diffusion usually resulted in an amorphous powder. Liquid diffusion was conducted by carefully layering the secondary solvent onto (or below, depending on density) the compound solution in an NMR tube. Over the course of days or weeks, amorphous solid appeared at the interface.

Slow cooling trials failed to produce any solid-phase product of *mop*-Tph-*n*, crystal or glass, in single solvents. It is likely that the initial solubility of the oligomers was was too high. In binary solvent systems, the oligomer solubility was extremely temperamental, and solid precipitated either as soon as the second solvent was added or when the solution was removed from the heat source.

Sublimation attempts were carried out using only *pop*-Tph-2 that had been prepared by Kleman.[43] Sublimation of this material under high vacuum resulted in crystals of a macrocyclic derivative (**Scheme 4.6**, **Figure 4.16**). If macrocyclization of this molecule is possible, it seems possible that the linear precursor has a low-energy conformation in which the C- and N- termini are near one another. Dihedral angles between rings in this macrocyclic structure are similar to those of the *mop*-Tph-1 structures in **Figure 4.14**.



Scheme 4.6 Conversion of *pop*-*R*-Tph-2 to the cyclic dimer proceeds under vacuum.

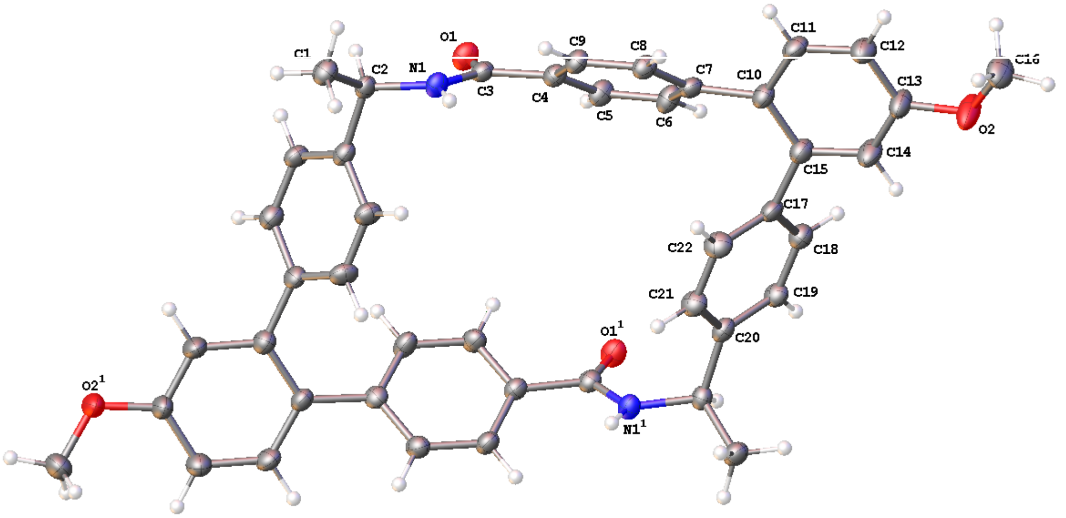


Figure 4.16 Crystal structure of *pop*-Tph-2 macrocycle from sublimation. Symmetry-dependent portion labelled. 50% probability ellipsoids. Resolution = 0.80 Å. Dihedral angles: C18-C17-C15-C14=-52.1°; C11-C10-C7-C8=-44.4°

4.5.2. Racemic Crystallization Attempts

Racemic crystallography is a strategy that has been employed to collect structural data for compounds that are difficult to crystallize as single enantiomers.[63–65] Racemates are sometimes more prone to crystallization because only 65 of the 230 possible crystallographic space groups are accessible to chiral compounds.[64] The remaining space groups have a center of symmetry accessible to either achiral molecules or pairs of enantiomers. Enantiomers can co-crystallize forming an inversion center that allows access to the other 165 space groups. Since the *mop*-terphenyl series oligomers did not crystallize as single enantiomers under the conditions tested, we synthesized the enantiomer of each *mop*-Tph-*n* compound and attempt crystallization of racemic mixtures.

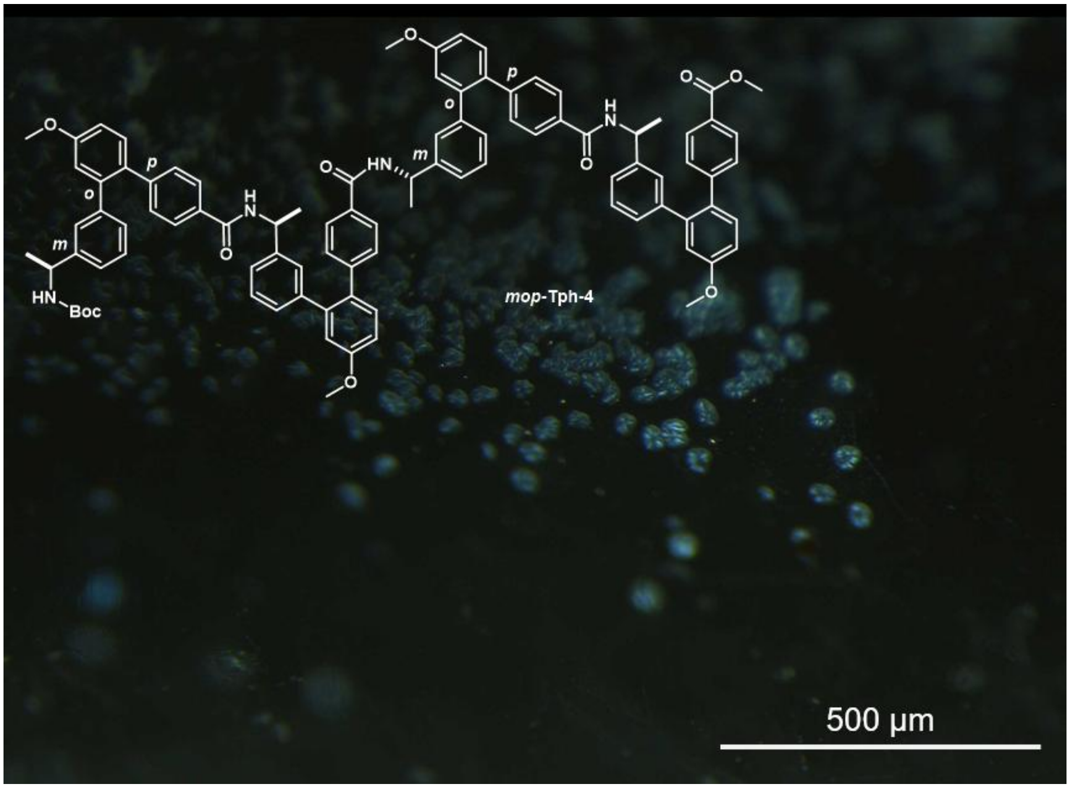


Figure 4.17 Microcrystals of *mop*-Tph-4 from vapor diffusion of pentane into ethyl acetate over 1.5 weeks.

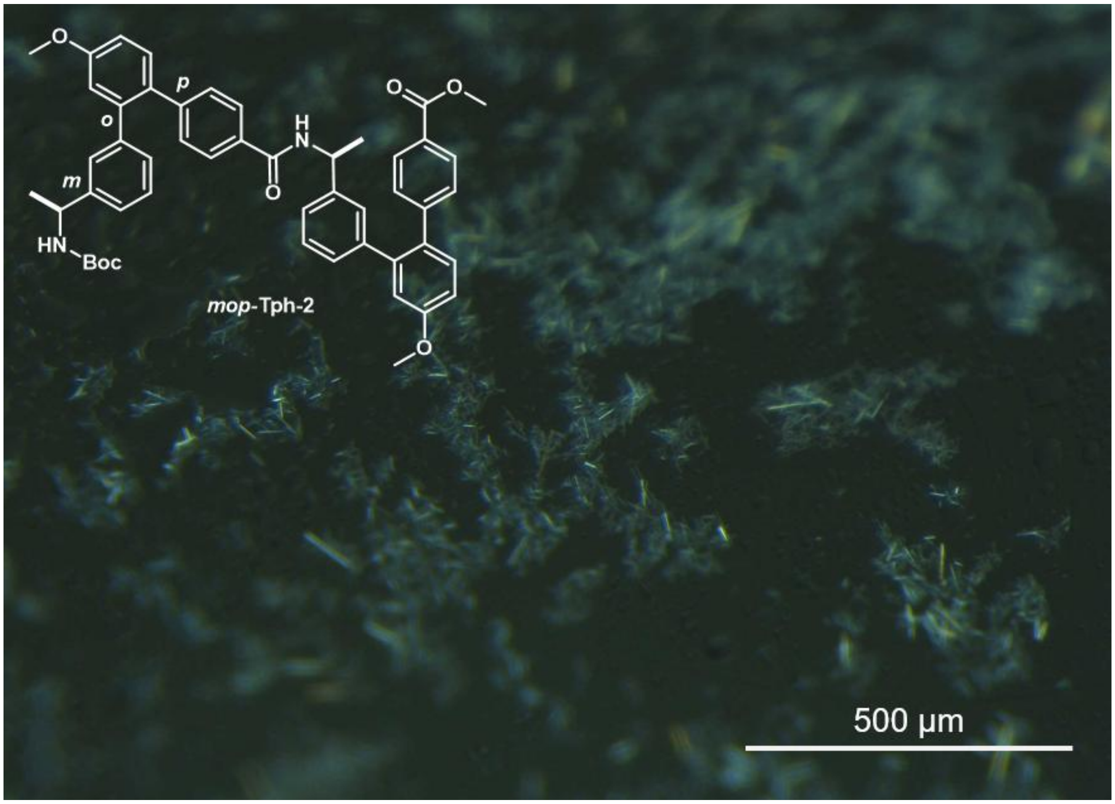


Figure 4.18 Racemic microcrystals of *mop*-Tph-2 from room temperature slow evaporation of ethyl acetate over the course of 1 week.

Although many of these attempts failed, microcrystals were observed in vapor diffusion trials of either toluene or pentane into racemic *mop*-Tph-2 or *mop*-Tph-4 (**Figure 4.17.**) dissolved in ethyl acetate for 1.5 weeks. Microcrystals of racemic *pop*-Tph-2 and *mop*-Tph-2 (**Figure 4.18**) were also observed via slow evaporation of diethyl ether or ethyl acetate, respectively. These microcrystals were unfortunately too small for x-ray crystallography, and attempts to grow larger crystals failed.

4.5.3. Microcrystal Electron Diffraction

We wondered whether the microcrystals from racemic crystallization trials might be suitable for analysis via a new cryogenic electron microscopy (cryo-EM) technique, microcrystal electron diffraction (microED).[66]MicroED was invented by the Gonen Lab in 2013 to determine the structure of molecules when crystals are not suitable for x-ray diffraction.[66–68] During the data acquisition, a microcrystalline sample is rotated in a beam of electrons. Diffraction patterns are acquired for 90° of sample rotation, and the structure can be solved using standard crystallography software. MicroED was attempted by Gonen, T., Unge, J., and Lin, J. at the University of California, Los Angeles on the racemic *mop* dimer and tetramer. Although many of the racemic tetramer microcrystals exhibited poor diffraction, and the best converged to only 1.88 Å resolution with 90.5% completeness, the racemic dimer microcrystals were solved with 0.8 Å resolution and 95.9% completeness (**Figures 4.19 – 4.20**).

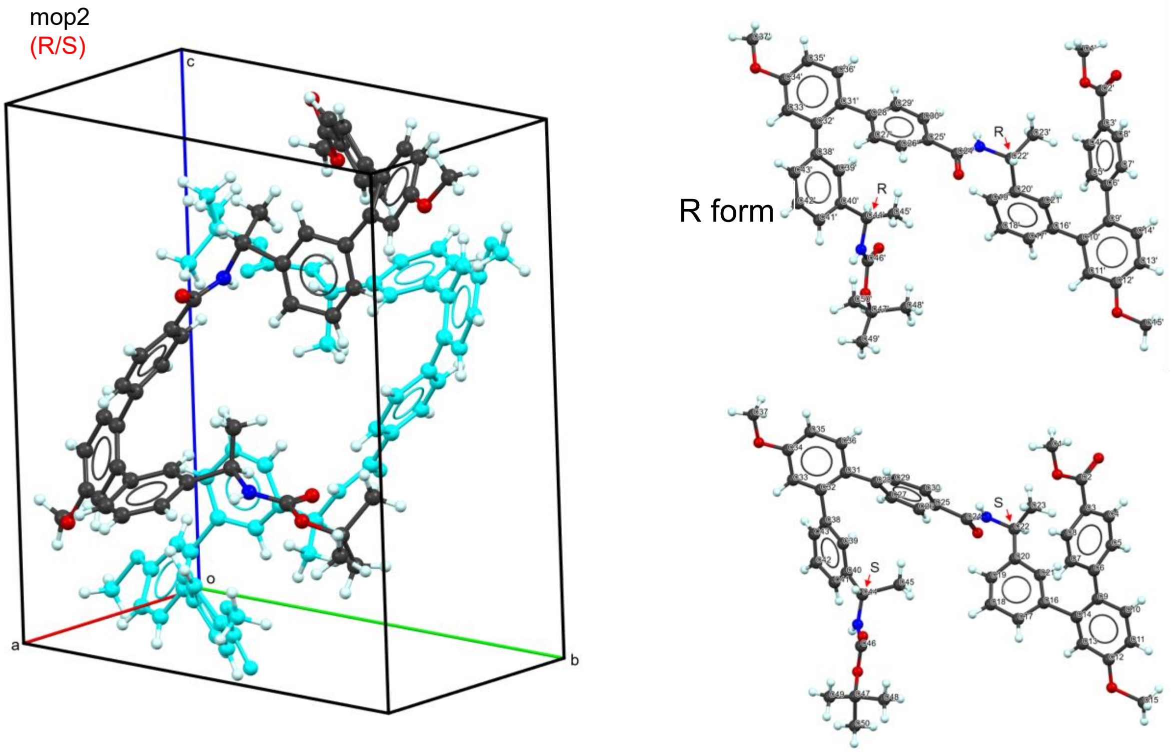


Figure 4.19 Crystal structures of *mop*-Tph-2 solved via microED with 0.8Å resolution. Both R- and S- forms are visible in the unit cell.

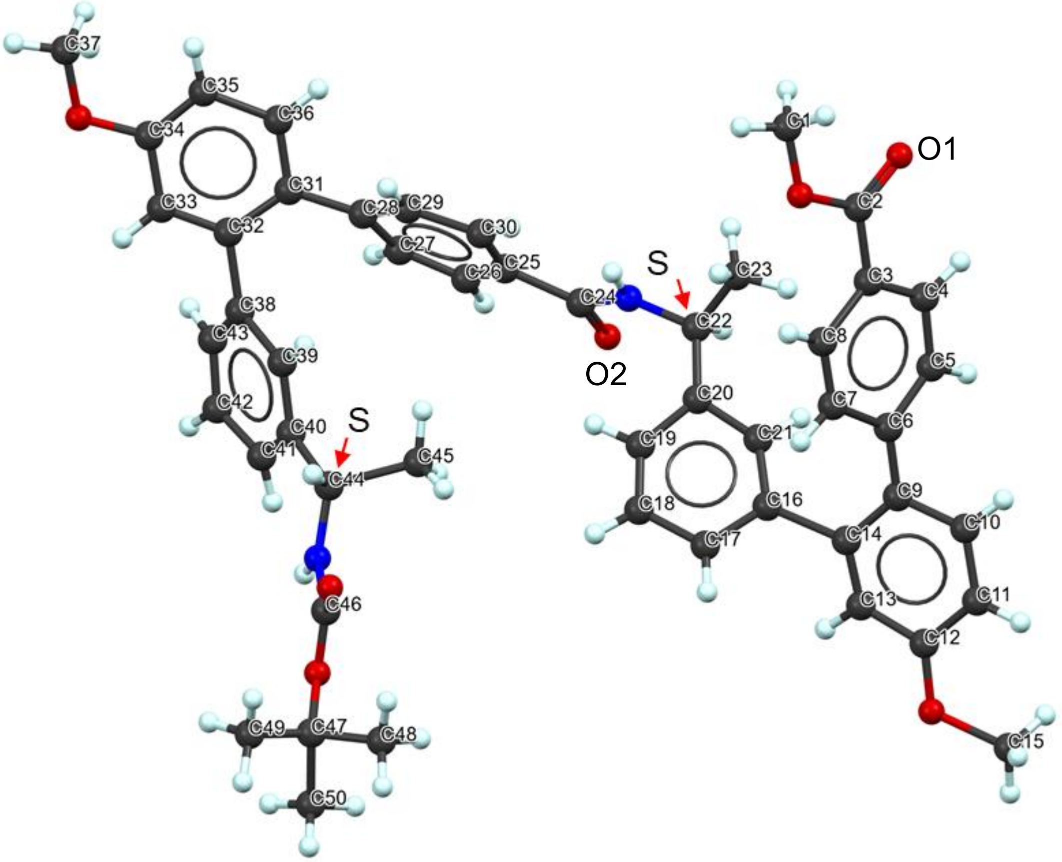


Figure 4.20 Crystal structure of symmetry-dependent *S,S* *mop*-Tph-2 extracted from the racemic crystal structure determined by microED. Torsional angles are measured as follows: O1--C2-C3-C4=3.1°; C5-C6-C9-C10=59.3°; C9-C14-C16-C21=47.4°, C21-C20-C22-C23=-59.1°; C20-C22-N-C24=143.6°; O2-C24-C25-C26=0.9°; C27-C28-C31-C32=44.8°; C31-C32-C38-C39=48.7°; C39-C40-C44-C45=-70°; C45-C44-N-C46=79°

The residues in the *mop*-Tph-2 structure show key differences in dihedral angles across analogous bonds. The C-terminal dihedral from C5-C6-C9-C10=59.3°, while the analogous bond on the other residue (C27-C28-C31-C32) is only 44.8°; the angle between the chiral methyl group and the proximal aryl groups (C21-C20-C22-C23; C39-C40-C44-C45) are 59.1° and -70°. The largest difference between these residues is the dihedral between the N-terminal aryl group and either the amide carbon (C20-C22-N-C24) or the carbamate (C45-C44-N-C46), which are 143.6° and 79° respectively. There are similarities between other analogous dihedrals angles: those between the central aryl group and the N-terminal phenylene groups (C9-C14-C16-C21 and C31-C32-C38-C39) are approximately 48° in both cases. Both the C-terminal ester and the amide are planar with the C-terminal phenylene.

Some dihedral angles between the central aryl group and the phenylene groups in the dimer are comparable to those in the monomer. The angles between the C-terminal phenylene and the central aryl group in the dimer structure are 59.3° and 44.8°; these angles are measured to be 46.2° and 49.6° in the monomer. The angles between the N-terminal phenylene and the central aryl group in the dimer are both 48°, while in the monomer they are closer to 55°.

The differences in these dihedral angles demonstrate that there is some amount of flexibility between the central aryl group and the phenylene substituents. These angles are prone to change as the polymer is elongated.

4.6. Conclusion

Drawing inspiration from initial studies by Kleman et al. that report folding and self-associating oligomers of *pop*-Tph-*n*,[41,43] we prepared several regioisomers—*mop-, pom-,* and *pmp*-Tph-*n*—hoping that the regiochemical modifications would enable some of these oligomers to adopt conformations more conducive to crystallization. The *mop*-Tph-*n* series seemed most interesting on account of its robust solubility, strong CD MRE, thermal stability, tolerance toward hydrogen bond donors, and resistance to aggregation at concentrations <1 mM. Although the *mop* oligomers were largely resistant to crystallization efforts, microcrystals of the *mop*-Tph-2 racemate allowed determination of a high-resolution structure based on data from microED. While the crystallographic structures of *mop*-Tph-1 and *mop*-Tph-2 do not tell us how higher order oligomers fold, we found that there is some flexibility between the aryl groups in the *o*-terphenyl structure. Most importantly, the introduction of a *meta*- linkage in our *o*-terphenyl structure allowed the *mop*-Tph-8 to fold without self-association at low concentrations.

4.7. Experimental Methods and Spectra

Starting materials and reagents were purchased from Ambeed, Chem-Impex International, or Sigma Aldrich.

Mass spectra obtained on a Thermo Q Exactive Plus instrument, NIH award 1S 10 OD020022-1. NMR obtained on a 500 MHz Bruker Avance III, a gift of Paul and Margaret Bender. DOSY NMR acquired on a 600 MHz Bruker Avance III instrument, NIH S10 OD012245. CD spectra obtained on a Jasco J1500, NIH 3R01 GM061238-21S1.

Crystal images acquired on a Leica M205 stereomicroscope.

X-ray crystallography performed on a Bruker D8 VENTURE Photon III Cu IμS 3.0 X-ray diffractometer that was partially funded by NSF Award CHE-1919350.

DOSY-NMR Information

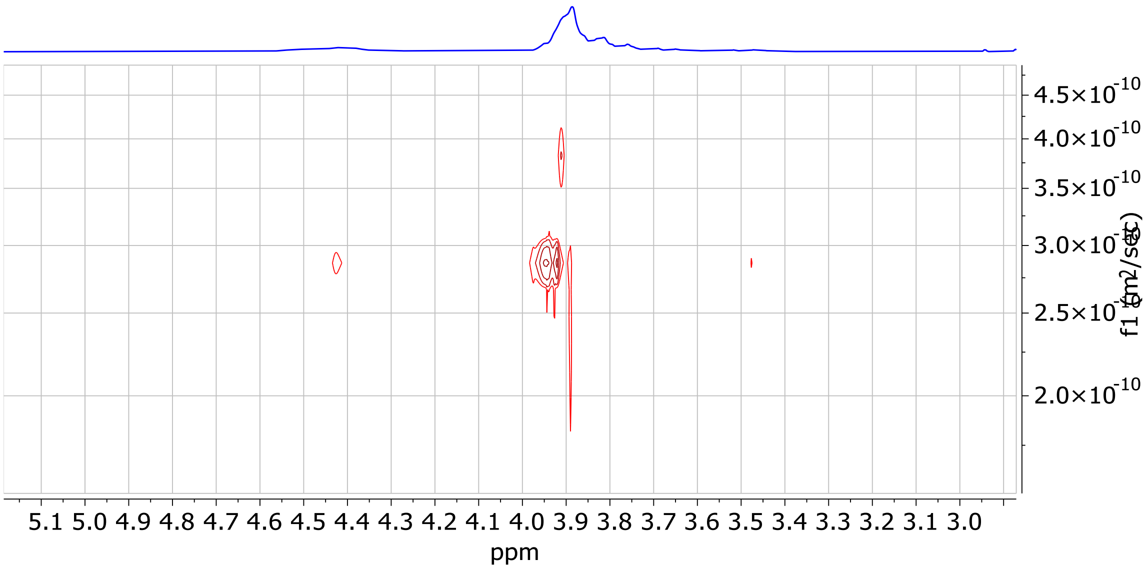
For DOSY experiments, 50 mM - 0.1 mM oligomer samples were prepared in ~500 µL $CDCl_3$ (with TMS) and were added to an NMR tube. The 600 MHz magnet with a TCI-F cryoprobe was locked and shimmed, and a proton NMR spectrum was first acquired. **The sample must remain in the magnet for ~10 minutes prior to the first DOSY experiment for temperature equilibration.**

A 1D DOSY (dstebpgp3s1d) was conducted for optimization. The signal of a 1D DOSY at 95% gradient strength should be ~1/20 of the signal at 5% strength. If not, adjust d20 (diffusion delay). Generally, diffusion delays were 0.07 s for the dimer, 0.12 s for the tetramer, and 0.15 s for the octamer. 2D DOSY (dstebpgp3s) was then acquired for 32 points from 5-95% gradient.

In a well-acquired DOSY, peaks of the same species should line up. Signals that are out-of-line are either other molecules, portions of the molecule of interest that are more dynamic (such as terminal groups) or artifacts. One of the methoxy- peaks (~4 ppm) was selected to compare with TMS.

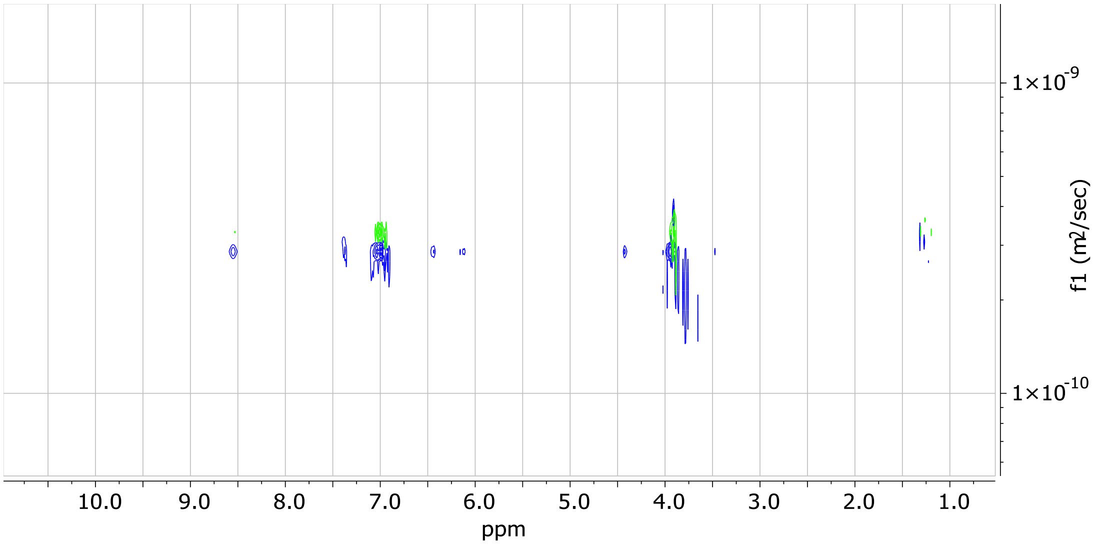
DOSY of *mop*-Tph-8. Green line shows how associated signals line up.

Because signals line up, it is usually not important which signal is selected for DOSY analysis as long as it is consistent with the choices made with spectra of other concentrations. Below is a closer look at the methoxy region. The signal with the higher diffusion coefficient is the terminal methyl ester that is more mobile than the aromatic methoxy groups. Several real signals from various methoxy groups line up at 2.9e-10. The narrow lines underneath are artifacts that do not correspond to real signals. It is therefore relatively easy to select a consistent signal.



DOSY of *mop*-Tph-8 methoxy- region.

For samples that are aggregating, increasing concentration will lower the diffusion coefficient. Below is an overlay of 1 mM and 10 mM mop-Tph-8. The 10 mM oligomer (blue) aggregates slightly and has a smaller diffusion coefficient relative to 1 mM (neon green).



DOSY overlay of *mop*-Tph-8 at 1 mM (neon green) and 10 mM (blue).

General Procedure 1: Dimerization of Terphenyl Amino Acids

Procedure adapted from Kleman et al.[41]

A portion of Tph-*n* was dissolved in minimal 1,4-dioxane and divided into two equal portions volumetrically.

Part A: N-terminal Boc deprotection was conducted on one portion by adding 4M HCl in 1,4-dioxane (5 mL per mmol terphenyl). Heated to 50°C with stirring. Bubbling indicated deprotection was progressing. After 2 hr, solution became cloudy. The volume was reduced under a gentle stream of nitrogen. The solution was concentrated *in vacuo* and coevaporated with ethyl acetate to yield the desired compound as an HCl salt.

Removal of the Boc group was confirmed by ^{1}H NMR and the compound was carried forward without purification.

Part B: C-terminal methyl ester deprotection was conducted on the second portion of terphenyl. Lithium hydroxide (10 eq) was dissolved in water (5 mL/mmol) and added to an equal volume of the Tph-*n* in 1,4-dioxane. The solution was stirred at 70°C. Disappearance of starting material was monitored by TLC.

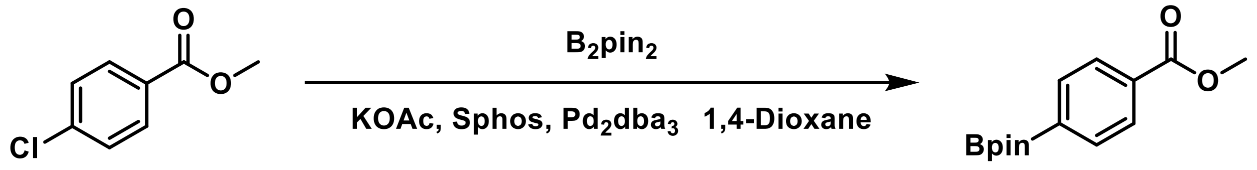
When complete conversion was achieved (usually several hours), most of the solvent was evaporated *in vacuo* and precipitated with 1 M HCl (aq) until neutral. The precipitate was dissolved in chloroform. The organic layer was extracted and washed with water before being dried with sodium sulfate and evaporated.

Removal of the methyl ester was confirmed by ^{1}H NMR and the compound was carried forward without purification.

Part C: The deprotected amino acids from parts **A** and **B** were dissolved in DMF (8 mL per mmol amino acid) and combined. 3 eq Hünig's base was added and the solution was cooled in an ice bath with stirring. COMU (1.2 eq) was finally added and the solution was stirred overnight at room temperature.

The Tph-2*n* was precipitated with water and centrifuged. The precipiate was dissolved in ethyl acetate and was washed in a separatory funnel with water, 0.5 M HCl, sodium bicarbonate, and twice with concentrated LiCl (aq) to remove DMF. The organic layer was dried and concentrated *in vacuo*.

1b - Methyl 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzoate



1a **1b**

Into a dry flask was added $Pd_2(dba)_3$ (1.37 g, 1.5 mmol, 0.02 eq), Sphos (1.23 g, 3 mmol, 0.04 eq) and a stir bar. Contents purged with nitrogen, and 75 mL anhydrous 1,4-dioxane were added. Stirred for 10 min. Potassium acetate (22.11 g, 225 mmol, 3 eq) and methyl 4-chlorobenzoate **1a** (12.79 g, 75 mmol, 1 eq) were added with stirring. B_2pin_2 (21.03 g, 82.5 mmol, 1.1 eq) was added with 75 mL more dioxane. Refluxed under nitrogen.

After 9 hours, TLC showed complete conversion (5% ethyl acetate/pentane, $R_f \approx 0.45$) and the reaction was allowed to cool to room temperature. Mixture was filtered through celite and washed with chloroform.

Isolated with column chromatography (3-4% gradient ethyl acetate in pentane) to yield 17.45 g (89% yield) **1b** as a white solid.

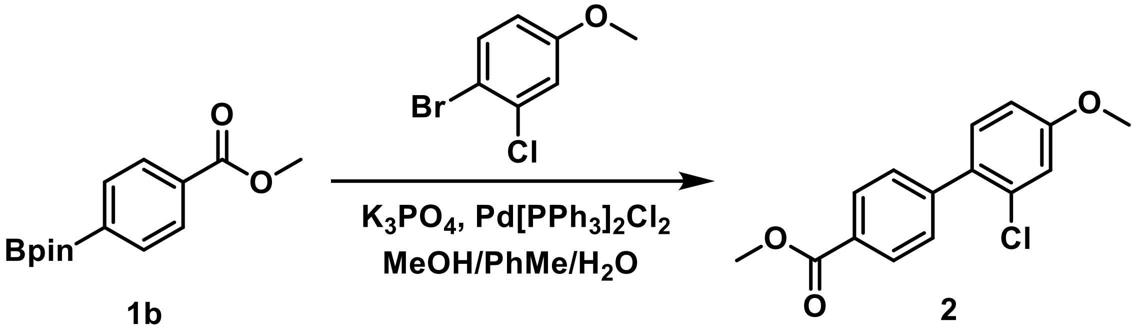
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.04 – 7.99 (m, 2H), 7.89 – 7.84 (m, 2H), 3.92 (s, 3H), 1.36 (s, 12H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 167.16, 134.68, 132.33, 128.61, 84.19, 52.15, 24.90.

HRMS (ASAP): $[M+H]^+$ 263.1449, found 263.1447 (0.8 ppm)

Procedure adapted from Kleman et al. with characterization that matched what was previously reported.[41]

2 - methyl 2'-chloro-4'-methoxy-[1,1'-biphenyl]-4-carboxylate



Boronic ester **1b** (10.01 g, 38.2 mmol, 1 eq) was dissolved in 500 mL 85:15 toluene/methanol with a stir bar and sparged with nitrogen. 4-bromo-2-chloroanisole (10.99 g, 49.6 mmol, 1.3 eq) was then added. Potassium phosphate tribasic (24.32 g, 115 mmol, 3 eq) was dissolved in water to make a 2 M solution, which was then added to the flask. While continuing to sparge, bis(triphenylphosphine)palladium chloride (1.87 g, 2.7 mmol, 7 mol%) were added. Mixture was refluxed under nitrogen for 5 hours, at which point TLC (5% ethyl acetate/pentane, $R_f \approx 0.3$) indicated complete conversion.

After cooling to room temperature, the mixture was filtered through celite and added to a separatory funnel. Water was drained out, and the organic layer was rinsed with equal volumes of water, sodium bicarbonate (sat.) and brine (sat.) and dried with sodium sulfate. The solution was concentrated under vacuum to yield a dark brown solid that was recrystallized in hot methanol to yield white crystals (8.76 g, 83% yield).

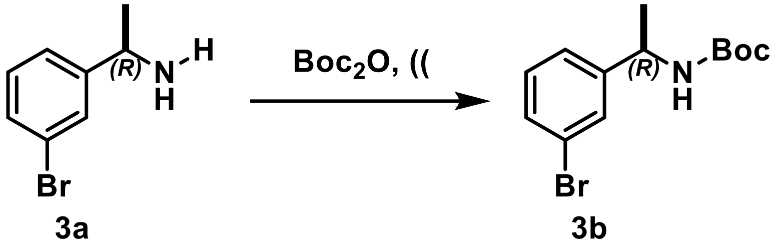
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.11 (d, *J* = 8.3 Hz, 1H), 7.52 (d, *J* = 8.3 Hz, 1H), 7.40 – 7.24 (m, 1H), 7.06 (d, *J* = 2.6 Hz, 1H), 6.91 (dd, *J* = 8.5, 2.6 Hz, 1H), 3.96 (s, 2H), 3.87 (s, 2H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.96, 159.83, 143.81, 132.91, 131.91, 131.76, 129.67, 129.32, 128.95, 115.33, 113.22, 55.63, 52.13.

HRMS (ASAP): $[M+H]^+$ Calculated 277.0626; found 277.0627 (0.4 ppm)

Characterization data matches that which was published by Kleman et al.[41]

3b - tert-butyl (*R*)-(1-(3-chlorophenyl)ethyl)carbamate



(*R*)-1-(3-bromophenyl)ethan-1-amine (**3a**) (2.00 g, 10.0 mmol, 1 eq) was measured in a vial and was added dropwise to fused di-tert-butyl dicarbonate (2.40 g, 11.0 mmol, 1.1 eq) with stirring. After the bubbling ceased and the reaction returned to room temperature, the remaining **3a** was transferred with 5 mL ethyl ether and the mixture was sonicated for 30 minutes. The oil was dried under vacuum and co-evaporated with 10 mL water to provide the title compound in quantitative yield (3.02 g), which was used for the next step without purification.

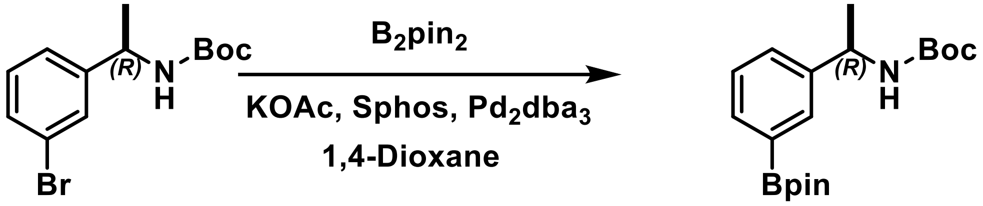
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.46 (d, *J* = 1.9 Hz, 1H), 7.39 (dt, *J* = 7.6, 1.7 Hz, 1H), 7.26 – 7.18 (m, 2H), 4.79 (d, 2H), 1.44 (m, 12H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 154.99, 146.58, 130.18, 130.16, 128.97, 124.60, 122.66, 79.75, 49.83, 28.36, 22.68.

HRMS (ESI): $[M+H]^+$ Calculated 300.0594, found 300.0587 (2.3 ppm)

Optical rotation: $[\alpha]^{22}_D$ = +46.9 (*c* 1.0, $CHCl_3$)

4 - tert-butyl (R)-(1-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)ethyl)carbamate



3b

4

Into a dry flask was weighed Sphos (415 mg, 1.01 mmol, 1.01 eq) and Pd$_2$(dba)$_3$ (460 mg, 0.501 mmol, 5 mol%) with a stir bar. The flask was purged with nitrogen before adding 100 mL of dry dioxane. Nitrogen was bubbled through the dioxane during the addition of **3b** (3.02 g), potassium acetate (2.96 g, 30.2 mmol, 3 eq), and B$_2$pin$_2$ (2.81 g, 11.1 mmol, 1.1 eq). The reaction was set up to reflux overnight.

The solution was cooled to room temperature and 500 mg *N*-acetylcysteine was mixed in before filtering through celite. The dioxane was removed with a rotovap and the mixture was redissolved in ethyl acetate. The mixture was washed with water followed by sodium bicarbonate and brine. The organic layer was dried with sodium sulfate and concentrated *in vacuo*. The yellow oil was purified using silica column chromatography (5-10% EtOAc/hexane, R$_f$ ≈ 0.2) to yield 1.71 g **4** (77% yield) as a pale yellow oil.

^{1}**H NMR** (500 MHz, CDCl$_3$) δ 7.73 (s, 1H), 7.70 (d, *J* = 7.2 Hz, 1H), 7.39 (d, *J* = 7.8 Hz, 1H), 7.33 (t, *J* = 7.5 Hz, 1H), 4.80 (s, 2H), 1.45 (d, *J* = 6.5 Hz, 3H), 1.42 (s, 9H), 1.34 (s, 112H).

