

BACKBONE CARBONYL INTERACTIONS IN PROTEINS

by

Robert William Newberry

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2016

Date of final oral examination: May 2, 2016

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

Ronald T. Raines, Professor, Biochemistry

Helen E. Blackwell, Professor, Chemistry

Silvia Cavagnero, Professor, Chemistry

Samuel H. Gellman, Professor, Chemistry

Alessandro Senes, Associate Professor, Biochemistry

ProQuest Number: 10189526

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10189526

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT**BACKBONE CARBONYL INTERACTIONS IN PROTEINS**

Robert W. Newberry

under the supervision of Professor Ronald T. Raines

at the University of Wisconsin–Madison

Proteins are nature’s principal molecular machines, and their many functions arise from their complex and varied three-dimensional structures. Understanding protein structure has therefore been one of the central scientific challenges of the twentieth and twenty-first centuries. Additionally, advances in DNA sequencing technology have far outpaced advances in experimental structural biology, motivating computational methods for predicting protein structures. Personalized medicine would likewise benefit from methods that predict changes in protein structure and function given sequence data. Moreover, a thorough understanding of protein folding should enable the design of new proteins *de novo*, which holds enormous promise for solving technical challenges in fields as diverse as medicine, materials, and energy. Yet, despite decades of research, the methods available for protein structure prediction and design still offer only limited success, suggesting an incomplete understanding of the forces governing protein folding. To address these limitations, I have undertaken a variety of studies that examine the potential contributions of previously unappreciated interactions to protein folding. This dissertation focuses specifically on the role of the carbonyl group in such interactions.

I begin with a foreword that serves as an introduction for general audiences; in it, I outline the fundamental physics of protein folding, thereby providing context for the original research I describe later. In Chapter 1, I review in detail a variety of secondary interactions that have been proposed to contribute to the folding of proteins and evaluate the relative importance of those contributions. The history of research regarding one such interaction in particular, the $n \rightarrow \pi^*$ interaction, becomes the focus of Chapter 2, as it is the subject of a large proportion of the original research presented in this dissertation. Finally, in Chapter 3, I introduce the utility of fluoroproline as probes of noncovalent interactions in proteins.

In Chapter 4, I describe original research that provides fundamental insights into the $n \rightarrow \pi^*$ interaction. In particular, spectroscopic, crystallographic, and computational analyses of proline model compounds bearing thioamides clarifies the nature of this interaction. These studies also provide an approximate lower bound for the energy of a typical interaction and highlight the utility of thioamides for modulating these interactions in peptides and proteins. In light of these insights, I employed thioamides to probe noncovalent interactions in the backbone of the collagen triple helix, which I describe in Chapter 5. The unique structure of collagen allows for selective perturbation of individual interactions, though biophysical analysis of backbone-modified collagen peptides highlights the challenges of studying weak interactions within the complex structure of proteins.

In Chapter 6, I describe computational analyses of polylactide, a polyester that has recently been demonstrated to adopt the same general secondary structure of the collagen triple helix: the polyproline II conformation. My computational analysis demonstrates the influence of ubiquitous $n \rightarrow \pi^*$ interactions in polylactide, and the ability of these interactions to guide structure in the absence of hydrogen bonding not only highlights their relevance, but also

suggests that they might play key roles in the early stages of protein folding. The broad influence of these interactions for protein folding is again underlined by a series of bioinformatics analyses I present in Chapter 7. Specifically, analysis of ultrahigh-resolution protein crystal structures demonstrates that residues in folded proteins bear the distinctive signature of the $n \rightarrow \pi^*$ interaction: carbonyl pyramidalization. I also present evidence that these interactions affect the electronics of the protein backbone. Chapter 8 then describes how an $n \rightarrow \pi^*$ interaction can contribute to the ability of protein to bind to a ligand; specifically, I show that free *N*-acyl homoserine lactones, which mediate quorum sensing in Gram-negative bacteria, engage in an $n \rightarrow \pi^*$ interaction that likely deters binding to their cognate receptors.

Arguably, the central role of the $n \rightarrow \pi^*$ interaction is to provide stability to helical secondary structures in proteins. Why then are β -sheets so common in proteins when they cannot benefit from stabilizing $n \rightarrow \pi^*$ interactions? To address this discrepancy, I performed a series of studies, outlined in Chapter 9, that demonstrate the existence of a common, yet heretofore-unrecognized hydrogen bond within single strands of β -sheets. Specifically, I hypothesize an attraction between the amide proton and carbonyl oxygen of individual amino acid residues in β -sheets, which should bias residues toward the “C5” geometry. Computations suggest that this geometry does permit meaningful hydrogen bonding, and experimental physical organic studies of a preorganized small-molecule model system demonstrate that these interactions bear many of the properties of canonical hydrogen bonds. Spectroscopic studies of backbone-modified peptides indicates that this interaction does operate in β -sheets, and biophysical analyses demonstrate that selective perturbation of this interaction causes changes in the global stability of β -sheets. Finally, bioinformatics analysis of folded proteins demonstrates that these interactions are quite common and can impart significant stabilizing energy. These interactions might also play a

particular role in the early stages of β -sheet folding, as well as in the formation of amyloid structures.

The conclusions and immediate future directions of this research are summarized in Chapter 10. In Appendix A, I explore the detection of $n \rightarrow \pi^*$ interactions in solution using nuclear magnetic resonance spectroscopy. In Appendix B, I characterize solvent effects on the conformations of proline derivatives, the results of which corroborate earlier studies as to the nature of the $n \rightarrow \pi^*$ interaction. Appendix C examines various thioesters as $n \rightarrow \pi^*$ acceptors, which could serve as probes of noncovalent interactions in proteins. Finally, Appendix D presents some examples of $n \rightarrow \pi^*$ interactions between proteins and their ligands.

ACKNOWLEDGMENTS

My success in completing the studies described in this dissertation is due in no small part to a great many individuals whom I would like to take this opportunity to thank.

I would like to begin by thanking my advisor, Prof. Ronald Raines, for his mentorship and guidance through my graduate studies. Ron fosters an environment that supports and values intellectual curiosity and creativity, which has contributed significantly to my independence. He also has a keen eye for clean experiments, and I hope to have developed some similar sense in my tenure at UW. Finally, Ron constantly sets high standards toward which to aspire, promoting excellence across the scientific enterprise.

I am also indebted to an excellent dissertation committee, all of whom have been a constant source of guidance throughout my graduate career. I would like to thank Prof. Helen Blackwell in particular for providing valuable feedback regarding my study concerning homoserine lactones and for helping to advise me through the chemical biology program here at UW. Prof. Silvia Cavagnero served as my minor advisor for the Biotechnology Training Program, through which she weighed in on my studies twice as often as she would have been obligated to otherwise. She also has a great eye for detail and promotes high standards of experimental soundness. I would like to thank Prof. Samuel Gellman for not only being kind enough to recommend me for graduate fellowships, but also for providing great scholarly context to the studies I have undertaken. Finally, I would like to thank Prof. Alessandro Senes, who has provided valuable perspective on the implications and applicability of my insights, and with whom I share many scientific interests.

I have also been very fortunate to work with many great collaborators. Perhaps my most important collaboration has been with Dr. Brett VanVeller, formerly of the Raines laboratory. In addition to helping me transition to an organic-focused laboratory, Brett engaged with me in many an intellectual debate regarding the $n \rightarrow \pi^*$ interaction. Our collaboration has resulted in four separate publications, making it my most important collaboration to date. I am also deeply indebted to Dr. Gail Bartlett and Prof. Derek Woolfson of the University of Bristol for serving as not only as wonderful collaborators, but also as mentors for my studies involving bioinformatics. Indeed, Gail helped me to learn how to perform bioinformatics studies much in the same way Brett taught me to run organic reactions. I believe that my work with Gail and Dek has brought great insight, which is evident in our publication record. I also need to give an emphatic thanks to Dr. Milo Westler, with whom I have worked on a wide variety of characterizations. Milo has taught me an enormous amount of practical NMR spectroscopy and has entertained many of my ideas, some of which have been more successful than others. I would also like to take this opportunity to thank Dr. Amit Choudhary, not only for his guidance on these projects, but for also giving me the opportunity to experience chemical biology research at the Broad Institute.

I have had the opportunity to work with a couple of very talented undergraduates, who have each helped to propel my research forward. First, I would like to thank Sam Orke, who has undertaken a wide variety of projects, including some that might have lacked proper precedent. Sam has proven resilient, energetic, and enthusiastic, making him a congenial coworker. Our studies on the interplay between $n \rightarrow \pi^*$ interactions and hydrogen bonds are primarily his achievement. I would also like to thank Kelly Wallin for her dedicated, independent pursuit of the cooperativity of $n \rightarrow \pi^*$ interactions. Kelly has been an important presence in our laboratory, and I am grateful to have had the chance to work with her.

I would like to thank the late Prof. Grant Krow of Temple University for his advice regarding proline model compounds. I thank Prof. James Fraser of the University of California, San Francisco for his help with characterization of the effects of $n \rightarrow \pi^*$ interactions on protein electronics. I thank Prof. Matthew Cremeens of Gonzaga University, with whom I collaborated on the infrared spectroscopy of proline aldehydes. I thank Prof. Bobby Arora of New York University for sharing with me his aldehyde capture auxiliaries and for advising me on issues related to interactions with α -helices. I would like to thank Prof. Katrina Forest and her associate, Dr. Kenneth Satyshur, for investigating the potential utility of restraints based on noncovalent interactions for the refinement of X-ray diffraction data of proteins. I thank Prof. Katherine Henzler-Wildman for helping me to apply solid-state NMR to detection of the $n \rightarrow \pi^*$ interaction. I would like to thank Prof. Thomas Record for advising me on thermodynamic experiments to characterize intermolecular interactions. I also thank Profs. Frank Weinhold and Clark Landis for helping me with the more sophisticated analyses available through NBO.

I have benefitted from a wealth of technical assistance during my time as a graduate student. I would like to thank Dr. Ilia Guzei and his associates at the UW Molecular Structure Laboratory, Brian Dolinar and Lara Spencer, for their assistance with X-ray diffraction. I would like to thank Dr. Craig Bingman of the Phillips laboratory for his efforts to solve the crystal structure of our backbone-modified collagen peptides. I would like to thank Dr. Desiree Bates for extensive advice regarding computations. I thank Dr. Melissa Boersma and Nina Porcaro of the UW Biotechnology Center for assistance with peptide synthesis. I would like to thank the Zanni laboratory for use of their peptide synthesizer, and specifically Tianqi Zhang for assistance with IR spectroscopy. I would also like to thank the National Magnetic Resonance Facility at

Madison, the UW Chemistry Mass Spectrometry Facility, and the Biophysics Instrumentation Facility for technical support.

I have benefitted substantially from a number of professional development opportunities at UW that have helped enable my success. I would like to recognize the DELTA program, and specifically Dr. Cheri Barta and Ryan Pakula, for their training in mentorship. I am particularly thankful to have been involved with the HHMI Teaching Fellows program at WISCIENCE, which has been central to my development as an educator. I would particularly like to thank Dr. Lillian Tong, Dr. Kristen Jenkins, Dr. Ruth Sullivan, and Prof. Ahna Skop for their guidance, as well as the other teaching fellows, Aayushi, Elaina, and especially Grace for being such great colleagues and for showing me so much about what excellent teaching can be. I would also like to thank Prof. Alan Attie for his mentorship in instructional approaches.

The Raines laboratory has been a fantastic place to work, learn, and grow over the past several years, and I am grateful to have had such wonderful colleagues. In addition to those already mentioned, I would like to thank Matt Aronoff and Thom Smith, with whom I have shared so much of this journey; undoubtedly one of the most important learning opportunities of my graduate career has been navigating it with these two gentlemen. I would also like to thank Kalie Mix, not only for tolerating me as a bay-mate, but for taking such constant interest in my science. Aubrey Ellison rounds out the core of the Raines lab chemists, and she is a wonderful coworker who does much to keep the lab functional. I would like to thank Dr. Caglar Tanrikulu for always asking the tough questions and never believing a word I say. I would like to thank Kristen Andersen and Trish Hoang not only for being my resident advisors on biology-related issues, but also for being a constant source of support. I would also like to thank James Vasta,

John Lukesh, Chelcie Eller, Joelle Lomax, Val Ressler, and Christine Bradford, among others, for making the lab such a fun place to work.

I owe my ability to succeed in the high caliber research environment at UW in great part to my former research advisors, Profs. Eric Anslyn, Lauren Webb, Eva Zubía, Henry Strobel, and Donovan Haines. Eric and Lauren in particular have been constant sources of guidance, and in fact, it was the intersection of my experiences in their labs that led me to undertake my chosen graduate projects. I would also like to thank Christina Krueckl for her support of my career in science.

I would like to close by acknowledging my friends and family for their support. I am very fortunate to have many friends from every stage of my life, and among them are some truly exceptional individuals who constantly motivate me to succeed at the highest level. I draw particular inspiration from Vicki, Sarah, Carissa, Rebecca, Shirin, Kate, Melissa, Danny, Brandon, Sean, Marlies, and many, many others. My family has been incredibly supportive throughout my Midwestern sojourn, and I am particularly grateful to Joe, Mary, Patty, John, Alex, Rachel, Meg, Ian, and Elizabeth for being there for me. I would also like to thank my brother Ben for putting up with me all these years. Finally, I want to thank my parents, Rose and Bill, who have given so much of themselves to make my life what it is. If I am anything good in this world, it is because of them.

TABLE OF CONTENTS

Abstract	i
Acknowledgments	v
Table of Contents	x
Table of Figures	xv
Foreword: Introduction for the General Public	xviii
Chapter I	
Secondary Forces in Protein Folding	1
Dominant Forces in Protein Folding	2
Current Limitations	5
Secondary Interactions of the Main Chain	6
Secondary Interactions Involving Side Chains	11
Relative Contributions of the Secondary Forces	16
Other Contributions	17
Outlook	18

Chapter II

The $n \rightarrow \pi^*$ Interaction	21
Conspectus	21
Introduction.....	23
Nature of Carbonyl Interactions.....	23
Contributions of $n \rightarrow \pi^*$ Interactions to Protein Structure	28
Contributions of $n \rightarrow \pi^*$ Interactions to Other Polymers	34
Contributions of $n \rightarrow \pi^*$ Interactions to Small Molecules	36
Outlook	38

Chapter III

4-Fluoroprolines: Conformational Analysis and Effects on Protein Stability and Folding.....	39
Introduction.....	40
Synthesis	41
Conformational Analysis	44
Effects on Collagen Stability	50
Effects on Peptide Structure.....	55
Effects on Protein Folding	61
Impact in Medicinal Chemistry	70
Outlook	73

Chapter IV

$n \rightarrow \pi^*$ Interactions of Amides and Thioamides: Implications for Protein Stability	75
Introduction.....	76
Results.....	78
Conclusions.....	82
Methods.....	83

Chapter V

Thioamides in the Collagen Triple Helix.....	131
Introduction.....	132
Results.....	133
Conclusions.....	141
Methods.....	142

Chapter VI

$n \rightarrow \pi^*$ Interactions in Poly(lactic acid) Suggest a Role in Protein Folding	159
Introduction.....	160
Results.....	162
Conclusions.....	164

Chapter VII

Signatures of $n \rightarrow \pi^*$ Interactions in Proteins	167
Introduction.....	168
Results.....	169
Conclusions.....	174
Methods.....	175

Chapter VIII

A Key $n \rightarrow \pi^*$ Interaction in <i>N</i> -Acyl Homoserine Lacones	179
Introduction.....	180
Results.....	182
Conclusions.....	186
Methods.....	187
Addendum.....	201

Chapter IX

Backbone C5 Hydrogen Bonds in Proteins	209
Introduction.....	210
Computational Analysis of C5 Hydrogen Bonding	211
Experimental Analysis of Individual C5 Hydrogen Bonds	213

C5 Hydrogen Bonding in β -Sheets	217
C5 Hydrogen Bonding in Proteins	221
Conclusions	225
Methods	226
Addendum	263
Chapter X	
Conclusions and Future Directions	289
Appendix A	
Correlation of $n \rightarrow \pi^*$ Interactions with NMR Observables	303
Appendix B	
Solvent Effects on Intermolecular Interactions in Proline Residues	309
Appendix C	
Thioesters: Alternative $n \rightarrow \pi^*$ Acceptors	325
Appendix D	
$n \rightarrow \pi^*$ Interactions between Proteins and their Ligands	335
References	341

TABLE OF FIGURES

Figure 1.1 C–H···O Hydrogen bonds in the backbone of the β -sheet.	8
Figure 1.2 $n \rightarrow \pi^*$ Interactions between adjacent carbonyl groups in the peptide backbone.	9
Figure 1.3 C5 Hydrogen bond in the peptide backbone.	10
Figure 1.4 Secondary interactions in the protein backbone.	11
Figure 2.1 Characterization of carbonyl interactions by torsion balance analysis.	24
Figure 2.2 Overlap of the n and π^* orbitals of <i>N</i> -acetyl proline dimethyl amide.	25
Figure 2.3 Geometric parameters characterizing an $n \rightarrow \pi^*$ interaction.	26
Figure 2.4 Orbital interactions between carbonyl groups.	27
Figure 2.5 Conformational preferences of substituted prolines.	29
Figure 2.6 Carbonyl lone pairs in the α -helix.	33
Figure 2.7 $n \rightarrow \pi^*$ Interactions in synthetic polymers.	35
Figure 2.8 Selection of small molecules reported to engage in $n \rightarrow \pi^*$ interactions.	37
Figure 3.1 Approaches to the chemical synthesis of fluoroproline.	42
Figure 3.2 Conformations of proline rings.	45
Figure 3.3 Stereoelectronic effects in fluoroproline.	46
Figure 3.4 Conformational preferences of fluoroproline.	49
Figure 3.5 Biosynthesis of proline from glutamate.	63
Figure 3.6 Protein crystal structures containing fluoroproline.	66
Figure 4.1 Example of an $n \rightarrow \pi^*$ interaction in a proline model system.	76
Figure 4.2 Compounds used to evaluate $n \rightarrow \pi^*$ interactions in torsion balance analyses	77
Figure 4.3 Characterization of $n \rightarrow \pi^*$ interactions in proline model compounds.	78

Figure 4.4	Conformational descriptors of proline model compounds.	80
Figure 4.5	Carbonyl pyramidalization in peptides containing thioamide bonds.	81
Figure 5.1	Conformation and intermolecular interactions of the collagen triple helix.....	132
Figure 5.2	Space-filling models of thioamide substitutions in the collage triple helix.	134
Figure 5.3	Synthesis of thioamide-substituted CMPs.....	136
Figure 5.4	CD spectra of thioamide-substituted CMPs.	137
Figure 5.5	CD spectra of thioamide-substituted CMPs at elevated temperatures.	138
Figure 5.6	Thermal denaturation of triple-helical CMPs monitored at various wavelengths....	139
Figure 5.7	Thermal denaturation of triple-helical, thioamide-substituted CMPs.	140
Figure 6.1	Synthesis of PLA.....	160
Figure 6.2	$n \rightarrow \pi^*$ Interactions in PLA.....	161
Figure 7.1	An $n \rightarrow \pi^*$ interaction in the protein backbone and its structural signature.	168
Figure 7.2	Calculated structural and electronic effects of $n \rightarrow \pi^*$ donation.	170
Figure 7.3	Pyramidalization of backbone carbonyl groups in proteins.	172
Figure 7.4	Electronic polarization of backbone carbonyl groups in proteins.	173
Figure 8.1	Structure of an <i>N</i> -acyl homoserine lactone and a proline residue.	180
Figure 8.2	Structural parameters characterizing $n \rightarrow \pi^*$ interactions in AHLs.....	181
Figure 8.3	Structures of AHLs.....	182
Figure 8.4	Molecular structure of <i>N</i> -(3-oxobutanoyl)-L-homoserine lactone.....	202
Figure 8.5	Packing of <i>N</i> -(3-oxobutanoyl)-L-homoserine lactone.	203
Figure 8.6	Structural parameters describing an $n \rightarrow \pi^*$ interaction.	203
Figure 8.7	Overlap of oxo-AHL amide lone pair (n) and ester π^* orbitals.	204
Figure 8.8	Overlap of oxo-AHL ketone lone pair (n) and amide π^* orbitals.	205

Figure 9.1 Carbonyl lone pairs in the peptide backbone of β -sheets,	210
Figure 9.2 Frequencies of C5 hydrogen-bond donor–acceptor distances in β -sheets.....	211
Figure 9.3 Computational characterization of AcGlyNHMe conformations.....	212
Figure 9.4 Conformational analysis of AcDegNHMe and AcDegOMe.....	214
Figure 9.5 C5 Interactions display properties typical of hydrogen bonds.	215
Figure 9.6 Fourier transform infrared spectra of AcDegNHMe and AcDegNH ₂	216
Figure 9.7 ¹ H NMR spectra of AcDegNHMe and AcDegOMe.	216
Figure 9.8 Perturbation of C5 hydrogen bonds impacts β -sheet stability.....	218
Figure 9.9 Far-UV circular dichroism spectra of TrpZip peptides.	219
Figure 9.10 Energy and frequency of C5 hydrogen bonds in proteins.	221
Figure 9.11 C5 hydrogen-bond distances in parallel and antiparallel β -sheets.	222
Figure 9.12 Solvent accessibility of residues engaged in C5 hydrogen bonds.....	222
Figure 9.13 Relationship of C5 hydrogen-bonding energy with β -sheet content in proteins....	223
Figure 9.14 Ramachandran plots of residues engaged in C5 hydrogen bonds.....	224
Figure 9.15 Relative frequency of amino acid residue-type engaged in C5 hydrogen bonds. ..	224
Figure 10.1 Two-dimensional infrared spectrum of AcProNMe ₂	291
Figure 10.2 Stapled α -helical turn for assaying contributions from $n \rightarrow \pi^*$ interactions.....	294
Figure 10.3 Helical peptide scaffold for assaying contributions from $n \rightarrow \pi^*$ interactions.....	295

FOREWORD

Decoding Protein Languages:

An Introduction for the General Public

Living organisms are organized into cells, each of which is capable of performing the basic processes of life: acquisition of nutrients and energy, reproduction, and self-defense. Within the cell, each of these activities is orchestrated by a team of complex machines called proteins. Like all molecules, proteins are particular combinations of atoms bonded together; however, unlike other molecules, such as water or carbon dioxide that have only a few component atoms, proteins can have up to hundreds of thousands of atoms each. The cell produces these enormous molecules by linking together smaller molecules, called amino acids, into long chains. Nature makes use of twenty specific amino acids, and together, these twenty building blocks provide all of the pieces necessary to create the 20,000 or so different proteins that enable life.

Even though proteins are constructed by linking amino-acid building blocks into long single chains, proteins do not look like long strings in the cell. Rather, those chains collapse down upon themselves into condensed particles. We call this process *protein folding* (Figure 1).

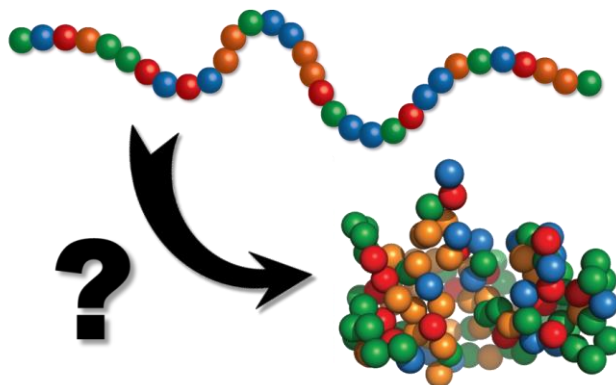


Figure 1. Proteins are chains of chemical building blocks that fold up into condensed particles.

While the final arrangement of the atoms may appear random, it is actually highly evolved to perform a specific function. Just like IKEA furniture, it's not enough for a protein to contain all of the correct components; those building blocks have to be correctly arranged, or the protein will not function. Even worse, sometimes when proteins fold into the wrong structure they cause disease. This is precisely what happens in Alzheimer's and Parkinson's diseases; specific proteins in the brain are prone to misfolding, and when they do, they obstruct neural activity. So what makes some proteins fold into the correct structure and what makes others go awry?

Unlike furniture, protein chains arrange themselves almost entirely on their own. Even if you manage to stretch a protein out into a single chain, it will usually go back the exact same, correct structure once you stop stretching it. It's as if, somehow, each of the pieces of the protein are communicating with one another so that each one knows exactly where to go. So, if we want to understand why proteins fold and misfold, we need to learn how to speak their language, or more precisely, languages. The trouble is that proteins are more like Europe than they are like America; many different languages are spoken, everyone seems to be able to speak multiple languages, some groups that speak one language like to cluster together, other groups like to reside in areas where lots of languages are spoken, and to top it all off, there are many forms of nontraditional communication that we do not fully comprehend. Understanding how each specific protein is able to arrange itself so perfectly has therefore been one of the greatest scientific challenges of the twentieth and twenty-first centuries, and it is a challenge I decided to pursue during my doctoral studies.

Over the course of several decades, the scientific community has identified several ways in which the amino acids on a protein communicate or interact with one another; these interactions depend on the chemical groups in each amino acid building block (Figure 2).

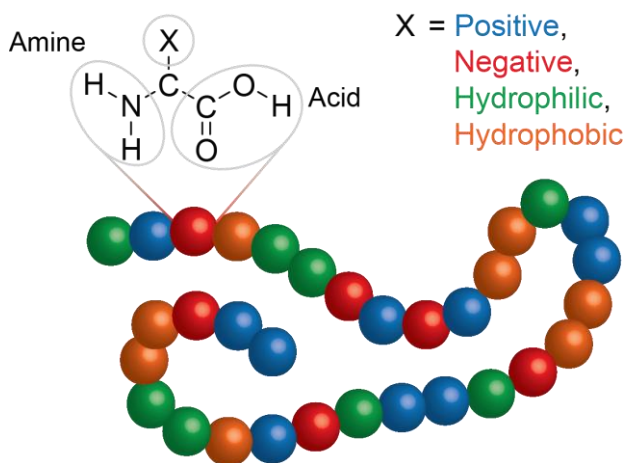


Figure 2. Proteins are comprised of building blocks called amino acids, each of which have different physical properties.

For example, several amino acids have areas of either positive or negative charge. As a result, amino acids with positive charge will form attractions with amino acids that have negative charge while avoiding other positively-charged amino acids. Another important interaction that occurs between amino acids involves water. Proteins are constantly bathed in water inside of the cell, but different amino acids have different affinities for water. Some amino acids form very favorable interactions with water, while others much prefer to avoid water. As a result, the “hydrophobic” amino acids will tend to congregate at the interiors of proteins, while “hydrophilic” amino acids will decorate the outside surface of proteins. Moreover, each of the different amino acids is a slightly different size and shape, so some combinations of amino acids can fit easily next to each other, akin to pieces of a jigsaw puzzle, whereas other combinations may not be able to.

However, despite everything we have learned about how amino acids interact with one another, we still do not understand how one protein folds into one structure while a different protein will fold into a different sequence. For example, the scientific community has a very low success rate in predicting the structure of a protein given information about which amino acids it

contains. Moreover, many scientists have tried to design new protein structures to solve important problems, but there have been very few successes; the sequences that scientists propose for new proteins seldom adopt the structures they want. One could envision designing new proteins that treat disease by acting as microscopic surgeons or proteins that help us to produce fuels from renewable resources by breaking down plant matter. Being able to design new proteins could aid in solving a wide variety of challenges, but doing so will require us to fully understand how each amino acid interacts with others in the chain, which has motivated an entire field of science devoted to the study of protein folding.

The atoms in proteins communicate entirely through physics, which means that a thorough understanding of physics is essential to understanding proteins. Unfortunately, most scientists who study proteins only understand classical physics, which treats atoms as small balls of fixed charge. While classical physics works well for large objects like pendulums or roller coasters, it breaks down for small objects like atoms. Pioneering experiments in the early twentieth century showed that atoms do not behave as one would expect for simple charged balls; rather, studies on the negatively-charged electrons that surround the positively-charged atomic nucleus showed that electrons have curious, wave-like properties. The most important wave-like property of electrons is that they can interfere with one another. Consider, for example, what happens when two cars pass one another versus when two pond ripples pass one another. Whereas the cars simply run by each other without any noticeable interaction, the two pond ripples will cause dramatic changes in one another as they go by. In order to account for these previously unappreciated properties, theoretical physicists developed a new model for the physics of small particles, which was later termed quantum mechanics. The development of this model was accompanied by several Nobel prizes, and it stands as one of the most successful theories in the

history of physics. I am one of few scientists studying proteins who can understand quantum mechanics, and I have been using that understanding to dissect proteins in ways that no one ever has before.

So how does quantum mechanics affect proteins? Again, like all molecules, proteins are comprised of atoms, each of which is surrounded by a cloud of electrons. Some of those electrons create the chemical bonds that hold the atoms together. Other electrons, however, are not involved in bonds, which leaves them free to form other interactions. What interactions they can form, depends of course on where they are located and what else is nearby. But one cannot accurately predict the locations of those “non-bonded” electrons without invoking the ideas of quantum mechanics. Appreciating the correct locations of non-bonded electrons leads to the discovery of previously unknown interactions that help to hold proteins into their correct topology. My dissertation concerns two examples of those previously unknown interactions.

One common chemical group in proteins is the carbonyl group (Figure 3). It consists of an oxygen atom that is chemically bonded to a carbon atom through two different electron pairs. Because the two atoms share two pairs of electrons, they are said to be “double-bonded.” Carbon and oxygen both prefer to have or share a total of eight electrons. Carbon atoms in carbonyl groups will often attain those remaining four electrons by forming two additional bonds to other atoms. Oxygen, on the other hand, generally prefers not to form more bonds, so it keeps its remaining four electrons largely to itself. However, because they are not involved in bonds, those remaining, non-bonded electrons can interact with other nearby atoms, though these interactions are much weaker than chemical bonds. The relatively weak interactions that the non-bonded electrons can form are therefore perfect for helping a protein to fold; they provide the metaphorical glue for protein structures.

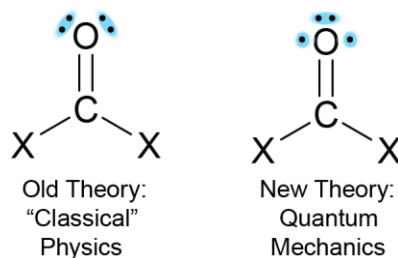


Figure 3. The carbonyl group features four non-bonded electrons (blue), but physical theories predict that they are in different locations.

For the carbonyl group, the location of the “non-bonded” electrons has been the subject of debate. Originally, the “non-bonded” electrons were predicted to be evenly spaced around the oxygen atom (Figure 3, left). This model implied that both pairs of electrons are identical; however, an experiment showed that those electrons actually have two different energies, proving that they cannot be the same. To account for this new information, chemical physicists performed sophisticated quantum mechanical calculations that showed a very different arrangement of non-bonded electrons (Figure 3, right).

This new model of the non-bonded electrons is essential for proteins because it shows where attractive interactions can occur within proteins. For example, one type of attraction non-bonded electrons can form is with a hydrogen atom. Hydrogen atoms are often slightly positively-charged, which makes them perfect partners for non-bonded electrons. The interactions between hydrogens and non-bonded electrons are termed *hydrogen bonds* because they draw the hydrogen atom close to the atom bearing the non-bonded electrons. In proteins, many hydrogen atoms are bonded to nitrogen atoms. If the hydrogen atom in this N–H group approaches a carbonyl oxygen, it will form an attraction; however, the angle at which the N–H and the carbonyl approach each other depends on the orientation of the non-bonded electrons on oxygen. For example, the two most common patterns in proteins, the alpha-helix (α -helix) and the beta-sheet (β -sheet), both form interactions that approach the carbonyl group along the C=O bond

(Figure 4). If you were only considering the old model for the location of the non-bonded electrons (Figure 3), the orientation of these hydrogen bonds would not make much sense. This shows why it is important for protein scientists to appreciate modern physics.

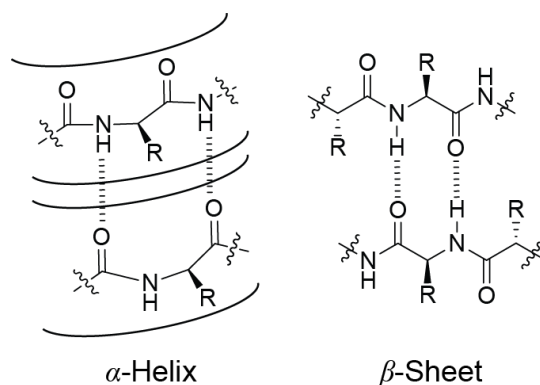


Figure 4. The α -helix and the β -sheet are the two most common structural patterns in proteins, but both feature hydrogen bonds that form along the carbonyl bond axis. Carbon atom labels are omitted for clarity.

We can also use quantum mechanics to predict other interactions that occur within proteins that no one has seen before. Let us examine the β -sheet (Figure 4, right). Notice that the carbonyl oxygen is actually close to two hydrogen atoms, one of which forms a hydrogen bond that was known previously, shown as short dashes. The fact that the oxygen is also close to a second hydrogen suggests that it may form a second attractive interaction. However, no one ever thought that the second hydrogen would form an interaction because they were mistaken about where the non-bonded electrons are. Without knowing that there are non-bonded electrons between the carbonyl oxygen and the adjacent hydrogen, no one would have expected there to be an interaction. Based on what I know about quantum mechanics, I was the first to predict that those two atoms would attract one another (Figure 5).

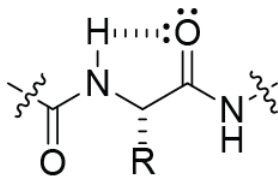


Figure 5. A new type of hydrogen bond that uses a second pair of non-bonded electrons from the carbonyl group. Carbon atom labels are omitted for clarity.

While this prediction seems reasonable, it is difficult to discern whether the C=O and N–H groups actively attract one another or if they are simply held close together by chemical bonds.

I therefore needed to determine if that interaction really is attractive. If it is, it should help to hold the β -sheet together, and I can measure how strongly the β -sheet is held together using a variety of methods. To test if the second hydrogen bond really does hold β -sheets together, I designed a β -sheet in which the second hydrogen bond, if it exists, would be weaker. To make the putative hydrogen bond weaker, I distracted the non-bonded electrons with a decoy interaction in the opposite direction. This should weaken the hydrogen bond and destabilize the β -sheet. There is only one problem with that strategy: if I distract the non-bonded electrons from engaging in this new hydrogen bond (horizontal in Figure 6), I will also distract them from forming the original hydrogen bond (vertical in Figure 6), which is already well-established. So, if I see change in the stability of the β -sheet, I cannot tell if it is due to weakening one hydrogen bond or the other.

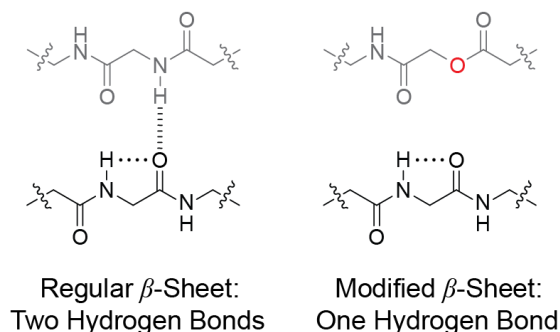


Figure 6. Oxygen atoms in β -sheets can form multiple hydrogen bonds. To probe my newly proposed hydrogen bond (horizontal dashes), I eliminated the original hydrogen bond by introducing an oxygen (red).

To get around that problem, I eliminated the original hydrogen bond by changing the N–H group to an oxygen (red in Figure 6). Unlike nitrogen, which prefers to make three bonds, oxygen prefer to form just two, which eliminates the hydrogen that forms the hydrogen bond. That leaves just the second hydrogen bond that I want to probe. When I then made these modified β -sheets, I found that distracting the non-bonded electrons away from the new hydrogen bond made the whole β -sheet less stable. That means that my new hydrogen bond really is attractive. Nobody knew that before this experiment. This might be one of the reasons that scientists cannot design new proteins: we have not yet been accounting for all of the interactions that hold proteins together. In fact, this second type of hydrogen bonds is very common: it happens in approximately 20% of β -sheets, meaning that there are a great many interactions that we have yet to account for.

The new picture of carbonyl lone pairs in Figure 3 has also inspired the discovery of another previously-unrecognized interaction. Non-bonded lone pair electrons on a carbonyl oxygen can also form an attraction to the carbon of another carbonyl group (Figure 7). Like hydrogen atoms, carbon atoms tend to have positive charge, so they can interact with the negatively-charged electrons on oxygen. These interactions have been named $n \rightarrow \pi^*$ interactions (read “n to pi star”),

n representing the non-bonded electrons and π^* representing the carbonyl carbon that the non-bonded electrons interact with.

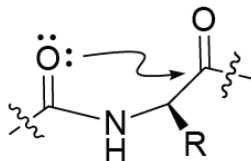


Figure 7. An attraction between the non-bonded electrons of a carbonyl oxygen and an adjacent carbonyl carbon, termed an $n \rightarrow \pi^*$ interaction. Carbon atom labels are omitted for clarity.

Proteins have an enormous number of carbonyl groups, meaning that they can potentially engage in numerous $n \rightarrow \pi^*$ interactions. In fact, based on how close the atoms have to be to engage in an $n \rightarrow \pi^*$ interaction, we predict that approximately one-third of all residues in folded proteins will engage in an $n \rightarrow \pi^*$ interaction. To test this hypothesis and determine if proteins really do engage in this interaction, I examined protein structures in detail, looking for signatures of the $n \rightarrow \pi^*$ interaction. Because the non-bonded electrons pull on the adjacent carbon atom, they pull it away from the atoms it is bonded to. We can see this effect by looking at how flat the carbonyl groups are; in the carbonyl group, the central carbon and all of the atoms to which it is bonded lie in one plane. So, when the non-bonded electrons pull on the central carbon of the carbonyl group, they pull that carbon out of the plane, and we can measure that distortion (Figure 8).

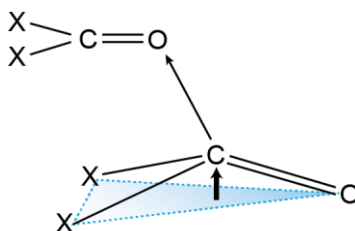


Figure 8. As non-bonded electrons pull on a carbonyl carbon, they pull it out of the plane formed by the atoms to which it is bonded.

When I looked at how flat the carbonyl groups in proteins are, I found that those carbonyl groups that were close to non-bonded electrons were usually less flat than carbonyl groups that were far away from non-bonded electrons. This result shows that the $n \rightarrow \pi^*$ interaction does occur in proteins, and in fact, occurs very commonly. Again, protein scientists, who do not usually appreciate modern physics, do not currently account for these interactions when predicting what a particular protein will look like. If we were to include the $n \rightarrow \pi^*$ interaction in our predictions, perhaps we will be better able to design our own proteins.

While the data I have generated so far demonstrate that $n \rightarrow \pi^*$ interactions are present in proteins, it is not yet clear what they do. To help determine what roles they might play, I developed technology that allows us to manipulate these interactions. Specifically, I substitute the oxygen in the carbonyl group with sulfur. Sulfur is directly below oxygen on the periodic table, so it forms most of the same bonds that oxygen does. However, because sulfur is larger than oxygen, its non-bonded electrons are farther from the nucleus, so they are not held as tightly. As a result, the non-bonded electrons in sulfur are more likely to go interact with other atoms. Replacing carbonyl oxygens with sulfur therefore enhances the $n \rightarrow \pi^*$ interaction. So, if we want to understand what an $n \rightarrow \pi^*$ interaction does in a protein, we can substitute the carbonyl oxygen with sulfur and look for what changes.

Based on the studies I have completed, we have learned much about how atoms communicate to coordinate protein structure. In particular, I identified a previously-unrecognized interaction, or “language,” that turns out to be very common in proteins. Still, proteins harbor many secrets that we have yet to unearth, but the results of my studies and the technology I have developed bring us one step closer to being able to control proteins and design new ones, with potentially revolutionary implications.

CHAPTER I

Secondary Forces in Protein Folding

Proteins are the principal molecular machines of the cell, capable of a myriad of activities that enable life. Each individual protein derives its function from the unique, three-dimensional arrangement of its chemical components.¹ Seminal experiments by Anfinsen² demonstrated that amino-acid sequences can contain all of the chemical information necessary to specify a particular, thermodynamically stable structure.³ That result implies that decoding the chemical information present within the polypeptide chain would, in principle, allow one to predict the structure from its sequence alone. Given the number of putative protein sequences generated from DNA sequencing data,⁴ such technology would prove invaluable for addressing many biological questions. Moreover, the same insight could eventually allow physicians to predict the effects of particular mutations, empowering personalized medicine; this prospect could be especially important for the treatment of diseases caused by protein misfolding.⁵ Finally, a complete understanding of the factors governing protein structure could be leveraged toward the design of new proteins with emergent functions, the limits of which are hard to conceive. Understanding the molecular basis for protein structure has thus become one of the central biophysical challenges of our age.

Dominant Forces in Protein Folding

We now appreciate that the folding of polypeptide chains is encouraged by key noncovalent interactions,⁶ excepting the formation of disulfide bridges. Heretofore, modeling of proteins has focused on a handful of interactions that are now accepted as canonical contributors to protein folding: the hydrophobic effect, van der Waals interactions, electrostatics, and hydrogen bonding.

Among these interactions, the hydrophobic effect has been shown to be the dominant contributor to the folding of globular proteins,⁷⁻⁸ where the release of ordered water molecules due to the burial of hydrophobic surface provides stability by increasing the overall entropy of the system. Various estimates suggest that over half of the thermal stability of proteins can be attributed to the hydrophobic effect.⁹ It is important to note, however, that the hydrophobic effect is relatively nonspecific; to a first approximation, hydrophobic surfaces discriminate between one another on the basis of sterics alone. Despite the lack of chemical specificity, steric interactions at hydrophobic interfaces can provide important specificity to protein folding; notably, the hydrophobic cores of proteins are found to be especially dense, even compared with crystalline solids, indicating substantial geometric complementarity in proteins that may specify particular structures.¹⁰ Moreover, in the arena of coiled coils, interfacial hydrophobic amino acids specify both oligomerization state and binding partners¹¹ due to so-called “knobs-into-holes” packing.¹² Continuing research on the hydrophobic effect attempts, among other things, to delineate the length scales on which the effect operates,¹³ as well as to refine the energy associated with burial of hydrophobic surface, estimates of which vary widely between systems and analytical approaches.¹⁴

Other nonspecific interactions in proteins include van der Waals interactions, such as the London dispersion forces that exist between the fluctuating induced dipoles of molecular electron clouds.¹⁵ These interactions become attractive only at short distances; for example, the Lennard–Jones potential indicates that the strength of attraction varies with an inverse-sixth dependence on the internuclear separation.¹⁶ At even shorter distances, these interactions become repulsive due to the interactions of filled electronic orbitals, which is referred to as the exchange repulsion or Pauli repulsion term. Ultimately, the degree of attraction depends on electronic polarizability, which in turn is related to atomic and molecular weight. Thus, though individually weak, these interactions can become significant for larger hydrocarbons, even to the point that they overcome seemingly insurmountable steric clashes,¹⁷ and thereby provide stabilizing energy to the hydrophobic cores of globular proteins.

Whereas nonspecific interactions are key for providing stability to proteins, the correct arrangement of chemical groups is essential for protein function and relies on the formation of specific interactions. The simplest class of specific interactions are the electrostatics. Experimental measurements indicate that typical, solvent-exposed salt bridges can impart 0.5–1.5 kcal/mol of stability each.¹⁸⁻¹⁹ Moreover, electrostatic interactions need not involve full charges, but may also involve the partial charges of permanent electric dipoles.²⁰ These dipolar interactions contribute to the formation of secondary structures²¹⁻²³ and lead to the macroscopic helix dipole moment,²⁴ though this notion is controversial.²⁵ Electrostatic interactions are particularly important for protein folding because they have relatively weak distance dependence,²⁶⁻²⁷ making them relevant at longer length scales and key determinants of nonlocal contacts and complex formation.²⁸⁻²⁹ The interaction of classical charges can often be described in a relatively simple manner by assuming that charges can be ascribed to infinitely small regions

of space. Nonetheless, there are two main complications in calculating the energy of electrostatic interactions. The first is estimating the dielectric constant of the medium, which for proteins has been the subject of considerable debate,³⁰ due in part to its heterogeneity. The second challenge is accounting for molecular polarization due to nearby charges. The latter complication has motivated the development of polarizable force fields, such as AMOEBA.³¹⁻³²

Perhaps even more important for protein structure than electrostatics is the hydrogen bond, which has been the subject of intense scrutiny. One of the first questions addressed by researchers was whether or not the hydrogen bond provides thermodynamic stability. Because proteins can alternatively form hydrogen bonds between peptide groups or to water, it was hypothesized that hydrogen bonds would not contribute significantly to the stability of the folded state. Indeed, this hypothesis has been supported spectroscopically.³³ There is also evidence, however, that many common types of hydrogen bonds in proteins, particularly amide–amide hydrogen bonds, are somewhat stronger than hydrogen bonds to water.³⁴ Perhaps the most convincing evidence for stabilization of the folded state by the hydrogen bond is the observation that polyalanine peptides form an α -helix³⁵ whose stability increases with increasing length.³⁶ Indeed, this paradigm was leveraged to provide accurate estimates of the energetics of individual hydrogen bonds. Conventional wisdom now holds that hydrogen bonds between peptide groups are slightly stabilizing to the folded state, and have expected energies on the order of one kilocalorie per mole.³⁴ Moreover, native protein structures are highly optimized to make complete use of their hydrogen bonding potential.³⁷

Hydrogen bonding energy can be substantially higher in the interior of a protein, where the dielectric constant is reduced and fewer competing donors and acceptors are present; indeed, membrane proteins are believed to benefit from enhanced hydrogen bonding within phospholipid

bilayers.³⁸ The dependence of hydrogen-bond energy on the dielectric of the environment has also raised the question as to the nature of the hydrogen bond. Such a dependence on environmental dielectric suggests a largely electrostatic contribution to the energy of a hydrogen bond. Yet, calculations,³⁹⁻⁴¹ NMR spectroscopy,⁴² and scattering experiments⁴³ have all shown important charge-transfer character to the hydrogen bond,⁴⁴ and accounting for this charge-transfer character has been shown to improve molecular force fields.⁴⁵ Regardless of their nature or stabilizing influence, hydrogen bonds are known to direct the folding of the polypeptide chain into various higher-order structures. The geometry of hydrogen bond formation is highly specific,⁴⁴ helping to give rise to the most common structural patterns in proteins. Indeed, based entirely on molecular modeling that was guided by insight into hydrogen bond geometry, Pauling and coworkers were successfully able to predict the atomic arrangements of both the α -helix⁴⁶ and the β -sheet,⁴⁷ the two most common elements of local structure in proteins.⁴⁸ Because of the importance of hydrogen bonds for imparting structure, their presence in proteins has been catalogued repeatedly.⁴⁸⁻⁵⁰

Current Limitations

The study of these interactions has enabled development of many important technologies, including force fields for molecular dynamics simulations⁵¹⁻⁵³ and automated methods for protein design,⁵⁴⁻⁵⁵ which in turn have yielded some exciting results.⁵⁶⁻⁵⁷ In order to assess the state of the art in understanding protein structure and folding, and therefore the limitations in these methods, the biophysics community has engaged for the past several years in a systematic evaluation called the Critical Assessment of Structure Prediction (CASP).⁵⁸ The premise of this biannual competition is simple: given only the target sequence, computational biophysicists

attempt to predict the three-dimensional structures of proteins that have recently been characterized by experimental structural biology. The contest has been held eleven times since 1994, and several insights have been gained.⁵⁸ Unfortunately, though significant improvements were made in the first several years of the CASP competition, results of late have stagnated.⁵⁹ Similar limitations have been reviewed in the arena of protein design.⁶⁰⁻⁶¹ These results underscore the sheer size of the problem. Proteins have access to an enormous number of conformations, each of which differs only slightly in energy, creating a massive computational challenge. Moreover, the accuracy of models for sequences bearing low similarity to proteins of known structure remains poor; even the most successful of these models rarely achieve accuracy at either the secondary or tertiary levels of structure.⁶² In the absence of homologous proteins or domains on which to base initial models, structure prediction relies almost entirely on molecular mechanics approaches, which can be problematic.⁶³ Our inability to predict protein structures or to improve our prediction methods therefore suggests an incomplete understanding of the forces that govern protein structure, stability, and folding.⁶³ Indeed, similar limitations have been noted in the prediction of organic crystal structures,⁶⁴⁻⁶⁵ a problem that is conceptually similar to that of protein structure prediction in that it requires an accurate inventory of intermolecular forces. It is apparent, therefore, that understanding of the now-canonical forces in protein folding is insufficient for properly describing their biophysics. To address this problem, a great deal of research has now identified a suite of additional interactions that also contribute.

Secondary Interactions of the Main Chain

For the sake of discussion, the term secondary interactions will encompass those interactions identified in proteins that are outside the four canonical forces identified above. Perhaps the most

ubiquitous secondary interactions articulated to date are noncanonical hydrogen bonds involving carbon-based donors. Though proteins generally feature weaker carbon acids than some other organic molecules, there are some protons that are sufficiently acidic to engage in hydrogen bonding. For example, there is substantial evidence that histidine side chains can donate hydrogen bonds from C^ε, as occurs in enzyme active sites.⁶⁶ By far the most common C–H hydrogen-bond donors, however, are the α protons of the amino-acid backbone.

C–H hydrogen bonds have been widely observed in crystal structures of small molecules, and in fact, were proposed to contribute to protein stability as early as the 1960s.⁶⁷⁻⁶⁸ Despite early debate, C–H hydrogen bonds are now well-accepted,⁶⁹⁻⁷⁰ as they share many properties with canonical hydrogen bonds, such as directionality and cooperativity,⁷¹ though they notably often induce blue shifts in vibrational spectra.⁷²⁻⁷³ Like canonical hydrogen bonds, they are predominantly electrostatic interactions, with smaller contributions from van der Waals attraction and charge transfer.⁷¹ Experimental characterizations of their energy within peptides or proteins remain scarce,⁷⁴⁻⁷⁵ owing in part to their small energies and in part to the challenge of probing the peptide backbone. Computations generally point to energies of 1–2 kcal/mol each,⁷⁶⁻⁷⁷ which is approximately half that computed for canonical hydrogen bonds.⁷⁸

Nevertheless, detailed analysis of the geometry of intermolecular contacts in proteins has shown a substantial propensity for carbon-based acids to engage with hydrogen-bond acceptors. Like other hydrogen bonds,⁷⁹ these interactions are identified by a short donor–acceptor distance (<2.5 Å typically), and relative linearity between the donor, acceptor, and their antecedents; however, some analyses also require that C–H hydrogen bonds approach the acceptor within the plane of the carbonyl acceptor.⁸⁰ By far the most common example in protein structure is the interstrand C–H···O hydrogen bond in β-sheets (Figure 1.1).⁸¹⁻⁸²

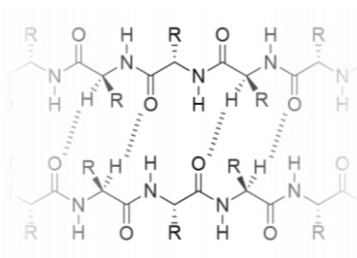


Figure 1.1 C–H···O Hydrogen bonds in the backbone of the β -sheet.

Derewenda and coworkers identified these contacts by observing that C–H···O contacts occur at distances significantly shorter than expected for repulsive van der Waals interactions, and in fact, the observed distances correspond closely to those observed in small-molecule crystal structures with validated interactions. Moreover, approach of the donor to the carbonyl acceptor occurs largely within the plane of the peptide group, where the carbonyl electron density is maximal. Later analyses have shown that these interactions affect some 35% of residues in β -sheets.⁸⁰

There are additional, albeit less frequent, examples of C–H···O hydrogen bonds in proteins. Contacts with carbonyl oxygens in the backbone of the α -helix have been noted,⁸³ though these interactions usually involve less acidic β -protons, so the energy contributed by such contacts is likely modest. Additionally, α -helices may benefit from C–H···O hydrogen bonds involving donation of a proline α -proton to carbonyl acceptors.⁸⁴ Indeed, this interaction has the potential to attenuate the strong helix-breaking tendencies of proline residues. C–H hydrogen bonds are also a feature of the collagen triple helix,⁸⁵ and can contribute to binding energy and discrimination at protein–protein interfaces.⁸⁶ One notable example of the latter occurs between transmembrane helices. Transmembrane helices, whether within individual proteins or at interfaces within complexes of multiple proteins, often contact one another along ridges of small amino acids, typified by the GXXXG motif.⁸⁷ These contacts are mediated by a network of C–H

hydrogen bonds between helices⁸⁸ and lead to a characteristic interaction geometry, which has been termed the GAS-right motif.⁸⁹

Recently, a distinct interaction within the peptide backbone has been posited to contribute to protein folding, termed the $n \rightarrow \pi^*$ interaction.⁹⁰ These weak interactions occur between adjacent carbonyl groups in the peptide backbone due to donation of lone-pair (n) electron density from a carbonyl oxygen into the π^* orbital of another carbonyl group (Figure 1.2).

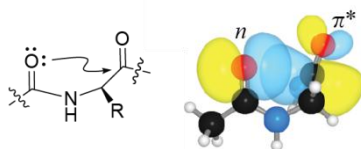


Figure 1.2 $n \rightarrow \pi^*$ Interactions between adjacent carbonyl groups in the peptide backbone.

Originally invoked to explain the correlation of proline ring pucker on the conformation of adjacent peptide bonds in collagen,⁹¹ this interaction has now been recognized in a variety of systems. It is identified⁹² by a sub-van der Waals contact of the donor oxygen on the acceptor carbon along the Bürgi-Dunitz trajectory.⁹³⁻⁹⁴ Though it was argued that these interactions are a particular example of dipolar interactions,⁹⁵⁻⁹⁷ extensive evidence⁹⁸⁻¹⁰⁰ indicates that these interactions harbor distinct, charge-transfer character, a view that has gained acceptance.¹⁰¹ The ensuing electronic donation can distort the acceptor, which has also been observed in ultra-high resolution protein crystal structures.¹⁰²⁻¹⁰³ Computational¹⁰⁰ and experimental⁹⁰ studies on small molecules have estimated the energy of a typical $n \rightarrow \pi^*$ interaction to be between 0.3 and 0.7 kcal/mol each; experimental estimates of the energy of an $n \rightarrow \pi^*$ interaction in a folded peptide or protein have yet to be achieved. Despite their modest energy, $n \rightarrow \pi^*$ interactions are predicted to contribute significantly to protein stability, thanks to their frequency; approximately one-third of all residues in folded proteins are poised to engage in $n \rightarrow \pi^*$ interactions.⁹² Moreover, they

likely contribute to secondary structure formation; over 70% of residues in α -helices are predicted to engage in $n \rightarrow \pi^*$ interactions, while less than 10% of β -sheet residues will.⁹² These interactions are also implicated in other helical conformations, such as 3_{10} ⁹² or PPII geometries.¹⁰⁴ Additional interactions are possible in amino-acid side chains,¹⁰⁵⁻¹⁰⁶ though this area remains underexplored.

An analogous, highly local interaction has recently been identified in β -sheets. Specifically, data shows that amide protons in β -strands can donate hydrogen bonds to their own carbonyl oxygens, forming what is termed a C5 hydrogen bond (Figure 1.3).

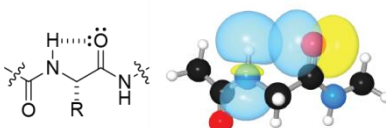


Figure 1.3 C5 Hydrogen bond in the peptide backbone.

These interactions appear to become significant at donor–acceptor distances below 2.5 Å, and despite their distorted geometry, they bear the hallmarks of traditional hydrogen bonding; their perturbation also causes predictable changes in the stability of model β -sheets. Though computation suggests that these interactions are significantly weaker than traditional hydrogen bonds, often affording only around 0.25 kcal/mol each, nearly 5% of residues in folded proteins are affected by such interactions, potentially making the C5 hydrogen bond a key contributor to protein structure and stability.

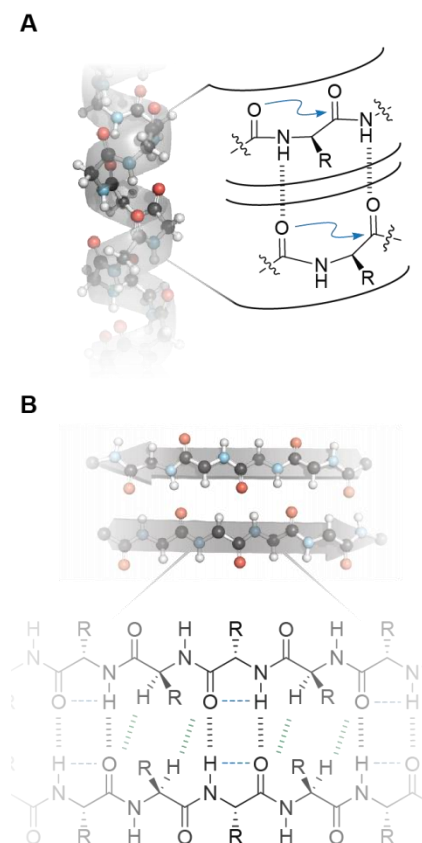


Figure 1.4 Secondary interactions in the protein backbone.

(A) Canonical hydrogen bonding (black) and secondary $n \rightarrow \pi^*$ interactions (blue) in the α -helix.

(B) Canonical (black), C-H...O (green), and C5 (blue) hydrogen bonding in the β -sheet.

Secondary Interactions Involving Side Chains

Among the secondary interactions possible between protein side chains, most important are those with aromatic rings. The unique electronic distribution in these side chains enables a number of possible interactions.¹⁰⁷ The facial π cloud bears significant partial negative charge and is nucleophilic, whereas the ring edges bear partial positive charge and are electrophilic. This unique charge arrangement also creates a significant permanent electric quadrupole that forms strong electrostatic interactions with both cations and anions. Indeed, perhaps the most important single example of noncanonical forces in protein folding is the cation- π interaction.

The importance of this interaction derives not only from its specific roles, but also from its energy, which is uniquely larger than other secondary interactions in proteins; individual cation- π interactions can contribute 2–5 kcal/mol each to the binding of ligands to their receptors.¹⁰⁸ Originally articulated in the supramolecular chemistry of organic cations,¹⁰⁹ these interactions are largely electrostatic attractions between electric monopoles and the electronically negative surfaces of aromatic rings, along with their corresponding quadrupole moments.¹¹⁰ Other contributions to the cation- π interaction, such as dispersion and charge transfer, exist and can be important, but predictions of binding affinity trends based on electrostatics alone are usually quite successful.¹¹¹⁻¹¹² Importantly, this energy is often sufficient to overcome the desolvation penalty for the binding of ions to protein pockets or cavities.¹¹³ Indeed, protein binding sites for a variety of organic cations feature an abundance of aromatic residues, making the cation- π interaction key for proteins to recognize small-molecule ligands¹⁰⁸ or post-translational modifications on proteins such as histones.¹¹⁴

In addition to these critical functional roles, the abundance of aromatic and cationic residues in protein side chains presents an important opportunity for their contribution to protein stability and folding. The cation- π interaction has been observed to perturb pK_a values of residues in proteins,¹¹⁵ demonstrating their influence unambiguously. Even under stringent criteria for identification, cation- π interactions are ubiquitous in protein structure, affecting approximately 1 in every 77 amino acids;¹¹⁶ for comparison, using similar criteria, canonical salt bridges are only approximately twice as common. Arginine forms cation- π interactions more often than does lysine, though this preference is not due to intrinsic interaction energies, but rather likely arises from the ability of arginine to participate in additional intermolecular interactions while near an aromatic ring. Aromatic residues in proteins form cation- π interactions with the relative

frequency of tryptophan more often than tyrosine more often than phenylalanine, which parallels the intrinsic interaction energies of the side chains. Analysis of cation- π distributions across structural motifs is much less advanced; however, study of designed peptides indicates that cation- π interactions can make significant contributions to secondary structure.¹¹⁷⁻¹¹⁸ In addition, preliminary work has shown that they may be a common feature of protein-protein interfaces.¹¹⁹ Importantly, though the contributions of individual cation- π interactions to ligand binding can lead to several kcal/mol of stabilization,¹⁰⁸ experimental measurements of contributions to protein folding provide somewhat lower values, suggesting that the energy of single cation- π interactions in peptides and proteins are generally in the range of 0.5–1.0 kcal/mol.^{107, 120}

Unsurprisingly, given their affinity for cations, aromatic rings are additionally capable of accepting hydrogen bonds¹²¹⁻¹²² even from weaker, carbon-based acids.¹²³⁻¹²⁴ Like the cation- π interaction, these interactions appear to be especially relevant for ligand binding, where they contribute in particular to carbohydrate recognition;¹²⁵⁻¹²⁶ the binding sites of lectins are often enriched in aromatic residues, particularly tryptophan,¹²⁷ which can direct the binding mode of the carbohydrate with exquisite specificity.¹²⁶ Within protein structure, hydrogen bonds to aromatic acceptors are identified by short contact of the donor heavy atom with the center of the aromatic ring at a steep angle of elevation to the plane of the ring.¹²² By these criteria, they appear sufficiently common to contribute to protein folding; surveys estimate that approximate 10% of aromatic residues accept hydrogen bonds from oxygen-, nitrogen-, or sulfur-based donors.¹²² In addition, interactions of aromatic rings with backbone donors, although relatively infrequent,¹²² can help to demark changes in the secondary structure pattern and stabilize structural termini. These interactions are, however, uniformly weaker than cation- π interactions, given the reduction in electrostatic attraction. Gas-phase studies indicate that energies of these

interactions are approximately 5- to 10- fold weaker than for analogous cations; for example, the interaction energy for ammonia and benzene in the gas phase is 1.8 kcal/mol,¹⁰⁷ whereas that for ammonium and benzene is 18 kcal/mol.¹¹⁰

The unique charge distribution in aromatic rings also allows them to interact favorably with one another. Inspired by the high aromatic contact of protein interiors, an early survey of aromatic–aromatic interactions in proteins by Burley and Petsko found not only that such pairs were more common than expected by chance, but also that particular short contact distances were strongly favored, suggesting an attraction.¹²⁸ Specific attraction between aromatic rings was first suggested by the dominance of enthalpy in the interaction, ruling out the previously hypothesized solvophobic nature.¹⁰⁷ Extensive characterizations have shown that aromatic rings interact primarily in two geometries: the T-shaped (or edge-to-face) and displaced-stacked (or offset-stacked);¹⁰⁷ Hunter and Sanders showed that both arrangements are well described by electrostatic considerations.¹²⁹ Geometries observed in crystal or gas-phase arenes are recapitulated in proteins, though the exact preference of aromatic residues for different geometries differs between estimates,¹³⁰ possibly due to expansions in the number of available protein structures.¹³¹ Nevertheless, these contacts can clearly offer stability to proteins. Interestingly, thermophilic proteins have significantly more aromatic–aromatic contacts than do homologues from mesophilic organisms, suggesting a contribution to protein stability.¹³² Experimental characterizations of individual aromatic–aromatic interactions in peptide and protein model systems have estimated the energy of a single interaction at approximately 0.5–1.5 kcal/mol each, lower values being observed for solvent-exposed residues in peptides¹³³⁻¹³⁴ and larger values for residues in proteins.¹³⁵ These values generally agree well with those from computation.¹³⁶⁻¹³⁷ Importantly, over half of aromatic residues in proteins have been predicted to

engage in attractive interactions, based on the simple criterion of a inter-centroid distance of less than 7 Å.¹²⁸

Aromatic rings can also interact with anions at their edges, which bear partial positive charge.¹³⁸ Indeed, initial analyses observed that carboxylates contact aromatic rings in proteins more frequently than would be expected by chance, and approach is predominantly edge-to-edge.¹³⁹ Energy deconvolution indicates that these interactions are likely dominated by electrostatics over other contributions, such as van der Waals interactions. Experimental measurements of the energy of a single anion- π interaction suggest that each such interaction contributes approximately 0.5 kcal/mol of stabilizing energy.¹⁴⁰ Using an energy-based criterion for identification, anion- π interactions were observed in approximately 70% of proteins examined.¹⁴¹ Though it is unclear how frequently they occur on a per-residue basis, they are somewhat less common than cation- π interactions. Sequence analysis suggests that anion- π interactions are predominately nonlocal, often involving residues outside secondary structure, though a small subset occur within single helices. No preference for aspartate or glutamate was observed, but importantly, inventories have only been collected for interactions with phenylalanine residues. Additional consideration of tyrosine and tryptophan residues may not only adjust our understanding of the overall frequency of these interactions, but could also illuminate particular amino-acid pairing preferences.

Surprisingly, aromatic rings are additionally capable of interacting with lone pairs, though these contacts generally involve electron-deficient rings, unlike those in proteins.¹⁴² Nevertheless, reports have documented an enrichment of sulfur atoms near aromatic rings in proteins,¹⁴³ leading to postulation of the so-called sulfur-arene interaction.¹⁰⁷ Though experimental perturbations of this interaction in peptides have found stabilizing energies on the

order of 0.5 kcal/mol each,¹⁴⁴⁻¹⁴⁵ experiments in proteins have provided contradictory results on whether these contacts are stabilizing.¹⁴⁶⁻¹⁴⁷ Moreover, the geometries of interaction observed in proteins are not the same as those found in small-molecule crystal structures;¹⁴⁸ specifically, whereas contacts in proteins occur at the face of the ring, contacts between small molecules favor interactions at the ring edge. In the former case, rather than acting as a lone-pair donor, the sulfur atoms can donate hydrogen bonds to the ring; in the latter cases, the sulfur can act as a hydrogen bond acceptor. These results raise the question as to the nature of this interaction, and results from peptide studies suggest that these interactions are dominated by the hydrophobic effect, rather than by a specific attraction.¹⁴⁵

Relative Contributions of the Secondary Forces

Upon consideration of this complex suite of interactions, it is clear that they can be divided largely into two groups. The first is the set of strong interactions that are relatively uncommon in proteins, either because of geometric constraints or amino-acid frequency, typified by the cation– π interaction. These interactions can contribute significant energy to the overall energy of folding; but more importantly, they direct specific contact formation, particularly at positions remote in sequence. Moreover, as these interactions pertain largely to side-chain functionalities, their appreciation is likely to improve methods for predicting protein structure from sequence.

Contrast these interactions with the weaker, yet more abundant interactions, such as C–H hydrogen bonds or the $n \rightarrow \pi^*$ interaction. In these cases, individual interactions are likely to be of little importance, given that their energies fall below that of thermal energy at ambient temperatures; however, their cumulative effects over a large number of residues can contribute to both the structure and stability of proteins. Due to their low energy, most are highly local

interactions, occurring often within single residues or between adjacent residues, and could thereby guide early events in the protein folding process. In addition, invoking the specific geometric preferences of these interactions may improve overall model accuracy and refinement. Finally, even crude estimates of the total contributions of pervasive, weak interactions suggest that they play critical roles in stabilizing the overall fold of proteins (Table 1.1), perhaps even more so than stronger interactions.

Table 1.1 Frequency and energy of various forces in protein folding.

Interaction	Approximate Frequency per 100 Residues	Approximate Energy (kcal/mol)	Total Energy per 100 Residues (kcal/mol)
$n \rightarrow \pi^*$ Interactions	33 ⁹²	0.25 ¹⁰⁰	8.5
C–H Hydrogen Bonds	13–15 ⁸¹	0.5 ⁷⁶⁻⁷⁸	6.5
C5 Hydrogen Bonds	5	>0.25	4.5
π – π Interactions	5 ¹²⁸	0.5–1.5 ¹³³⁻¹³⁵	2.5–7.5
Salt Bridges	2–4 ¹¹⁶	0.5–1.5 ¹⁸⁻¹⁹	2–6
Cation– π Interactions	1–2 ¹¹⁶	1–2 ¹⁰⁷	1–4
Anion– π Interactions	<1 ¹⁴¹	0.5 ¹⁴⁰	<0.5
X–H $\cdots\pi$ Hydrogen Bonds	1 ¹²²	<0.25 ¹⁰⁷	<0.25

Other Contributions

Though this discussion focuses on chemical interactions within proteins themselves, it would be remiss not to acknowledge the importance of environmental and biological contributions to protein structure, including the effects of cellular crowding,¹⁴⁹⁻¹⁵² as well as interactions of proteins with the ribosome,¹⁵³⁻¹⁵⁴ chaperones,¹⁵⁵⁻¹⁵⁶ and biological membranes.^{38, 157-158} Nature has evolved an exquisite set of post-translational modifications¹⁵⁹ and cellular machinery¹⁶⁰⁻¹⁶² to control the folding and function of proteins.¹⁶³⁻¹⁶⁴ These factors, whose regulation is the subject of intense study, undoubtedly contribute significantly to the observed structures of proteins. This area will likely be a source of immense insight into protein structure and folding over the coming decades.

Outlook

While significant progress has been made in inventorying the noncovalent interactions available to proteins, additional interactions undoubtedly remain to be identified. In addition, many known interactions remain undercharacterized, either in terms of their nature, as in the case of sulfur–arene interactions, their individual energetic contributions, as in the case of C–H hydrogen bonds or $n \rightarrow \pi^*$ interactions, or their frequency, as in the case of the anion– π interaction. Probing the contributions of backbone interactions is particularly challenging given the lack of genetic approaches to perturbation. There is also limited data on the distributions of these secondary interactions across both secondary and tertiary structural motifs. The interplay of these interactions with one another, as well as with canonical hydrogen bonding or electrostatic interactions, also remains largely uncharacterized. In addition, relatively little work has been done to characterize secondary contributions involving post-translational modifications. Although strong electrostatic effects of phosphorylation are well described, the effects of other modifications, such as acetylation or oxidation, remain largely unexplored. Moreover, as protein design and engineering efforts continue, consideration of interactions not possible in natural, proteinogenic amino acids, such as halogen bonding,¹⁶⁵⁻¹⁶⁷ might also be warranted.

A comprehensive understanding of these secondary contributions to protein structure will most directly benefit the accuracy of computational force fields. Understanding the nature of each interaction is particularly important, since electrostatic or dispersive contributions are likely to be better modeled by current force fields than are contributions from charge transfer, polarization, or multipolar interactions. Given the intimacy of the secondary interactions, they could help to achieve the dense packing observed commonly in folded proteins; proper packing densities have previously been attained by, for example, arbitrary scaling of atomic radii.¹⁶⁸ In

addition, because some of these interactions correlate with secondary structure, such as the $n \rightarrow \pi^*$ interaction or C5 hydrogen bonds, including these parameters could improve secondary structure prediction or refinement. Most importantly, assigning proper energies to individual contributions allows them to be sampled accurately and independently. As is, force fields reduce a variety of interactions into relatively few terms. Take for example the hydrogen bonds in the α -helix. There is significant evidence that the α -helix is stabilized by both hydrogen bonds and $n \rightarrow \pi^*$ interactions; however, a force field only accounts for the hydrogen bonds. Thus, in order to achieve agreement with experimental results, hydrogen bonding potentials have absorbed the energy that should be attributed to the $n \rightarrow \pi^*$ interaction. Although this approach might suffice for modeling the α -helix, it distorts hydrogen-bonding energies within other areas of the protein. Similar arguments can be made for each interaction discussed above. By properly dissecting the individual contributions of specific interactions, one likely improves the generality and robustness of these methods, making them more suitable for use in a wider variety of applications. The integration of additional interactions into computational methods therefore holds great promise for improving these vital technologies.

CHAPTER II

The $n \rightarrow \pi^*$ Interaction

Conspectus

The carbonyl group holds a prominent position in chemistry and biology not only because it allows diverse transformations, but also because it supports key intermolecular interactions, including hydrogen bonding. More recently, carbonyl groups have been found to interact with a variety of nucleophiles, including other carbonyls, in what we have termed an $n \rightarrow \pi^*$ interaction. In an $n \rightarrow \pi^*$ interaction, a nucleophile donates lone pair (n) electron density into the empty π^* antibonding orbital of a nearby carbonyl group. Mixing of these orbitals releases energy, resulting in an attractive interaction. Hints of such interactions were evident in small molecule crystal structures as early as the 1970s, but it was not until 2001 that the role of such an interaction was clearly articulated.

These noncovalent interactions were first discovered during investigations into the thermal stability of the proline-rich protein collagen, which achieves a robust structure despite a relatively low potential for hydrogen bonding. It was found that by modulating the distance between two carbonyl groups in the peptide backbone, one could alter the conformational preferences of a peptide bond to proline. Specifically, only the *trans* conformation of a peptide bond to proline allows for an attractive interaction with an adjacent carbonyl group, so when one increases the proximity of the two carbonyls, one enhances the carbonyl-carbonyl interaction and promotes the *trans* conformation of the peptide bond, which increases the thermal stability of collagen.

Since then, much attention has been paid to the nature of these interactions. Some have argued that, rather than resulting from charge transfer, carbonyl interactions are a particular example of dipolar interactions that are well approximated by classical mechanics. However, experimental evidence has demonstrated otherwise. Numerous examples now exist where an increase in the dipole moment of a carbonyl group decreases the strength of its interactions with other carbonyls, demonstrating unequivocally that a dipolar mechanism is insufficient for describing these interactions. Rather, these interactions have important quantum mechanical character that can be evaluated through careful experimental analysis and judicious use of computation.

While individual $n \rightarrow \pi^*$ interactions are relatively weak (~ 0.27 kcal/mol), the ubiquity of carbonyl groups across chemistry and biology gives the $n \rightarrow \pi^*$ interaction broad impact. In particular, the $n \rightarrow \pi^*$ interaction is likely to play an important role in dictating protein structure. Indeed, bioinformatics analysis suggests that approximately one-third of residues in folded proteins satisfy the geometric requirements to engage in an $n \rightarrow \pi^*$ interaction, which is likely to be of particular importance for the α -helix. Other carbonyl-dense polymeric materials like polyesters and peptoids are also influenced by $n \rightarrow \pi^*$ interactions, as are a variety of small molecules, some with particular medicinal importance. Research will continue to identify molecules whose conformation and activity is affected by the $n \rightarrow \pi^*$ interaction and will clarify their specific contributions to the structures of biomacromolecules.

Introduction

Study of the carbonyl group has justifiably received great attention, thanks in part to its varied reactivity and ubiquity across chemistry and biology. In addition to the enormous catalogues of chemical transformations supported by carbonyls, their intermolecular interactions play a paramount role in the organization of biological systems. For example, carbonyls participate in the hydrogen bonding that governs the structure of nucleic acids¹⁶⁹ and proteins.⁴⁶⁻⁴⁷ More recently, it has been recognized that carbonyl groups can additionally form attractive interactions with one another. This interaction was first studied in detail by Allen, who identified intimate intermolecular contacts between carbonyl groups in crystal structures.¹⁷⁰ These interactions were reminiscent of other nucleophilic interactions with carbonyl groups that had been identified by Bürgi and Dunitz.⁹³⁻⁹⁴ However, it was not until 2001 that a definitive example was presented wherein perturbations to a carbonyl-carbonyl interaction was shown to affect the stability of a protein.⁹¹ Since then, substantial effort has been devoted to studying the nature of these interactions and their contributions to chemical and biological phenomena.

Nature of Carbonyl Interactions

The charge distribution of the carbonyl group creates the potential for a variety of attractive interactions to exist between two carbonyl groups. In particular, one can envisage (a) an electrostatic interaction between opposite point charges on the carbon of one moiety and the oxygen of another, (b) a dipolar interaction between the permanent electric dipoles of the two groups, or (c) a charge transfer interactions where electrons from the nucleophilic end of one carbonyl donate into electron deficient orbitals of another. Purely electrostatic and dipolar contributions to carbonyl interactions are likely to be well approximated by molecular force

fields used for interrogating biological phenomena; however, contributions from charge transfer are not represented in even sophisticated force fields, and thus represent a potential point of contention with the behavior of real biomolecules.

Contributions of each of these mechanisms to the interaction of carbonyl groups have been explored using a proline dipeptide model system (Figure 2.1). Not only does the pyrrolidine ring preorganize the *i-1* and *i* carbonyl groups for interaction, but these molecules also provide a convenient readout for perturbations to the interactions. Specifically, peptide bonds to proline populate both the *cis* and *trans* conformations, but because an attractive carbonyl interaction only exists in the *trans* conformation, changes in the conformer populations can report on changes to the strength of a carbonyl interaction. For example, perturbations that enhance the strength of carbonyl interactions should cause an increase in the population of the *trans* conformer, since this geometry allows for attractive interactions.

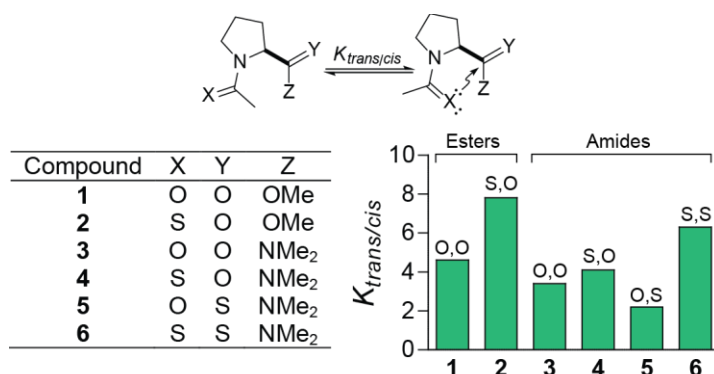


Figure 2.1 Characterization of carbonyl interactions by torsion balance analysis. Experimental $K_{trans/cis}$ values of proline derivatives are measured by NMR spectroscopy in D₂O at 25°C.

To probe the nature of carbonyl interactions, thioamides were strategically incorporated into either **1** or **3**.^{98, 100} Thioamides bear less partial negative charge on sulfur than oxoamides do at oxygen, so substituting the acetyl proline peptide bond with a thioamide should attenuate electrostatic interactions.¹⁷¹ However, an increase in the population of the *trans* conformation

was observed upon thioamide substitution of the *i*-1 carbonyl, indicating a stronger carbonyl interaction in **2** relative to **1** and **4** relative to **3** (Figure 2.1). Carbonyl interactions are therefore not well described by simple electrostatics. To interrogate the relative contributions from dipolar interactions and charge transfer, the methyl ester of **1** was exchanged for the dimethyl amide of **3**. Amides have higher dipole moments than esters,¹⁷¹ so one would predict that if a dipolar interaction is dominant, **3** would have a stronger carbonyl interaction than **1**, but the opposite was observed experimentally. Moreover, the thioamide has a still larger dipole moment than oxoamides,¹⁷¹ and yet **5** has a weaker carbonyl interaction than **3**. These data clearly demonstrate that carbonyl interactions cannot be described as purely dipolar.

Instead, the data are more consistent with a charge transfer interaction. For example, **3** shows a weaker carbonyl interaction than **1** because amides are less electrophilic than esters. Moreover, the divergent effects of thioamide substitution of the two amides of **3** (compare **4** and **5** relative to **3**) also demonstrates that each carbonyl has a unique role in these interactions: one as a lone pair donor, and the other as the electronic acceptor. These interactions are reminiscent of the approach of a nucleophile to a carbonyl during an addition reaction. This trajectory of approach maximizes overlap of the lone pair donor (*n*) of the nucleophile, which need not be a carbonyl, with the π^* orbital of the acceptor carbonyl and is named for Bürgi and Dunitz (Figure 2.2).⁹³ Carbonyl interactions have thus garnered the term Bürgi-Dunitz interactions, or more commonly, $n \rightarrow \pi^*$ interactions to reflect participation of the relevant electronic orbitals.

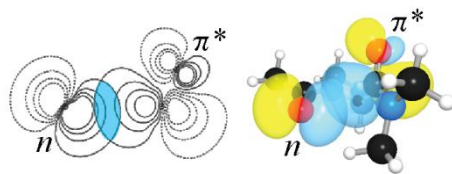


Figure 2.2 Overlap of the *n* and π^* orbitals of *N*-acetyl proline dimethyl amide.

Several additional lines of evidence support the quantum mechanical character of carbonyl interactions. It has been observed in analyses of both small molecule⁹⁹ and protein⁹² crystal structures that while the angle between the two carbonyl dipoles varies between interacting pairs, approach of the nucleophile to its carbonyl partner occurs strictly along the Bürgi-Dunitz trajectory, highlighting the importance of the molecular orbitals of these groups in dictating the interaction. Moreover, since the $n \rightarrow \pi^*$ interaction involves population of the π^* antibonding orbital, it should weaken the strength of the carbonyl π -bond, which should manifest itself multiple ways. First, a red shift in the acceptor carbonyl stretching frequency, corresponding to a weaker carbonyl bond, has been noted repeatedly.^{91, 172-174} Second, weakening of the π -bond should reduce the planarity of the carbonyl group, thus engendering pyramidalization (Figure 2.3) that can be observed in high-resolution crystal structures. Indeed, such signatures of the $n \rightarrow \pi^*$ interaction have been reported for many years in a wide variety of systems,^{98-100, 102, 175-178} including polymers¹⁷⁹ and proteins,¹⁰³ giving strong credence to the notion of carbonyl interactions as fundamentally electronic in nature.

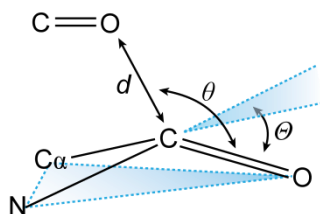


Figure 2.3 Geometric parameters characterizing an $n \rightarrow \pi^*$ interaction and the ensuing pyramidalization.

As the $n \rightarrow \pi^*$ interaction relies on orbital mixing, two fundamental quantities govern the energy of an $n \rightarrow \pi^*$ interaction: the degree of orbital overlap and the energy difference between donor and acceptor orbitals. Greater orbital overlap, generally corresponding to shorter donor-acceptor distances, increases the strength of an $n \rightarrow \pi^*$ interaction.⁹⁸ However, as donor-acceptor

distances decrease, the filled donor orbital will also experience Pauli repulsion by filled orbitals on the acceptor. It is therefore important that the acceptor group be highly polarized, which allows preferential interaction of the electron pair donor with the unfilled antibonding orbital of the acceptor over the filled bonding orbital (Figure 2.4).

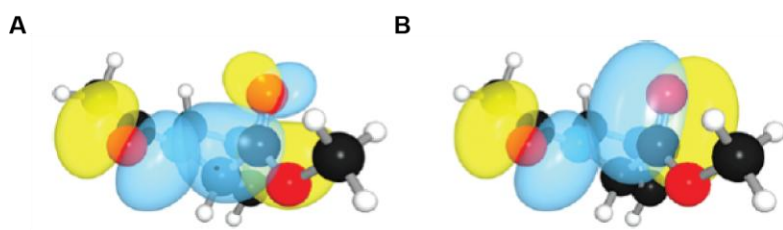


Figure 2.4 Orbital interactions between carbonyl groups. Overlap of the n donor orbital with the (A) π^* and (B) π orbitals of *N*-acetyl proline methyl ester.

For example, while carbonyl groups are effective $n \rightarrow \pi^*$ acceptors, isosteric alkenes and fluoroalkenes do not engage in substantial $n \rightarrow \pi^*$ interactions, due to the lack of sufficient orbital polarization.¹⁸⁰ As to the effect of the donor-acceptor energy gap, it is known from second-order perturbation theory that the energy released upon the mixing of a filled orbital with an empty one is inversely proportional to the energy gap between the donor and acceptor orbitals. Pairs of thioamides, like those in **6** for example, form especially strong $n \rightarrow \pi^*$ interactions because the lone pair donor orbital is higher in energy than the corresponding amide while the acceptor antibonding orbital is lower in energy.¹⁰⁰ Alkenes, on the other hand, have particularly high energy π^* orbitals, again making them poor $n \rightarrow \pi^*$ acceptors.¹⁸⁰

Because of the quantum mechanical nature of carbonyl interactions, computational methods have proven especially important in evaluating $n \rightarrow \pi^*$ interactions. In particular, natural bond orbital (NBO) analysis is often used to estimate the energy of $n \rightarrow \pi^*$ interactions.³⁹ NBO protocols partition electron density from diffuse molecular orbitals into localized, Lewis-type

orbitals, from which the energy of mixing can be computed. Extensive calculations have revealed the energy of the $n \rightarrow \pi^*$ interaction for a wide variety of amide-amide geometries⁹² and have placed a lower bound on the energy of a typical $n \rightarrow \pi^*$ interaction between amides at approximately 0.27 kcal/mol,¹⁰⁰ though numerous examples of stronger interactions exist (see below). These calculations and others¹⁸¹⁻¹⁸² highlight the modest energy of common $n \rightarrow \pi^*$ interactions. Nevertheless, because of the ubiquity of carbonyl groups across chemistry and biology, these interactions make important contributions to a variety of systems.

Contributions of $n \rightarrow \pi^*$ Interactions to Protein Structure

Because the pyrrolidine ring of proline preorganizes adjacent carbonyl groups for formation of an $n \rightarrow \pi^*$ interaction, these interactions are particularly important for the conformations of proline-rich peptides and proteins. In fact, the first observation that an interaction of this type contributes to the structure of proteins was made on collagen mimetic peptides.⁹¹ Collagen is the main structural protein of the skin and extracellular matrix and consists of three polypeptide type II helices wrapped around one another. This unique structure is enabled by its distinctive amino acid sequence: a Xaa-Yaa-Gly repeat, where Xaa is often proline and Yaa is often 4(*R*)-hydroxyproline.¹⁸³ The presence of a 4(*R*) electron-withdrawing substituent on the Yaa proline residue had been found to be important for the thermal stability of collagen because it enforces the *exo* pucker of the pyrrolidine ring through the gauche effect (Figure 2.5A).¹⁸⁴⁻¹⁸⁵ Curiously, incorporation of prolines with a 4(*S*) electron-withdrawing substituent, which enforces the *endo* pucker, was found to be destabilizing.⁹¹ When molecular modeling showed no obvious clashes in either case, it was hypothesized that the different conformations of the proline ring modulated the attraction between adjacent carbonyl groups. This hypothesis was confirmed by examining

the *trans/cis* preferences of AcProOMe derivatives (Figure 2.5B). Indeed, prolines with 4(*R*) electron withdrawing substituents, either hydroxy or fluoro, had higher preferences for the *trans* conformation of the acetyl-proline peptide bond, indicating a stronger carbonyl interaction. Conversely, prolines with 4(*S*) electron withdrawing substituents had weaker preferences for the *trans* conformation. Similar results were later obtained for prolines with 4-azido substituents.¹⁷²

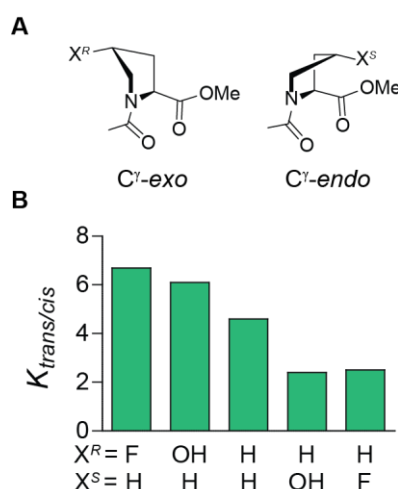


Figure 2.5 Conformational preferences of substituted prolines. (A) C^γ-*exo* and C^γ-*endo* puckers of the pyrrolidine ring. (B) Experimental $K_{trans/cis}$ of proline derivatives measured by NMR spectroscopy in D₂O at 25 °C.

Crystallographic and computational analyses,⁹⁸ microwave spectroscopy,¹⁸⁶ and study of constrained methanoprolines¹⁸⁷⁻¹⁸⁸ have since established that the *endo* pucker of the pyrrolidine ring increases the distance between the donor oxygen and the acceptor carbon, leading to weaker $n \rightarrow \pi^*$ interactions. These results explain the destabilization of collagen mimetic peptides by prolines with 4(*S*) electron withdrawing substituents at the Yaa position: because collagen requires exclusively *trans* peptide bonds, weakening the $n \rightarrow \pi^*$ interaction that enforces the *trans* conformation decreases the overall thermal stability of collagen. This insight has now been applied to the design of collagens with a variety of physical and chemical properties.^{173-174, 189-194}

The importance of the $n \rightarrow \pi^*$ interaction for controlling the conformational preferences of peptide bonds to proline is magnified in polyproline peptides, which, because of the lack of backbone hydrogen bonding, rely heavily on the $n \rightarrow \pi^*$ interaction. Polyproline adopts two distinct helical conformations: the polyproline type I helix with exclusively *cis* peptide bonds and the aforementioned type II helix with exclusively *trans* peptide bonds. In polymers of 4(*R*)-hydroxy-, methoxy-, fluoro,¹⁰⁴ or azidoproline,¹⁹⁵ there is a strong preference for the PPII conformation relative to polymers of proline. Conversely, polymers of the diastereomeric prolines inverted at C4 show a weaker preference for the PPII conformation than does polyproline. Moreover, a later study also showed that the presence of a strong $n \rightarrow \pi^*$ interaction, enforced by 4(*R*) electron withdrawing substituents, also increases the barrier to interconversion between the PPI and PPII helices.¹⁹⁶ The $n \rightarrow \pi^*$ interaction has been further implicated in the PPII structures of other sequences,¹⁹⁷⁻¹⁹⁹ demonstrating its ability to control peptide conformation. Recently, a crystal structure of an oligoproline reported by Wennemers and coworkers demonstrated unequivocally that $n \rightarrow \pi^*$ interactions contribute significantly to the PPII secondary structure.²⁰⁰ Specifically, this structure bears the hallmarks of the $n \rightarrow \pi^*$ interaction, namely short donor-acceptor distances and significant pyramidalization. Moreover, the lack of water in this structure precludes the importance of hydration for stabilizing PPII structure, underscoring the important stabilizing effect of the $n \rightarrow \pi^*$ interaction.

The immense carbonyl density of proteins suggests that the impact of the $n \rightarrow \pi^*$ interaction could be quite broad. Molecular modeling suggested early that many conformations of the peptide backbone would allow for close contact of adjacent carbonyl groups,⁹⁰ which was confirmed later through detailed computational and bioinformatics analysis.²⁰¹ In a key study,⁹² the energy of the $n \rightarrow \pi^*$ interaction was calculated for the entire conformational space of the

peptide backbone, which showed clear areas of Ramachandran space with significant (>0.5 kcal/mol) $n \rightarrow \pi^*$ interactions. In addition, a survey of high resolution crystal structures from the Protein Data Bank found a large fraction ($\sim 34\%$) of residues were properly oriented ($O_{i-1} \cdots C'_i$ distance within the sum of the van der Waals radii and $O_{i-1} \cdots C'_i = O_i$ angle approximately along the Bürgi-Dunitz trajectory) for an $n \rightarrow \pi^*$ interaction. Importantly, those residues found to be properly oriented for an $n \rightarrow \pi^*$ interaction had backbone dihedral angles consistent with those predicted computationally to engage in $n \rightarrow \pi^*$ interactions with significant energy. A later study of sub-Å protein crystal structures showed that residues engaging in $n \rightarrow \pi^*$ interactions showed greater pyramidalization of the acceptor carbonyl than those that do not, demonstrating that these interactions have measurable consequences for protein structure.¹⁰³ Even considering only the reported lower bound to the energy of a typical $n \rightarrow \pi^*$ interaction (0.27 kcal/mol),¹⁰⁰ the fact that approximately one-third of residues in folded proteins engage in an $n \rightarrow \pi^*$ interaction means their contributions could be on the order of 10 kcal/mol for a 100 residue protein. Given that the thermal stability of most proteins is estimated at 5–10 kcal/mol,²⁰² this contribution could be key.

Many of the residues engaged in $n \rightarrow \pi^*$ interactions are located within helical secondary structures, particularly the α -helix. Initial analysis of protein crystal structures demonstrated that almost 90% of residues in α -helices are aligned for participating in the $n \rightarrow \pi^*$ interaction.⁹² Dissecting the energetic contributions of the $n \rightarrow \pi^*$ interaction to the stability of the α -helix remains ongoing,²⁰³ but strong evidence for their effect came from analysis of high-resolution crystal structures, which showed pyramidalization of residues in α/β peptides that adopt helical conformations similar to the α -helix.¹⁰² While α -amino acids can position adjacent amide carbonyls within close proximity, β -amino acids, which contain an extra methylene group between the amine and carboxylate, cannot achieve the close contact of carbonyl groups

necessary for formation of an $n \rightarrow \pi^*$ interaction. Interestingly, only those carbonyls from α -amino acids were pyramidalized toward their putative $n \rightarrow \pi^*$ donors in these helices. These results provide convincing evidence that the $n \rightarrow \pi^*$ interaction does result in attraction between carbonyls in α -helices. There is also evidence that $n \rightarrow \pi^*$ interactions similarly contribute to the stability of 3_{10} helices,⁹² as α -aminoisobutyric acid residues (Aib), which strongly enforce the 3_{10} conformation, also induce strong $n \rightarrow \pi^*$ interactions.²⁰⁴

Amino acid side chains are additionally capable of forming $n \rightarrow \pi^*$ interactions. For example, Pal *et al.* investigated the ability of aspartate residues to interact with their own main chain carbonyls.¹⁰⁵ These interactions were first identified by their relatively short oxygen-oxygen distances, which seemed counterintuitive. The authors examined the geometries of these interactions and found not only that the aspartate oxygen approaches the backbone carbonyl along the Bürgi-Dunitz trajectory, but also that the arrangement of the carbonyl dipoles was likely to be destabilizing, suggesting that the $n \rightarrow \pi^*$ interaction stabilizes what would otherwise be an unfavorable self-contact; this hypothesis was supported by extensive computational evidence.

A more recent study examined similar self-contacts made by asparagine residues.¹⁰⁶ In addition to confirming the ability of asparagine side chains to donate $n \rightarrow \pi^*$ interactions to the protein backbone, this paper examined how hydrogen bonds to a carbonyl oxygen affect its ability to serve as an $n \rightarrow \pi^*$ donor. Self-contacting $n \rightarrow \pi^*$ interactions were much more common when the asparagine side chain carbonyl receives a hydrogen bond along the carbonyl bond axis. When the hydrogen bond donor approaches at $\sim 120^\circ$ to the carbonyl bond axis, the prevalence and calculated energy of self-contacting $n \rightarrow \pi^*$ interactions were diminished. It was shown that the geometry of a hydrogen bond to an $n \rightarrow \pi^*$ donor affects the ensuing $n \rightarrow \pi^*$ interaction by

controlling the demixing of the carbonyl lone pairs. When hydrogen bond donors approach at $\sim 120^\circ$ to the carbonyl bond axis, they encourage mixing of the orthogonal lone pair orbitals, which reduces the ability of the lone pairs to function independently in separate interactions. Conversely, when hydrogen bond donors approach along the carbonyl bond axis, they encourage demixing of the carbonyl lone pairs into *s*- and *p*-type orbitals (Figure 2.6A-B), which allows them to function independently, corresponding to higher incidences and energies of self-contacting $n \rightarrow \pi^*$ interactions. In this latter case, the $n \rightarrow \pi^*$ interaction seems to stabilize what would otherwise be nonideal hydrogen bonding geometries. This effect of hydrogen bonding on orbital demixing likely also contributes to the stability of the α -helix, where the canonical $i \rightarrow i+4$ hydrogen bond occurs along the carbonyl bond axis (Figure 2.6C-D).

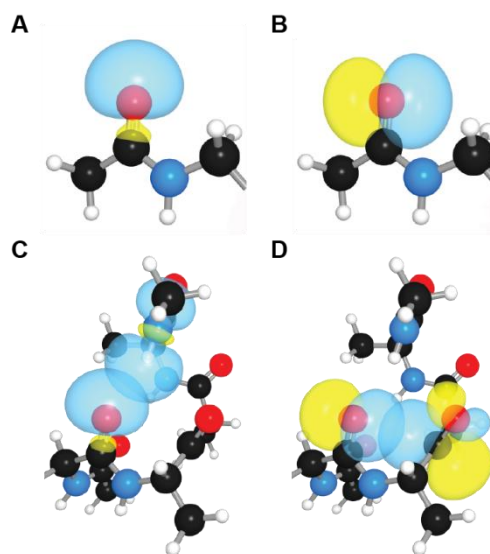


Figure 2.6 Carbonyl lone pairs in the α -helix.

(A) *s*-Type and (B) *p*-type lone pairs of a carbonyl oxygen. (C) Hydrogen bonding to the *s*-type carbonyl lone pair and (D) an $n \rightarrow \pi^*$ interaction with the *p*-type carbonyl lone pair in an α -helix.

We note that $n \rightarrow \pi^*$ interactions can also occur between proteins and their ligands. In an intriguing example,²⁰⁵ an $n \rightarrow \pi^*$ interaction was found to be conserved between the protein backbone of fluorescent proteins and the imidazolidine chromophore. The presence of this $n \rightarrow \pi^*$

interaction is consistent with the red shift in the vibrational frequency of the imidazolidine carbonyl in the protein-bound chromophore relative to small molecule mimics in solution. Moreover, it was hypothesized based on analysis of protein crystal structures with premature chromophores that this $n \rightarrow \pi^*$ interaction preorganizes the chromophore for cyclization and precludes bond rotations that would lower quantum yield.

Contributions of $n \rightarrow \pi^*$ Interactions to Other Polymers

While proteins have received the most attention, the $n \rightarrow \pi^*$ interaction contributes to other polymers as well. For example, it was recently reported that the $n \rightarrow \pi^*$ interaction contributes to poly(lactic acid),¹⁷⁹ a biodegradable polyester (Figure 2.7A). Fiber diffraction has shown that the backbone dihedral angles in PLA resemble those of the PPII helix of peptides, which takes advantage of numerous $n \rightarrow \pi^*$ interactions (see above). Computation placed the average energy of an $n \rightarrow \pi^*$ interaction in PLA at 0.44 kcal/mol, and analysis of small molecule crystal structures demonstrated characteristic pyramidalization of the acceptor carbonyl resulting from $n \rightarrow \pi^*$ donation. Like polyproline, PLA has no potential for hydrogen bonding, so the observation of $n \rightarrow \pi^*$ interactions in this polymer demonstrates not only that the $n \rightarrow \pi^*$ interaction can operate independently of hydrogen bonding, but also that it is sufficient for dictating molecular conformation, even in the absence of preorganizing rings. Moreover, because the PPII conformation has been observed in the unfolded states of some peptides and proteins,²⁰⁶ one could imagine that the $n \rightarrow \pi^*$ interaction, which operates between adjacent residues, may direct the peptide chain toward folding into the PPII conformation prior to the formation of the native hydrogen bonding pattern, which can involve contacts between residues relatively distant in sequence.

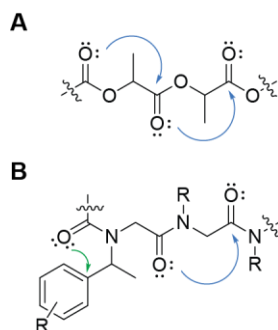


Figure 2.7 $n \rightarrow \pi^*$ Interactions in synthetic polymers.

Amide-amide (blue) and amide-aryl (green) $n \rightarrow \pi^*$ interactions in (A) poly(lactic acid) and (B) polypeptoids.

$n \rightarrow \pi^*$ Interactions are especially important for controlling the conformation of polymers of *N*-substituted glycines, or peptoids (Figure 2.7B).²⁰⁷ Analogously to polyproline and PLA, the lack of backbone hydrogen bonding places an enormous burden on amide $n \rightarrow \pi^*$ interactions for determining the molecule's overall conformation. In addition to the typical backbone $n \rightarrow \pi^*$ interactions, however, peptoids are additionally capable of forming $n \rightarrow \pi^*$ interaction with aryl rings on some peptoid side chains. Antibonding orbitals in aromatic π -systems can be potent electron pair acceptors, and this observation has been used to tune peptoid structure. Specifically, while backbone amide-amide $n \rightarrow \pi^*$ interactions favor the *trans* conformation of the tertiary amide, the amide-aryl backbone-side chain $n \rightarrow \pi^*$ interaction favors the *cis* conformation.²⁰⁸ These two interactions can therefore be exploited for exquisite control of peptoid structure, with the amide-aryl $n \rightarrow \pi^*$ interaction offering an exciting new avenue. For example, decoration of side chain phenyl rings with electron withdrawing fluoro or nitro groups increases the electrophilicity of the aromatic π^* orbitals and therefore encourages the *cis* conformation of the tertiary amide, while adding electron donating hydroxy groups can reverse this preference.²⁰⁹ One can also construct peptoid side chains using click chemistry, which affords the electron deficient triazolium ring, a potent $n \rightarrow \pi^*$ acceptor enforcing *cis* geometry of the tertiary amide.²¹⁰

Alternatively, one can modulate the ability of the carbonyl to donate an $n \rightarrow \pi^*$ interaction, either enhancing it with thioamides or attenuating it with electron withdrawing groups.²¹¹⁻²¹²

Contributions of $n \rightarrow \pi^*$ Interactions to Small Molecules

In principle, multitudinous compounds (Figure 2.8) are likely to engage in an $n \rightarrow \pi^*$ interaction, especially given that the lone pair donor need not be a carbonyl group. For example, Lesarri *et al.* observed $n \rightarrow \pi^*$ interactions between the hydroxyl group of 4(*S*)-hydroxyproline and the C1 carbonyl using microwave spectroscopy in the gas phase. This result has now been observed in crystal structures of a wide variety of prolines, including thiaproline,¹⁷⁶ acetoxypoline, and methoxypoline. It is now clear that many groups can interact with carbonyls in an $n \rightarrow \pi^*$ fashion, including halide ions,⁹⁹ thiols/thioethers/disulfides,²¹³ and a variety of nitrogen heterocycles.²¹⁴⁻²¹⁵ Similar types of electronic interactions involving carbonyls have been studied computationally and include complexes of SO₂ with carbon dioxide²¹⁶ or formaldehyde.²¹⁷

$n \rightarrow \pi^*$ Interactions have also been detected in other amino acids besides proline. In particular, gas phase microwave spectroscopy identified conformations of β -alanine that are consistent with the presence of an $n \rightarrow \pi^*$ interaction between the amino nitrogen and the carboxylic acid. The population of this conformer was similar to those for hydrogen bonded conformations, demonstrating that the energy of the $n \rightarrow \pi^*$ interaction may be similar to a typical hydrogen bond. Similar results were obtained for an analysis of γ -aminobutyric acid (GABA), an important neurotransmitter, though in this case the $n \rightarrow \pi^*$ interaction was observed in the conformation predicted as the global minimum, demonstrating the profound impact that the $n \rightarrow \pi^*$ interaction can have on the global topology of small molecules.²¹⁸

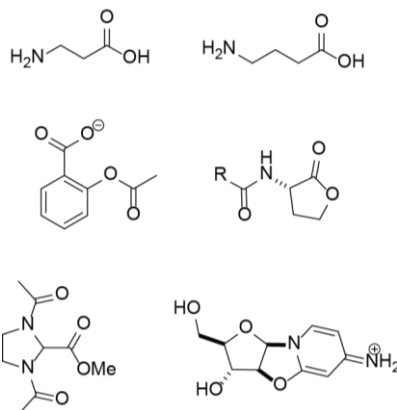


Figure 2.8 Selection of small molecules reported to engage in $n \rightarrow \pi^*$ interactions.

Recently, medicinal implications of $n \rightarrow \pi^*$ interactions have been noted. For example, an $n \rightarrow \pi^*$ interaction in aspirin was revealed through both crystallographic¹⁷⁷ and spectroscopic²¹⁹ investigations. Donation of electron density from the negatively-charged carboxylate into the ester carbonyl is predicted to shield the negative charge on the molecule and improve its entry into cells. An $n \rightarrow \pi^*$ interaction was also observed in *N*-acyl homoserine lactones (AHLs), which mediate quorum sensing in Gram-negative bacteria.¹⁷⁸ Crystallography and computation established the presence of an $n \rightarrow \pi^*$ interaction in the free AHL, while examination of protein crystal structures with bound AHLs demonstrated that AHL receptors break that $n \rightarrow \pi^*$ interaction upon binding. One can hypothesize, therefore, that attenuating the $n \rightarrow \pi^*$ interaction, which can be accomplished by appending electron-withdrawing substituents to the acyl group, could preorganize the ligand for receptor binding and increase potency.

These interactions also affect carbonyl reactivity. Using an imidazolidine-based model system, Choudhary *et al.* observed that the presence of one $n \rightarrow \pi^*$ interaction to a particular carbonyl discourages its interactions with a second donor; they concluded, therefore, that $n \rightarrow \pi^*$ donation reduces the electrophilicity of the acceptor carbonyl. This has important consequences for the reactivity of carbonyls that receive $n \rightarrow \pi^*$ interactions. For example, Houk and coworkers

have identified carbonyl interactions as determinants of stereoselectivity, such as in the dihydroxylation of *cis*-bicyclo[3.0.0]octenes,²²⁰ as well as the kinetic resolution of azlactones by benztetramisole catalysis.²²¹ An analogous stereoelectronic interaction was also proposed to explain the observed preference of anhydroarabinonucleosides for phosphorylation at the 3' position rather than the less sterically encumbered 5'-position.¹⁷⁵ This reaction is a key step in a proposed prebiotic route toward nucleotide synthesis, and regioselectivity in phosphorylation is essential for generating cyclic phosphates of cytidine for polymerization. A similar interaction was observed in cycloaddition of 3-hydroxyflavones.²²² The $n\rightarrow\pi^*$ interaction induced by proline was also shown to be the cause of slow ligation at proline thioesters during native chemical ligation.²²³ There, the presence of a putative donor carbonyl actually increased the reactivity of the thioester, preorganization of that carbonyl toward the $n\rightarrow\pi^*$ interaction by the pyrrolidine ring accounted for the dramatically reduced ligation rates observed during NCL.

Outlook

We anticipate that $n\rightarrow\pi^*$ interactions will be found in an ever-expanding array of molecules, particularly those with high carbonyl density such as proteins. Their impact on the structure of proteins has only begun to be studied and will require both ingenious experimental studies and judicious use of computation. Given the current limitations in protein structure prediction and design, it is likely that a thorough understanding of these ubiquitous interactions will improve efforts across protein science.

CHAPTER III

4-Fluoroprolines: Conformational Analysis and Effects on Peptide and Protein Stability and Folding

Proline is unique among proteinogenic amino acids thanks to the pyrrolidine ring connecting the side chain to the amino nitrogen; as such, it plays an indispensable role in restricting conformational freedom of the peptide backbone and templating secondary structure. Proline residues are frequently subject to post-translational modification at the 4- or γ -position to yield 4(*R*)-hydroxyproline, which has been observed in several proteins, most notably collagen. Interest in understanding this important modification led to the development of 4-fluoroprolines, which approximate the inductive effects of the hydroxyl group of 4-hydroxyprolines while eliminating complications from hydrogen bonding. The inductive effect of the fluoro substituent has three main effects: it enforces a particular pucker of the pyrrolidine ring, biases conformation of the preceding peptide bond, and accelerates *cis/trans* isomerization. These effects have important consequences for stability and folding of a number of proteins, and fluoroprolines have found great utility in this area due to the limitations in mutagenic approaches to assessing the many roles of proline. We examine in detail these effects on a wide variety of peptides and proteins.

This chapter has been published, in part, under the same title: Newberry, R.W.; Raines, R. T. *Top. Heterocy. Chem.* **2016**, DOI: 10.1007/7081_2015_196.

R.W.N. performed the literature review and drafted the manuscript, which was edited by both authors.

Introduction

Proline is unique among proteinogenic amino acids because the α -amino group is constrained by the side chain within a pyrrolidine ring. This cyclization of the amino acid restricts backbone dihedral angles, making proline an important determinant of polypeptide structure.²²⁴ Moreover, the pyrrolidine ring enables population of both the *cis* and *trans* conformations of the preceding peptide bond, thereby enabling structural diversity; this step in particular has been implicated as rate-limiting in the folding of many proteins.²²⁵ It is not surprising, given the conformational restriction of proline relative to other amino acids, that proline is highly enriched in structural proteins, particularly collagen.¹⁸³ It was the study of collagen, the main protein component of the skin and extracellular matrix, that first highlighted the ability of proline residues to be modified post-translationally. Specifically, within collagen strands, proline residues can become stereospecifically hydroxylated at the 4-position. Attempts to understand this modification eventually led to the development of a wide variety of proline analogues.²²⁶ Among these, 4-fluoroprolines have emerged as a powerful tool, not only for the study of collagen, but for protein engineering in general. 4-Fluoroprolines were first reported in 1965,²²⁷ when they were used to investigate the mechanism of collagen hydroxylation and were found to be successfully incorporated into collagen proteins.²²⁸⁻²³¹ Though early studies focused mainly on collagen hydroxylation,^{227, 232-233} 4-fluoroprolines have now been applied to a wide variety of systems, small and large, some of which have been reviewed previously.²³⁴⁻²³⁶ In this review, we introduce the utility of these important heterocycles with a thorough discussion of their synthesis and conformational properties, before examining in detail their impact on the stability and folding of peptides and proteins. We conclude with an overview of the growing interest of employing fluoroprolines for medicinal applications.

Synthesis

While significant improvements in fluoroproline synthesis have been made since the first report, nearly all take the same general approach of displacing the hydroxyl group of 4-hydroxyproline with a fluoride source. The original report by Witkop²²⁷ activated the hydroxyl group of Cbz-4-hydroxyproline as the tosylate before displacement by inorganic fluoride at elevated temperatures (Figure 3.1A). Yields were modest (56-62%), and while stereochemical inversion to 4(*S*)-fluoroproline (flp) from 4(*R*)-hydroxyproline (Hyp) was complete, mixed stereochemistry was observed in pursuit of the 4(*R*)-fluoroproline (Flp) starting from 4(*S*)-hydroxyproline (hyp). Moreover, subsequent reports found difficulty in generalizing the procedure in the presence of alternative protecting schemes.²³⁷ Most importantly, demand existed for syntheses that could provide both stereoisomers starting from naturally-occurring Hyp.

Preliminary work to bypass the alcohol activation step by employing organic fluorinating agents generally resulted in inseparable mixtures of fluoroproline stereoisomers.²³⁸⁻²⁴¹ Successful syntheses of both fluoroprolines from a single hydroxyproline were first reported by Dugave and coworkers.²³⁷ Employing diethylaminosulfur trifluoride (DAST) as the fluorine source, they were able to generate both Boc- and Fmoc-flp in approximately 80% yield from Boc-Hyp-OMe (Figure 3.1B). Moreover, from the same starting material, they were able generate Boc-hyp-OMe through a Mitsunobu procedure with hydroxide as the nucleophile, allowing access to Boc- and Fmoc-Flp. However, yield for the alcohol inversion was less than 25%, and conditions were incompatible with carbamate protection of the amine, instead requiring temporary trityl protection before introducing the fluoro substituent using DAST.

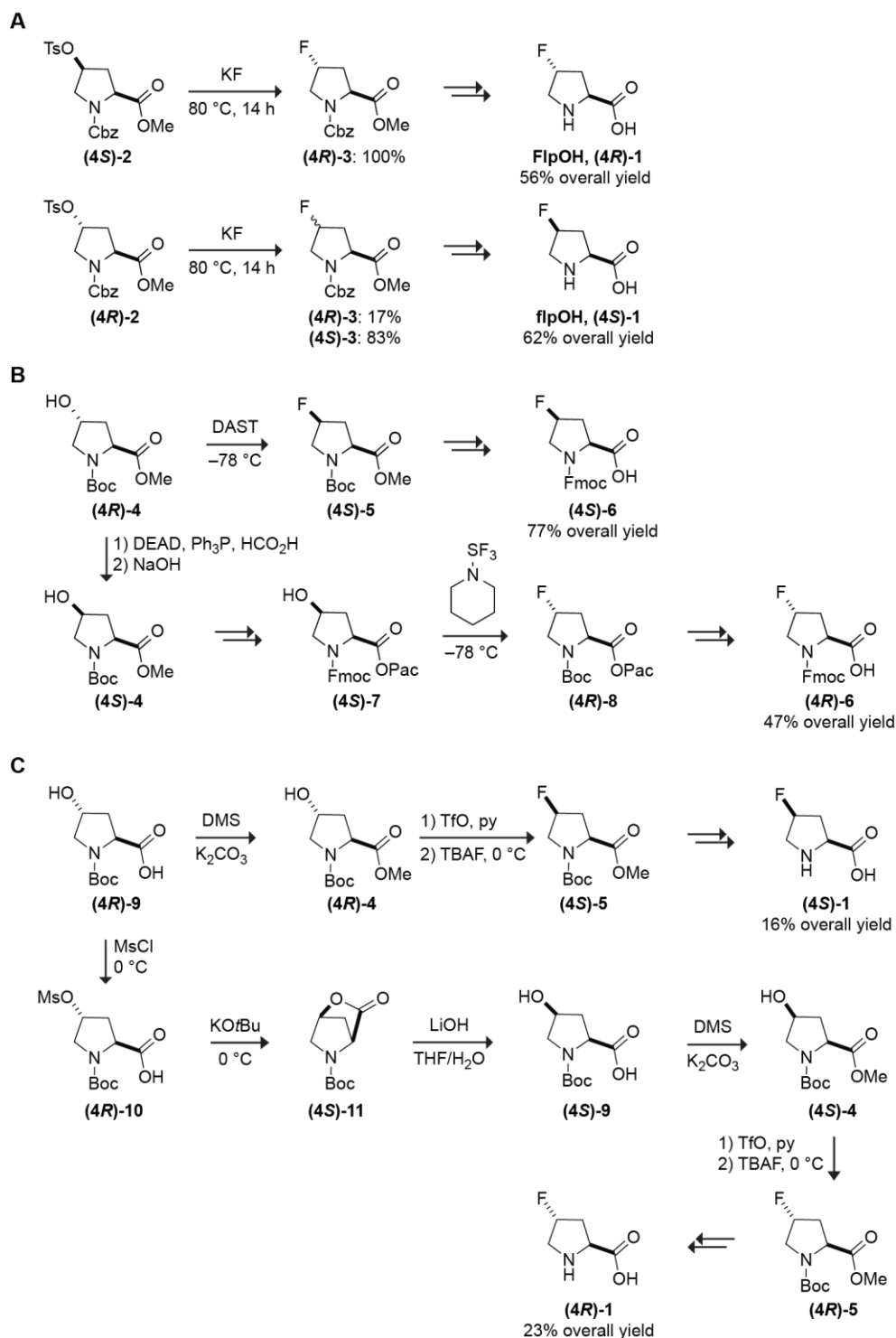


Figure 3.1 Approaches to the chemical synthesis of fluoroprolines.

(A) Original syntheses by Witkop and coworkers employing inorganic fluoride.²²⁷

(B) Fluoroproline synthesis from a single hydroxyproline stereoisomer using sulfur trifluoride reagents.^{237, 242}

(C) Fluoroproline synthesis from a single hydroxyproline stereoisomer using organic fluoride.²⁴³

Improvements on this procedure were reported later by the Kobayashi group.²⁴² By substituting formic acid for the benzoic acid used previously, they were able to conduct the Mitsunobu reaction in the presence of Boc-amine protection to generate Boc-hyp-OMe in substantially better yield (78%, Figure 3.1B). The authors then chose to exchange methyl ester protection for phenylacetylcarboxy (Pac) protection, which cleaves by reduction with Zn/AcOH, to improve orthogonality with the Boc and Fmoc amine protecting groups, though the authors admit that this complication likely is not necessary. Fluorination with morpholinosulfur trifluoride converted the Pac esters of hydroxyproline to the inverted fluoroprolines in excellent yield (79-94%), and X-ray diffraction and chiral HPLC established that the synthesis produces only a single stereoisomer. Protecting group manipulation then provides the starting materials for solid-phase synthesis in good yields (Figure 3.1B).