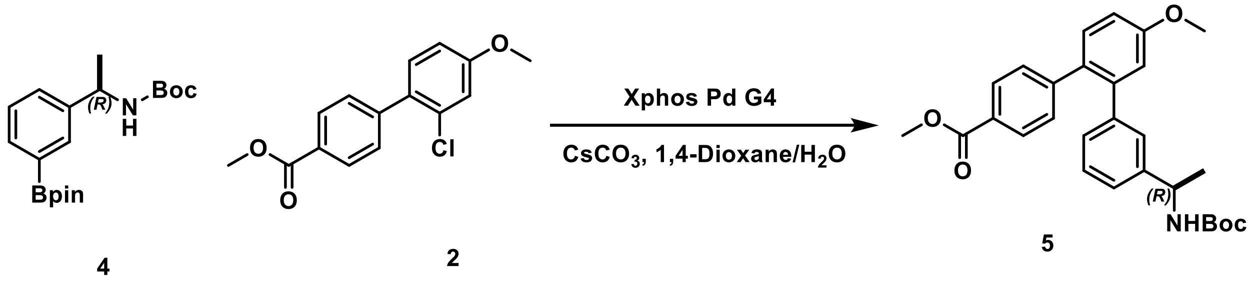
^{13}**C NMR** (126 MHz, CDCl$_3$) δ 155.07, 143.29, 133.66, 132.01, 129.07, 127.99, 83.83, 83.51, 50.40, 28.40, 24.87, 22.86.

HRMS (ESI): $[M+H]^+$ Calculated 348.2341, found 348.2332 (2.6 ppm)

Optical rotation: $[\alpha]^{20}_{D}$ = +17.5 (*c* 1.0, CHCl$_3$)

Procedure adapted from Kleman et al.[41]

5 - methyl (R)-3''-(1-((tert-butoxycarbonyl)amino)ethyl)-4'-methoxy-[1,1':2',1''-terphenyl]-4-carboxylate *or* ***mop-R*-Tph-1**



Starting materials **2** (3.48 g, 12.56 mmol, 1 eq) and **4** (4.80 g, 13.82 mmol, 1.1 eq) were dissolved in 80 mL 4:1 dioxane/water with a stir bar. Cesium carbonate (12.28 g, 37.7 mmol, 3 eq) were added and the solution was sparged with nitrogen for ten minutes with gentle heating. When the mixture approached 80°C, Xphos Pd G_4 (540 mg, 5 mol%) was added and the mixture was set up to reflux under nitrogen overnight.

The next day, the solution was allowed to cool to room temperature, and most of the solvent was removed *in vacuo*. The remainder was diluted with ethyl acetate and washed with water, sodium bicarbonate, and brine before drying with sodium sulfate. The solution was concentrated into a black oil *in vacuo*. Column chromatography (10-20% EtOAc/hexane; $R_f \approx 0.3$ in 20%) yielded **5** (4.92 g, 84%) as a white foam.

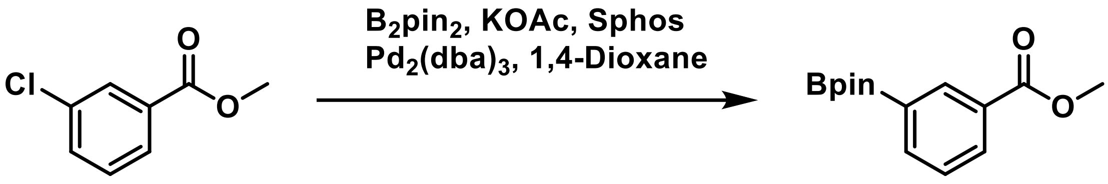
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.87 (d, *J* = 8.4 Hz, 1H), 7.36 (d, *J* = 8.3 Hz, 1H), 7.23 – 7.14 (m, 2H), 7.08 – 7.03 (m, 1H), 7.03 – 6.96 (m, 1H), 4.79 (d, *J* = 7.7 Hz, 1H), 4.73 – 4.63 (m, 1H), 3.88 (d, 3H), 1.42 (s, 6H), 1.24 (d, 2H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.99, 159.43, 155.07, 146.27, 141.88, 141.17, 132.16, 131.58, 129.92, 129.16, 128.40, 128.34, 127.70, 127.57, 124.59, 115.93, 113.25, 79.30, 55.43, 51.97, 49.95, 28.38, 22.50.

HRMS (ESI): $[M+H]^+$ Calculated 348.2341, found 348.2332 (2.6 ppm)

Optical rotation: $[\alpha]^{20}_{D}$ = +22.5 (*c* 1.0, $CHCl_3$)

6b - Methyl 3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzoate



6a **6b**

Into a dry flask was added $Pd_2(dba)_3$ (1.61 g, 1.76 mmol, 0.02 eq), Sphos (1.44 g, 3.52 mmol, 0.04 eq) and a stir bar. Contents purged with nitrogen, and 100 mL anhydrous 1,4-dioxane were added. Stirred for 10 min. Potassium acetate (25.91 g, 264 mmol, 3 eq) and methyl 3-chlorobenzoate **6a** (15.00 g, 88 mmol, 1 eq) were added with stirring. B_2pin_2 (24.54 g, 96.7 mmol, 1.1 eq) was added with 75 mL more dioxane. Refluxed under nitrogen.

After 16 hours, TLC showed complete conversion (5% ethyl acetate/pentane, $R_f \approx 0.45$) and the reaction was allowed to cool to room temperature. Mixture was filtered through celite and washed with chloroform.

Isolated with column chromatography (3-4% gradient ethyl acetate in pentane) to yield 21.12 g (92% yield) **6b** as a white solid.

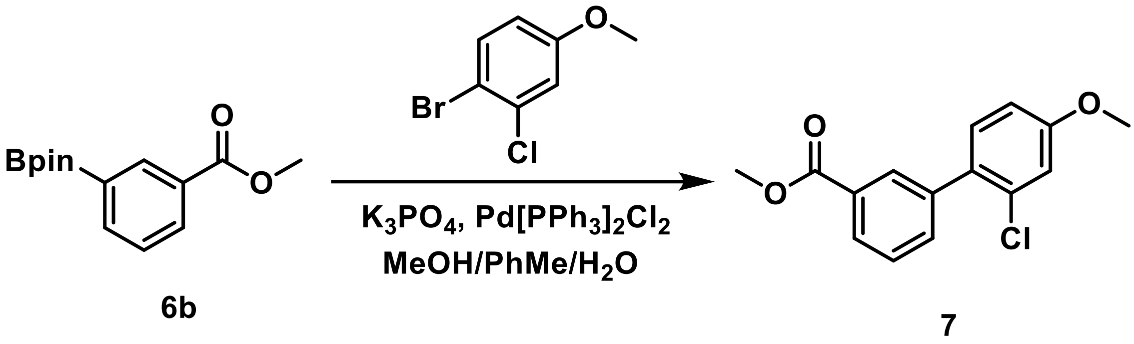
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.47 (t, *J* = 1.5 Hz, 1H), 8.13 (dt, *J* = 7.8, 1.6 Hz, 1H), 7.98 (dt, *J* = 7.5, 1.3 Hz, 1H), 7.45 (t, *J* = 7.6 Hz, 1H), 3.92 (s, 3H), 1.36 (s, 12H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 167.18, 139.16, 135.85, 132.31, 129.60, 127.82, 84.11, 52.03, 24.89.

HRMS (ASAP): $[M+H]^+$ Calculated 136.0519, found 136.0524 (3.6 ppm)

Procedure adapted from Kleman et al. with characterization that matched what was previously reported.[41]

7 - methyl 2'-chloro-4'-methoxy-[1,1'-biphenyl]-3-carboxylate



Boronic ester **6b** (20.00 g, 76.3 mmol, 1 eq) was dissolved in 1000 mL 85:15 toluene/methanol with a stir bar and sparged with nitrogen. 4-bromo-2-chloroanisole (21.97 g, 99.19 mmol, 1.3 eq) was then added. Potassium phosphate tribasic (48.6 g, 229 mmol, 3 eq) was dissolved in water to make a 2 M solution, which was then added to the flask. While continuing to sparge, bis(triphenylphosphine)palladium chloride (3.75 g, 5.34 mmol, 7 mol%) were added. Mixture was refluxed under nitrogen for 16 hours, at which point TLC (5% ethyl acetate/pentane, $R_f \approx 0.3$) indicated complete conversion.

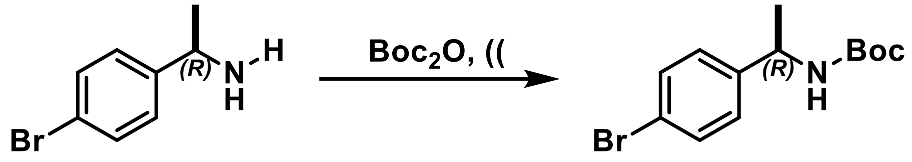
After cooling to room temperature, the mixture was filtered through celite and added to a separatory funnel. Water was drained out, and the organic layer was rinsed with equal volumes of water, sodium bicarbonate (sat.) and brine (sat.) and dried with sodium sulfate. The solution was concentrated under vacuum to yield a dark brown solid that was recrystallized in hot methanol to yield white crystals (17.36 g, 83%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.09 (t, *J* = 1.8 Hz, 1H), 8.02 (dt, *J* = 7.8, 1.5 Hz, 1H), 7.62 (dt, *J* = 7.8, 1.5 Hz, 1H), 7.47 (t, *J* = 7.7 Hz, 1H), 7.25 (d, *J* = 8.5 Hz, 1H), 7.03 (d, *J* = 2.6 Hz, 1H), 6.87 (dd, *J* = 8.5, 2.6 Hz, 1H), 3.92 (s, 3H), 3.83 (s, 3H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.96, 159.68, 139.44, 134.15, 132.99, 131.90, 131.83, 130.72, 130.10, 128.43, 128.07, 115.22, 113.16, 55.59, 52.15.

HRMS (ASAP): $[M+H]^+$ calculated 277.0626, found 277.0625 (0.4 ppm)

8b - tert-butyl (*R*)-(1-(4-bromophenyl)ethyl)carbamate



8a **8b**

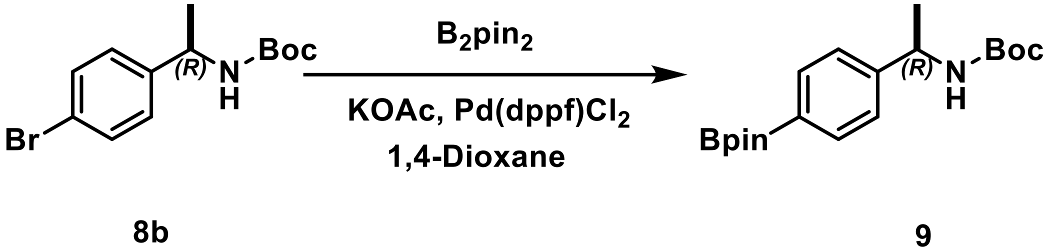
(*R*)-1-(4-bromophenyl)ethan-1-amine (**8a**) (2.00 g, 10 mmol, 1 eq) was measured in a vial and was added dropwise to fused di-tert-butyl dicarbonate (2.4 g, 11 mmol, 1.1 eq) with stirring. After the bubbling ceased and the reaction returned to room temperature, the remaining **8a** was transferred with 5 mL ethyl ether and the mixture was sonicated for 30 minutes. The oil was dried under vacuum and co-evaporated with 10 mL water to provide the title compound in quantitative yield (3.02 g), which was used for the next step without purification.

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.43 (d, *J* = 8.5 Hz, 1H), 7.17 (d, *J* = 8.1 Hz, 1H), 4.88 (s, 1H), 4.73 (s, 1H), 1.40 (d, 9H).

^{13}C NMR (126 MHz, CDCl$_3$) δ 155.01, 143.29, 131.58, 127.61, 120.78, 79.56, 49.69, 28.36, 22.61.

HRMS (ASAP): [M+H]$^+$ Calculated 300.0594, found 300.0592 (0.7 ppm)

9 - tert-butyl (R)-(1-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)ethyl)carbamate



Pd(dppf)Cl$_2$ (610 mg, 5 mol%) was transferred into a dry flask with a stir bar and sparged with nitrogen. 70 mL dry 1,4-dioxane was cannulated into the flask. Aryl chloride **8b** (5.00 g, 16.66 mmol, 1 eq) was added followed by potassium acetate (4.90 g, 50 mmol, 3 eq) and B$_2$pin$_2$ (4.66 g, 18.3 mmol, 1.1 eq). Continued to sparge for 10 minutes. Set up for reflux under nitrogen. After 16 hours, more Pd(dppf)Cl$_2$ (300 mg, 2.5 mol%) and B$_2$pin$_2$ (2.5 g, 0.6 eq) were added. Continued reflux for 2 hours. Crude NMR confirmed completion of reaction at this time. Filtered mixture through silica plug and evaporated to dryness *in vacuo*.

Conducted flash column chromatography (10-25% EtOAc in pentane, $R_f \approx 0.4$). Material co-elutes with excess B$_2$pin$_2$. Dissolved contaminated product in ethyl acetate and washed with 25% triethanolamine in water twice. Isolated 3.70 g (64% yield) **9** as a pale yellow oil.

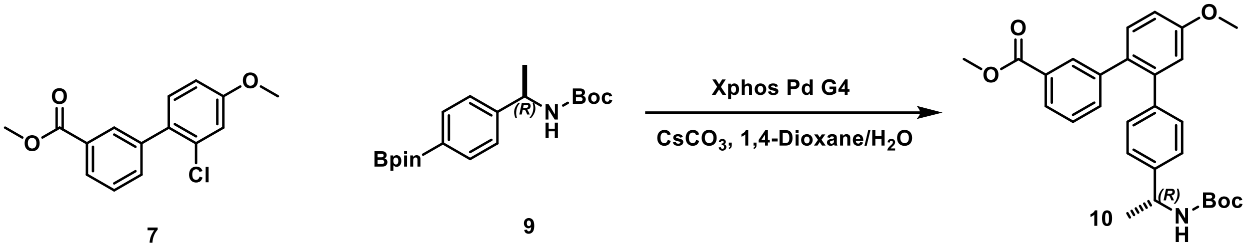
^{1}H NMR (500 MHz, CDCl$_3$) δ 7.77 (d, *J* = 8.0 Hz, 2H), 7.30 (d, *J* = 7.6 Hz, 2H), 4.79 (s, 2H), 1.42 (d, 12H), 1.34 (s, 12H).

^{13}C NMR (126 MHz, CDCl$_3$) δ 171.10, 154.95, 147.31, 135.06, 125.02, 83.58, 79.30, 60.33, 50.26, 28.17, 24.78.

HRMS (ESI): [M+H]$^+$ Calculated 348.2341, found 348.2330 (3.2 ppm)

Adapted from the procedure published by Kleman et al. Characterization data match what was reported.[41]

10 - methyl (R)-4''-(1-((tert-butoxycarbonyl)amino)ethyl)-4'-methoxy-[1,1':2',1''-terphenyl]-3-carboxylate or ***pom-R*-Tph-1**



Starting materials **7** (2.68 g, 9.69 mmol, 1 eq) and **9** (3.70 g, 10.65 mmol, 1.1 eq) were dissolved in 40 mL 4:1 dioxane/water with a stir bar. Solution was degassed using freeze-pump-thaw method. Cesium carbonate (9.47 g, 29.06 mmol, 3 eq) were added and the solution was sparged with nitrogen for ten minutes with gentle heating. When the mixture approached 80°C, Xphos Pd G_4 (440 mg, 5.3 mol%) was added and the mixture was set up to reflux under nitrogen.

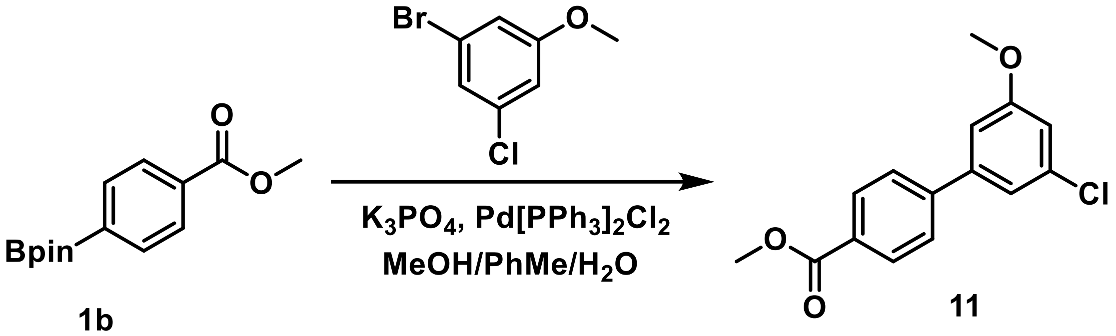
The reaction was complete after 4 hours, and the solution was allowed to cool to room temperature. Most of the solvent was removed *in vacuo*. The remainder was diluted with chloroform and washed with water, sodium bicarbonate, and brine before drying with sodium sulfate. The solution was concentrated into a black oil *in vacuo*. Column chromatography (20-25% EtOAc/hexane; $R_f \approx 0.25$ in 20%) yielded **10** (3.91 g, 87%) as a white foam.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.90 (s, 1H), 7.83 (dt, *J* = 7.2, 1.8 Hz, 1H), 7.37 – 7.33 (m, 1H), 7.20 – 7.11 (m, 4H), 7.11 – 7.06 (m, 2H), 6.98 – 6.94 (m, 2H), 4.79 (d, 2H), 3.86 (d, 6H), 1.41 (d, 13H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.96, 158.85, 155.04, 141.40, 139.90, 134.81, 132.04, 131.49, 130.61, 129.63, 127.67, 127.03, 125.57, 116.09, 113.26, 79.20, 67.27, 60.57, 55.23, 52.13, 49.88, 28.38, 22.44, 14.07.

HRMS (ESI): $[M+NH_4]^+$ Calculated 479.2540, found 479.2538 (0.4 ppm)

11 - methyl 3'-chloro-5'-methoxy-[1,1'-biphenyl]-4-carboxylate



Boronic ester **1b** (4.5 g, 17.36 mmol, 1 eq) was dissolved in 200 mL 85:15 toluene/methanol with a stir bar and sparged with nitrogen. Aryl chloride (5.00 g, 22.6 mmol, 1.3 eq) was then added. Potassium phosphate tribasic (11.15, 52.1 mmol, 3 eq) was dissolved in water to make a 2 M solution, which was then added to the flask. While continuing to sparge, bis(triphenylphosphine)palladium chloride (0.85 g, 7 mol%) were added. Mixture was refluxed under nitrogen for 16 hours, at which point TLC (5% ethyl acetate/pentane, $R_f \approx 0.3$) indicated complete conversion.

After cooling to room temperature, the mixture was filtered through celite and added to a separatory funnel. Water was drained out, and the organic layer was rinsed with equal volumes of water, sodium bicarbonate (sat.) and brine (sat.) and dried with sodium sulfate. The solution was concentrated under vacuum to yield a dark brown solid that was purified by flash column chromatography (7.5% EtOAc/Hexanes) to yield 3.65 g (75%) **11**.

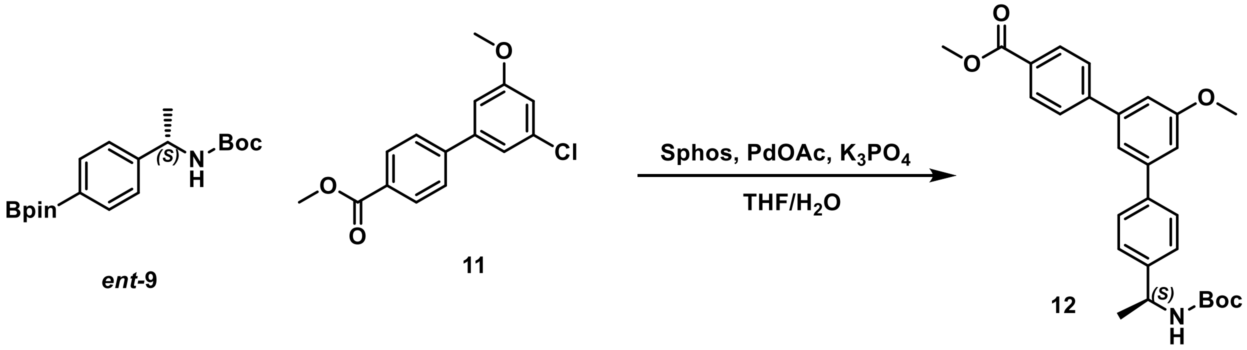
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.10 (d, *J* = 8.5 Hz, 1H), 7.62 (d, *J* = 8.4 Hz, 1H), 7.19 (t, *J* = 1.7 Hz, 1H), 7.02 (dd, *J* = 2.3, 1.6 Hz, 1H), 6.93 – 6.92 (m, 1H), 3.94 (s, 2H), 3.86 (s, 2H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.81, 160.67, 144.21, 142.63, 135.45, 130.18, 129.62, 127.09, 119.92, 113.65, 111.88, 55.67, 52.20.

HRMS (ASAP): $[M+H]^+$ Calculated 277.0626, found 277.0621 (1.8 ppm)

Experiment conducted by Schimmelpfennig, L.

12 - methyl (S)-4''-(1-((tert-butoxycarbonyl)amino)ethyl)-5'-methoxy-[1,1':3',1''-terphenyl]-4-carboxylate or ***pmp-S*-Tph-1**



PdOAc (221 mg, 7.5 mol%) and Sphos (0.81 g, 15 mol%) were combined in a flask and purged with nitrogen. A mixture of 50 mL 3:1 THF/water was added and solution was sonicated for 5 minutes. Solution was sparged with nitrogen. Biaryl **11** (3.6 g, 13 mmol) and boronic ester ***ent*-9** (5 g, 14.5 mmol) were added followed by potassium phosphate tribasic (8.24 g, 39 mmol). Solution was heated at 80°C overnight.

The reaction was complete after 24 hours, and the solution was allowed to cool to room temperature. Most of the solvent was removed *in vacuo*. The remainder was diluted with chloroform and washed with water, sodium bicarbonate, and brine before drying with sodium sulfate. The solution was concentrated into a black oil *in vacuo*. Column chromatography yielded **12** (2.3 g, 24%) as a white foam.

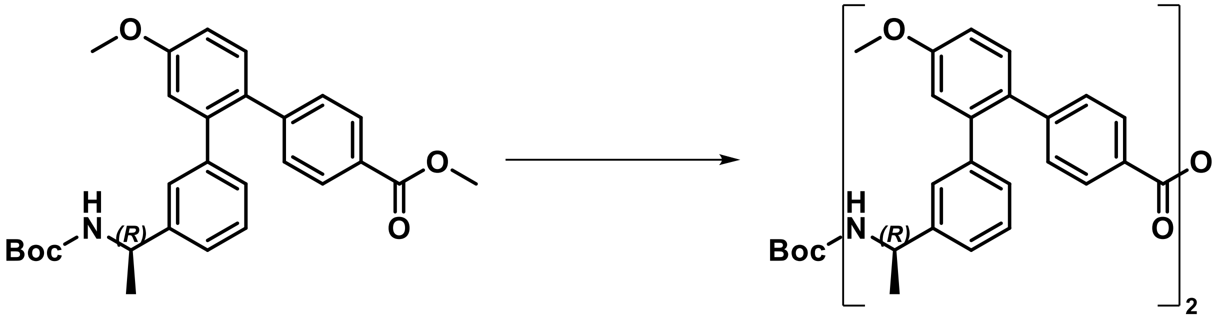
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.10 (d, *J* = 8.4 Hz, 2H), 7.68 (d, *J* = 8.4 Hz, 2H), 7.58 (d, *J* = 8.3 Hz, 2H), 7.41 – 7.36 (m, 3H), 7.11 (ddd, *J* = 9.7, 2.4, 1.5 Hz, 2H), 4.99 (s, 1H), 4.85 (s, 1H), 3.93 (s, 2H), 3.90 (s, 3H), 1.46 (d, 12 H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.97, 160.40, 155.19, 145.53, 143.75, 143.06, 141.87, 139.65, 130.11, 129.14, 127.44, 127.18, 126.36, 118.85, 112.49, 111.81, 79.47, 58.28, 55.47, 52.13, 49.93, 28.41, 22.69, 18.39.

HRMS (ESI): $[2M+Na]^+$ Calculated 940.4743, found 940.4733 (1.1 ppm)

Experiment conducted by Schimmelpfennig, L.

***mop-R*-Tph-2**



Synthesized according to **General Procedure 1** using two 5.63 mmol portions of *mop-R*-Tph-1**.**

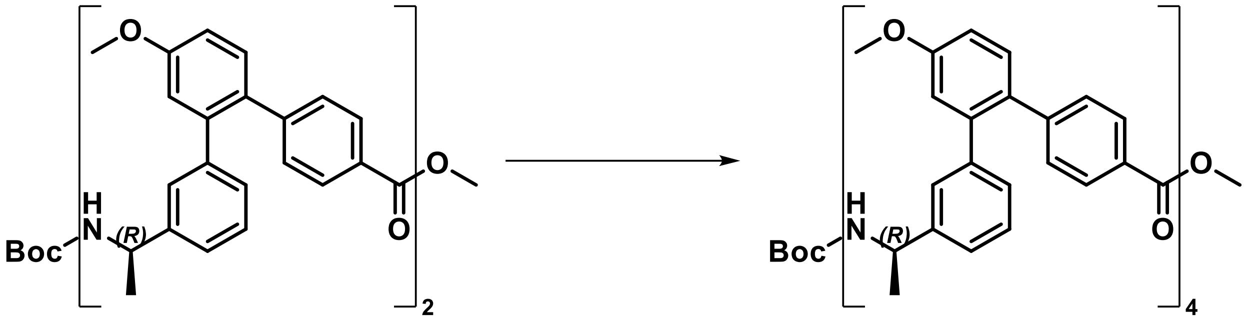
Purification with column chromatography (40% ethyl acetate/hexane) afforded 3.96 g (89% yield) *mop-R*-Tph-1.

^{1}H NMR (500 MHz, DMSO) δ 8.55 (d, *J* = 8.0 Hz, 1H), 7.73 (d, *J* = 8.1 Hz, 2H), 7.68 (d, *J* = 8.1 Hz, 2H), 7.38 (dd, *J* = 8.5, 1.6 Hz, 2H), 7.27 – 7.02 (m, 15H), 6.97 (dd, *J* = 5.9, 2.7 Hz, 2H), 6.93 (t, *J* = 4.7 Hz, 2H), 5.02 (p, *J* = 7.1 Hz, 1H), 4.52 (t, *J* = 7.5 Hz, 1H), 3.85 (d, *J* = 2.5 Hz, 7H), 3.79 (s, 3H), 1.35 (s, 9H), 1.23 (d, 3H), 1.11 (d, 3H).

^{13}C NMR (126 MHz, DMSO) δ 166.49, 165.44, 159.56, 159.37, 155.24, 146.29, 145.64, 145.03, 144.16, 141.96, 141.89, 141.03, 140.87, 132.36, 132.22, 132.07, 131.88, 130.23, 129.69, 129.20, 128.52, 128.25, 128.14, 128.10, 128.00, 127.80, 127.63, 127.44, 125.19, 124.82, 116.18, 113.77, 113.70, 78.08, 55.75, 52.38, 49.90, 48.63, 28.70, 23.05, 22.39.

HRMS (ASAP): $[M+H]^+$ Calculated 791.3691, found 791.3689 (0.3 ppm)

***mop-R*-Tph-4**



Synthesized according to **General Procedure 1** using two 1.73 mmol portions of *mop-R*-Tph-2**.**

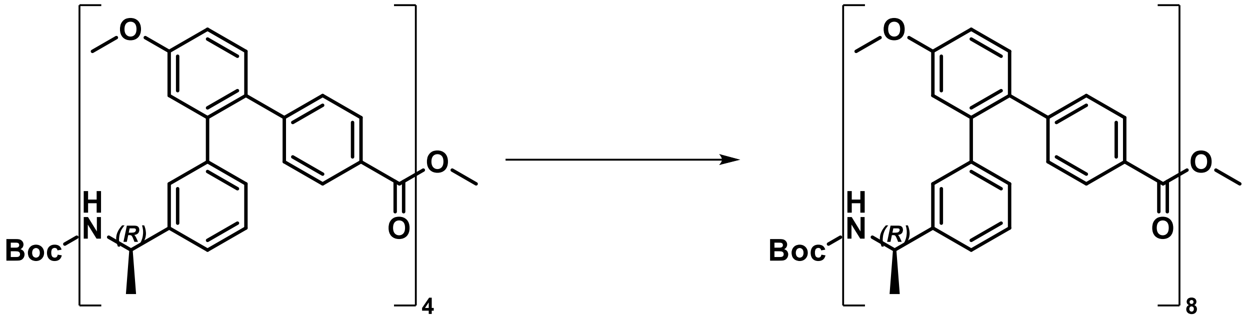
Purification with column chromatography (5% methanol in DCM) afforded 1.96 g (80% yield) *mop-R*-Tph-4.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.78 (d, J = 8.0 Hz, 2H), 7.69 – 6.89 (m, 37H), 6.64 – 6.16 (m, 4H), 5.09 (dt, J = 18.0, 8.7 Hz, 2H), 4.85 – 4.50 (m, 1H), 4.05 – 3.74 (m, 12H), 1.48 – 1.15 (m, 15H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.99, 166.75, 166.73, 159.40, 159.36, 159.34, 159.32, 155.02, 146.21, 144.29, 143.19, 141.81, 141.79, 141.77, 141.74, 141.29, 141.26, 141.23, 141.19, 132.45, 132.41, 132.23, 132.16, 132.11, 131.62, 131.58, 131.52, 131.50, 129.98, 129.95, 129.92, 129.90, 129.86, 129.83, 129.81, 129.15, 129.12, 128.57, 128.54, 128.49, 128.45, 128.41, 128.33, 127.78, 127.71, 127.71, 127.66, 126.68, 126.62, 126.55, 124.95, 124.61, 124.56, 115.93, 115.86, 113.27, 113.24, 55.54, 55.52, 55.47, 55.45, 55.43, 55.42, 51.98, 48.89, 48.86, 28.64, 28.52, 28.38, 22.22, 21.82.

HRMS (ASAP): $[M+H]^+$ Calculated 1449.6522, found 1449.6536 (1.0 ppm)

mop-R-Tph-8



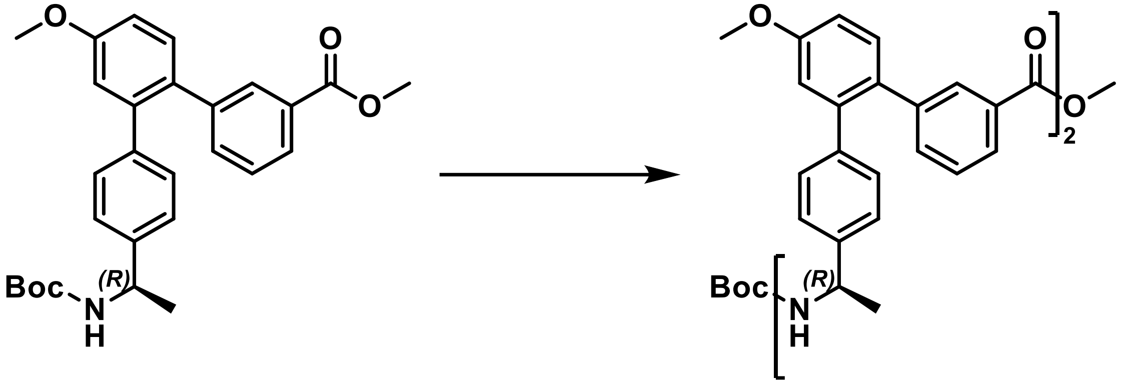
Synthesized according to **General Procedure 1** using two 0.51 mmol portions of *mop-R*-Tph-4**.**

Purification with column chromatography (2% MeOH in DCM) afforded 0.98 g (69% yield) *mop-R*-Tph-8.

^{1}H NMR (500 MHz, DMSO) δ 8.53 (p, *J* = 8.1 Hz, 7H), 7.75 – 7.56 (m, 16H), 7.42 – 6.85 (m, 75H), 4.98 (p, *J* = 7.6 Hz, 7H), 4.54 – 4.40 (m, 1H), 3.90 – 3.73 (m, 27H), 1.41 – 1.11 (m, 48H).

HRMS (ESI): $[M+2H]^{2+}$ Calculated 1383.6129, found 1383.6133 (0.3 ppm)

***pom-R*-Tph-2**



Synthesized according to **General Procedure 1** using two 4.23 mmol portions of *pom-R*-Tph-1**.**

Purification with column chromatography (40% EtOAc/pentane) afforded 2.62 g (78% yield) *pom-R*-Tph-2.

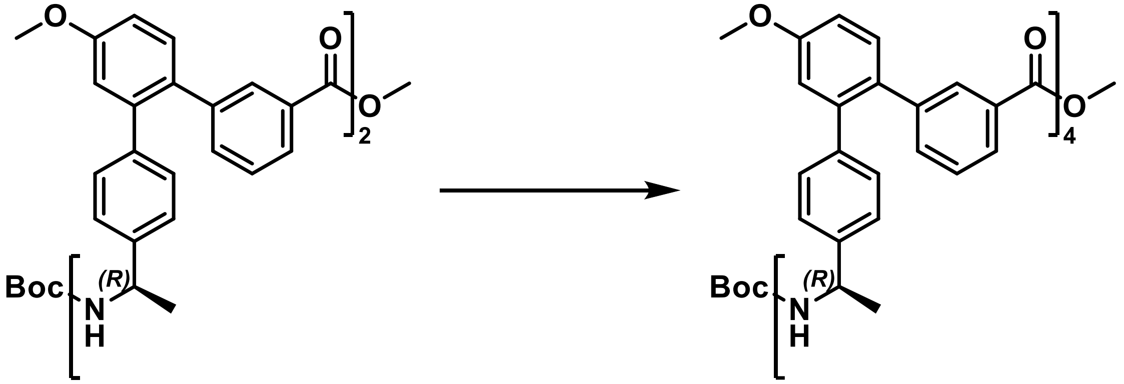
HRMS (ESI): $[M+2H]^{2+}$ Calculated 791.3691, found 791.3695 (0.5 ppm)

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.91 (dt, *J* = 2.5, 1.0 Hz, 1H), 7.84 (ddd, *J* = 5.5, 3.6, 1.7 Hz, 1H), 7.60 (dt, *J* = 7.6, 1.6 Hz, 1H), 7.54 (s, 1H), 7.37 (ddd, *J* = 7.6, 6.7, 0.9 Hz, 2H), 7.25 – 7.07 (m, 11H), 7.02 – 6.95 (m, 4H), 6.21 – 6.15 (m, 1H), 4.80 (d, *J* = 54.0 Hz, 2H), 3.89 (d, *J* = 3.1 Hz, 6H), 3.84 (s, 3H), 1.53 (d, *J* = 6.9 Hz, 3H), 1.50 – 1.35 (m, 12H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 206.93, 171.14, 167.07, 166.64, 159.23, 159.22, 155.09, 141.70, 141.53, 141.50, 141.40, 140.26, 140.05, 134.65, 134.41, 133.02, 132.15, 132.10, 131.72, 131.62, 130.80, 130.13, 130.00, 129.99, 128.20, 128.09, 127.82, 127.36, 125.94, 125.60, 124.94, 115.99, 115.93, 113.26, 113.24, 79.39, 77.35, 77.10, 76.84, 60.39, 55.44, 52.03, 48.76, 28.39, 22.50, 21.65, 21.04, 14.31, 14.21, 14.06.

HRMS (ESI): $[M+H]^+$ Calculated 791.3691, found 791.3682 (1.1 ppm)

***pom-R*-Tph-4**



Synthesized according to **General Procedure 1** using two 1.32 mmol portions of *pom-R*-Tph-2**.**

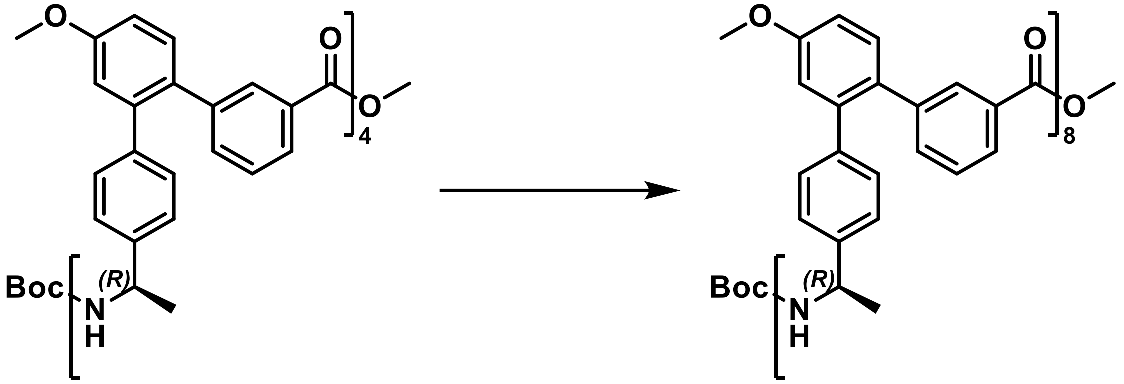
Afforded 1.81 g crude *pom-R*-Tph-4 that was purified using normal phase HPLC (5% MeOH in DCM).

^{1}H NMR (500 MHz, CDCl3) δ 7.96 – 6.89 (m, 44H), 5.19 (q, *J* = 6.8 Hz, 2H), 4.70 (s, 1H), 3.92 – 3.85 (m, 12H), 1.49 – 1.31 (m, 20H).

^{13}C NMR (126 MHz, CDCl3) δ 171.16, 167.09, 166.62, 166.60, 159.27, 159.25, 159.23, 159.20, 141.76, 141.70, 141.66, 141.55, 141.52, 141.50, 141.49, 141.48, 141.40, 141.38, 140.34, 140.33, 140.23, 140.04, 134.64, 134.42, 134.35, 133.04, 133.02, 132.99, 132.14, 132.08, 131.76, 131.71, 131.66, 131.64, 131.62, 130.83, 130.80, 130.20, 130.17, 130.13, 130.11, 130.00, 129.96, 129.95, 129.93, 128.36, 128.35, 128.21, 128.16, 128.12, 128.07, 127.87, 127.82, 127.34, 126.06, 125.95, 125.89, 125.82, 125.63, 125.61, 125.60, 125.58, 124.95, 124.90, 124.86, 115.99, 115.97, 115.95, 115.94, 115.91, 113.30, 113.27, 113.24, 113.22, 113.20, 79.43, 79.42, 79.41, 60.40, 55.51, 55.48, 55.46, 55.44, 52.07, 52.03, 48.71, 48.68, 48.59, 28.39, 21.64, 21.61, 21.59, 21.05, 14.21.