A later report sought to improve the scalability and cost of fluoroproline synthesis while bypassing the often explosive aminosulfur trifluoride reagents used previously.²⁴³ To generate flp, Hyp was first Boc and methyl ester protected before generating the triflate and treating with tetrabutylammonium fluoride (TBAF). Acidic reflux provided deprotected flp in 16% overall yield (Figure 3.1C). Though considerably lower-yielding than previous approaches, largely due to inefficiency in the safer fluorination step, this synthesis was successfully scaled to kilogram quantities. To generate Flp, rather than employing the Mitsunobu approach, Boc-Hyp was first activated as the mesylate before displacement of the hydroxyl group by the carboxyl group was induced using potassium *t*-butoxide. Opening of this ensuing lactone with lithium hydroxide provided the isomeric Boc-hyp in good yield (78%), which could then be subjected to fluorination as before. A similar lactonization approach was also used to prepare the D-isomers through treatment of Hyp with acetic anhydride prior to subsequent manipulation.

Of particular importance for peptide chemists is the recent advent of an approach termed proline editing.²⁴⁴⁻²⁴⁵ Zondlo and coworkers have demonstrated that the reactions necessary for generating fluoroprolines are amenable to solid-phase synthesis. In this approach, rather than preparing an appropriately protected fluoroproline monomer, Hyp is incorporated into a growing peptide by traditional solid phase methods, after which it can be elaborated. Synthetic chemists can elect to functionalize the Hyp residue either immediately following coupling or just prior to cleavage by employing selective protection and deprotection of the hydroxyl group, generally using trityl groups, though Alloc and TBS groups have also been used. Fluoroproline residues have successfully been incorporated by both strategies using the DAST fluorinating reagent. Moreover, Flp can be accessed following Mitsunobu inversion of the C4 alcohol, analogously to solution-phase approaches. These methods circumvent the labor required for solution-phase purification.

Conformational Analysis

Eclipsing interactions generally prevent population of the planar conformer of saturated five-membered rings, which instead adopt one of two predominant conformations: the envelope conformation, where four atoms lie in plane with the fifth distorted away; or the half-chair conformation, where two adjacent atoms distort in opposite directions out of the plane of the other three (Figure 3.2A).

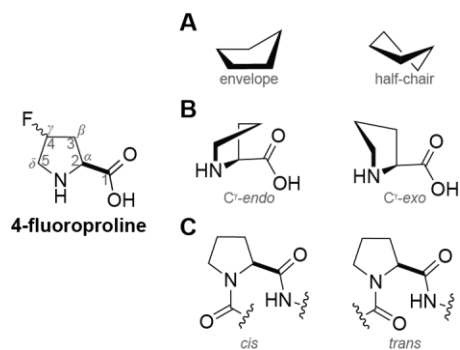


Figure 3.2 Conformations of proline rings.

(A) Conformations of saturated five-membered rings. (B) Pyrrolidine ring puckers. (C) Peptide bond configurations.

Early crystallographic studies,²⁴⁶⁻²⁴⁸ as well as data from NMR,²⁴⁹ indicated that proline generally adopts the envelope conformation. Many envelope conformations are possible, but crystallographic analysis demonstrates that two predominate: one where C4 (also called C γ) distorts out of plane to the same side of the ring as C1; and another where it distorts out of plane to the opposite side of the ring as C1 (Figure 3.2B).²⁵⁰ These conformations can appropriately be termed C γ -*endo* and C γ -*exo*, respectively, to reflect the relative positions of the out-of-plane C γ -atom and C1. It is worth noting, however, that these designations are approximate. For example, many molecules labeled as “C γ -*exo*” are better described as C β -*endo*. Moreover, many proline crystal structures display some half-chair character. Nevertheless, because C γ experiences the largest changes in orientation, and because this is often the site of modification, the C γ -*endo/exo* terminology has become customary and indeed is immensely useful for discussing stereoelectronic effects on proline conformations (see below).

The energy difference between the *endo* and *exo* puckers is small for proline (approximately 0.5 kcal/mol), and thus the two forms interconvert rapidly at room temperature.¹⁸⁵ NMR spectroscopy has estimated the equilibrium population of the two forms to be approximately 2:1 *endo:exo* for Ac-Pro-OMe.¹⁸⁵ Fluorination of C4 serves to bias this equilibrium.

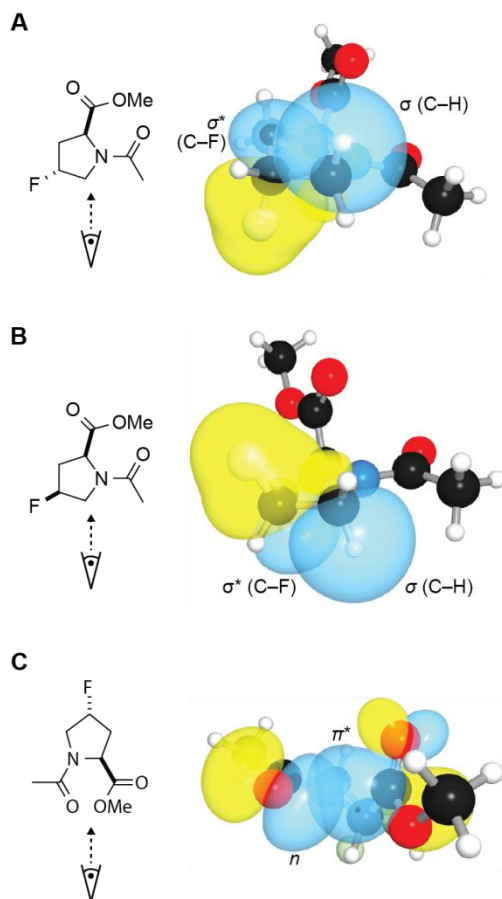


Figure 3.3 Stereoelectronic effects in fluoroprolines.

$\sigma(C-H) \rightarrow \sigma^*(C-F)$ Gauche interactions stabilizing the (A) C^γ -exo or (B) C^γ -endo conformations of Ac-Flp-OMe or Ac-flp-OMe, respectively. (C) $n \rightarrow \pi^*$ Interaction between adjacent carbonyl groups stabilizing the *trans* conformation of the Ac-Flp peptide bond in Ac-Flp-OMe.

Specifically, C–F bonds in fluoroprolines will orient antiperiplanar to adjacent C–H single bonds thanks to a gauche effect. The polarity of the C–F bond makes the ipso carbon electron deficient and therefore a potent electronic acceptor. The C–F σ^* orbital in particular is highly electrophilic and will accept electron density from C–H bonding orbitals when oriented closely in space.²⁵¹ The C–F bond in Flp is satisfied by C–H donor orbitals when the fluoro group orients toward the opposite side of the ring as the carbonyl carbon, thereby enforcing the C^γ -exo pucker (Figure 3.3A). Conversely, flp exists largely in the C^γ -endo conformation due to similar hyperconjugation (Figure 3.3B). These preferences were first established using ^1H NMR

spectroscopy,²⁵² where the fluoro substituent is valuable for dispersing the condensed proline proton signals, thereby allowing for accurate Karplus calculations. Refinements of this analysis quantified the preference of flp for the *endo* conformation to be approximately 20:1; similarly, Flp prefers the *exo* conformation at a 6:1 ratio.¹⁸⁵

Table 3.1 Conformational preferences of proline residues.					
Compound	Ring Pucker	φ (°)	ψ (°)	$K_{trans/cis}$ ⁹¹	Reference
Ac-Pro-OMe	<i>endo</i>	−79	177	4.6	253
Ac-Hyp-OMe	<i>exo</i>	−57	151	6.1	253
Ac-Flp-OMe	<i>exo</i>	−55	141	6.7	253
Ac-flp-OMe	<i>endo</i>	−76	172	2.5	185

Importantly, the reorientation of the pyrrolidine ring between the two puckers causes changes in the main chain dihedral angles, where the *exo* pucker generally reduces absolute backbone torsions (Table 3.1).²⁵¹ Though the φ ($C'_{i-1}-N_i-C^{\alpha}_i-C'_i$) angle is constrained by the pyrrolidine ring, it is still found to vary significantly between ring puckers, from around -60° in the *exo* pucker to approximately -75° for the *endo* conformation;¹⁸⁵ while these changes may appear modest, they can have a profound effect on the structure and stability of proline-rich peptides (see below). Two general orientations exist for the ψ ($N_i-C^{\alpha}_i-C'_i-N_{i+1}$) rotation: one around 150° , corresponding to the polyproline type II (PPII) peptide secondary structure; and the other around -30° , which is consistent with α -helical secondary structure. In isolated proline molecules and derivatives, the former is more common, and calculations suggest that it is indeed lower in energy.¹⁸⁰ Within the PPII rotamer, the influence of fluorination on proline ring puckering has been shown to affect the ψ dihedral angle by as much as 30° , from approximately 140° in the *exo* pucker of Flp to around 170° in the *endo* pucker of flp.¹⁸⁵

As a result of the influence of ring pucker on the backbone dihedral angles, fluorination of C4 can modulate interactions between the carbonyl groups on either side of the proline residue.

Adjacent peptide carbonyl groups have been shown to engage in attraction through so-called $n \rightarrow \pi^*$ interactions.⁹⁰ In an $n \rightarrow \pi^*$ interaction, lone pair (n) electron density delocalizes from the oxygen of one carbonyl group into the π^* antibonding orbital of an adjacent carbonyl (Figure 3.3C).⁹¹ Donation typically occurs from the $i-1$ residue to the i residue within a polypeptide.⁹² Effective orbital mixing requires close, sub-van der Waals contact of the donor oxygen with the acceptor carbon along the Bürgi-Dunitz trajectory for nucleophilic addition. These interactions have energies generally greater than 0.27 kcal/mol per interaction,¹⁰⁰ and have been shown to be ubiquitous in proteins.^{92, 103, 106} $n \rightarrow \pi^*$ Interactions are particularly common at proline residues because the pyrrolidine ring preorganizes adjacent carbonyls for interaction.⁹² As such, the effect of fluorination on proline ring pucker, and therefore on the backbone dihedral angles, has important consequences for the $n \rightarrow \pi^*$ interaction.¹⁸⁷ Specifically, the oxygen-carbon donor-acceptor distance is generally longer in the *endo* pucker, which leads to weaker $n \rightarrow \pi^*$ interactions.⁹⁸ Conversely, the *exo* pucker is associated with shorter donor-acceptor distances and stronger $n \rightarrow \pi^*$ interactions. In turn, the $n \rightarrow \pi^*$ interaction biases the conformation of the preceding peptide bond. An attractive $n \rightarrow \pi^*$ interaction is only possible in the *trans* conformation of the peptide bond (ω , $C_{i-1}^\alpha - C'_{i-1} - N_i - C_i^\alpha$, Figure 3.2C), and as such, *trans* peptide bonds are highly correlated with population of the *exo* pucker of the proline ring that can be enforced by a 4(*R*)-fluoro group (Table 3.1, Figure 3.4). Modification of the proline ring with fluoro groups therefore has important structural consequences that are transduced to all three dihedral angles of the protein main chain. We will see later that this principle can be exploited to modulate peptide and protein stability with exquisite control.

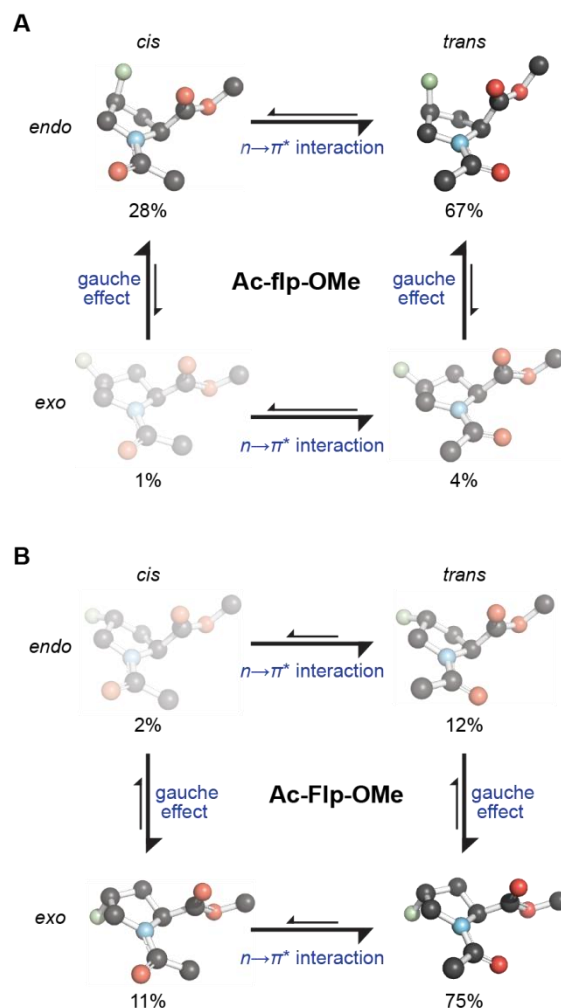


Figure 3.4 Conformational preferences of fluoroprolines.

(A) Conformational preferences of Ac-flp-OMe. (B) Conformational preferences of Ac-Flp-OMe.

The inductive effect of 4-fluoro substituents has additional consequences for peptide bonds to proline residues. Specifically, electron withdrawal by fluoro groups reduces C–N double bond character of peptide bonds to proline;²⁵³ this in turn reduces the rotational barrier between the *cis* and *trans* conformation of the peptide bond.²⁵⁴ As isomerization of peptide bonds to proline has been hypothesized to be a rate-determining step in protein folding,²²⁵ fluoroprolines have emerged as powerful probes of protein folding kinetics. In addition to the monofluorination discussed above, 4,4-difluoroproline (Diflp) has also been developed to exacerbate this effect to

study protein folding (see below). Though a detailed conformational analysis has not yet been achieved as have been for the monofluorinated prolines, we can likely expect the conformational preferences of Diflp to echo those of proline itself.

Effects on Collagen Stability

Fluoroprolines have found no greater utility than in the study of collagen stability.¹⁸³ Collagen is the principle structural protein in animals, forming a large portion of the dry weight of the skin. It is also prevalent in the extracellular matrix, again fulfilling a structural role. Collagen consists of a distinct triple helical structure comprised of three polypeptide strands that each have a characteristic Xaa-Yaa-Gly amino acid repeat.²⁵⁵ This sequence allows the three individual strands to each adopt PPII helices that wrap around one another with a single amino acid register offset; a slice through a collagen strand will therefore contain one each of Xaa, Yaa, and Gly residues on different strands. Interestingly, the PPII secondary structure of the individual strands lacks *intrastrand* hydrogen bonding, and each slice of the triple helix contains only a single *interstrand* hydrogen bond. The relatively low prevalence of hydrogen bonding raised the question as to the source of collagen's thermal and mechanical stability.

One clue to the source of this stability came from the amino acid content of collagen. Collagen has a high prevalence of Pro and Hyp residues, which appear at 28% and 38% frequencies in the Xaa and Yaa positions, respectively.²⁵⁶ The importance of the hydroxyl modification was shown early by thermal denaturation studies of both synthetic peptides²⁵⁷ and procollagen,²⁵⁸ which is the non-hydroxylated precursor to mature collagen. A crystal structure of a collagen mimetic peptide (CMP) showed bridging water molecules between the hydroxyl substituent and backbone carbonyl groups that were proposed to be stabilizing.²⁵⁹⁻²⁶⁰ However, it

was speculated that the entropic penalty associated with the ordering of those water molecules would be destabilizing overall. The stabilization of collagen by bridging water molecules between Hyp residues therefore remained unclear until replacement with the equivalent fluoroprolines settled debate. To eliminate the ability of the 4-substituent to participate in hydrogen bonding, the Hyp residues in the Yaa position of a CMP were substituted for Flp; circular dichroism (CD) spectroscopy then demonstrated that the substitution increased the thermal stability of the resulting CMPs relative to peptides containing Hyp (Table 3.2).^{184, 261} Thermal stabilization of CMPs by hydrogen-bonded water networks would be disrupted by the fluoro substitution, so the observed data emphatically discredited the hypothesis that such networks were responsible for collagen stability. Molecular dynamics simulations of CMPs containing Hyp later confirmed that bridging water molecules in CMPs organize only transiently.²⁶² This also marked the first synthetic incorporation of fluoroproline into a protein-mimetic.

Table 3.2 CMP melting temperatures.		
$(\text{Xaa-Yaa-Gly})_n$	T_m (°C)	Reference
(Pro-Pro-Gly) ₇	No Helix	263
(Pro-Hyp-Gly) ₇	36	91
(Pro-Flp-Gly) ₇	45	91
(Pro-flp-Gly) ₇	No Helix	91
(flp-Pro-Gly) ₇	33	264
(Flp-Pro-Gly) ₇	No Helix	264
(flp-Flp-Gly) ₇	No Helix	263
(Pro-Pro-Gly) ₁₀	31-41	261
(Pro-Hyp-Gly) ₁₀	61-69	261
(Pro-Flp-Gly) ₁₀	91	261
(flp-Pro-Gly) ₁₀	58	265
(Flp-Pro-Gly) ₁₀	No Helix	265
(flp-Flp-Gly) ₁₀	30	266

Instead, thorough conformational analysis of fluoroproline monomers, as described above, provides convincing evidence that collagen is stabilized by stereoelectronic effects that are

exaggerated by fluoroprolines.⁹¹ For example, while Flp incorporation at the Yaa position increased thermal stability relative to peptides with Pro at that position, flp incorporation caused a decrease in melting temperature (Table 3.2), clearly demonstrating the importance of stereochemistry at the proline 4-position. Inspection of available crystal structures indicated a prevalence of *exo* ring puckers in the Yaa position of collagen;²⁵⁹ as Flp serves to enforce this ring pucker, it likely serves to stabilize CMPs at the Yaa position through templating of the correct local dihedral angles. Moreover, because Flp encourages higher population of the *trans* configuration of the peptide bond (Table 3.1), it better templates the ω dihedral angle for the collagen triple helix, which features only *trans* peptide bonds.²⁵⁵ The ability of Flp residues to affect all three local dihedral angles therefore enables its unique stabilizing effect. These effects are evident even with a single Flp/flp substitution of a CMP²⁶⁷ and were shown to affect the biological activity of collagen by modulating the ability of cells to adhere to collagenous substrates.²⁶⁸

Further fluoroproline substitutions have revealed additional stereoelectronic influences on collagen stability. Specifically, whereas Flp incorporation at the Yaa position caused an increase in melting temperature of CMPs relative to Pro-Pro-Gly repeats, it caused a decrease in melting temperature when incorporated in the Xaa position.²⁶⁴ Conversely, flp demonstrated a stabilizing effect on CMPs when incorporated at the Xaa position, unlike its effect at the Yaa position.²⁶⁵ Again, inspection of available crystal structures demonstrated a prevalence of the *endo* ring pucker at the Xaa position; flp likely templates this structure better than Pro itself, even though Pro does have a slight preference for the *endo* pucker. The slight preference of Pro for the *endo* pucker likely also explains why the effect of Flp at the Yaa position is stronger than the effect of flp at the Xaa position (Table 3.2).

By templating the appropriate backbone dihedral angles of both the Xaa and Yaa residues, the incorporation of flp and Flp, respectively, into the same peptide is predicted to confer greater stability than either substitution alone. Surprisingly, simultaneous incorporation of fluoroprolines in this manner dramatically destabilized the collagen triple helix, even relative to repeats of Pro-Pro-Gly (Table 3.2).^{263, 266} Molecular modeling then revealed a potential steric clash between the two different fluoroprolines on adjacent strands that would prevent self-association of this peptide. These clashes could in principle be attenuated by incorporating fluoroprolines into only a single strand of the triple helix. By mixing different ratios of (flp-Flp-Gly)₇ and (Pro-Pro-Gly)₇, stable triple helices could be formed (Table 3.2). This makes flp-Flp-Gly polymers unique: they are capable of binding to other collagen strands while avoiding associations with other molecules of themselves. This principle has been applied to the probing of biological collagen samples. Collagen helices can often be disrupted by damage to the skin, potentially providing binding opportunities for probe peptides like flp-Flp-Gly polymers. This concept, termed “strand invasion,” based on its conceptual similarity to analogous processes in nucleic acids, has successfully been applied to specifically visualize collagen damage in wounds.²⁶⁹ Moreover, these invasive strands can be conjugated with a variety of factors designed to affect the biology of the surrounding environment, such as growth factors that aid in wound healing.²⁷⁰

Differential scanning calorimetry later provided complicating data regarding the stability conferred upon collagen by fluorination.²⁷¹ Specifically, melting of (Pro-Hyp-Gly)₁₀ indicated a strong enthalpic contribution to stability compared to (Pro-Pro-Gly)₁₀, while (Pro-Flp-Gly)₁₀ showed a stronger entropic contribution. Analysis of molecular volumes indicated that the enthalpic contribution to the stability of (Pro-Hyp-Gly)₁₀ was mediated significantly by solvating water, the release of which likely explains the entropic stabilization of (Pro-Flp-Gly)₁₀. In this

sense, the stabilization of CMPs by Flp could be due to the hydrophobic effect. Indeed, the Flp fluoro group is a much weaker hydrogen bond acceptor than the corresponding alcohol of Hyp,²⁷² while also not possessing any hydrogen bond donor. In an interesting follow-up, Diflp was employed to further test this hypothesis.²⁷³ Whereas Diflp retains the hydrophobicity of Flp, it lacks the strong preference for the *exo* ring pucker, so if the hydrophobic effect is dominant in the contributions of Flp to CMP stability, Diflp should have a similar effect. However, Diflp was shown to decrease CMP stability when incorporated into the Yaa position of CMPs, relative to CMPs with Flp in the Yaa position. This result argues against a strong contribution from the hydrophobic effect in driving CMP assembly.

Despite limited attention heretofore,²⁷⁴ fluoroprolines also make excellent NMR probes due to the utility of the ¹⁹F nucleus. This technology was recently exploited by the Kobayashi group to shed light on the mechanism of collagen strand association and dissociation.²⁷⁵ Specifically, they synthesized a Pro-Hyp-Gly repeat peptide featuring a single Flp residue at the center Yaa position. They were then able to use ¹⁹F-NMR, and in particular ¹H-¹⁹F and ¹⁹F-¹⁹F exchange experiments, to study the populations of various native and nonnative structures in solution. As the temperature of the solution increased, approximating the denaturation of native collagen strands, they observed the presence of multiple intermediates, contradicting the two-state model often assumed for the denaturation of CMPs. Using information from their exchange experiments, they were able to conclude not only that isolated monomer strands feature both *cis* and *trans* peptide bonds, but also that collagen assembly takes place only from monomers with strictly *trans* peptide bonds. Moreover, they characterized off-pathway intermediates as misaligned triple helices featuring more than a single residue offset; these misaligned helices converted to the native state only through the all-*trans* monomer. These data therefore

demonstrate not only an additional power of the fluoroproline substitution, but also shed new light on the detailed mechanisms of collagen folding.

In addition to imparting thermal stability and binding specificity to collagen strands, fluoroprolines have also been used to investigate the mechanism by which collagen strands are hydroxylated by prolyl-4-hydroxylase (P4H), an essential enzyme in animals.²⁷⁶⁻²⁷⁷ By using a model P4H substrate, it was shown that prolines like flp that prefer the *endo* pucker and the *cis* peptide bond conformation were better substrates than residues like Flp that prefer the opposite conformations; in fact, while flp is a substrate for P4H, Flp is not even bound by the enzyme.²⁷⁸ These results, along with studies of other proline modifications, helped to determine that P4H discriminates its substrates during the binding event and not during the chemical turnover. These results not only shed light on the mechanism of collagen hydroxylation, but also provide valuable information for designing inhibitors of P4H that could potentially serve to treat fibrotic diseases that feature an overabundance of collagen. Moreover, the oxidation of flp to 4-ketoproline (Kep) by P4H has been developed into a facile probe of P4H activity, since enzymatic activity can be measured by fluoride ion release, allowing for direct and continuous assaying.²⁷⁹

Effects on Peptide Structure

Fluoroproline incorporation has benefitted peptide studies beyond the triple helical domain of collagen. Perhaps most important among these are the studies of polyproline helices. Two major conformations of the oligoproline peptide are known: the polyproline type I (PPI) geometry, which is marked by the *cis* orientation of the peptide bonds, and the aforementioned PPII geometry, which features *trans* peptide bond; the PPII conformation in particular is quite

common in both the folded²⁸⁰ and unfolded²⁰⁶ states of proteins; polyproline helices also find utility as molecular rulers.²⁰⁰ PPII structure is generally favored in more polar solvents, such as water, while the PPI conformation is formed preferentially in less polar solvents like *n*-propanol.²⁸¹ As the transition between these structures involves isomerization of the peptide bond, it was predicted that modulating the conformation of the proline ring could have a profound effect on the observed structures of oligoproline, given the connection between the proline ring pucker and backbone dihedral angles (see above). To evaluate this hypothesis, Pro₁₀, Flp₁₀, and flp₁₀ were synthesized by solid-phase methods and subjected to analysis by CD spectroscopy.¹⁰⁴ Unlike Pro₁₀, Flp₁₀ demonstrated PPII structure in both aqueous and organic solutions. Moreover, the observed melting temperature of Flp₁₀ was over 25 °C higher than that of Pro₁₀. Conversely, flp₁₀ demonstrated only slight PPII content in water and high population of the PPI geometry in propanol. These results are readily explained given the conformational preferences of the amino acid monomers; for example, because of the *exo* ring pucker enforced by Flp, a stronger *n*→ π^* interaction forms between adjacent carbonyls, leading to increased population of the *trans* conformation of the peptide bonds, thereby resulting in a PPII helix. A subsequent report also demonstrated that the stereoelectronic effects of fluorine incorporation on proline structure also affect the rate of interconversion between PPI and PPII geometries.¹⁹⁶ By transferring host-guest peptides from aqueous solution into propanol, it is possible to observe polyproline structural transitions over time by monitoring changes in the CD spectra. Results showed that Flp incorporation slowed the rate of PPII→PPI conversion, while flp incorporation accelerated it. This is again consistent with Flp stabilizing the *trans* conformation of the peptide bond. A follow-up study then demonstrated that this effect is stronger when the fluoroproline is placed at the C terminus of the peptide, relative to placement at the N terminus.²⁸² These data are

consistent with a proposed mechanism of PPI \leftrightarrow PPII transition occurring by a mechanism that initiates at the C terminus.²⁸³

Polyproline helices have recently gained additional attention as motifs in important protein-protein interactions.²⁸⁴ For example, SH3 domains, a ubiquitous peptide module for protein-protein interactions, binds to Pro-rich protein substrates, and structural analysis has demonstrated that these substrates often adopt a PPII geometry.²⁸⁵ In an effort to design high-affinity ligands for SH3 domains, possibly as a step toward therapeutics for cancer chemotherapy, among other applications, Ruzza *et al.* incorporated fluoroprolines into various Pro-rich peptides (P2 peptide) derived from hematopoietic progenitor kinase 1 (HPK1).²⁸⁶ These peptides were then assayed for binding affinity to the SH3 domain of the hematopoietic-lineage cell-specific (HS1) protein. Interestingly, while the effect of fluoroproline substitution on the peptide structure followed hypotheses based on conformational analysis of fluoroproline monomers (see above), this relationship did not translate to potency of binding: despite the observation that SH3 ligands have high PPII structural content, encouraging PPII structure did not improve binding. These data suggest that SH3-ligand interactions involve recognition of features beyond secondary structure. Alternatively, the individual fluoro groups may perturb specific binding interactions, though no evidence of this possibility was presented.

The Horng lab has expanded the utility of fluoroprolines beyond modulating PPII secondary structures, beginning with a study on the C-terminal subdomain of the villin headpiece (HP36), a small, mostly α -helical miniprotein.²⁸⁷ It has been noted that a proline-aromatic interaction likely forms between Pro62 and Trp64 at the N-terminus of its C-terminal helix, possibly contributing to the stability of this peptide. Structural investigations of this peptide have assigned an *exo* ring pucker to Pro62,²⁸⁸ so it was hypothesized that the substitution of this residue with flp would

weaken the proline-aromatic interaction and depress the thermal stability of this peptide. However, flp substitution was found to have little effect on the structure or stability of HP36; rather, Flp, which enforced the *exo* pucker, was found to be significantly destabilizing. The incorporation of flp was predicted based on molecular modeling to position the fluoro group toward the hydrophobic core of this protein, while the positioning of the fluoro group toward the exterior of the protein by Flp was suggested as a source of its destabilizing effect. Regardless of the exact nature of the effect, these results highlight the fact that even subtle perturbations can have unexpected effects in the context of complex systems like proteins.

A subsequent report by the Horng lab on fluoroproline incorporation into a β -rich peptide underscores this point.²⁸⁹ They sought to determine the role of a key proline residue on the stability of the Pin1 WW domain, a small, three-stranded β -sheet protein of about 40 residues. In particular, the crystal structure of this miniprotein shows Pro37 packed into the hydrophobic core of the structure and adopting an *endo* pucker.²⁹⁰ To examine the role of this residue in establishing the stability of this peptide, they synthesized variants that contained various substituted prolines, among them both fluoroproline diastereomers. Consistent with the hypothesis that the native *endo* pucker is preferred for Pin1 WW domain stability, prolines that favor the *exo* pucker of Pro37, including Flp, decreased the thermal stability of this peptide. Interestingly, most *endo*-favoring prolines, including hyp and 4(*S*)-methoxyproline (mop), were also destabilizing at this position; however, flp increased the thermal stability of this protein. The authors rationalized this result by pointing out that the hydrophobicity of flp makes it more likely to be well tolerated in the hydrophobic core of this peptide than hyp or mop. Moreover, flp incorporation was also shown to increase the affinity of the protein for a phosphorylated peptide ligand. This result again highlights the advantages of fluoroprolines for probing peptide

structure: complications from hydrogen bonding to hyp produce a potentially misleading result, while flp incorporation allows for more selective perturbation of the proline ring conformation.

The Zondlo lab has further extended the use of fluoroproline as structural probes of key structural motifs in peptides through their study of the Trp cage miniprotein.²⁹¹ This 20-residue peptide displays some of the physical characteristics associated with full proteins, namely tertiary structure and cooperative folding,²⁹² which is unique for a peptide of this size. The Trp cage consists of a C-terminal polyproline helix that folds back on the N-terminal α -helix thanks to a structured loop. Critical for this loop is a single proline residue, Pro12, and based on the observation of an *exo* ring pucker of this residue in the NMR structure reported for this peptide, the hypothesis was formed that the local structure of this proline residue could be important for the stability of this small protein. Employing the “proline editing” approach described earlier, they were able to generate Trp cage peptides bearing flp or Flp residues, among others, at the key position. In accord with the hypothesis, they found that Flp incorporation increased the thermal stability of the Trp cage relative to native proline, while flp incorporation caused a decrease in stability. Moreover, the destabilizing effect of flp was much greater than the stabilizing effect of Flp, which is consistent with the native proline being puckered in the *exo* conformation. The authors point out that these subtle effects in conformation, even of a single residue, are particularly important for small proteins which lack a hydrophobic core, the formation of which drives the folding of larger proteins.

In another example of the influence of fluoroproline on conformations of key turns, Moroder and coworkers examined the folding of cysteine-rich terminal regions of minicollagen-1, an important structural protein in lower animals.²⁹³ Specifically, both the N- and C-terminal domains of this protein contain cysteine-rich sequences with central prolines that fold into

unique disulfide-bonded conformations, despite significant sequence homology. In an NMR structure of the N-terminal domain, the central proline was found to exist in the *cis* conformation of the preceding peptide bond, whereas similar analysis of the C-terminal domain indicated a *trans* peptide bond to the central proline, thereby enabling a Type I β -turn.²⁹⁴ Moreover, while the C-terminal domain was found to adopt its native fold within one hour under experimental conditions, the N-terminal fragment folded orders of magnitude more slowly. Considering the relatively slow rate of *cis/trans* isomerization of peptide bonds to proline, it was hypothesized that this could present the rate-determining step in folding of these domains. To test this hypothesis, the authors substituted the key proline residue of the N-terminal domain with both stereoisomers of fluoroproline. The incorporation of Flp, which, through its stabilization of the *exo* ring pucker favors the *trans* conformation of the preceding peptide bond, was found to further slow the kinetics of folding, while the substitution of flp enhanced folding kinetics. The authors were able to show through a combination of molecular dynamics simulations and kinetic analysis that the effect of flp is to decrease the population of the *trans* isomer of the key peptide bond, which prevents it from being trapped in nonnative disulfide bonds. Their lab also employed fluoroprolines to investigate the contributions of similar disulfide bonds to CMP stability and folding.²⁹⁵

Beyond modulating peptide stability, fluoroprolines have also been used to successfully alter peptide function. Specifically, fluoroprolines were recently used to modulate the DNA-binding properties of peptides derived from integration host factor (IHF).²⁹⁶ Crystal structures of this protein have demonstrated that the so-called α -arm senses DNA sequence in part by intercalating an *endo*-puckered proline residue between base pairs.²⁹⁷ To probe its role, this proline was substituted with a variety of analogues in the context of a cyclic peptide mimicking the α arm;

this peptide was subsequently conjugated to a lysine dendrimer, which provides the necessary electrostatic environment to allow DNA binding. Replacement of the key proline by flp was then shown to increase the binding of this construct to specific DNA sequences in an electrophoretic mobility assay. These data indicate that fluoroprolines can be used successfully to modulate not only peptide structure, but also the emergent functions of those peptides.

Effects on Protein Folding

As the power of fluoroproline substitutions has gained broader appreciation, the incorporation of this monomer into larger protein systems has been achieved. Important for this work has been the development of complimentary approaches to the incorporation of fluoroprolines into full-length proteins. The first approach, termed selective pressure incorporation (SPI), exploits protein biosynthesis in proline auxotrophs and results in global fluoroproline substitution for proline. Importantly, this approach has enabled the simultaneous incorporation of multiple unnatural amino acids by combining auxotrophic mutations.²⁹⁸⁻³⁰³ The second method employs chemical ligation of fluoroproline-substituted peptides with protein fragments expressed biologically.³⁰⁴ Though operationally challenging, this latter approach allows for site-specific fluoroproline substitution in proteins bearing multiple proline residues. Both approaches have now found use in protein chemistry.

The first complete replacement of the proline residues in a protein with fluoroprolines was performed on an engineered barstar variant, the cognate inhibitor of bacterial ribonuclease (barnase).³⁰⁵ Wild-type barstar has two prolines, one in each peptide bond conformation. Genetic manipulation yielded a variant containing only the *cis*-configured proline. The authors showed that the protein was expressed from a proline auxotroph at approximately equal efficiency

regardless of whether the bacteria were fed proline or either fluoroproline. This result was profound for the utility of fluoroprolines in protein science because it demonstrated that fluoroprolines could be recognized efficiently by the endogenous proline amino-acyl-tRNA synthetase, thereby allowing future investigations. In accord with the hypothesis that flp incorporation would favor the *cis* peptide bond conformation, the authors observed that the barstar variant containing flp had a higher melting temperature than the parent protein; conversely, the variant containing Flp demonstrated a lower melting temperature. Moreover, these fluorinated barstar variants were subsequently employed to study the mechanism of peptide bond isomerization by peptidyl-proline isomerases.³⁰⁶ These results highlight not only the importance of proline residues in protein stability but also the power of fluoroprolines for probing these effects; no other method, genetic or chemical, provides such selective perturbation of peptide conformation at proline residues.

Improvement of bacterial strains by the Conticello group enabled even more efficient fluoroproline incorporation, particularly into elastin derivatives.³⁰⁷ While the initial report of fluoroproline incorporation by an auxotrophic bacterium employed a pan-auxotroph, Conticello and coworkers were able to express elastin derivatives efficiently from strains lacking one of two genes: either 1) the *proA* gene that encodes 5-glutamylphosphate reductase to generate glutamate-5-semialdehyde, which spontaneously cyclizes to a pyrroline intermediate, or 2) the *proC* gene that encodes pyrroline carboxylate reductase, which is responsible for reducing the pyrroline to generate free proline (Figure 3.5).

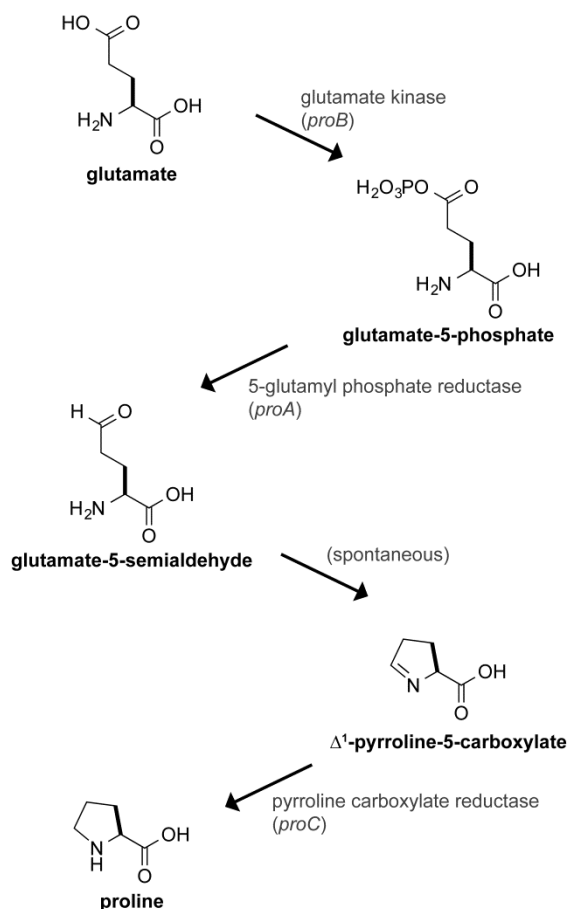


Figure 3.5 Biosynthesis of proline from glutamate.
Cofactors omitted for clarity.

Knockout of the *proC* gene, followed by supplementation of fluoroproline to the growth media resulted in efficient fluoroproline incorporation throughout the expressed protein, which contains eighty total proline sites. Elastin consists of concatenated oligopeptide repeats, most commonly VPGVG, making elastin highly enriched in proline residues. Structural investigations have indicated a prevalence of β -turn structure within this peptide repeat, for which proline is critical; these turns are believed to form above a critical transition temperature, below which the protein likely lacks definite structure. Moreover, it is believed that the opening and closing of these short hairpins gives elastin its elastic properties.³⁰⁸ As such, the ability of the proline residue to participate in the formation of these hairpin structures was hypothesized to be of critical

importance. Using the auxotroph strategy developed previously, the Conticello group was able to produce elastins uniformly substituted with either Flp or flp.³⁰⁹ As Flp enforces the *exo* conformation commonly observed in type II β -turns, it was predicted to stabilize the structured state and thereby lower the critical transition temperature. Differential scanning calorimetry and CD spectroscopy studies both confirmed this hypothesis. Conversely, elastins incorporating flp not only transitioned at a higher temperature, indicating weaker β -turn stability, but they also had weak thermal transitions. In addition, ^{19}F NMR analysis suggested significant population of the *cis* isomer of the valyl-proline peptide bond, also consistent with flp incorporation relative to Pro, which could exacerbate the destabilizing effect of this modification. These results provide dramatic new insight into the role of proline in controlling the structure of this protein. Importantly, this insight is unlikely to be gleaned from mutagenic experiments; rather, it is the exquisite control of proline conformation by diastereomeric fluoroproline incorporation that allows for this key advance.

The simplicity of this method has made it broadly applicable, as evidenced from studies of ubiquitin.³¹⁰ Ubiquitylation serves as a key post-translational modification with a wide variety of cellular effects, notably targeting of proteins for proteasomal degradation.³¹¹ Ubiquitin has three native prolines that were replaced with fluoroproline *via* the SPI method. The three prolines in native ubiquitin all are found in the *trans-exo* conformation, according to high-resolution crystal structures.³¹² As such, ubiquitin was predicted to better tolerate Flp incorporation than flp. In accord with this hypothesis, ubiquitin incorporating Flp was efficiently expressed from a bacterial auxotroph, resulting in ubiquitin nearly uniformly labeled with fluoroproline in comparable yields to wild-type, despite the use of two different proline codons in the ubiquitin gene. Conversely, no expression of ubiquitin was observed during attempts to incorporate flp,

likely due to aggregation or degradation of unfolded protein, demonstrating the importance of proper proline conformation for ubiquitin folding. Both wild-type and Flp-ubiquitin demonstrated native ubiquitin structure by CD spectroscopy, and stopped-flow fluorescence measurements indicated similar folding pathways involving a single intermediate. Rigorous analysis of these data demonstrated that Flp-ubiquitin has a lower barrier between the unfolded state and the intermediate and slower unfolding of the native state to the intermediate, both data establishing that Flp accelerates ubiquitin folding. Thermal denaturation at low pH also demonstrated an increase in melting temperature upon fluorination by approximately 7 °C; Flp-ubiquitin was also significantly more tolerant of GdnCl denaturant. Finally, Flp-ubiquitin was shown to be competent for self-ubiquitylation by ubiquitin-processing enzymes. These results again highlight that fluoroproline incorporation can increase both the thermal stability and folding rate of proteins while maintaining their biological function, suggesting important applications of this strategy in biotechnology.

Fluoroprolines were also used successfully to improve the physical properties of antibody fragments for potential uses in therapeutics, diagnostics, and biotechnology.³¹³ Antibody fragments such as single chain Fv (scFv), which are prepared by linking the variable domains of an IgG, generally lack the thermal stability of their full-length parents, preventing their application. To address this issue, Lee and coworkers produced an anti-c-Met scFv substituted with fluoroprolines using the (SPI) method. As was observed with ubiquitin, flp incorporation resulted in the expression of misfolded, insoluble protein, whereas as properly folded protein was obtained following incorporation of Flp. Mass spectrometry analysis demonstrated greater than 90% substitution efficiency of the five proline residues, which are generally conserved between scFvs. Interestingly, this scFv maintained higher activity in ELISA assays at elevated

temperature compared to wild-type, suggesting that this strategy may be of use for improving the physical properties of important bioreagents.

Another intriguing example of this technology was its application to the fluorescent protein EGFP.³¹⁴ Upon expressing EGFP from a proline auxotroph fed either Flp or flp, it was found that while EGFP incorporating Pro or flp expressed as soluble, fluorescent protein, EGFP incorporating Flp was detected only in inclusion bodies. Attempts to refold this protein failed, indicating that Flp incorporation has an irrecoverable effect on the structure of this protein. This result was unexpected because Flp is known to promote the *trans* conformation of the preceding peptide bond, as is common in GFP. Conversely, expression of EGFP containing flp resulted in a well-behaved protein whose structure was solved by X-ray diffraction experiments (Figure 3.6A).

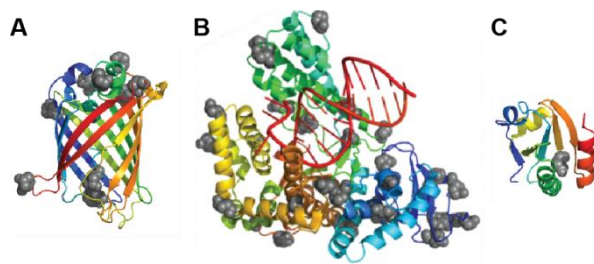


Figure 3.6 Protein crystal structures containing fluoroproline. Fluoroproline residues are shown in grey spheres. (A) EGFP (2q6p). (B) *Taq* DNA polymerase (4dle). (C) Thioredoxin (4hua).

This structure shows no noticeable differences with that of the parent structure, and the additional fluorine atoms are not found to disrupt local contacts. Moreover, kinetic refolding experiments demonstrated that flp incorporation accelerated the folding of EGFP relative to EGFP incorporating Pro. Though not articulated by the authors, it is possible that the reduced ω rotational barrier in flp relative to Pro allowed EGFP to escape from folding intermediates faster than the wild-type protein.

A later study performed the same perturbation to another fluorescent protein, mRFP1, with remarkably different results.³¹⁵ Whereas flp incorporation into EGFP resulted in a well-behaved protein, flp incorporation into mRFP1 produced only insoluble aggregates. Flp incorporation, which was deleterious to the folding of EGFP, resulted in a soluble mRFP1 that lacked fluorescence. By modeling the fluorine modification around the chromophore, the authors discovered an unfavorable steric clash between the chromophore and the fluorine of Pro63. By genetically mutating that position to alanine before incorporating Flp globally, the authors were able to express a fluorescent mRFP1 with enhanced thermal stability and faster maturation. Interestingly, mRFP1 is the only protein reported to date that will fold when substituted with flp but not with Flp.

In an impressive extension of the SPI method, global fluoroproline substitution has been achieved on the nearly 60 kDa KlenTaq DNA polymerase.³¹⁶ Using appropriate conditions, Holzberger, *et al.* were able to replace the 32 native proline residues with Flp at 92% efficiency. In accord with previous results on other proteins, attempts to incorporate flp did not result in protein expression. The resulting fluorinated enzyme was less thermostable than its wild-type counterpart, but otherwise functioned normally, displaying similar replication kinetics and error rates. Fortuitously, a refined preparation of the fluorinated protein with 98% substitution efficiency produced crystals of sufficient quality for diffraction to high resolution (2.4 Å, Figure 3.6B);³¹⁷ the authors suggest that the presence of fluoro groups on the exterior of the protein mediate contacts between individual protein molecules within the crystal lattice, thereby improving crystallinity. A rigorous approach to the structure solution and refinement provided unprecedented insight into the effect of fluoroproline substitution on protein structure. Reliable pucker assignments were possible for 28 out of 32 of the native prolines, which demonstrated

that 89% of these residues adopt the *exo* pucker. A 1.9 Å resolution structure of wild-type KlenTaq DNA polymerase showed only 43% of prolines adopting the *exo* pucker, demonstrating that the strong *gauche* effect in Flp can significantly alter the conformation of proline residues, even in a context with strong structural biases from nonlocal interactions. While the incorporation of Flp did not alter the conformation of the preceding peptide bonds (KlenTaq DNA polymerase contains two *cis* peptide bonds to proline), both of these residues showed significant population of the *endo* pucker, accounting for half of the *endo*-puckered residues. Moreover, whereas electron density for 36% of the prolines in the wild-type polymerase were consistent with a superposition of puckers, only 7% of Flp residues in the fluorinated protein were consistent with multiple puckers in the electron density map, indicating that fluoroproline substitution dramatically reduces conformational flexibility of the proline ring.

Because SPI necessarily results in global substitution, dissecting the impact of a modification at a given position can be challenging. To highlight the use of fluoroprolines as selective probes of a single position, Rubini *et al.* engineered a variant of *E. coli* thioredoxin containing only a single proline;³¹⁸ specifically, they mutated the four *trans*-configured prolines to alanine, leaving only the single *cis*-configured proline, the structure of which is conserved among thioredoxins. They were then able to generate both Flp- and flp-substituted variants of their single proline mutant, both of which were stable, well-folded proteins. Interestingly, in chemical denaturation experiments, both modifications were shown to have identical effects on protein stability: they both stabilize the reduced form of the protein while destabilizing the oxidized form. Both substitutions also improved the cooperativity of folding, which had been compromised by the four proline to alanine mutations. Measures of catalytic activity demonstrated little effect of fluorination. The lack of significant differences in the effect of these diastereomeric substitutions

was explained by high-resolution crystal structures of the various single proline proteins (Figure 3.6C). Although Flp and flp generally template the pyrrolidine ring for the *exo* and *endo* puckers, respectively, both were observed in the *endo* conformation in the crystal structure. Steric clashes of the 4-fluoro substituent in Flp were predicted to prevent adoption of the *exo* pucker. The authors concluded, therefore, that in this case, the tertiary structure of the protein nullified the usual effect of fluoroproline conformation, resulting in nearly identical proteins. Compared to the dramatic results obtained with *Taq* polymerase, these results demonstrate the complexities of even conservative single-atom replacements in the context of an entire protein.

Further characterization of the role of a single proline residue to protein folding was accomplished by Hilvert and coworkers in the first fully synthetic production of a fluoroproline-substituted protein, which they employed to study the folding and aggregation of β 2-microglobulin (β 2m), the causative agent of dialysis-related amyloidosis.³¹⁹ Structural characterization has shown that one of this protein's five prolines, Pro32, is involved in a *cis* peptide bond in the native structure.³²⁰ Interestingly, this bond isomerizes to the *trans* conformation in the amyloid, as shown by solid-state NMR spectroscopy.³²¹ As such, this peptide bond isomerization was predicted to be key in the mechanism of amyloid formation. However, traditional mutagenic experiments yielded contradictory results: substitution with glycine induced fibril formation, while substitution with valine or alanine did not.³²² Fluoroproline (Flp, flp, Diflp) substitution was therefore selected as a more subtle and selective approach to evaluating the role of this isomerization in the mechanism of β 2m amyloid formation. To site specifically incorporate fluoroprolines, β 2m variants were produced by native chemical ligation of three peptide segments generated through solid-phase synthesis. Temporary protection of Cys25 as a thiazolidine allowed for successful synthesis of the full-length protein,

which could then be oxidatively folded into its native structure. Thermal denaturation of these variants demonstrated that the *cis*-favoring flp substitution increased stability of the folded state, while Flp incorporation was destabilizing relative to wild-type. Interestingly, despite the similar conformational preferences of Diflp to Pro, the Diflp variant was the least stable of the series. The Diflp variant also displayed the least cooperative unfolding, the highest association with the ANS fluorophore (a probe of unstructured hydrophobic patches on proteins), and the most facile amyloidogenesis; this variant spontaneously formed fibrils after two weeks at neutral pH, whereas wild-type and the monofluoroproline variants did not. These results demonstrated that the *cis-trans* populations alone are insufficient for dictating amyloid behavior in this protein. Instead, the rate of this isomerization appears to be key. The authors speculate that the nature of this dependence is related to the ensuing flexibility of the BC loop instigated by rapid *cis-trans* isomerization at position 32. Consistent with their observations employing ANS, this flexibility could expose hydrophobic regions of the protein and increase the rate of amyloid formation. These results again highlight the power of substituting key proline residues with judicious choices of fluoroprolines to probe biophysical processes, especially in light of the ambiguous results from traditional genetic approaches.

Impact in Medicinal Chemistry

Fluorination has emerged as a powerful, general strategy for combating the oxidative metabolism of various pharmaceutical candidates.³²³ As such, fluoroprolines provide a convenient strategy for fluorination of peptide-based pharmaceuticals.³²⁴ One great example comes from the research laboratories at Merck, which were able to employ fluoroprolines to improve thrombin inhibitors.³²⁵ As thrombin is an important protease for blood clotting, effective

inhibitors are attractive drug candidates for clotting diseases. Working on a tripeptide scaffold with a central proline residue, they were able to show that though substitution with flp reduced potency to thrombin approximately 200-fold, substitution with Flp retained potency while eliminating oxidative metabolism of the central proline residue. The differential effect of fluoro stereochemistry is explained by comparison to available crystal structures of thrombin:inhibitor complexes, which show a preference for the *exo* pucker of the central proline residue³²⁶. These results enabled production of compounds with improved metabolic stability and selectivity toward thrombin over related proteases like trypsin.

Similar results have been obtained for the development of peptide inhibitors of Stat3 phosphorylation.³²⁷ Phosphorylation of Stat3 occurs following binding of cytokines to their receptors and results in dimerization and translocation of Stat3 to the nucleus, where it effects changes in gene expression; high activation of Stat3 has been observed in a number of disease states, including cancers, making it an important pharmaceutical target. Because phosphorylation occurs on the SH2 domain of Stat3, mimics of the SH2 phosphopeptide hold promise as inhibitors of dimerization. Mandal *et al.* found that replacing a key proline residue within a phosphopeptide prodrug with flp or Diflp increased the potency of the compound in cellular assays, despite decreasing affinity of the compound for full-length Stat3 by *in vitro* fluorescence polarization assays. Again, these data suggest that fluoroproline substitution presents an attractive strategy for increasing compound potency by slowing deleterious metabolism.

Recent crystal structures of the complex between neurotensin (NT) and its receptor NST1 have inspired structure-based drug development programs to yield new analgesics.³²⁸ NT features a key proline residue that binds to a tight pocket of its GPCR receptor; importance of this residue had been previously confirmed by mutagenic studies. To further develop the lead

compound, as well as to pursue selective inhibition of the related receptor NST2, which also binds neurotensin, NT8-13 analogues were produced involving a variety of proline modifications.³²⁹ A cellular assay was then employed to determine the ability of these analogues to inhibit binding of the native ligand. These results demonstrated a preference of the receptor for *exo*-puckered prolines, including Flp. Further elaboration of this peptide with a single peptoid moiety produced a fluoroproline-based ligand with nearly 10,000-fold selectivity of NST2 over NST1, thereby providing a valuable new probe, which the authors suggest may have additional utility in ¹⁹F MRI imaging.

While the effect of the fluoro substituent on ring conformation or peptide bond isomerization is usually the dominant contribution to systems incorporating *N*-acyl-fluoroprolines, the story can be quite different in *N*-alkyl-fluoroproline derivatives. For example, in a medicinal chemistry optimization of tiagabine-related GABA uptake inhibitors, fluorination of the amine heterocycle was shown to drastically decrease potency, regardless of stereochemistry.³³⁰ In this case, as the pyrrolidine amine is ionizable, fluorination likely decreases drug potency by decreasing basicity of the amine, thereby reducing binding to the GABA uptake protein.

Finally, though not a focus of this review, we note that fluoroprolines are also in the early stages of development as PET probes. ¹⁸F is an attractive PET nucleus due to its half-life (110 min) and lack of necessary chelators. As fluoroprolines are proteinogenic, they offer an opportunity to incorporate a PET nucleus directly into a protein, notably collagen. However, the combined time to radiosynthesize the fluoroproline, administer it to a subject, allow for trafficking and biosynthetic incorporation, and image the results make this strategy a difficult one to implement. Nevertheless, several radiosyntheses have been reported,³³¹⁻³³⁴ as have

pharmacokinetic studies.³³⁵⁻³³⁶ While results of this approach to imaging tumors³³⁷⁻³⁴⁰ or collagen synthesis³⁴¹⁻³⁴² have been mixed, we anticipate that further advance may be possible.

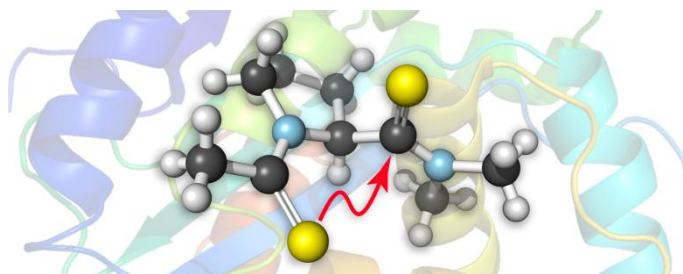
Outlook

Research to date has clearly demonstrated the utility of fluoroprolines for rationally perturbing peptides and proteins, both small and large. While conformational analysis of fluoroproline monomers is usually predictive of their effects when incorporated into biopolymers, complications have been observed due to the chemical and structural constraints imposed by these complex molecules. We expect fluoroprolines to continue to find utility as probes of key proline residues in a wide variety of proteins thanks to their subtle, yet reliable modulation of polypeptide structure, making fluoroproline substitution an important compliment to traditional genetic approaches for exploring protein structure and function. Moreover, as fluoroprolines have successfully improved the stability and folding kinetics of several important proteins, it is likely that fluoroprolines will be exploited for improving the physical, chemical, and biological properties of protein-based technologies and therapeutics.

Acknowledgements. R.W.N. was supported by Biotechnology Training Grant T32 GM008349 (NIH), the Nelson J. Leonard Graduate Fellowship of the ACS Division of Organic Chemistry sponsored by *Organic Syntheses*, and an Eastman Summer Research Award from the Eastman Chemical Company. Work on 4-fluoroprolines in the Raines laboratory was supported by Grant R01 AR044276 (NIH).

CHAPTER IV

$n \rightarrow \pi^*$ Interactions of Amides and Thioamides: Implications for Protein Stability



Carbonyl–carbonyl interactions between adjacent backbone amides have been implicated in the conformational stability of proteins. By combining experimental and computational approaches, we show that relevant amidic carbonyl groups associate through an $n \rightarrow \pi^*$ donor–acceptor interaction with an energy of at least 0.27 kcal/mol. The $n \rightarrow \pi^*$ interaction between two thioamides is 3-fold stronger than between two oxoamides due to increased overlap and reduced energy difference between the donor and acceptor orbitals. This result suggests that backbone thioamide incorporation could stabilize protein structures. Finally, we demonstrate that intimate carbonyl interactions are described more completely as donor–acceptor orbital interactions rather than dipole–dipole interactions.

This chapter has been published, in part, under the same title: Newberry, R. W.; VanVeller, B.; Guzei, I. A.; Raines, R. T. *J. Am. Chem. Soc.* **2013**, *135*, 7843–7846.

R.W.N. and B.V. conceived the project and designed the experiments and computational analyses with input from R.T.R. R.W.N. performed the experiments and computations. I.A.G. performed X-ray diffraction analysis. R.W.N. drafted the manuscript, which was edited by all authors.

Introduction

Protein architecture is mediated by a suite of noncovalent interactions within and between polypeptide chains, including the hydrophobic effect, hydrogen bonding, Coulombic interactions, and van der Waals interactions.^{3, 6} We have put forth an $n \rightarrow \pi^*$ interaction as an additional means by which peptide bonds themselves interact.^{90-91, 181, 185} In this $n \rightarrow \pi^*$ interaction, a carbonyl oxygen donates lone pair (n) electron density into another carbonyl group (Figure 4.1).

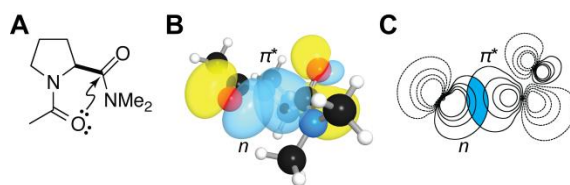


Figure 4.1 Example of an $n \rightarrow \pi^*$ interaction in a proline model system.

(A) Notion of a carbonyl–carbonyl $n \rightarrow \pi^*$ interaction in amide **3**. (B) Three-dimensional orbital rendering and (C) contour plot showing overlap of n and π^* orbitals in the *trans* *exo* conformation of **3**.

Such donation occurs when the donor and acceptor form a short contact along the Bürgi–Dunitz trajectory for nucleophilic addition.⁹⁴ These interactions have been implicated in many systems, including small molecules,^{99, 177} peptides,^{102, 104, 172-173, 187, 192-193, 195-196, 223, 343-345} proteins,^{92, 201, 203} peptoids,²⁰⁸⁻²⁰⁹ and nucleic acids.¹⁷⁵

We and others have used a torsion balance in a proline model system to characterize energetic relationships of importance to peptide and protein structure (Figure 4.2).^{90-91, 98, 172-173,}

181, 185, 196, 245, 305, 345-357

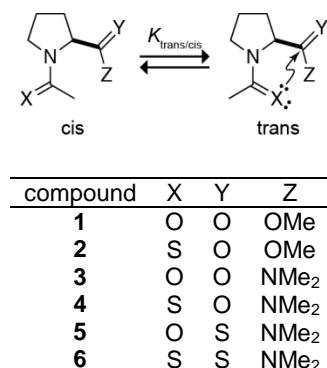


Figure 4.2 Compounds used to evaluate $n \rightarrow \pi^*$ interactions in torsion balance analyses

Both the *cis* and *trans* isomers of the *N*-acetyl proline peptide bond are populated at room temperature and can be distinguished by NMR spectroscopy due to their slow interconversion. As the $n \rightarrow \pi^*$ interaction is only possible in the *trans* isomer, the ratio of isomers ($K_{trans/cis}$) reports on the energy of the interaction. All previous studies employed esters as the $n \rightarrow \pi^*$ acceptor (**1**); because esters are more electrophilic than the amides found in proteins, those studies overestimated the strength of $n \rightarrow \pi^*$ interactions at 0.7 kcal/mol.^{90, 98} Hence, we sought to determine the energy of a prototypical $n \rightarrow \pi^*$ interaction between two amides. Primary and secondary amides can both donate hydrogen bonds to the acetyl group, obscuring the $n \rightarrow \pi^*$ interaction in our analysis;^{249, 251, 358-359} thus, we elected to examine the tertiary dimethyl amide (**3**).

Results

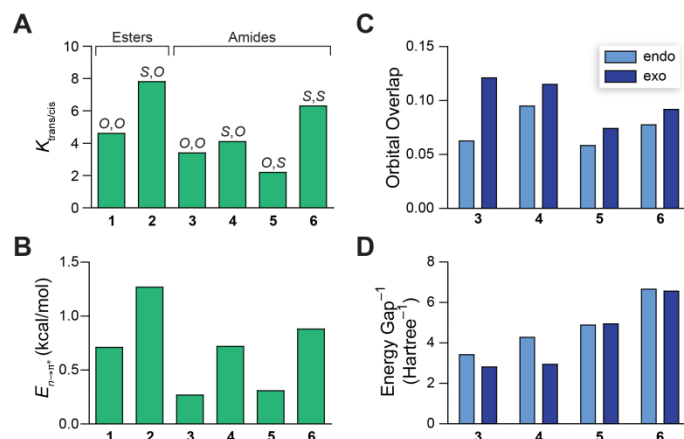


Figure 4.3 Characterization of $n \rightarrow \pi^*$ interactions in proline model compounds. (A) Values of $K_{trans/cis}$ for compounds **1–6** in D_2O at 25 °C. (B) Energy of $n \rightarrow \pi^*$ interactions in **1–6** from second-order perturbation theory. (C) Overlap integrals between the n and π^* orbitals of **3–6**. (D) Reciprocal of the energy gap between the n and π^* orbitals of **3–6**.

In D_2O , the value of $K_{trans/cis}$ for amide **3**, like that for ester **1**, is greater than unity (Figure 4.3A). We then employed density functional theory (DFT) calculations at the B3LYP/6-311+G(2d,p) level of theory with natural bond orbital (NBO) analysis to estimate the energy of the $n \rightarrow \pi^*$ interaction.³⁹ We optimized the geometry of **3** in both the C^γ -*endo* and C^γ -*exo* puckers of its pyrrolidine ring (Figure 4.4) and found the corresponding $n \rightarrow \pi^*$ energies to be 0.14 and 0.53 kcal/mol, respectively. At room temperature, proline exists ~66% in the *endo* pucker and ~34% in the *exo* pucker.¹⁸⁵ Based on this ratio, we estimate the energy of the $n \rightarrow \pi^*$ interaction in **3** to be $E_{n \rightarrow \pi^*} = 0.27$ kcal/mol (Figure 4.3B). This interaction is weaker than that in **1**, which is consistent with the lower electrophilicity of the amide acceptor relative to the ester. Importantly, because the tertiary amide is less electrophilic and more sterically encumbered than the secondary amides common in proteins, the values we report here are likely to *underestimate* the energy of an $n \rightarrow \pi^*$ interaction between most peptide bonds. Thus, we expect that $n \rightarrow \pi^*$ interactions in proteins contribute ≥ 0.27 kcal/mol of stabilization per interaction.

Previously, we demonstrated that the substitution of an amide donor (*i.e.*, **1**) with a thioamide (**2**) increases $n \rightarrow \pi^*$ donation to an ester carbonyl (Figure 4.3A).^{98, 180} This increase arises from sulfur being a better electron-pair donor than its oxygen congener. This finding suggested to us that backbone thioamide substitution could enhance the $n \rightarrow \pi^*$ interaction between carbonyl groups and stabilize the folded structures of proteins. Still, the quality of a thioamide as an $n \rightarrow \pi^*$ acceptor has been predicted only computationally.¹⁷¹ Hence, we synthesized **4–6** to evaluate thioamides as both donors and acceptors of the $n \rightarrow \pi^*$ interaction for the first time.

Replacing the donor of **3** with the larger thioamide to yield **4** increased the value of $K_{trans/cis}$ despite the added steric clash confirming that sulfur is a stronger $n \rightarrow \pi^*$ donor than oxygen (Figure 4.3A). Interestingly, whereas replacing the acceptor of **3** with a thioamide to yield **5** reduced $K_{trans/cis}$, replacing the acceptor of **4** with a thioamide to yield **6** led to an increase in $K_{trans/cis}$. Our NBO analyses show that replacing the acceptor with a thioamide reduces orbital overlap with the donor (Figure 4.3C), providing a rational basis for the value of $K_{trans/cis}$ for **5** being lower than that for **3**.

NBO analysis of thioamides **3** and **5** revealed another pertinent quantum mechanical attribute. The π^* orbital of a thioamide is lower in energy than that of an amide, reducing the energy gap between donor and acceptor orbitals (Figure 4.3D). From second-order perturbation theory, the energy released by the mixing of two orbitals is proportional to the reciprocal of the energy gap between those orbitals. Thus, though a thioamide acceptor overlaps less with an $n \rightarrow \pi^*$ donor (Figure 4.3C), a smaller energy gap between the donor and acceptor can produce more effective orbital mixing and a stronger interaction overall. The consequences are apparent in thioamide **6**, in which the $n \rightarrow \pi^*$ interaction is particularly strong at 0.88 kcal/mol

(Figure 4.3B), demonstrating that pairs of thioamides engage in significantly stronger $n \rightarrow \pi^*$ interactions than do pairs of analogous amides.

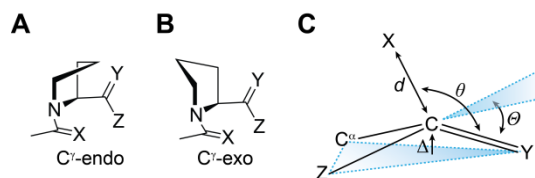


Figure 4.4 Conformational descriptors of proline model compounds.

(A) C^γ -endo and (B) C^γ -exo pyrrolidine ring puckers. (C) Parameters that denote pyramidalization of carbonyl groups due to $n \rightarrow \pi^*$ donation.

As the $n \rightarrow \pi^*$ interaction populates the π^* orbital of the acceptor, it induces pyramidalization of the acceptor toward the donor (Figure 4.4C). This distortion is detectable by X-ray diffraction analysis and can provide a signature of $n \rightarrow \pi^*$ interactions in small molecules and peptides.^{98-99, 102, 176-177, 180} Hence, we conducted X-ray diffraction analysis of crystalline **4–6** to search for this signature of an $n \rightarrow \pi^*$ interaction. Thioamide **4** crystallized as its *cis* isomer and therefore does not show an $n \rightarrow \pi^*$ interaction, leaving the acceptor nearly planar (Table 4.1).

Table 4.1 Conformational parameters of thioamides^a

compound	conformation	d (Å)	θ (deg)	Δ (Å)	Θ (deg)
4	<i>cis, endo</i>	4.6158(14)	66.96(7)	0.0035(14)	0.43(17)
5	<i>trans, endo</i>	3.2529(12)	92.19(4)	0.0237(8)	2.61(9)
6	<i>trans, endo</i>	3.4248(16)	96.11(6)	0.0243(17)	2.70(19)
6	<i>trans, exo</i>	3.2433(15)	101.92(7)	0.0392(16)	4.36(18)

^aFrom X-ray diffraction analysis of the crystalline compound. Parameters are defined in Figure 4.4.

Thioamides **5** and **6** both crystallized as their *trans* isomer, with **6** crystallizing in both pyrrolidine ring puckers. In both **5** and **6**, the acceptor carbon is pyramidalized toward the donor significantly more than in **4**, denoting a stronger $n \rightarrow \pi^*$ interaction. In addition, pyramidalization of the acceptor in both conformations of **6** is greater than in **5**, which is consistent with the

stronger $n \rightarrow \pi^*$ interaction in **6**. Moreover, the greater pyramidalization of the acceptor in **6-*exo*** than in **6-*endo*** confirms that the *exo* ring pucker of proline promotes stronger $n \rightarrow \pi^*$ interactions.¹⁸³ Indeed, the pyramidalization in **6-*exo*** is among the largest observed to date in this proline model system.^{98, 360} These observations are also consistent with the pyramidalization in crystal structures of thioamide-containing peptides (Figure 4.5).

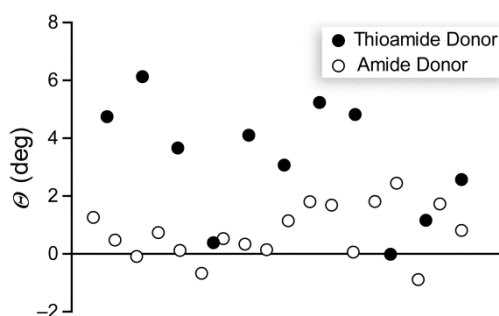


Figure 4.5 Carbonyl pyramidalization in peptides containing thioamide bonds.

The abscissa is used to distribute the values randomly. The parameter Θ is defined in Figure 4.4. In the Cambridge Structural database (CSD) were found the structures of 23 peptides in which at least one peptide bond was replaced with a thioamide. Residues that received an $n \rightarrow \pi^*$ interaction were identified by applying a geometric operational definition: a donor chalcogen contacts an acceptor carbonyl group within the sum of their van der Waals radii (3.22 Å for $O \cdots C$; 3.50 Å for $S \cdots C$) and along the Bürgi–Dunitz trajectory ($99\text{--}119^\circ X \cdots C=O$).

These data further establish the quantum mechanical nature of intimate carbonyl–carbonyl interactions. Some have argued that these interactions are primarily dipolar in nature.^{20, 95, 97} The dipole moment of an amide is greater than that of an ester,³⁶¹ so if a dipolar interaction is dominant, replacing the ester of **1** with the amide in **3** should cause an increase in $K_{trans/cis}$, which is contrary to observation (Figure 4.3A). Moreover, as a thioamide has a still larger dipole moment than an amide,^{362–363} a dipolar origin would predict a larger value of $K_{trans/cis}$ in thioamide **5** than in amide **3**. That was not observed in our experiments. Finally, in a dipolar interaction, neither of the participating groups has a defined role; rather, they interact symmetrically. Thus, if intimate carbonyl interactions are dipolar in nature, then substituting

either amide with a thioamide should have a comparable effect on $K_{trans/cis}$. The conformational preferences of **4** and **5** suggest otherwise: substituting the $n \rightarrow \pi^*$ donor amide with a thioamide increases $K_{trans/cis}$, whereas replacing the acceptor decreases $K_{trans/cis}$. These data affirm that a dipolar mechanism is insufficient to describe intimate interactions between carbonyl groups. Instead, the data are more consistent with an electronic donor–acceptor effect like the $n \rightarrow \pi^*$ interaction.

Conclusions

Individual $n \rightarrow \pi^*$ interactions between amides are relatively weak. In abundance, however, they could make a significant contribution to the conformational stability of a protein. We note another implication as well. Shifts in the equilibrium between α -helices and β -sheets have been implicated in amyloid fibrillogenesis.^{5, 364-365} Hydrogen bonding, which is operative in both α -helices and β -sheets,^{37, 46-47} is unlikely to affect this equilibrium decisively. In contrast, $n \rightarrow \pi^*$ interactions are common in α -helices but not β -sheets,⁹² and thus could play a critical role in the maintenance of protein homeostasis. In addition, our finding that the $n \rightarrow \pi^*$ interaction between two thioamides is 3-fold stronger than that between two amides (Figure 4.3B) encourages efforts to exploit thioamides to enhance conformational stability in peptides and proteins. Finally, as these interactions are not included in conventional force fields, we argue that accounting for the $n \rightarrow \pi^*$ interaction could improve the accuracy of computational investigations of proteins.

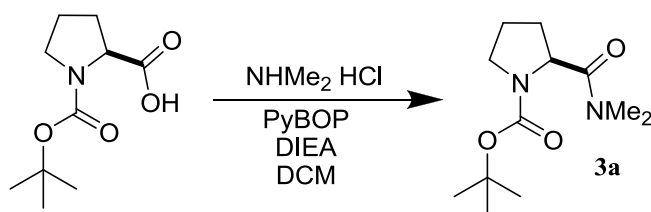
Acknowledgements. We are grateful to B. S. Dolinar for assistance with X-ray diffraction analysis, and to A. Choudhary and G. R. Krow for contributive discussions. R.W.N. was supported by NIH Biotechnology Training Grant T32 GM008349. This work was funded by

grants R01 AR044276 (NIH) and CHE-1124944 (NSF). NMR spectroscopy was supported by grants P41 GM103399 (NIH) and P41 GM66326 (NIH). High-performance computing was supported by grant CHE-0840494 (NSF).

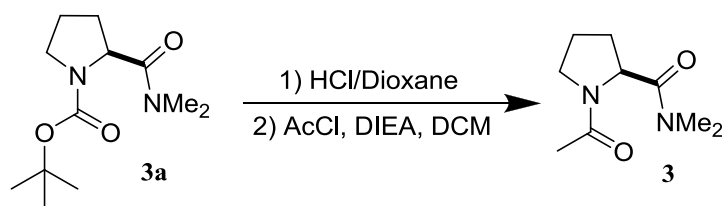
Methods

General Experimental. Commercial chemicals were of reagent grade or better, and were used without further purification. Proline starting materials and PyBOP were purchased from Chem-Impex International (Wood Dale, IL). All other chemicals were obtained from Sigma–Aldrich. Anhydrous THF and CH₂Cl₂ were obtained from CYCLE-TAINER[®] solvent delivery systems (J. T. Baker, Phillipsburg, NJ). Reactions were monitored by thin-layer chromatography with visualization by UV light or staining with KMnO₄. Flash chromatography was performed with columns of silica gel 60, 230–400 mesh (Silicycle, Québec City, Canada). The removal of solvents and other volatile materials “under reduced pressure” refers to the use of a rotary evaporator at water-aspirator pressure (<20 torr) and a water bath of <45 °C. All reported yields are unoptimized.

Instrumentation. NMR spectra were acquired at ambient temperature with a Bruker Avance III 500 MHz spectrometer (¹H, 500 MHz; ¹³C, 125 MHz) in the National Magnetic Resonance Facility at Madison (NMRFAM). ¹³C spectra were proton-decoupled. Mass Spectrometry was performed with a Micromass LCT (electrospray ionization, ESI) instrument in the Mass Spectrometry Facility of the Department of Chemistry at the University of Wisconsin–Madison. X-Ray data were collected in the Molecular structure Laboratory of the Department of Chemistry at the University of Wisconsin–Madison.



***N*-Boc-(2*S*)-proline dimethyl amide (3a).** *N*-Boc-(2*S*)-proline (3.0 g, 13.9 mmol) and PyBOP (8.0 g, 15.3 mmol) were dissolved in DCM (100 mL). *N,N*-diisopropylethylamine (7.3 mL, 41.8 mmol) and dimethylamine hydrochloride (1.1 g, 13.9 mmol) were added, and the resulting solution was stirred overnight. DCM was removed under reduced pressure, and the resulting oil was purified by flash chromatography in EtOAc to yield **3a** as a yellow oil (1.8 g, 53%). ¹H NMR (CDCl₃, 1:1 mixture of two rotamers) δ 4.69 (dd, *J* = 8.4, 3.2 Hz, 0.5H), 4.56 (dd, *J* = 8.2, 4.3 Hz, 0.5H), 3.66–3.40 (m, 2H), 3.11 (s, 1.5H), 3.08 (s, 1.5H), 2.99 (s, 1.5H), 2.97 (s, 1.5H), 2.25–2.00 (m, 2H), 1.85 (m, 2H), 1.47 (s, 4.5H), 1.41 (s, 4.5H); ¹³C NMR δ 172.8, 172.4, 154.6, 153.9, 79.5, 79.4, 56.5, 56.4, 46.8, 46.6, 37.0, 36.0, 30.4, 29.6, 28.6, 28.4, 24.2, 23.7; ESI-MS: [M + Na]⁺ calculated 265.1523, observed 265.1536.

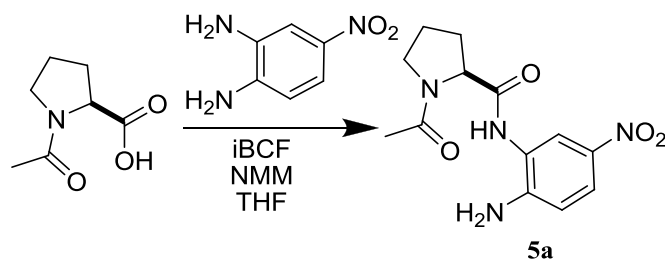


***N*-acetyl-(2*S*)-proline dimethyl amide (3).** *N*-Boc-(2*S*)-proline dimethyl amide (**3a**, 0.90 g, 3.7 mmol) was dissolved in HCl (4 M in dioxane, 5 mL) and stirred for 1 h. Solvent was removed under reduced pressure. The resulting oil (0.63 g, 2.8 mmol) was dissolved in DCM (5 mL) with *N,N*-diisopropylethylamine (1 mL, 5.6 mmol). Acetyl chloride (0.35 mL, 4.8 mmol) was added dropwise, and the reaction mixture was stirred overnight. Solvent was removed under reduced pressure, and the resulting oil was partitioned between EtOAc/hexane and 1 M NaOH.

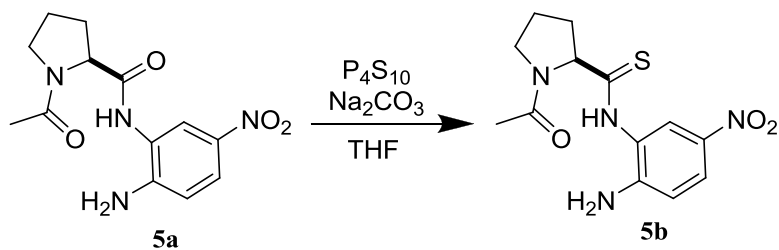
The aqueous portion was evaporated, and the residue was extracted with DCM. The organic portion was evaporated to dryness, and the resulting oil was purified by flash chromatography in EtOAc/Hex to yield **3** as a clear oil. ^1H NMR (D_2O , 3.4:1 mixture of two rotamers) δ 4.89 (dd, J = 9.0, 3.2 Hz, 0.23H), 3.53 (t, J = 6.9 Hz, 1.54H) 3.41 (m, 0.46H), 3.03 (s, 0.69H), 3.01 (s, 2.31H), 2.86 (s, 0.69H), 2.82 (s, 2.31H), 2.31 (m, 0.23H), 2.22 (m, 0.77H), 1.99 (s, 2.31H), 1.79 (s, 0.69H), 1.95–1.72 (m, 3H); ^{13}C NMR δ 173.3, 173.1, 172.9, 172.4, 59.1, 57.2, 48.7, 47.2, 36.9, 36.8, 30.3, 28.7, 22.8, 22.4, 21.2, 21.0; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 185.1285, observed 185.1285.



***N*-thioacetyl-(2*S*)-proline dimethyl amide (4).** A solution of **3** (0.1 g, 0.5 mmol) and the Lawesson's reagent (0.22 g, 0.5 mmol) were refluxed in THF (5 mL) overnight. Solvent was removed under reduced pressure, and the resulting oil was purified by flash chromatography with EtOAc/hexane to afford **4** as a white solid. ^1H NMR (D_2O , 4.1:1 mixture of two rotamers) δ 5.13 (m, 1H), 3.84–3.68 (m, 2H) 3.04 (s, 2.41H), 3.02 (s, 0.59H), 2.85 (s, 0.59H), 2.82 (s, 2.41H), 2.46 (s, 2.41H), 2.44–2.22 (m, 1H), 2.08–1.82 (m, 3H); ^{13}C NMR δ 197.7, 197.6, 172.0, 171.4, 64.2, 62.8, 55.3, 52.9, 36.9, 36.9, 35.8, 35.6, 31.6, 31.4, 30.6, 28.7, 24.4, 22.5; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 201.1057, observed 201.1054.

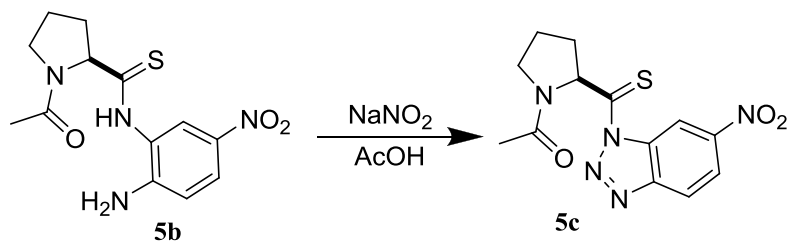


***N*-acetyl-(2*S*)-proline 2-amine-5-nitroanilide (5a).**³⁶⁶ *N*-methylmorpholine (2.8 mL, 25.4 mmol) was added to a solution of *N*-acetyl-(2*S*)-proline (2.0 g, 12.7 mmol) in THF (100 mL) at $-20\text{ }^{\circ}\text{C}$. Isobutyl chloroformate (1.6 mL, 12.7 mmol) was added dropwise, and the resulting solution was stirred for 10 min. 4-Nitro-1,2-phenylenediamine (1.96 g, 12.7 mmol) was added, and the reaction mixture was stirred at $-20\text{ }^{\circ}\text{C}$ for 3 h before warming to room temperature and stirring overnight. THF was removed under reduced pressure, and the resulting yellow solid was washed with 1 M NaH_2PO_4 , followed by EtOAc. Trituration in EtOAc/hexane afforded **5a** as a yellow solid (2.0 g, 51%). ^1H NMR (CDCl_3) δ 8.83 (s, 1H), 8.39 (d, $J = 2.5$ Hz, 1H), 7.93 (dd, $J = 8.9, 2.5$ Hz, 1H), 6.68 (d, $J = 8.9$ Hz, 1H), 4.80 (s, 2H), 4.65 (dd, $J = 8.6, 2.6$ Hz, 1H), 3.63 (m, 1H), 3.52 (m, 1H), 2.54 (m, 1H), 2.22 (m, 1H), 2.18 (s, 3H), 2.04 (m, 2H); ^{13}C NMR (DMSO) δ 171.5, 169.22, 150.5, 135.2, 123.7, 122.8, 121.1, 113.4, 60.0, 47.9, 29.7, 24.6, 22.5; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 293.1245, observed 293.1258.

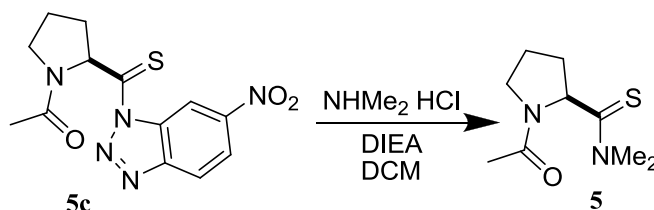


***N*-acetyl-(2*S*)-proline 2-amine-5-nitrothioanilide (5b).**³⁶⁶ P_4S_{10} (1.5 g, 3.3 mmol) and Na_2CO_3 (0.36 g, 3.3 mmol) were added to THF (100 mL), and the resulting solution was stirred for 1 h before cooling to $0\text{ }^{\circ}\text{C}$. compound **5a** (2.0 g, 6.5 mmol) was added, and the reaction mixture was

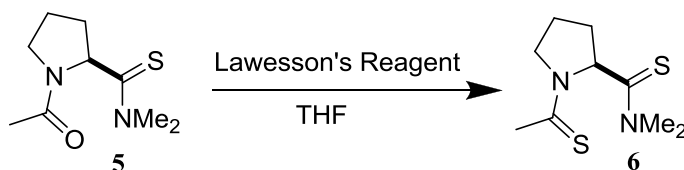
stirred at 0 °C for 3 h. The reaction mixture was filtered through Celite, evaporated to dryness, and dissolved in 3:1 EtOAc:hexane (100 mL). The solution was washed with saturated NaHCO₃(aq). The aqueous portion was back extracted with 3:1 EtOAc:hexane (100 mL), the organic portions were combined, dried over Na₂SO₄(s), and evaporated to dryness, yielding **5b** as a yellow solid (1.0 g, 47%). ¹H NMR (DMSO) δ 11.07 (s, 1H), 7.98 (dd, *J* = 9.1, 2.7 Hz, 1H), 7.83 (d, *J* = 2.7 Hz, 1H), 6.77 (d, *J* = 9.1 Hz, 1H), 6.47 (s, 2H), 4.75 (dd, *J* = 8.5, 4.8 Hz, 1H), 3.69 (m, 1H), 3.60 (m, 1H), 2.32 (m, 1H), 2.10 (m, 1H), 1.99 (s, 3H), 1.91 (m, 2H); ¹³C NMR δ 206.5, 170.1, 151.1, 135.2, 125.2, 122.3, 113.61, 67.3, 48.5, 32.6, 24.4, 22.8; ESI-MS: [M + Na]⁺ calculated 331.0836, found 331.0850.



1-(*N*-acetyl-(2*S*)-thiopropyl)-6-nitrobenzotriazole (5c**).**³⁶⁶ To a solution of **5b** (1.0 g, 3.1 mmol) in 95% AcOH/H₂O at 0 °C was added NaNO₂ (0.32 g, 4.6 mmol) portionwise. The reaction was stirred for 30 min before adding 100 mL water. The aqueous solution was extracted with DCM (150 mL), and the organic portion was washed with saturated NaHCO₃(aq), dried over Na₂SO₄(s), and evaporated to dryness, yielding **5c** as an orange solid (0.62 g, 62%). ¹H NMR (CDCl₃) δ 9.70 (d, *J* = 2.1 Hz, 1H), 8.44 (dd, *J* = 9.0, 2.1 Hz, 1H), 8.30 (d, *J* = 9.0 Hz, 1H), 6.37 (dd, *J* = 9.3, 2.8 Hz, 1H), 3.84 (m, 1H), 3.73 (m, 1H), 2.64 (m, 1H), 2.20 (m, 2H), 2.17 (s, 3H), 2.14 (m, 1H); ¹³C NMR δ 206.9, 169.4, 149.6, 148.9, 132.2, 122.2, 121.5, 113.1, 67.5, 48.6, 33.2, 24.4, 22.8; ESI-MS: [M + Na]⁺ calculated 342.0632, found 342.0636.



***N*-acetyl-(2*S*)-proline dimethyl thioamide (5).** compound **5c** (0.37 g, 1.2 mmol) was dissolved in DCM (20 mL) with *N,N*-diisopropylethylamine (0.40 mL, 2.4 mmol). Dimethylamine hydrochloride (0.094 g, 1.2 mmol) was added, and the reaction mixture was stirred overnight. Solvent was removed under reduced pressure, and the resulting oil was purified by flash chromatography, first with EtOAc/MeOH and then with DCM/MeOH. The resulting product was decolorized over charcoal to yield **5** as a white solid (0.10 g, 43%). ^1H NMR (D_2O , 2.2:1 mixture of two rotamers) δ 5.09 (dd, $J = 8.4, 2.9$ Hz, 0.31H), 4.94 (dd, $J = 8.7, 4.4$ Hz, 0.69H), 3.66–3.40 (m, 2H) 3.35 (s, 0.94H), 3.33 (s, 3H), 3.31 (s, 2.06H), 2.40–2.20 (m, 1H), 1.98 (s, 2.06H), 1.77 (s, 0.94H), 2.08–1.68 (m, 3H); ^{13}C NMR δ 201.7, 201.4, 173.3, 172.3, 64.8, 62.9, 49.0, 47.4, 45.3, 45.1, 41.8, 41.7, 31.9, 30.5, 23.6, 21.9, 21.7, 21.6; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 201.1057, observed 201.1056.



***N*-Thioacetyl-(2*S*)-proline dimethyl thioamide (6).** Lawesson's reagent (0.10 g, 0.25 mmol) and **5** (0.10 g, 0.50 mmol) were heated to reflux in THF (10 mL) overnight. Solvent was removed under reduced pressure, and the resulting oil was purified by flash chromatography in DCM/MeOH to yield **6** as a white solid. ^1H NMR (D_2O , 6.3:1 mixture of two rotamers) δ 5.34 (dd, $J = 8.7, 4.2$ Hz, 0.86H), 5.27 (dd, $J = 8.7, 2.5$ Hz, 0.14H), 3.77 (m, 0.27H) 3.73 (t, $J = 7.0$

Hz, 1.73H), 3.29 (s, 2.59H), 3.8 (s, 0.41H), 3.27 (s, 2.59H), 3.26 (s, 0.41H), 2.40 (s, 2.59H), 2.21 (s, 0.41H), 2.30–2.15 (m, 2H), 1.95–1.75 (m, 2H); ^{13}C NMR δ 199.6, 198.9, 197.6, 197.3, 69.5, 68.4, 55.5, 53.1, 44.9, 44.6, 41.7, 41.6, 32.2, 31.8, 31.8, 30.3, 24.0, 21.9; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 217.0828, observed 217.0837.

Determination of $K_{trans/cis}$ for compounds 3–6. Each compound (5–10 mg) was dissolved in D_2O with enough CD_3OD to solubilize the compound (less than 10% of total volume), and equilibrated for a day. ^1H NMR spectra were acquired and analyzed with the Mestrenova software package.³⁶⁷ Proton assignments for *cis* and *trans* isomers were confirmed by NOEs observed between the C^δ protons and acetyl protons in ^1H – ^1H NOESY spectra. Amide methyl signals for the *cis* and *trans* isomers were fitted, and the value of $K_{trans/cis}$ was determined from the relative areas of the calculated fits.

Crystal structure determination for compound 4. Compound **4** was dissolved in water, and slow evaporation afforded diffraction quality crystals after approximately one week. A colorless crystal with approximate dimensions $0.175 \times 0.154 \times 0.131 \text{ mm}^3$ was selected under oil under ambient conditions and attached to the tip of a MiTeGen MicroMount[®]. The crystal was mounted in a stream of cold nitrogen at 100(1) K and centered in the X-ray beam by using a video camera. The crystal evaluation and data collection were performed on a Bruker SMART APEXII diffractometer with Cu K_α ($\lambda = 1.54178 \text{ \AA}$) radiation and a diffractometer-to-crystal distance of 4.03 cm. The initial cell constants were obtained from three series of ω scans at different starting angles. Each series consisted of 41 frames collected at intervals of 0.6° in a 25° range about ω with an exposure time of 2 s per frame. The reflections were indexed successfully

by an automated indexing routine built in the APEXII program. The final cell constants were calculated from a set of 9561 strong reflections from the actual data collection. data were collected by using the full sphere data collection routine to survey the reciprocal space to the extent of a full sphere to a resolution of 0.82 Å. A total of 12276 data were harvested by collecting 14 sets of frames with 0.6° scans in ω with an exposure time 15–40 s per frame. These highly redundant datasets were corrected for Lorentz and polarization effects. The absorption correction was based on fitting a function to the empirical transmission surface as sampled by multiple equivalent measurements.³⁶⁸ The systematic absences in the diffraction data were uniquely consistent for the space group $P2_12_12_1$ that yielded chemically reasonable and computationally stable results of refinement.³⁶⁹ A successful solution by the direct methods provided most non-hydrogen atoms from the *E*-map. The remaining non-hydrogen atoms were located in an alternating series of least-squares cycles and difference Fourier maps. All non-hydrogen atoms were refined with anisotropic displacement coefficients. All hydrogen atoms were included in the structure factor calculation at idealized positions and were allowed to ride on the neighboring atoms with relative isotropic displacement coefficients. The molecule has a stereocenter at atom C1. The absolute configuration of the molecule was unequivocally established through anomalous dispersion effects. The molecule is the *S* enantiomer. The final least-squares refinement of 122 parameters against 1917 data resulted in residual *R* (based on F^2 for $I \geq 2\sigma$) and *wR* (based on F^2 for all data) of 0.0236 and 0.0636, respectively. The final difference Fourier map was featureless.

Crystal structure determination of compound 5. Compound **5** was dissolved in water, and slow evaporation afforded diffraction quality crystals after approximately one week. A colorless

crystal with approximate dimensions $0.40 \times 0.21 \times 0.12 \text{ mm}^3$ was selected under oil under ambient conditions and attached to the tip of a MiTeGen MicroMount[®]. The crystal was mounted in a stream of cold nitrogen at 100(1) K and centered in the X-ray beam by using a video camera. The crystal evaluation and data collection were performed on a Bruker Quazar SMART APEXII diffractometer with Mo K_α ($\lambda = 0.71073 \text{ \AA}$) radiation and the diffractometer-to-crystal distance of 4.96 cm. The initial cell constants were obtained from three series of ω scans at different starting angles. Each series consisted of 12 frames collected at intervals of 0.5° in a 6° range about ω with an exposure time of 1 s per frame. The reflections were indexed successfully by an automated indexing routine built in the APEXII program suite. The final cell constants were calculated from a set of 9696 strong reflections from the actual data collection. The data were collected by using the full sphere data collection routine to survey the reciprocal space to the extent of a full sphere to a resolution of 0.70 \AA . A total of 28497 data were harvested by collecting 6 sets of frames with 0.4° scans in ω and ϕ with exposure times of 10 s per frame. These highly redundant datasets were corrected for Lorentz and polarization effects. The absorption correction was based on fitting a function to the empirical transmission surface as sampled by multiple equivalent measurements.³⁶⁸ The systematic absences in the diffraction data were uniquely consistent for the space group $P2_12_12_1$ that yielded chemically reasonable and computationally stable results of refinement.³⁶⁹ A successful solution by the direct methods provided most non-hydrogen atoms from the E -map. The remaining non-hydrogen atoms were located in an alternating series of least-squares cycles and difference Fourier maps. All non-hydrogen atoms were refined with anisotropic displacement coefficients. All hydrogen atoms were located in the difference map and refined independently. The structure is chiral, the compound is chiral, the configuration of the chiral center C1 is *S*. The final least-squares

refinement of 167 parameters against 3029 data resulted in residuals R (based on F^2 for $I \geq 2\sigma$) and wR (based on F^2 for all data) of 0.0212 and 0.0577, respectively. The final difference Fourier map was featureless.

Crystal structure determination of compound 6. Compound **6** was dissolved in water, and slow evaporation afforded diffraction quality crystals after approximately one week. A colorless crystal with approximate dimensions $0.777 \times 0.541 \times 0.464 \text{ mm}^3$ was selected under oil under ambient conditions and attached to the tip of a MiTeGen MicroMount[®]. The crystal was mounted in a stream of cold nitrogen at 100(1) K and centered in the X-ray beam by using a video camera. The crystal evaluation and data collection were performed on a Bruker SMART APEXII diffractometer with Cu K_α ($\lambda = 1.54178 \text{ \AA}$) radiation and the diffractometer-to-crystal distance of 4.03 cm. The initial cell constants were obtained from three series of ω scans at different starting angles. Each series consisted of 41 frames collected at intervals of 0.6° in a 25° range about ω with an exposure time of 1 s per frame. The reflections were successfully indexed by an automated indexing routine built in the APEXII program. The final cell constants were calculated from a set of 9875 strong reflections from the actual data collection. The data were collected by using the full sphere data collection routine to survey the reciprocal space to the extent of a full sphere to a resolution of 0.8 \AA . A total of 35128 data were harvested by collecting 19 sets of frames with 0.5° scans in ω and ϕ with an exposure time 2–5 s per frame. These highly redundant datasets were corrected for Lorentz and polarization effects. The absorption correction was based on fitting a function to the empirical transmission surface as sampled by multiple equivalent measurements.³⁶⁸ The systematic absences in the diffraction data were uniquely consistent for the space group $P2_12_12_1$ that yielded chemically reasonable and computationally

stable results of refinement.³⁶⁹ A successful solution by the direct methods provided most non-hydrogen atoms from the *E*-map. The remaining non-hydrogen atoms were located in an alternating series of least-squares cycles and difference Fourier maps. All non-hydrogen atoms were refined with anisotropic displacement coefficients. All hydrogen atoms were included in the structure factor calculation at idealized positions and were allowed to ride on the neighboring atoms with relative isotropic displacement coefficients. There are two symmetry-independent molecules with identical composition and opposite handedness in the asymmetric unit. There are two chiral centers: C4 (*R*) and C4a (*S*). The absolute configuration was unequivocally established based on established by anomalous-dispersion effects in diffraction measurements on the crystal. The final least-squares refinement of 242 parameters against 4228 data resulted in residuals *R* (based on F^2 for $I \geq 2\sigma$) and *wR* (based on F^2 for all data) of 0.0254 and 0.0660, respectively. The final difference Fourier map was featureless.

Computational methodology. Conformational preferences, orbital energies, orbital overlaps, and $n \rightarrow \pi^*$ energies were determined using density functional theory implemented in Gaussian 09.³⁷⁰ Geometry optimizations and frequency calculations were performed at the B3LYP/6-311+G (2d,p) level of theory for *endo* and *exo* conformers of *cis* and *trans* geometries of compounds **3-6**. Frequency calculations of optimized structures yielded no imaginary frequencies, indicating a true minimum on the potential energy surface. Energies were corrected by the zero-point vibrational energy. optimized geometries were subjected to analysis by NBO 5.0 at the B3LYP/6-311+G(2d,p) level of theory implemented in Gaussian 09.³⁷¹

Table 4.S1 Data depicted in Figure 4.3.

	$K_{trans/cis}$	$E_{n \rightarrow \pi^*}$ (kcal/mol)		Overlap Integral		Energy Gap (Hartree)	
		<i>endo</i>	<i>exo</i>	<i>endo</i>	<i>exo</i>	<i>endo</i>	<i>exo</i>
1 ^a	4.6	0.42	1.29	ND	ND	ND	ND
2 ^a	7.8	0.86	2.08	ND	ND	ND	ND
3	3.4	0.14	0.53	0.0626	0.1212	0.29179	0.35461
4	4.1	0.64	0.89	0.0949	0.1150	0.23381	0.33908
5	2.2	0.23	0.48	0.0584	0.0743	0.20480	0.20217
6	6.3	0.72	1.19	0.0775	0.0918	0.15013	0.15244

^aData from Choudhary, et al. *J. Am. Chem. Soc.* **2009**, *131*, 7244-7246.

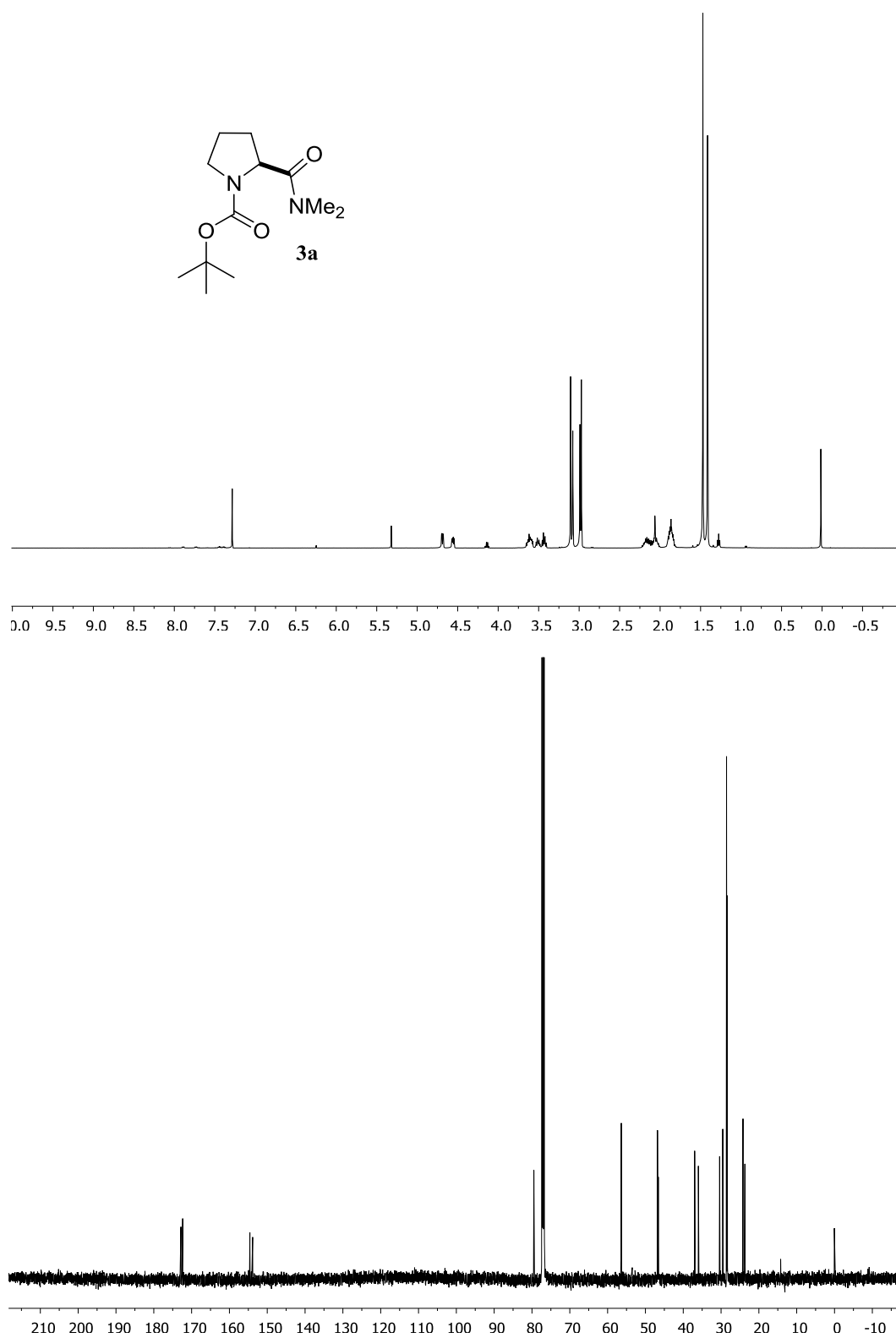


Figure 4.S1 ^1H and ^{13}C NMR spectra for compound **3a**.

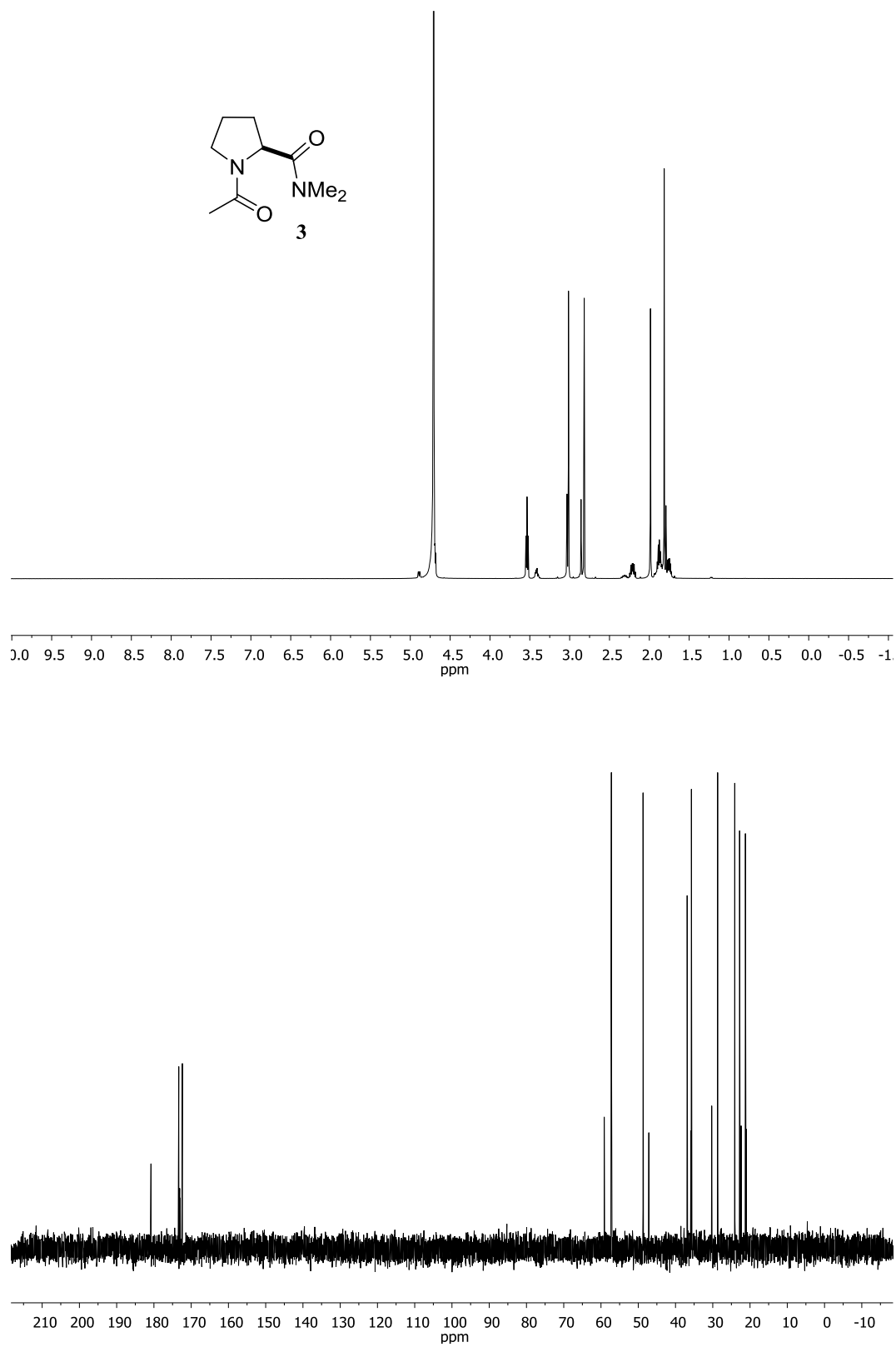


Figure 4.S2 ^1H and ^{13}C NMR spectra for compound **3**.

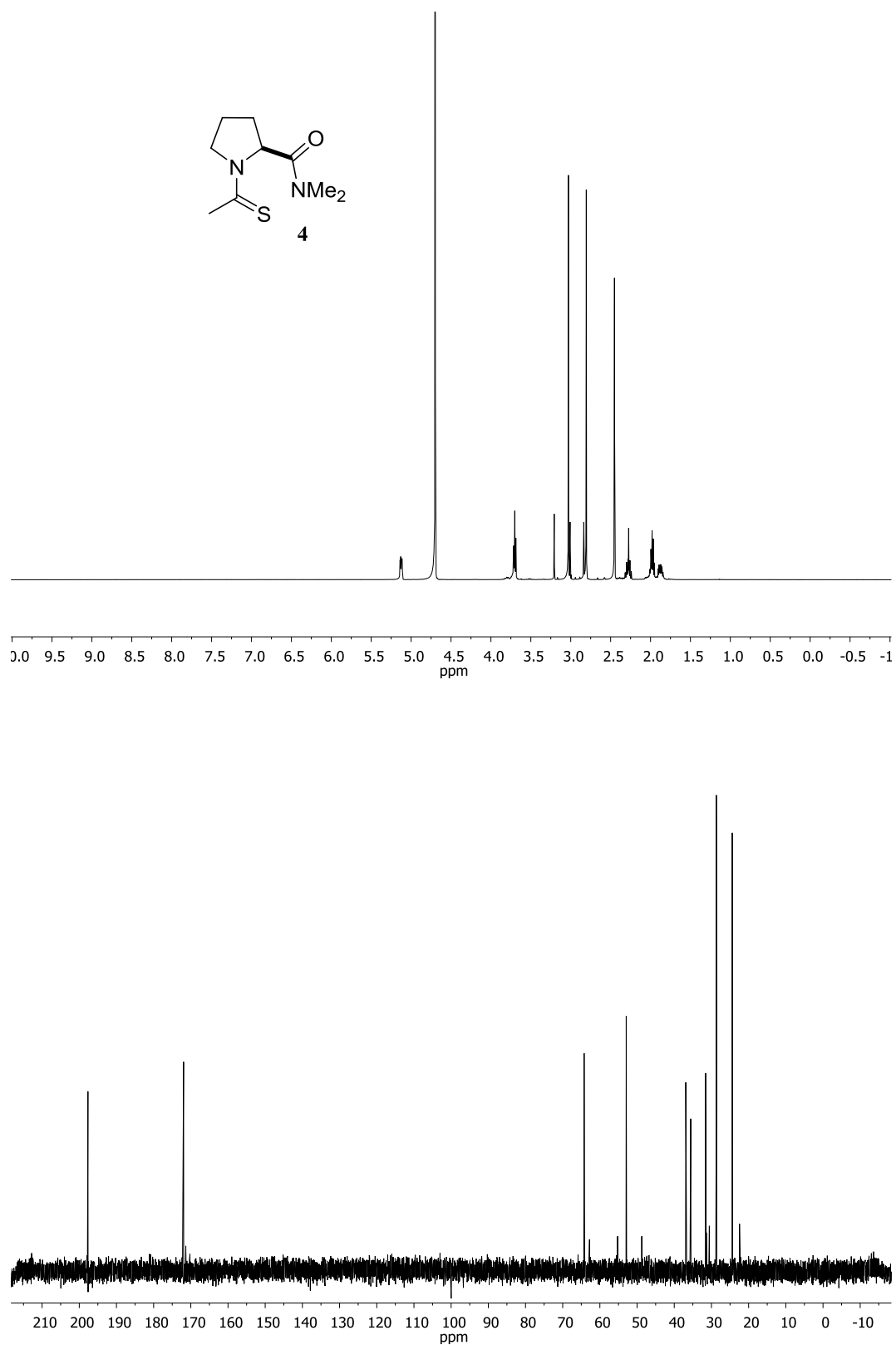


Figure 4.S3 ^1H and ^{13}C NMR spectra for compound **4**.

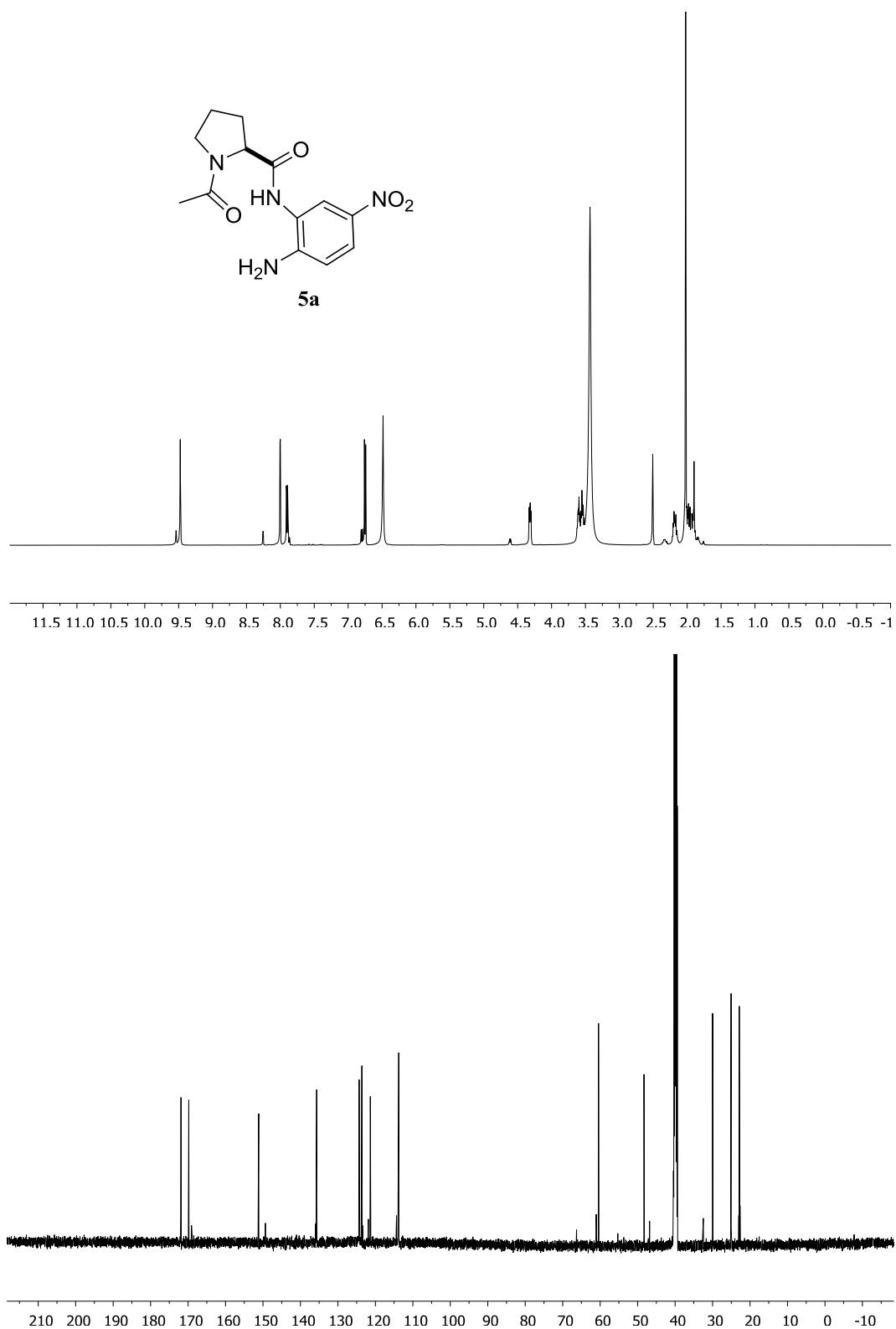


Figure 4.S4 ^1H and ^{13}C NMR spectra for compound **5a**.

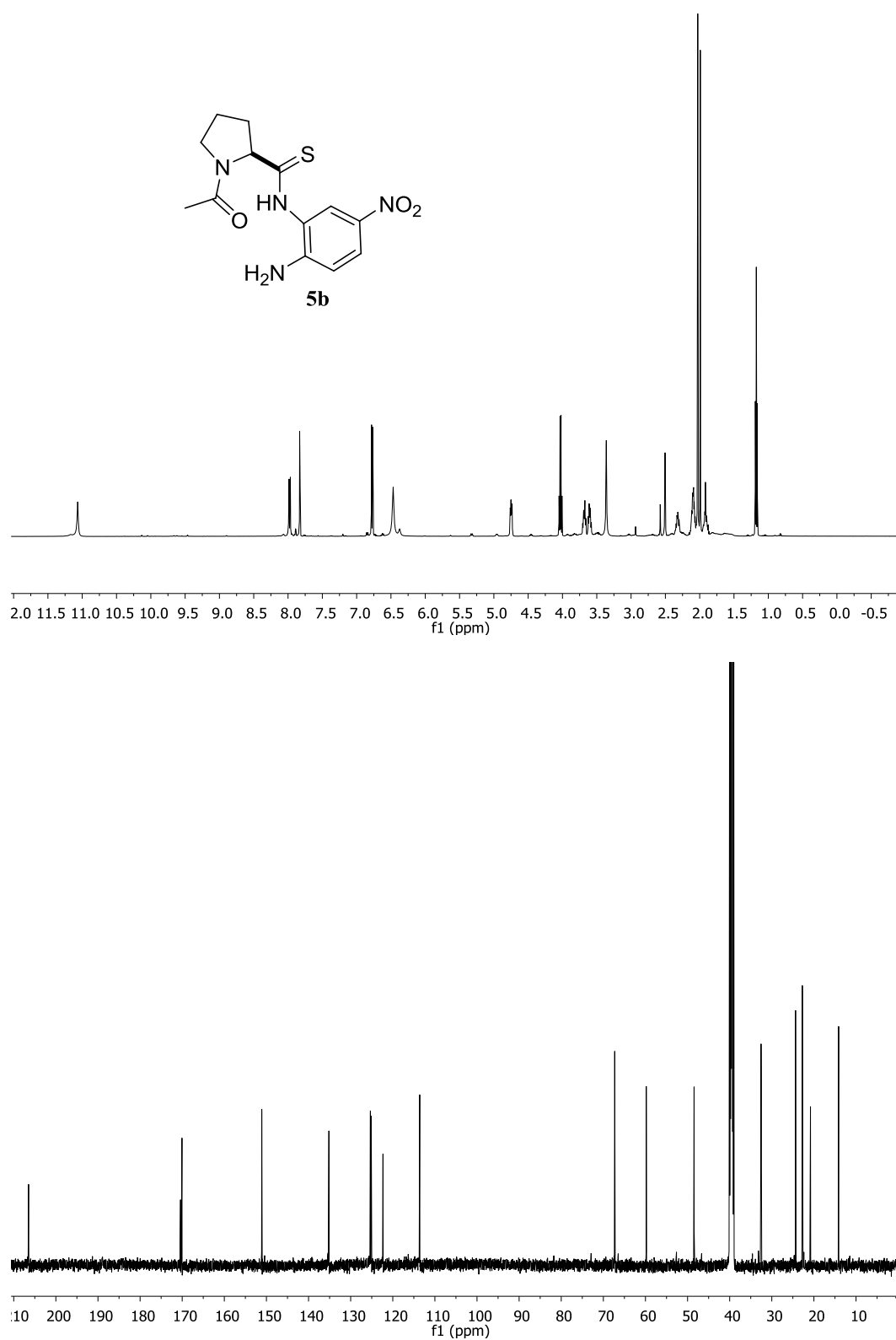


Figure 4.S5 ^1H and ^{13}C NMR spectra for compound **5b**. Residual EtOAc.