HRMS (ESI): $[M+H]^+$ Calculated 1449.6522, found 1449.6530 (0.5 ppm)**HRMS** (ESI): $[M+H]^+$ Calculated 1449.6522, found 1449.6489 (2.3 ppm)

***pom-R*-Tph-8**



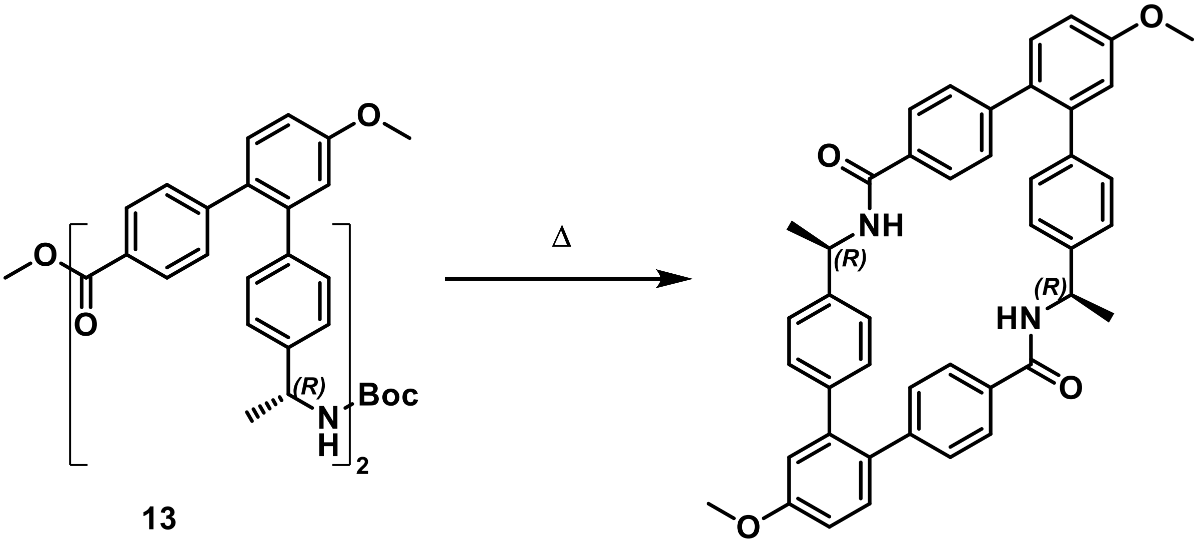
Synthesized according to **General Procedure 1** using two 0.499 mmol portions of *pom-R*-Tph-4**.**

Afforded 1.21 g crude *pom-R*-Tph-8 that was purified using normal phase HPLC (5% MeOH in DCM).

This compound was insoluble in $CDCl_3$ and complete characterization was not performed.

HRMS (ESI): $[M+H]^+$ Calculated 2766.2186, found 2766.24

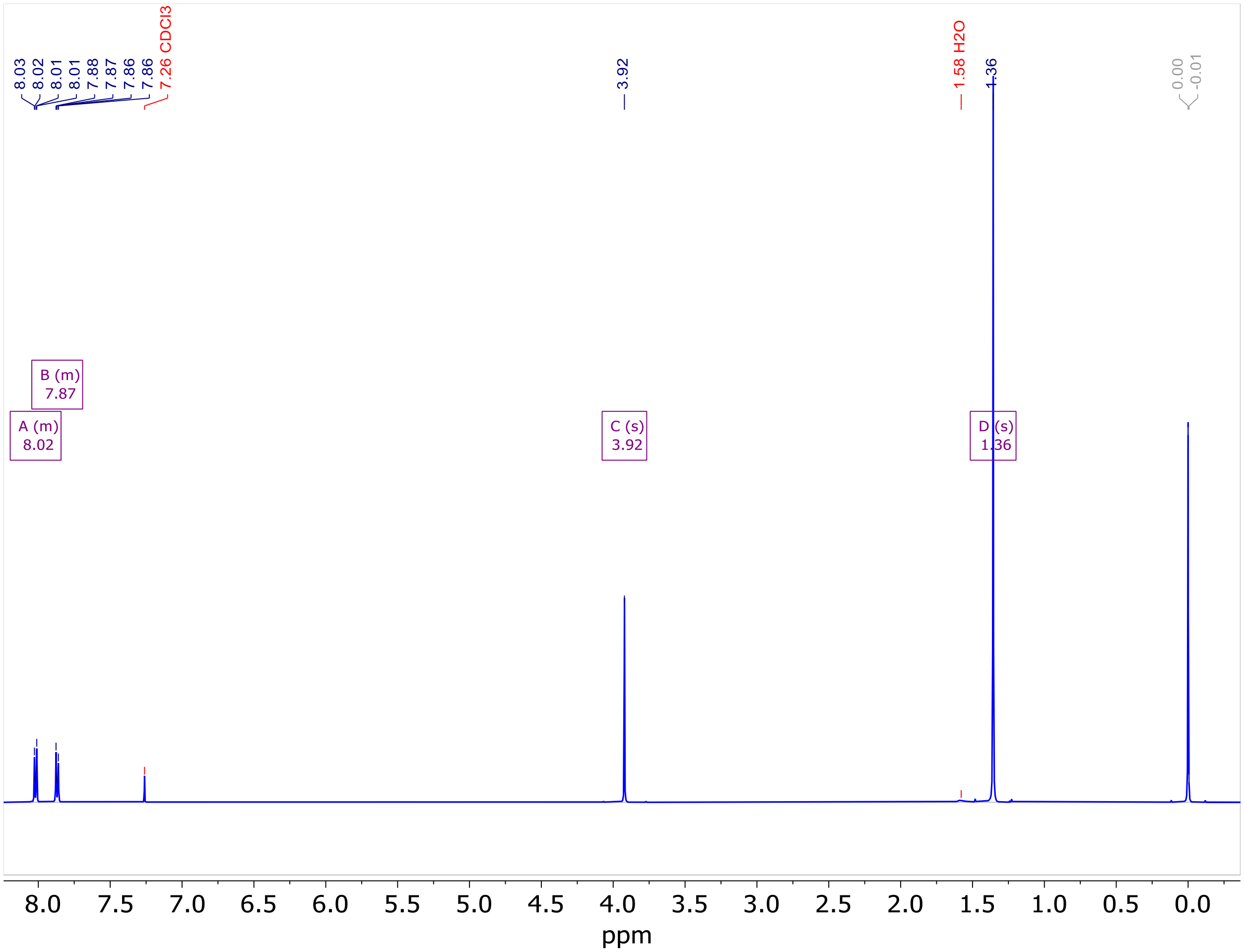
pop-R-Tph Cyclic Dimer



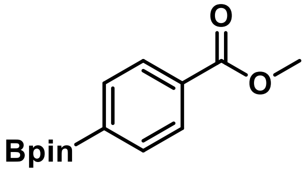
13

14

pop-R-Tph-2 (~500 mg) was placed in a vacuum sublimation apparatus. The apparatus was evacuated using an oil vacuum pump with a liquid nitrogen cold trap. Tap water was used to cool the cold finger. The sublimation apparatus was placed in a sand bath and the temperature was raised to ~200°C. After 2 hours, the apparatus was disassembled and the crystals were characterized with x-ray crystallography (data acquired by Ilia Guzei).

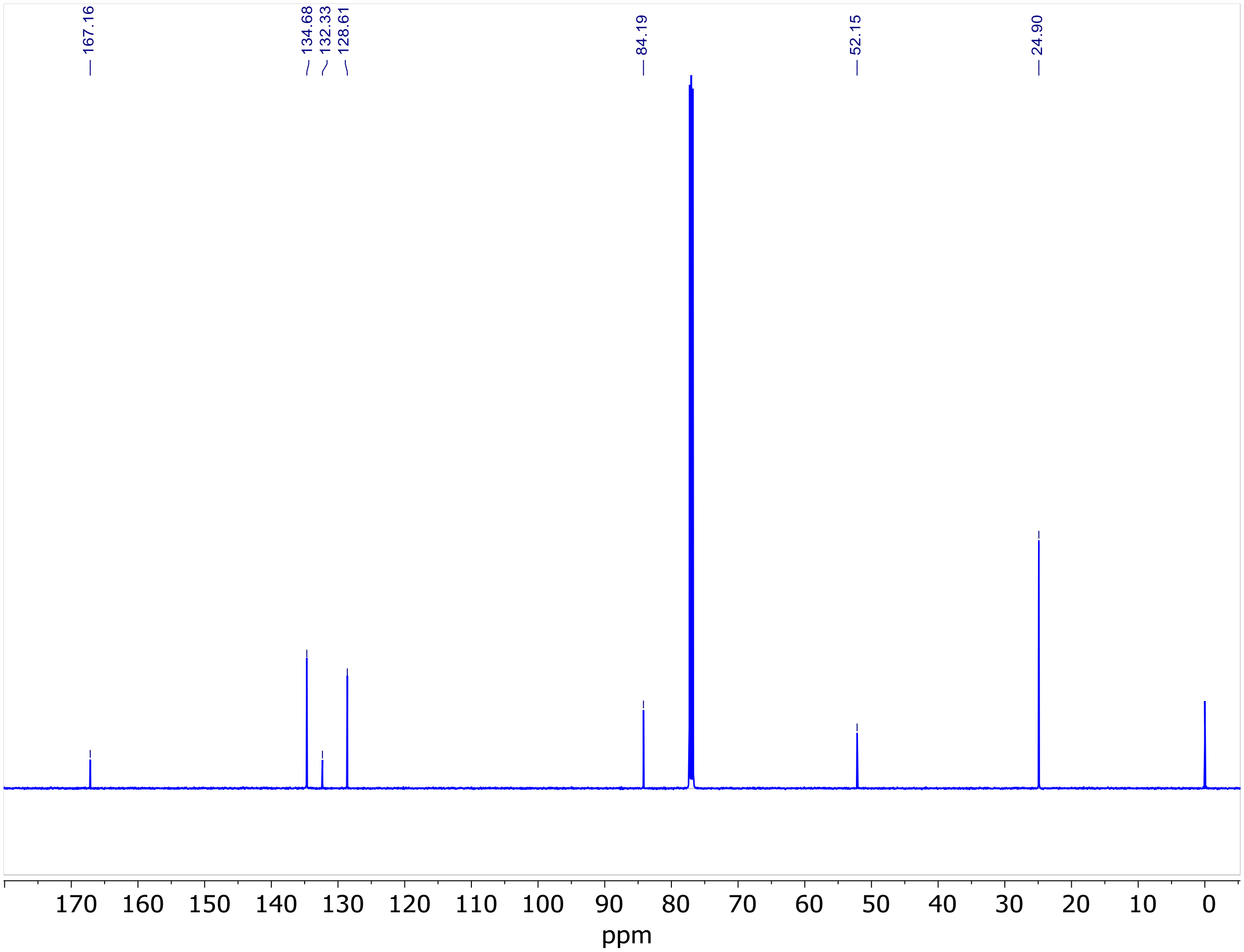


1b - Methyl 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzoate

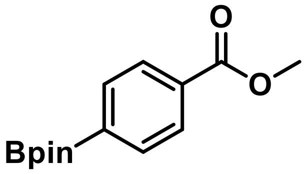


1b

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.03, 8.02, 8.01, 8.01, 7.88, 7.87, 7.86, 7.86, 7.26, 3.92, 1.58, 1.36, 0.00, -0.01.

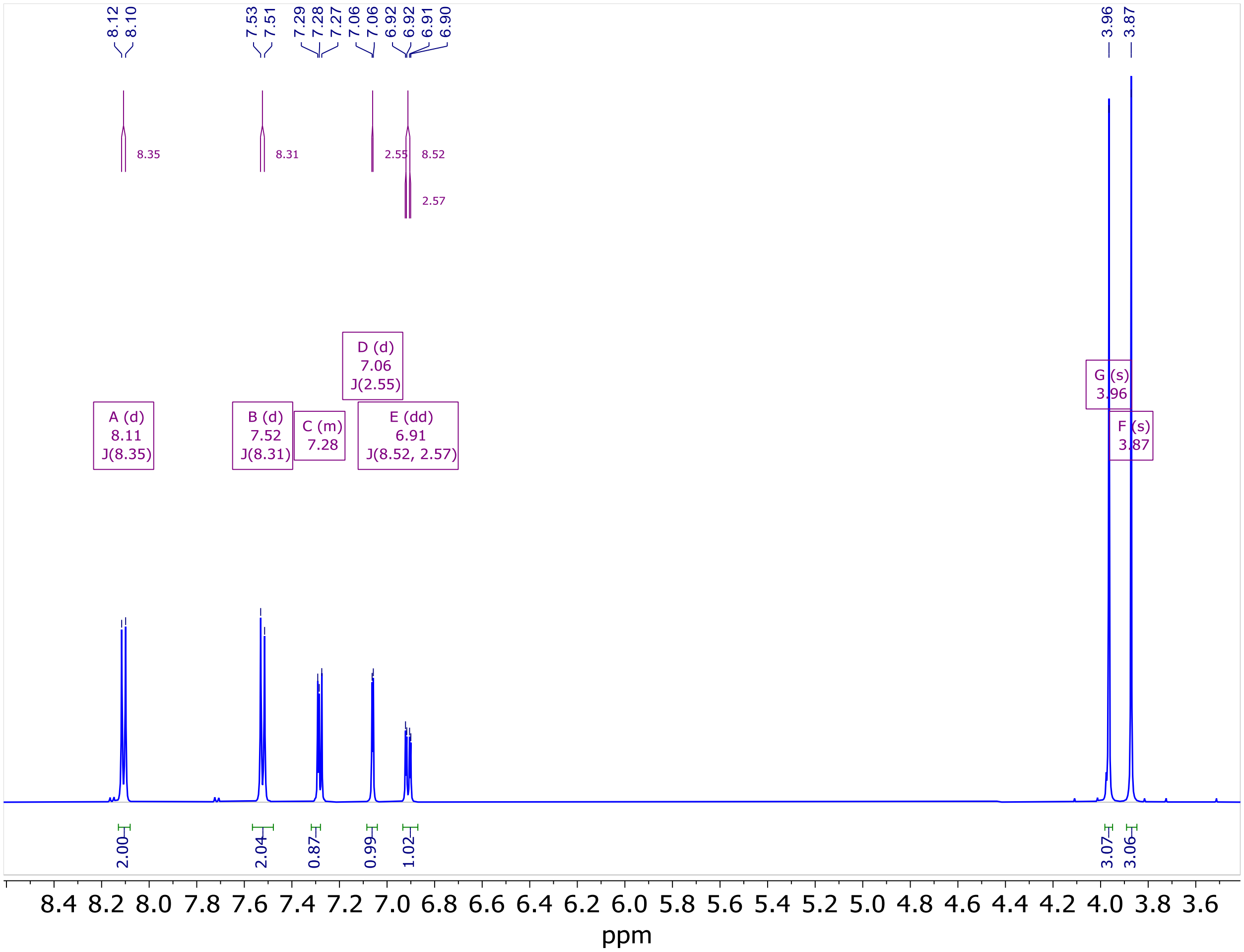


1b - Methyl 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzoate

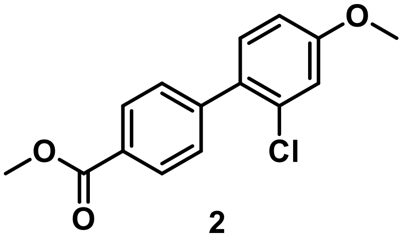


1b

^{13}C NMR (126 MHz, CDCl$_3$) δ 167.16, 134.68, 132.33, 128.61, 84.19, 52.15, 24.90.

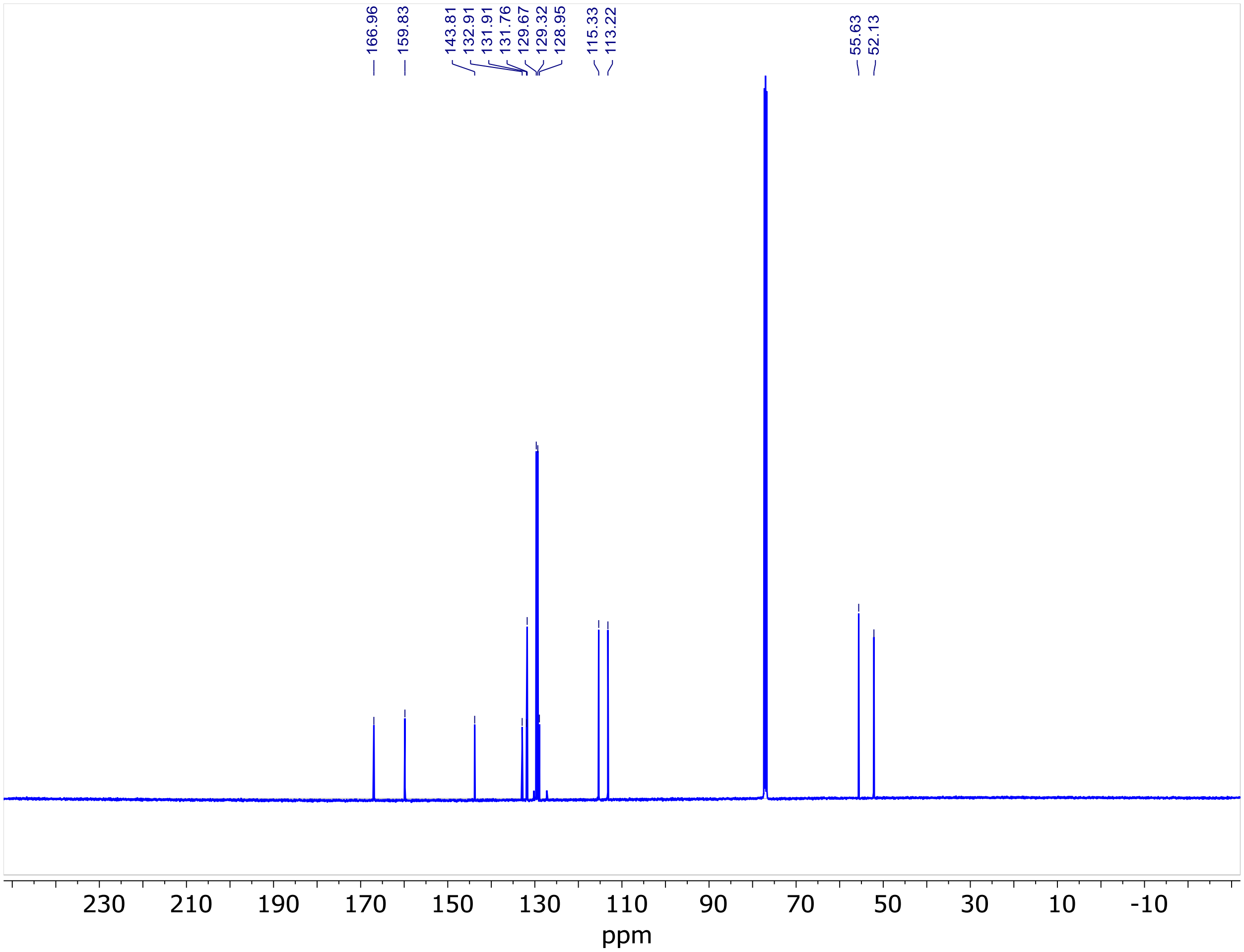


2 - methyl 2'-chloro-4'-methoxy-[1,1'-biphenyl]-4-carboxylate

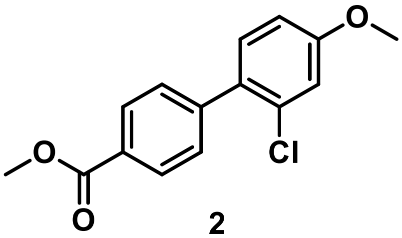


2

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.11 (d, *J* = 8.3 Hz, 1H), 7.52 (d, *J* = 8.3 Hz, 1H), 7.40 – 7.24 (m, 1H), 7.06 (d, *J* = 2.6 Hz, 1H), 6.91 (dd, *J* = 8.5, 2.6 Hz, 1H), 3.96 (s, 2H), 3.87 (s, 2H).

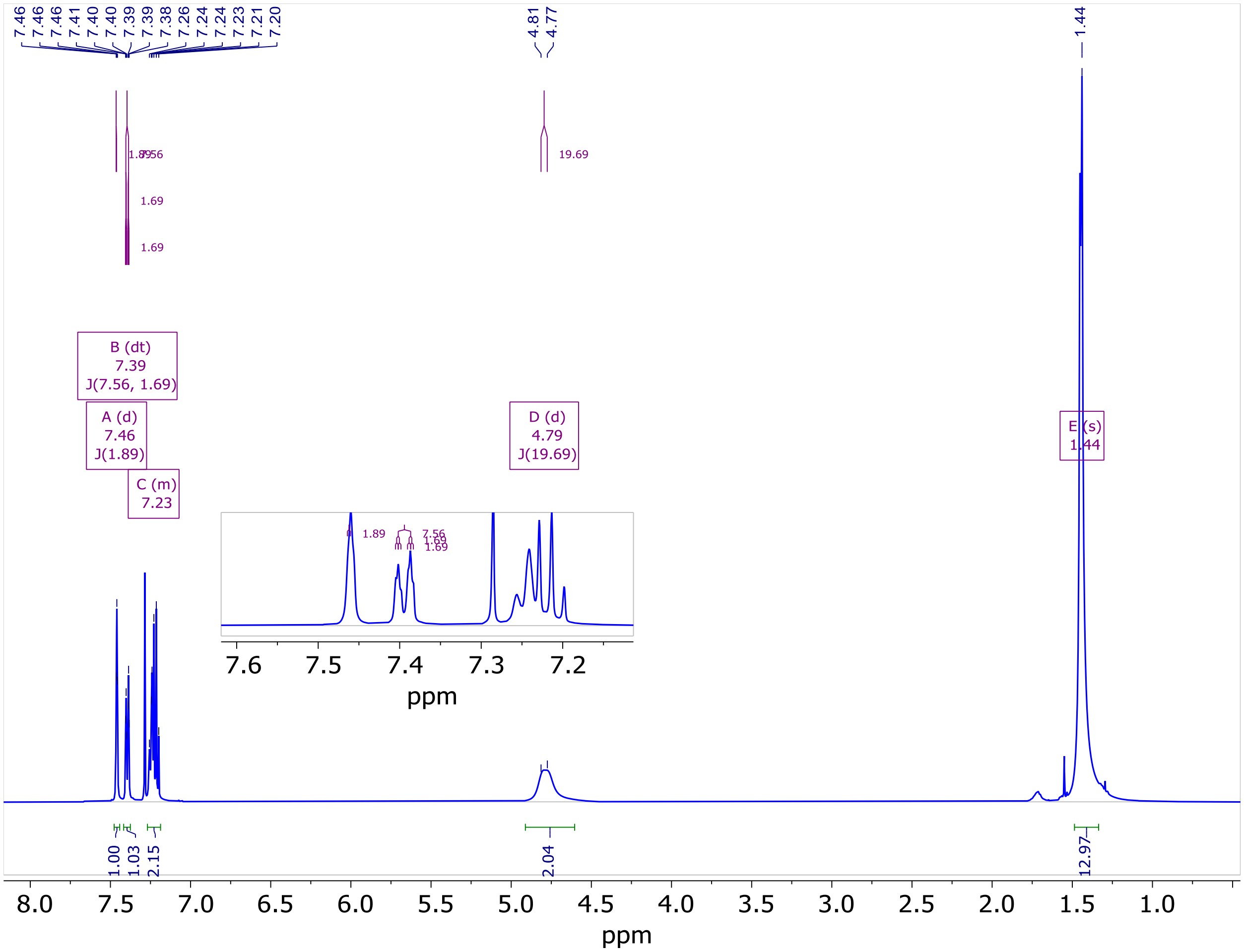


2 - methyl 2'-chloro-4'-methoxy-[1,1'-biphenyl]-4-carboxylate

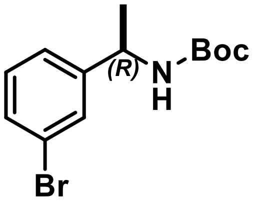


2

^{13}C NMR (126 MHz, CDCl$_3$) δ 166.96, 159.83, 143.81, 132.91, 131.91, 131.76, 129.67, 129.32, 128.95, 115.33, 113.22, 55.63, 52.13.

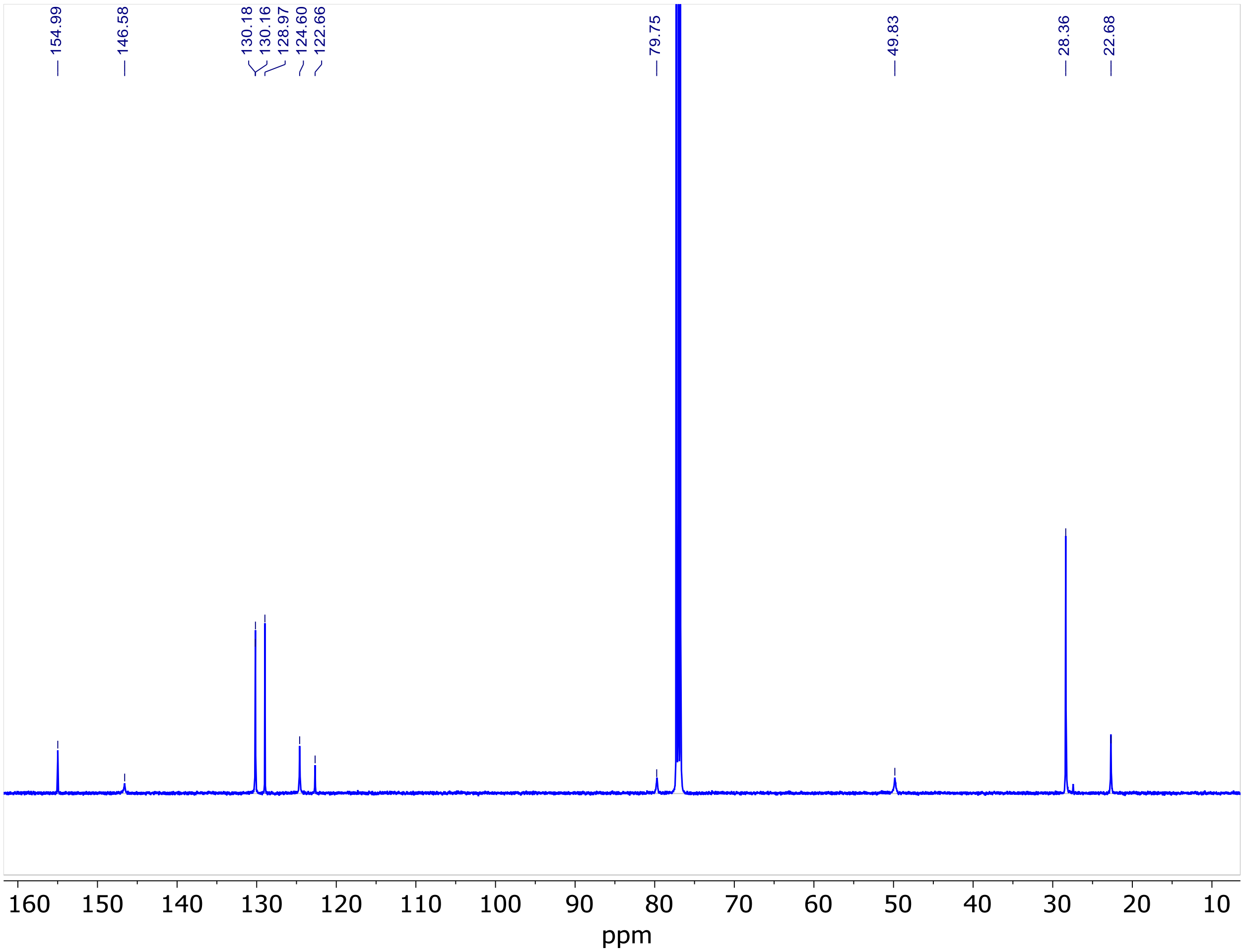


3b - tert-butyl (*R*)-(1-(3-chlorophenyl)ethyl)carbamate

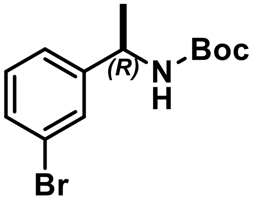


3b

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.46 (d, *J* = 1.9 Hz, 1H), 7.39 (dt, *J* = 7.6, 1.7 Hz, 1H), 7.26 – 7.18 (m, 2H), 4.79 (d, 2H), 1.44 (s, 12H).

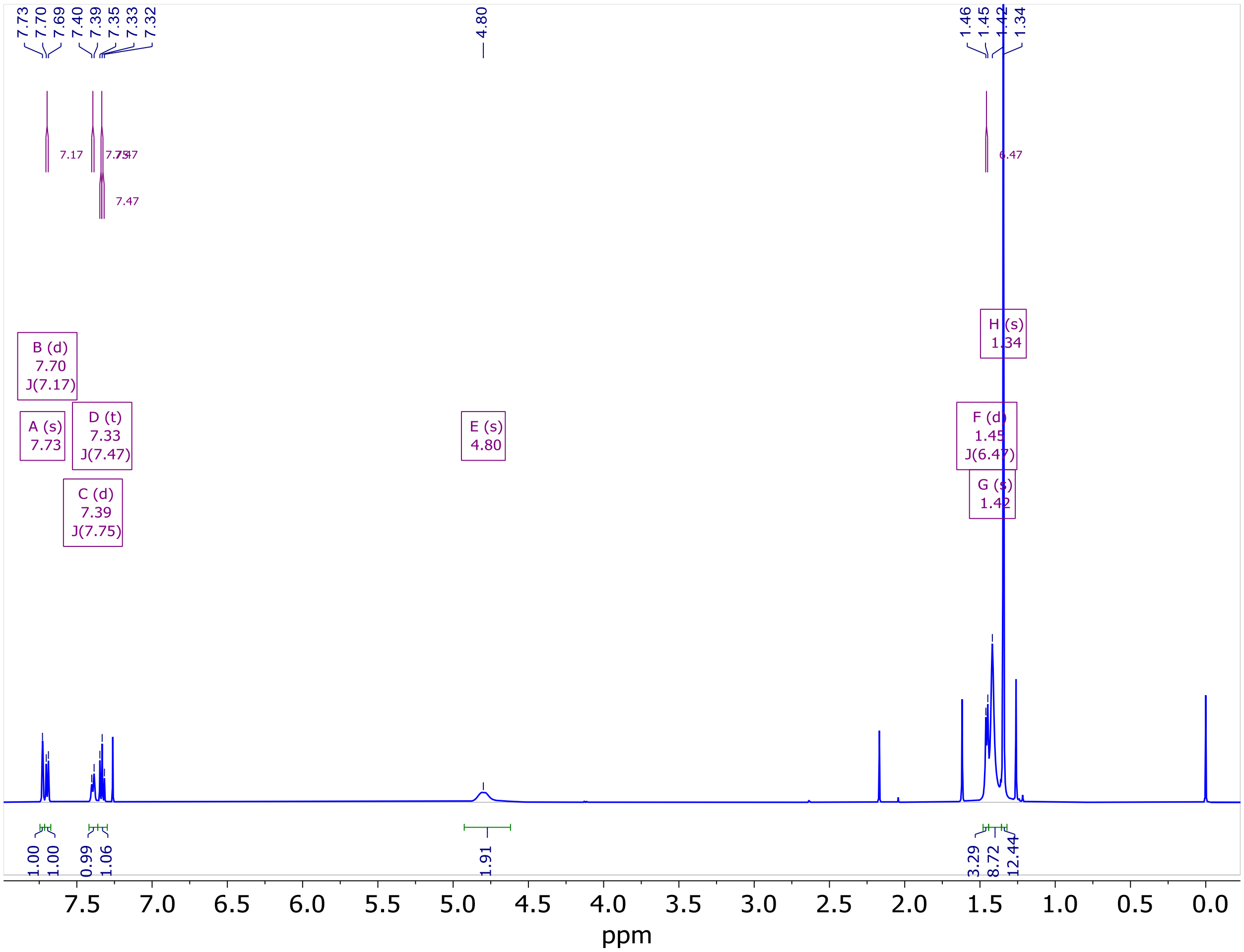


3b - tert-butyl (*R*)-(1-(3-chlorophenyl)ethyl)carbamate

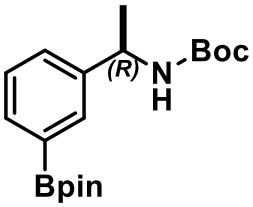


3b

^{13}C NMR (126 MHz, CDCl$_3$) δ 154.99, 146.58, 130.18, 130.16, 128.97, 124.60, 122.66, 79.75, 49.83, 28.36, 22.68.

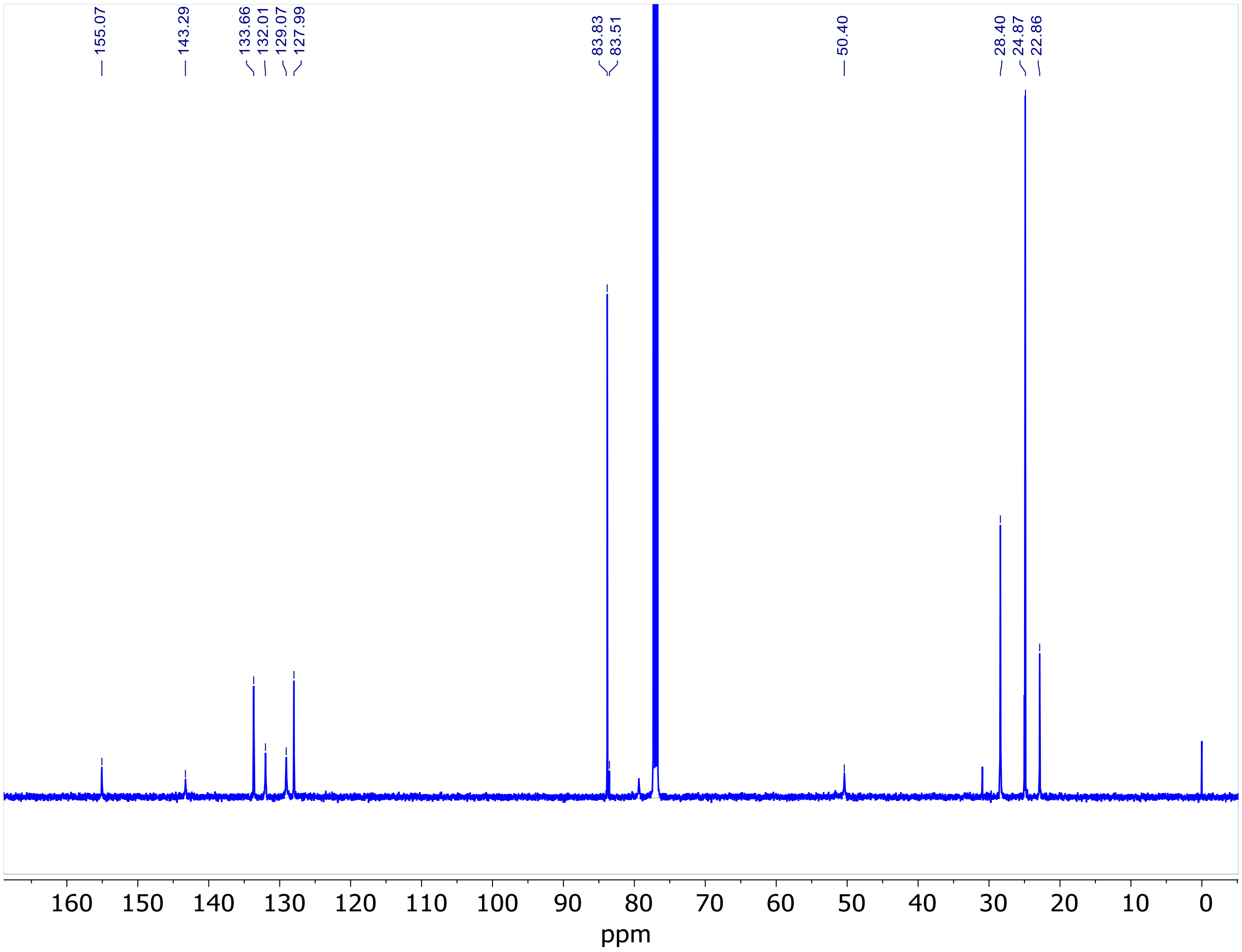


4 - tert-butyl (R)-(1-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)ethyl)carbamate

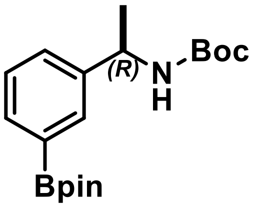


4

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.73 (s, 1H), 7.70 (d, *J* = 7.2 Hz, 1H), 7.39 (d, *J* = 7.8 Hz, 1H), 7.33 (t, *J* = 7.5 Hz, 1H), 4.80 (s, 2H), 1.45 (d, *J* = 6.5 Hz, 3H), 1.42 (s, 9H), 1.34 (s, 112H).

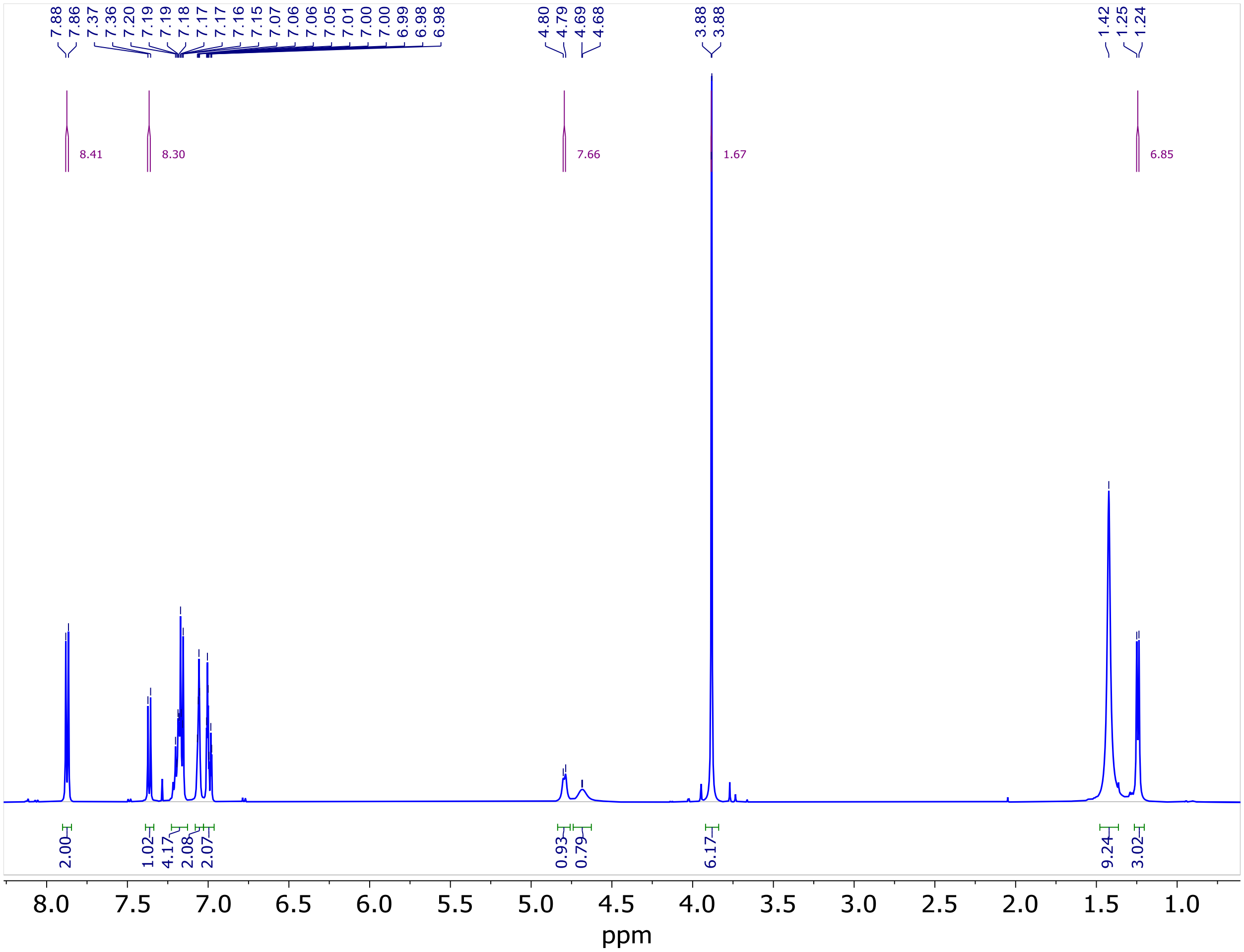


4 - tert-butyl (R)-(1-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)ethyl)carbamate

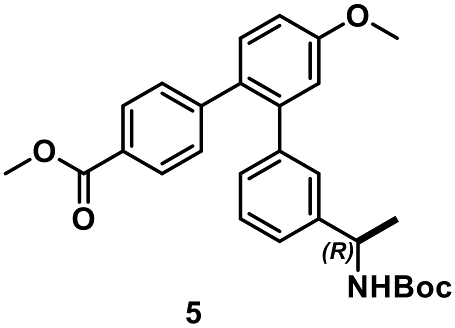


4

^{13}C NMR (126 MHz, CDCl$_3$) δ 155.07, 143.29, 133.66, 132.01, 129.07, 127.99, 83.83, 83.51, 50.40, 28.40, 24.87, 22.86.

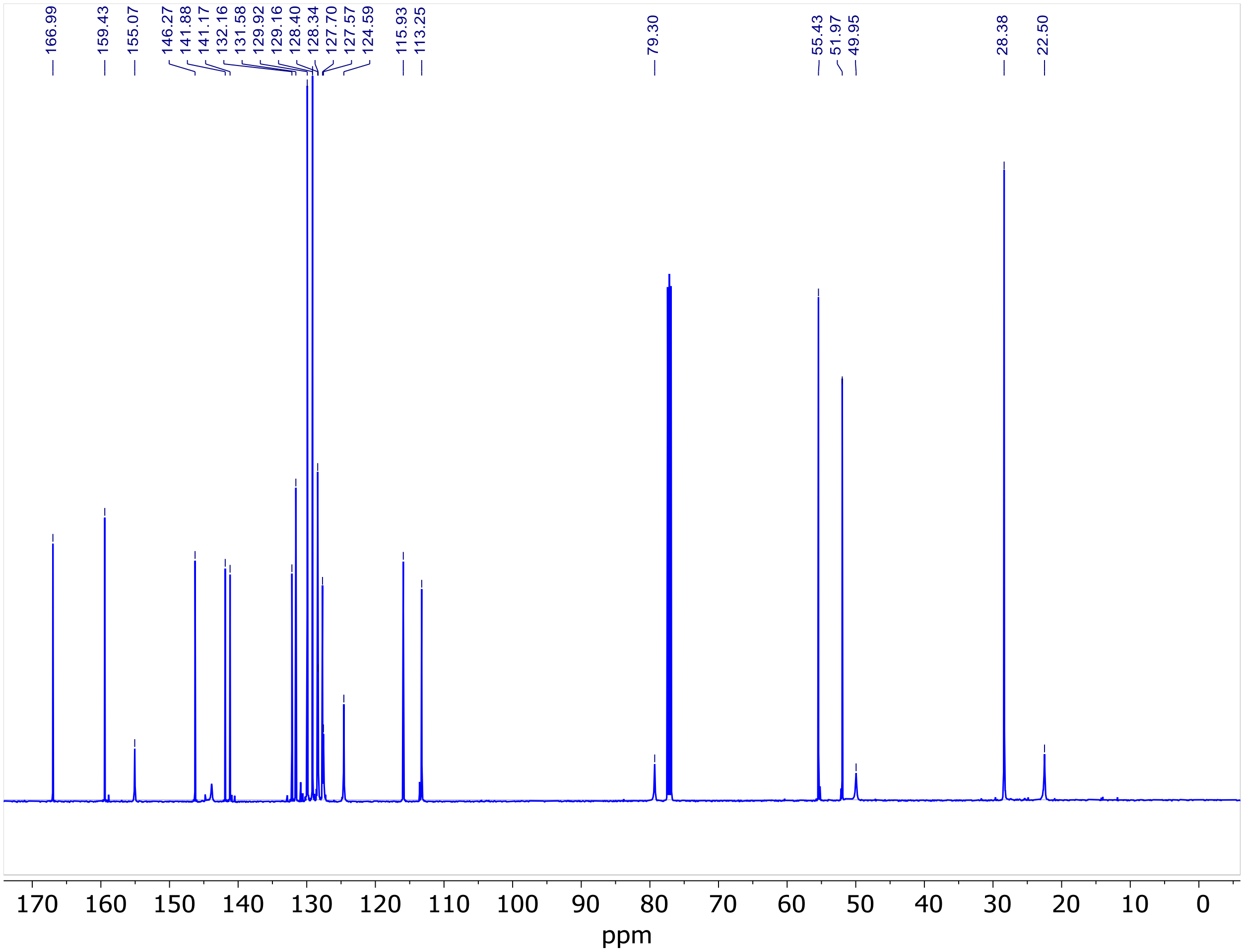


5 – *mop*-*R*-Tph-1

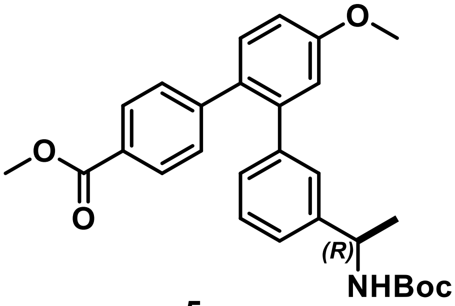


5

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.87 (d, *J* = 8.4 Hz, 1H), 7.36 (d, *J* = 8.3 Hz, 1H), 7.23 – 7.14 (m, 2H), 7.08 – 7.03 (m, 1H), 7.03 – 6.96 (m, 1H), 4.79 (d, *J* = 7.7 Hz, 1H), 4.73 – 4.63 (m, 1H), 3.88 (d, 3H), 1.42 (s, 6H), 1.24 (d, 2H).

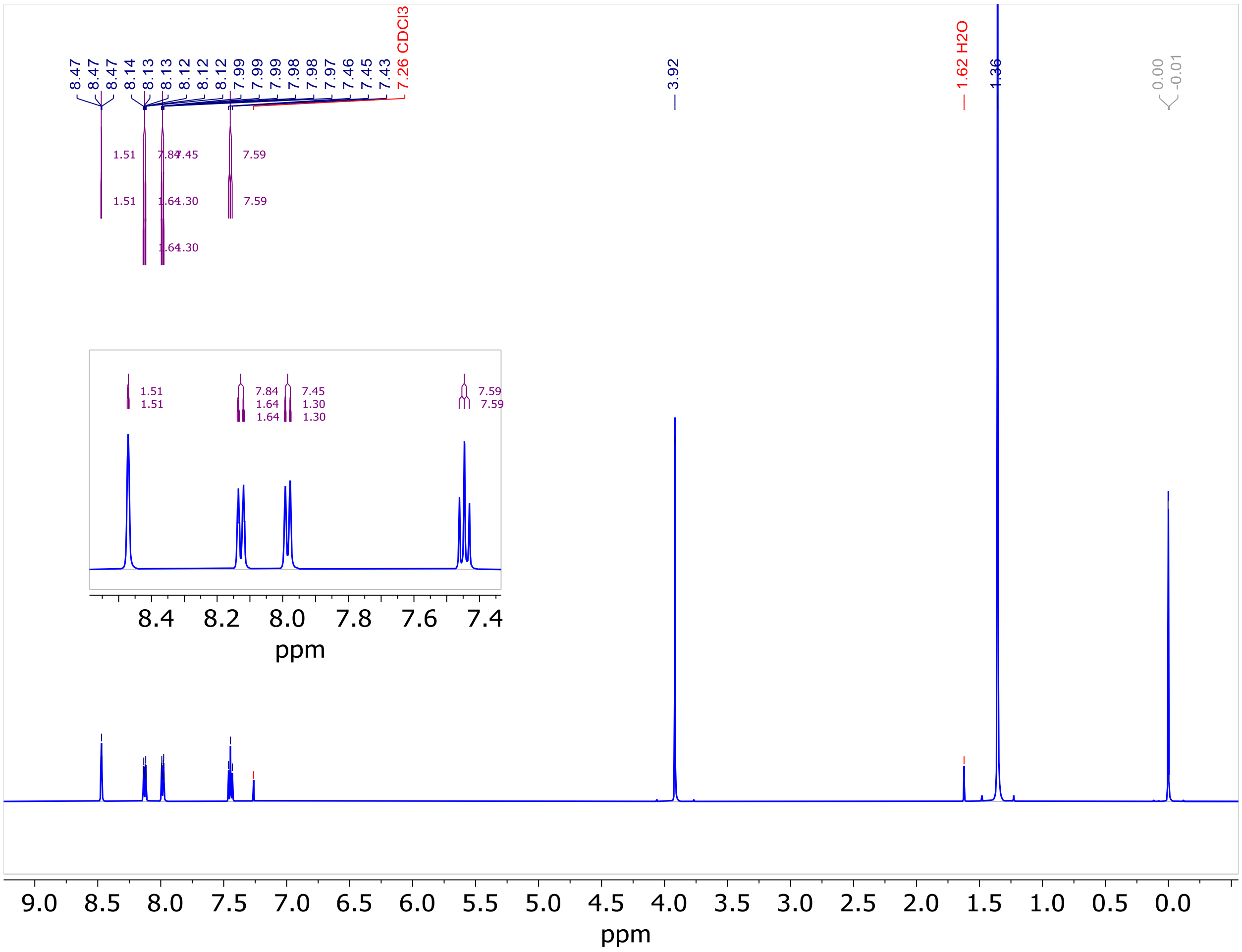


5 – *mop*-*R*-Tph-1

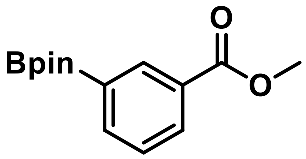


5

^{13}C NMR (126 MHz, CDCl$_3$) δ 166.99, 159.43, 155.07, 146.27, 141.88, 141.17, 132.16, 131.58, 129.92, 129.16, 128.40, 128.34, 127.70, 127.57, 124.59, 115.93, 113.25, 79.30, 55.43, 51.97, 49.95, 28.38, 22.50.

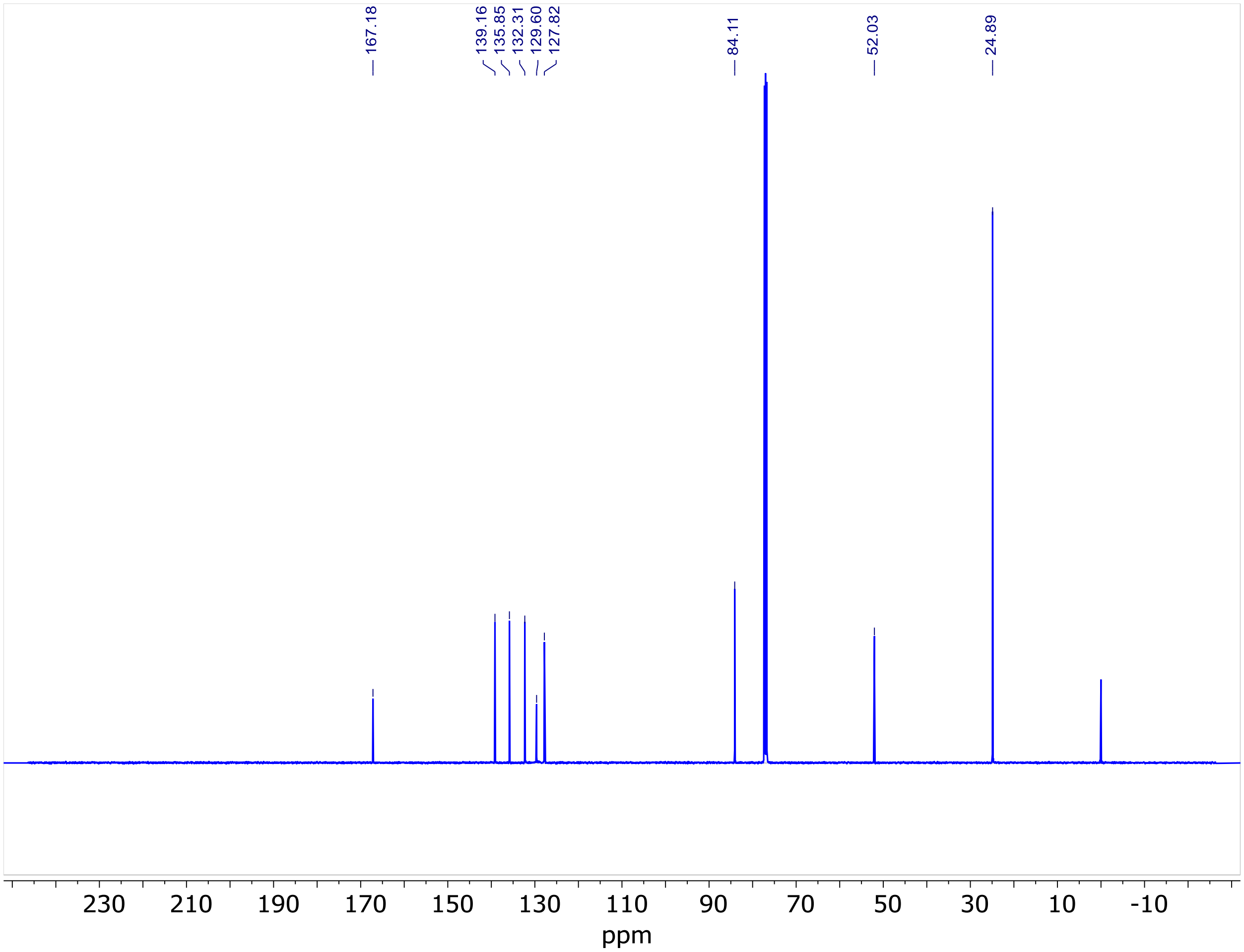


6b - Methyl 3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzoate

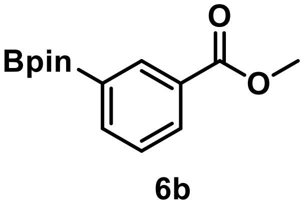


6b

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.47 (t, *J* = 1.5 Hz, 1H), 8.13 (dt, *J* = 7.8, 1.6 Hz, 1H), 7.98 (dt, *J* = 7.5, 1.3 Hz, 1H), 7.45 (t, *J* = 7.6 Hz, 1H), 3.92 (s, 3H), 1.36 (s, 12H).

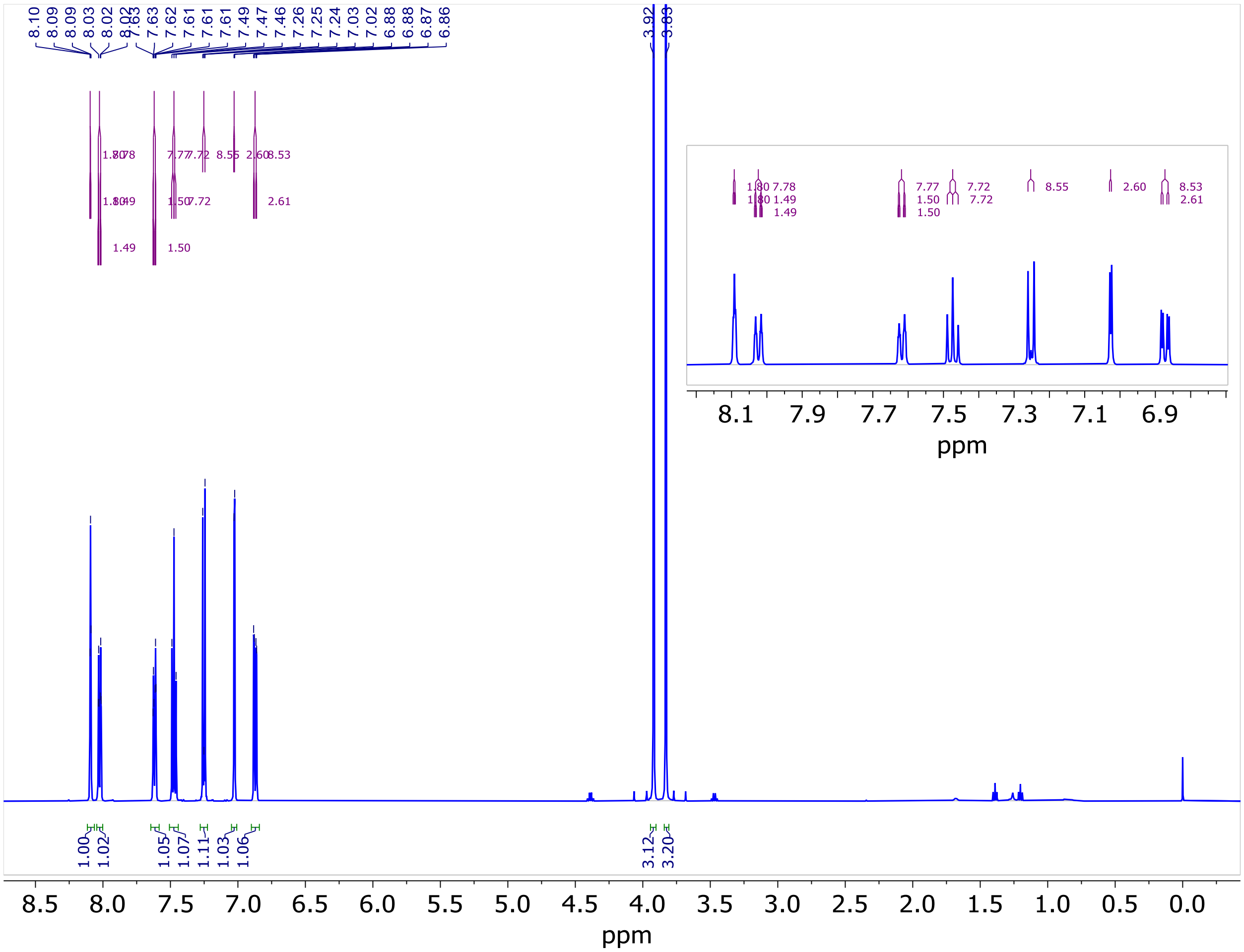


6b - Methyl 3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzoate

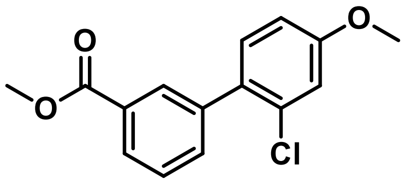


6b

^{13}C NMR (126 MHz, $CDCl_3$) δ 167.18, 139.16, 135.85, 132.31, 129.60, 127.82, 84.11, 52.03, 24.89.

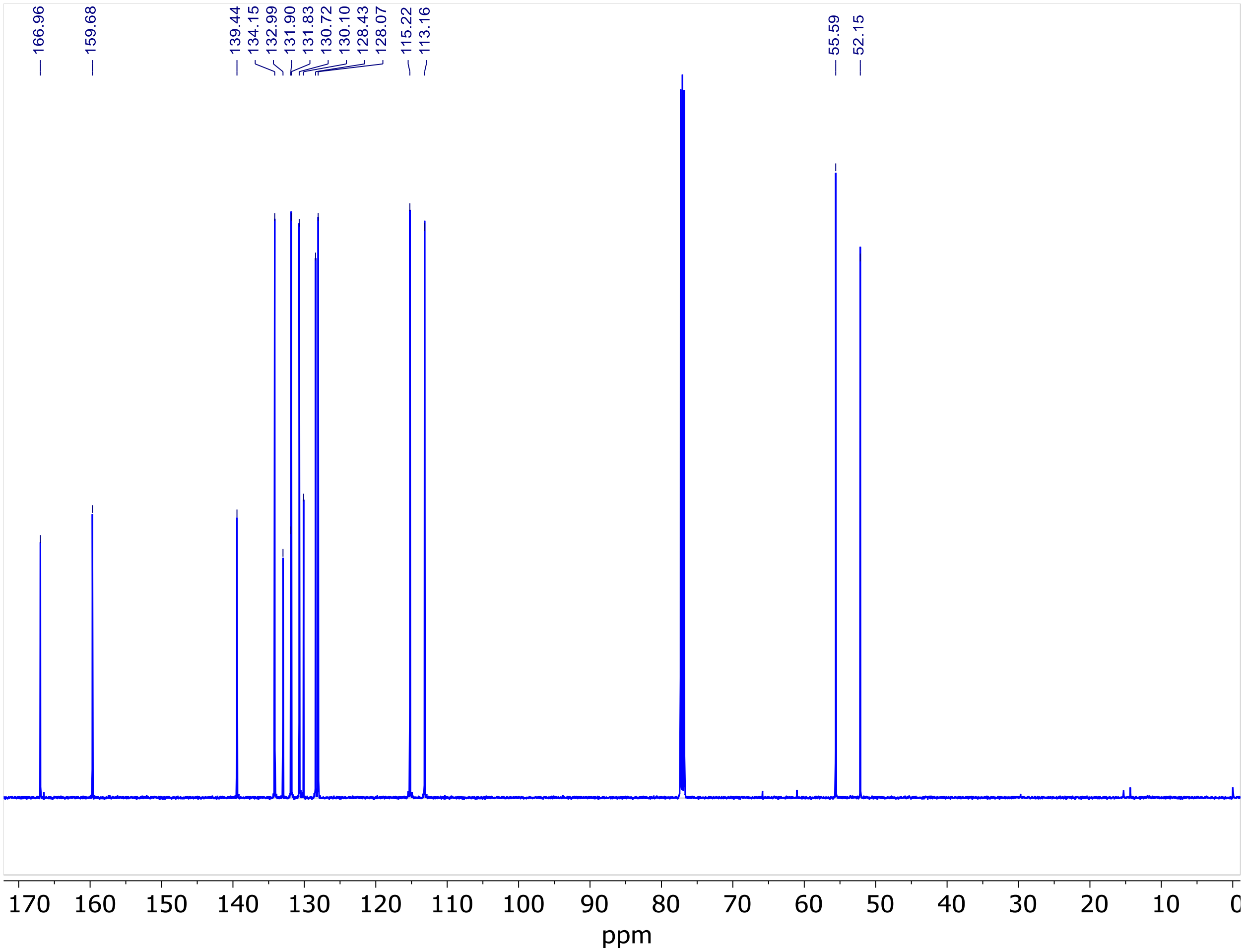


7 - methyl 2'-chloro-4'-methoxy-[1,1'-biphenyl]-3-carboxylate

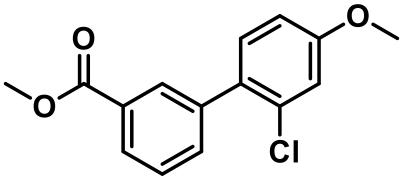


7

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.09 (t, *J* = 1.8 Hz, 1H), 8.02 (dt, *J* = 7.8, 1.5 Hz, 1H), 7.62 (dt, *J* = 7.8, 1.5 Hz, 1H), 7.47 (t, *J* = 7.7 Hz, 1H), 7.25 (d, *J* = 8.5 Hz, 1H), 7.03 (d, *J* = 2.6 Hz, 1H), 6.87 (dd, *J* = 8.5, 2.6 Hz, 1H), 3.92 (s, 3H), 3.83 (s, 3H).

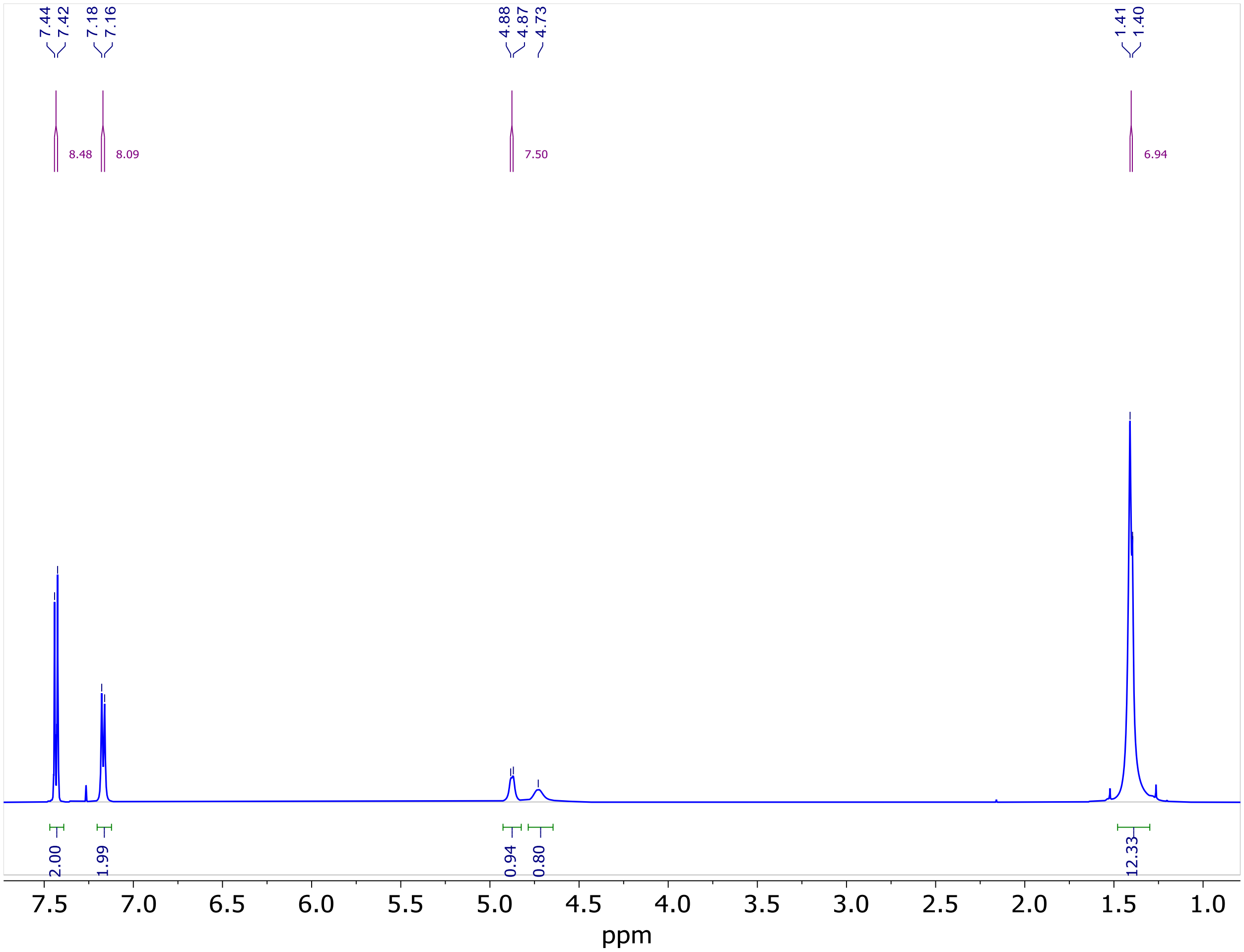


7 - methyl 2'-chloro-4'-methoxy-[1,1'-biphenyl]-3-carboxylate

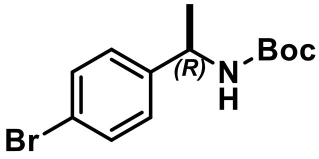


7

^{13}C NMR (126 MHz, CDCl$_3$) δ 166.96, 159.68, 139.44, 134.15, 132.99, 131.90, 131.83, 130.72, 130.10, 128.43, 128.07, 115.22, 113.16, 55.59, 52.15.

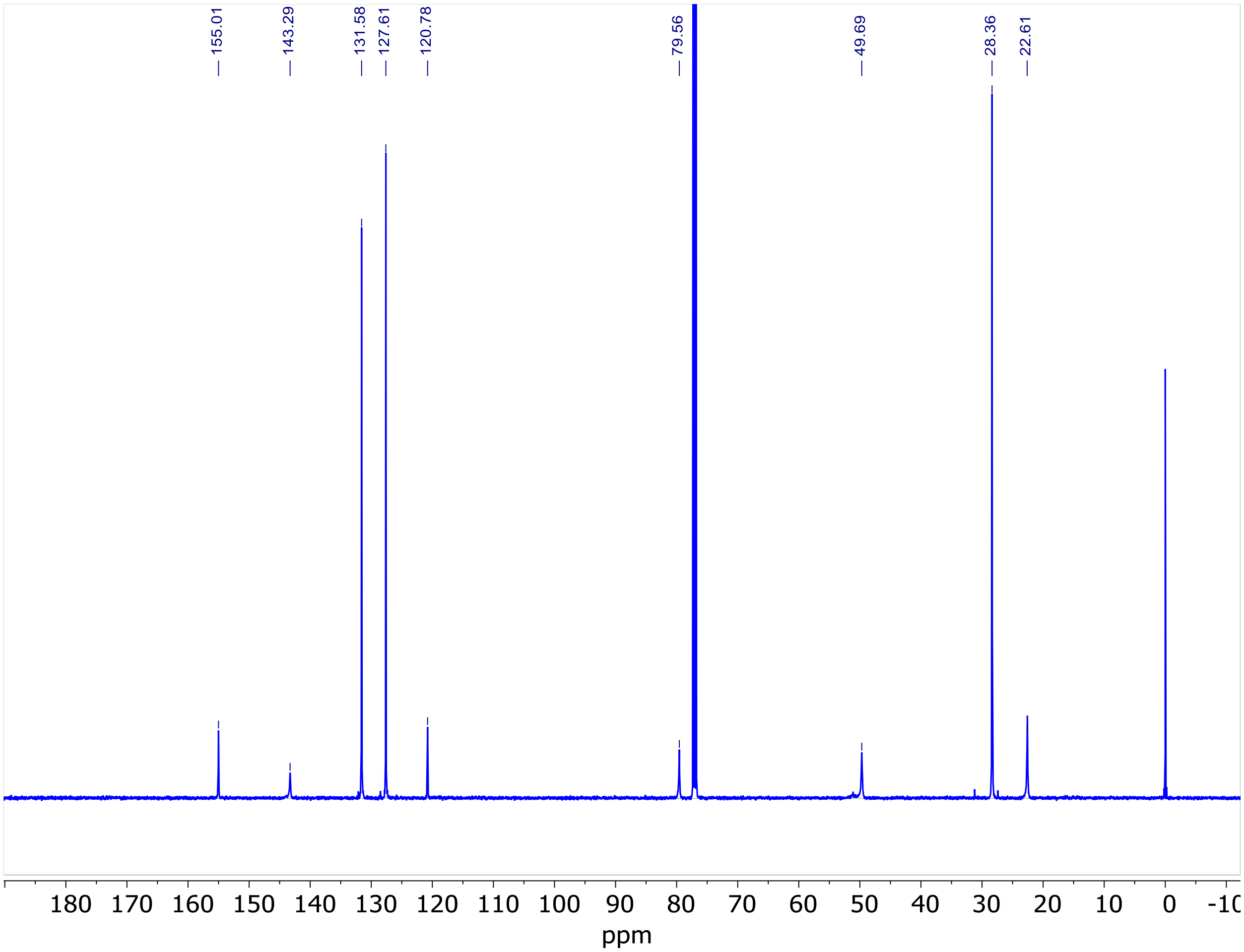


8b - tert-butyl (*R*)-(1-(4-bromophenyl)ethyl)carbamate

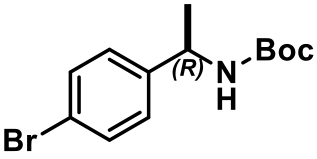


8b

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.43 (d, *J* = 8.5 Hz, 1H), 7.17 (d, *J* = 8.1 Hz, 1H), 4.88 (s, 1H), 4.73 (s, 1H), 1.40 (d, 9H).

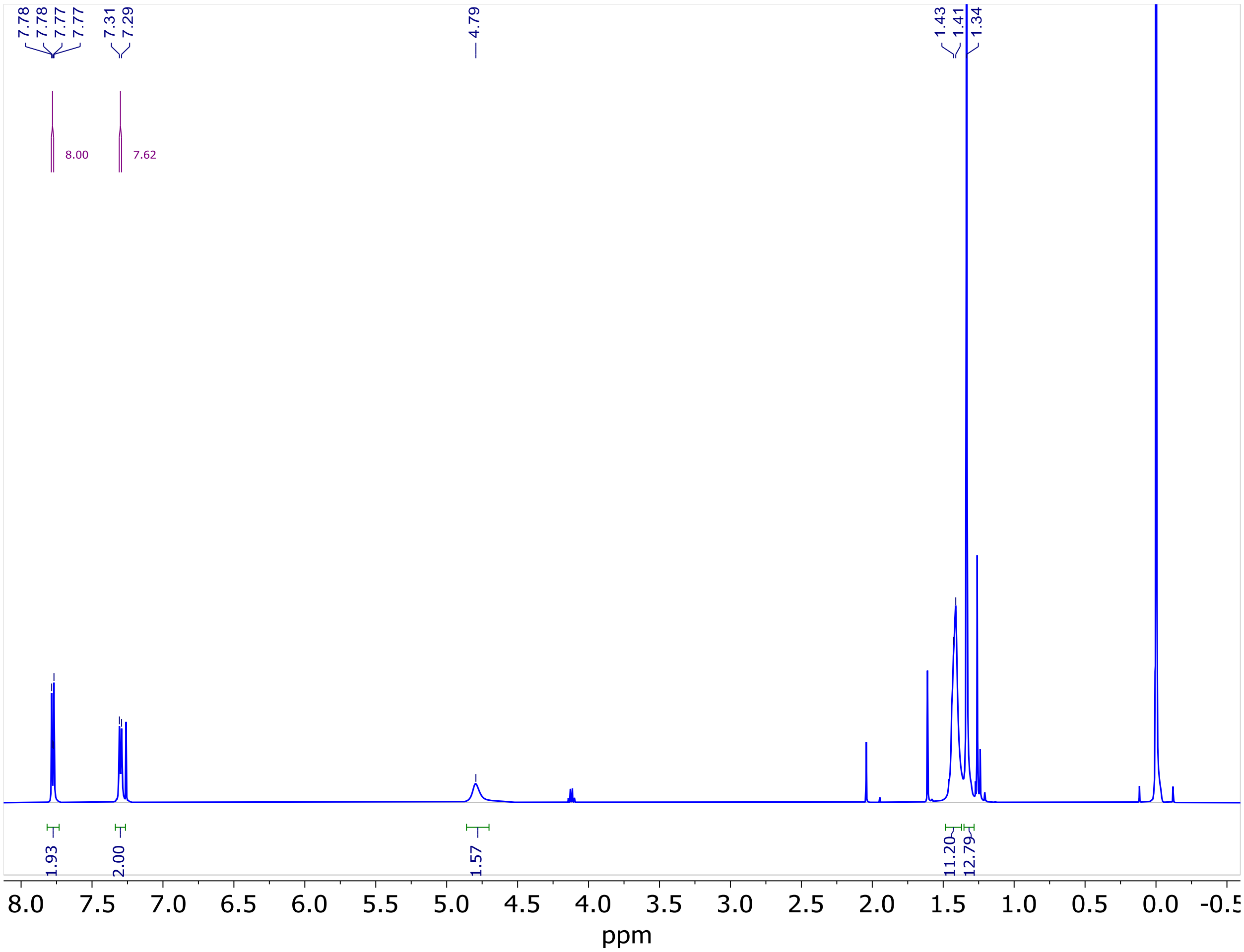


8b - tert-butyl (*R*)-(1-(4-bromophenyl)ethyl)carbamate

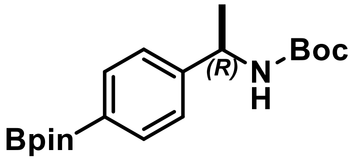


8b

^{13}C NMR (126 MHz, CDCl$_3$) δ 155.01, 143.29, 131.58, 127.61, 120.78, 79.56, 49.69, 28.36, 22.61.