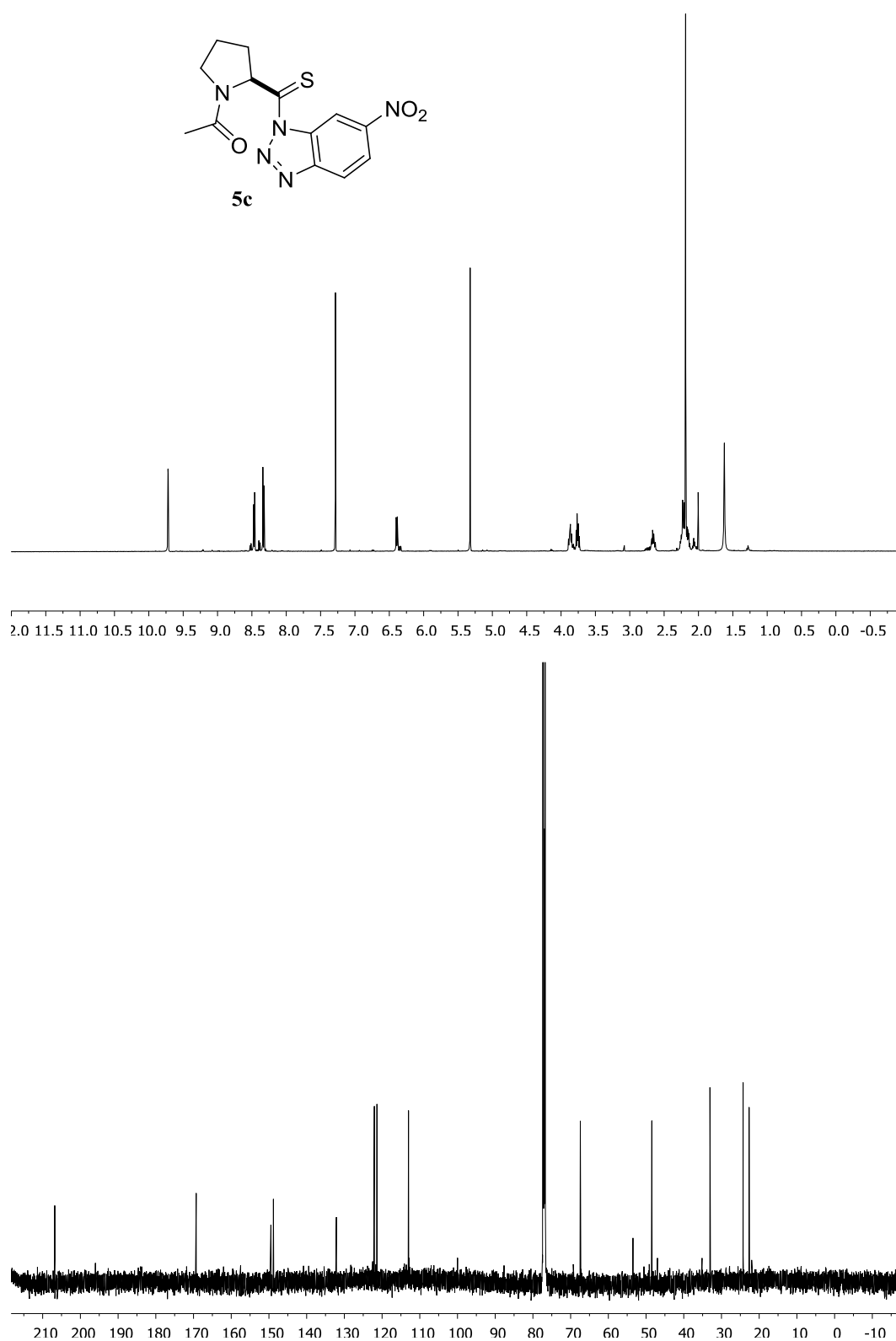


Figure 4.S6 ^1H and ^{13}C NMR spectra for compound **5c**. Residual DCM.

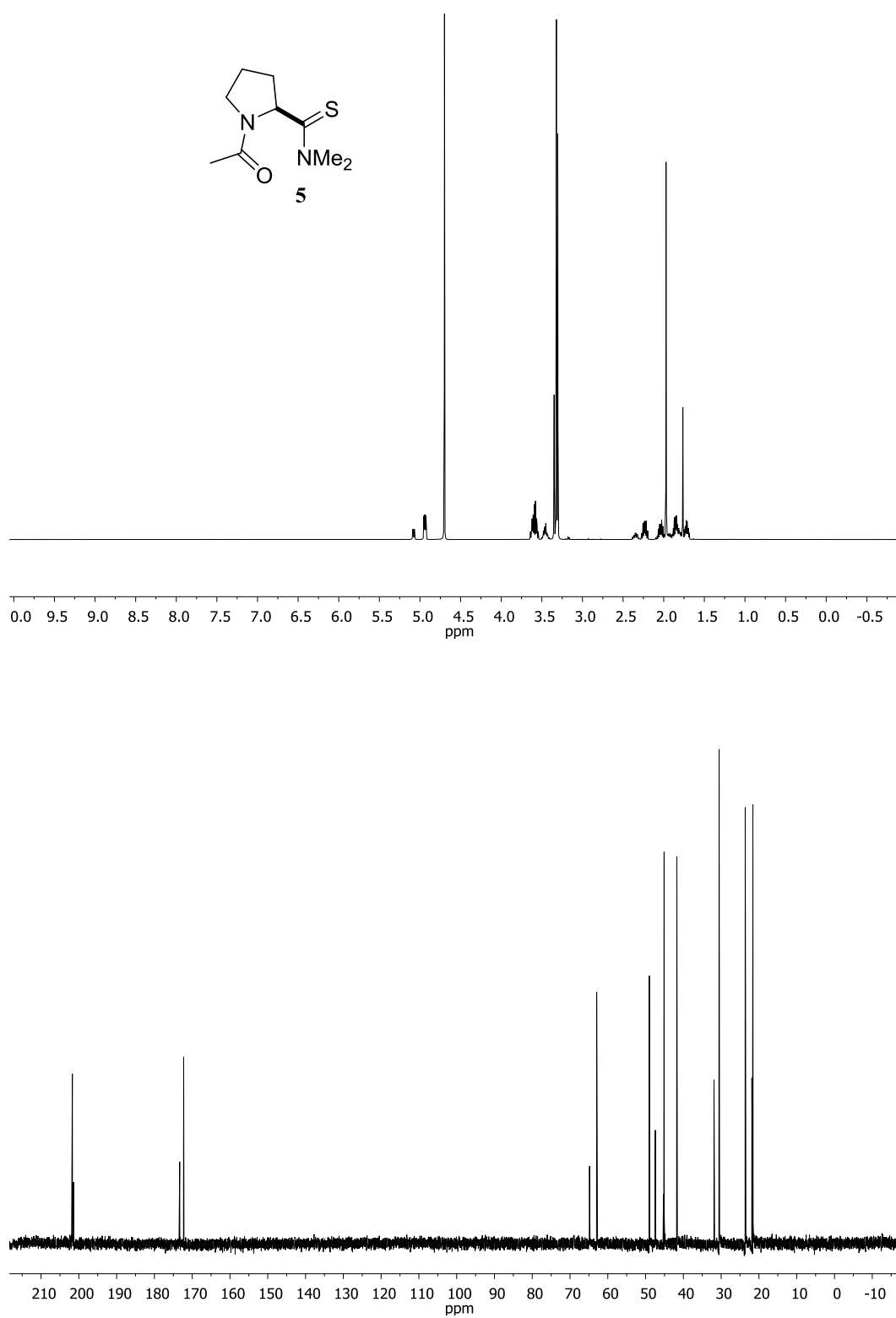


Figure 4.S7 ^1H and ^{13}C NMR spectra for compound **5**.

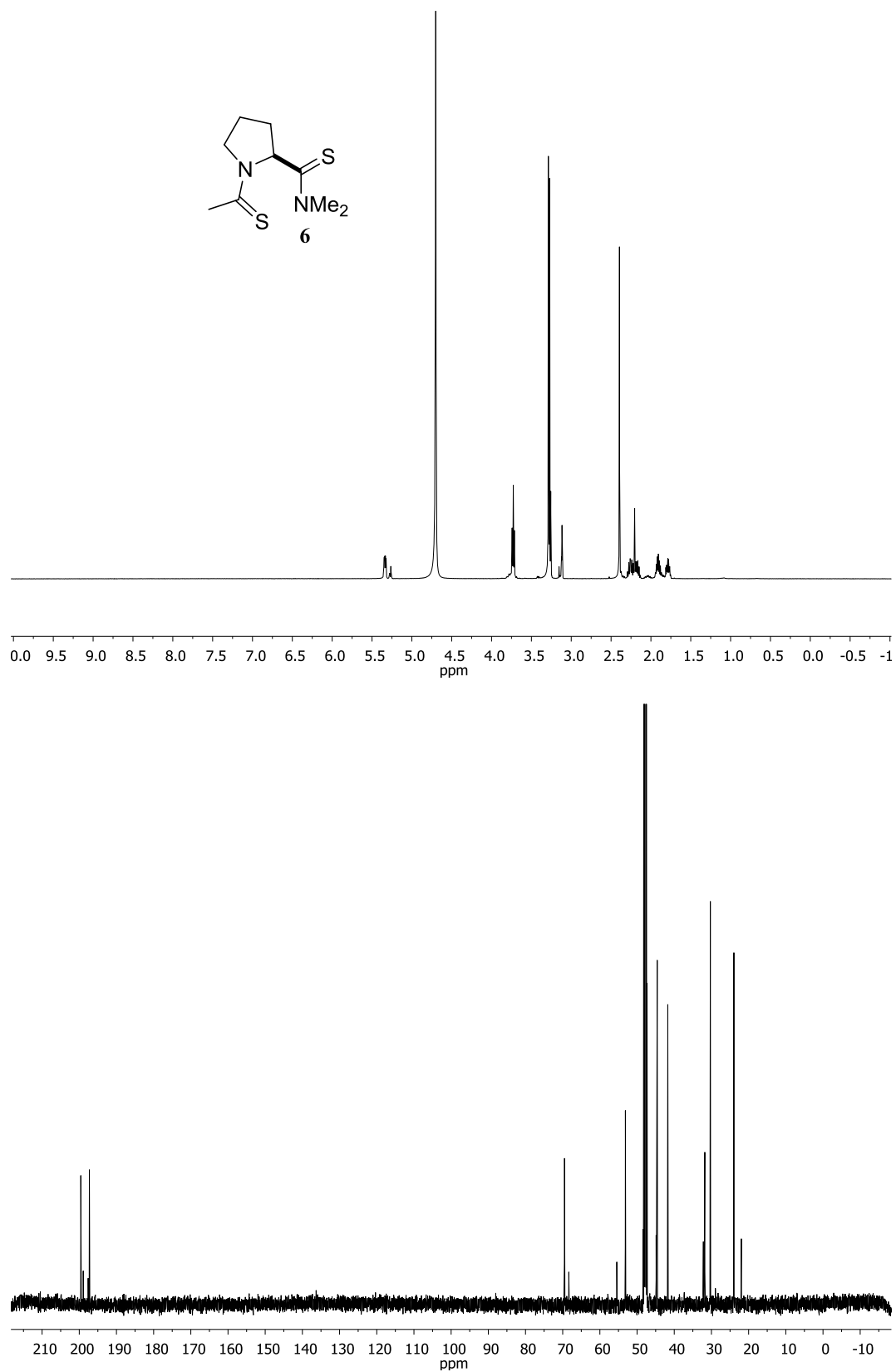


Figure 4.S8 ^1H and ^{13}C NMR spectra for compound **6**.

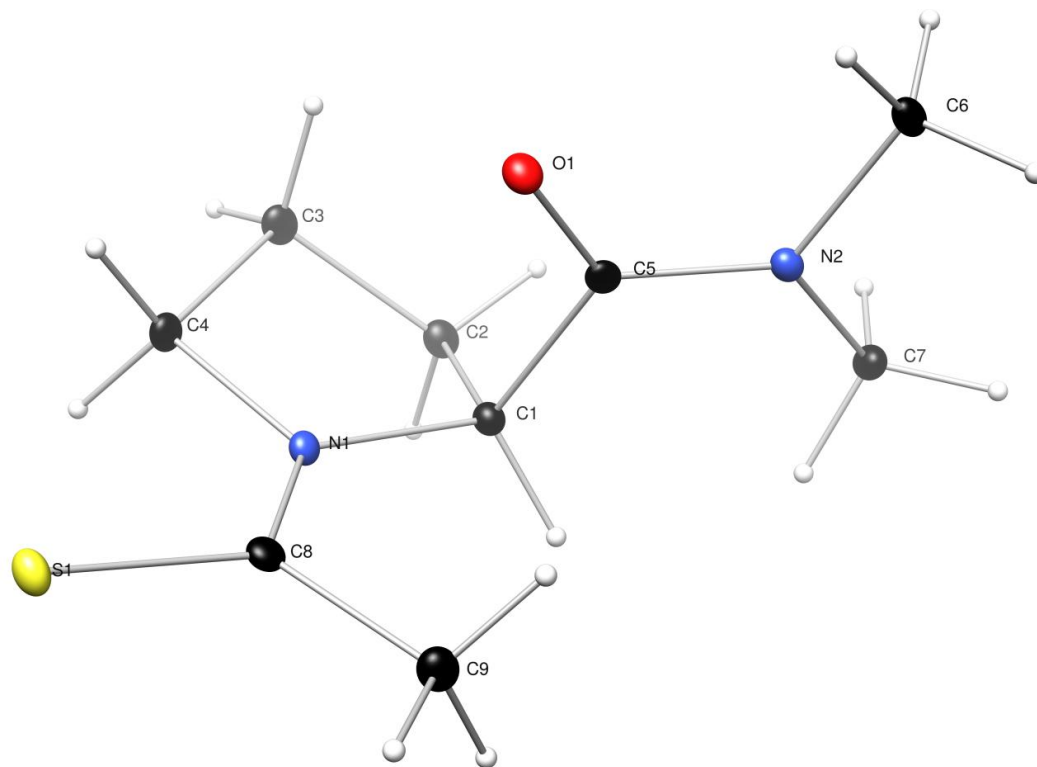


Figure 4.S9 Molecular drawing of compound **4** showing 50% probability ellipsoids.

Table 4.S2 Crystal data and structure refinement for compound **4**.

Identification code	Raines64
Empirical formula	C ₉ H ₁₆ N ₂ OS
formula weight	200.30
Temperature/K	100(1)
$\lambda/\text{\AA}$	1.54178
Crystal system	orthorhombic
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁
<i>a</i> /Å	5.2933(5)
<i>b</i> /Å	13.513(2)
<i>c</i> /Å	14.3901(18)
$\alpha/^\circ$	90
$\beta/^\circ$	90
$\gamma/^\circ$	90
Volume/Å ³	1029.3(2)
<i>Z</i>	4
ρ_{calc} mg/mm ³	1.293
μ/mm^{-1}	2.503
<i>F</i> (000)	432.0
Crystal size/mm ³	0.175 × 0.154 × 0.131
2 Θ range for data collection	8.978 to 139.44°
Index ranges	−6 ≤ <i>h</i> ≤ 6, −16 ≤ <i>k</i> ≤ 14, −17 ≤ <i>l</i> ≤ 17
Reflections collected	12276
Independent reflections	1917 [<i>R</i> (int) = 0.0202]
data/restraints/parameters	1917/0/122
Goodness-of-fit on <i>F</i> ²	1.061
Final <i>R</i> indexes [<i>I</i> ≥ 2 σ (<i>I</i>)]	<i>R</i> ₁ = 0.0236, <i>wR</i> ₂ = 0.0634
Final <i>R</i> indexes [all data]	<i>R</i> ₁ = 0.0240, <i>wR</i> ₂ = 0.0636
Largest diff. peak/hole / e Å ^{−3}	0.24/−0.14
Flack <i>x</i> parameter	0.028(19)
Hooft <i>y</i> parameter	0.032(6)

Table 4.S3 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for compound **4**. U_{eq} is defined as $1/3$ of the trace of the orthogonalized U_{ij} tensor.

atom	x	y	z	U_{eq}
S1	2200.2 (9)	8356.7 (3)	8490.6 (3)	21.03 (14)
O1	3494 (3)	10846.6 (9)	10274.4 (9)	19.7 (3)
N1	1030 (3)	9135.3 (10)	10114.6 (10)	14.5 (3)
N2	408 (3)	11655.3 (11)	11056.8 (10)	15.9 (3)
C1	-49 (4)	9864.6 (13)	10758.1 (11)	14.9 (4)
C2	396 (4)	9380.0 (14)	11712.2 (12)	18.3 (4)
C3	2936 (4)	8863.6 (13)	11581.5 (12)	19.1 (4)
C4	2821 (4)	8465.8 (13)	10585.1 (12)	18.1 (4)
C5	1445 (3)	10841.9 (13)	10673.6 (11)	14.5 (4)
C6	1820 (4)	12584.0 (12)	11053.5 (12)	19.1 (4)
C7	-1897 (3)	11668.8 (14)	11614.2 (13)	21.5 (4)
C8	713 (3)	9149.9 (13)	9201.1 (12)	16.2 (4)
C9	-1071 (4)	9915.4 (14)	8814.6 (13)	22.3 (4)

Table 4.S4 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for compound **4**. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11} + \dots + 2hka \times b \times U_{12}]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
S1	25.5 (2)	19.8 (2)	17.8 (2)	-5.13 (17)	4.43 (17)	-0.32 (19)
O1	19.3 (7)	18.4 (6)	21.3 (6)	-0.8 (5)	4.5 (5)	-3.3 (5)
N1	15.4 (7)	11.9 (7)	16.4 (7)	-0.5 (5)	0.8 (6)	0.1 (6)
N2	17.4 (7)	14.2 (7)	16.2 (7)	-0.8 (6)	-2.4 (6)	0.1 (6)
C1	14.9 (8)	15.2 (8)	14.5 (8)	-1.7 (6)	1.9 (7)	-0.6 (7)
C2	22.5 (10)	17.2 (8)	15.2 (8)	1.5 (6)	3.0 (7)	-1.4 (7)
C3	24.7 (9)	18.0 (8)	14.4 (8)	2.9 (7)	-0.7 (8)	1.9 (8)
C4	20.6 (9)	15.3 (8)	18.4 (9)	0.9 (7)	-1.4 (7)	4.0 (7)
C5	17.0 (9)	15.8 (8)	10.7 (7)	2.2 (6)	-3.6 (7)	0.5 (7)
C6	25.1 (10)	12.7 (8)	19.6 (8)	0.4 (7)	-6.1 (7)	0.1 (7)
C7	17.0 (9)	21.3 (9)	26.3 (10)	-7.4 (8)	-1.4 (8)	3.4 (8)
C8	15.4 (8)	14.7 (8)	18.4 (8)	-0.6 (6)	1.4 (7)	-5.0 (7)
C9	27.2 (10)	21.7 (9)	17.8 (8)	-0.1 (7)	-3.3 (8)	3.1 (8)

Table 4.S5 Bond lengths for compound **4**.

atom	atom	length/Å	atom	atom	length/Å
S1	C8	1.6775 (18)	N2	C7	1.460 (2)
O1	C5	1.227 (2)	C1	C2	1.539 (2)
N1	C1	1.468 (2)	C1	C5	1.544 (2)
N1	C4	1.475 (2)	C2	C3	1.526 (3)
N1	C8	1.325 (2)	C3	C4	1.532 (3)
N2	C5	1.347 (2)	C8	C9	1.507 (3)
N2	C6	1.461 (2)			

Table 4.S6 Bond angles for compound **4**.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
C1	N1	C4	111.85 (13)	C3	C2	C1	102.68 (14)
C8	N1	C1	124.50 (15)	C2	C3	C4	103.92 (15)
C8	N1	C4	123.08 (15)	N1	C4	C3	103.89 (14)
C5	N2	C6	119.43 (15)	O1	C5	N2	123.20 (16)
C5	N2	C7	125.14 (15)	O1	C5	C1	119.59 (16)
C7	N2	C6	114.75 (15)	N2	C5	C1	117.21 (15)
N1	C1	C2	102.56 (14)	N1	C8	S1	122.39 (14)
N1	C1	C5	108.98 (14)	N1	C8	C9	117.10 (16)
C2	C1	C5	110.83 (14)	C9	C8	S1	120.51 (13)

Table 4.S7 Torsion angles for compound **4**.

A	B	C	D	angle/°
N1	C1	C2	C3	35.51 (17)
N1	C1	C5	O1	-14.5 (2)
N1	C1	C5	N2	166.04 (14)
C1	N1	C4	C3	-3.34 (19)
C1	N1	C8	S1	174.44 (13)
C1	N1	C8	C9	-5.8 (3)
C1	C2	C3	C4	-38.20 (18)
C2	C1	C5	O1	97.68 (19)
C2	C1	C5	N2	-81.82 (19)
C2	C3	C4	N1	25.91 (17)
C4	N1	C1	C2	-20.26 (18)
C4	N1	C1	C5	97.25 (16)
C4	N1	C8	S1	3.7 (2)
C4	N1	C8	C9	-176.48 (16)
C5	C1	C2	C3	-80.67 (17)
C6	N2	C5	O1	-3.7 (2)
C6	N2	C5	C1	175.77 (14)
C7	N2	C5	O1	-173.76 (16)
C7	N2	C5	C1	5.7 (2)
C8	N1	C1	C2	168.13 (17)
C8	N1	C1	C5	-74.4 (2)
C8	N1	C4	C3	168.41 (16)

Table 4.S8 Hydrogen atom coordinates ($\text{\AA}\times 10^4$) and isotropic displacement parameters ($\text{\AA}^2\times 10^3$) for compound **4**.

atom	<i>x</i>	<i>y</i>	<i>z</i>	<i>U</i> _{eq}
H1	-1890	9969	10637	18
H2A	484	9883	12211	22
H2B	-954	8899	11861	22
H3A	3145	8318	12034	23
H3B	4351	9336	11656	23
H4A	4504	8491	10287	22
H4B	2199	7775	10575	22
H6A	2472	12716	11679	29
H6B	702	13125	10863	29
H6C	3234	12534	10616	29
H7A	-2990	11118	11427	32
H7B	-2786	12296	11517	32
H7C	-1461	11601	12273	32
H9A	-1112	9864	8135	33
H9B	-491	10577	8993	33
H9C	-2770	9803	9064	33

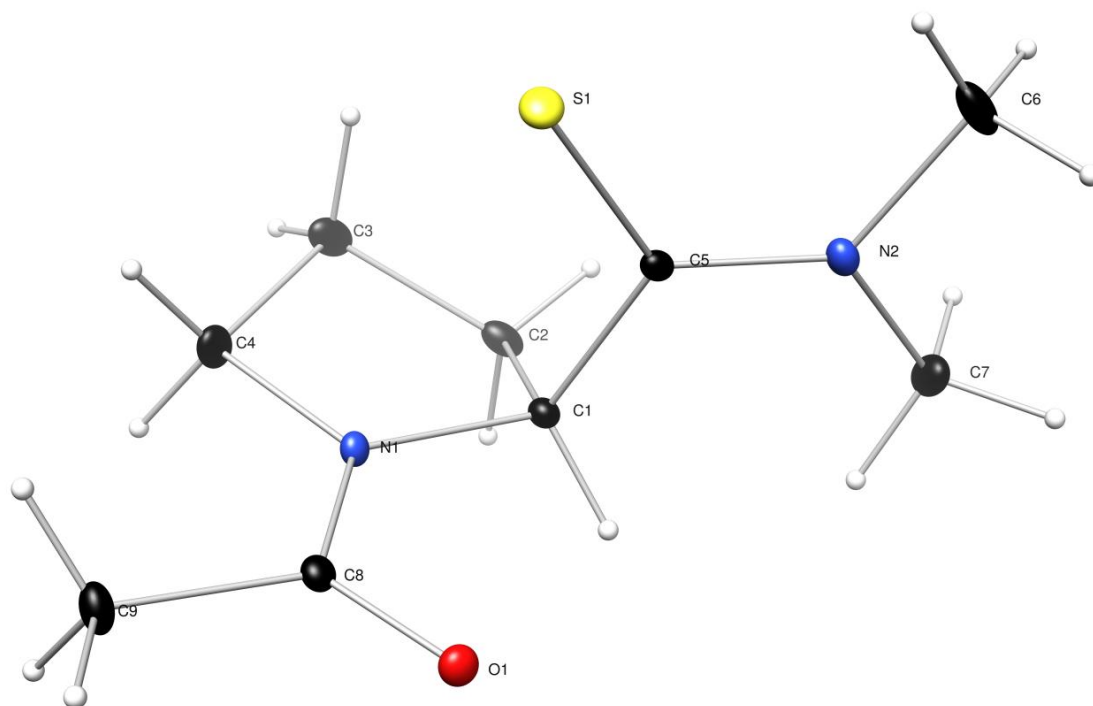


Figure 4.S10 Molecular drawing of compound **5** shown with 50% probability ellipsoids.

Table 4.S9 Crystal data and structure refinement for compound **5**.

Identification code	Raines63
Empirical formula	C ₉ H ₁₆ N ₂ OS
formula weight	200.30
Temperature/K	100.0
Crystal system	orthorhombic
Space group	P2 ₁ 2 ₁ 2 ₁
<i>a</i> /Å	7.360(3)
<i>b</i> /Å	7.513(2)
<i>c</i> /Å	18.801(6)
α /°	90
β /°	90
γ /°	90
Volume/Å ³	1039.6(6)
<i>Z</i>	4
ρ_{calc} /mg/mm ³	1.280
<i>m</i> /mm ⁻¹	0.276
<i>F</i> (000)	432.0
Crystal size/mm ³	0.4 × 0.21 × 0.12
2 Θ range for data collection	4.332 to 60.022°
Index ranges	-10 ≤ <i>h</i> ≤ 10, -10 ≤ <i>k</i> ≤ 10, -26 ≤ <i>l</i> ≤ 26
Reflections collected	28497
Independent reflections	3029 [<i>R</i> (int) = 0.0227]
data/restraints/parameters	3029/0/167
Goodness-of-fit on <i>F</i> ²	0.974
Final <i>R</i> indexes [<i>I</i> ≥ 2σ(<i>I</i>)]	<i>R</i> ₁ = 0.0212, <i>wR</i> ₂ = 0.0574
Final <i>R</i> indexes [all data]	<i>R</i> ₁ = 0.0215, <i>wR</i> ₂ = 0.0577
Largest diff. peak/hole / e Å ⁻³	0.27/-0.16
Flack parameter	-0.02(5)

Table 4.S10 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for compound **5**. U_{eq} is defined as $1/3$ of the trace of the orthogonalized U_{ij} tensor.

atom	x	y	z	U_{eq}
S1	4366.7 (4)	5993.0 (4)	6018.55 (16)	22.70 (8)
O1	512.3 (13)	6498.9 (11)	7277.6 (4)	18.91 (16)
N1	380.9 (12)	5487.5 (12)	6151.9 (5)	14.15 (17)
N2	4388.2 (15)	2864.0 (14)	6675.4 (5)	18.52 (19)
C1	1368.3 (14)	3867.6 (14)	6323.8 (5)	13.96 (18)
C2	894.8 (16)	2662.1 (16)	5686.8 (7)	19.9 (2)
C3	744.1 (17)	3964.9 (18)	5067.3 (6)	23.5 (2)
C4	-124.2 (18)	5620.7 (17)	5397.7 (6)	21.7 (2)
C5	3430.2 (14)	4167.1 (15)	6371.8 (5)	13.89 (18)
C6	6368.5 (19)	2898 (2)	6714.2 (8)	28.7 (3)
C7	3623 (2)	1214.4 (18)	6964.9 (7)	28.3 (3)
C8	25.4 (15)	6714.0 (14)	6654.9 (6)	14.48 (19)
C9	-1019.1 (18)	8336.6 (16)	6423.8 (7)	21.4 (2)

Table 4.S11 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for compound **5**. The anisotropic displacement factor exponent takes the form: $-\pi^2 [h^2 a^{*2} U_{11} + \dots + 2hka \times b \times U_{12}]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
S1	16.68 (13)	22.08 (13)	29.33 (14)	5.29 (11)	0.37 (10)	5.94 (11)
O1	21.4 (4)	20.6 (3)	14.8 (3)	-1.6 (3)	-1.5 (3)	2.2 (3)
N1	14.2 (4)	14.7 (4)	13.6 (4)	0.8 (3)	-2.0 (3)	2.4 (3)
N2	17.3 (4)	21.5 (4)	16.7 (4)	-0.1 (3)	-0.6 (4)	5.8 (4)
C1	12.5 (4)	13.3 (4)	16.1 (4)	0.8 (4)	1.7 (3)	1.2 (4)
C2	14.1 (5)	18.6 (5)	27.1 (5)	-8.0 (4)	0.8 (4)	-0.7 (4)
C3	19.8 (5)	33.4 (6)	17.3 (4)	-8.1 (4)	-2.6 (4)	0.9 (5)
C4	24.3 (5)	26.2 (6)	14.5 (4)	-0.4 (4)	-5.5 (4)	4.6 (5)
C5	12.6 (4)	17.3 (4)	11.8 (4)	-1.1 (4)	0.3 (3)	0.9 (4)
C6	17.5 (6)	37.5 (7)	31.1 (6)	-10.1 (6)	-7.5 (5)	10.9 (5)
C7	33.4 (7)	23.8 (6)	27.9 (6)	9.9 (5)	6.5 (5)	11.9 (5)
C8	11.8 (4)	13.5 (4)	18.1 (4)	0.5 (4)	-0.6 (4)	-1.0 (4)
C9	22.3 (6)	14.1 (4)	27.9 (6)	-0.7 (4)	-6.0 (4)	4.3 (4)

Table 4.S12 Bond lengths for compound **5**.

atom	atom	length/Å	atom	atom	length/Å
S1	C5	1.6727 (12)	N2	C7	1.4661 (18)
O1	C8	1.2350 (14)	C1	C2	1.5414 (16)
N1	C1	1.4539 (14)	C1	C5	1.5368 (16)
N1	C4	1.4693 (14)	C2	C3	1.5255 (18)
N1	C8	1.3462 (14)	C3	C4	1.5304 (18)
N2	C5	1.3347 (14)	C8	C9	1.5053 (16)
N2	C6	1.4595 (19)			

Table 4.S13 Bond angles for compound **5**.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
C1	N1	C4	113.46 (9)	C3	C2	C1	103.45 (10)
C8	N1	C1	120.93 (9)	C2	C3	C4	104.00 (9)
C8	N1	C4	125.61 (9)	N1	C4	C3	103.34 (9)
C5	N2	C6	122.42 (12)	N2	C5	S1	123.62 (9)
C5	N2	C7	125.17 (11)	N2	C5	C1	116.06 (10)
C6	N2	C7	112.31 (12)	C1	C5	S1	120.24 (8)
N1	C1	C2	101.90 (9)	O1	C8	N1	121.34 (10)
N1	C1	C5	112.58 (9)	O1	C8	C9	121.86 (10)
C5	C1	C2	110.79 (8)	N1	C8	C9	116.79 (10)

Table 4.S14 Torsion angles for compound **5**.

A	B	C	D	angle/°	A	B	C	D	angle/°
N1	C1	C2	C3	33.76 (10)	C4	N1	C1	C5	99.77 (11)
N1	C1	C5	S1	-17.13 (12)	C4	N1	C8	O1	-179.49 (11)
N1	C1	C5	N2	165.89 (8)	C4	N1	C8	C9	1.56 (16)
C1	N1	C4	C3	-3.65 (13)	C5	C1	C2	C3	-86.22 (10)
C1	N1	C8	O1	-0.54 (16)	C6	N2	C5	S1	-1.30 (15)
C1	N1	C8	C9	-179.49 (10)	C6	N2	C5	C1	175.56 (10)
C1	C2	C3	C4	-36.87 (11)	C7	N2	C5	S1	-177.45 (9)
C2	C1	C5	S1	96.24 (10)	C7	N2	C5	C1	-0.59 (16)
C2	C1	C5	N2	-80.73 (12)	C8	N1	C1	C2	161.98 (10)
C2	C3	C4	N1	25.12 (12)	C8	N1	C1	C5	-79.30 (12)
C4	N1	C1	C2	-18.95 (12)	C8	N1	C4	C3	175.37 (10)

Table 4.S15 Hydrogen atom coordinates ($\text{\AA}\times 10^4$) and isotropic displacement parameters ($\text{\AA}^2\times 10^3$) for compound **5**.

atom	<i>x</i>	<i>y</i>	<i>z</i>	<i>U</i> _{eq}
H1	900 (20)	3440 (20)	6769 (8)	17
H2A	1770 (20)	1770 (20)	5598 (9)	24
H2B	-250 (20)	2090 (20)	5787 (9)	24
H3A	30 (30)	3470 (30)	4687 (10)	28
H3B	1960 (30)	4300 (20)	4889 (9)	28
H4A	310 (30)	6700 (20)	5185 (9)	26
H4B	-1400 (30)	5620 (30)	5339 (10)	26
H6A	6810 (30)	1870 (30)	6507 (12)	43
H6B	6660 (30)	2730 (30)	7197 (12)	43
H6C	6840 (30)	3860 (30)	6525 (12)	43
H7A	2300 (30)	1260 (30)	7024 (11)	43
H7B	4210 (30)	990 (30)	7433 (11)	43
H7C	3990 (30)	210 (30)	6632 (12)	43
H9A	-2230 (30)	8030 (30)	6286 (10)	32
H9B	-550 (30)	8890 (30)	6025 (10)	32
H9C	-1090 (30)	9180 (30)	6797 (10)	32

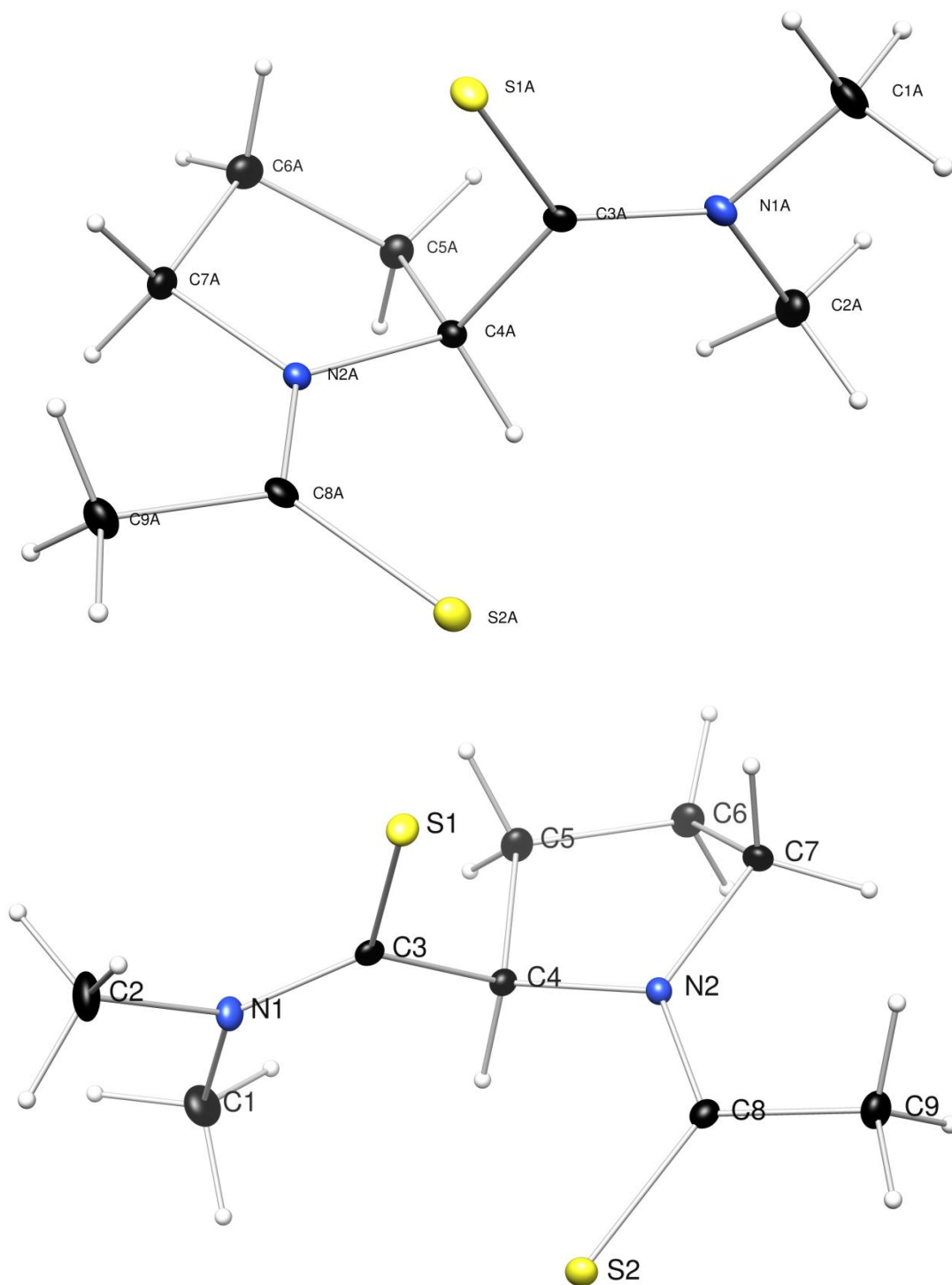


Figure 4.S11 Molecular drawing of compound **6** showing 50% probability ellipsoids.

Table 4.S16 Crystal data and structure refinement for compound **6**.

Identification code	Raines65
Empirical formula	C ₉ H ₁₆ N ₂ S ₂
Formula weight	216.36
Temperature/K	99.97
Crystal system	orthorhombic
Space group	P2 ₁ 2 ₁ 2 ₁
<i>a</i> /Å	11.5526(6)
<i>b</i> /Å	13.3416(8)
<i>c</i> /Å	14.1333(8)
α /°	90
β /°	90
γ /°	90
Volume/Å ³	2178.4(2)
<i>Z</i>	8
ρ_{calc} /mg/mm ³	1.319
<i>m</i> /mm ⁻¹	4.078
<i>F</i> (000)	928.0
Crystal size/mm ³	0.777 × 0.541 × 0.464
2 θ range for data collection	9.114 to 144.92°
Index ranges	-14 ≤ <i>h</i> ≤ 14, -16 ≤ <i>k</i> ≤ 16, -17 ≤ <i>l</i> ≤ 17
Reflections collected	35128
Independent reflections	4228 [<i>R</i> (int) = 0.0200]
Data/restraints/parameters	4228/0/242
Goodness-of-fit on <i>F</i> ²	1.100
Final <i>R</i> indexes [<i>I</i> ≥ 2σ(<i>I</i>)]	<i>R</i> ₁ = 0.0254, <i>wR</i> ₂ = 0.0660
Final <i>R</i> indexes [all data]	<i>R</i> ₁ = 0.0254, <i>wR</i> ₂ = 0.0660
Largest diff. peak/hole / e Å ⁻³	0.40/-0.33
Flack parameter	0.010(15)

Table 4.S17 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for compound **6**. U_{eq} is defined as $1/3$ of the trace of the orthogonalized U_{ij} tensor.

atom	<i>x</i>	<i>y</i>	<i>z</i>	<i>U</i>
S1	3773.3 (4)	7888.1 (4)	4788.2 (4)	17.51 (13)
S2	3488.4 (4)	9620.4 (4)	2535.4 (4)	17.93 (13)
N1	4761.7 (17)	9671.5 (14)	4988.4 (13)	18.0 (4)
N2	4921.1 (16)	8128.6 (13)	2877.6 (13)	13.9 (4)
C1	5583 (2)	10478.7 (18)	4764.9 (19)	26.9 (5)
C2	4018 (2)	9881 (2)	5806.8 (17)	29.6 (6)
C3	4652.8 (18)	8838.2 (16)	4488.6 (15)	13.9 (4)
C4	5458.1 (18)	8715.4 (15)	3637.4 (14)	13.8 (4)
C5	6540 (2)	8099.0 (17)	3917.7 (16)	20.4 (5)
C6	6737 (2)	7394.1 (18)	3085.5 (17)	23.2 (5)
C7	5521 (2)	7161.6 (17)	2740.2 (16)	18.5 (4)
C8	4047.4 (18)	8475.6 (16)	2362.1 (14)	14.1 (4)
C9	3583 (2)	7797.7 (18)	1600.1 (15)	20.1 (5)
S1A	3464.9 (5)	4741.2 (4)	6862.7 (4)	20.21 (13)
S2A	3226.2 (5)	2921.1 (4)	4647.4 (4)	19.34 (13)
N1A	4465.1 (16)	3037.6 (15)	7373.0 (12)	17.7 (4)
N2A	4750.3 (16)	4319.8 (13)	5100.5 (13)	13.8 (4)
C1A	3745 (2)	3001 (2)	8227.9 (16)	27.3 (5)
C2A	5255 (2)	2174.2 (18)	7293.4 (17)	23.1 (5)
C3A	4366.2 (18)	3762.1 (16)	6738.7 (15)	15.1 (4)
C4A	5174.8 (18)	3708.3 (16)	5879.9 (14)	14.0 (4)
C5A	6364.3 (19)	4182.1 (16)	6095.8 (17)	19.7 (5)
C6A	6186 (2)	5286.9 (19)	5859.7 (17)	24.3 (5)
C7A	5392 (2)	5277.5 (17)	5010.9 (16)	18.9 (5)
C8A	3886.4 (18)	4036.4 (16)	4545.5 (14)	15.7 (4)
C9A	3513 (2)	4783.6 (18)	3810.8 (16)	22.3 (5)

Table 4.S18 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for compound **6**. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11}+...+2hka \times b \times U_{12}]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
S1	18.3 (3)	17.6 (2)	16.6 (2)	1.90 (19)	2.95 (19)	-4.5 (2)
S2	17.8 (3)	18.7 (2)	17.3 (2)	1.03 (19)	0.5 (2)	6.1 (2)
N1	20.6 (9)	17.7 (9)	15.6 (9)	-2.5 (7)	2.2 (7)	-4.0 (8)
N2	15.5 (9)	13.5 (8)	12.7 (8)	-0.5 (7)	1.2 (7)	1.0 (7)
C1	28.1 (12)	21.3 (11)	31.3 (13)	-3.5 (10)	4.0 (11)	-8.7 (10)
C2	36.6 (14)	30.6 (13)	21.6 (12)	-11 (1)	11.2 (11)	-6.0 (11)
C3	12.5 (9)	16.5 (10)	12.7 (9)	2.0 (8)	-2.5 (8)	1.4 (8)
C4	13.4 (9)	15.2 (10)	12.8 (9)	0.8 (8)	0.1 (8)	-0.5 (8)
C5	15.7 (10)	25.3 (11)	20.1 (10)	-1.6 (9)	-2.0 (9)	4.6 (9)
C6	20.9 (11)	28.0 (11)	20.8 (11)	-1.0 (9)	2.0 (9)	9.6 (10)
C7	22.4 (11)	15.8 (10)	17.3 (10)	0.8 (8)	0.1 (8)	6.4 (9)
C8	12.7 (9)	17.4 (10)	12.4 (9)	2.3 (8)	2.0 (8)	-1.3 (8)
C9	18.7 (10)	24.7 (11)	16.8 (10)	-3.3 (9)	-1.6 (8)	-0.3 (9)
S1A	21.8 (3)	21.8 (3)	17.1 (3)	-0.5 (2)	4.1 (2)	7.6 (2)
S2A	17.8 (2)	22.8 (3)	17.4 (2)	-1.2 (2)	-0.3 (2)	-5.5 (2)
N1A	21.0 (9)	19.9 (9)	12.2 (8)	0.3 (7)	-0.1 (7)	1.5 (8)
N2A	15.3 (8)	14.4 (8)	11.8 (8)	-0.1 (7)	0.4 (7)	0.1 (7)
C1A	33.2 (13)	34.1 (13)	14.6 (10)	4.3 (10)	4.8 (10)	1.0 (11)
C2A	26.6 (12)	19.3 (11)	23.2 (11)	4.0 (9)	-2.8 (9)	3.7 (10)
C3A	15.8 (10)	16.9 (10)	12.6 (9)	-2.5 (8)	-2.8 (8)	0.3 (8)
C4A	14.2 (10)	14.4 (10)	13.4 (10)	-0.4 (8)	-0.9 (8)	2.6 (8)
C5A	14.8 (10)	20.3 (11)	23.8 (11)	0.5 (9)	-2.5 (9)	1.3 (9)
C6A	23.2 (12)	23.7 (12)	26.0 (12)	-1.9 (10)	0.8 (10)	-3.6 (10)
C7A	21.1 (11)	15.6 (10)	19.9 (10)	1.0 (8)	2.1 (9)	-1.6 (9)
C8A	15.3 (10)	20.6 (10)	11.2 (9)	-1.4 (8)	2.3 (8)	1.8 (8)
C9A	22.3 (12)	27.6 (12)	16.9 (10)	4.5 (9)	-3.4 (9)	2.2 (10)

Table 4.S19 Bond lengths for compound **6**.

atom	atom	length/Å	atom	atom	length/Å
S1	C3	1.679 (2)	S1A	C3A	1.680 (2)
S2	C8	1.676 (2)	S2A	C8A	1.678 (2)
N1	C1	1.469 (3)	N1A	C1A	1.468 (3)
N1	C2	1.467 (3)	N1A	C2A	1.474 (3)
N1	C3	1.323 (3)	N1A	C3A	1.323 (3)
N2	C4	1.467 (3)	N2A	C4A	1.456 (3)
N2	C7	1.477 (3)	N2A	C7A	1.483 (3)
N2	C8	1.328 (3)	N2A	C8A	1.325 (3)
C3	C4	1.530 (3)	C3A	C4A	1.533 (3)
C4	C5	1.547 (3)	C4A	C5A	1.543 (3)
C5	C6	1.523 (3)	C5A	C6A	1.525 (3)
C6	C7	1.520 (3)	C6A	C7A	1.510 (3)
C8	C9	1.505 (3)	C8A	C9A	1.503 (3)

Table 4.S20 Bond angles for compound **6**.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
C2	N1	C1	114.06 (19)	C1A	N1A	C2A	112.82 (19)
C3	N1	C1	124.2 (2)	C3A	N1A	C1A	122.21 (19)
C3	N1	C2	121.67 (19)	C3A	N1A	C2A	124.93 (19)
C4	N2	C7	111.36 (17)	C4A	N2A	C7A	112.28 (17)
C8	N2	C4	122.48 (18)	C8A	N2A	C4A	122.81 (18)
C8	N2	C7	126.06 (18)	C8A	N2A	C7A	124.91 (18)
N1	C3	S1	123.87 (17)	N1A	C3A	S1A	123.43 (17)
N1	C3	C4	116.88 (18)	N1A	C3A	C4A	116.71 (19)
C4	C3	S1	119.05 (16)	C4A	C3A	S1A	119.78 (16)
N2	C4	C3	112.07 (17)	N2A	C4A	C3A	111.56 (17)
N2	C4	C5	104.20 (17)	N2A	C4A	C5A	102.71 (17)
C3	C4	C5	110.29 (17)	C3A	C4A	C5A	111.52 (17)
C6	C5	C4	104.56 (18)	C6A	C5A	C4A	103.42 (18)
C7	C6	C5	103.62 (18)	C7A	C6A	C5A	104.35 (19)
N2	C7	C6	102.31 (17)	N2A	C7A	C6A	104.06 (18)
N2	C8	S2	122.02 (16)	N2A	C8A	S2A	123.01 (17)
N2	C8	C9	116.99 (19)	N2A	C8A	C9A	115.9 (2)
C9	C8	S2	120.98 (16)	C9A	C8A	S2A	121.14 (17)

Table 4.S21 Torsion angles for compound **6**.

A	B	C	D	angle/°	A	B	C	D	angle/°
S1	C3	C4	N2	35.8 (2)	S1A	C3A	C4A	N2A	-22.0 (2)
S1	C3	C4	C5	-79.8 (2)	S1A	C3A	C4A	C5A	92.2 (2)
N1	C3	C4	N2	-149.16 (19)	N1A	C3A	C4A	N2A	161.07 (18)
N1	C3	C4	C5	95.2 (2)	N1A	C3A	C4A	C5A	-84.7 (2)
N2	C4	C5	C6	17.8 (2)	N2A	C4A	C5A	C6A	32.1 (2)
C1	N1	C3	S1	175.93 (18)	C1A	N1A	C3A	S1A	3.1 (3)
C1	N1	C3	C4	1.2 (3)	C1A	N1A	C3A	C4A	179.9 (2)
C2	N1	C3	S1	-6.6 (3)	C2A	N1A	C3A	S1A	-179.26 (17)
C2	N1	C3	C4	178.6 (2)	C2A	N1A	C3A	C4A	-2.5 (3)
C3	C4	C5	C6	138.26 (19)	C3A	C4A	C5A	C6A	-87.5 (2)
C4	N2	C7	C6	-26.7 (2)	C4A	N2A	C7A	C6A	-5.7 (2)
C4	N2	C8	S2	0.3 (3)	C4A	N2A	C8A	S2A	-2.9 (3)
C4	N2	C8	C9	179.58 (18)	C4A	N2A	C8A	C9A	176.60 (19)
C4	C5	C6	C7	-34.0 (2)	C4A	C5A	C6A	C7A	-36.4 (2)
C5	C6	C7	N2	36.7 (2)	C5A	C6A	C7A	N2A	26.1 (2)
C7	N2	C4	C3	-113.59 (19)	C7A	N2A	C4A	C3A	102.9 (2)
C7	N2	C4	C5	5.7 (2)	C7A	N2A	C4A	C5A	-16.6 (2)
C7	N2	C8	S2	-175.81 (16)	C7A	N2A	C8A	S2A	177.24 (16)
C7	N2	C8	C9	3.5 (3)	C7A	N2A	C8A	C9A	-3.3 (3)
C8	N2	C4	C3	69.8 (2)	C8A	N2A	C4A	C3A	-77.0 (2)
C8	N2	C4	C5	-170.96 (19)	C8A	N2A	C4A	C5A	163.45 (19)
C8	N2	C7	C6	149.8 (2)	C8A	N2A	C7A	C6A	174.2 (2)

Table 4.S22 Hydrogen atom coordinates ($\text{\AA}\times 10^4$) and isotropic displacement parameters ($\text{\AA}^2\times 10^3$) for compound **6**.

atom	<i>x</i>	<i>y</i>	<i>z</i>	<i>U</i> _{eq}
H1A	5223	10949	4321	40
H1B	6280	10192	4477	40
H1C	5793	10833	5347	40
H2A	4450	9752	6391	44
H2B	3335	9446	5784	44
H2C	3774	10584	5791	44
H4	5696	9388	3393	17
H5A	7217	8542	4013	24
H5B	6398	7717	4507	24
H6A	7199	7724	2584	28
H6B	7140	6776	3290	28
H7A	5522	6961	2066	22
H7B	5159	6624	3122	22
H9A	2864	8080	1346	30
H9B	3429	7133	1867	30
H9C	4155	7739	1091	30
H1AA	3224	2422	8194	41
H1AB	3287	3617	8275	41
H1AC	4243	2939	8786	41
H2AA	5612	2042	7910	35
H2AB	5860	2324	6828	35
H2AC	4819	1583	7090	35
H4A	5272	2997	5671	17
H5AA	6575	4093	6770	24
H5AB	6977	3884	5695	24
H6AA	5823	5645	6397	29
H6AB	6931	5614	5705	29
H7AA	4856	5856	5025	23
H7AB	5841	5297	4415	23
H9AA	4159	4918	3379	33
H9AB	3278	5408	4122	33
H9AC	2859	4512	3452	33

Table 4.S23 SCF Energies (atomic units) of compounds **3–6** Calculated at the B3LYP/6-311+G(2d,p) level of theory.

Compound	Amide Conformation	Ring Pucker	Energy	ZPE Correction	Energy (Corrected)
3	<i>trans</i>	<i>endo</i>	–612.7666889	0.250153	–612.5165359
		<i>exo</i>	–612.7661143	0.250001	–612.5161133
	<i>cis</i>	<i>endo</i>	–612.7627137	0.249532	–612.5131817
		<i>exo</i>	–612.7616812	0.249632	–612.5120492
4	<i>trans</i>	<i>endo</i>	–935.7221726	0.247932	–935.4742406
		<i>exo</i>	–935.7211053	0.247981	–935.4731243
	<i>cis</i>	<i>endo</i>	–935.7182725	0.247612	–935.4706605
		<i>exo</i>	–935.7164957	0.247507	–935.4689887
5	<i>trans</i>	<i>endo</i>	–935.7189804	0.248119	–935.4708614
		<i>exo</i>	–935.7194686	0.248284	–935.4711846
	<i>cis</i>	<i>endo</i>	–935.7144801	0.247581	–935.4668991
		<i>exo</i>	–935.7132949	0.247558	–935.4657369
6	<i>trans</i>	<i>endo</i>	–1258.674434	0.246079	–1258.428355
		<i>exo</i>	–1258.674022	0.246152	–1258.427870
	<i>cis</i>	<i>endo</i>	–1258.669679	0.245564	–1258.424115
		<i>exo</i>	–1258.667992	0.245704	–1258.422288

Table 4.S24 Orbital Energies (atomic units) of $n \rightarrow \pi^*$ Donors and Acceptors as Calculated with NBO 5.0

Compound	Ring Pucker	Orbital Energies		
		Donor	Acceptor	Difference
3	<i>endo</i>	–0.24811	0.04368	0.29179
	<i>exo</i>	–0.24652	0.10809	0.35461
4	<i>endo</i>	–0.19233	0.04148	0.23381
	<i>exo</i>	–0.19332	0.14576	0.33908
5	<i>endo</i>	–0.25003	–0.04523	0.20480
	<i>exo</i>	–0.24956	–0.04739	0.20217
6	<i>endo</i>	–0.19549	–0.04536	0.15013
	<i>exo</i>	–0.19623	–0.04379	0.15244

Table 4.S25 Cartesian coordinates of compound **3** in the optimized *trans endo* conformation.

C	0.05747600	0.45226800	0.71045200
H	-0.19320800	-0.12304400	1.59434800
C	2.27587100	0.95473300	-0.20585300
H	2.61097000	0.86578200	-1.23827200
H	3.15090500	0.86464900	0.44398000
C	1.52684100	2.27005500	0.04193900
H	1.04954900	2.59970700	-0.87670500
H	2.19338600	3.05583300	0.39385900
C	0.45374000	1.89776000	1.07546400
H	0.87180100	1.90276500	2.08328700
H	-0.39625800	2.57618800	1.06031300
N	1.28952500	-0.08506800	0.12270800
C	1.45162200	-1.42461600	-0.02460500
O	0.57675000	-2.21788300	0.31585200
C	2.75979300	-1.89298200	-0.62860500
H	2.86272700	-1.53449300	-1.65378500
H	3.61696100	-1.52433400	-0.06373200
H	2.76568700	-2.97801400	-0.62739500
C	-1.09223700	0.44868900	-0.31683800
N	-2.25644800	-0.17168700	0.01497900
C	-3.39446300	-0.04943900	-0.88787100
H	-3.55993900	-0.98165400	-1.43358300
H	-4.29386700	0.18229700	-0.31549700
H	-3.19834900	0.74359700	-1.60052300
C	-2.42229000	-1.08645700	1.14425600
H	-3.20822200	-1.79602300	0.89194900
H	-1.51355800	-1.65543900	1.31508100
H	-2.72314300	-0.56285900	2.05530300
O	-0.94739000	1.06066600	-1.37116100

Table 4.S26 Cartesian coordinates of compound **3** in the optimized *cis endo* conformation.

C	-0.04187700	-0.35840600	0.60677500
H	0.12809900	0.13149500	1.56291700
C	-2.06190800	-1.30557300	-0.38355500
H	-2.44283500	-1.10444200	-1.38241600
H	-2.91324500	-1.57373500	0.24514200
C	-0.96170100	-2.37070900	-0.36117000
H	-0.43707500	-2.38505600	-1.31553200
H	-1.35456600	-3.36869300	-0.17683500
C	-0.01451600	-1.89840700	0.75139800
H	-0.40874400	-2.16983800	1.73176100
H	0.98774100	-2.31332200	0.66458700
N	-1.40350400	-0.10836500	0.15789300
C	-2.05384000	1.09014100	0.14988200
O	-3.20666300	1.17846200	-0.25045200
C	-1.29395000	2.29739900	0.66641200
H	-0.98774100	2.16463800	1.70559900
H	-0.40084900	2.48120200	0.06900100
H	-1.95101900	3.15845900	0.60039600
C	1.00749800	0.12263200	-0.42531000
N	2.29958700	0.27228900	0.00270600
C	3.31561500	0.69701700	-0.95169100
H	3.83241800	1.58300600	-0.57709000
H	4.05431900	-0.09397100	-1.10347700
H	2.83692900	0.92673100	-1.89649700
C	2.80190900	-0.01764300	1.33575100
H	3.26462700	0.87284500	1.76842700
H	2.01669800	-0.35106200	2.00233000
H	3.56179600	-0.80267600	1.29215800
O	0.67634900	0.34131800	-1.58025100

Table 4.S27 Cartesian coordinates of compound **3** in the optimized *trans* *exo* conformation.

C	0.09898500	-0.65183100	-0.49786600
C	0.56892700	-2.12218200	-0.40412200
C	2.10239500	-2.04072200	-0.40611600
C	2.38043300	-0.74400300	0.35588900
H	-0.13746600	-0.37528400	-1.52130200
H	0.22190000	-2.54550000	0.53670400
H	0.17316000	-2.72394600	-1.21984600
H	2.57016200	-2.90627700	0.05987200
H	2.48326300	-1.95782800	-1.42525200
H	2.35054500	-0.90999300	1.43629500
H	3.34033300	-0.30195300	0.10001500
C	1.24846500	1.47446100	-0.09790400
C	-1.08889000	-0.40287200	0.45207500
O	0.26408100	2.08054600	-0.51466500
O	-0.92914300	-0.59174700	1.65246600
C	2.48132100	2.20528700	0.39545700
H	3.33779300	2.01417100	-0.25321100
H	2.26624600	3.26887600	0.39042000
H	2.75189700	1.89367800	1.40455500
C	-3.40987000	0.17907500	0.81061000
H	-3.19948100	-0.29237800	1.76391900
H	-3.56627400	1.24836300	0.97534000
H	-4.31710300	-0.24637600	0.38141000
N	1.27436500	0.11850500	-0.06583200
N	-2.28592000	-0.05171000	-0.08991500
C	-2.44228800	0.41155200	-1.47025100
H	-2.27570700	-0.39119400	-2.18929700
H	-3.46586600	0.75404100	-1.59691600
H	-1.76888000	1.24105500	-1.68152000

Table 4.S28 Cartesian coordinates of compound **3** in the optimized *cis* *exo* conformation.