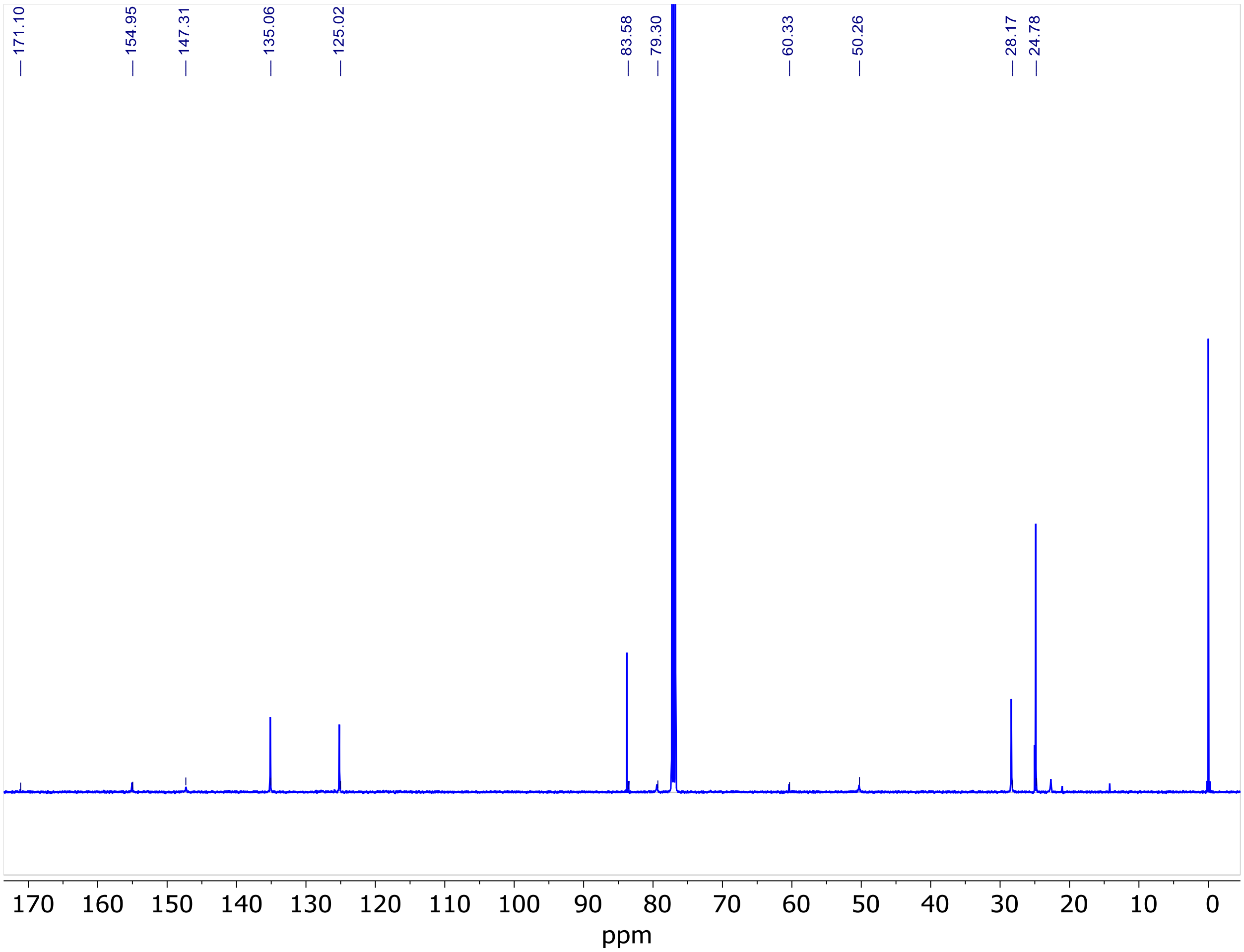


9 - tert-butyl (R)-(1-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)ethyl)carbamate

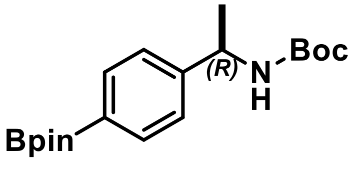


9

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.77 (d, *J* = 8.0 Hz, 2H), 7.30 (d, *J* = 7.6 Hz, 2H), 4.79 (s, 2H), 1.42 (d, 12H), 1.34 (s, 12H).

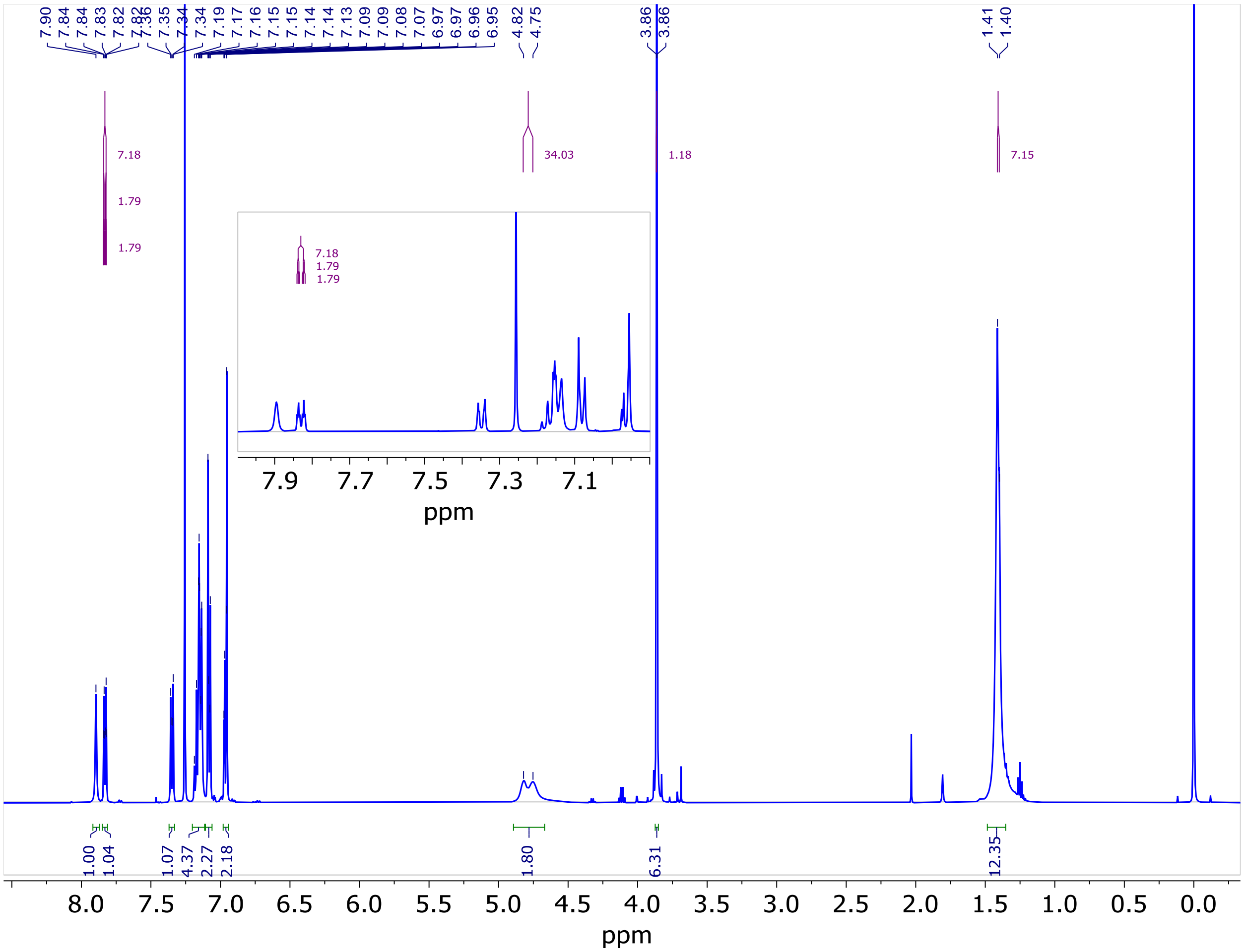


9 - tert-butyl (R)-(1-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)ethyl)carbamate

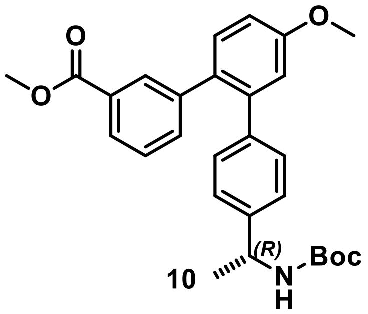


9

^{13}C NMR (126 MHz, CDCl$_3$) δ 171.10, 154.95, 147.31, 135.06, 125.02, 83.58, 79.30, 60.33, 50.26, 28.17, 24.78.

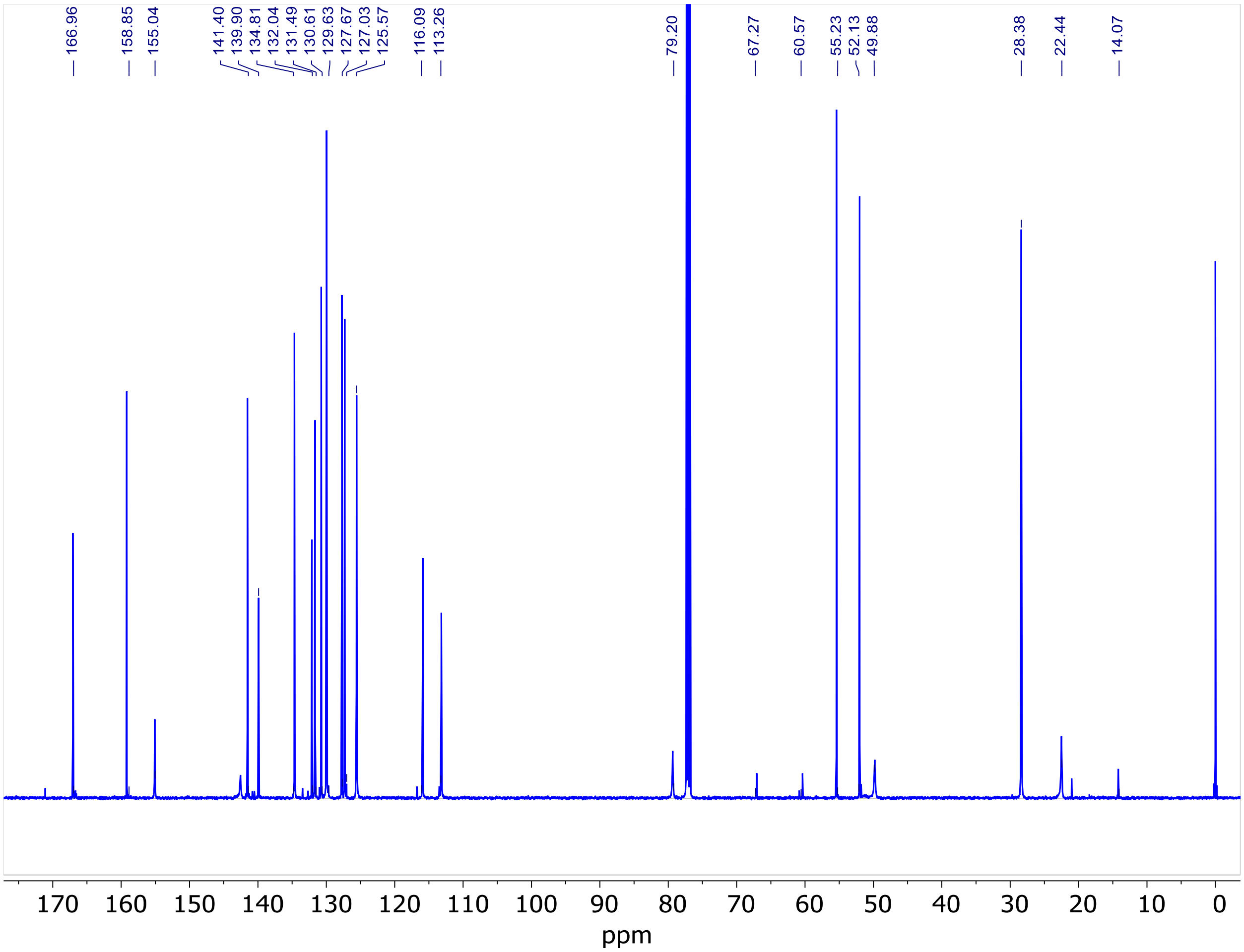


10 - methyl (R)-4''-(1-((tert-butoxycarbonyl)amino)ethyl)-4'-methoxy-[1,1':2',1''-terphenyl]-3-carboxylate or ***pom-R*-Tph-1**

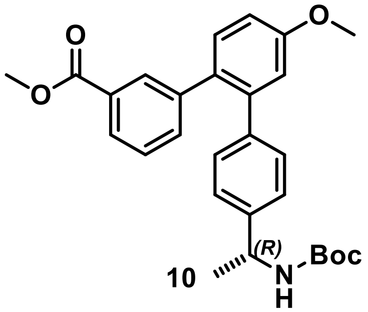


10

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.90 (s, 1H), 7.83 (dt, *J* = 7.2, 1.8 Hz, 1H), 7.37 – 7.33 (m, 1H), 7.20 – 7.11 (m, 4H), 7.11 – 7.06 (m, 2H), 6.98 – 6.94 (m, 2H), 4.79 (d, 2H), 3.86 (d, 6H), 1.41 (d, 13H).

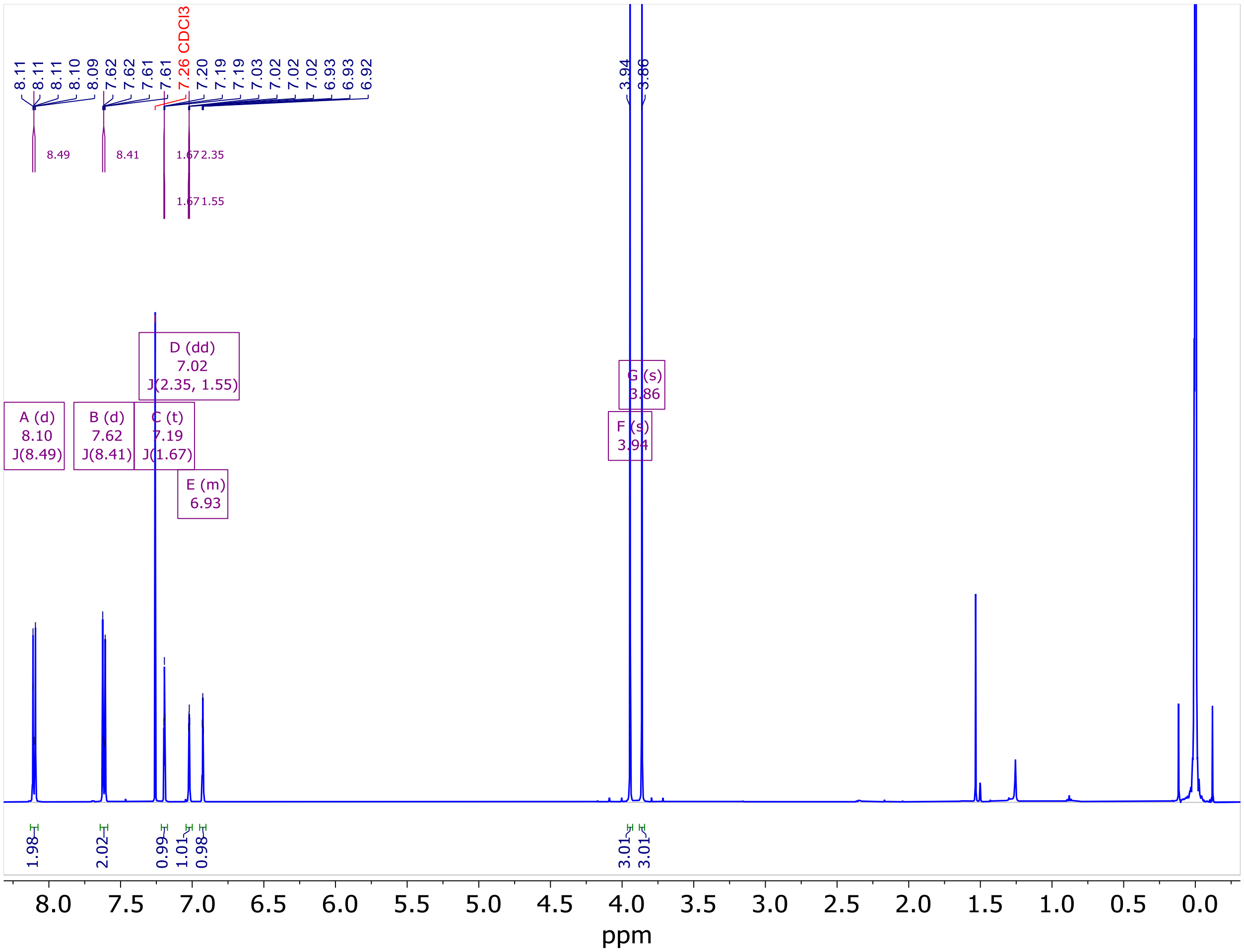


10 - methyl (R)-4''-(1-((tert-butoxycarbonyl)amino)ethyl)-4'-methoxy-[1,1':2',1''-terphenyl]-3-carboxylate or ***pom-R*-Tph-1**

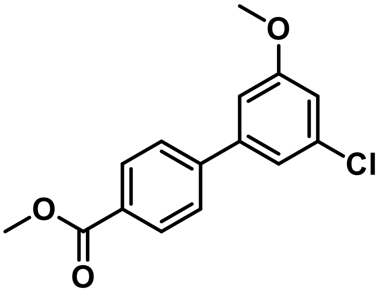


10

^{13}C NMR (126 MHz, CDCl$_3$) δ 166.96, 158.85, 155.04, 141.40, 139.90, 134.81, 132.04, 131.49, 130.61, 129.63, 127.67, 127.03, 125.57, 116.09, 113.26, 79.20, 67.27, 60.57, 55.23, 52.13, 49.88, 28.38, 22.44, 14.07.



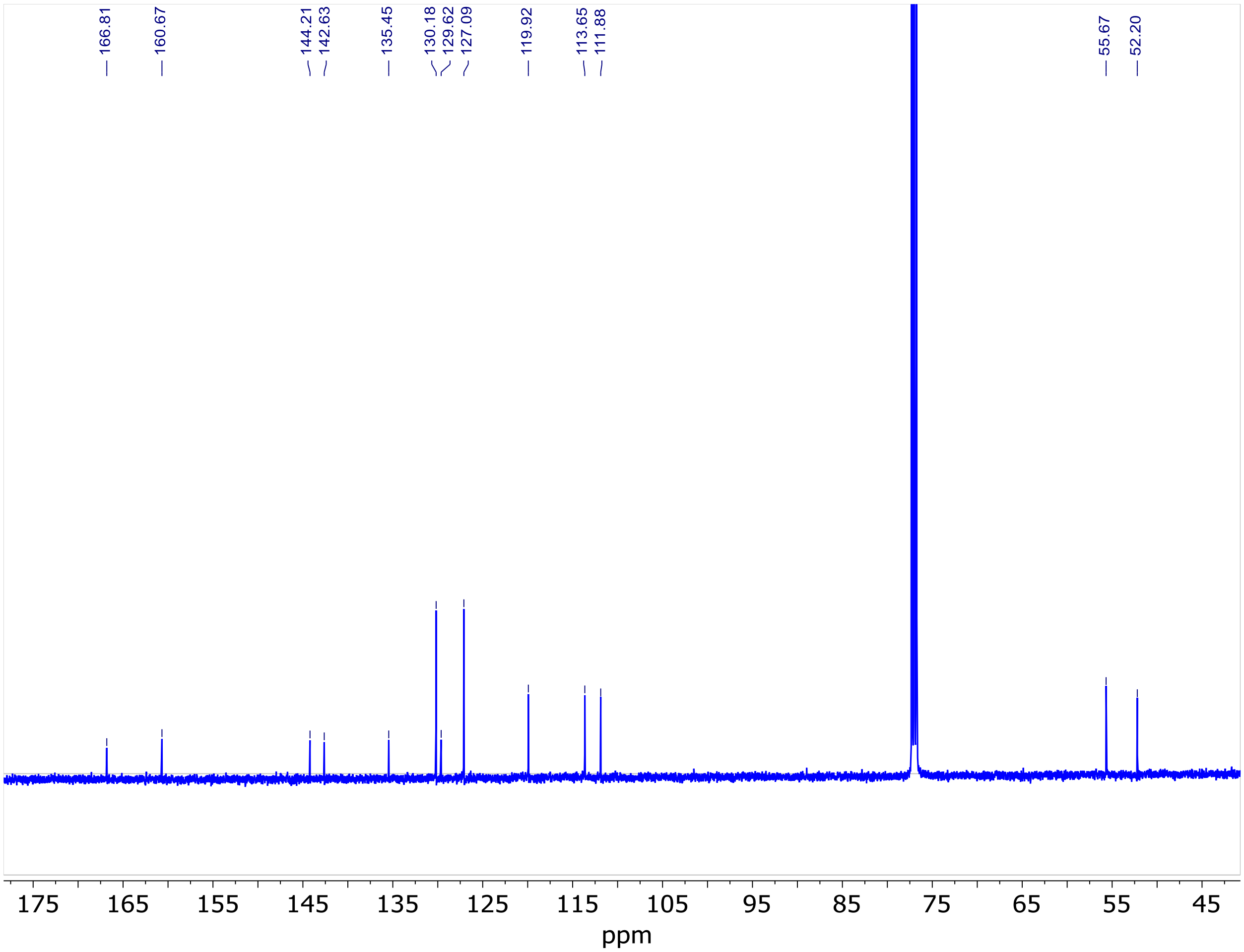
11 - methyl 3'-chloro-5'-methoxy-[1,1'-biphenyl]-4-carboxylate



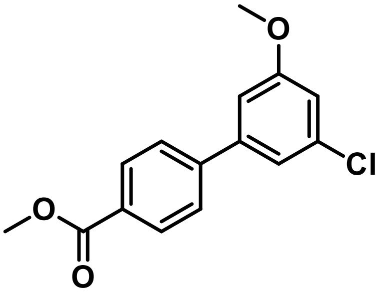
11

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.10 (d, *J* = 8.5 Hz, 1H), 7.62 (d, *J* = 8.4 Hz, 1H), 7.19 (t, *J* = 1.7 Hz, 1H), 7.02 (dd, *J* = 2.3, 1.6 Hz, 1H), 6.93 – 6.92 (m, 1H), 3.94 (s, 2H), 3.86 (s, 2H).

Data acquired by Schimmelpfennig, L.



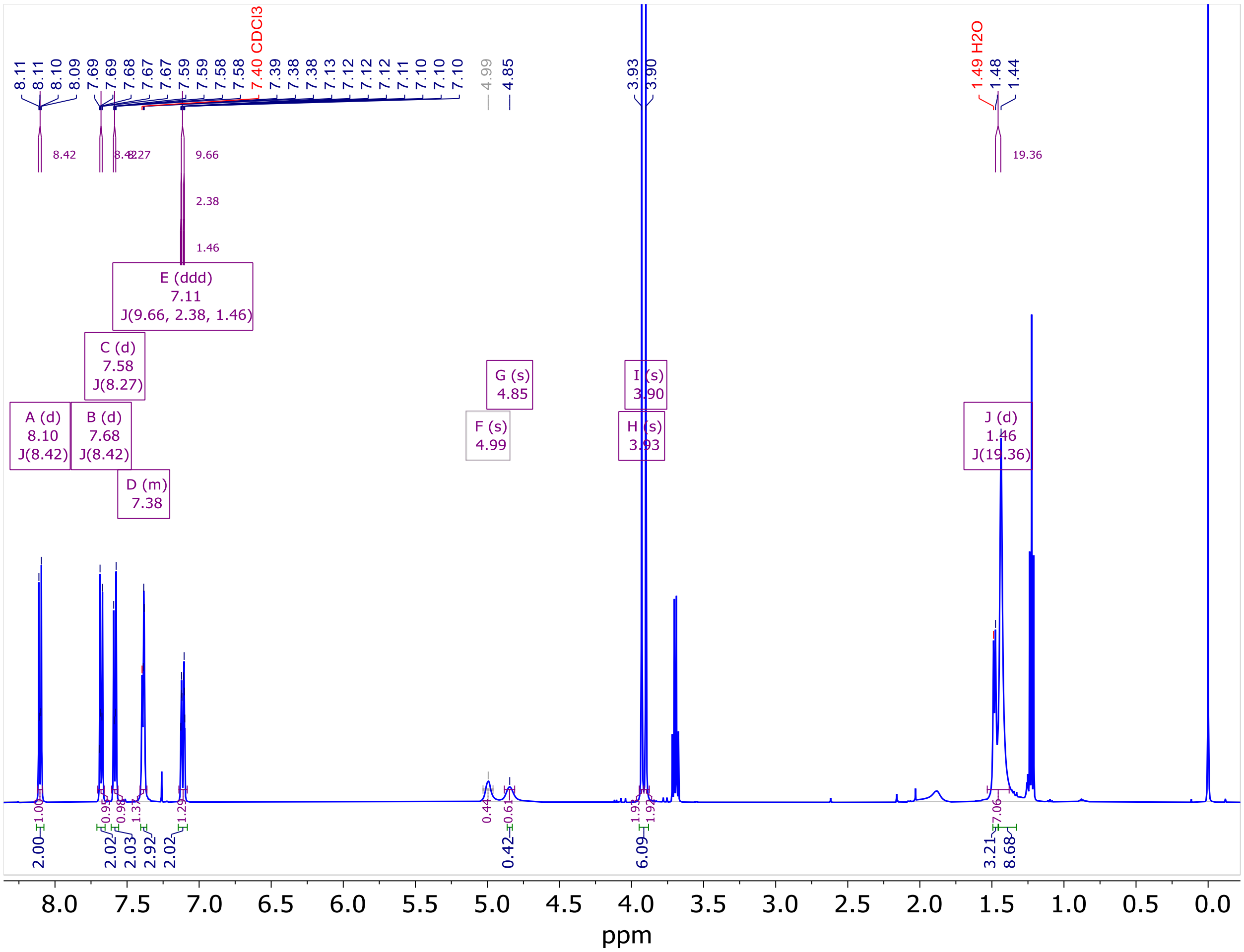
11 - methyl 3'-chloro-5'-methoxy-[1,1'-biphenyl]-4-carboxylate



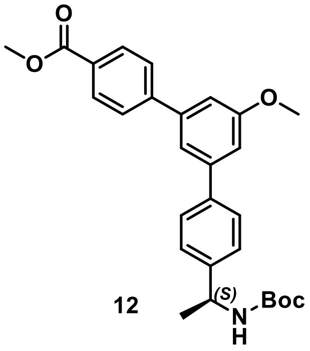
11

^{13}C NMR (126 MHz, $CDCl_3$) δ 166.81, 160.67, 144.21, 142.63, 135.45, 130.18, 129.62, 127.09, 119.92, 113.65, 111.88, 55.67, 52.20.

Data acquired by Schimmelpfennig, L.



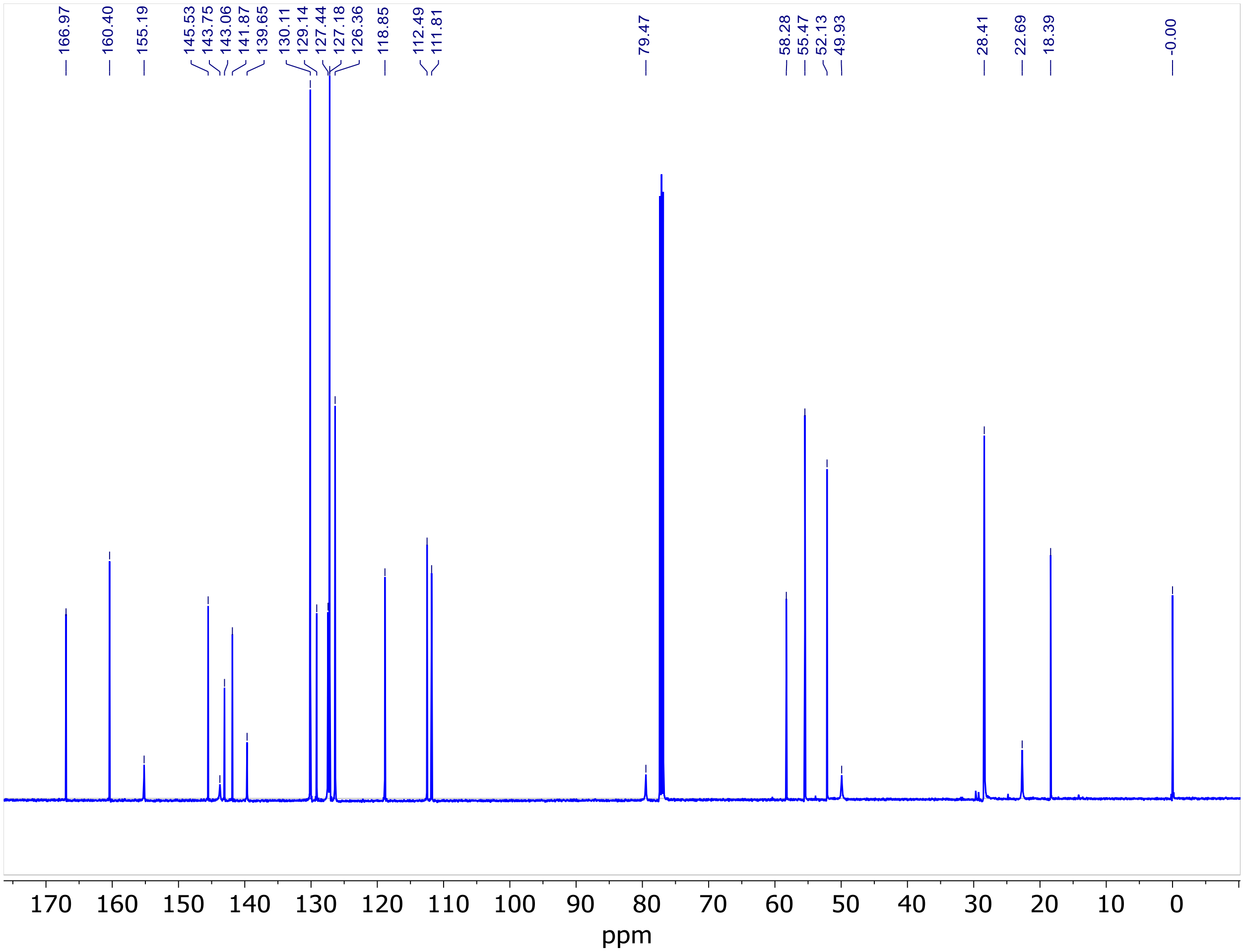
12 - methyl (S)-4''-(1-((tert-butoxycarbonyl)amino)ethyl)-5'-methoxy-[1,1':3',1''-terphenyl]-4-carboxylate or ***pmp-S*-Tph-1**



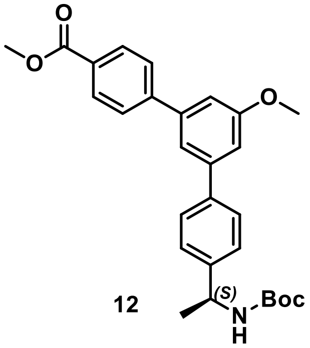
12

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.10 (d, *J* = 8.4 Hz, 2H), 7.68 (d, *J* = 8.4 Hz, 2H), 7.58 (d, *J* = 8.3 Hz, 2H), 7.41 – 7.36 (m, 3H), 7.11 (ddd, *J* = 9.7, 2.4, 1.5 Hz, 2H), 4.99 (s, 1H), 4.85 (s, 1H), 3.93 (s, 2H), 3.90 (s, 3H), 1.46 (d, 12 H).

Data acquired by Schimmelpfennig, L.



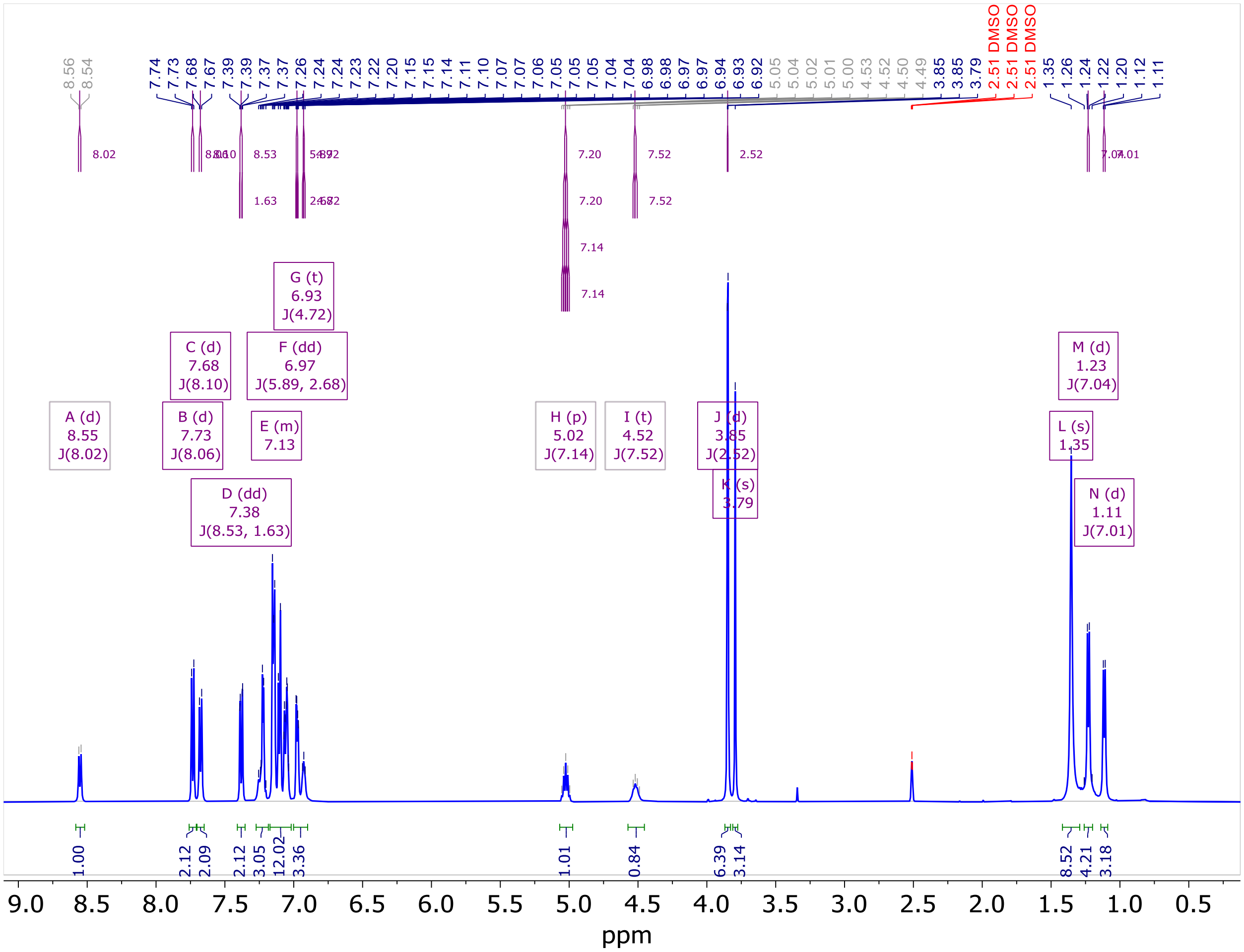
12 - methyl (S)-4''-(1-((tert-butoxycarbonyl)amino)ethyl)-5'-methoxy-[1,1':3',1''-terphenyl]-4-carboxylate or ***pmp-S*-Tph-1**



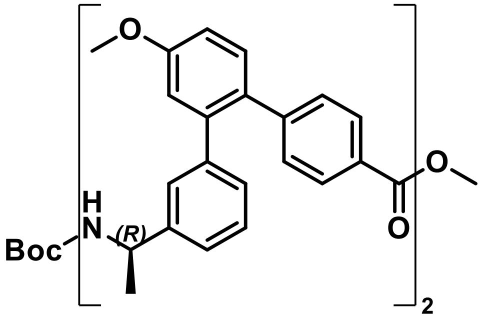
12

^{13}C NMR (126 MHz, CDCl$_3$) δ 166.97, 160.40, 155.19, 145.53, 143.75, 143.06, 141.87, 139.65, 130.11, 129.14, 127.44, 127.18, 126.36, 118.85, 112.49, 111.81, 79.47, 58.28, 55.47, 52.13, 49.93, 28.41, 22.69, 18.39.

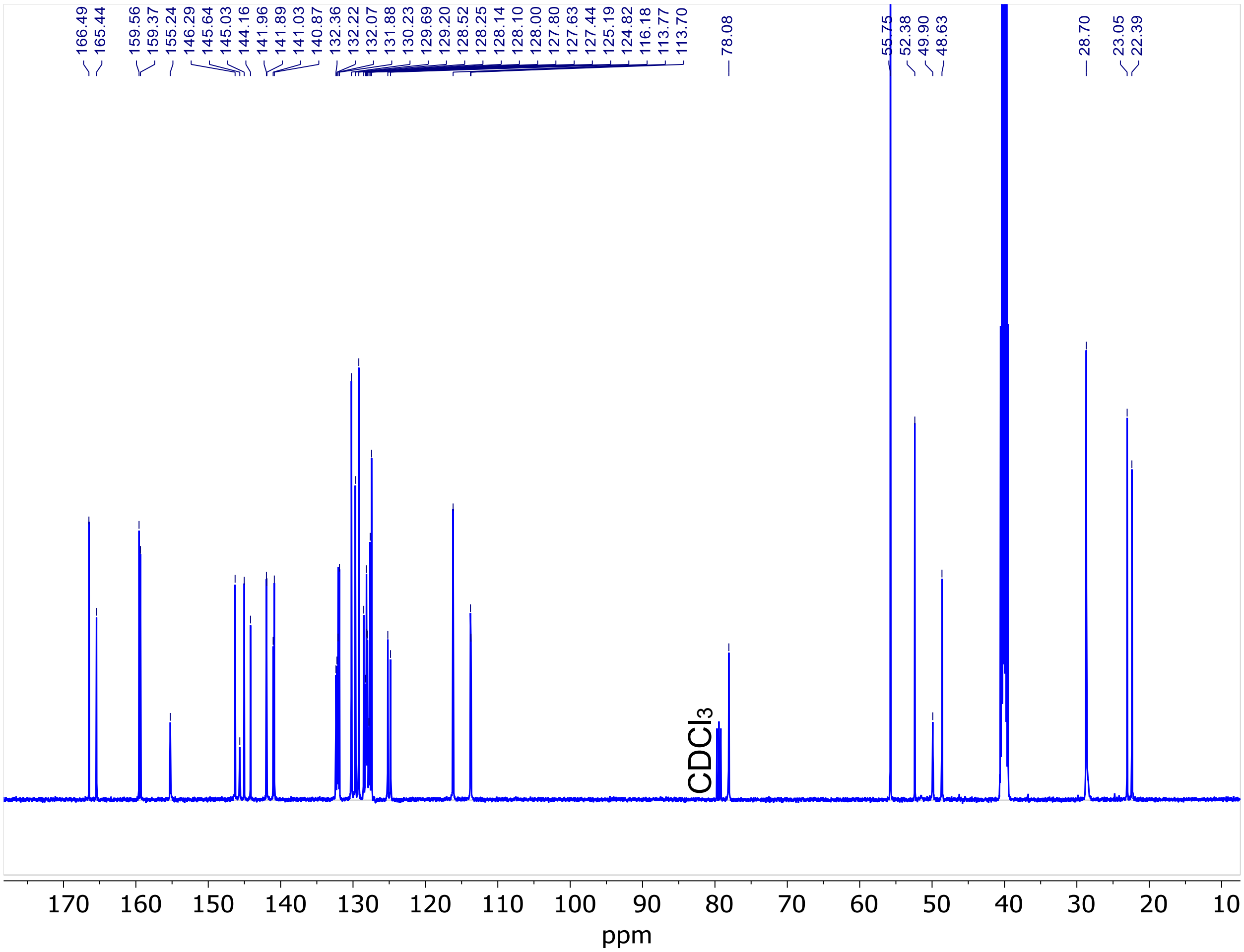
Data acquired by Schimmelpfennig, L.



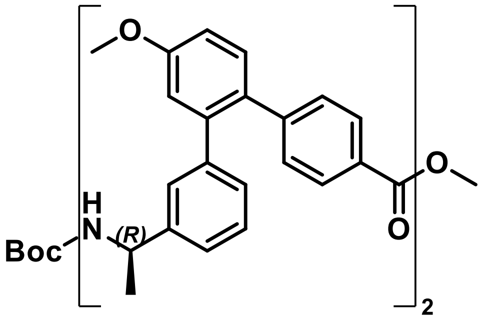
mop-*R*-Tph-2



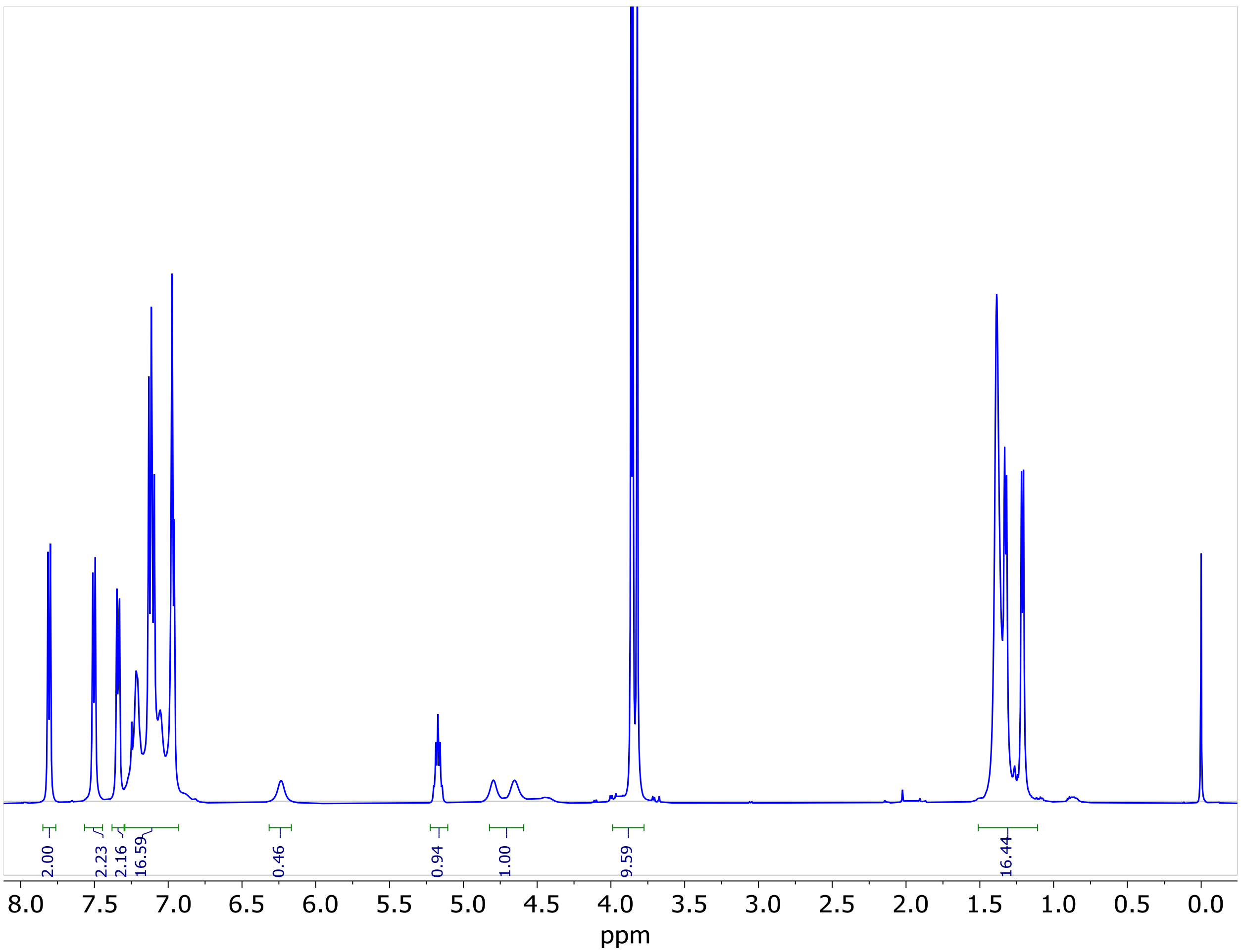
^{1}H NMR (500 MHz, DMSO) δ 8.56, 8.54, 7.74, 7.73, 7.68, 7.67, 7.39, 7.39, 7.37, 7.37, 7.26, 7.24, 7.24, 7.23, 7.22, 7.20, 7.15, 7.15, 7.14, 7.11, 7.10, 7.07, 7.07, 7.06, 7.05, 7.05, 7.05, 7.04, 7.04, 6.98, 6.98, 6.97, 6.97, 6.94, 6.93, 6.92, 5.05, 5.04, 5.02, 5.01, 5.00, 4.53, 4.52, 4.50, 4.49, 3.85, 3.85, 3.79, 2.51, 2.51, 2.51, 1.35, 1.26, 1.24, 1.22, 1.20, 1.12, 1.11.



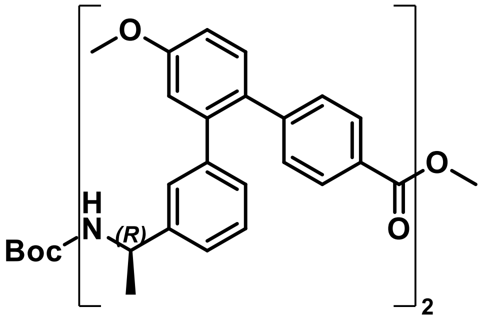
mop-*R*-Tph-2



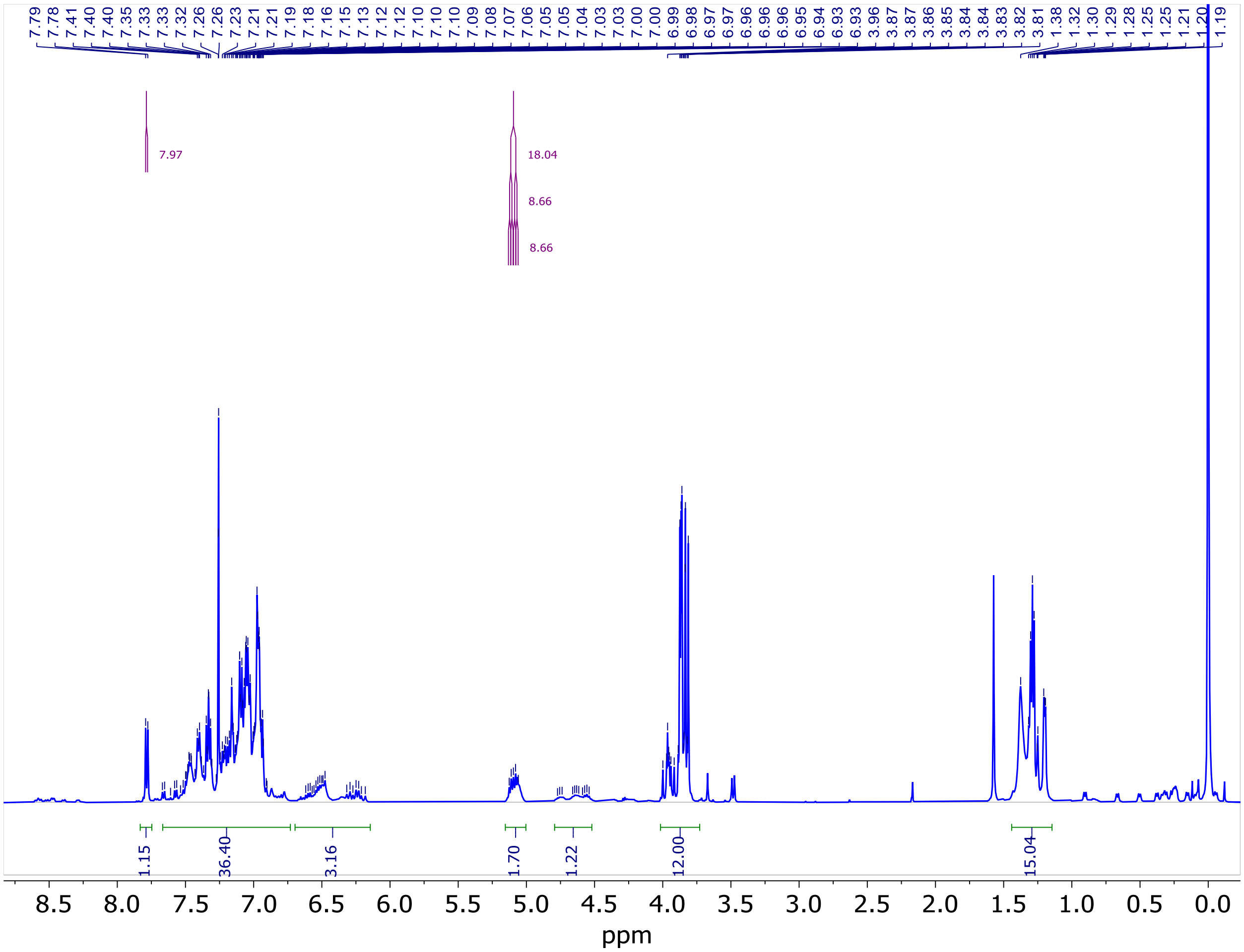
^{13}C NMR (126 MHz, DMSO) δ 166.49, 165.44, 159.56, 159.37, 155.24, 146.29, 145.64, 145.03, 144.16, 141.96, 141.89, 141.03, 140.87, 132.36, 132.22, 132.07, 131.88, 130.23, 129.69, 129.20, 128.52, 128.25, 128.14, 128.10, 128.00, 127.80, 127.63, 127.44, 125.19, 124.82, 116.18, 113.77, 113.70, 78.08, 55.75, 52.38, 49.90, 48.63, 28.70, 23.05, 22.39.



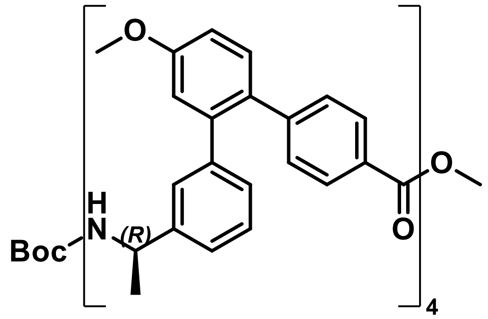
***mop-R*-Tph-2**



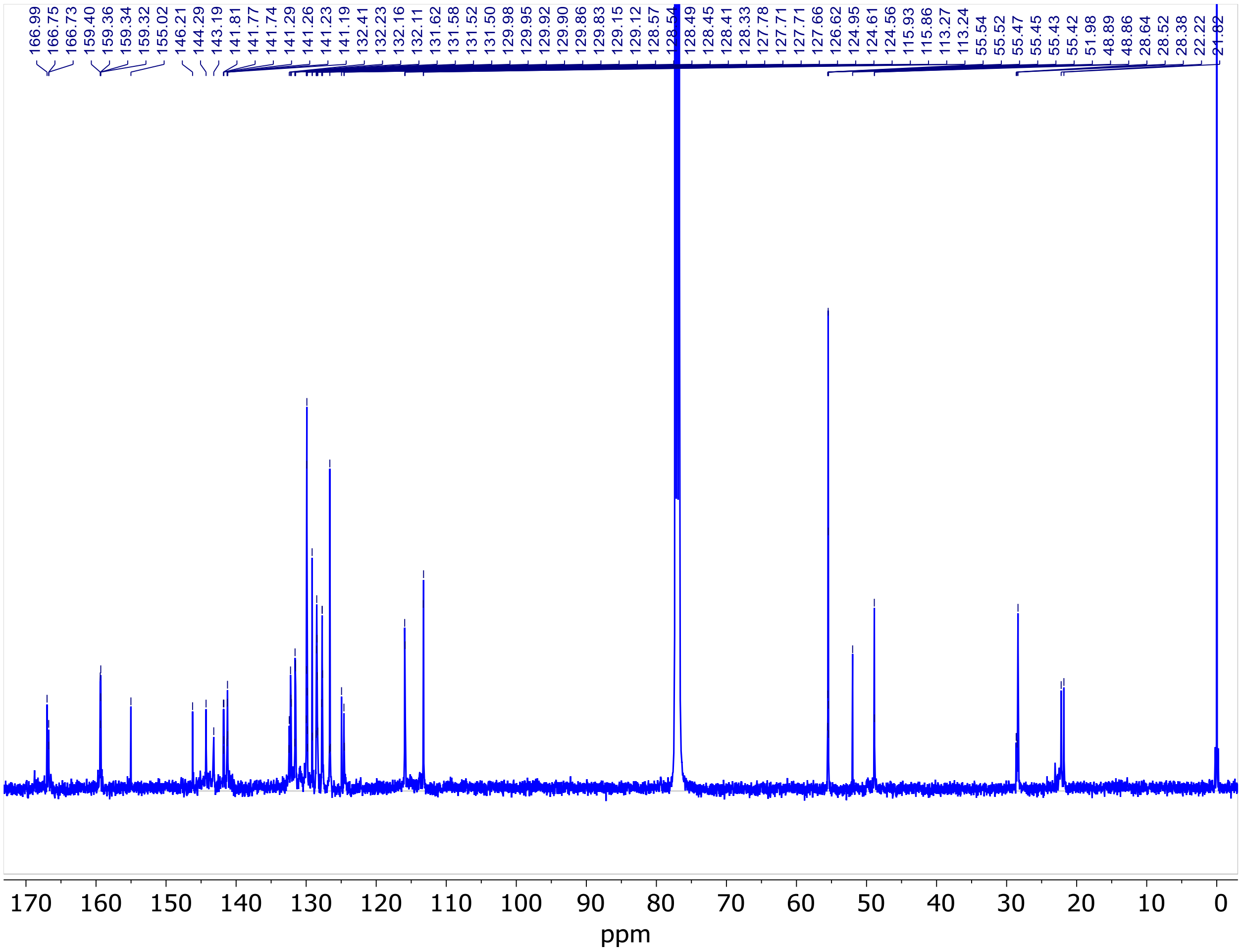
^{1}H NMR (500 MHz, CDCl$_3$)



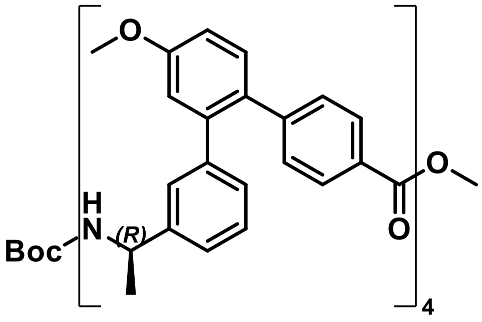
***mop-R*-Tph-4**



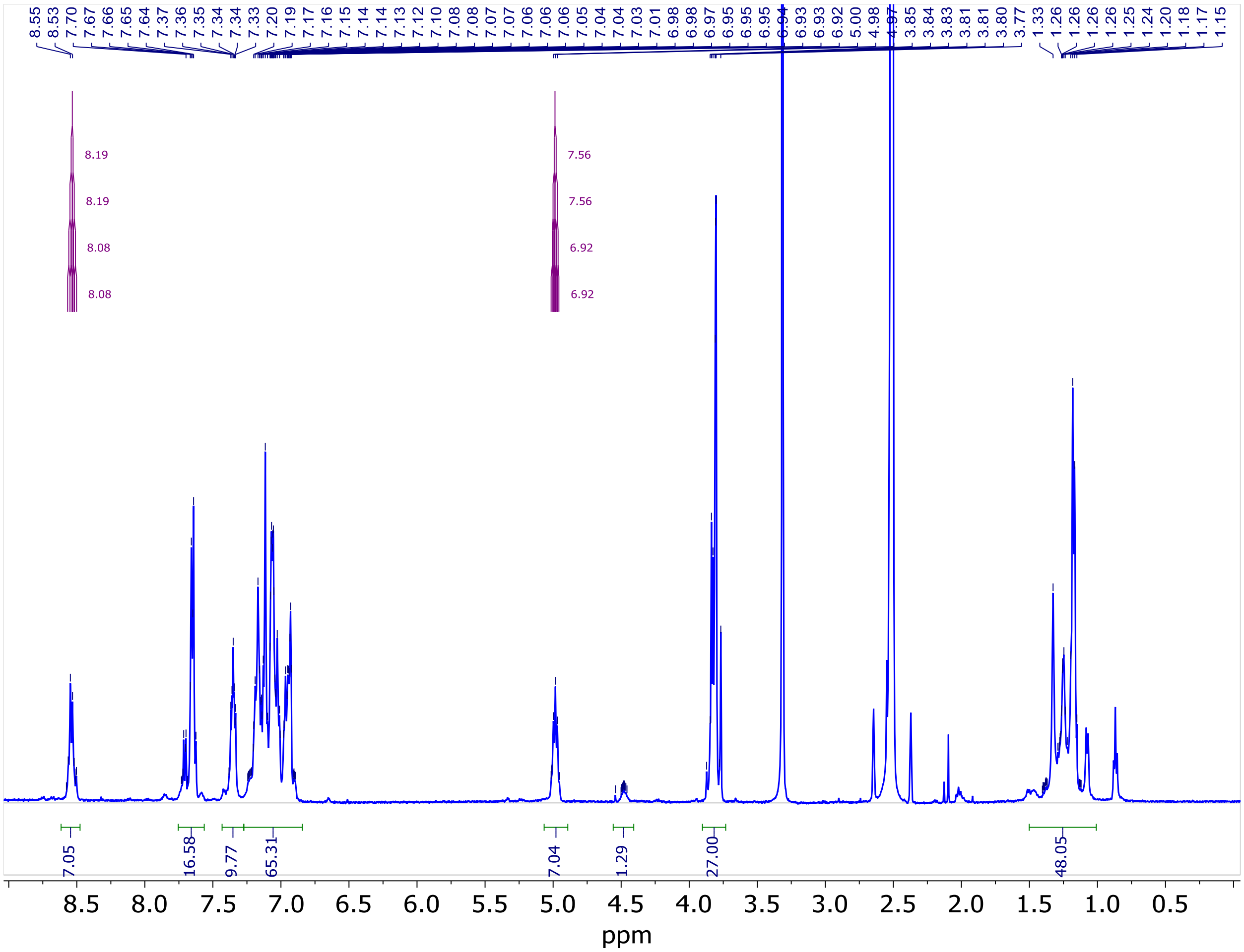
^{1}H NMR (500 MHz, CDCl$_3$) δ 7.78 (d, *J* = 8.0 Hz, 1H), 7.69 – 6.89 (m, 31H), 6.64 – 6.16 (m, 2H), 5.09 (dt, *J* = 18.0, 8.7 Hz, 1H), 4.85 – 4.50 (m, 0H), 4.05 – 3.74 (m, 10H), 1.48 – 1.15 (m, 11H).



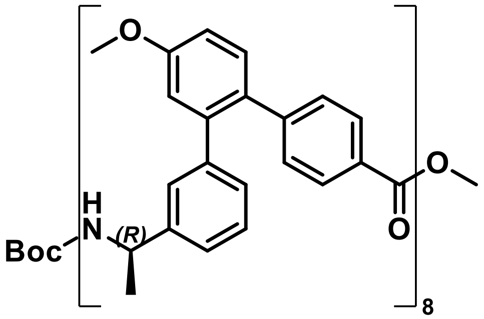
***mop-R*-Tph-4**



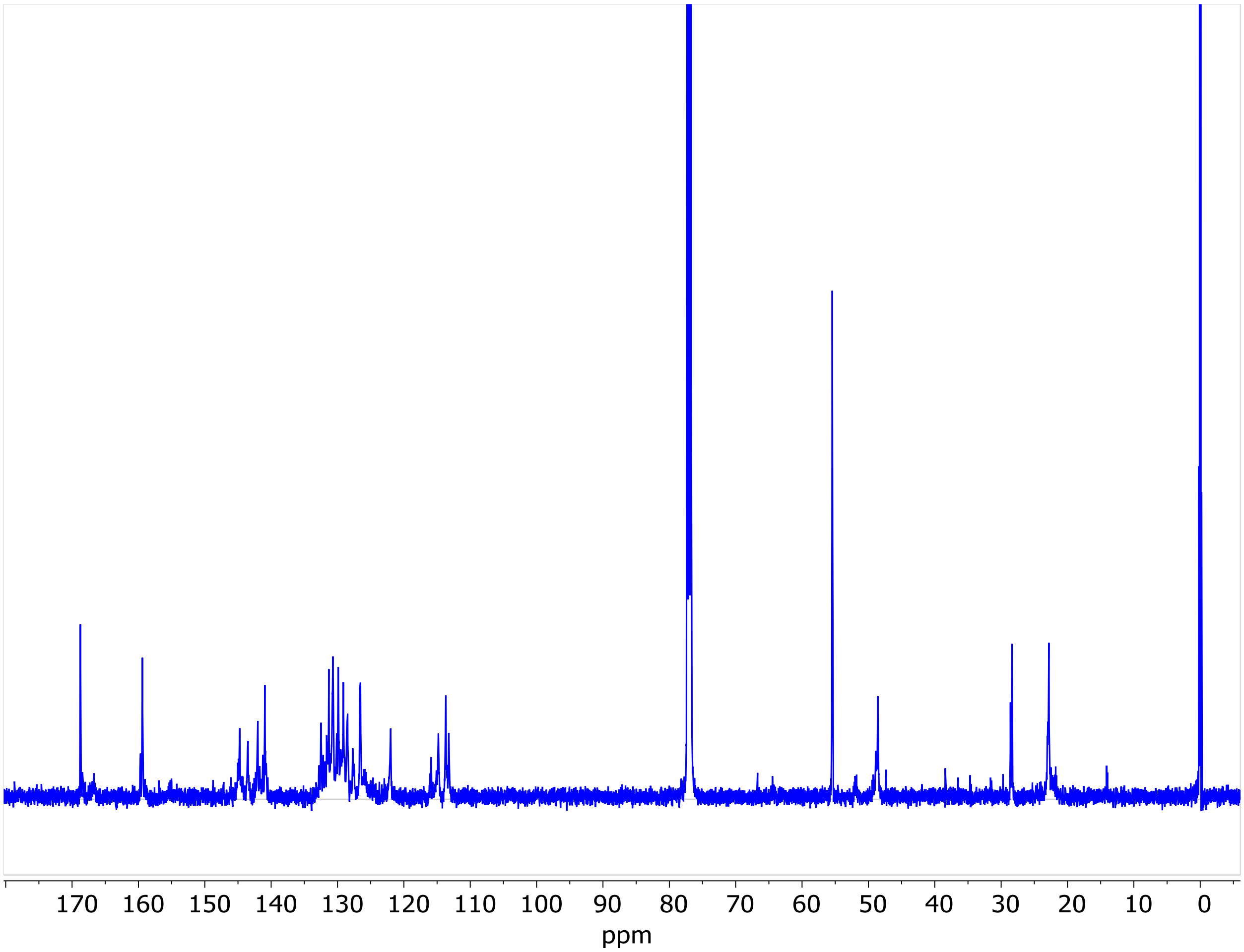
^{13}C NMR (126 MHz, CDCl$_3$) δ 166.99, 166.75, 166.73, 159.40, 159.36, 159.34, 159.32, 155.02, 146.21, 144.29, 143.19, 141.81, 141.79, 141.77, 141.74, 141.29, 141.26, 141.23, 141.19, 132.45, 132.41, 132.23, 132.16, 132.11, 131.62, 131.58, 131.52, 131.50, 129.98, 129.95, 129.92, 129.90, 129.86, 129.83, 129.81, 129.15, 129.12, 128.57, 128.54, 128.49, 128.45, 128.41, 128.33, 127.78, 127.71, 127.71, 127.66, 126.68, 126.62, 126.55, 124.95, 124.61, 124.56, 115.93, 115.86, 113.27, 113.24, 55.54, 55.52, 55.47, 55.45, 55.43, 55.42, 51.98, 48.89, 48.86, 28.64, 28.52, 28.38, 22.22, 21.82.



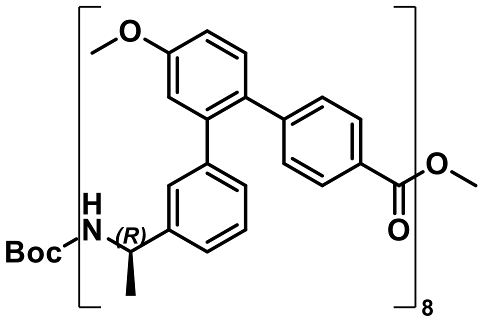
***mop-R*-Tph-8**



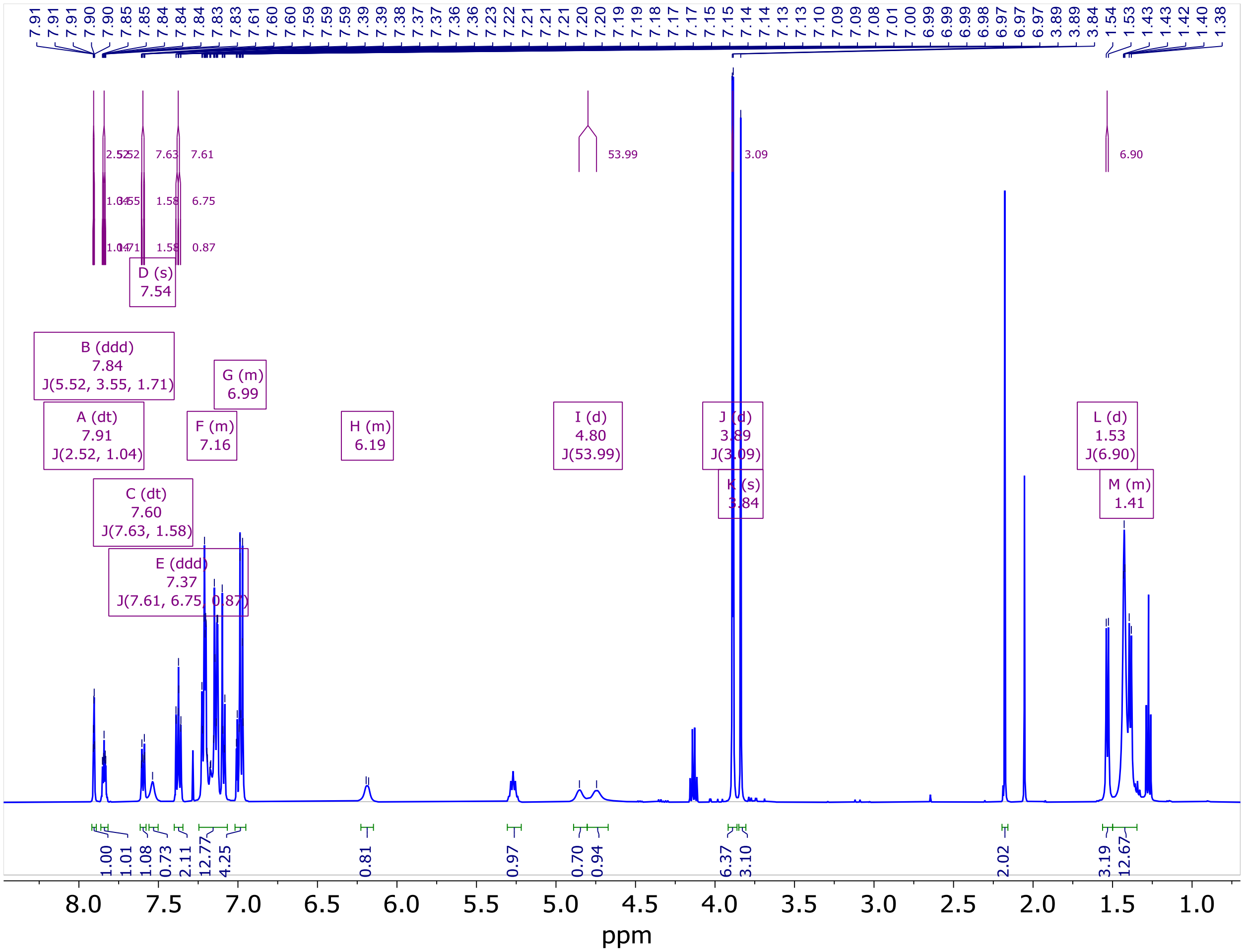
^{1}H NMR (500 MHz, DMSO) δ 8.53 (p, *J* = 8.1 Hz, 7H), 7.75 – 7.56 (m, 16H), 7.42 – 6.85 (m, 75H), 4.98 (p, *J* = 7.6 Hz, 7H), 4.54 – 4.40 (m, 1H), 3.90 – 3.73 (m, 27H), 1.41 – 1.11 (m, 48H).



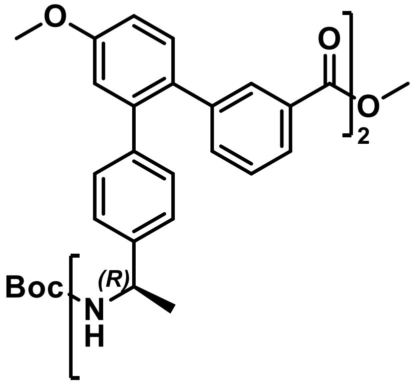
***mop-R*-Tph-8**



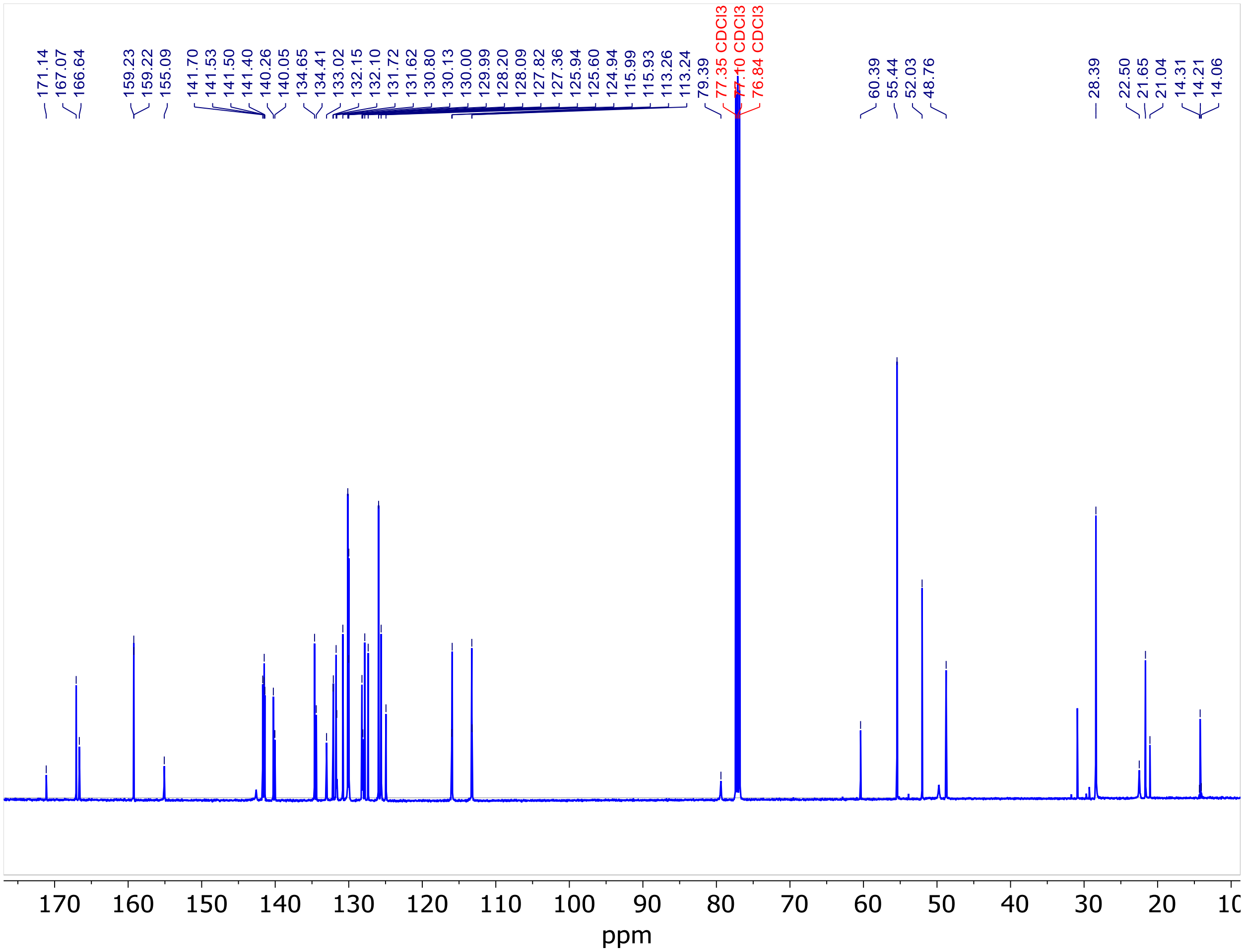
^{13}**C NMR** (125 MHz, CDCl$_3$)



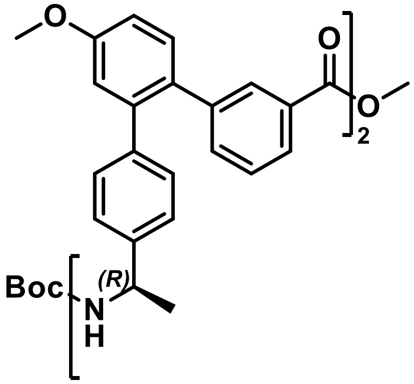
***pom-R*-Tph-2**



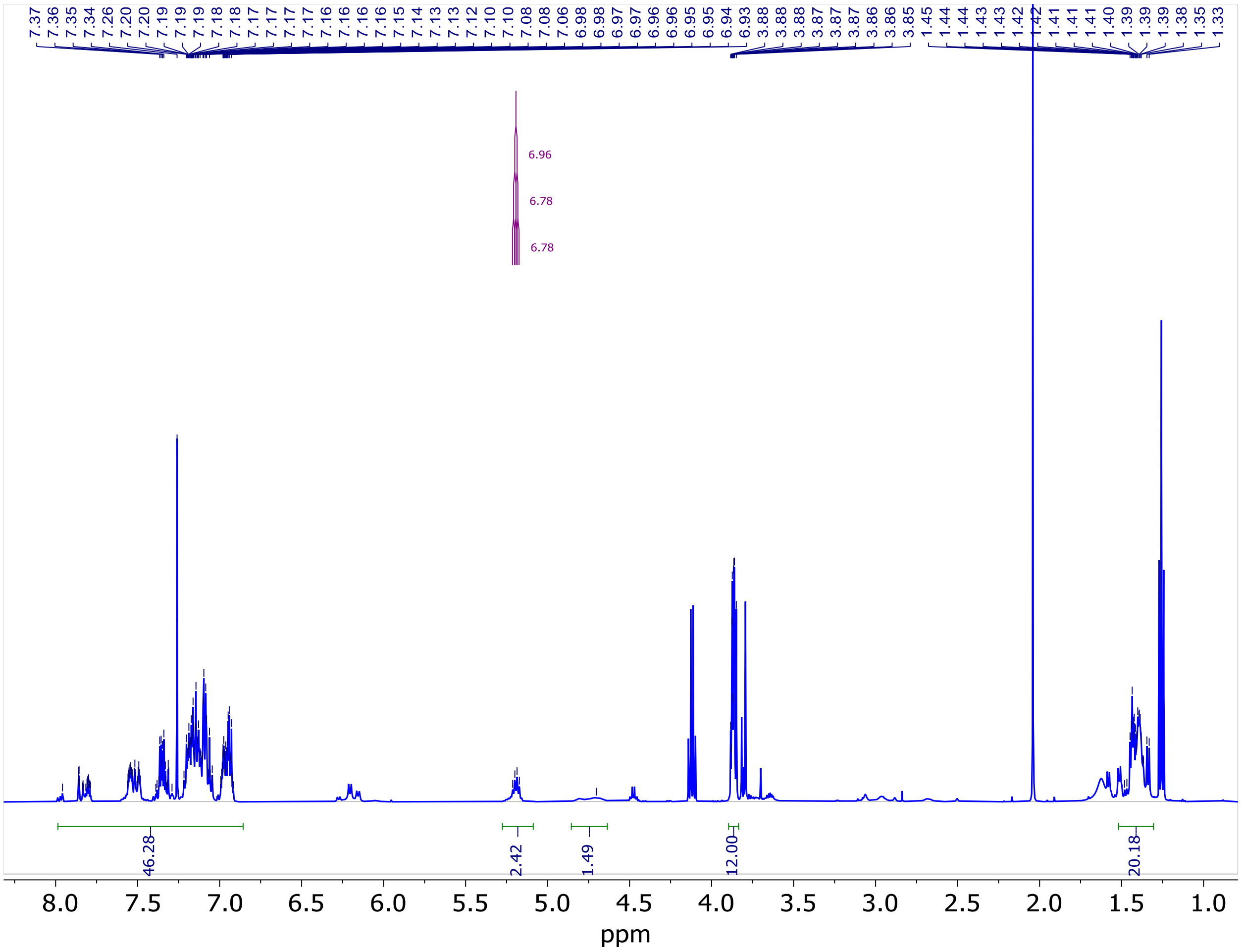
^{1}H NMR (500 MHz, CDCl$_3$) δ 7.91 (dt, *J* = 2.5, 1.0 Hz, 1H), 7.84 (ddd, *J* = 5.5, 3.6, 1.7 Hz, 1H), 7.60 (dt, *J* = 7.6, 1.6 Hz, 1H), 7.54 (s, 1H), 7.37 (ddd, *J* = 7.6, 6.7, 0.9 Hz, 2H), 7.25 – 7.07 (m, 11H), 7.02 – 6.95 (m, 4H), 6.21 – 6.15 (m, 1H), 4.80 (d, *J* = 54.0 Hz, 2H), 3.89 (d, *J* = 3.1 Hz, 6H), 3.84 (s, 3H), 1.53 (d, *J* = 6.9 Hz, 3H), 1.50 – 1.35 (m, 12H).