C	0.05423800	0.50593100	0.40725800
C	0.09880300	2.03107200	0.10701400
C	1.59398600	2.36959400	0.03756500
C	2.21132200	1.09939700	-0.54343300
H	-0.11498800	0.32835100	1.46903100
H	-0.36908400	2.22276400	-0.85896900
H	-0.43931000	2.60904300	0.85594200
H	1.79194500	3.25097700	-0.56983200
H	1.99165200	2.55191600	1.03703900
H	2.13660700	1.07666800	-1.63408400
H	3.25200000	0.94433000	-0.27516900
C	1.89577300	-1.21558700	0.23002800
C	-1.05243500	-0.16583200	-0.43401400
O	3.03557200	-1.49742600	-0.11440800
O	-0.77750200	-0.67057000	-1.51112700
C	0.99276200	-2.24309200	0.88403000
H	0.22493500	-2.56455000	0.18028800
H	1.60386800	-3.09806300	1.15518500
H	0.49633900	-1.85803600	1.77573800
C	-3.40259600	-0.71327700	-0.75297200
H	-2.97070600	-1.20291500	-1.61788600
H	-3.96738400	-1.43802200	-0.16344900
H	-4.08861900	0.06945400	-1.08654500
N	1.39056700	0.04061700	0.04837800
N	-2.33093300	-0.14403400	0.05473200
C	-2.77664400	0.51660500	1.27065100
H	-3.43530700	1.35646600	1.03235300
H	-3.33818000	-0.18465800	1.89191600
H	-1.94701900	0.89529700	1.85419600

Table 4.S29 Cartesian coordinates of compound **4** in the optimized *trans endo* conformation.

C	0.07836200	0.55073100	-0.66828600
H	0.24768300	-0.10976900	-1.51057600
C	-1.99184100	1.51572600	0.19284000
H	-2.33338700	1.49524200	1.22579200
H	-2.86611600	1.60928600	-0.45582300
C	-0.97908200	2.63812700	-0.05820500
H	-0.45438100	2.87123600	0.86514600
H	-1.45929700	3.54470100	-0.42201000
C	-0.00393500	2.03388700	-1.07800900
H	-0.40941300	2.10326600	-2.08847300
H	0.96914500	2.52102200	-1.07274000
N	-1.25843100	0.27783900	-0.13612300
C	-1.72688600	-0.96529700	0.03230800
C	-3.12145500	-1.07668400	0.59508100
H	-3.14964100	-0.69178800	1.61700200
H	-3.83696000	-0.49760600	0.00747500
H	-3.42712500	-2.11578900	0.60534100
C	1.13759000	0.33739500	0.43354000
N	2.36239200	-0.11840800	0.05013600
C	3.35860200	-0.38700600	1.08113700
H	3.46195900	-1.46211200	1.24870600
H	4.32540900	0.01488300	0.77651300
H	3.04208000	0.08346600	2.00495700
C	2.65054500	-0.64404700	-1.28192700
H	3.72710000	-0.76887900	-1.37185000
H	2.17054900	-1.61071300	-1.44896600
H	2.33991300	0.04730000	-2.06223400
O	0.87386100	0.65276100	1.58730100
S	-0.84234600	-2.33680300	-0.34410200

Table 4.S30 Cartesian coordinates of compound **4** in the optimized *cis endo* conformation.

C	0.31301500	0.41666200	-0.62095200
H	0.42766200	-0.11183900	-1.56257800
C	-1.58541900	1.59652900	0.35028500
H	-1.98578100	1.44725200	1.35048400
H	-2.40704500	1.94132100	-0.27990700
C	-0.37385100	2.53134100	0.30984200
H	0.15052900	2.50628000	1.26379800
H	-0.65796700	3.56177900	0.10662100
C	0.51007200	1.93801900	-0.79468200
H	0.14555800	2.23406700	-1.77933900
H	1.55299400	2.23929700	-0.71734400
N	-1.07938700	0.31875900	-0.17673300
C	-1.84201200	-0.78506200	-0.24440800
C	-1.17987900	-2.03417600	-0.77500100
H	-0.86812900	-1.90344200	-1.81452100
H	-0.29651100	-2.29117600	-0.18762200
H	-1.88332200	-2.85672500	-0.73049700
C	1.28469200	-0.15363200	0.43960800
N	2.54869800	-0.48273800	0.03200300
C	3.49827800	-0.98321500	1.01867500
H	3.94360900	-1.91456700	0.66432400
H	4.29926000	-0.25833100	1.18425300
H	2.97720400	-1.15940200	1.95244700
C	3.09961600	-0.29061800	-1.29974600
H	3.93832200	0.41002600	-1.26604500
H	3.46974900	-1.23908300	-1.69576700
H	2.36481800	0.10274200	-1.99090900
O	0.91879700	-0.27771400	1.59764800
S	-3.44709900	-0.79710800	0.21455000

Table 4.S31 Cartesian coordinates of compound **4** in the optimized *trans* *exo* conformation.

C	0.00461300	0.72564000	-0.49472900
C	-0.22360600	2.25706400	-0.53495400
C	-1.74830900	2.42791700	-0.48081100
C	-2.19788100	1.24866500	0.37899000
H	0.20036400	0.31227800	-1.47930200
H	0.23036600	2.71480600	0.34250500
H	0.22281200	2.70535500	-1.42016600
H	-2.04946700	3.38586600	-0.06077900
H	-2.18305300	2.34393600	-1.47778600
H	-2.06960200	1.46202300	1.44327200
H	-3.22986300	0.96244500	0.20158100
C	-1.52913000	-1.13677400	0.00542000
C	1.12507500	0.38474000	0.50850600
O	0.88329500	0.46826900	1.70553800
C	-2.87296800	-1.55301600	0.55240200
H	-3.68334000	-1.20175000	-0.09070900
H	-2.91968500	-2.63392800	0.60530300
H	-3.03607600	-1.13797300	1.54876800
C	3.41443700	-0.25707600	0.95822700
H	3.11977300	0.05612200	1.95329900
H	3.56774400	-1.33946700	0.96049500
H	4.34912000	0.23003000	0.68020100
N	-1.27633200	0.17700400	-0.02615100
N	2.36529000	0.11577900	0.01580500
C	2.63218900	-0.16039800	-1.39303200
H	2.25400700	0.63228200	-2.03526900
H	3.70977200	-0.20179100	-1.53257600
H	2.19835100	-1.11289800	-1.70396100
S	-0.43730100	-2.28709600	-0.53168000

Table 4.S32 Cartesian coordinates of compound **4** in the optimized *cis* *exo* conformation.

C	-0.32102800	0.61930500	0.40799300
C	-0.49679000	2.12658200	0.08423300
C	0.93090000	2.68164800	0.05114600
C	1.74706900	1.51624800	-0.49858300
H	-0.47205900	0.43166400	1.46953000
H	-0.95972300	2.23329100	-0.89684100
H	-1.13695000	2.62404100	0.81023400
H	1.01552600	3.57644500	-0.56246500
H	1.26985300	2.93009800	1.05793800
H	1.70799300	1.46367300	-1.58938600
H	2.79134800	1.51984000	-0.20232100
C	1.72197200	-0.82361600	0.30345000
C	-1.29689000	-0.21605100	-0.44651800
O	-0.93781100	-0.64095000	-1.53276200
C	0.92007300	-1.92110700	0.95904400
H	0.18111700	-2.31438400	0.25945100
H	1.58676400	-2.72505700	1.24646200
H	0.38968400	-1.56706200	1.84606900
C	-3.51565500	-1.15030400	-0.78573100
H	-2.99281500	-1.59547800	-1.62404400
H	-3.99296100	-1.93113100	-0.19142500
H	-4.29206400	-0.47960600	-1.16278400
N	1.08446000	0.33959500	0.07761700
N	-2.55913600	-0.42075100	0.03883800
C	-3.11931200	0.15384000	1.25169000
H	-3.90099500	0.87844500	1.00738400
H	-3.56768200	-0.63165800	1.86333600
H	-2.36757400	0.65727200	1.84638400
S	3.32148300	-1.08087300	-0.10568400

Table 4.S33 Cartesian coordinates of compound **5** in the optimized *trans endo* conformation.

C	-0.13732700	-0.25176200	0.85191800
H	0.05485400	0.46110000	1.64579400
C	-2.34561600	-0.94538100	0.03310000
H	-2.64397000	-1.06598900	-1.00748100
H	-3.24110400	-0.73156400	0.62237700
C	-1.60941300	-2.18646400	0.55638700
H	-1.17447200	-2.73169900	-0.27562400
H	-2.27832500	-2.85354000	1.09801500
C	-0.49335800	-1.62954300	1.45493800
H	-0.86125300	-1.47493500	2.47020100
H	0.36769000	-2.29123600	1.50614300
N	-1.37282800	0.14665500	0.17725400
C	-1.55507400	1.44154700	-0.18764000
C	-2.85904000	1.78177600	-0.87763000
H	-2.93417300	1.26047400	-1.83301100
H	-3.72050400	1.49132100	-0.27501700
H	-2.88358700	2.85256500	-1.05132300
C	1.06706600	-0.30643600	-0.10296000
N	2.16983800	0.36583900	0.26494100
C	3.38937500	0.25434500	-0.53442200
H	3.32362600	0.87056100	-1.43352000
H	4.23267100	0.58011400	0.06892300
H	3.53138300	-0.77536000	-0.84842100
C	2.18597900	1.37570000	1.33610900
H	3.07931800	1.98029700	1.21543500
H	1.31909500	2.02638300	1.25293700
H	2.22110500	0.91457500	2.32451500
S	0.98410600	-1.22852500	-1.49978800
O	-0.69658400	2.29415000	0.03031200

Table 4.S34 Cartesian coordinates of compound **5** in the optimized *cis endo* conformation.

C	-0.11153800	-0.40934000	0.68075100
H	0.02460800	0.04648700	1.65934400
C	-2.13808100	-1.26817000	-0.38205000
H	-2.46607200	-1.03813000	-1.39354400
H	-3.02580200	-1.50091200	0.20841800
C	-1.08954300	-2.38367100	-0.33422900
H	-0.54992100	-2.43029900	-1.27809100
H	-1.53397800	-3.35979700	-0.14936000
C	-0.13883800	-1.95623300	0.79301400
H	-0.55112200	-2.22838700	1.76592200
H	0.84878400	-2.40464200	0.70659700
N	-1.45127900	-0.10856500	0.20410300
C	-2.10698700	1.08294100	0.31635900
C	-1.35121100	2.24145900	0.93971100
H	-1.09929500	2.03548600	1.98199800
H	-0.42737200	2.44816600	0.39981100
H	-1.99306200	3.11562400	0.90193600
C	1.02551900	0.07067300	-0.24593300
N	2.24834500	0.18894400	0.31144300
C	3.37419600	0.65788400	-0.49661300
H	3.10420000	1.56878200	-1.02547500
H	4.21756300	0.84774100	0.16301700
H	3.65330400	-0.08663600	-1.24294900
C	2.59561700	-0.16162800	1.68920300
H	3.46132800	-0.82498900	1.68487000
H	2.85368900	0.73355600	2.25931100
H	1.78957500	-0.67621800	2.19361600
S	0.74166000	0.39631600	-1.85576200
O	-3.26340600	1.20087200	-0.06416700

Table 4.S35 Cartesian coordinates of compound **5** in the optimized *trans exo* conformation.

C	0.17115100	-0.45783800	-0.72730300
C	0.58887600	-1.92911500	-0.97651900
C	2.12132200	-1.92430600	-0.86618400
C	2.39107600	-0.85132000	0.19063400
H	0.01111000	0.05118300	-1.67284200
H	0.16190100	-2.55774500	-0.19909500
H	0.23371200	-2.28325600	-1.94264100
H	2.52172000	-2.89655600	-0.58418900
H	2.57799600	-1.63195600	-1.81306400
H	2.26807600	-1.25356900	1.19912900
H	3.38261100	-0.41415400	0.10542700
C	1.41253500	1.48317900	0.09887800
C	-1.08171100	-0.34660600	0.15658900
C	2.64985400	2.03477600	0.77760700
H	3.52768100	1.92512100	0.13850800
H	2.48573400	3.09021600	0.96908200
H	2.85377000	1.52053200	1.71664500
C	-3.41737500	0.26348400	0.38547800
H	-3.56290200	-0.67384000	0.91442000
H	-3.37859500	1.06263700	1.12847200
H	-4.24533900	0.43979800	-0.29593600
N	1.35241400	0.14168900	-0.09208200
N	-2.17599700	0.20870800	-0.38541600
C	-2.16050000	0.95042300	-1.65618700
H	-2.08571300	0.27919900	-2.51291500
H	-3.09385500	1.49767000	-1.73984200
H	-1.33993500	1.66385800	-1.66278400
S	-1.03698400	-0.95992100	1.71501100
O	0.49843500	2.21884400	-0.26800400

Table 4.S36 Cartesian coordinates of compound **5** in the optimized *cis exo* conformation.

C	-0.14658700	0.58364600	-0.47023400
C	-0.29907100	2.10031700	-0.14439100
C	-1.81037000	2.31870500	0.01347000
C	-2.29886300	0.98726900	0.58067700
H	-0.00731600	0.43963000	-1.54090600
H	0.20927900	2.31523500	0.79429900
H	0.14489300	2.72662300	-0.91673700
H	-2.03875200	3.16278200	0.66195800
H	-2.27555100	2.50199700	-0.95614500
H	-2.15520800	0.93235200	1.66241400
H	-3.33748500	0.76029000	0.36136400
C	-1.90189500	-1.22779900	-0.42403600
C	1.05920700	-0.00840500	0.28222200
C	-0.97099700	-2.13146600	-1.21027400
H	-0.09604400	-2.39429800	-0.61578300
H	-1.51796700	-3.03496800	-1.45923000
H	-0.62611000	-1.65827800	-2.13135800
C	3.42767500	-0.56042700	0.28859700
H	3.19711000	-1.55143500	0.67118600
H	4.22594700	-0.62379200	-0.44669000
H	3.75205000	0.05131300	1.13119600
N	-1.43700700	0.01448800	-0.09427000
N	2.24767100	0.02089100	-0.35284700
C	2.48867600	0.62553300	-1.66512400
H	3.44559200	1.14414600	-1.64117000
H	2.52966700	-0.13361000	-2.45007000
H	1.73050900	1.35548700	-1.91772600
S	0.87872700	-0.60125400	1.82955100
O	-3.02519300	-1.59316300	-0.10688100

Table 4.S37 Cartesian coordinates of compound **6** in the optimized *trans endo* conformation.

C	0.00468000	0.39471100	-0.81908400
H	0.10104000	-0.37254300	-1.57842600
C	-2.03639600	1.52044900	-0.07863700
H	-2.33309800	1.68225200	0.95580800
H	-2.93619600	1.50165700	-0.69735600
C	-1.03952700	2.57987700	-0.56020000
H	-0.51672200	3.00449600	0.29202400
H	-1.53577200	3.38553400	-1.09860900
C	-0.05627300	1.80388400	-1.44870700
H	-0.44284100	1.71562500	-2.46499000
H	0.92284900	2.27469700	-1.50120500
N	-1.31831600	0.23803000	-0.21337200
C	-1.82189300	-0.96122700	0.10715500
C	-3.19931400	-0.95821300	0.71956000
H	-3.18973100	-0.42145100	1.67089600
H	-3.92266600	-0.46014200	0.07036000
H	-3.52234200	-1.97724100	0.89385900
C	1.14341900	0.26289700	0.20405400
N	2.27044800	-0.32676200	-0.22850700
C	3.41649200	-0.43029000	0.67517300
H	3.20400400	-1.12333200	1.49078100
H	4.27744600	-0.77917900	0.11290900
H	3.63050200	0.54006600	1.11656600
C	2.35756000	-1.03392700	-1.51355100
H	3.34150500	-1.48309800	-1.59281600
H	1.60866000	-1.82467900	-1.56685100
H	2.23133700	-0.35361600	-2.35564000
S	-0.99765800	-2.39933800	-0.13482000
S	0.98544500	0.91453300	1.73677500

Table 4.S38 Cartesian coordinates of compound **6** in the optimized *cis endo* conformation.

C	0.23261100	0.46022900	-0.71055100
H	0.32690900	-0.06281600	-1.65806100
C	-1.70004400	1.58221300	0.27240000
H	-2.06369500	1.43909900	1.28744100
H	-2.55512900	1.86514700	-0.34355700
C	-0.54101200	2.57710700	0.17684900
H	-0.00772700	2.62457200	1.12386800
H	-0.88375100	3.57977000	-0.07087300
C	0.36594900	1.98965200	-0.91163100
H	-0.00914700	2.24662400	-1.90329200
H	1.39562900	2.33370100	-0.83806300
N	-1.14575900	0.31709300	-0.23750400
C	-1.88363300	-0.80331800	-0.33013700
C	-1.19542100	-2.02702800	-0.88645500
H	-0.90955500	-1.87501600	-1.93081900
H	-0.29363500	-2.26425900	-0.31941100
H	-1.87462500	-2.86956800	-0.83914400
C	1.29535900	-0.07687800	0.26643600
N	2.49387300	-0.38266900	-0.27291500
C	3.54579900	-0.93333300	0.58384700
H	3.18951900	-1.82330500	1.10049600
H	4.40332700	-1.18522400	-0.03342900
H	3.83751700	-0.20970600	1.34401200
C	2.80021000	-0.28989400	-1.70355100
H	3.86943600	-0.13641800	-1.82127300
H	2.52520700	-1.20481300	-2.23506400
H	2.29854500	0.55187600	-2.16793700
S	-3.48718300	-0.86145900	0.12763000
S	0.97440000	-0.22748600	1.89329500

Table 4.S39 Cartesian coordinates of compound **6** in the optimized *trans exo* conformation.

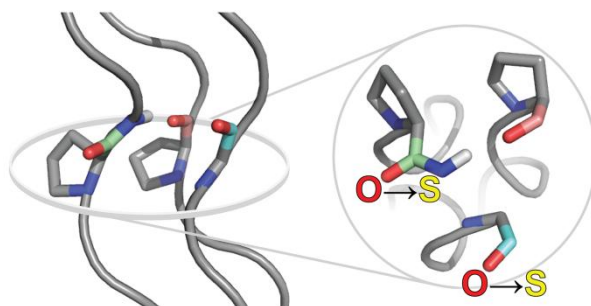
C	0.04905300	-0.60106100	-0.68015100
C	0.22438800	-2.11925400	-0.94263800
C	1.73867700	-2.35733400	-0.86197100
C	2.19720500	-1.33686100	0.17820300
H	-0.06530500	-0.05490200	-1.61143000
H	-0.28367400	-2.68095400	-0.16200800
H	-0.19964900	-2.40602000	-1.90319900
H	1.98676000	-3.37819100	-0.57724200
H	2.21608600	-2.14763100	-1.82011300
H	2.01919100	-1.69603900	1.19415800
H	3.24382900	-1.06869600	0.07621500
C	1.65759200	1.10731100	0.07180900
C	-1.13979300	-0.33130300	0.25359500
C	2.99140800	1.38589900	0.72082600
H	3.81456100	1.06464400	0.07830700
H	3.08950400	2.45016800	0.89733100
H	3.08225900	0.85377500	1.66945700
C	-3.45136400	0.33068500	0.52845400
H	-3.62751000	-0.53769200	1.15860900
H	-3.29727200	1.18900300	1.18424300
H	-4.31202200	0.50711300	-0.10977100
N	1.33003400	-0.17982900	-0.09357200
N	-2.27514600	0.10569500	-0.31207600
C	-2.35332400	0.53747700	-1.71328100
H	-2.15193100	-0.28701500	-2.39686900
H	-3.35935300	0.89278300	-1.90800800
H	-1.65182500	1.35081100	-1.90125500
S	0.66695300	2.37173900	-0.40264100
S	-1.00263500	-0.69833700	1.88026800

Table 4.S40 Cartesian coordinates of compound **6** in the optimized *cis exo* conformation.

C	0.22592200	0.71135300	-0.48780800
C	0.30403100	2.23231300	-0.17803600
C	-1.15756300	2.68491400	-0.08659000
C	-1.86956900	1.46301000	0.48519200
H	0.35182700	0.54038600	-1.55511500
H	0.80063000	2.37423400	0.78012000
H	0.87276200	2.76803300	-0.93644200
H	-1.27824800	3.56766900	0.53840400
H	-1.55368100	2.91575000	-1.07648600
H	-1.77564300	1.40233700	1.57154800
H	-2.92266300	1.39702900	0.23142500
C	-1.73591200	-0.84010200	-0.41504500
C	1.30716700	-0.04991400	0.29538900
C	-0.88871900	-1.85568400	-1.14287100
H	-0.06563800	-2.19083500	-0.50975100
H	-1.50152200	-2.71093400	-1.40020300
H	-0.46169000	-1.44111200	-2.05968000
C	3.53185600	-1.01920500	0.33302700
H	3.14870500	-1.96056600	0.72175800
H	4.33362300	-1.21245600	-0.37358600
H	3.91396900	-0.44779700	1.17889700
N	-1.15346200	0.34001000	-0.13293400
N	2.46710400	-0.28132300	-0.35049000
C	2.73964100	0.15633800	-1.72404500
H	3.81501600	0.18140300	-1.87103700
H	2.31028300	-0.52618500	-2.46111400
H	2.36391800	1.15863700	-1.90700400
S	-3.31277800	-1.20454200	-0.00249000
S	1.04476200	-0.49074200	1.87979900

CHAPTER V

Thioamides in the Collagen Triple Helix



To probe noncovalent interactions within the collagen triple helix, backbone amides were replaced with a thioamide isostere. This subtle substitution is the first in the collagen backbone that does not compromise thermostability. A triple helix with a thioamide as a hydrogen bond donor was found to be more stable than triple helices assembled from isomeric thiopeptides.

This chapter has been published, in part, under the same title: Newberry, R.W.; VanVeller, B.; Raines, R. T. *Chem. Commun.* **2015**, 51, 9624-9627.

B.V. and R.T.R. conceived the project. All authors contributed to the design of the experiments, which were performed by R.W.N. and B.V. R.W.N. drafted the manuscript, which was edited by all authors.

Introduction

Collagen is the main structural protein in animals, forming a significant portion of the extracellular matrix and the dry weight of skin. At the core of collagen is a three amino-acid repeat, Xaa-Yaa-Gly,²⁵⁶ that imparts a unique three-dimensional structure in which three individual polypeptide strands wind into a helix with a single-residue offset (Figure 5.1).²⁵⁵

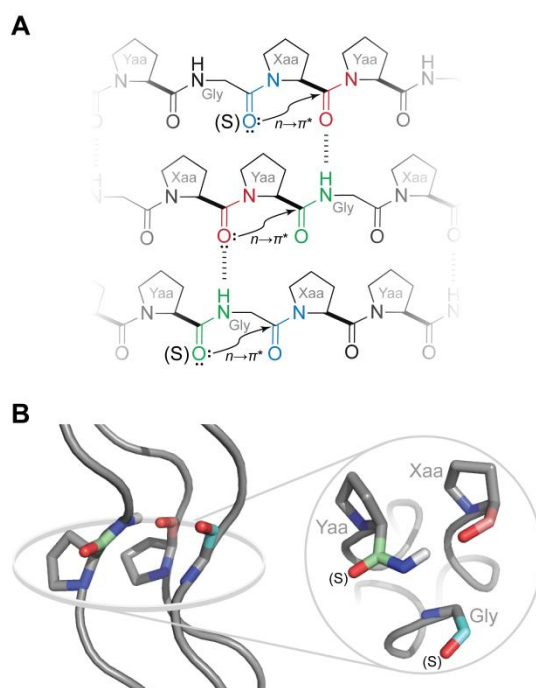


Figure 5.1 Conformation and intermolecular interactions of the collagen triple helix.

(A) Schematic diagram and (B) molecular model of the collagen triple helix. $n \rightarrow \pi^*$ Interactions across an Xaa, Yaa, and Gly residue and backbone amides serving as hydrogen bond acceptors (red), hydrogen bond donors (green), or neither (blue) are indicated explicitly. The image in panel B was derived from PDB entry 1v4f.³⁷² Thioamides at the peptide bonds indicated with an “(S)” were examined in this work.

Each individual strand adopts a polyproline type II conformation. This conformation is devoid of intramolecular hydrogen bonding within the peptide backbone, thereby allowing for intermolecular hydrogen bonds to form between strands.²⁵⁹ Each of those hydrogen bonds contributes approximately 2 kcal/mol to stability.¹⁸⁹ The unique structure of collagen is enforced further by the high (2*S*)-proline and (2*S*,4*R*)-4-hydroxyproline content of collagen proteins.

These amino acids serve not only to template the ideal backbone dihedral angles of the collagen triple helix, but also to effect attraction between adjacent backbone carbonyl groups through an $n \rightarrow \pi^*$ interaction.⁹¹ The interplay of these steric and electronic contributions gives collagen its robust thermal and mechanical stability, and has thus received significant attention.¹⁸³

Most previous studies on collagen stability have relied on the modification of protein side chains;^{183-184, 190-191, 194, 264, 343, 345, 349, 373-375} however, the core triple-helical structure of collagen is enforced almost exclusively through backbone–backbone hydrogen bonds and $n \rightarrow \pi^*$ interactions (Figure 5.1), which have been probed directly in only a few cases^{189, 192-193} and indirectly in others.^{174, 345, 374}

Those previous studies have employed either esters¹⁸⁹ or alkenes^{189, 192-193} as surrogates for the peptide bond, and examined the role of hydrogen bonding as well as *cis/trans* isomerization within the backbone as contributors to collagen stability. Though informative, the use of ester or alkene surrogates necessarily results in the deletion of donors or acceptors of hydrogen bonds, the alteration of $n \rightarrow \pi^*$ interactions, or the complete restriction of bond rotations.¹⁷¹ Here we sought to use the most subtle of isosteres to interrogate the backbone of collagen—a thioamide.

Results

Replacing the oxygen of a peptide bond with sulfur has several consequences.¹⁷¹ First, the thiocarbonyl bond is longer than that of a carbonyl group, which could disturb close atomic packing within the triple helix. Second, the thioamide has a higher rotational barrier than does an oxoamide, owing to greater C–N double bond character.^{362-363, 376-377} Third, thioamides are weaker hydrogen bond acceptors than are oxoamides.³⁷⁸ Fourth, a thioamide is a stronger hydrogen bond donor than is an oxoamide.³⁷⁸ Finally, thioamides are known to perturb backbone

$n \rightarrow \pi^*$ interactions.^{98, 100} Thioamides therefore provide a strategy for subtly modulating a variety of effects on collagen stability.

To deconvolute these effects, we have pursued site-specific thioamide-substitution within a collagen mimetic peptide (CMP): (Pro-Pro-Gly)₁₀ (**O** in Table 5.1).

Table 5.1 Sequence of CMPs and thermostability of their triple helices		
CMP	Sequence	$T_m \pm \text{SE } (^{\circ}\text{C})^a$
O	(PPG) ₁₀	30.5 ± 1.0
Y	(PPG) ₄ -PP ^S GPPG-(PPG) ₄	31.4 ± 1.1
G	(PPG) ₄ -PPG ^S PPG-(PPG) ₄	24.3 ± 0.7
G-N	PPG-PPG ^S PPG-(PPG) ₇	24.7 ± 1.9
G-C	(PPG) ₇ -PPG ^S PPG-PPG	29.5 ± 0.8

^aValues were obtained in 50 mM sodium phosphate buffer, pH 7.0, as described in Figure 5.7.

Collagen has only three types of peptide bonds: Xaa-Yaa, Yaa-Gly, and Gly-Xaa (Figure 5.1). The carbonyl of the Xaa-Yaa peptide bond (red in Figure 5.1) is directed toward the center of the collagen triple helix, where it accepts a hydrogen bond from another strand. Accordingly, this position is likely to be especially sensitive to the steric effects of thioamide-incorporation. Indeed, our molecular modeling suggested significant steric clashes of the thioamide with the backbone of adjacent strands (Figure 5.2). Accordingly, a thioamide is unlikely to be tolerated in this position.

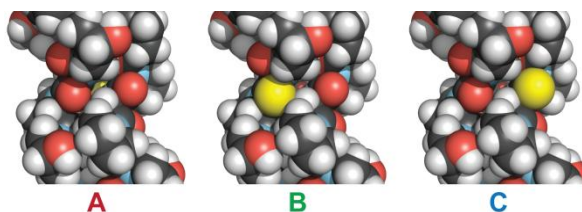


Figure 5.2 Space-filling models of thioamide substitutions in the collagen triple helix.

Space-filling models of collagen triple helices in which a thioamide replaces an (A) Xaa-Yaa, (B) Yaa-Gly, or (C) Gly-Xaa peptide bond. The thioamide sulfur is rendered in yellow in the CPK coloring scheme. Models were made by replacing an oxygen with sulfur in PDB entry 1v4f.³⁷² Thioamides at the (B) Yaa-Gly and (C) Gly-Xaa peptide bonds were examined in this work.

On the other hand, the Yaa-Gly peptide bond (green in Figure 5.1) serves as a hydrogen bond donor, and its carbonyl group points away from the center of the triple helix (Figure 5.2). This position is likely to be relatively insensitive to the steric effects of thioamide-incorporation while being sensitive to electronic effects on hydrogen bonding and the $n \rightarrow \pi^*$ interaction. The Gly-Xaa peptide bond (blue in Figure 5.1) is not engaged in a hydrogen bond. Thus, this position should be insensitive to the effect of thioamide-incorporation on hydrogen bonding. Moreover, the carbonyl group of the Gly-Xaa peptide bond is oriented away from the center of the triple helix, which should reduce the effect of thioamide sterics on collagen stability (Figure 5.2). Consequently, the Gly-Xaa peptide bond appears to isolate the effect of the thioamide on the $n \rightarrow \pi^*$ interaction.

The synthesis of thiopeptides presents numerous technical challenges. Two routes to thiopeptides are pursued commonly: thioacylation of free amines^{366, 379-382} and direct thionation of oxoamides.³⁸³⁻³⁸⁶ Thioacylation offers advantages in regiospecificity while suffering from lower yields and the difficulty of synthesizing activated thioacylating reagents. Thionation, while easier to implement and often affording higher yields, suffers from idiosyncratic dependencies on steric effects near the thionation site, complicating site-selective incorporation. Ultimately, we combined these synthetic approaches to access our targets.

Motivated by our predictions (*vide supra*), we established four thiopeptides as targets (Table 5.1). Peptides **Y** and **G** have thioamides replacing the Yaa-Gly and Gly-Xaa peptide bonds, respectively. Peptides **G-N** and **G-C** are controls (*vide infra*). We constructed these four target peptides by first introducing the thioamide into protected amino-acid trimers and then condensing those trimers on a solid phase using Fmoc-based chemistry. We discovered early-on that successful thioamide-incorporation was predicated on placing the thioamide distal to the

coupling site of the trimer. Peptide synthesis of trimers containing internal thioamides failed, likely due to attack of the thiocarbonyl sulfur on the activated ester generated during peptide coupling; similar results were observed for the incorporation of thioamides into proteins using native chemical ligation.³⁸⁷⁻³⁸⁹ Thus, to realize our CMPs, we pursued the synthesis of tripeptides **1** and **2** as building blocks for peptide synthesis (Figure 5.3).

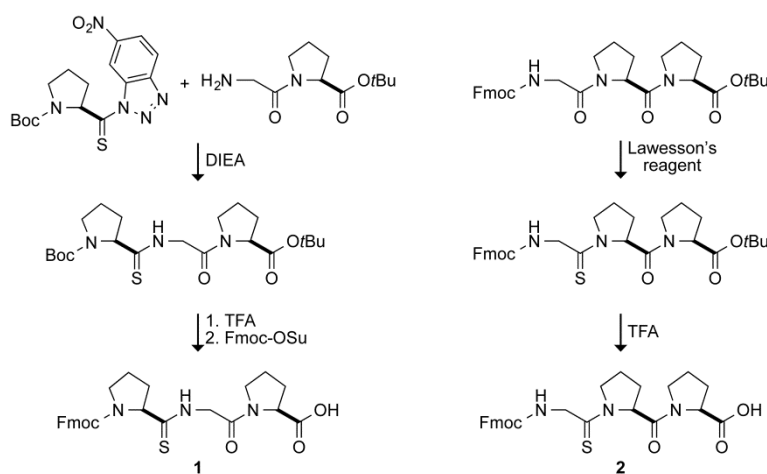


Figure 5.3 Synthesis of thioamide-substituted CMPs.

Synthetic route to tripeptide units **1** and **2**, which were used in the synthesis of thiopeptides **Y** and **G**, respectively.

To access tripeptide **1**, we first synthesized the thioacylated 6-nitrobenzotriazole derivative of Fmoc-ProOH.³⁶⁶ Test reactions with amine nucleophiles indicated, however, the presence of both thioamide and oxoamide products. In contrast, when we performed the same reaction with a thioacylated triazole prepared from Boc-ProOH, we observed quantitative yield of the thioamide.³⁹⁰ Hence, we used this route to access to trimer **1** (Figure 5.3).

To access tripeptide **2**, we used the known preference of Lawesson's reagent for amides over esters or carbamates,³⁹¹ along with previous observations on the inefficiency of thionation of proline residues.³⁹²⁻³⁹³ Accordingly, we employed direct thionation, treating Fmoc-Gly-Pro-Pro-OtBu with excess Lawesson's reagent at elevated temperatures (Figure 5.3). We observed only a

single, monothionated product. We confirmed its regiochemistry unambiguously with NMR spectroscopy, which demonstrated coupling between the thiocarbonyl carbon and the glycine methylene protons in heteronuclear multiple-bond correlation experiments.

With tripeptides **1** and **2** in hand, we synthesized 30-mer thiopeptides **Y** and **G**, respectively, each of which features a single thioamide near its center. We allowed thiopeptides **Y** and **G** to self-assemble overnight in neutral phosphate buffer before examining thermostability with circular dichroism spectroscopy. For comparison, we performed similar analyses on peptide **O**, which does not contain a thioamide. The spectra of all three peptides show the characteristic CD spectra of PPII helices and collagen, featuring a maximum at 226 nm (Figure 5.4).

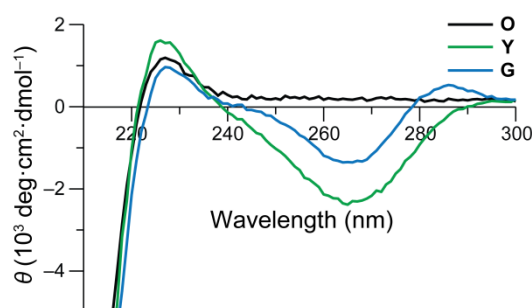


Figure 5.4 CD spectra of triple-helical, thioamide-substituted CMPs.

Representative far-UV circular dichroism spectra of (Pro-Pro-Gly)₁₀ peptides containing only oxoamides (**O**) or with a thioamide replacing a central Yaa-Gly (**Y**) or Gly-Xaa (**G**) peptide bond. Spectra were recorded at 4 °C in 50 mM sodium phosphate buffer, pH 7.0.

In addition, those peptides containing thioamides had additional features in the 250–300 nm range, corresponding to the thioamide electronic absorption. We note three interesting characteristics of these spectra. First, despite outnumbering the thioamide nearly thirty to one, the absorption band produced by the oxoamides is only marginally more intense than that of the thioamide. Second, the thioamide region of the CD spectra of **Y** and **G** differ qualitatively in the folded state, with the spectrum of **G** displaying a maximum at 287 nm in addition to the minimum that both spectra share at 265 nm. Third, no qualitative differences are observed in the

spectra of **Y** and **G** in the unfolded state (Figure 5.5). These observations suggest that the CD spectrum of the peptide in the thioamide region might report on local structural changes around the chromophore, as proposed by others.³⁹⁴⁻³⁹⁵

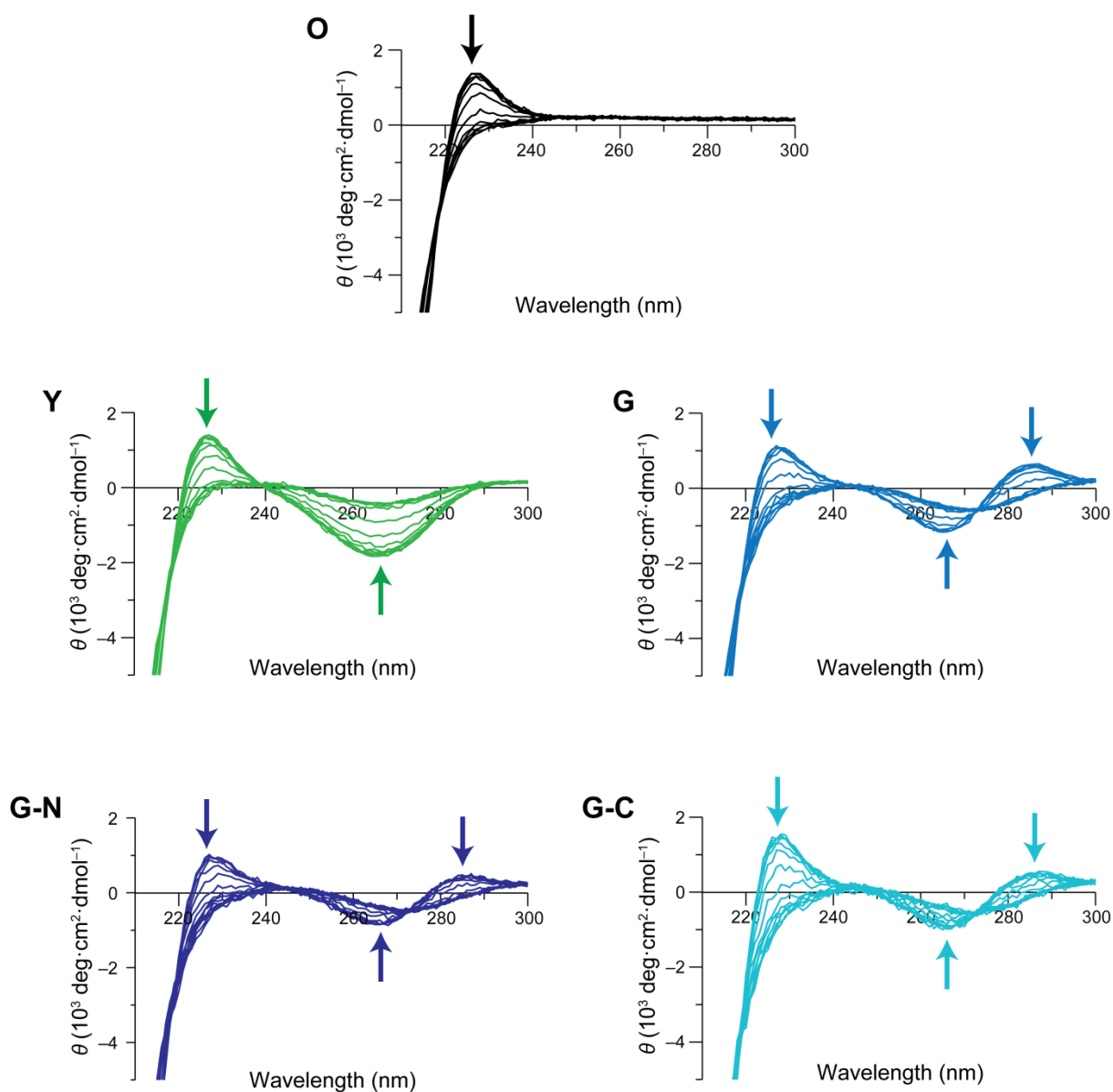


Figure 5.5 CD spectra of triple-helical, thioamide-substituted CMPs at elevated temperatures. Arrows indicate wavelengths that have large temperature-dependent changes in ellipticity.

Upon thermal denaturation, we observed that self-assembled thiopeptides **Y** and **G** exhibit cooperative denaturation, a hallmark of triple helices that distinguishes them from other PPII structures. Moreover, the transition temperatures of these peptides were consistent regardless of which spectral feature we tracked (Figure 5.6).

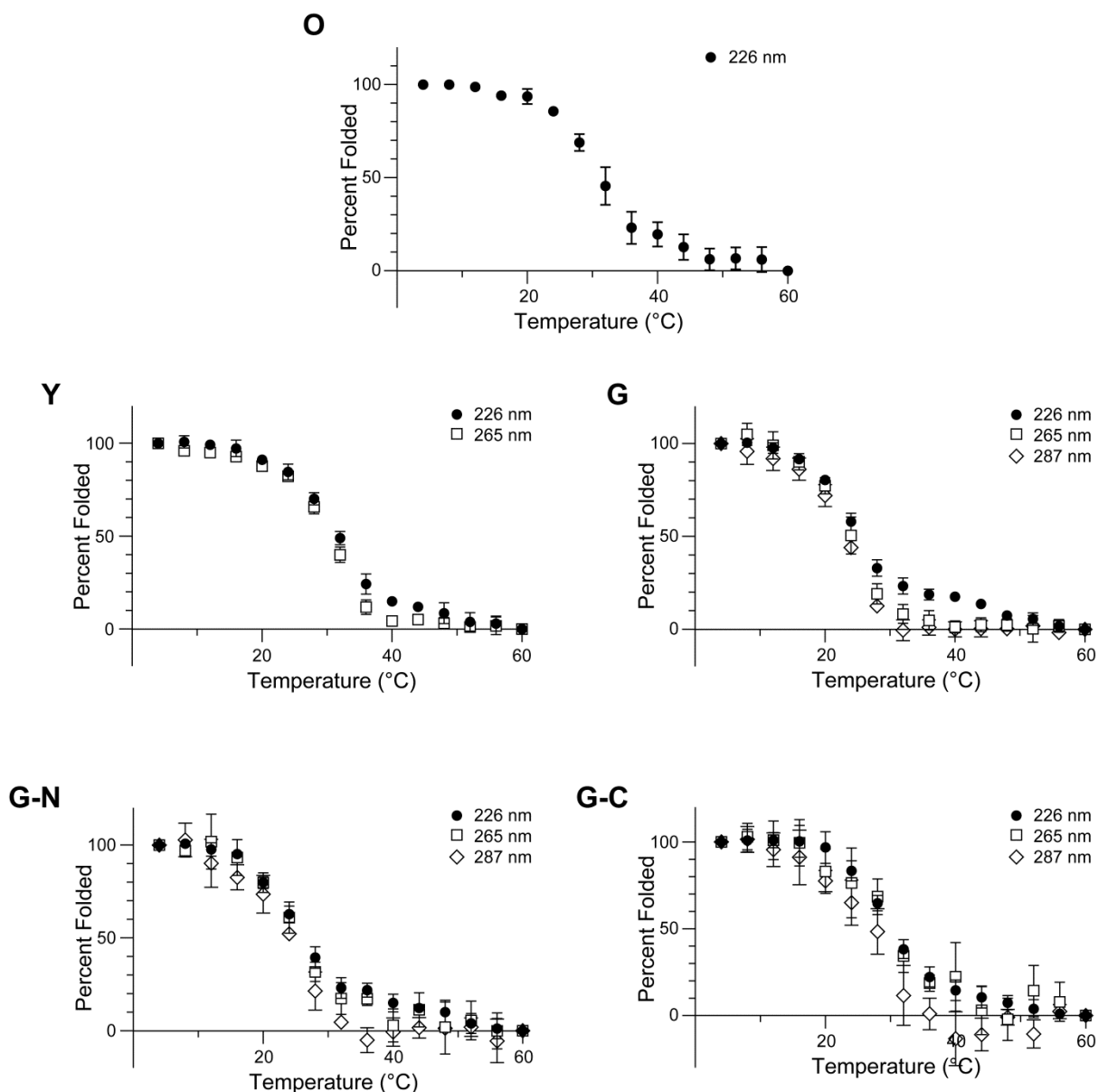


Figure 5.6 Thermal denaturation of triple-helical CMPs as monitored at various wavelengths.

In accord with our predictions, triple-helical **Y** denatured at a temperature 7 °C higher than that of triple-helical **G** (Figure 5.7A and Table 5.1). Again, the thioamide is a stronger hydrogen bond donor than is the oxoamide, so replacing the hydrogen bond donor at the Yaa–Gly position should increase stability of CMPs relative to replacement at the Gly–Xaa position, which does not participate in interstrand hydrogen bonding. Moreover, the thermostability of triple-helical **Y** was greater than that of triple-helical **O** (though not significantly so), making a thioamide the first backbone modification of a CMP that does not compromise the stability of a collagen triple helix. This tolerance is meaningful given the extreme constraints imposed by the tight packing of the backbone atoms in the center of the triple helix (Figure 5.1).

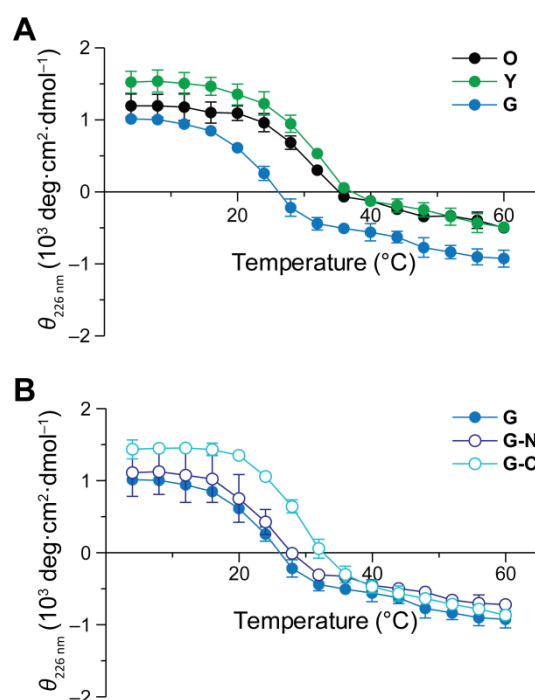


Figure 5.7 Thermal denaturation of triple-helical, thioamide-substituted CMPs.

(A) Thermal denaturation data of (Pro-Pro-Gly)₁₀ containing only oxoamides (**O**) or with a thioamide replacing a central Yaa-Gly (**Y**) or Gly-Xaa (**G**) peptide bond. (B) Thermal denaturation of (Pro-Pro-Gly)₁₀ with a thioamide replacing a central (**G**), N-terminal (**G-N**), or C-terminal (**G-C**) Gly-Xaa peptide bond. Denaturation was done in 50 mM sodium phosphate buffer, pH 7.0. Values of θ at 226 nm were recorded every 4 °C from 4–60 °C following a 5-min equilibration. Experiments were performed in triplicate. Data were fitted to a two-state model to obtain the values of T_m (\pm SE) listed in Table 5.1. For comparison, the data for peptide **G** (blue) are depicted in both panels A and B.

Why does a thioamide at the Gly-Xaa position diminish triple-helical stability despite the absence of a direct effect on interstrand hydrogen bonding? We reasoned that this thioamide could impart a steric clash. To test this hypothesis, we moved the thioamide from the center of the peptide toward the C terminus, resulting in thiopeptide **G-C** (Table 5.1). Upon thermal denaturation of this peptide, we found that, triple-helical **G-C** had a T_m value close to that of **O** (Figure 5.7B and Table 5.1). Other backbone modifications have been found to be less destabilizing at the C terminus than elsewhere,³⁹⁶ consistent with our results. In contrast, moving the thioamide from the center toward the N terminus, as in thiopeptide **G-N** (Table 5.1), led to no measurable change in thermostability (Figure 5.7). CMPs are thought to fold in a C- to N-terminal direction.³⁹⁷ We suspect that a disruption in the backbone near the N terminus might reduce the extent of folding, consistent with the lower values of θ observed for **G-N** relative to **G-C** (Figure 5.7). Regardless, these findings motivate additional study of end-effects on helix stability.

Conclusions

Our results demonstrate that thioamide-substitution is a conservative approach to the rational modification of the backbone of CMPs. The effect of thioamide-incorporation at different positions within the triplet amino-acid repeat is well predicted from known hydrogen-bonding properties of thioamides. Importantly, thioamide-substitution has yielded the first backbone-modified collagen triple helix that does not suffer a loss in thermostability. These thiopeptides inspire and enable a variety of future studies. For example, the *cis/trans* prolyl peptide bond isomerization during collagen folding could be modulated by exploiting the tendency of thioamides to isomerize upon UV radiation.^{377, 398-399} In addition, thioamide-containing CMPs

could be useful in high-throughput assays of collagen binding by accessing the ability of thioamides to quench pendant fluorophores.⁴⁰⁰ Finally, thioamides are protease-resistant,⁴⁰¹ making thioamide-incorporation a potential strategy for producing robust collagen-based biomaterials.²⁷⁰ Importantly, backbone modification of a CMP allows for the full suite of side-chain modifications, which are well-known to be effective in tuning various properties of collagens.^{183, 402-403}

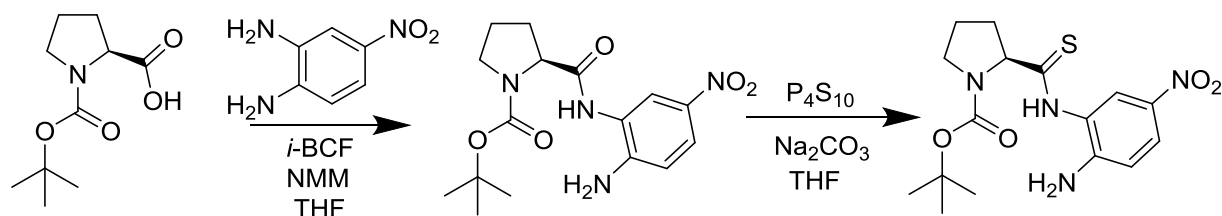
Acknowledgements. This work was supported by grant R01 AR044276 (NIH). R.W.N. was supported by NIH Biotechnology Training Grant T32 GM008349, an ACS Division of Organic Chemistry Graduate Fellowship, and an Eastman Summer Research Award. B.V. was supported by postdoctoral fellowship 289613 (CIHR). This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grant P41 GM103399 (NIH). This work also makes use of the Biophysics Instrumentation Facility at UW–Madison, which was established by grants BIR-9512577 (NSF) and S10 RR13790 (NIH).

Methods

General Experimental. Commercial chemicals were of reagent grade or better, and were used without further purification. Proline starting materials were obtained from Chem-Impex (Wood Dale, IL). All other chemicals were obtained from Sigma–Aldrich. Anhydrous THF and DMF were obtained from CYCLE-TAINER solvent delivery systems (J. T. Baker, Phillipsburg, NJ). Reactions were monitored by thin-layer chromatography with visualization by UV light or staining with KMnO_4 . Flash chromatography was performed with columns of silica gel 60, 230–400 mesh (Silicycle, Québec City, Canada). The removal of solvents and other volatile materials

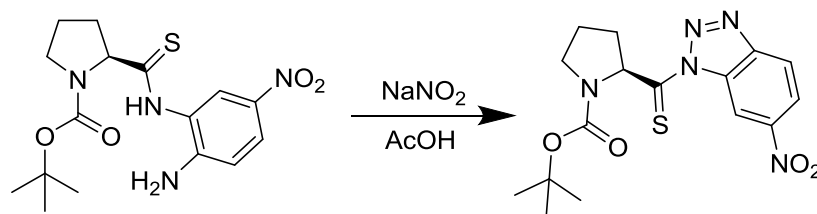
“under reduced pressure” refers to the use of a rotary evaporator at water-aspirator pressure (<20 torr) and a water bath of <45 °C. All yields are unoptimized.

Instrumentation. NMR spectra were acquired at ambient temperature with a Bruker Avance III 500 MHz spectrometer (^1H , 500 MHz; ^{13}C , 125 MHz) in the National Magnetic Resonance Facility at Madison (NMRFAM). ^{13}C spectra were proton-decoupled. Peptide synthesis was performed with a Protein Technologies Prelude automated synthesizer in the University of Wisconsin–Madison Biotechnology Center. Peptide purification was accomplished on a Shimadzu LC-20 HPLC. ESI mass spectrometry was performed with a Micromass LCT instrument in the Mass Spectrometry Facility of the Department of Chemistry at the University of Wisconsin–Madison or on a Shimadzu LCMS 2020. MALDI mass spectrometry was performed on a Voyager DE-Pro MALDI–TOF mass spectrometer in the Biophysics Instrumentation Facility at the University of Wisconsin–Madison. Circular dichroism spectra were collected using an AVIV Model 420 circular dichroism spectrometer in the Biophysics Instrumentation Facility at the University of Wisconsin–Madison.



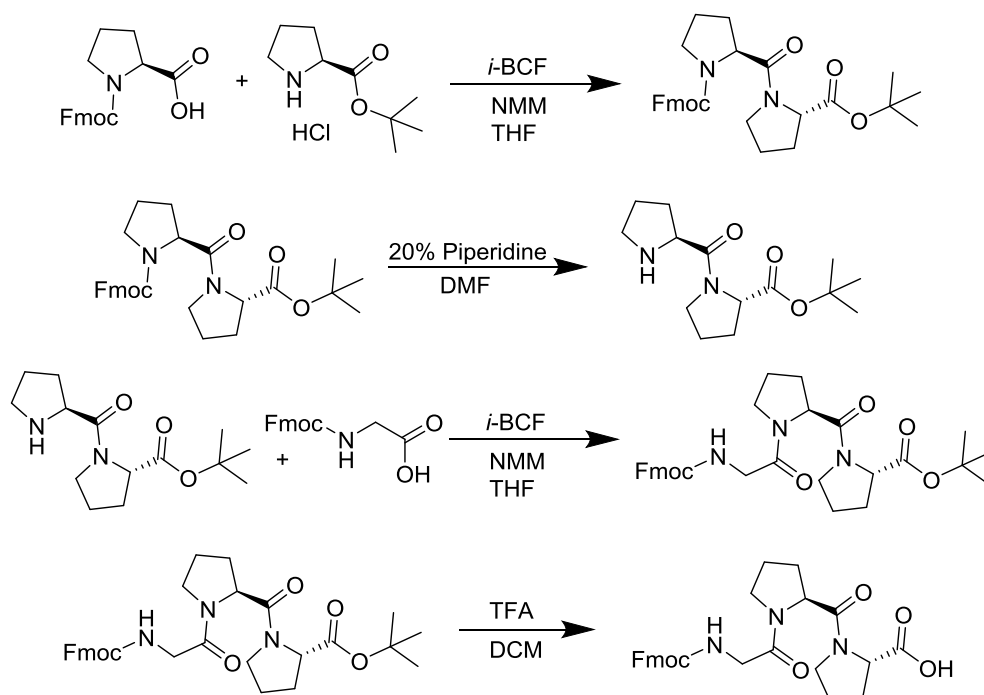
***N*-Boc-L-proline 2-amino-5-nitrothioanilide.** *N*-Methylmorpholine (4.1 mL, 37.2 mmol) was added to a solution of *N*-Boc-L-proline (4.0 g, 18.6 mmol) in THF (100 mL) at $-20\text{ }^{\circ}\text{C}$. Isobutyl chloroformate (2.4 mL, 18.6 mmol) was added dropwise and stirred for 10 min. 4-Nitro-1,2-phenylenediamine (2.85 g, 18.6 mmol) was added, and the reaction mixture was stirred at $-20\text{ }^{\circ}\text{C}$

for 3 h before warming to room temperature and stirring overnight. THF was removed under reduced pressure, and the resulting yellow solid was dissolved in EtOAc. This solution was washed with 1 M NaH_2PO_4 and saturated aqueous NaHCO_3 . The organic portion was dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$ and concentrated under reduced pressure. The resulting yellow oil was dissolved in THF, and this solution was added to a solution of P_4S_{10} (4.13 g, 9.3 mmol) and Na_2CO_3 (1.00 g, 9.3 mmol) in THF at 0 °C. The reaction mixture was allowed to warm to room temperature and stir overnight. The solvent was removed under reduced pressure. The residue was dissolved in EtOAc and washed with saturated aqueous NaHCO_3 . The aqueous portion was back-extracted with EtOAc. The organic portions were combined, dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$, and concentrated under reduced pressure. The resulting yellow oil was purified by chromatography on silica gel with an eluent of 4% v/v MeOH in DCM, affording a yellow solid (5.75 g, 85%). ^1H NMR (500 MHz, DMSO, mixture of two rotamers, δ): 11.22/11.09 (s, 1H), 8.01/7.83 (d, J = 2.5 Hz, 1H), 7.97 (dd, J = 9.1, 2.5 Hz, 1H), 6.85/6.78 (d, J = 9.1 Hz, 1H), 6.49/6.23 (s, 2H), 4.70/4.65 (dd, J = 8.4, 4.5 Hz, 1H), 3.52/3.42 (m, 2H), 2.32 (m, 1H), 2.06 (m, 2H), 1.84 (m, 1H), 1.41/1.39 (s, 9H); ^{13}C NMR (125 MHz, DMSO, δ): 207.7, 207.1, 154.4, 154.2, 151.0, 150.3, 135.5, 135.2, 125.2, 124.9, 124.8, 124.7, 122.6, 122.4, 114.5, 113.6, 79.5, 79.0, 47.3, 47.11, 33.9, 32.7, 28.2, 28.1, 24.0, 23.1; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 367.1435, found 367.1432.