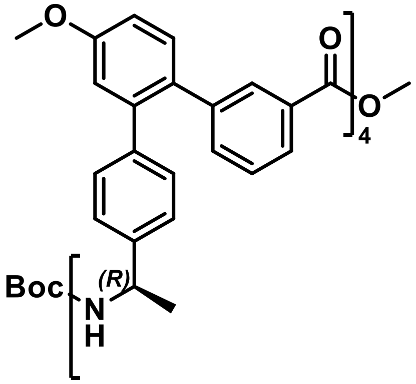
***pom-R*-Tph-2**



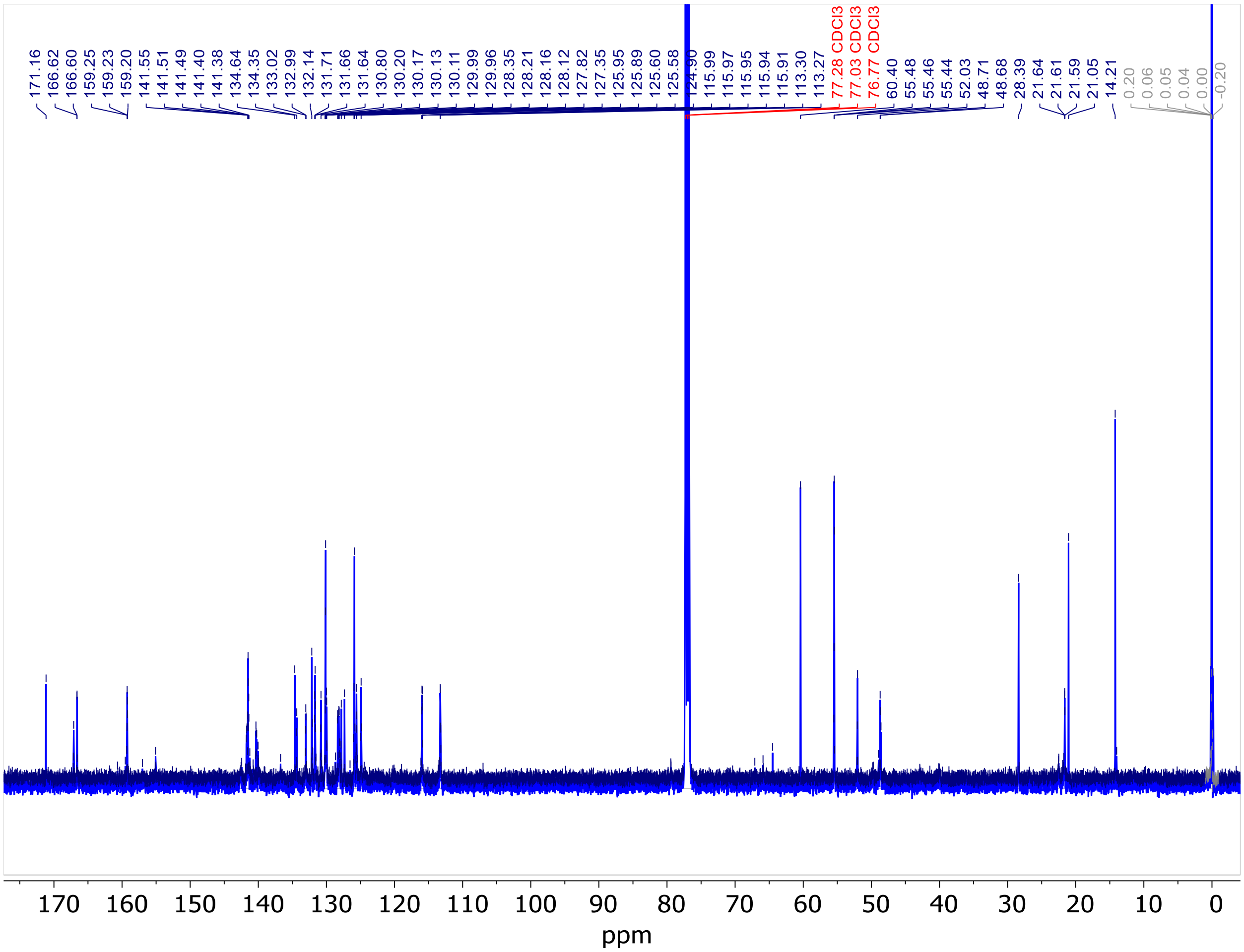
^{13}C NMR (126 MHz, CDCl$_3$) δ 206.93, 171.14, 167.07, 166.64, 159.23, 159.22, 155.09, 141.70, 141.53, 141.50, 141.40, 140.26, 140.05, 134.65, 134.41, 133.02, 132.15, 132.10, 131.72, 131.62, 130.80, 130.13, 130.00, 129.99, 128.20, 128.09, 127.82, 127.36, 125.94, 125.60, 124.94, 115.99, 115.93, 113.26, 113.24, 79.39, 77.35, 77.10, 76.84, 60.39, 55.44, 52.03, 48.76, 28.39, 22.50, 21.65, 21.04, 14.31, 14.21, 14.06.



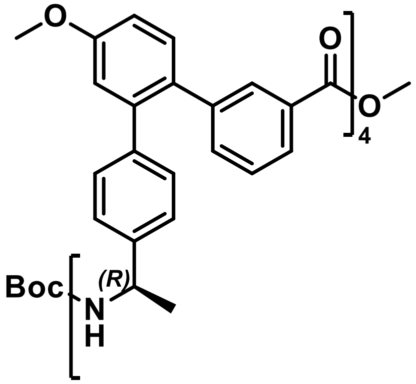
***pom-R*-Tph-4**



^{1}H NMR (500 MHz, CDCl$_3$) δ 7.96 – 6.89 (m, 44H), 5.19 (q, *J* = 6.8 Hz, 2H), 4.70 (s, 1H), 3.92 – 3.85 (m, 12H), 1.49 – 1.31 (m, 20H).



***pom-R*-Tph-4**



^{13}C NMR (126 MHz, CDCl$_3$) δ 171.16, 167.09, 166.62, 166.60, 159.27, 159.25, 159.23, 159.20, 141.76, 141.70, 141.66, 141.55, 141.52, 141.50, 141.49, 141.48, 141.40, 141.38, 140.34, 140.33, 140.23, 140.04, 134.64, 134.42, 134.35, 133.04, 133.02, 132.99, 132.14, 132.08, 131.76, 131.71, 131.66, 131.64, 131.62, 130.83, 130.80, 130.20, 130.17, 130.13, 130.11, 130.00, 129.96, 129.95, 129.93, 128.36, 128.35, 128.21, 128.16, 128.12, 128.07, 127.87, 127.82, 127.34, 126.06, 125.95, 125.89, 125.82, 125.63, 125.61, 125.60, 125.58, 124.95, 124.90, 124.86, 115.99, 115.97, 115.95, 115.94, 115.91, 113.30, 113.27, 113.24, 113.22, 113.20, 79.43, 79.42, 79.41, 60.40, 55.51, 55.48, 55.46, 55.44, 52.07, 52.03, 48.71, 48.68, 48.59, 28.39, 21.64, 21.61, 21.59, 21.05, 14.21.

4.8. References

(1) Benkovic, S. J.; Hammes-Schiffer, S. A Perspective on Enzyme Catalysis. *Science* **2003**, *301* (5637), 1196–1202. https://doi.org/10.1126/science.1085515.

(2) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426* (6968), 884–890. https://doi.org/10.1038/nature02261.

(3) Gething, M.-J.; Sambrook, J. Protein Folding in the Cell. *Nature* **1992**, *355* (6355), 33–45. https://doi.org/10.1038/355033a0.

(4) Menger, F. M. Enzyme Reactivity from an Organic Perspective. *Acc. Chem. Res.* **1993**, *26* (4), 206–212. https://doi.org/10.1021/ar00028a011.

(5) Wolfenden, R.; Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Acc. Chem. Res.* **2001**, *34* (12), 938–945. https://doi.org/10.1021/ar000058i.

(6) Sarodaya, N.; Suresh, B.; Kim, K.-S.; Ramakrishna, S. Protein Degradation and the Pathologic Basis of Phenylketonuria and Hereditary Tyrosinemia. *Int. J. Mol. Sci.* **2020**, *21* (14), 4996. https://doi.org/10.3390/ijms21144996.

(7) Shakhnovich, E. Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry, and Biology Meet. *Chem. Rev.* **2006**, *106* (5), 1559–1588. https://doi.org/10.1021/cr040425u.

(8) Matthews, C. R. PATHWAYS OF PROTEIN FOLDING. *Annu. Rev. Biochem.* **1993**, *62* (1), 653–683. https://doi.org/10.1146/annurev.bi.62.070193.003253.

(9) Baker, D.; Agard, D. A. Kinetics versus Thermodynamics in Protein Folding. *Biochemistry* **1994**, *33* (24), 7505–7509. https://doi.org/10.1021/bi00190a002.

(10) Dupuis, N. F.; Holmstrom, E. D.; Nesbitt, D. J. Molecular-Crowding Effects on Single-Molecule RNA Folding/Unfolding Thermodynamics and Kinetics. *Proc. Natl. Acad. Sci.* **2014**, *111* (23), 8464–8469. https://doi.org/10.1073/pnas.1316039111.

(11) Tokuriki, N.; Kinjo, M.; Negi, S.; Hoshino, M.; Goto, Y.; Urabe, I.; Yomo, T. Protein Folding by the Effects of Macromolecular Crowding. *Protein Sci.* **2004**, *13* (1), 125–133. https://doi.org/10.1110/ps.03288104.

(12) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29* (31), 7133–7155. https://doi.org/10.1021/bi00483a001.

(13) Walter, S.; Buchner, J. Molecular Chaperones—Cellular Machines for Protein Folding. *Angew. Chem. Int. Ed.* **2002**, *41* (7), 1098–1113. https://doi.org/10.1002/1521-3773(20020402)41:7<1098::AID-ANIE1098>3.0.CO;2-9.

(14) Konermann, L. Protein Unfolding and Denaturants. In *Encyclopedia of Life Sciences*; Wiley, 2003. https://doi.org/10.1038/npg.els.0003004.

(15) Prakash, S.; Matouschek, A. Protein Unfolding in the Cell. *Trends Biochem. Sci.* **2004**, *29* (11), 593–600. https://doi.org/10.1016/j.tibs.2004.09.011.

(16) Rader, A. J.; Hespenheide, B. M.; Kuhn, L. A.; Thorpe, M. F. Protein Unfolding: Rigidity Lost. *Proc. Natl. Acad. Sci.* **2002**, *99* (6), 3540–3545. https://doi.org/10.1073/pnas.062492699.

(17) Gellman, S. H. Foldamers: A Manifesto. *Acc. Chem. Res.* **1998**, *31* (4), 173–180. https://doi.org/10.1021/ar960298r.

(18) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. A Field Guide to Foldamers. *Chem. Rev.* **2001**, *101* (12), 3893–4012. https://doi.org/10.1021/cr990120t.

(19) Girvin, Z. C.; Gellman, S. H. Foldamer Catalysis. *J. Am. Chem. Soc.* **2020**, *142* (41), 17211–17223. https://doi.org/10.1021/jacs.0c07347.

(20) Martinek, T. A.; Fülöp, F. Peptidic Foldamers: Ramping up Diversity. *Chem. Soc. Rev.* **2012**, *41* (2), 687–702. https://doi.org/10.1039/C1CS15097A.

(21) Nakano, T.; Okamoto, Y. Synthetic Helical Polymers: Conformation and Function. *Chem. Rev.* **2001**, *101* (12), 4013–4038. https://doi.org/10.1021/cr0000978.

(22) Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci.* **1951**, *37* (4), 205–211. https://doi.org/10.1073/pnas.37.4.205.

(23) Eldirany, S. A.; Ho, M.; Hinbest, A. J.; Lomakin, I. B.; Bunick, C. G. Human Keratin 1/10-1B Tetramer Structures Reveal a Knob-pocket Mechanism in Intermediate Filament Assembly. *EMBO J.* **2019**, *38* (11), e100741. https://doi.org/10.15252/embj.2018100741.

(24) Eldirany, S. A.; Lomakin, I. B.; Bunick, C. G. Crystal Structure of the Wild-Type Heterocomplex between Coil 1B Domains of Human Intermediate Filament Proteins Keratin 1 (KRT1) and Keratin 10 (KRT10): 6ec0, 2019. https://doi.org/10.2210/pdb6ec0/pdb.

(25) Horne, W. S.; Price, J. L.; Gellman, S. H. Interplay among Side Chain Sequence, Backbone Composition, and Residue Rigidification in Polypeptide Folding and

Assembly. *Proc. Natl. Acad. Sci.* **2008**, *105* (27), 9151–9156. https://doi.org/10.1073/pnas.0801135105.

(26) Giuliano, M. W.; Maynard, S. J.; Almeida, A. M.; Guo, L.; Guzei, I. A.; Spencer, L. C.; Gellman, S. H. A γ-Amino Acid That Favors 12/10-Helical Secondary Structure in α/γ-Peptides. *J. Am. Chem. Soc.* **2014**, *136* (42), 15046–15053. https://doi.org/10.1021/ja5076585.

(27) Horne, W. S.; Gellman, S. H. Foldamers with Heterogeneous Backbones. *Acc. Chem. Res.* **2008**, *41* (10), 1399–1408. https://doi.org/10.1021/ar800009n.

(28) Pilsl, L. K. A.; Reiser, O. α/β-Peptide Foldamers: State of the Art. *Amino Acids* **2011**, *41* (3), 709–718. https://doi.org/10.1007/s00726-011-0894-2.

(29) Andrews, M. K.; Liu, X.; Gellman, S. H. Tailoring Reaction Selectivity by Modulating a Catalytic Diad on a Foldamer Scaffold. *J. Am. Chem. Soc.* **2022**, *144* (5), 2225–2232. https://doi.org/10.1021/jacs.1c11542.

(30) Girvin, Z. C.; Gellman, S. H. Exploration of Diverse Reactive Diad Geometries for Bifunctional Catalysis via Foldamer Backbone Variation. *J. Am. Chem. Soc.* **2018**, *140* (39), 12476–12483. https://doi.org/10.1021/jacs.8b05869.

(31) Ferrand, Y.; Huc, I. Designing Helical Molecular Capsules Based on Folded Aromatic Amide Oligomers. *Acc. Chem. Res.* **2018**, *51* (4), 970–977. https://doi.org/10.1021/acs.accounts.8b00075.

(32) Gong, B. Hollow Crescents, Helices, and Macrocycles from Enforced Folding and Folding-Assisted Macrocyclization. *Acc. Chem. Res.* **2008**, *41* (10), 1376–1386. https://doi.org/10.1021/ar700266f.

(33) John, E. A.; Massena, C. J.; Berryman, O. B. Helical Anion Foldamers in Solution. *Chem. Rev.* **2020**, *120* (5), 2759–2782. https://doi.org/10.1021/acs.chemrev.9b00583.

(34) Jiang, H.; Léger, J.-M.; Huc, I. Aromatic δ-Peptides. *J. Am. Chem. Soc.* **2003**, *125* (12), 3448–3449. https://doi.org/10.1021/ja029887k.

(35) Zhang, D.-W.; Zhao, X.; Li, Z.-T. Aromatic Amide and Hydrazide Foldamer-Based Responsive Host–Guest Systems. *Acc. Chem. Res.* **2014**, *47* (7), 1961–1970. https://doi.org/10.1021/ar5000242.

(36) Juwarker, H.; Lenhardt, J. M.; Pham, D. M.; Craig, S. L. 1,2,3-Triazole CH···Cl^- Contacts Guide Anion Binding and Concomitant Folding in 1,4-Diaryl Triazole Oligomers. *Angew. Chem. Int. Ed.* **2008**, *47* (20), 3740–3743. https://doi.org/10.1002/anie.200800548.

(37) Juwarker, H.; Jeong, K.-S. Anion-Controlled Foldamers. *Chem. Soc. Rev.* **2010**, *39* (10), 3664. https://doi.org/10.1039/b926162c.

(38) Juwarker, H.; Suk, J.; Jeong, K.-S. Foldamers with Helical Cavities for Binding Complementary Guests. *Chem. Soc. Rev.* **2009**, *38* (12), 3316. https://doi.org/10.1039/b909034g.

(39) Outlaw, V. K.; Cheloha, R. W.; Jurgens, E. M.; Bovier, F. T.; Zhu, Y.; Kreitler, D. F.; Harder, O.; Niewiesk, S.; Porotto, M.; Gellman, S. H.; Moscona, A. Engineering Protease-Resistant Peptides to Inhibit Human Parainfluenza Viral Respiratory Infection. *J. Am. Chem. Soc.* **2021**, *143* (15), 5958–5966. https://doi.org/10.1021/jacs.1c01565.

(40) Outlaw, V. K.; Bovier, F. T.; Mears, M. C.; Cajimat, M. N.; Zhu, Y.; Lin, M. J.; Addetia, A.; Lieberman, N. A. P.; Peddu, V.; Xie, X.; Shi, P.-Y.; Greninger, A. L.; Gellman, S. H.; Bente, D. A.; Moscona, A.; Porotto, M. Inhibition of Coronavirus Entry *In Vitro* and *Ex Vivo* by a Lipid-Conjugated Peptide Derived from the SARS-CoV-2 Spike Glycoprotein HRC Domain. *mBio* **2020**, *11* (5), e01935-20. https://doi.org/10.1128/mBio.01935-20.

(41) Kleman, A. F.; Dufek, D. L.; Fobe, T. L.; McCaslin, D. R.; Cary, B. P.; Shirts, M. R.; Gellman, S. H. Potential Foldamers Based on an *Ortho-* Terphenyl Amino Acid. *Org. Lett.* **2021**, *23* (12), 4855–4859. https://doi.org/10.1021/acs.orglett.1c01592.

(42) Mathew, S.; Crandall, L. A.; Ziegler, C. J.; Hartley, C. S. Enhanced Helical Folding of *Ortho* -Phenylenes through the Control of Aromatic Stacking Interactions. *J. Am. Chem. Soc.* **2014**, *136* (47), 16666–16675. https://doi.org/10.1021/ja509902m.

(43) Kleman, A. F. *Potential Foldamers Based on an Ortho-Terphenyl Amino Acid*; Ph.D. Thesis; University of Wisconsin - Madison, 2021.

(44) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2006**, *1* (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.

(45) Linhares, L. A.; Ramos, C. H. I. Unlocking Insights into Folding, Structure, and Function of Proteins through Circular Dichroism Spectroscopy—A Short Review. *Appl. Biosci.* **2023**, *2* (4), 639–655. https://doi.org/10.3390/applbiosci2040040.

(46) Stone, M. T.; Heemstra, J. M.; Moore, J. S. The Chain-Length Dependence Test. *Acc. Chem. Res.* **2006**, *39* (1), 11–20. https://doi.org/10.1021/ar0501267.

(47) Li, A. D. Q.; Wang, W.; Wang, L. Folding versus Self-Assembling. *Chem. – Eur. J.* **2003**, *9* (19), 4594–4601. https://doi.org/10.1002/chem.200305025.

(48) Crutchfield, C. A.; Harris, D. J. Molecular Mass Estimation by PFG NMR Spectroscopy. *J. Magn. Reson.* **2007**, *185* (1), 179–182. https://doi.org/10.1016/j.jmr.2006.12.004.

(49) Cabrita, E. J.; Berger, S. DOSY Studies of Hydrogen Bond Association: Tetramethylsilane as a Reference Compound for Diffusion Studies. *Magn. Reson. Chem.* **2001**, *39* (S1), S142–S148. https://doi.org/10.1002/mrc.917.

(50) Waldeck, A. R.; Kuchel, P. W.; Lennon, A. J.; Chapman, B. E. NMR Diffusion Measurements to Characterise Membrane Transport and Solute Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *30* (1–2), 39–68. https://doi.org/10.1016/S0079-6565(96)01034-5.

(51) Semmelhack, H. C.; Esquinazi, P. Observation of Metastable Ordered Structures and the Kinetics of Crystallization of O-Terphenyl. *Phys. B Condens. Matter* **1998**, *254* (1–2), 14–20. https://doi.org/10.1016/S0921-4526(98)00415-3.

(52) Urawa, Y.; Naka, H.; Miyazawa, M.; Souda, S.; Ogura, K. Investigations into the Suzuki–Miyaura Coupling Aiming at Multikilogram Synthesis of E2040 Using (o-Cyanophenyl)Boronic Esters. *J. Organomet. Chem.* **2002**, *653* (1–2), 269–278. https://doi.org/10.1016/S0022-328X(02)01175-0.

(53) Inoue, M.; Nagamori, H.; Morita, T.; Kobayashi, S.; Suzawa, K.; Kitao, Y.; Saito, T.; Kawahara, I.; Orita, T.; Akai, S.; Adachi, T.; Motomura, T. Design and Synthesis of Novel Fluorene Derivatives as Inhibitors of Pyruvate Dehydrogenase Kinase.

Bioorg. Med. Chem. Lett. **2024**, *109*, 129839. https://doi.org/10.1016/j.bmcl.2024.129839.

(54) Chen, H.; Wang, K.; Yang, Y.; Huang, X.; Dai, X.; Feng, Z. Design, Synthesis, and Structure-Activity Relationship of Programmed Cell Death-1/Programmed Cell Death-Ligand 1 Interaction Inhibitors Bearing a Benzo[d]Isothiazole Scaffold. *Eur. J. Med. Chem.* **2021**, *217*, 113377. https://doi.org/10.1016/j.ejmech.2021.113377.

(55) Lu, T.; Zhang, J.; Chen, Q.; Ni, M.; Zhang, J.; Wu, Y.; Jia, R.; Wang, Y. Design, Synthesis, Evaluation, and SAR of 5-Phenylisoindoline Derivatives, a Potent Class of Small-Molecule Inhibitors Targeting the Programmed Cell Death-1/Programmed Cell Death-Ligand 1 (PD-1/PD-L1) Interaction. *J. Med. Chem.* **2025**, *68* (7), 7291–7312. https://doi.org/10.1021/acs.jmedchem.4c02206.

(56) Stambuli, J. P.; Kuwano, R.; Hartwig, J. F. Unparalleled Rates for the Activation of Aryl Chlorides and Bromides: Coupling with Amines and Boronic Acids in Minutes at Room Temperature. *Angew. Chem. Int. Ed.* **2002**, *41* (24), 4746–4748. https://doi.org/10.1002/anie.200290036.

(57) Upadhyaya, D. J.; Barge, A.; Stefania, R.; Cravotto, G. Efficient, Solventless N-Boc Protection of Amines Carried out at Room Temperature Using Sulfamic Acid as Recyclable Catalyst. *Tetrahedron Lett.* **2007**, *48* (47), 8318–8322. https://doi.org/10.1016/j.tetlet.2007.09.126.

(58) Carole, W. A.; Colacot, T. J. Understanding Palladium Acetate from a User Perspective. *Chem. – Eur. J.* **2016**, *22* (23), 7686–7695. https://doi.org/10.1002/chem.201601450.

(59) Carole, W. A.; Bradley, J.; Sarwar, M.; Colacot, T. J. Can Palladium Acetate Lose Its “Saltiness”? Catalytic Activities of the Impurities in Palladium Acetate. *Org. Lett.* **2015**, *17* (21), 5472–5475. https://doi.org/10.1021/acs.orglett.5b02835.

(60) Bruno, N. C.; Niljianskul, N.; Buchwald, S. L. *N* -Substituted 2-Aminobiphenylpalladium Methanesulfonate Precatalysts and Their Use in C–C and C–N Cross-Couplings. *J. Org. Chem.* **2014**, *79* (9), 4161–4166. https://doi.org/10.1021/jo500355k.

(61) Billingsley, K. L.; Barder, T. E.; Buchwald, S. L. Palladium-Catalyzed Borylation of Aryl Chlorides: Scope, Applications, and Computational Studies. *Angew. Chem. Int. Ed.* **2007**, *46* (28), 5359–5363. https://doi.org/10.1002/anie.200701551.

(62) El-Faham, A.; Albericio, F. COMU: A Third Generation of Uronium-type Coupling Reagents. *J. Pept. Sci.* **2010**, *16* (1), 6–9. https://doi.org/10.1002/psc.1204.

(63) Matthews, B. W. Racemic Crystallography—Easy Crystals and Easy Structures: What’s Not to Like? *Protein Sci.* **2009**, *18* (6), 1135–1138. https://doi.org/10.1002/pro.125.

(64) Yeates, T. O.; Kent, S. B. H. Racemic Protein Crystallography. *Annu. Rev. Biophys.* **2012**, *41* (1), 41–61. https://doi.org/10.1146/annurev-biophys-050511-102333.

(65) Honegger, P.; Gajic, N.; Prado-Roller, A.; Widhalm, M. Racemic and Meso Crystal Structures of an Axial-Chiral Spirobi-(Dinaphthoazepin)Ium Salt: Emergence of an S4-Symmetric Molecule. *Symmetry* **2021**, *13* (8), 1365. https://doi.org/10.3390/sym13081365.

(66) Nannenga, B. L.; Gonen, T. The Cryo-EM Method Microcrystal Electron Diffraction (MicroED). *Nat. Methods* **2019**, *16* (5), 369–379. https://doi.org/10.1038/s41592-019-0395-x.

(67) Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T. The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent. Sci.* **2018**, *4* (11), 1587–1592. https://doi.org/10.1021/acscentsci.8b00760.

(68) Kunde, T.; Schmidt, B. M. Microcrystal Electron Diffraction (MicroED) for Small-Molecule Structure Determination. *Angew. Chem. Int. Ed.* **2019**, *58* (3), 666–668. https://doi.org/10.1002/anie.201813215.

Appendix A

Additional LLPS Systems

In collaboration with Charles Lenz (UW-Madison)

While investigating liquid-liquid phase separation (LLPS), several new coacervate-forming systems have been developed. Although these systems do not appear in the previous chapters of this thesis, each will be mentioned in this chapter.

Three categories of LLPS systems are described in this appendix: peptidic LLPS systems, LLPS systems based on synthetic molecules, and LLPS systems of ions. Each system will be briefly discussed.

Peptidic LLPS Systems

In addition to the systems tested in Chapter 3, other peptidic systems include those listed below.

- *Poly-L-arginine and polyphosphate*

  Poly-L-Arginine and polyphosphate (P45) were tested in TRIS buffer (10 mM TRIS, 15 mM KCl, 0.5 mM $MgCl_2$). Under some conditions, droplets were observed (**Figure A.1**). Other conditions yielded solid aggregates or gels.

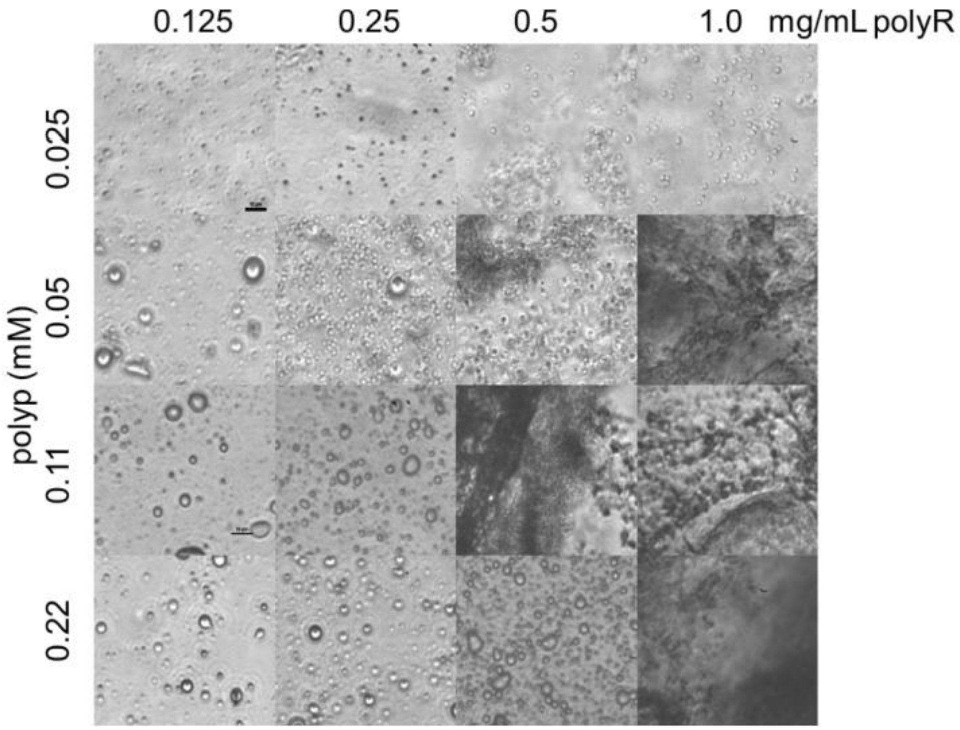
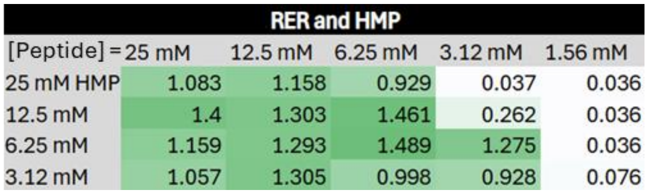


Figure A.1 Droplet formation and other forms of assembly observed with microscopy between P45 polyphosphate and poly-L-arginine in TRIS buffer.

- *RER and HMP*

L-Arg-L-Glu-L-Arg (RER) formed droplets with hexametaphosphate (HMP) at peptide concentrations between 1 mM and 25 mM in TRIS buffer (**Figure A.2**). Droplets seemed large. LLPS was easily disrupted by free amino acids (notably glycine). RER did not form droplets with ATP.



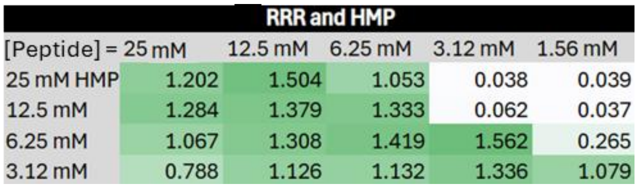
RER and HMP

[Peptide] =	25 mM	12.5 mM	6.25 mM	3.12 mM	1.56 mM
25 mM HMP	1.083	1.158	0.929	0.037	0.036
12.5 mM	1.4	1.303	1.461	0.262	0.036
6.25 mM	1.159	1.293	1.489	1.275	0.036
3.12 mM	1.057	1.305	0.998	0.928	0.076

Figure A.2 OD_{600} values of coacervation between peptide RER and HMP in TRIS buffer.

- *RRR and HMP*

  L-Arg-L-Arg-L-Arg (RRR) also underwent LLPS with HMP, and had a lower critical droplet concentration than RER and RR (see chapter 3) in TRIS buffer (**Figure A.3**).



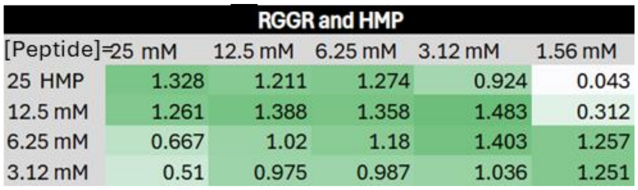
RRR and HMP

[Peptide] =	25 mM	12.5 mM	6.25 mM	3.12 mM	1.56 mM
25 mM HMP	1.202	1.504	1.053	0.038	0.039
12.5 mM	1.284	1.379	1.333	0.062	0.037
6.25 mM	1.067	1.308	1.419	1.562	0.265
3.12 mM	0.788	1.126	1.132	1.336	1.079

Figure A.3 OD_{600} values of coacervation between peptide RRR and HMP in TRIS buffer.

- *RGGR and HMP*

  L-Arg-Gly-Gly-L-Arg (RGGR) showed the greatest LLPS propensity of RR, RER, and RRR in TRIS buffer with HMP (**Figure A.4**).



RGGR and HMP

[Peptide]=	25 mM	12.5 mM	6.25 mM	3.12 mM	1.56 mM
25 HMP	1.328	1.211	1.274	0.924	0.043
12.5 mM	1.261	1.388	1.358	1.483	0.312
6.25 mM	0.667	1.02	1.18	1.403	1.257
3.12 mM	0.51	0.975	0.987	1.036	1.251

Figure A.4 OD_{600} values of coacervation between peptide RGGR and HMP in TRIS buffer.

- RHR and rHR

  L-Arg-L-His-L-Arg (RHR) and its diastereomer D-Arg-L-His-L-Arg (rHR) were prepared. Histidine was chosen as a chromophore to allow for easy separation. An LLPS phase diagram was not generated for these tripeptides, however, both diastereomers underwent LLPS at similar concentrations to RER.

Peptidic LLPS systems like the ones tested above were sensitive to free amino acids in solution. Many of these systems remained homogeneous in glycine buffer but formed coacervates in TRIS buffer. The ability of free amino acids to modulate LLPS has been characterized previously in protein-containing systems where it is thought that amino acids disrupt hydrogen bonding between the amide groups in a peptide/protein backbone.[1,2]

Synthetic Molecules

- *Alkyl-diguanidinium Compounds and HMP*

  Cheng et al. reported an LLPS system utilizing an alkyl-diguanidinium compound with a 7-methylene linker (**Figure A.5**) in hexametaphosphate:[3]

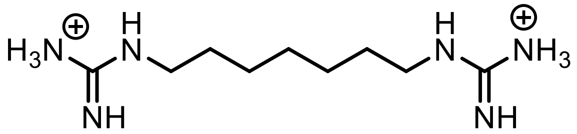
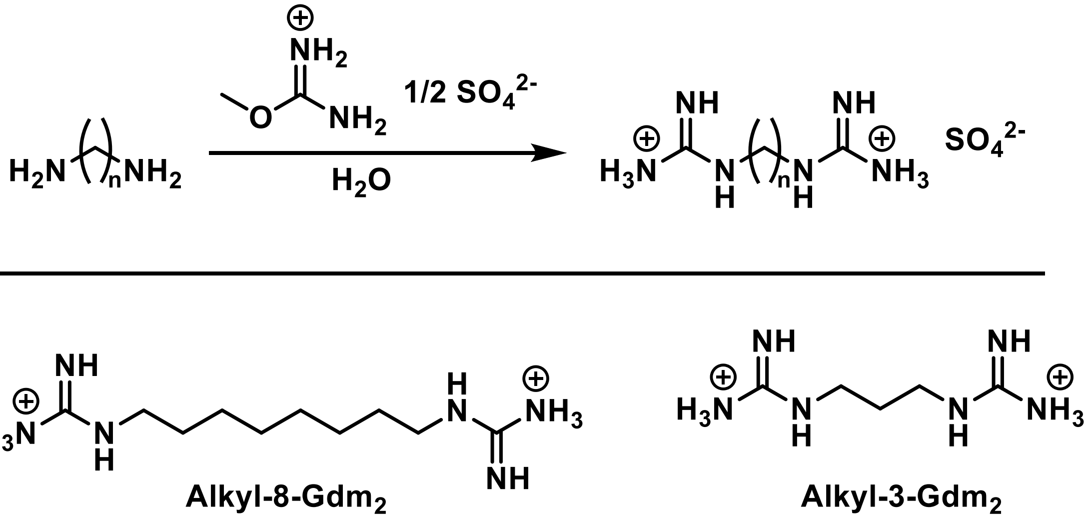


Figure A.5 Alkyl-diguanidinium compound described in the literature.[3]

We prepared similar compounds with 3 and 8 methylene groups through guanylation of alkyl diamines (**Scheme A.1**):



Alkyl-8-Gdm$_2$

Alkyl-3-Gdm$_2$

Scheme A.1 Alkyl-diguanidinium compound synthesis.

These compounds were facile to make: the diamine was simply stirred with 2 eq methylcarbamimidate in minimum water. The desired compound precipitated after several hours and was washed with water and ethanol to produce the pure compound.

Unlike peptidic systems, alkyl-diguanidinium systems were resistant to high concentrations (>500 mM) of free amino acids (**Figure A.6**). These synthetic LLPS systems might therefore be utilized in future experiments requiring free amino acids. However, they are sensitive to the addition of magnesium. Mg^{2+} concentrations in excess of 10 mM were found to inhibit droplet formation completely.

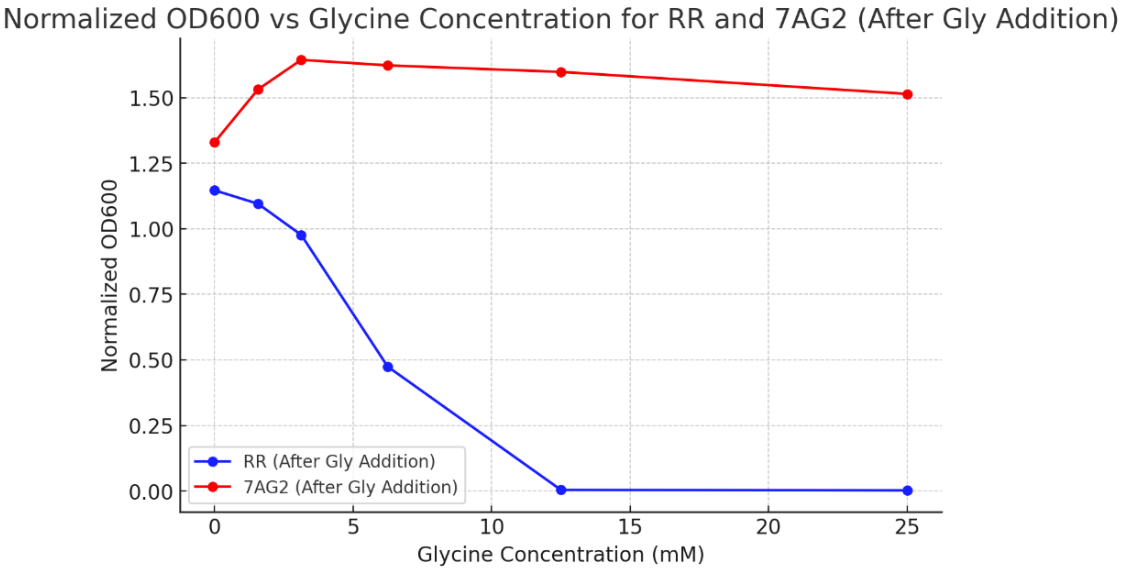


Figure A.6 Titration of glycine into droplets made from HMP with either alkyl-7-Gdm_2 or arginylarginine. Vertical axis represents optical density.