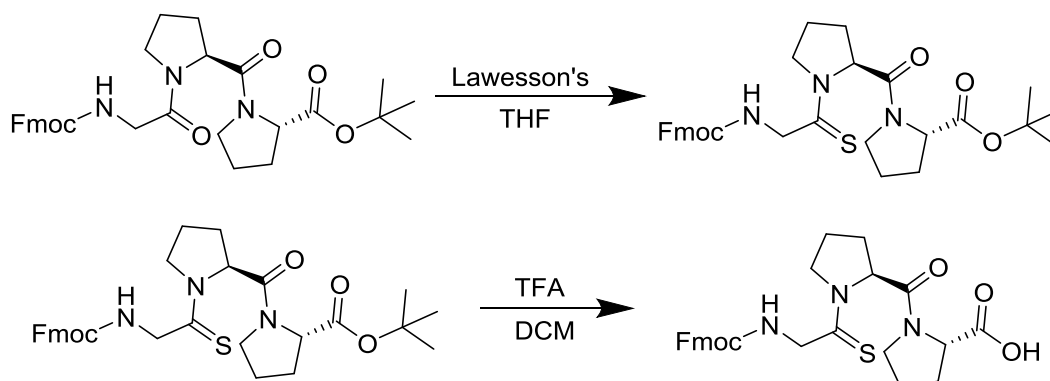
1-(*N*-Boc-*L*-thiopropyl)-6-nitrobenzotriazole. *N*-Boc-*L*-proline 2-amino-5-nitrothioanilide

(5.75 g, 15.7 mmol) was dissolved in glacial acetic acid diluted with 5% v/v H₂O, and the resulting solution was cooled to 0 °C. NaNO₂ (1.63 g, 23.6 mmol) was added portionwise over 5 min, and the reaction mixture was stirred for an additional 30 min at 0 °C. The solution was diluted with water and extracted twice with DCM. The organic portions were washed with saturated aqueous NaHCO₃, dried over anhydrous Na₂SO₄(s), and evaporated to dryness to yield an orange solid (5.00 g, 84%). ¹H NMR (500 MHz, CDCl₃, mixture of two rotamers, δ): 9.74 (d, J = 1.9 Hz, 1H), 8.48/8.45 (dd, J = 6.9, 1.9 Hz, 1H), 8.35/8.30 (d, J = 6.9 Hz, 1H), 6.22 (m, 1H), 3.71 (m, 2H), 2.67 (m, 1H), 2.12 (m, 3H), 1.48/1.24 (s, 9H); ¹³C NMR (125 MHz, CDCl₃, δ): 209.5, 208.0, 154.2, 153.3, 149.6, 148.7, 132.1, 131.9, 122.2, 122.0, 121.5, 121.3, 113.1, 112.8, 80.2., 80.2, 67.8, 67.6, 47.2, 46.9, 34.5, 33.5, 28.5, 28.2, 23.8, 23.1; ESI-MS: [M + H]⁺ calculated 378.1231, found 378.1224.

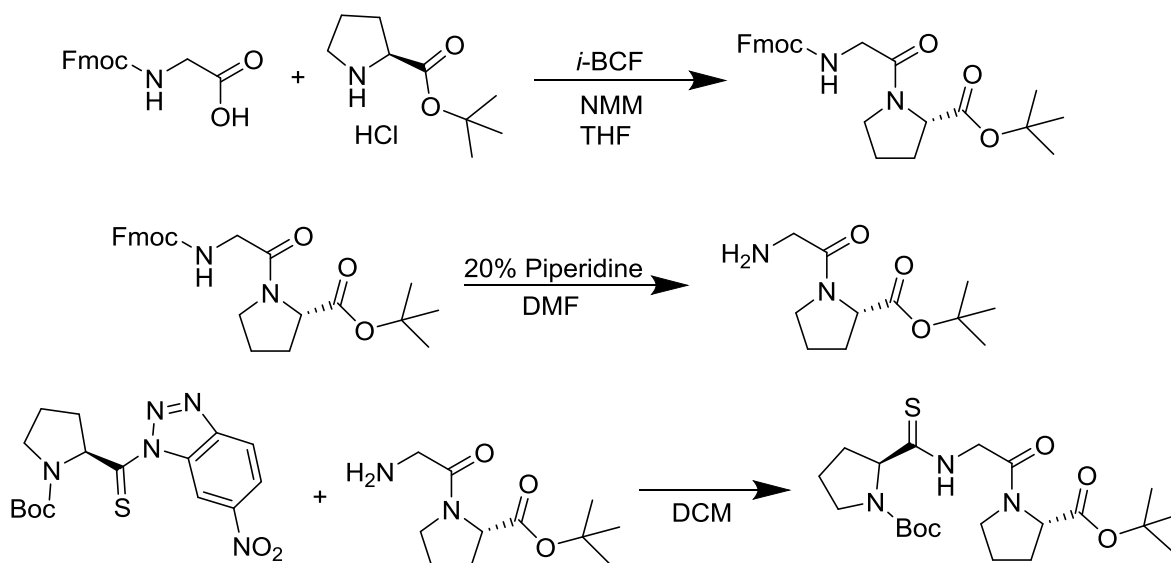


Fmoc-glycyl-prolyl-proline. Fmoc-proline (4.0 g, 11.8 mmol) was dissolved in 100 mL of anhydrous THF containing *N*-methylmorpholine (3.9 mL, 23.6 mmol). The solution was cooled to $-20\text{ }^{\circ}\text{C}$ before isobutyl chloroformate (1.54 mL, 11.8 mmol) was added dropwise. The solution was stirred for 10 min before the addition of proline *tert*-butyl ester hydrochloride (2.46 g, 11.8 mmol). The reaction mixture was stirred overnight and then allowed to warm to room temperature. The solvent was removed under reduced pressure, and the resulting yellow solid was dissolved in EtOAc and washed with 1 M NaH_2PO_4 and saturated aqueous NaHCO_3 . The organic portion was dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$ and concentrated under reduced pressure. The resulting solid (5.25 g, 10.7 mmol, 91% yield) was dissolved in 20 mL of 20% v/v piperidine in DMF and stirred for 1 h at room temperature before the solvent was removed under reduced pressure. Meanwhile, Fmoc-glycine (3.19, 10.7 mmol) dissolved in 100 mL anhydrous THF with *N*-methylmorpholine (3.9 mL, 21.4 mmol). The solution was cooled to $-20\text{ }^{\circ}\text{C}$ before isobutyl chloroformate (1.54 mL, 10.7 mmol) was added dropwise. The reaction stirred for ten minutes before the addition of prolyl-proline *tert*-butyl ester. The reaction was stirred overnight

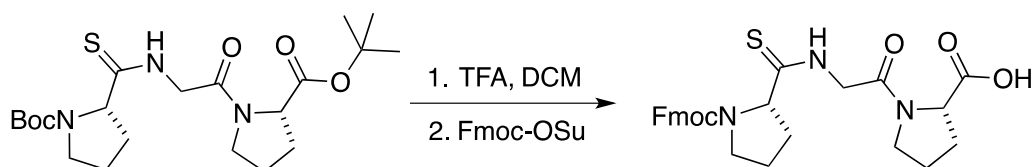
and allowed to warm to room temperature. THF was removed under reduced pressure, and the resulting yellow solid was dissolved in EtOAc and washed with 1 M NaH_2PO_4 and saturated aqueous NaHCO_3 . The organic portion was dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$, and solvent was removed under reduced pressure. Purification of the resulting solid by chromatography on silica gel in 4% v/v MeOH in DCM afforded Fmoc-glycyl-prolyl-proline *tert*-butyl ester as a crystalline solid, which was then dissolved in 1:1 TFA/DCM for 1 h. Removal of solvent under reduced pressure afforded Fmoc-glycyl-prolyl-proline as an off-white solid (3.15 g, 6.4 mmol, 55% overall yield).



Fmoc-thioglycyl-prolyl-proline. Fmoc-glycyl-prolyl-proline *tert*-butyl ester (2.0 g, 3.6 mmol) and Lawesson's reagent (0.74 g, 1.8 mmol) were dissolved in 100 mL THF, and the resulting solution was heated to reflux overnight. Solvent was removed under reduced pressure, and the residue was purified by chromatography on silica gel in 7:3 EtOAc/hexanes to afford Fmoc-thioglycyl-prolyl-proline *tert*-butyl ester as a crystalline solid (1.0 g, 1.8 mmol, 50% yield). Fmoc-thioglycyl-prolyl-proline *tert*-butyl ester (1.0 g, 1.8 mmol) was dissolved in 1:1 TFA/DCM for 1 h. Removal of solvent under reduced pressure afforded Fmoc-thioglycyl-prolyl-proline as an off-white solid (0.9 g, 1.8 mol, quantitative yield).



Boc-thiopropyl-glycyl-proline *tert*-butyl ester. Fmoc-glycine (2.97 g, 10.0 mmol) was dissolved in 100 mL anhydrous THF containing *N*-methymorpholine (3.3 mL, 20.0 mmol). The solution was cooled to $-20\text{ }^{\circ}\text{C}$ before isobutyl chloroformate (1.30 mL, 10.0 mmol) was added dropwise. The solution was stirred for 10 min before the addition of proline *tert*-butyl ester hydrochloride (2.07 g, 10.0 mmol). The reaction mixture was stirred overnight and then allowed to warm to room temperature. Solvent was removed under reduced pressure, and the resulting yellow solid was dissolved in EtOAc and washed with 1 M NaH₂PO₄ and saturated aqueous NaHCO₃. The organic portion was dried over anhydrous Na₂SO₄(s) and evaporated to dryness. The resulting solid was dissolved in 20 mL 20% v/v piperidine in DMF and stirred for 1 h at room temperature before the solvent was removed under reduced pressure. The resulting solid was dissolved in 100 mL DCM and 1-(*N*-Boc-L-thiopropyl)-6-nitrobenzotriazole (3.9 g, 10.0 mmol) was added. The reaction mixture was stirred at room temperature overnight. Following removal of solvent under reduced pressure, purification by chromatography on silica gel in 2% v/v MeOH in DCM afforded Boc-thiopropyl-glycyl-proline *tert*-butyl ester as a yellow solid (3.09 g, 7.0 mmol, 70% overall yield).



Fmoc-thiopropyl-glycyl-proline. Boc-thiopropyl-glycyl-proline *tert*-butyl ester (1.3 g, 2.94 mmol) was dissolved in 20 mL DCM and 20 mL TFA (2.5% H₂O, 1% TIPSH), and the resulting solution was stirred for 1 h at room temperature. Solvent was removed under reduced pressure and the residue was dissolved in 100 mL DCM. Fmoc-NHS ester (1 g, 3.2 mmol) and DIEA (3 mL, 17 mmol) were added, and the reaction mixture was stirred for 90 min at room temperature. Saturated aqueous NaHCO₃ was added, and the mixture was stirred for 90 min at room temperature. The solution was acidified and washed with 1 M HCl, and the organic layer was dried over MgSO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel in 99:1 EtOAc/AcOH, to give Fmoc-thiopropyl-glycyl-proline (81%).

Peptide Synthesis. Peptide synthesis was accomplished on a 25- μ mol scale using standard Fmoc chemistry protocols on NovaSyn Fmoc-Gly TGT resin from EMD Millipore (Darmstadt, Germany). Briefly, peptide bond formation was accomplished by treatment of deprotected resin with 4 equiv each of protected amino acid and HATU, and with 8 equiv of NMM. Cleavage was performed for 1 h in TFA containing 2.5% v/v H₂O and 1% v/v TIPSH. Peptides were precipitated with diethyl ether, isolated, and dissolved in acetonitrile/water before purification by reverse-phase HPLC on a preparative NucleoSil C18 column from Macherey–Nagel (Düren, Germany) using 0.1% v/v TFA in H₂O (A) and 0.1% v/v TFA in MeCN (B) as eluents. Lyophilization afforded white solids. Analytical HPLC was performed with an analytical NucleoSil C18 column from Macherey–Nagel with a 5–95% B gradient over 25 min.

Circular Dichroism Spectroscopy. Peptide samples were dissolved in 50 mM sodium phosphate buffer, pH 7.0, to a final concentration of 200 μ M and allowed to equilibrate overnight at 4 °C. Circular dichroism spectra were collected between 200 and 300 nm (1-nm bandwidth, 3-s averaging time) every 4 °C from 4 °C to 60 °C with a 5-min equilibration between temperature steps. Denaturation experiments were performed in triplicate. The mean residue ellipticity as a function of temperature was fitted to a two-state model to extract the value of T_m , which is the temperature at the midpoint of the thermal transition between the triple-helical and single-stranded states.

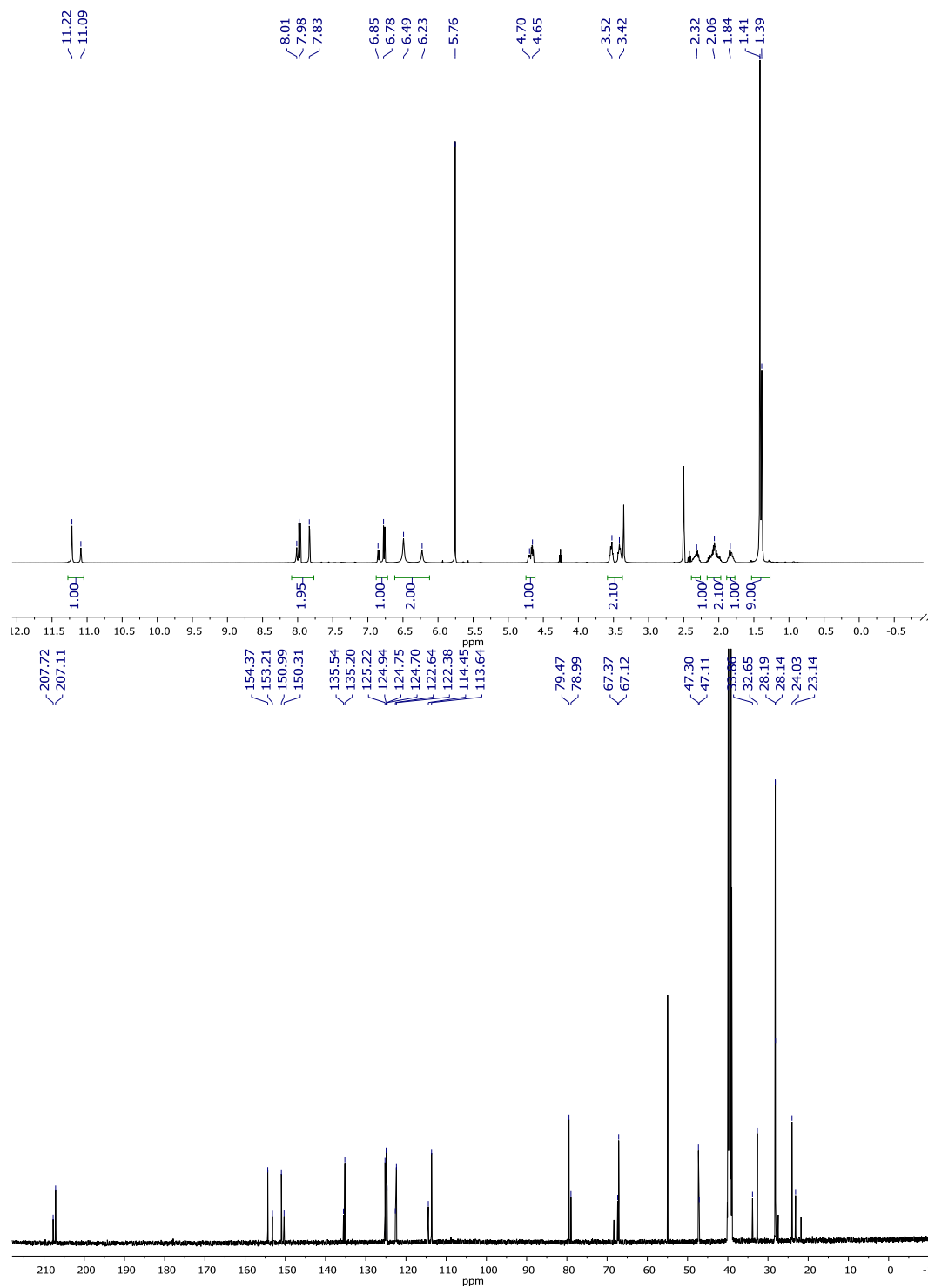


Figure 5.S1 ¹H and ¹³C NMR spectra of *N*-Boc-L-proline-2-amino-5-nitrothioanilide in DMSO/DCM.

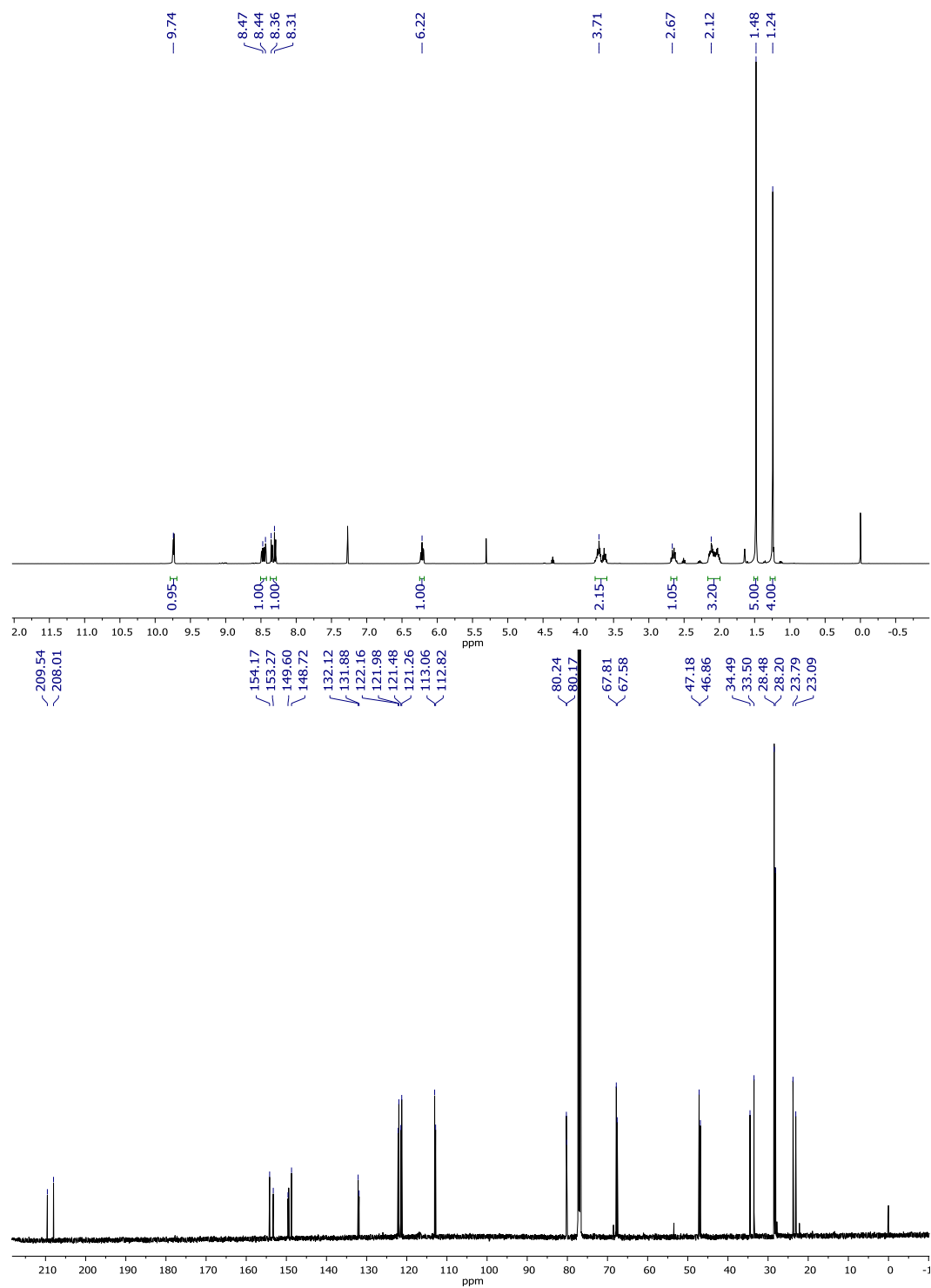


Figure 5.S2 ^1H and ^{13}C NMR spectra of 1-(*N*-Boc-L-thiopropyl)-6-nitrobenzotriazole in CDCl_3/DCM .

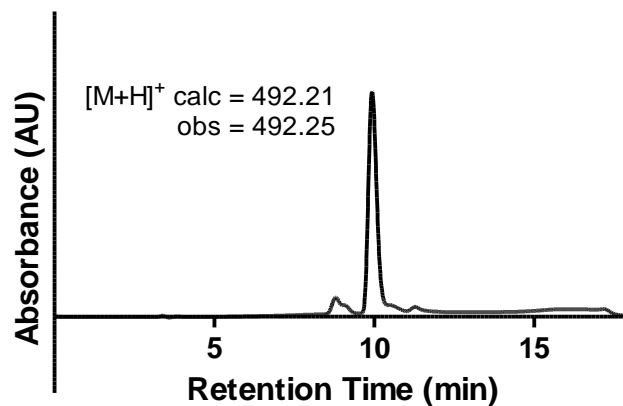


Figure 5.S3 LC-MS analysis of Fmoc-glycyl-prolyl-proline

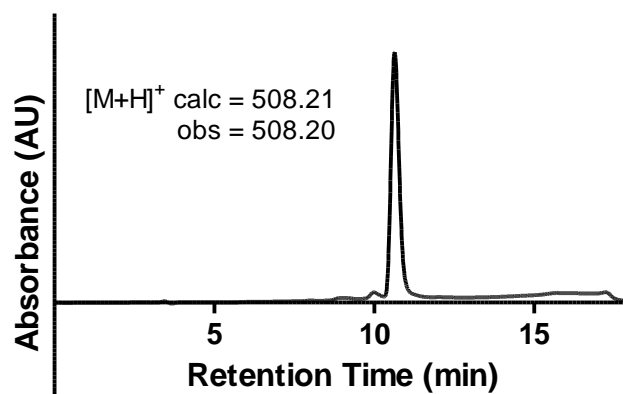


Figure 5.S4 LC-MS analysis of Fmoc-thioglycyl-prolyl-proline

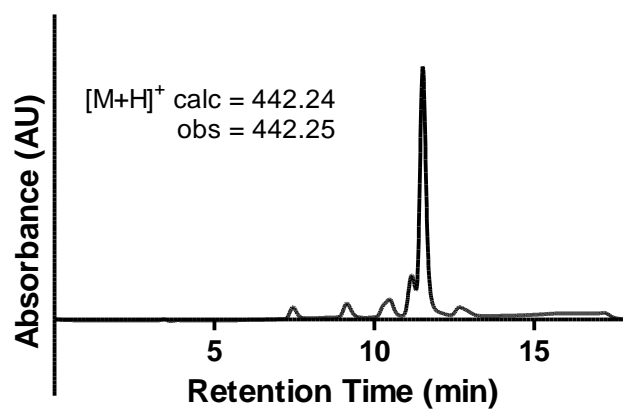


Figure 5.S5 LC-MS analysis of Boc-thiopropyl-glycyl-proline *tert*-butyl ester

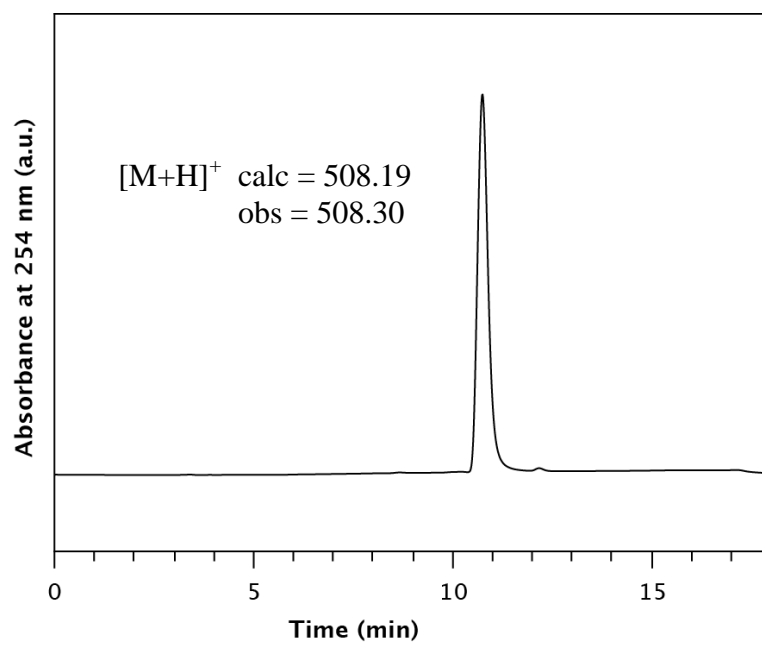


Figure 5.S6 LC-MS analysis of Fmoc-thiopropyl-glycyl-proline

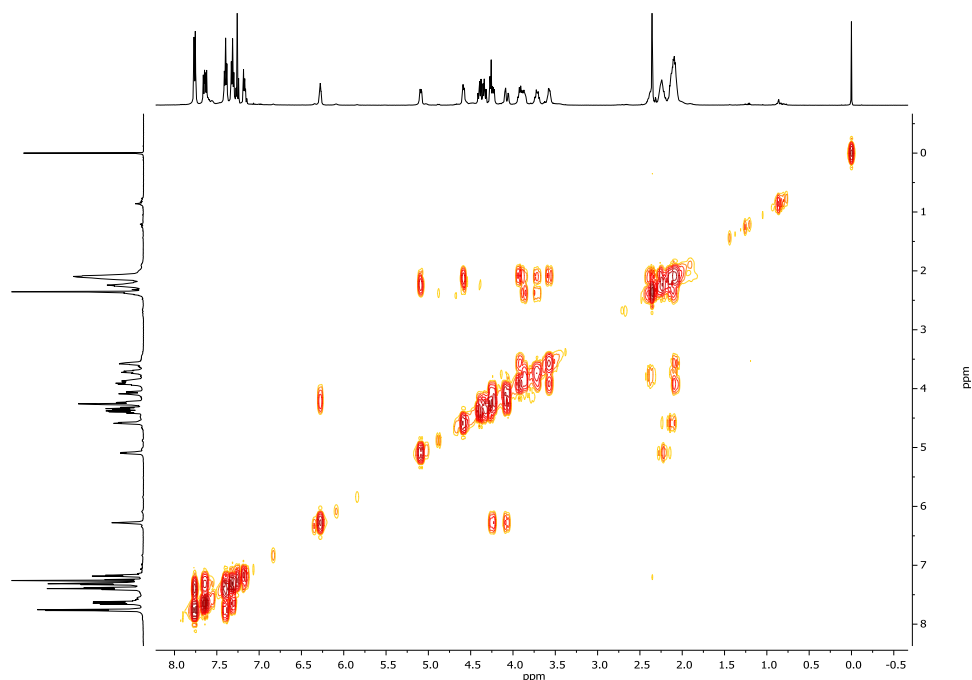


Figure 5.S7 ^1H , ^1H -COSY spectrum of Fmoc-thioglycyl-prolyl-proline in CDCl_3 at an ^1H frequency of 500 MHz. Coupling to the carbamate proton (6.3 ppm) identifies multiplets at 4.0–4.3 ppm as glycine methylene protons.

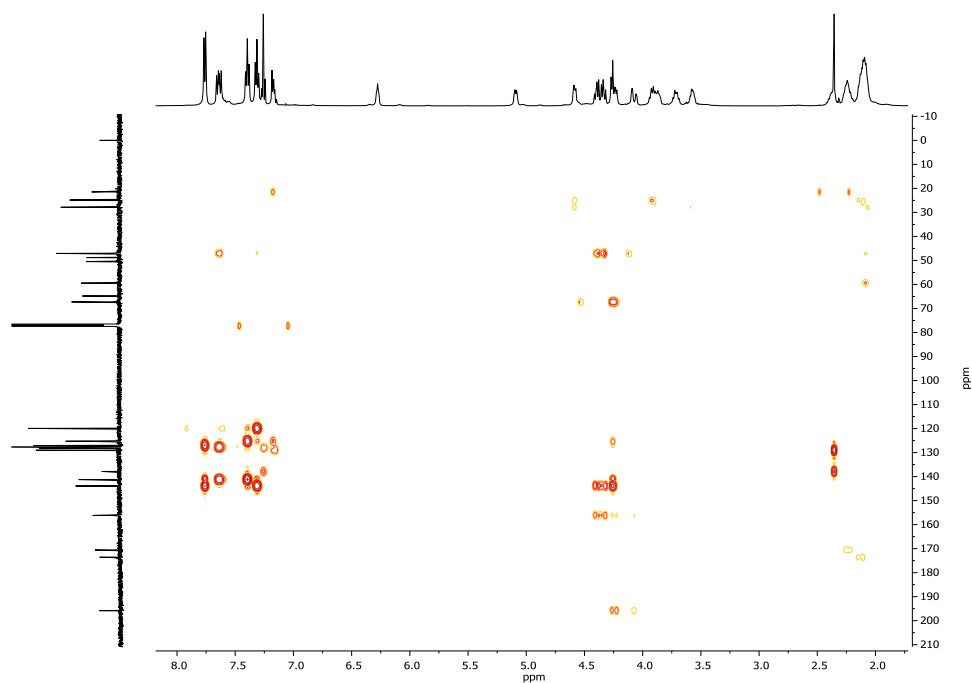


Figure 5.S8 ^1H , ^{13}C -HMBC spectrum of Fmoc-thioglycyl-prolyl-proline in CDCl_3 at an ^1H frequency of 500 MHz. Coupling of the thiocarbonyl carbon (195.8 ppm) to the glycine methylene protons (4.0–4.3 ppm) confirms thionation regiochemistry. Note also the absence of coupling of the thiocarbonyl carbon to the proline α -protons (5.0 and 4.5 ppm).

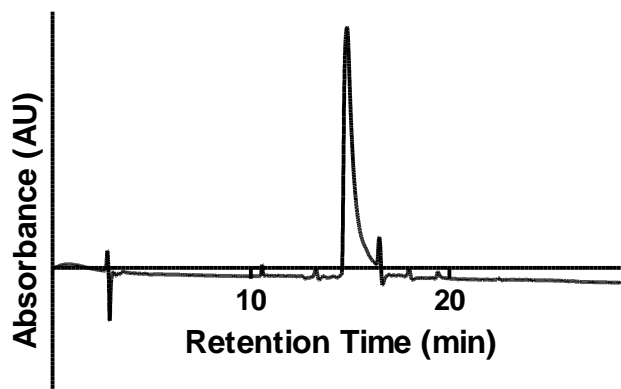


Figure 5.S9 Analytical HPLC trace of purified thiopeptide **Y**. MALDI-MS: $[M+H]^+$ calculated 2546.3, observed 2547.2.

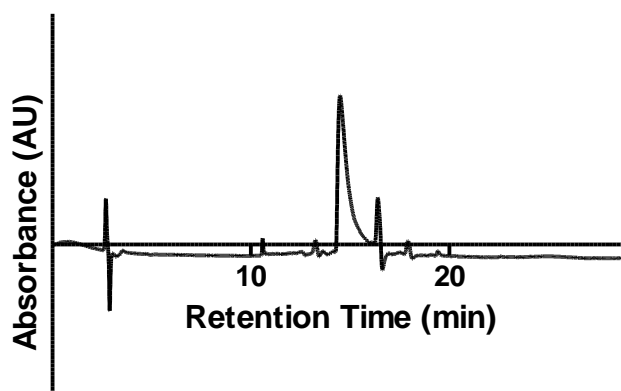


Figure 5.S10 Analytical HPLC trace of purified thiopeptide **G**. MALDI-MS: $[M+H]^+$ calculated 2546.3, observed 2546.7.

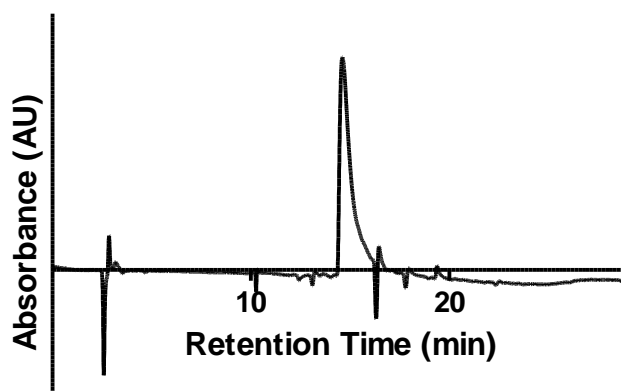


Figure 5.S11 Analytical HPLC trace of purified thiopeptide **G-N**. MALDI-MS: $[M+H]^+$ calculated 2546.3, observed 2546.9.

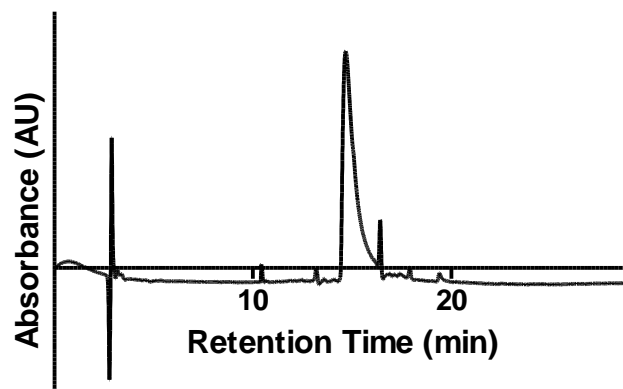
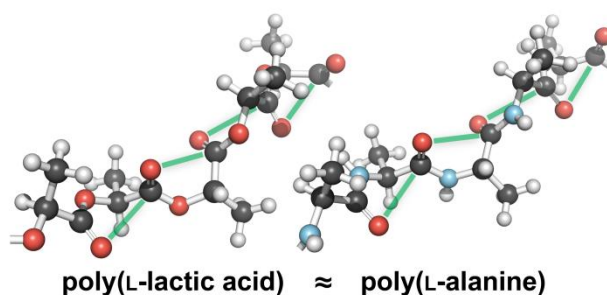


Figure 5.S12 Analytical HPLC trace of purified thiopeptide **G-C**. MALDI-MS: $[M+H]^+$ calculated 2546.3, observed 2547.0.

CHAPTER VI

 $n \rightarrow \pi^*$ Interactions in Poly(lactic acid) Suggest a Role in Protein Folding

Poly(lactic acid) (PLA) is a versatile synthetic polyester. We noted that this depsipeptide analog of polyalanine has a helical structure that resembles a polyproline II helix. Using natural bond orbital analysis, we find that $n \rightarrow \pi^*$ interactions between sequential ester carbonyl groups contribute 0.44 kcal/mol per monomer to the conformational stability of PLA helices. We conclude that analogous $n \rightarrow \pi^*$ interactions could direct the folding of a polypeptide chain into a polyproline II helix prior to the formation of any hydrogen bonds between backbone amides.

This chapter has been published, in part, under the same title: Newberry, R.W.; Raines, R. T. *Chem. Commun.* **2013**, 49, 7699-7701.

R.T.R. conceived the project. R.W.N. designed and performed the computational analysis and drafted the manuscript, which was edited by both authors.

Introduction

Polyesters have found widespread utility due to their inexpensive preparation and structural integrity.⁴⁰⁴ Poly(lactic acid) (PLA) is a recyclable polyester that can be prepared by condensing lactic acid, a renewable resource (Figure 6.1).⁴⁰⁵⁻⁴⁰⁷

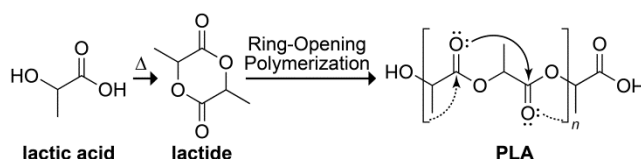


Figure 6.1 Synthesis of PLA.
Curved arrows indicate putative $n \rightarrow \pi^*$ interactions.

The thermal and structural properties of PLA can be adjusted by varying the ratio of L- and D-lactic acid monomers, or by altering the polymer processing conditions. These ensuing materials have received significant attention amongst macromolecular scientists, especially for use in biocompatible and biodegradable devices.^{405-407,408}

Interest in tuning properties of PLA has motivated the determination of its structure in atomic detail. These analyses have revealed the existence of conformational isomers, which arise from different preparation conditions.⁴⁰⁹⁻⁴¹⁰ The α form of PLA has received the most attention, due to its high stability. Fiber diffraction of α -PLA has proven challenging, and the ensuing structure has been revised several times since its first report.⁴¹¹⁻⁴¹⁵ Recently, data from neutron diffraction and NMR spectroscopy have been used to complement X-ray diffraction data.⁴¹⁶⁻⁴¹⁷ Though refinement of the structural model continues, the overall topology appears to be consistent between studies (Figure 6.2A).

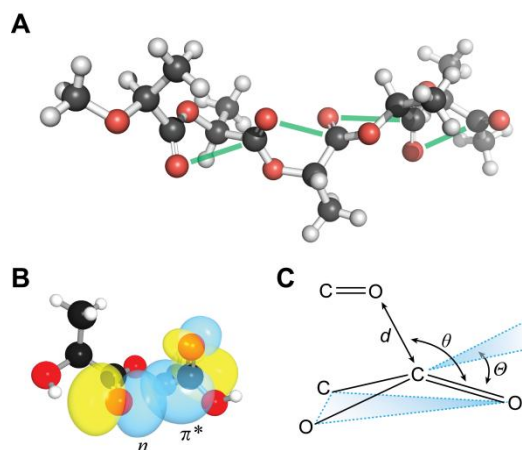


Figure 6.2 $n \rightarrow \pi^*$ Interactions in PLA.

(A) Five-residue segment of α -PLA from L-lactic acid (*i.e.*, isotactic PLLA).⁴¹⁵ Green lines show putative $n \rightarrow \pi^*$ interactions. (B) Overlap of the n and π^* orbitals in di(L-lactic acid) (CCDC Refcode: DUZMER). (C) Structural parameters describing the $n \rightarrow \pi^*$ interaction and the resulting pyramidalization of the acceptor carbonyl.

We became interested in PLA due to its similarity to polyalanine. Indeed, PLA is the depsipeptide counterpart to polyalanine wherein each amide linkage is replaced by an ester. The amide-to-ester modification has proven useful for revealing the contribution of hydrogen bonds to the structure and stability of peptides and proteins because incorporation of the ester linkage deletes a backbone hydrogen-bond donor and reduces the strength of a hydrogen bond with the carbonyl oxygen.^{171, 203, 418-421}

By examining the structure of a peptide-like polymer that is incapable of forming intramolecular hydrogen bonds, we sought to isolate other interactions that bias the conformation of peptide chains. In particular, we wished to determine the role of the $n \rightarrow \pi^*$ interaction in dictating the conformational geometry of PLA. In an $n \rightarrow \pi^*$ interaction, the filled p -type lone pair (n) of a carbonyl oxygen overlaps with the empty π^* antibonding orbital of a nearby carbonyl group (Figure 6.1 and Figure 6.2B). This overlap allows for orbital mixing and the subsequent release of energy. Such an interaction occurs when the donor oxygen contacts the acceptor carbonyl carbon within the sum of their van der Waals radii ($r_{\text{O}} + r_{\text{C}} = 3.22 \text{ \AA}$), and

along the Bürgi–Dunitz trajectory for nucleophilic addition ($\angle \text{O}\cdots\text{C}=\text{O} = \sim 109^\circ$).⁹⁴ We have estimated that $n \rightarrow \pi^*$ interactions between amides likely contribute 0.27 kcal/mol of stabilization energy per interaction,¹⁰⁰ and we have shown that these interactions are present in protein structures, especially helices.⁹² The question remains, however, does the $n \rightarrow \pi^*$ interaction bias a peptide toward a particular conformation, or is the interaction an artifact of a particular structural motif? By examining the structure of a polymer devoid of hydrogen bonds, we hoped to ascertain the relevance of the $n \rightarrow \pi^*$ interaction to macromolecular conformation.

Results

Upon inspection of the structure of α -PLA,⁴¹⁵ we observed that its backbone torsion angles (Table 6.1) bear striking similarity to those of the polyproline II helix, which has backbone torsion angles of ϕ ($\text{C}'_{i-1}-\text{N}_i-\text{C}^\alpha_i-\text{C}'_i$) = -75° and ψ ($\text{N}_i-\text{C}^\alpha_i-\text{C}'_i-\text{N}_i$) = 150° , and the strands of a collagen triple helix.¹⁸³ We showed previously that these torsion angles allow for effective $n \rightarrow \pi^*$ interactions.⁹² Indeed, the average $\text{O}\cdots\text{C}$ distance in the α -PLA structure is 2.98 Å, which is 0.24 Å less than the sum of the van der Waals radii; moreover, the average $\angle \text{O}\cdots\text{C}=\text{O}$ is 94° , which is consistent with an $n \rightarrow \pi^*$ interaction.⁹²

Table 6.1 Structural parameters and $n \rightarrow \pi^*$ energies of lactic acid polymers.

Polymer	ϕ ($^\circ$)	ψ ($^\circ$)	d (Å) ^d	θ ($^\circ$) ^d	$E_{n \rightarrow \pi^*}$ (kcal/mol) ^e	Θ ($^\circ$) ^d
α -PLA ^a	-63.7	154.4	2.98	93.6	0.44	ND
Di(L-lactic acid) ^b	-69.2	148.0	2.90	102.0	0.67	2.62
Tri(L-lactic acid) ^c	-69.2	163.5	2.98	92.7	0.41	3.40

^aValues are the mean over five consecutive L-lactic acid residues.⁴¹⁵ For an image, see: Figure 6.2A. ND, not determined. ^bValues are the mean from two molecules in the unit cell of CCDC Refcode DUZMER. For an image, see: Figure 6.2B. ^cValues are the mean from three residues in CCDC Refcode DUZMIV. ^dFor definitions, Figure 6.2C. ^eValues are from second-order perturbation theory.

To evaluate whether or not an $n \rightarrow \pi^*$ interaction is operative in the structure of α -PLA, we conducted natural bond orbital (NBO) analysis of its crystalline structure at the B3LYP/6-311+G(2d,p) level of theory.^{39, 370-371} We observed an average $n \rightarrow \pi^*$ energy of 0.44 kcal/mol per interaction. This value is consistent with a strong $n \rightarrow \pi^*$ interaction between the carbonyl groups of adjacent backbone esters in α -PLA.

To establish further the presence of an $n \rightarrow \pi^*$ interaction, we searched for a structural signature. As the $n \rightarrow \pi^*$ interaction populates the π^* orbital of the acceptor carbonyl, it induces a pyramidalization of the carbonyl group from planar sp^2 geometry (Figure 6.2C), which can be detected in high-resolution crystal structures.^{98, 100, 176-177} Unfortunately, the initial α -PLA structure-determination assumed planarity of the ester bond,⁴¹⁵ thereby obscuring the most definitive signature of an $n \rightarrow \pi^*$ interaction. Later structures do not provide enough resolution to determine pyramidalization accurately.⁴¹⁶ Accordingly, we sought high-resolution structures of lactic-acid oligomers.

We analyzed structures of di(L-lactic acid) and tri(L-lactic acid), which were obtained from the Cambridge Structural Database.²⁵⁰ To ensure that these short oligomers are appropriate models for the structure of α -PLA, we compared their backbone torsion angles to those observed in α -PLA and found gratifying agreement (Table 6.1). We also employed DFT calculations and NBO analysis to estimate the energy of the $n \rightarrow \pi^*$ interaction in these molecules and found that the $n \rightarrow \pi^*$ energies are consistent with those observed in the polymer. Confident that these structures are an accurate reflection of the structure of α -PLA, we then determined the distortion of the backbone esters from planarity, as measured by the angle Θ . In both structures, we observed substantial pyramidalization of the putative $n \rightarrow \pi^*$ acceptor toward the putative donor. In the absence of an attractive interaction, one would expect distortion to occur in the opposite

direction, so as to reduce unfavorable Pauli repulsion.¹⁸⁰ Accordingly, the observed pyramidalization is strong evidence of an attractive $n\rightarrow\pi^*$ interaction between the monomeric units in α -PLA.

Conclusions

These observations have broad implications. First, they imply a new means to modulate the structure of organic polymers. We found previously that the nucleophilicity of sulfur in thioamides can be exploited to increase the strength of an $n\rightarrow\pi^*$ interaction^{98, 100} and that surrogate alkenes and fluoroalkenes can be used to attenuate an $n\rightarrow\pi^*$ interaction.¹⁸⁰ These isosteres¹⁷¹ could be used to produce polymeric materials with tailored structural and thermal properties. Secondly, as the $n\rightarrow\pi^*$ interaction is likely to reduce the electrophilicity of the acceptor carbonyl by contributing additional electron density,^{175, 223} it could contribute to the observed hydrolytic stability of PLA. Thirdly, because $n\rightarrow\pi^*$ interactions are extant in PLA even without the potential for intramolecular hydrogen bonding, we conclude that the $n\rightarrow\pi^*$ interaction can operate independently of the geometric constraints imposed by hydrogen-bonding patterns. Finally, the observation of polyproline-like structure in PLA is itself significant, given that this structural motif is prevalent in the unfolded state of proteins.⁴²²⁻⁴²⁷ During their folding, polypeptide chains are likely to sample highly local interactions sooner than less local ones. Operating between adjacent residues (that is, $i\rightarrow i+1$), the $n\rightarrow\pi^*$ interaction is considerably more local than common hydrogen-bonding patterns such as that in the α -helix ($i\rightarrow i+4$). Thus, the presence of $n\rightarrow\pi^*$ interactions in the structure of PLA suggests that—before hydrogen bonds can form—the conformation of polypeptide chains can be guided by $n\rightarrow\pi^*$ interactions.⁴²⁸

Acknowledgements. We thank M. K. Mahanthappa and M. R. Aronoff for contributive discussions. R.W.N. was supported by NIH Biotechnology Training Grant T32 GM008349. This work was supported by grants R01 AR04476 (NIH) and CHE-1124944 (NSF). High-performance computing was supported by grant CHE-0840494 (NSF).

CHAPTER VII

Signatures of $n \rightarrow \pi^*$ Interactions in Proteins

The folding of proteins is directed by a variety of interactions, including hydrogen bonding, electrostatics, van der Waals' interactions, and the hydrophobic effect. We have argued previously that an $n \rightarrow \pi^*$ interaction between carbonyl groups be added to this list. In an $n \rightarrow \pi^*$ interaction, the lone pair (n) of one carbonyl oxygen overlaps with the π^* antibonding orbital of another carbonyl group. The tendency of backbone carbonyl groups in proteins to engage in this interaction has consequences for the structures of folded proteins that we unveil herein. First, we employ density functional theory to demonstrate that the $n \rightarrow \pi^*$ interaction causes the carbonyl carbon to deviate from planarity. Then, we detect this signature of the $n \rightarrow \pi^*$ interaction in high-resolution structures of proteins. Finally, we demonstrate through natural population analysis that the $n \rightarrow \pi^*$ interaction causes polarization of the electron density in carbonyl groups and detect that polarization in the electron density map of cholesterol oxidase, further validating the existence of $n \rightarrow \pi^*$ interactions. We conclude that the $n \rightarrow \pi^*$ interaction is operative in folded proteins.

This chapter has been published, in part, under the same title: Newberry, R.W.; Bartlett, G.J.; VanVeller, B.; Woolfson, D. N.; Raines, R. T. *Protein Sci.* **2014**, *23*, 284-288.

R.W.N. and B.V. conceived the project. R.W.N., G.J.B., and B.V. designed the analyses, with input from D.N.W. and R.T.R. R.W.N. performed the analyses and drafted the manuscript, which was edited by all authors.

Introduction

The three-dimensional structures of proteins enable their specific functions and arise largely from noncovalent interactions within and between polypeptide chains.³ These interactions include hydrogen bonding, electrostatics, van der Waals' interactions, and the hydrophobic effect.⁶ The current challenges in *de novo* structure prediction and protein design demonstrate that the understanding of these interactions is incomplete.⁵⁹⁻⁶¹

We have argued previously that an $n \rightarrow \pi^*$ interaction between two carbonyl groups can play a role in dictating protein conformation.^{90-91, 98, 180-181, 185} In an $n \rightarrow \pi^*$ interaction, the lone pair (n) of a carbonyl oxygen overlaps with the π^* antibonding orbital of another carbonyl group (Figure 7.1A).

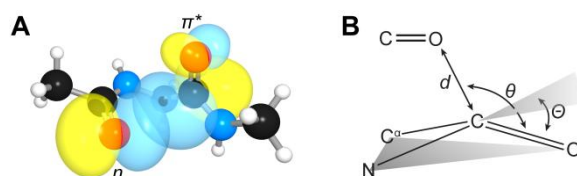


Figure 7.1 Example of an $n \rightarrow \pi^*$ interaction in the protein backbone and its structural signature. (A) Overlap of the n and π^* orbitals in the backbone of bitter melon trypsin inhibitor (PDB: 1vbw), residues 5–7. Image rendered with NBOView 1.1. (B) Geometric parameters that characterize an $n \rightarrow \pi^*$ interaction.

This orbital overlap is possible when the putative donor forms a sub-van der Waals' contact with the acceptor ($d < 3.22 \text{ \AA}$) along the Bürgi–Dunitz trajectory for nucleophilic addition ($\theta \approx 109^\circ$). The result of overlap between the n orbital of the donor and the π^* orbital of the acceptor is the release of energy due to orbital mixing. The energy associated with this interaction varies with the geometry of the interacting groups, but we anticipate a typical $n \rightarrow \pi^*$ interaction between amide bonds to contribute at least 0.27 kcal/mol.¹⁰⁰ We have shown that numerous residues in folded proteins are oriented to take advantage of this energy release, suggesting that $n \rightarrow \pi^*$

interactions could contribute significantly to the three-dimensional structure and conformational stability of proteins.^{92, 106, 201}

An $n \rightarrow \pi^*$ interaction results in population of the π^* orbital of the acceptor carbonyl. A distinctive signature should arise in the structures of proteins. Specifically, as population of the π^* orbital weakens the carbonyl π -bond, the acceptor should distort from a planar sp^2 geometry toward a pyramidal sp^3 -like geometry (Figure 7.1B). Such deviations can be represented by the angle θ at which the carbonyl group rises out of the plane of its substituents; as computed according to the method of Mazzarella and coworkers.⁴²⁹ Previously, we examined the pyramidalization of carbonyl groups in short, helical peptides from the Cambridge Structural Database.¹⁷⁷ We observed that $n \rightarrow \pi^*$ interactions engender pyramidalization of the carbonyl group toward its respective donor; helical residues not engaged in an $n \rightarrow \pi^*$ interaction distort away from the putative donor so as to reduce Pauli repulsion. We have also detected the distortion instilled by an $n \rightarrow \pi^*$ interaction in small molecules.^{98, 100, 176-177} Here, we ask the question: Is pyramidalization a signature in proteins as well?

Results

We began our investigation with a foundational computational analysis. To evaluate the propensity of an amide carbonyl group to pyramidalize in the presence of an $n \rightarrow \pi^*$ interaction, we employed density functional theory (DFT) at the B3LYP/6-311+G(2d,p) level of theory to optimize the structure of a model amide as a putative $n \rightarrow \pi^*$ donor approaches along the Bürgi–Dunitz trajectory.⁹⁴ To reduce the degrees of freedom in our analysis, we studied the structure of formamide as it was approached by formaldehyde along the Bürgi–Dunitz trajectory ($d = 2.75$ – 3.50 Å, $\theta = 110^\circ$); to simplify our analysis further, we restricted the geometry of the complex to

ensure parallel orientation of the carbonyl groups. Then, we plotted the observed pyramidalization as a function of the $n \rightarrow \pi^*$ interaction energy ($E_{n \rightarrow \pi^*}$) determined from second-order perturbation theory in the natural bond orbital analysis program NBO 5.9 (Figure 7.2A).³⁹ As expected, at high, stabilizing values of $E_{n \rightarrow \pi^*}$, which correspond to shorter donor–acceptor contacts, we observed greater pyramidalization. This correlation validates pyramidalization of backbone carbonyl carbon atoms as a signature of an $n \rightarrow \pi^*$ interaction that could enable its detection in proteins.

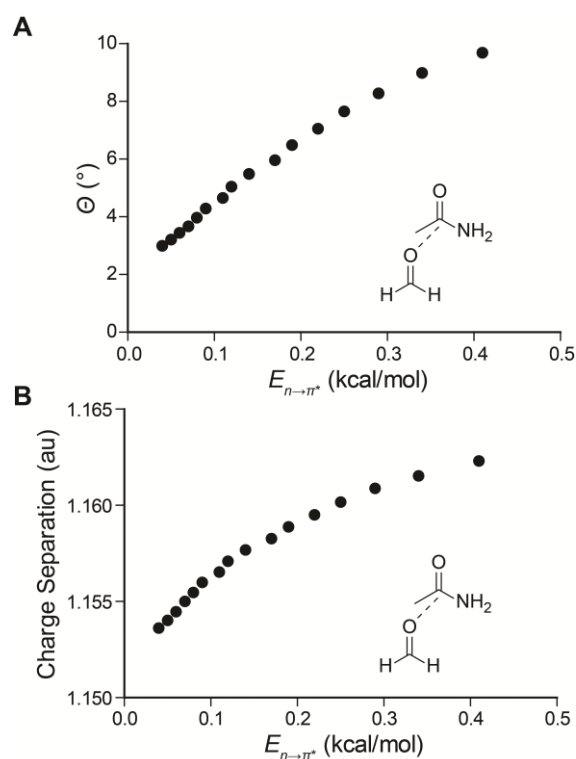


Figure 7.2 Calculated structural and electronic effects of $n \rightarrow \pi^*$ donation.

Calculated pyramidalization (A) and carbonyl charge separation (B) of formamide resulting from the approach of formaldehyde along the Bürgi–Dunitz trajectory. The value of $E_{n \rightarrow \pi^*}$ was determined by second-order perturbation theory and charge separation is determined from natural population analysis (NPA), both as implemented by NBO 5.9.

Structural refinement can bias the atomic coordinates of peptide bonds toward planarity. Hence, we examined a nonredundant set (<25% pairwise sequence identity) of 192 protein

crystal structures that were determined to sub-Å resolution to minimize the bias introduced by refinement.⁴³⁰⁻⁴³¹ We identified residues that receive a backbone $n \rightarrow \pi^*$ interaction using a geometric operational definition (Figure 7.1B). Specifically, when a carbonyl oxygen forms a sub-van der Waals' contact ($d < 3.22$ Å) with a carbonyl carbon along the Bürgi–Dunitz trajectory ($95^\circ < \theta < 125^\circ$), we scored the carbonyl carbon as positive for receiving an $n \rightarrow \pi^*$ interaction; those residues not meeting this criteria were scored as negative for receiving an $n \rightarrow \pi^*$ interaction.

We sought to control for the effect of secondary structure in our analysis of carbonyl pyramidalization, recognizing that local conformation could contribute to pyramidalization,^{429, 432} In particular, the prevalence of $n \rightarrow \pi^*$ interactions appears to vary dramatically between different secondary structures, being observed with great frequency in α -helices but rarely in β -sheets.⁹² To remove bias from local conformation, we examined the pyramidalization of the 3,759 carbonyl groups from our high-resolution set of protein structures that were not assigned to any particular secondary structure by Kabsch and Sander criteria⁴⁸ as implemented by PROMOTIF.⁴³³ Of these residues, 24% receive an $n \rightarrow \pi^*$ interaction, a fraction that enables us to make an effective comparison between the geometries of residues that receive an $n \rightarrow \pi^*$ interaction and those that do not.

We were able to observe a difference in the absolute pyramidalization of these two populations (Figure 7.3A). The absolute pyramidalization was found to be 0.32° higher on average in the presence of an $n \rightarrow \pi^*$ interaction, indicating that the two populations are distorted differently ($p < 0.00001$). Thus, we conclude that $n \rightarrow \pi^*$ interactions cause a distortion of the peptide bond in proteins. Moreover, for those carbonyl groups that accept an $n \rightarrow \pi^*$ interaction, we observed that pyramidalization tends to occur toward the donor (Figure 7.3B). In the absence

of an attractive interaction, one would expect the carbonyl group to distort away from the incoming oxygen so as to reduce unfavorable van der Waals' contacts and Pauli repulsion.¹⁸⁰ We note that, even in these high-resolution structures, structural refinement likely enforces planarity upon the peptide bond; the true effect of the $n \rightarrow \pi^*$ interaction on pyramidalization of protein carbonyl groups could be even greater than that observed herein.

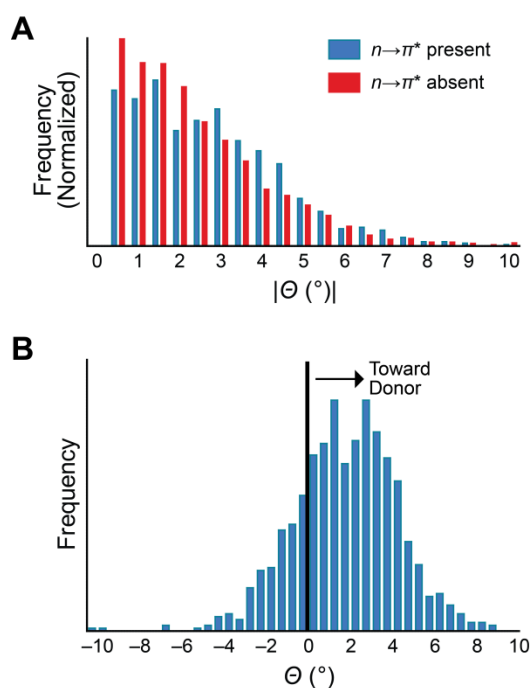


Figure 7.3 Pyramidalization of backbone carbonyl groups in proteins.

(A) Absolute pyramidalization of residues receiving an $n \rightarrow \pi^*$ interaction (blue, $n = 896$) versus those that do not (red, $n = 2863$). (B) Pyramidalization of residues receiving an $n \rightarrow \pi^*$ interaction measured relative to the location of the $n \rightarrow \pi^*$ donor, with $\Theta > 0$ indicating pyramidalization toward the donor.

As the $n \rightarrow \pi^*$ interaction involves donation of electron density to a carbonyl carbon, another signature of the $n \rightarrow \pi^*$ interaction should be evident. Specifically, the approach of the nucleophilic $n \rightarrow \pi^*$ donor oxygen should polarize the electron density of the acceptor amide carbonyl. To evaluate this hypothesis, we subjected formamide to natural population analysis (NPA) upon approach by the formaldehyde $n \rightarrow \pi^*$ donor.⁴³⁴ NPA is a reliable method for

assigning atomic charges and thus provides a measure of charge separation in the acceptor carbonyl group, which we plotted against the corresponding $E_{n\rightarrow\pi^*}$ (Figure 7.2B). As expected, the degree of charge separation in the carbonyl group increases as the $n\rightarrow\pi^*$ interaction grows stronger.