Alkyl-diguanidinium droplets were very concentrated—no excess HMP or alkyl-8-Gdm_2 was observed in the dilute phase via ^{31}P or ^{1}H NMR (respectively). The most robust droplet formation occurred when [alkyl-8-Gdm_2]=10 mM and [HMP]=3 mM.

Alkyl-3-Gdm$_2$ required higher concentrations of both components relative to the octyl derivative. LLPS of alkyl-3-Gdm$_2$ occurred when the [alkyl-3-Gdm$_2$]=20 mM and [HMP]=3-6 mM (**Figure A.7**).

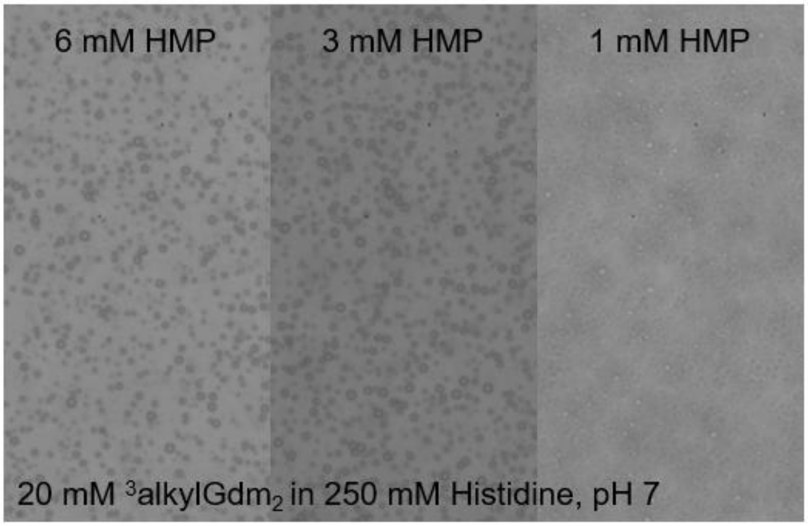


Figure A.7 Optical microscopy of coacervates formed between HMP and alkyl-3-Gdm$_2$ in histidine buffer.

After the evaporation of buffer (either 0.1 M glycine, or phosphate-buffered saline containing 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4) from the LLPS experiments, it was typical of alkyl-diguanidinium systems to leave behind crystals in each well. X-ray crystallography (with Ilia Guzei and Leah Garman in the Chemical Instrumentation Center) revealed that the crystals consisted of the alkyl-diguanidinium compound with a sulfate counterion (**Figure A.8**).

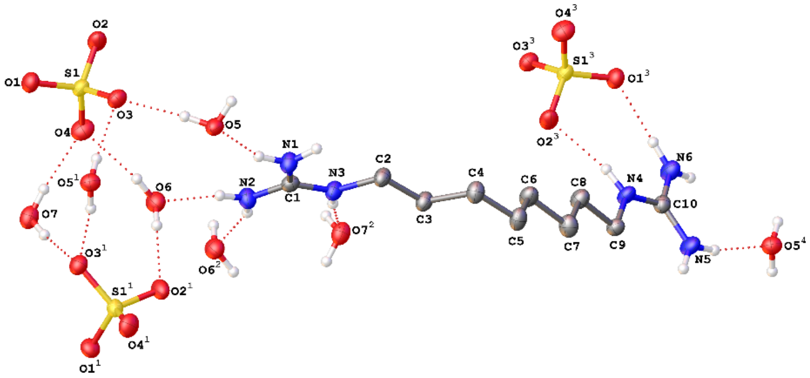


Figure A.8 Crstal structure of alkyl-8-Gdm$_2$ with sulfate counterion.

Multiple unit cells stacked together to form crystals with pores assembled from the hydrophobic alkyl portion of the molecule (**Figure A.9**).

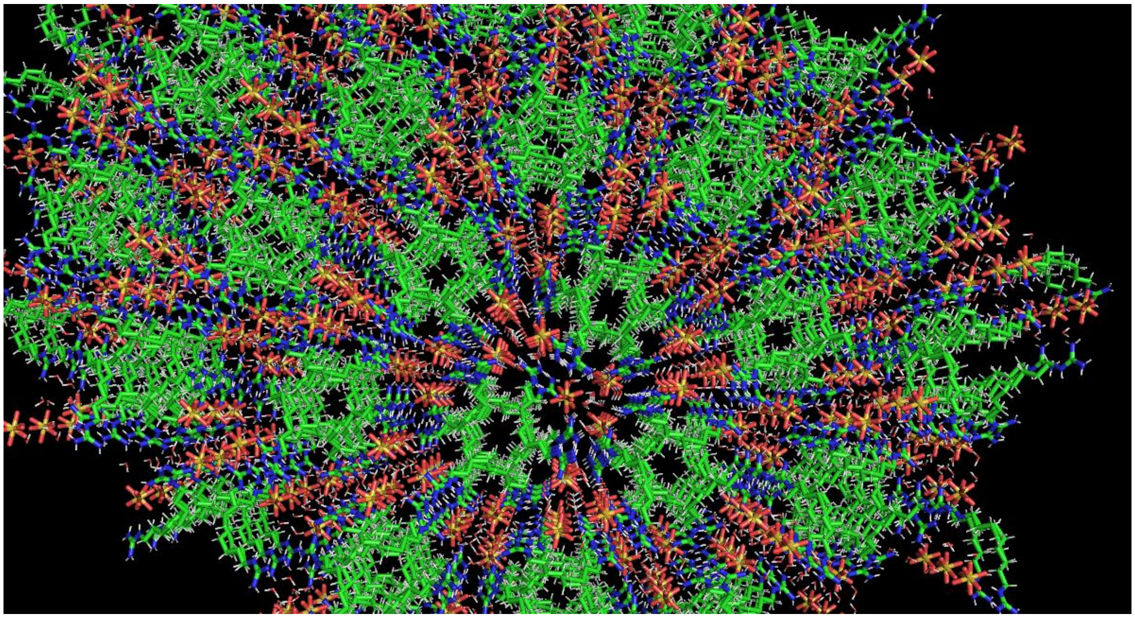


Figure A.9 Crstal packing of alkyl-8-Gdm$_2$ with sulfate counterion.

Ionic Systems

Aevi van der Stok previously noted that calcium ions will spontaneously form coacervates with medium-chain polyphosphate. This helped us rationalize that small ions, not just medium or large molecules, can also coacervate under the right conditions. We were successful in forming droplets with calcium, magnesium, and guanidinium ions with HMP. Other researchers reported similar findings with manganese.[4]

- *Calcium or magnesium and HMP*

  Both calcium and magnesium (125-500 mM) underwent LLPS with HMP (30-60 mM) to form large coacervates in phosphate-buffered saline (containing 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4) (**Figure A.10**).

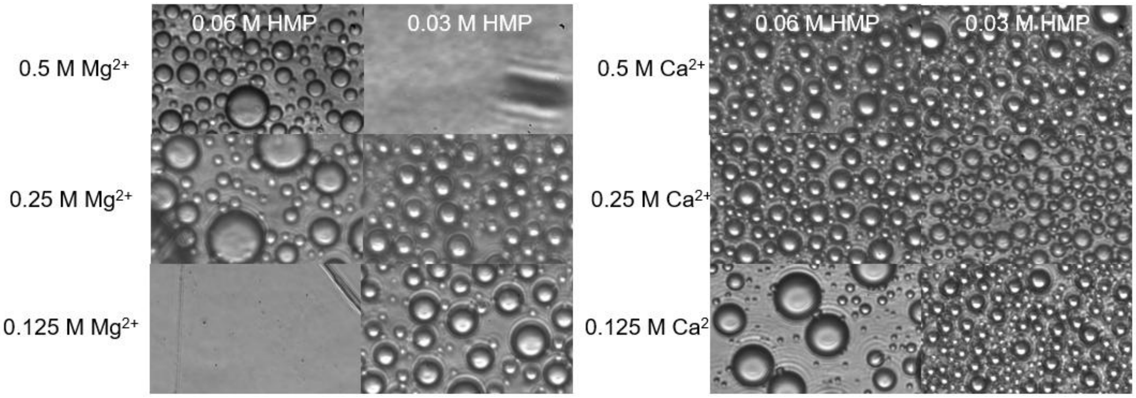
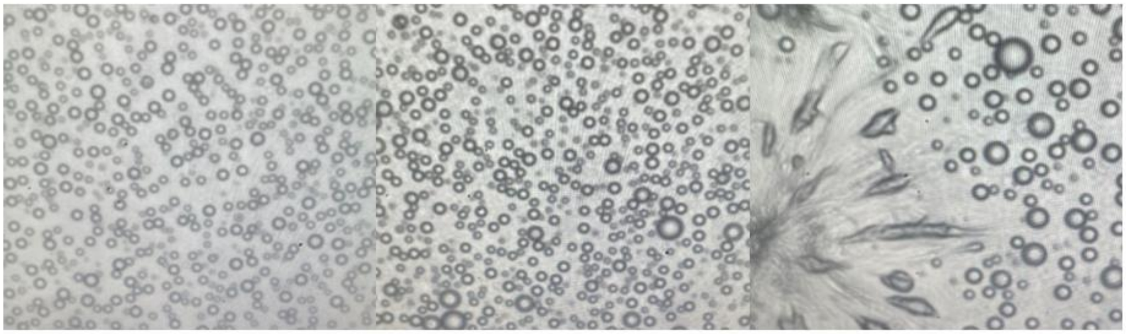


Figure A.10 Optical microscopy of coacervates formed between HMP and either calcium or magnesium chloride in phosphate-buffered saline.

It has been reported previously that calcium ions can catalyze the hydrolysis of HMP.[5] We did not conduct these studies in this droplet system.

- *Guanidinium and HMP*

  Guanidinium chloride or sulfate underwent coacervation with HMP in either distilled water or TRIS buffer (**Figure A.11**). Droplets were very large and concentrated. A large molar excess of guanidinium was required for coacervation. These droplets matured into fibrils after ~1 hr.



10 mM HMP; 0.25 M GdmCl 10 mM HMP; 0.5 M GdmCl 10 mM HMP; 2 M GdmCl

Figure A.11 Optical microscopy of coacervates formed between HMP and guanidinium chloride in distilled water.

References:

(1) Winkler, P. M.; Siri, C.; Buczkowski, J.; Silva, J. V. C.; Bovetto, L.; Schmitt, C.; Stellacci, F. Modulating Weak Protein–Protein Cross-Interactions by the Addition of Free Amino Acids at Millimolar Concentrations. *J. Phys. Chem. B* **2024**, *128* (29), 7199–7207. https://doi.org/10.1021/acs.jpcb.4c01086.

(2) Xu, X.; Rebane, A. A.; Roset Julia, L.; Rosowski, K. A.; Dufresne, E. R.; Stellacci, F. Amino Acids Modulate Liquid–Liquid Phase Separation in Vitro and in Vivo by Regulating Protein–Protein Interactions. *Proc. Natl. Acad. Sci.* **2024**, *121* (50), e2407633121. https://doi.org/10.1073/pnas.2407633121.

(3) Cheng, Y.; Hirano, E.; Wang, H.; Kuwayama, M.; Meijer, E. W.; Huang, H.; Aida, T. Mechanically Strong yet Metabolizable Supramolecular Plastics by Desalting upon Phase Separation. *Science* **2024**, *386* (6724), 875–881. https://doi.org/10.1126/science.ado1782.

(4) Dai, S.; Xie, Z.; Wang, B.; Ye, R.; Ou, X.; Wang, C.; Yu, N.; Huang, C.; Zhao, J.; Cai, C.; Zhang, F.; Buratto, D.; Khan, T.; Qiao, Y.; Hua, Y.; Zhou, R.; Tian, B. An Inorganic Mineral-Based Protocell with Prebiotic Radiation Fitness. *Nat. Commun.* **2023**, *14* (1), 7699. https://doi.org/10.1038/s41467-023-43272-5.

(5) Avila Salazar, D. A.; Bellstedt, P.; Miura, A.; Oi, Y.; Kasuga, T.; Brauer, D. S. Unravelling the Dissolution Mechanism of Polyphosphate Glasses by^{31} P NMR Spectroscopy: Ligand Competition and Reactivity of Intermediate Complexes. *Dalton Trans.* **2021**, *50* (11), 3966–3978. https://doi.org/10.1039/D0DT03381B.

Appendix B

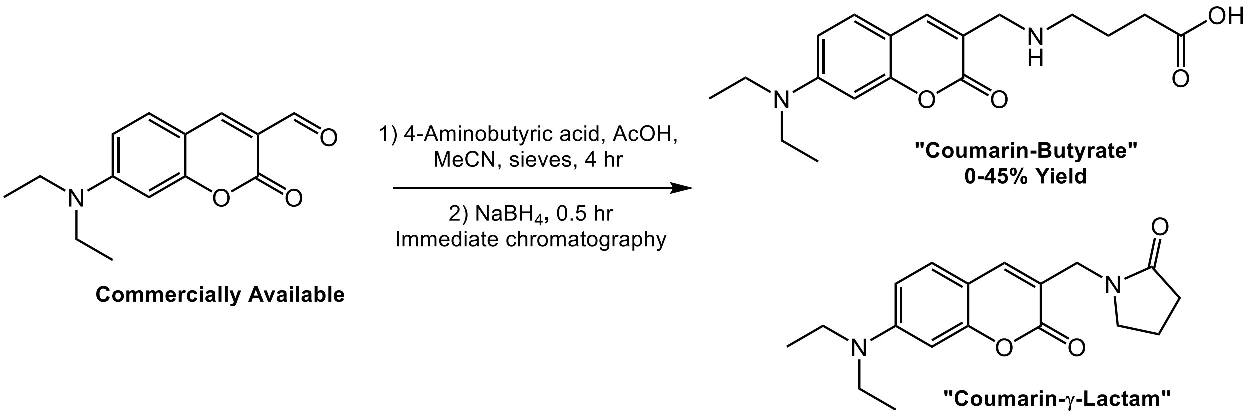
Attempts to Generate Fluorescent Sensors for Amide Formation

The mechanism by which peptides were elongated prior to the evolution of the ribosome remains a mystery. Amino acid condensation is unfavorable in aqueous media, especially at pH 7, where amino acids are zwitterionic. It seems most likely that some other molecule activated the carboxylic acid to make the carbonyl sufficiently electrophilic.[1,2] Extant life phosphorylates amino acids so they can be appended to tRNA.[3] The aminoacyl-tRNA is subsequently used in the ribosome to build peptides and proteins. Here we propose that a similar mechanism might have occurred abiotically.[2]

Previous studies demonstrated a plausible prebiotic path towards amide bond formation using activation with polyphosphate.[4,5] Most notably, trimetaphosphate (TMP) has been shown to form aminoacyl phosphate esters with α-amino acids or simple dipeptides.[6–8] These phosphate esters can polymerize into peptides in low yield. Because polyphosphates can drive coacervation of cations and short peptides (see Appendix A), and droplets can catalyze reactions due to high local concentrations of reactant molecules and partitioning of starting materials and products, we reason that polyphosphate-containing droplets could have behaved as prebiotic microreactors for amide bond formation.

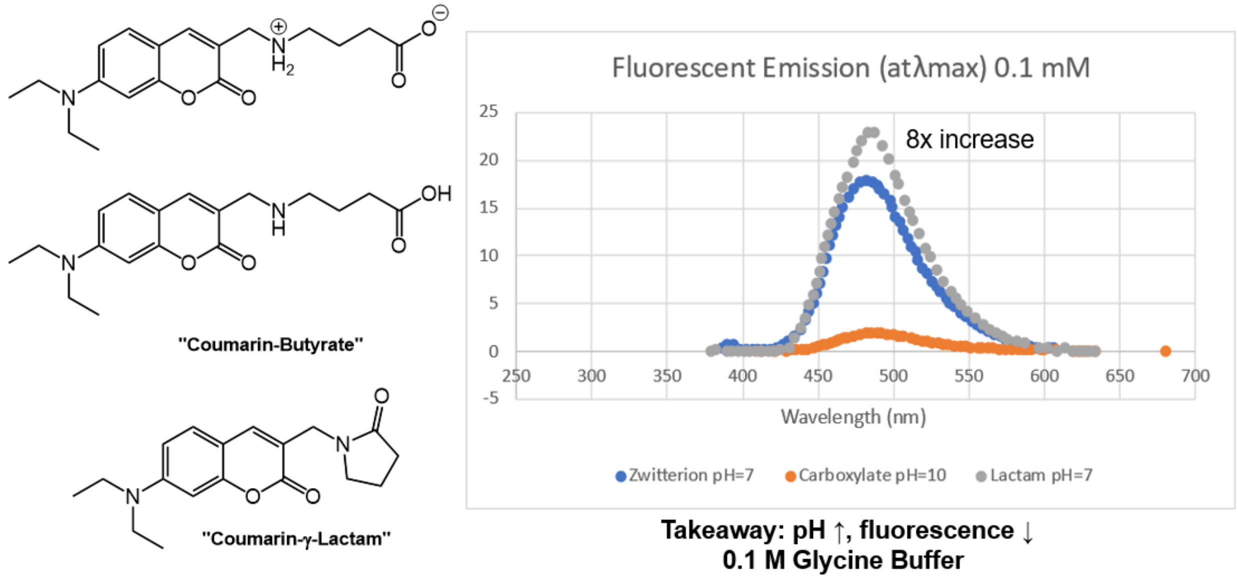
To detect amide bond formation in the condensed phase, we desired a turn-on fluorescent probe that would allow high-throughput screening and sensitive detection of conditions conducive to amide bond formation. Our efforts to develop this probe are chronicled in this appendix; these efforts were not successful.

Our initial design focused on a coumarin-based fluorescent probe. We proposed a secondary amine near the coumarin that could participate in photoinduced electron transfer (PET) quenching of the coumarin. Forming an amide between the amine and carboxylic acid would remove the ability for quenching. Starting with the aldehyde, we formed an amino acid and the corresponding lactam via reductive amination (**Scheme B.1**). Too little reaction time resulted in incomplete conversion, while long reaction time resulted in spontaneous cyclization. Both compounds were formed after a 30-minute reduction and separated with chromatography.



Scheme B.1 Synthesis of coumarin-butyrate fluorescent probe using reductive amination.

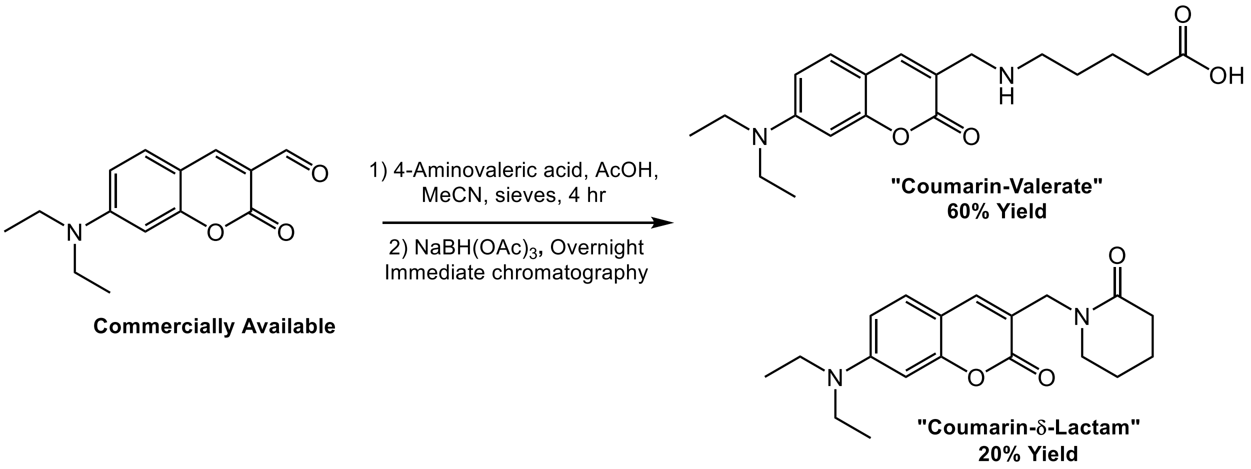
The carboxylate “coumarin-butyrate” showed a baseline level of fluorescence at both neutral and basic pH. We expected that the protonation of the nitrogen at pH=7 would attenuate PET, and at basic pH values deprotonation of the nitrogen would result in more efficient quenching. Fluorescence measurements supported this hypothesis (**Figure B.1**). The fluorescence of the protonated coumarin-butyrate had a similar magnitude as that of the lactam. This result meant that this probe design could not be used in LLPS systems at neutral pH.



Takeaway: pH ↑, fluorescence ↓
0.1 M Glycine Buffer

Figure B.1 Fluorescence of coumarin-butyrate and lactam in glycine buffer.

A similar "coumarin-valerate" compound was made in the same way (**Scheme B.2**). This compound was not as prone to spontaneous cyclization under reductive amination conditions as the butyrate analogue.



Scheme B.2 Synthesis of coumarin-valerate fluorescent probe using reductive amination.

The valerate showed a considerable color change upon lactamization, from pale yellow to violet (**Figure B.2**).

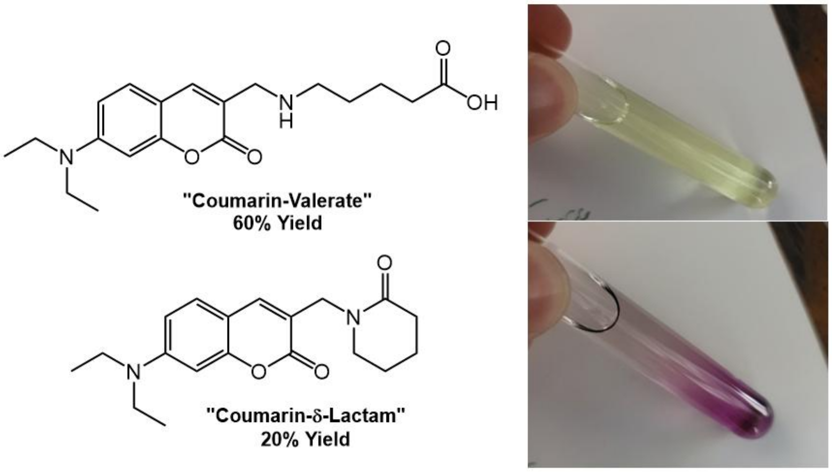


Figure B.2 Coumarin valerate changes color upon lactamization.

Despite this color difference, no conditions were discovered in which fluorescence is significantly different between open-form coumarin valerate and its δ-lactam under various dilutions and pH ranges (100 nM – 10 µM; pH 6 - 9).

Another coumarin derivative was used to form an alternative probe (**Figure B.3**) that did not exhibit a fluorescence turn-on after lactamization.

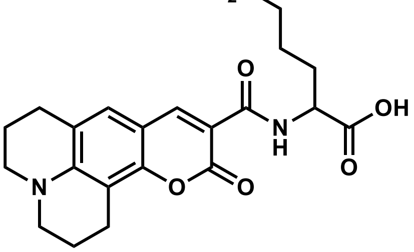
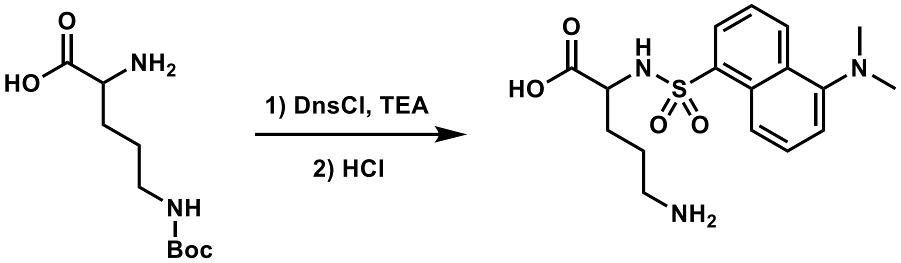


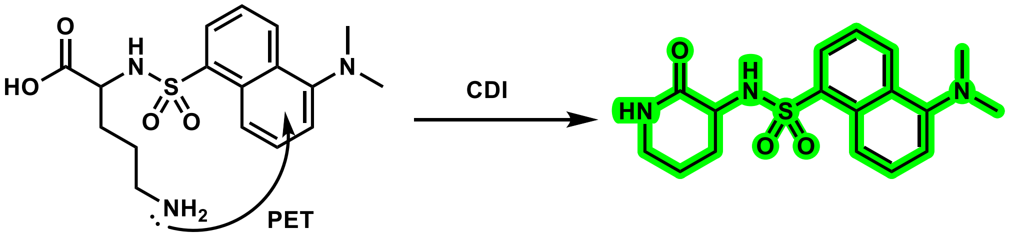
Figure B.3 Alternate coumarin fluorophore does not exhibit turn-on after lactamization.

Dansyl-ornithine was another possibility. The side-chain amine might allow for PET quenching and cyclization with the carboxylic acid while the α-amino group is tethered to the fluorescent dansyl group (**Scheme B.3**).



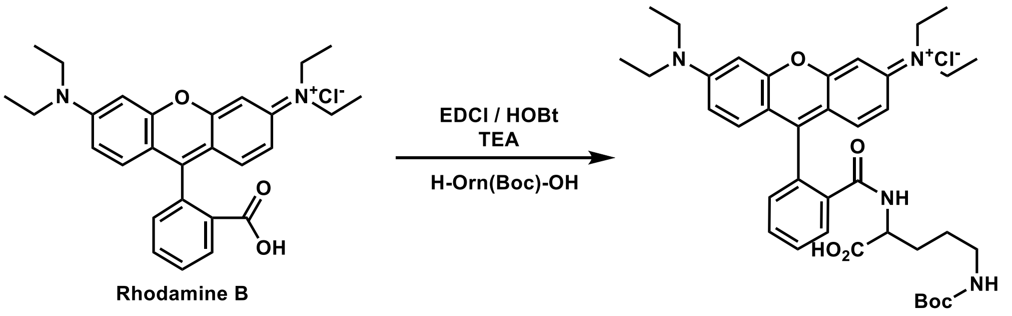
Scheme B.3 Synthesis of dansyl-ornithine probe.

This molecule excited at 325 nm and exhibited broad fluorescence peaks at 500 nm in nM concentrations. Cyclization of dansyl-ornithine with carbonyldiimidazole (CDI) (**Scheme B.4**) resulted in a 2x fluorescence turn-on, which was determined to be insufficient for sensitive experiments.



Scheme B.4 Cyclization with CDI results in a 2x turn-on in with dansyl ornithine.

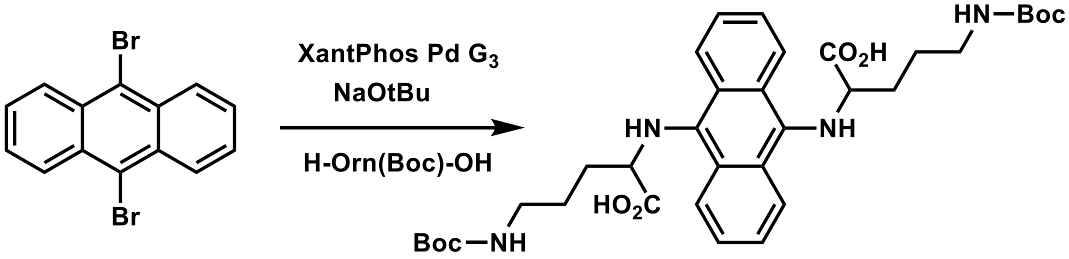
We then tried to make a similar rhodamine-B-derived sensor using an amide coupling with ornithine (**Scheme B.5**).



Scheme B.5 Synthesis of rhodamine-B ornithine sensor.

The rhodamine reaction resulted in a mixture of numerous pink products. The desired compound was present in a very small amount.

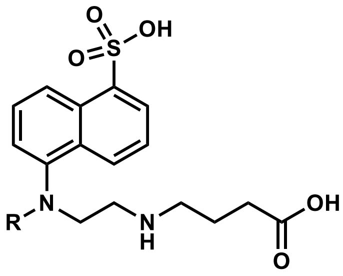
An anthracene-based probe (**Scheme B.6**) was designed that would allow for dual PET-quenching that we hoped would result in a more powerful turn-on:



Scheme B.6 Synthesis of anthracene-ornathine fluorescent probe using a Buchwald-Hartwig amination. Boc deprotection would yield the sensor molecule.

Qualitatively, it seemed that the anthracene construct might have worked. The fluorescence difference before and after Boc deprotection was apparent. However, the material was totally insoluble in water.

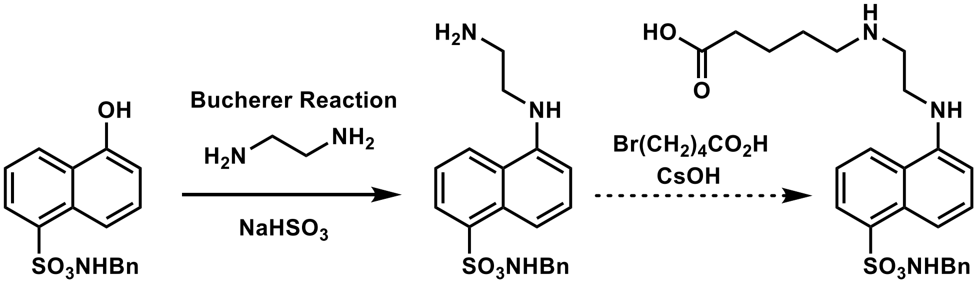
We figured there were two likely problems involved with fluorophore development. Perhaps the insufficient turn-on was from the amine being too distant from the fluorophore, or there might be too weak of a dipole for internal charge transfer. We therefore were interested in another design based on a known PET-quenching probe (**Figure B.4**).[9]



R=Me or $(CH)_2NH(CH_2)_3CO_2H$

Figure B.4 Possible PET-based sensor with dipole separating amine from fluorophore.

We rationalized that this probe could be made with the following reaction sequence (**Scheme B.7**):[10]

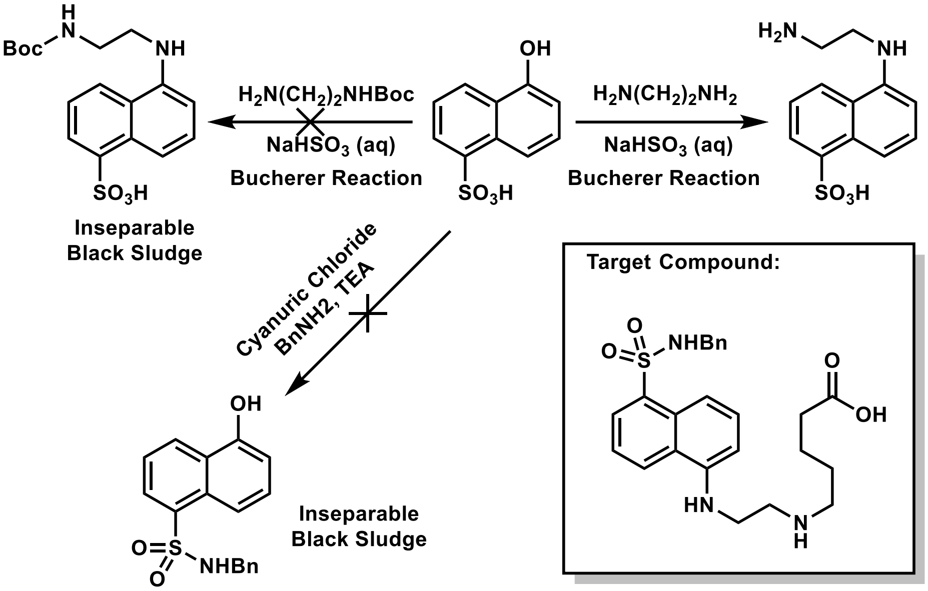


(With CsOH, primary nitrogen is alkylated *once)*

Org. Lett. **1999, *1*(12), 1893–1896**

Scheme B.7 Proposed synthesis of alternative dansyl- sensor.

There were several issues with the synthesis of the target molecule of **Scheme B.7**. These problems are summarized in **Scheme B.8**. Conducting the Bucherer reaction with Boc-protected ethylenediamine resulted in a thick black sludge that could not be tested or purified. Using unprotected ethylenediamine seemed to work without undergoing a second Bucherer substitution. However, the literature precedent to mono-alkylate the amine only once failed with this substrate.[10] Attempting to form the benzyl sulfonamide first also resulted in an inseparable mixture.



Scheme B.8 Problems occurred during the synthesis of the second generation dansyl probe.

Since the Bucherer reaction seemed to work well with some substrates, we designed and synthesized a pyrene-based ‘highlighter molecule’ sensor (**Figure B.5**) using a Bucherer reaction:

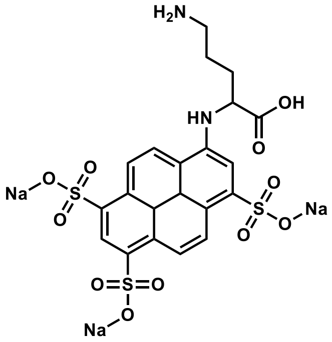
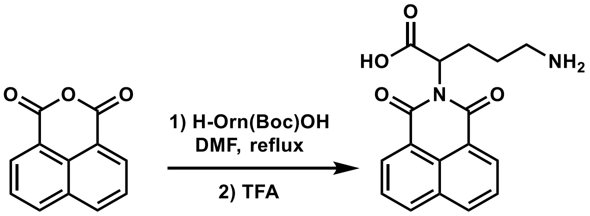


Figure B.5 Possible ornithine-containing fluorescent probe based on the pyrene sulfonic acid used in highlighters.

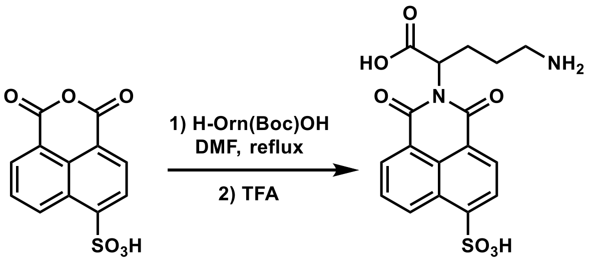
The ‘highlighter molecule’ probe in **Figure B.5** is very soluble in base on account of the multiple sulfonate groups. However, it is completely insoluble in acid.

We also designed several possible fluorescence probes based on naphthalic anhydride and the corresponding imide. Firstly, we tried forming an imide of sidechainprotected ornithine and naphthalic anhydride by refluxing in DMF (**Scheme B.9**). After deprotection, the imide product was very clean. The resulting compound was initially soluble in water, but precipitated upon standing. NMR analysis showed that the precipitate is the imide. The lactam after reaction with CDI was completely insoluble in water.



Scheme B.9 Synthesis of imide-based fluorescent probe.

To address the poor solubility, we synthesized a derivative bearing a sulfonic acid group (**Scheme B.10**). The product is readily soluble. The corresponding lactam was also water-soluble and has a threefold fluorescence turn-on at pH 6-10. However, the lactam degraded into a clumpy precipitate in water after several hours.



Scheme B.10 Synthesis of imide-based fluorescent probe with sulfonic acid.

Our efforts to invent a turn-on fluorescent sensor for amide formation were therefore unsuccessful. Many of our syntheses resulted in inseparable mixtures or products with poor solubility. The lactam compounds that were most promising showed only a modest fluorescence turn-on compared to the open-chain amino acids. An ideal fluorescent probe for future experiments will need to have an obvious fluorescent turn-on upon lactamization (5-fold or better) across a wide range of concentrations and pH.

Sensors quenched via PET have the inherent limitation that they are sensitive to differences in pH. Future designs must overcome all of these challenges.

References

(1) Sharma, A.; Dai, K.; Pol, M. D.; Thomann, R.; Thomann, Y.; Roy, S. K.; Pappas, C. G. Selective Peptide Bond Formation via Side Chain Reactivity and Self-Assembly of Abiotic Phosphates. *Nat. Commun.* **2025**, *16* (1), 1306. https://doi.org/10.1038/s41467-025-56432-6.

(2) Pol, M. D.; Thomann, R.; Thomann, Y.; Pappas, C. G. Abiotic Acyl Transfer Cascades Driven by Aminoacyl Phosphate Esters and Self-Assembly. *J. Am. Chem. Soc.* **2024**, *146* (43), 29621–29629. https://doi.org/10.1021/jacs.4c10082.

(3) Rubio Gomez, M. A.; Ibba, M. Aminoacyl-tRNA Synthetases. *RNA* **2020**, *26* (8), 910–936. https://doi.org/10.1261/rna.071720.119.

(4) Keefe, AnthonyD.; Miller, StanleyL. Are Polyphosphates or Phosphate Esters Prebiotic Reagents? *J. Mol. Evol.* **1995**, *41* (6). https://doi.org/10.1007/BF00173147.

(5) Yuhao, Z.; Chen, Y.; Jiecheng, L.; Yan, L.; Yufen, Z. Prebiotic Template-Directed Peptide Formation Mediated by Polyphosphates and Peptide Promoters. *Org. Biomol. Chem.* **2025**, *23* (36), 8205–8211. https://doi.org/10.1039/D5OB01149C.

(6) Chi, Y.; Li, X.; Chen, Y.; Zhang, Y.; Liu, Y.; Gao, X.; Zhao, Y. Prebiotic Formation of Catalytically Active Dipeptides via Trimetaphosphate Activation. *Chem. – Asian J.* **2022**, *17* (24), e202200926. https://doi.org/10.1002/asia.202200926.

(7) Gan, D.; Ying, J.; Zhao, Y. Prebiotic Chemistry: The Role of Trimetaphosphate in Prebiotic Chemical Evolution. *Front. Chem.* **2022**, *10*, 941228. https://doi.org/10.3389/fchem.2022.941228.

(8) Sibilska, I.; Feng, Y.; Li, L.; Yin, J. Trimetaphosphate Activates Prebiotic Peptide Synthesis across a Wide Range of Temperature and pH. *Orig. Life Evol. Biospheres* **2018**, *48* (3), 277–287. https://doi.org/10.1007/s11084-018-9564-7.

(9) O’Connor, N. A.; Sakata, S. T.; Zhu, H.; Shea, K. J. Chemically Modified Dansyl Probes: A Fluorescent Diagnostic for Ion and Proton Detection in Solution and in Polymers. *Org. Lett.* **2006**, *8* (8), 1581–1584. https://doi.org/10.1021/ol060130y.

(10) Salvatore, R. N.; Nagle, A. S.; Schmidt, S. E.; Jung, K. W. Cesium Hydroxide Promoted Chemoselective *N* -Alkylation for the Generally Efficient Synthesis of Secondary Amines. *Org. Lett.* **1999**, *1* (12), 1893–1896. https://doi.org/10.1021/ol9910417.