Finally, we searched for evidence of this signature in the electron density of a particular protein. Our analysis is based on a previous report of the electron density in carbonyl groups of cholesterol oxidase. In that work, Lario and Vrielink classified backbone carbonyl groups by the degree of charge separation observed between the carbonyl carbon and carbonyl oxygen in the electron density map (Figure 7.4A).⁴³⁵

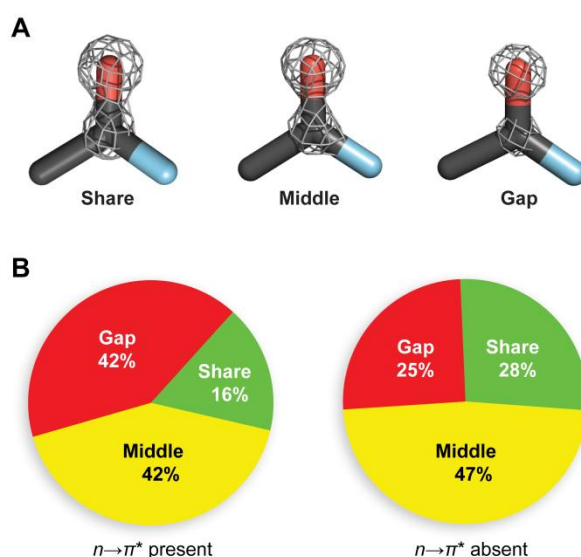


Figure 7.4 Electronic polarization of backbone carbonyl groups in proteins.

(A) Prototypical examples of “gap”, “middle”, and “share” electronic distributions from residues Leu155, Val108, and Leu287 of cholesterol oxidase (PDB: 1n1p), respectively. (B) Relative proportion of “gap”, “middle”, and “share” electronic distributions for residues in cholesterol oxidase that receive an $n\rightarrow\pi^*$ interaction ($n = 50$) and those that do not ($n = 97$).⁴³⁵

Here, if a clear separation in electron density is visible at 4.5σ , the carbonyl group is classified as a “gap”. In contrast, if no separation is evident at 5.5σ , the carbonyl group is classified as a “share”. Carbonyl groups for which separation of electron density was evident at 5.5σ but not

4.5 σ were classified as “middle.” To control for the influence of secondary structure, we exclude residues assigned to helices or sheets by PROMOTIF. We then tabulated the relative abundance of each of these categories for residues in cholesterol oxidase that receive an $n\rightarrow\pi^*$ interaction versus those that do not, according to our geometric criteria (Figure 7.4B). We observe that for carbonyl groups that receive an $n\rightarrow\pi^*$ interaction, there is a lower proportion of “share” carbonyl groups and a higher proportion of “gap” carbonyl groups, compared to those residues that do not receive an $n\rightarrow\pi^*$ interaction. The greater proportion of “gap” carbonyl groups in the $n\rightarrow\pi^*$ positive set is consistent with the notion that the $n\rightarrow\pi^*$ interaction causes a polarization of the electron density of carbonyl groups.

Conclusions

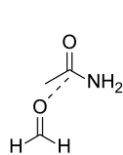
We note two important implications of these observations. First, the polarization of electron density by the $n\rightarrow\pi^*$ interaction could serve to increase the strength of hydrogen bonds that are critical for dictating secondary structure. Secondly, because polarization increases the nucleophilicity of a carbonyl oxygen, the polarization of carbonyl groups caused by $n\rightarrow\pi^*$ donation should increase the strength of subsequent $n\rightarrow\pi^*$ interactions, leading to cooperativity. These effects on carbonyl polarization could conspire to increase the thermal stability of the α -helix, which is stitched together by the action of both hydrogen bonding and $n\rightarrow\pi^*$ interactions.^{92, 102} The impact of $n\rightarrow\pi^*$ interactions on the structure and the stability of proteins should thus extend beyond the direct effect of carbonyl attraction. Considering their sheer abundance, the impact of $n\rightarrow\pi^*$ interactions on protein folding could be profound.

Acknowledgements. R.W.N. was supported by NIH Biotechnology Training Grant T32 GM008349. This work was supported by grants R01 AR04476 (NIH), CHE-1124944 (NSF), and EP/J001430 (EPSRC). High-performance computing was supported by grant CHE-0840494 (NSF).

Methods

Computational Chemistry. Geometry optimizations of the formaldehyde–formamide complex were conducted at the B3LYP/6-311+G(2d,p) level of theory as implemented in Gaussian 09.³⁷⁰ The distance between the formaldehyde oxygen and the formamide carbon was varied from 2.75 Å to 3.50 Å with the angle of approach of the formaldehyde oxygen to the formamide carbonyl constrained to 110° and the dihedral angle between the two carbonyl groups constrained to 180°. Optimized geometries were then subjected to NBO analysis at the B3LYP/6-311+G(2d,p) as implemented in NBO 5.9.³⁷¹

Bioinformatics. A nonredundant set (<25% pairwise sequence identity) of 192 protein crystal structures (>40 residues, *R* <20%) with resolution of 1.0 Å or better was culled from the PDB on 28 November 2012 using the PISCES server.⁴³¹ Secondary structure assignments were made using Kabsch and Sander criteria as implemented in PROMOTIF.⁴³³ Residues modeled in multiple conformations of the peptide backbone were excluded from analysis.



Parameter	Value
O _{aldehyde} -C _{amide} distance (<i>d</i>)	2.5–3.5 Å
O _{aldehyde} -C _{amide} -O _{amide} angle	110°
C _{aldehyde} -O _{aldehyde} -C _{amide} -O _{amide} dihedral	180°
H _{aldehyde} -C _{aldehyde} -O _{aldehyde} -C _{amide} dihedral	180°

Figure S1 Optimization parameters of formamide as approached by formaldehyde (**1**).

Table 7.S1 Cartesian coordinates of 1 at <i>d</i> = 2.75 Å				Table 7.S2 Cartesian coordinates of 1 at <i>d</i> = 2.80 Å			
C	2.21955600	0.13416200	0.12859200	C	2.24354900	0.13741800	0.13139500
O	1.46735700	-0.70820000	-0.27323400	O	1.49615000	-0.70793200	-0.27340800
H	3.31315700	0.06878400	-0.02535400	H	3.33797500	0.07643000	-0.01847800
H	1.86483300	1.02939700	0.67981700	H	1.88243600	1.03092800	0.68131200
C	-1.12407700	-0.05808800	0.37817000	C	-1.14749300	-0.05652500	0.37991000
N	-0.89655000	1.16483000	-0.18088700	N	-0.90929300	1.16184800	-0.18381700
H	-0.68059200	-0.16989000	1.38216400	H	-0.70718100	-0.16613100	1.38568000
H	-1.23029300	1.32754000	-1.11901700	H	-1.23337100	1.32152200	-1.12587000
H	-0.21152900	1.79416100	0.20047400	H	-0.22213300	1.78762700	0.19970700
O	-1.88643200	-0.87433000	-0.08832300	O	-1.90477700	-0.87565100	-0.08952500

Table 7.S3 Cartesian coordinates of 1 at <i>d</i> = 2.85 Å				Table 7.S4 Cartesian coordinates of 1 at <i>d</i> = 2.90 Å			
C	2.26746800	0.14030900	0.13392600	C	2.29195500	0.14384600	0.13652700
O	1.52494500	-0.70758700	-0.27352900	O	1.55407200	-0.70713100	-0.27354600
H	3.36314500	0.08434200	-0.01108400	H	3.38821500	0.09167500	-0.00495400
H	1.90117000	1.03317900	0.68296300	H	1.92011500	1.03498900	0.68391500
C	-1.17079300	-0.05470600	0.38163700	C	-1.19389700	-0.05337200	0.38316000
N	-0.92187900	1.15855300	-0.18669100	N	-0.93597200	1.15574500	-0.18933700
H	-0.73377200	-0.16324500	1.38924400	H	-0.75990500	-0.15947100	1.39241400
H	-1.23712600	1.31432200	-1.13245000	H	-1.24114700	1.30741800	-1.13910600
H	-0.23328800	1.78223700	0.19808500	H	-0.24464700	1.77559600	0.19661600
O	-1.92332300	-0.87670300	-0.09063300	O	-1.94146900	-0.87827800	-0.09165900

Table 7.S5 Cartesian coordinates of 1 at <i>d</i> = 2.95 Å				Table 7.S6 Cartesian coordinates of 1 at <i>d</i> = 3.00 Å			
C	2.31651600	0.14736500	0.13890600	C	2.34130300	0.15095100	0.14113000
O	1.58330700	-0.70669200	-0.27341300	O	1.61274100	-0.70616500	-0.27322100
H	3.41343200	0.09908700	0.00113000	H	3.43888300	0.10668800	0.00708000
H	1.93903200	1.03710700	0.68448100	H	1.95830900	1.03943900	0.68485300
C	-1.21691800	-0.05212700	0.38445000	C	-1.23973600	-0.05099000	0.38565900
N	-0.95003600	1.15295000	-0.19204000	N	-0.96469400	1.15018100	-0.19466600
H	-0.78593900	-0.15582100	1.39544700	H	-0.81170100	-0.15219400	1.39832900
H	-1.24634600	1.30120300	-1.14516400	H	-1.25240500	1.29476900	-1.15100000
H	-0.25751000	1.77010900	0.19620400	H	-0.27125500	1.76506200	0.19569200
O	-1.95955900	-0.88002900	-0.09258200	O	-1.97753700	-0.88193500	-0.09340700

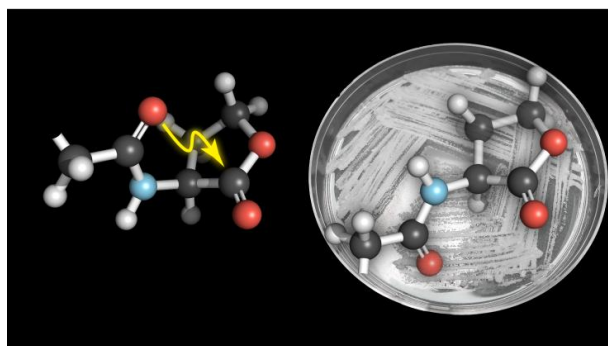
Table 7.S7 Cartesian coordinates of 1 at <i>d</i> = 3.05 Å				Table 7.S8 Cartesian coordinates of 1 at <i>d</i> = 3.10 Å			
C	2.36635400	0.15466200	0.14320700	C	2.39167000	0.15850100	0.14514500
O	1.64241400	-0.70552900	-0.27299100	O	1.67232700	-0.70479100	-0.27272900
H	3.46457100	0.11443200	0.01281200	H	3.49048200	0.12227900	0.01829500
H	1.97796900	1.04196200	0.68501500	H	1.99798100	1.04464000	0.68496800
C	-1.26232700	-0.05003200	0.38678500	C	-1.28469300	-0.04927300	0.38782300
N	-0.98003100	1.14746000	-0.19719500	N	-0.99604200	1.14481800	-0.19961500
H	-0.83715800	-0.14864700	1.40106400	H	-0.86231400	-0.14519900	1.40365200
H	-1.25982300	1.28836600	-1.15644400	H	-1.26866300	1.28210100	-1.16147200
H	-0.28610300	1.76065500	0.19505500	H	-0.30202200	1.75689800	0.19431700
O	-1.99533900	-0.88406800	-0.09414500	O	-2.01295500	-0.88643500	-0.09480400

Table 7.S9 Cartesian coordinates of 1 at $d = 3.15$ Å				Table 7.S10 Cartesian coordinates of 1 at $d = 3.20$ Å			
C	2.41722800	0.16243100	0.14695600	C	2.44300400	0.16641800	0.14865100
O	1.70244400	-0.70396800	-0.27244200	O	1.73273100	-0.70308100	-0.27212900
H	3.51659000	0.13018000	0.02354100	H	3.54287500	0.13809000	0.02856700
H	2.01831900	1.04743000	0.68474000	H	2.03896600	1.05030300	0.68434400
C	-1.30685800	-0.04869400	0.38878100	C	-1.32884700	-0.04827100	0.38965800
N	-1.01267200	1.14228000	-0.20191800	N	-1.02983500	1.13985400	-0.20410900
H	-0.88719800	-0.14185100	1.40610200	H	-0.91184200	-0.13860600	1.40842100
H	-1.27869200	1.27595900	-1.16615100	H	-1.28970600	1.26996000	-1.17052900
H	-0.31877200	1.75363300	0.19346400	H	-0.33620400	1.75076200	0.19254500
O	-2.03041500	-0.88899800	-0.09539500	O	-2.04775500	-0.89171600	-0.09592500
Table 7.S11 Cartesian coordinates of 1 at $d = 3.25$ Å				Table 7.S12 Cartesian coordinates of 1 at $d = 3.30$ Å			
C	2.46898000	0.17044000	0.15023500	C	2.49525800	0.17462600	0.15166200
O	1.76316200	-0.70214400	-0.27179100	O	1.79390400	-0.70110500	-0.27141700
H	3.56932200	0.14598200	0.03338400	H	3.59603600	0.15406900	0.03791900
H	2.05991000	1.05324000	0.68379100	H	2.08120000	1.05637100	0.68297400
C	-1.35067900	-0.04798400	0.39045400	C	-1.37223200	-0.04796300	0.39114200
N	-1.04746500	1.13753900	-0.20619300	N	-1.06588600	1.13531100	-0.20818800
H	-0.93627200	-0.13546500	1.41061200	H	-0.96035000	-0.13241500	1.41266500
H	-1.30157800	1.26413400	-1.17463100	H	-1.31523000	1.25874800	-1.17828700
H	-0.35423400	1.74823100	0.19160800	H	-0.37412800	1.74685600	0.19094000
O	-2.06500000	-0.89455900	-0.09640100	O	-2.08196400	-0.89774300	-0.09679800
Table 7.S13 Cartesian coordinates of 1 at $d = 3.35$ Å				Table 7.S14 Cartesian coordinates of 1 at $d = 3.40$ Å			
C	2.52177300	0.17881300	0.15300300	C	2.54836000	0.18302700	0.15418300
O	1.82478600	-0.69997800	-0.27112200	O	1.85576600	-0.69894600	-0.27058600
H	3.62294100	0.16205500	0.04218700	H	3.64990800	0.17005200	0.04643500
H	2.10302300	1.05956000	0.68223200	H	2.12460600	1.06283700	0.68087600
C	-1.39359200	-0.04808200	0.39188000	C	-1.41490600	-0.04829000	0.39224700
N	-1.08442600	1.13292300	-0.21017600	N	-1.10388300	1.13119600	-0.21187800
H	-0.98422100	-0.12974800	1.41456500	H	-1.00798200	-0.12678900	1.41636900
H	-1.33154000	1.25435700	-1.18109900	H	-1.34495400	1.24903200	-1.18479100
H	-0.39556700	1.74691000	0.19031200	H	-0.41594100	1.74538300	0.19005300
O	-2.09887800	-0.90101900	-0.09716000	O	-2.11566300	-0.90446800	-0.09746200
Table 7.S15 Cartesian coordinates of 1 at $d = 3.45$ Å				Table 7.S16 Cartesian coordinates of 1 at $d = 3.50$ Å			
C	2.57520300	0.18724500	0.15530800	C	2.60220200	0.19143700	0.15635100
O	1.88689300	-0.69778800	-0.27017100	O	1.91808100	-0.69661500	-0.26975500
H	3.67708200	0.17800500	0.05036100	H	3.70439800	0.18580800	0.05418000
H	2.14684400	1.06611200	0.67969200	H	2.16931200	1.06935700	0.67840600
C	-1.43600500	-0.04865300	0.39275200	C	-1.45700000	-0.04906300	0.39319700
N	-1.12331800	1.12909400	-0.21370000	N	-1.14292800	1.12699300	-0.21543200
H	-1.03146500	-0.12404500	1.41803600	H	-1.05479700	-0.12155700	1.41969300
H	-1.36167000	1.24481300	-1.18751600	H	-1.37911700	1.24084500	-1.18998700
H	-0.43887300	1.74602900	0.19002800	H	-0.46247700	1.74719500	0.18995600
O	-2.13237800	-0.90797800	-0.09771000	O	-2.14908600	-0.91149100	-0.09793300

Table 7.S17 Energies of **1** optimized at various values of d .

d (Å)	Energy (Hartree)
2.75	-284.4950067740
2.80	-284.4957078650
2.85	-284.4963193460
2.90	-284.4968531970
2.95	-284.4973185180
3.00	-284.4977250820
3.05	-284.4980812470
3.10	-284.4983937100
3.15	-284.4986676930
3.20	-284.4989073750
3.25	-284.4991165440
3.30	-284.4992991080
3.35	-284.4994588180
3.40	-284.4995995940
3.45	-284.4997246160
3.50	-284.4998364470

CHAPTER VIII

A Key $n \rightarrow \pi^*$ Interaction in *N*-Acyl Homoserine Lactones

Many Gram-negative bacteria employ *N*-acyl homoserine lactones (AHLs) as signal molecules for quorum sensing. The binding of AHLs to their target LuxR-type receptor proteins can effect changes in growth, virulence, and other phenotypes. Here, we present X-ray crystallographic and computational evidence that the conformation of free AHLs is biased away from the conformation observed when bound to their cognate receptor due to the influence of an $n \rightarrow \pi^*$ interaction. In this $n \rightarrow \pi^*$ interaction, the p -type lone pair (n) of the *N*-acyl oxygen overlaps with the π^* orbital of the lactone carbonyl group, thereby releasing approximately 0.64 kcal/mol of energy. We also show that this interaction can be attenuated by installing electron-withdrawing groups on the *N*-acyl chain. Modulating this previously unappreciated interaction could present a new avenue towards effective inhibitors of bacterial quorum sensing.

This chapter has been published, in part, under the same title: Newberry, R.W.; Raines, R. T. *ACS Chem. Biol.* **2014**, 9, 880-883.

R.T.R. conceived the project. R.W.N. designed, synthesized, and analyzed the compounds and drafted the manuscript, which was edited by both authors.

Introduction

Despite being unicellular organisms, bacteria have evolved mechanisms of chemical communication that regulate various physiological processes in response to cell density, a phenomenon known as quorum sensing.⁴³⁶⁻⁴³⁷ This process is receiving much attention because of its influence on biofilm formation and virulence. The principle mediators of these communication events in Gram-negative bacteria are the *N*-acyl homoserine lactones (AHLs, Figure 8.1).⁴³⁸⁻⁴³⁹

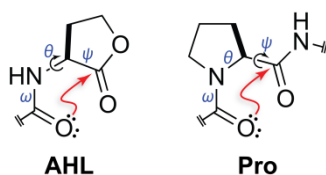


Figure 8.1 Structure of an *N*-acyl homoserine lactone and a proline residue. Arrows indicate freely rotatable bonds (black) and putative $n \rightarrow \pi^*$ interactions that would constrain those bonds (red).

AHLs act by binding to intracellular LuxR-type receptors, which are then activated as transcription factors.⁴⁴⁰⁻⁴⁴¹ Physiological transitions in bacterial colonies are induced when AHL concentrations cross particular thresholds. The strong influence of AHLs on bacterial behavior has attracted the attention of chemical biologists, as modulators of AHL-binding could impart exquisite control of bacterial pathogenesis.⁴⁴² Towards this end, we noticed that AHLs have proximal carbonyl groups (Figure 8.1). We have shown that the conformations of molecules with proximal carbonyl moieties can be influenced by an $n \rightarrow \pi^*$ interaction (Figure 8.2).^{90, 98} Here, we sought to determine if an $n \rightarrow \pi^*$ interaction could influence the conformation of an AHL.

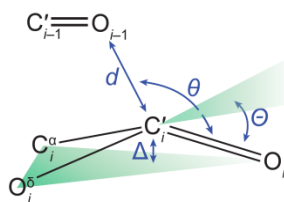


Figure 8.2 Structural parameters characterizing $n \rightarrow \pi^*$ interactions in AHLs.

In an $n \rightarrow \pi^*$ interaction, the filled lone pair (n) of one carbonyl group interpenetrates the empty π^* orbital of another. The mixing of these orbitals releases energy, thereby causing attraction between the two groups. This overlap is most effective when the oxygen of the electron-pair donor forms a sub-van der Waals contact ($d < 3.22 \text{ \AA}$) with the carbon of the acceptor carbonyl group along the Bürgi–Dunitz trajectory for nucleophilic addition ($95^\circ < \theta < 125^\circ$).⁹⁴ We have estimated that such an interaction between adjacent amides in a polypeptide contributes 0.27 kcal/mol of stabilizing energy per occurrence.¹⁰⁰ As these are relatively weak interactions, their influence is often only realized in systems in which carbonyl groups are in close proximity, as they are in proteins,^{92, 103, 106, 201} peptides,¹⁰² peptoids,^{208-209, 211, 443} polyesters,¹⁷⁹ and some small molecules.^{99, 102}

The preorganization of two carbonyl groups due to the constraint of an intervening ring can enhance an $n \rightarrow \pi^*$ interaction.^{90, 92, 98-99, 102, 106, 172-173, 177, 180, 196, 214, 220-221, 223, 374} We realized that the γ -lactone of an AHL restricts its ψ dihedral angle ($\text{N}_i\text{-C}_i^\alpha\text{-C}'_i\text{-N}_{i+1}$) and that amidic resonance restricts its ω dihedral angle ($\text{C}_{i-1}^\alpha\text{-C}'_{i-1}\text{-N}_i\text{-C}_i^\alpha$),⁴⁴⁴⁻⁴⁴⁵ leaving only a single unconstrained bond between the two carbonyl groups. In this sense, an AHL is analogous to a proline residue, which has a restricted θ dihedral angle ($\text{C}'_{i-1}\text{-N}_i\text{-C}_i^\alpha\text{-C}'_i$) and has a strong tendency to form an $\text{O}_{i-1} \cdots \text{C}'_i=\text{O}_i$ $n \rightarrow \pi^*$ interaction (Figure 8.1). Thus, we suspected that AHLs, like proline residues, could be predisposed to form an $n \rightarrow \pi^*$ interaction.

Results

To begin, we sought evidence for a putative $n \rightarrow \pi^*$ interaction by using X-ray diffraction analysis. Natural AHLs can have long alkyl chains that are resistant to crystallization, so we began by examining *N*-acetyl homoserine lactone. When this molecule did not afford crystals of sufficient quality for diffraction, we appended minimal precipitating groups to the *N*-acyl chain in attempts to drive crystal formation. Gratifyingly, we found the *N*-trimethylacetyl appendage to be suitable, affording crystals that diffracted to high resolution (Figure 8.3A).

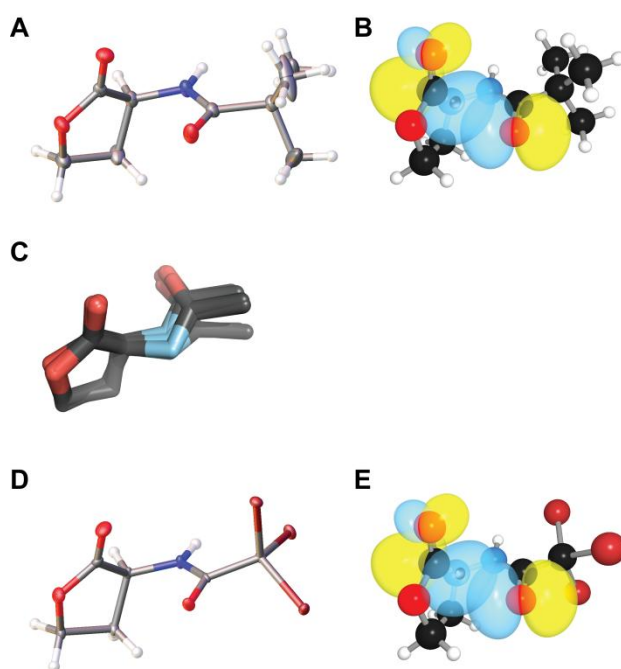


Figure 8.3 Structures of AHLs.

(A) Crystal structure of *N*-trimethylacetyl homoserine lactone drawn with 50% probability ellipsoids. (B) Rendering of the n and π^* orbitals of *N*-trimethylacetyl homoserine lactone in its optimized geometry. (C) Overlap of the crystal structures of the acetyl homoserine lactone moieties in the 14 ligands listed in Table 2. (D) Crystal structure of *N*-tribromoacetyl homoserine lactone drawn with 50% probability ellipsoids. (E) Rendering of the n and π^* orbitals of *N*-tribromoacetyl homoserine lactone in its optimized geometry.

In this molecular structure, we noted the presence of a short contact between the *N*-acyl carbonyl oxygen and the lactone carbonyl carbon ($d = 2.73 \text{ \AA}$), which is 15% below the sum of the van der Waals radii (Table 8.1).

Table 8.1 Conformational parameters of *N*-acyl homoserine lactones.

<i>N</i> -Acyl group	ϕ ($^\circ$) ^a	ψ ($^\circ$) ^a	d (\AA) ^a	θ ($^\circ$) ^a	Θ ($^\circ$) ^a	Δ (\AA) ^a	$E_{n \rightarrow \pi^*}$ (kcal/mol) ^b
<i>N</i> -trimethylacetyl	51.51(18)	-144.65(15)	2.732(2)	90.62(12)	2.7(2)	0.023(2)	0.64
<i>N</i> -tribromoacetyl	48.50(30)	-148.20(30)	2.801(4)	88.00(20)	1.6(5)	0.014(4)	0.55

^aFrom X-ray diffraction analysis of the crystalline compound. Structural parameters are defined in Figures 8.1 and 8.2. ^bFrom second-order perturbation theory.

We also found the angle of approach of the donor oxygen to the acceptor carbonyl ($\theta = 90.6^\circ$) to be consistent with an $n \rightarrow \pi^*$ interaction.⁹² As these structural features were strongly suggestive of an $n \rightarrow \pi^*$ interaction, we were motivated to search for characteristic structural deviations that result from $n \rightarrow \pi^*$ donation. In particular, we have shown that the presence of $n \rightarrow \pi^*$ interaction engenders pyramidalization of the acceptor carbonyl carbon toward the donor oxygen, as measured by the distortion parameter Θ .^{92, 98, 102-103, 106, 176-177} In *N*-trimethylacetyl homoserine lactone, we observed substantial pyramidalization ($\Theta = 2.7^\circ$) of the acceptor carbonyl toward the donor, in accord with that observed for other molecules with confirmed $n \rightarrow \pi^*$ interactions. Distortion of the carbonyl carbon toward the $n \rightarrow \pi^*$ is strong evidence of an attractive interaction; otherwise, distortion would likely occur away from the short contact so as to reduce unfavorable Pauli repulsion.¹⁸⁰

Confident that AHLs have a preference for forming an $n \rightarrow \pi^*$ interaction, we wished to evaluate the energy of this interaction. To do so, we performed natural bond orbital (NBO) analysis of the *N*-trimethylacetyl homoserine lactone structure optimized by density functional theory (DFT) calculations *in vacuo* at the B3LYP/6-311+G(2d,p) level of theory.³⁹ Using

second-order perturbation theory, as implemented by NBO 5.9,³⁷⁰⁻³⁷¹ we observed significant overlap of the n and π^* orbitals (Figure 8.3B), with an estimated energy of $E_{n \rightarrow \pi^*} = 0.64$ kcal/mol. This value is larger than that observed with a proline residue,¹⁰⁰ consistent with the carbonyl group of an ester being a better acceptor than that of an amide (Figure 8.1).

As the biological activity of an AHL relies upon its binding to its target LuxR-type receptor, we sought to compare the structure of a free AHL to that observed in a receptor·AHL complex. The Protein Data Bank (PDB)⁴³⁰ currently houses the atomic coordinates of ten LuxR-type receptor structures with bound AHLs, reflecting four distinct receptors as well as four structures of two different AHL-lactonases with bound AHLs. Remarkably, the conformation of the bound AHL ligand is nearly identical in all of these complexes (Table 8.2; Figure 8.3C), and that conformation differs dramatically from the conformation in the unbound state.

Table 8.2 Conformational parameters of protein-bound <i>N</i> -acyl homoserine lactones					
PDB entry	Resolution (Å)	Protein	<i>N</i> -Acyl group	ϕ (°)	ψ (°)
4g8b ^a	1.30	AidH	butyryl	−110	−136
3qp1	1.55	CviR	hexanoyl	−103	−139
3qp2	1.64	CviR	octanoyl	−108	−138
3qp4	1.55	CviR	decanoyl	−106	−139
3qp6	2.00	CviR	hexanoyl	−102	−140
3qp8 ^b	1.60	CviR	decanoyl	−108	−139
3ojg	1.60	GKL	butyryl	−159	−137
4h9t ^a	2.10	GKL	butyryl	−139	−135
4h9x ^a	2.20	GKL	butyryl	−100	−145
2uv0 ^b	1.80	LasR	3-oxo-dodecanoyl	−114	−144
3ix3 ^a	1.40	LasR	3-oxo-dodecanoyl	−110	−145
3szt ^a	2.55	QscR	3-oxo-dodecanoyl	−109	−125
1l3l ^b	1.66	traR	3-oxo-octanoic	−107	−129
2q0o ^a	2.00	traR	3-oxo-octanoic	−108	−128
^a Mean for two molecules in the asymmetric unit. ^b Mean for four molecules in the asymmetric unit.					

In particular, each of these proteins prefers to bind the AHL ligand with a ϕ dihedral angle between -100° and -160° , a nearly 180° -reorientation from that observed in the unbound state ($\phi \sim 50^\circ$; Table 8.1). The dichotomy in the conformation of the free and bound forms indicates that the receptor must reorganize the ligand for binding. The conformation of a bound AHL is enforced by hydrogen bonds with its receptor. In particular, the amide oxygen of a bound AHL forms a hydrogen bond with the phenolic hydroxyl group of a conserved tyrosine residue.⁴⁴⁰ This $C'_{i-1}=O_{i-1}\cdots H$ hydrogen bond competes with the $O_{i-1}\cdots C'_i=O_i$ $n\rightarrow\pi^*$ interaction of the free ligand. Accordingly, attenuating the basal $n\rightarrow\pi^*$ interaction could preorganize an AHL for binding to its receptor.

We reasoned that electron-withdrawing groups in the *N*-acyl chain would reduce the nucleophilicity of the donor oxygen and thereby reduce the influence of the $n\rightarrow\pi^*$ interaction.^{211, 446} To control for the influence of sterics, we replaced the three methyl groups in *N*-trimethylacetyl homoserine lactone with nearly isosteric bromo groups. Again using X-ray diffraction analysis, we found *N*-tribromoacetyl homoserine lactone adopts a conformation nearly identical to that of *N*-trimethylacetyl homoserine lactone (Table 8.1; Figure 8.3D). Moreover, the *N*-tribromoacetyl compound has a longer oxygen–carbon distance and diminished acceptor pyramidalization than does its *N*-trimethylacetyl analogue. These attributes are indicative of a weaker $n\rightarrow\pi^*$ interaction in the *N*-tribromoacetyl compound.^{98, 106} We confirmed this conclusion with NBO analysis of the optimized geometry of this compound, which reported that the energy associated with this $n\rightarrow\pi^*$ interaction was 0.55 kcal/mol (Table 8.1; Figure 8.3E), approximately 14% lower than that in the parent compound. Although the *N*-trimethylacetyl and *N*-tribromoacetyl AHLs *per se* might not be ideal quorum sensing inhibitors due to steric concerns, our observations demonstrate not only that the $n\rightarrow\pi^*$ interaction

contributes to the conformation of these important signal molecules, but also that the conformation can be modulated by an appropriate choice of *N*-acyl substituents.

Conclusions

We note that the presence of an $n \rightarrow \pi^*$ interaction could have another important pharmacological implication. γ -Lactones are susceptible to hydrolysis,⁴⁴⁷ which eliminates the activity of an AHL.⁴⁴⁸ An $n \rightarrow \pi^*$ interaction increases the energy of the acceptor π^* orbital^{99, 102} and thereby reduces the electrophilicity of the carbonyl group of the γ -lactone.^{175, 223, 449} Thus, the $n \rightarrow \pi^*$ interaction of AHLs could protect their γ -lactones against hydrolysis.

We conclude that an $O_{i-1} \cdots C'_i=O_i$ $n \rightarrow \pi^*$ interaction plays a key role in the conformation and, potentially, the biological activity of AHLs. Modifications that weaken this $n \rightarrow \pi^*$ interaction should increase the affinity of AHLs to their cognate receptors. Indeed, the propensity of electron-withdrawing substituents to attenuate the $n \rightarrow \pi^*$ interaction in AHLs could be contributing to the efficacy observed for certain synthetic AHLs as modulators of quorum sensing.⁴⁵⁰ Conversely, modifications that strengthen this interaction should decrease the rate of hydrolysis and endow AHLs with a longer biological half-life. We encourage exploration of this strategy for modulating bacterial quorum sensing with tailored small molecules.

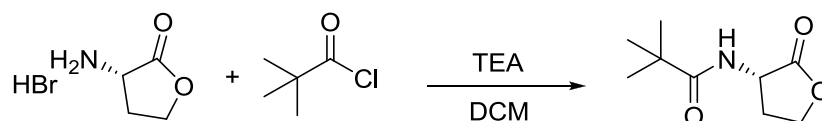
Acknowledgements. We are grateful to H. E. Blackwell for valuable advice and discussions, and I. A. Guzei for assistance with X-ray diffraction analysis. R. W. N. was supported by NIH Biotechnology Training Grant T32 GM008349. This work was supported by grants R01 AR044276 (NIH) and CHE-1124944 (NSF). Nuclear magnetic resonance spectroscopy was

supported by grants P41 RR002301 and P41 GM066326 (NIH). High-performance computing was supported by grant CHE-0840494 (NSF).

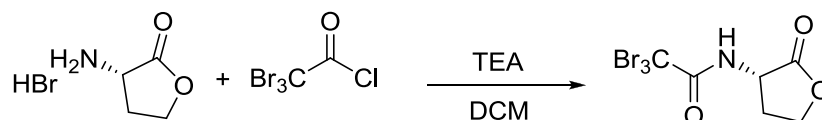
Methods

General Experimental. Commercial chemicals were of reagent grade or better, and were used without further purification. All chemicals were obtained from Sigma–Aldrich (St. Louis, MO). Anhydrous CH_2Cl_2 was obtained from CYCLE-TAINER[®] solvent delivery systems (J. T. Baker, Phillipsburg, NJ). Reactions were monitored by thin-layer chromatography with visualization by UV light or staining with KMnO_4 . Flash chromatography was performed with columns of silica gel 60, 230–400 mesh (Silicycle, Québec City, Canada). The removal of solvents and other volatile materials “under reduced pressure” refers to the use of a rotary evaporator at water-aspirator pressure (<20 torr) and a water bath of <45 °C.

Instrumentation. NMR spectra were acquired at ambient temperature with a Bruker DMX 400 MHz spectrometer (^1H , 400 MHz; ^{13}C , 100 MHz) in the National Magnetic Resonance Facility at Madison (NMRFAM). ^{13}C spectra were proton-decoupled. Mass spectrometry was performed with a Micromass LCT (electrospray ionization, ESI) instrument in the Mass Spectrometry Facility of the Department of Chemistry at the University of Wisconsin–Madison. X-Ray diffraction data were collected in the Molecular Structure Laboratory of the Department of Chemistry at the University of Wisconsin–Madison on a Bruker Quazar SMART APEXII diffractometer.



Synthesis of *N*-trimethylacetyl homoserine lactone (1). (*S*)-(-)- α -Amino- γ -butyrolactone hydrobromide (0.10 g, 0.55 mmol) was dissolved in 10 mL of anhydrous DCM with TEA (0.15 mL, 1.1 mmol). Trimethylacetylchloride (0.65 mL, 0.55 mmol) was added dropwise, and the resulting mixture was stirred at room temperature overnight. The mixture was then extracted with aqueous hydrochloric acid, followed by saturated aqueous sodium bicarbonate. The organic portion was then dried over anhydrous sodium sulfate, and solvent was removed under reduced pressure. Purification was achieved using silica gel chromatography with 4% v/v methanol in DCM. ^1H NMR (CDCl_3) δ 6.21 (s, 1H), 4.47 (m, 2H), 4.28 (m, 1H), 2.84 (m, 1H), 2.11 (m, 1H), 1.22 (s, 9H); ^{13}C NMR δ 179.1, 175.6, 66.0, 49.3, 38.6, 30.5, 27.4; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 186.1125, observed 186.1131.



Synthesis of *N*-tribromoacetyl homoserine lactone (2). (*S*)-(-)- α -Amino- γ -butyrolactone hydrobromide (0.50 g, 2.8 mmol) was dissolved in 20 mL of anhydrous DCM containing TEA (1.15 mL, 8.25 mmol). Tribromoacetylchloride (0.53 mL, 2.75 mmol) was added dropwise, and the resulting mixture was stirred at room temperature overnight. The mixture was then extracted with aqueous hydrochloric acid, followed by saturated aqueous sodium bicarbonate. The organic portion was then dried over anhydrous sodium sulfate, and solvent was removed under reduced pressure. Purification was achieved using silica gel chromatography with 4% v/v methanol in DCM. ^1H NMR (CDCl_3) δ 7.35 (s, 1H), 4.55 (m, 2H), 4.36 (m, 1H), 2.98 (m, 1H), 2.27 (m, 1H);

^{13}C NMR δ 174.1, 163.1, 66.4, 51.2, 33.9, 29.8; ESI-MS: $[\text{M} + \text{NH}_4]^+$ calculated 394.8237, found 394.8229.

Structure determination by X-ray diffraction analysis. Compounds were dissolved in hexane with a minimal amount of DCM. Slow evaporation afforded diffraction quality crystals in approximately three days. Crystals were selected under oil under ambient conditions and attached to the tip of a MiTeGen MicroMount. The crystal was mounted in a stream of cold nitrogen at 100(1) K and centered in the X-ray beam by using a video camera. Crystal evaluation and data collection were performed on a Bruker SMART APEXII diffractometer with Cu K_α ($\lambda = 1.54178 \text{ \AA}$) radiation for **1** and Mo K_α ($\lambda = 0.71073 \text{ \AA}$) for **2**. Data were collected by using the full-sphere data collection routine to a resolution of 0.80 \AA . The intensity data was then corrected for Lorentz and polarization effects. Absorption was corrected analytically. Structure solution by direct methods was carried out using SHELXS.⁴⁵¹ Refinement was performed using SHELXL.³⁶⁹ All non-hydrogen atoms were refined anisotropically. The final difference Fourier maps were featureless.

Computational methodology. Preferred conformations were determined by optimizing the respective compounds at the B3LYP/6-311+G(2d,p) level of theory as implemented by Gaussian 09.³⁷⁰ Frequency calculations of optimized structures yielded no imaginary frequencies, indicating a true minimum on the potential energy surface. Energies were corrected by the zero-point vibrational energy. Optimized geometries were subjected to analysis by NBO 5.9 at the B3LYP/6-311+G(2d,p) level of theory implemented in Gaussian 09.³⁷¹

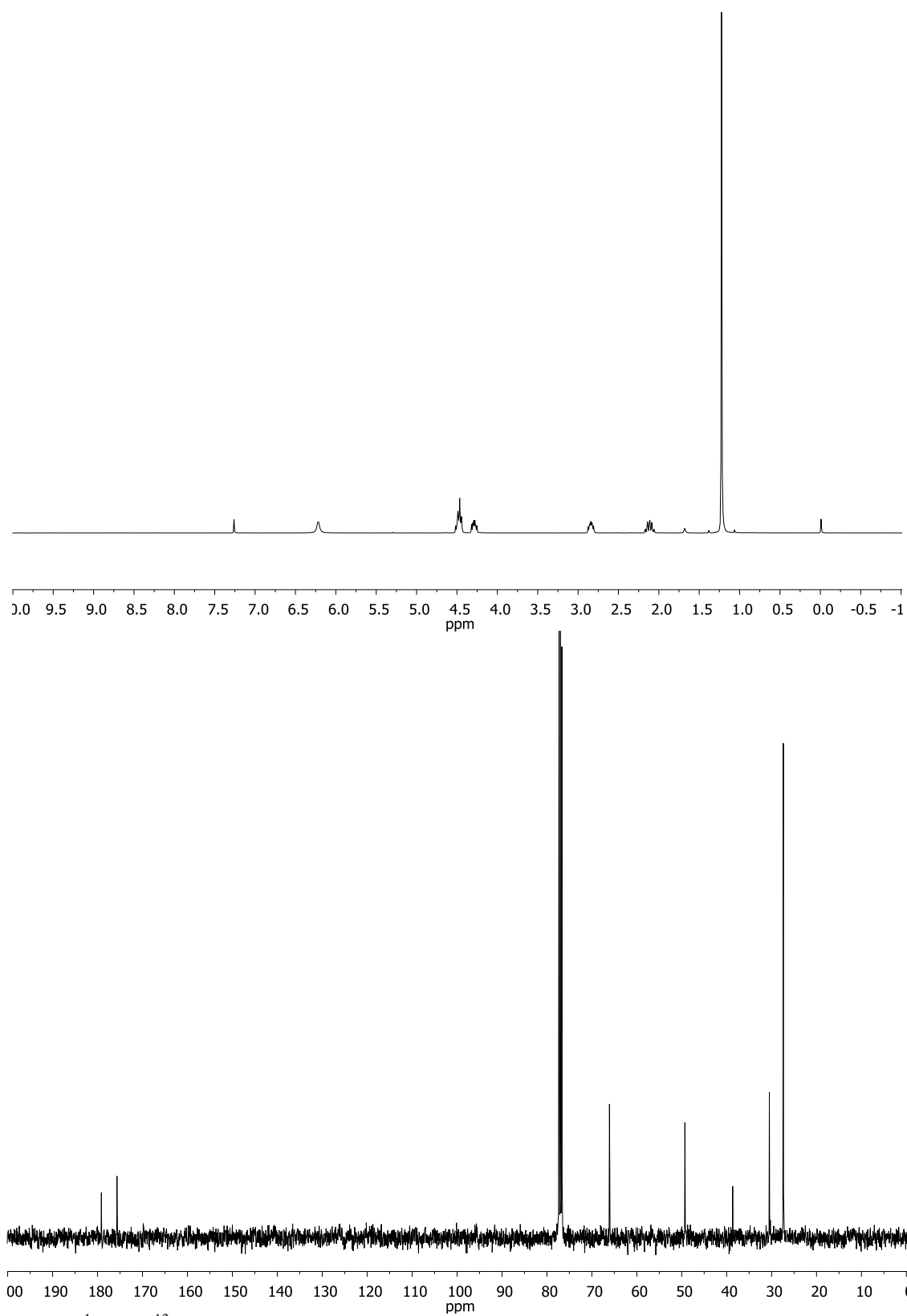


Figure 8.S1 ^1H and ^{13}C NMR spectra of *N*-trimethylacetyl homoserine lactone (**1**) in CDCl_3 .

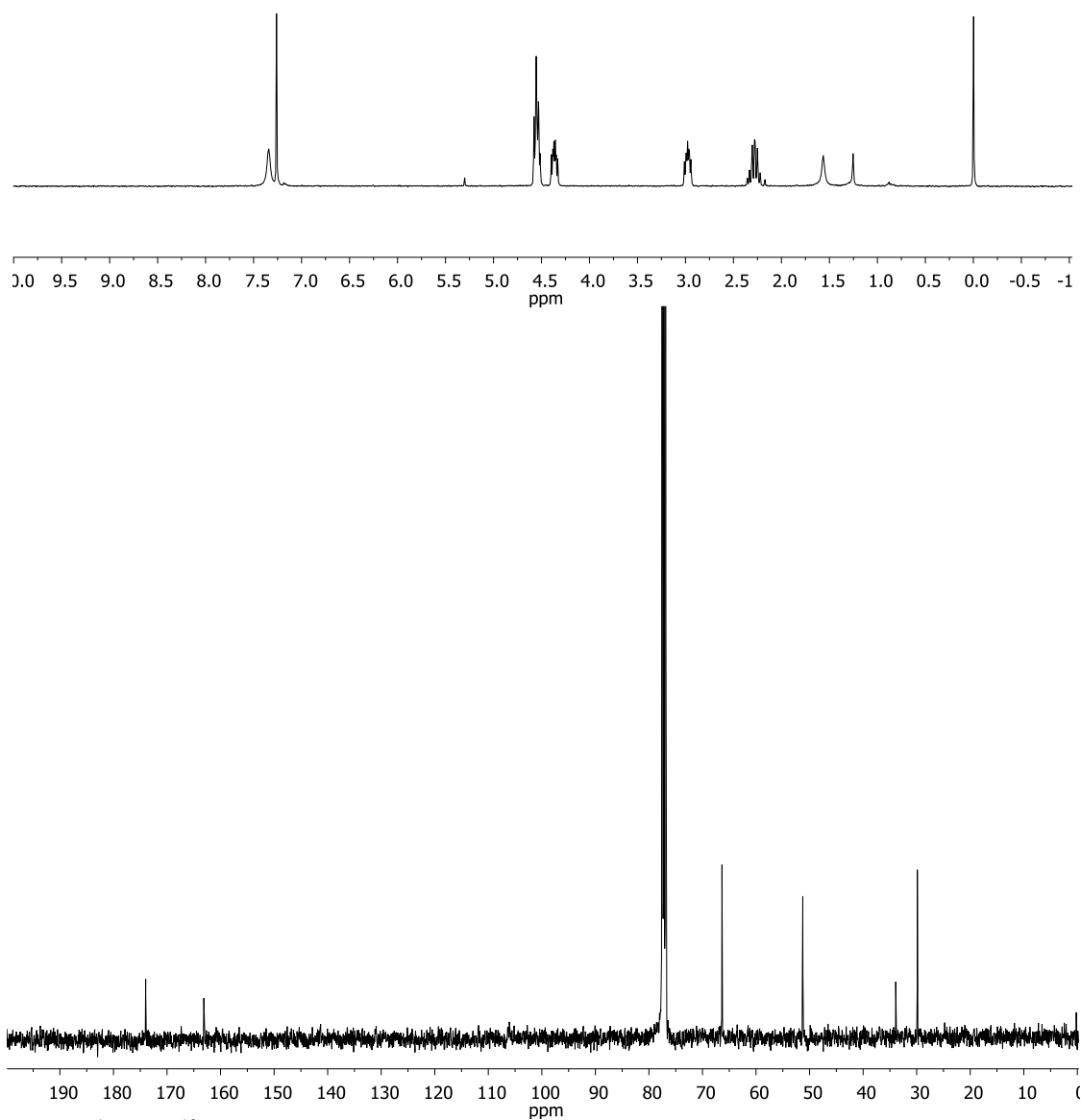


Figure 8.S2 ¹H and ¹³C NMR spectra of *N*-tribromoacetyl homoserine lactone (**2**) in CDCl₃/H₂O.

Table 8.S1 Crystal data and structure refinement for **1**

Identification code	raines66
Empirical formula	C ₉ H ₁₀ NO ₃
Formula weight	185.22
Temperature/K	100.01
Crystal system	monoclinic
Space group	P2 ₁
<i>a</i> /Å	5.9595(4)
<i>b</i> /Å	9.2235(10)
<i>c</i> /Å	9.2066(6)
α /°	90
β /°	102.719(4)
γ /°	90
Volume/Å ³	493.65(7)
<i>Z</i>	2
ρ_{calc} (mg/mm ³)	1.2460
m/mm ⁻¹	0.771
<i>F</i> ₀₀₀	200.7
Crystal size/mm ³	1.942 × 0.371 × 0.348
2 θ range for data collection	9.84 to 145.38°
Index ranges	−7 ≤ <i>h</i> ≤ 7, −8 ≤ <i>k</i> ≤ 10, −11 ≤ <i>l</i> ≤ 11
Reflections collected	10374
Independent reflections	1721 [<i>R</i> _{int} = 0.0277]
Data/restraints/parameters	1721/0/120
Goodness-of-fit on <i>F</i> ²	1.613
Final <i>R</i> indexes [<i>I</i> ≥ 2 σ _{<i>I</i>}]	<i>R</i> ₁ = 0.0583
Final <i>R</i> indexes [all data]	<i>R</i> ₁ = 0.0584, w <i>R</i> ₂ = 0.1605
Largest diff. peak/hole / e Å ⁻³	0.35/−0.67
Flack parameter	−0.0(2)

Table 8.S2 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for **1**. U_{eq} is defined as $1/3$ of the trace of the orthogonalized U_{ij} tensor.

atom	x	y	z	U_{eq}
O1	1617 (2)	5131.8 (17)	8378.8 (15)	22.9 (4)
O2	4374 (2)	4238.6 (17)	11934.2 (14)	19.6 (4)
O3	679 (2)	4873.8 (17)	11438.7 (15)	20.9 (4)
N1	243 (3)	2940 (2)	8815.7 (16)	17.2 (4)
C1	2178 (3)	4079 (2)	11210.1 (18)	14.5 (4)
C6	-1304 (3)	4122 (2)	6399 (2)	17.8 (4)
C4	5810 (3)	3099 (3)	11505 (2)	22.5 (5)
C2	1936 (3)	2756 (2)	10196.3 (19)	16.4 (4)
C5	319 (3)	4115 (2)	7953 (2)	16.2 (4)
C3	4404 (3)	2486 (2)	10051 (2)	24.2 (5)
C7	-2932 (4)	2842 (3)	6089 (2)	32.0 (6)
C8	239 (5)	4154 (5)	5288 (2)	58.8 (12)
C9	-2765 (7)	5496 (4)	6285 (4)	70.4 (15)

Table 8.S3 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for **1**. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11}+...+2hka \times b \times U_{12}]$.

atom	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
O1	26.0 (7)	20.1 (9)	19.9 (6)	-8.5 (6)	-0.6 (5)	0.5 (6)
O2	16.1 (6)	19.3 (8)	21.5 (6)	0.2 (5)	-0.2 (5)	-3.6 (6)
O3	18.6 (6)	20.0 (9)	24.9 (7)	2.0 (6)	6.6 (5)	-1.5 (6)
N1	17.2 (7)	15.4 (9)	16.5 (7)	-2.7 (6)	-1.7 (6)	0.9 (6)
C1	14.9 (8)	11.7 (11)	16.1 (8)	-2.8 (7)	1.9 (6)	1.9 (7)
C6	18.3 (8)	16.1 (10)	17.1 (8)	1.4 (8)	-0.1 (7)	0.8 (7)
C4	17.5 (8)	28.1 (13)	20.5 (9)	4.0 (8)	1.5 (7)	-3.0 (8)
C2	16.9 (8)	15.1 (11)	15.4 (8)	0.6 (7)	-0.3 (6)	1.0 (7)
C5	13.4 (7)	16.8 (11)	17.1 (8)	1.0 (7)	0.7 (6)	-0.2 (7)
C3	23.3 (9)	27.2 (13)	19.9 (9)	12.8 (9)	0.2 (7)	-5.8 (9)
C7	32.3 (9)	38.4 (15)	19.8 (9)	-16.5 (11)	-6.3 (7)	4.6 (9)
C8	37.9 (12)	121 (4)	17.6 (10)	-37.3 (19)	6.9 (9)	-5.7 (15)
C9	88 (2)	34.8 (19)	58 (2)	35.2 (19)	-50.1 (19)	-21.9 (15)

Table 8.S4 Bond lengths for **1**.

atom	atom	length/Å	atom	atom	length/Å
O1	C5	1.224 (3)	C6	C5	1.541 (2)
O2	C1	1.340 (2)	C6	C7	1.515 (3)
O2	C4	1.464 (3)	C6	C8	1.518 (3)
O3	C1	1.209 (2)	C6	C9	1.528 (3)
N1	C2	1.448 (2)	C4	C3	1.523 (3)
N1	C5	1.350 (3)	C2	C3	1.526 (2)
C1	C2	1.524 (3)			

Table 8.S5 Bond angles for **1**.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
C4	O2	C1	110.61 (15)	C9	C6	C8	111.0 (3)
C5	N1	C2	120.19 (16)	C3	C4	O2	105.21 (14)
O3	C1	O2	121.68 (17)	C1	C2	N1	113.43 (16)
C2	C1	O2	110.06 (15)	C3	C2	N1	116.16 (14)
C2	C1	O3	128.17 (15)	C3	C2	C1	102.78 (15)
C7	C6	C5	114.62 (17)	N1	C5	O1	121.84 (16)
C8	C6	C5	106.04 (15)	C6	C5	O1	121.16 (18)
C8	C6	C7	110.2 (2)	C6	C5	N1	117.01 (17)
C9	C6	C5	107.40 (16)	C2	C3	C4	102.55 (15)
C9	C6	C7	107.6 (2)				

Table 8.S6 Hydrogen atom coordinates ($\text{\AA} \times 10^4$) and isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for **1**.

atom	<i>x</i>	<i>y</i>	<i>z</i>	<i>U</i> _{eq}
H1	−838 (3)	2284 (2)	8538.1 (16)	20.6 (5)
H4a	7282 (3)	3505 (3)	11358 (2)	26.9 (6)
H4b	6146 (3)	2336 (3)	12278 (2)	26.9 (6)
H2	1450 (3)	1918 (2)	10743.6 (19)	19.7 (5)
H3a	4739 (3)	3003 (2)	9181 (2)	29.0 (6)
H3b	4704 (3)	1438 (2)	9959 (2)	29.0 (6)
H7a	−2052 (5)	1939 (3)	6290 (20)	48.0 (8)
H7b	−4040 (20)	2898 (12)	6733 (17)	48.0 (8)
H7c	−3760 (30)	2861 (13)	5044 (7)	48.0 (8)
H8a	1280 (50)	4980 (20)	5500 (30)	88.2 (18)
H8c	1140 (50)	3256 (19)	5370 (30)	88.2 (18)
H8b	−709 (6)	4240 (40)	4277 (4)	88.2 (18)
H9a	−3510 (50)	5544 (19)	7130 (20)	106 (2)
H9b	−1780 (12)	6348 (4)	6290 (40)	106 (2)
H9c	−3940 (40)	5477 (17)	5356 (18)	106 (2)

Table 8.S7 Crystal data and structure refinement for **2**.

Identification code	Raines67
Empirical formula	C ₆ H ₆ NO ₃ Br ₃
Formula weight	379.83
Temperature/K	100.01
Crystal system	orthorhombic
Space group	P2 ₁ 2 ₁ 2 ₁
<i>a</i> /Å	9.559(4)
<i>b</i> /Å	9.933(4)
<i>c</i> /Å	10.094(4)
α /°	90
β /°	90
γ /°	90
Volume/Å ³	958.4(6)
<i>Z</i>	4
ρ_{calc} (mg/mm ³)	2.6321
m/mm ⁻¹	12.598
<i>F</i> ₀₀₀	709.4
Crystal size/mm ³	0.463 × 0.37 × 0.048
Radiation	MoK α (λ = 0.71073)
2 θ range for data collection	5.76 to 61.04°
Index ranges	−13 ≤ <i>h</i> ≤ 13, −14 ≤ <i>k</i> ≤ 14, −14 ≤ <i>l</i> ≤ 14
Reflections collected	22866
Independent reflections	2931 [<i>R</i> _{int} = 0.0611]
Data/restraints/parameters	2931/0/117
Goodness-of-fit on <i>F</i> ²	0.602
Final <i>R</i> indexes [<i>I</i> ≥ 2 σ ₁]	<i>R</i> ₁ = 0.0229
Final <i>R</i> indexes [all data]	<i>R</i> ₁ = 0.0280, <i>wR</i> ₂ = 0.0494
Largest diff. peak/hole / e Å ⁻³	1.44/−0.78
Flack parameter	0.033(15)

Table 8.S8 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for **2**. U_{eq} is defined as $1/3$ of the trace of the orthogonalized U_{ij} tensor.

atom	x	y	z	U_{eq}
Br1	-583.9 (3)	-967.5 (3)	-3272.6 (4)	11.56 (7)
Br2	-2246.3 (4)	-1198.7 (4)	-5951.1 (4)	13.19 (7)
Br3	-3792.9 (3)	-1710.5 (3)	-3241.8 (4)	12.12 (7)
O1	-1277 (3)	-3972 (2)	-3136 (2)	11.7 (4)
O2	-3692 (3)	-5885 (3)	-3465 (3)	19.3 (6)
N1	-2065 (3)	-4169 (3)	-5233 (3)	10.3 (5)
O3	-2004 (3)	-7403 (3)	-3748 (3)	15.5 (5)
C3	-1975 (4)	-5613 (3)	-5262 (3)	9.8 (6)
C5	-514 (4)	-6238 (3)	-5248 (4)	11.4 (6)
C1	-2039 (4)	-1947 (3)	-4195 (3)	8.4 (6)
C2	-1731 (3)	-3466 (3)	-4154 (4)	8.5 (6)
C4	-794 (4)	-7602 (3)	-4606 (4)	14.1 (7)
C6	-2679 (4)	-6267 (4)	-4052 (4)	13.3 (6)

Table 8.S9 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for **2**. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11}+...+2hka \times b \times U_{12}]$.

atom	U_{11}	U_{22}	U_{33}	U_{12}	U_{13}	U_{23}
Br1	11.58 (14)	9.94 (14)	13.15 (16)	-1.59 (11)	-2.62 (14)	0.08 (13)
Br2	19.54 (17)	11.21 (14)	8.82 (15)	2.69 (13)	-1.42 (13)	2.63 (13)
Br3	9.75 (14)	13.06 (14)	13.54 (16)	2.52 (12)	1.67 (14)	1.22 (14)
O1	15.5 (11)	11.7 (10)	8.0 (11)	1.3 (10)	-3.8 (10)	-0.9 (10)
O2	15.6 (12)	22.4 (13)	19.9 (15)	-3.6 (11)	7.2 (12)	-0.7 (11)
N1	15.4 (14)	7.1 (12)	8.4 (13)	-1.1 (10)	-2.3 (11)	0.6 (10)
O3	18.6 (14)	14.3 (12)	13.7 (13)	-1.8 (10)	3.2 (10)	3.6 (10)
C3	13.5 (15)	7.8 (13)	8.2 (15)	-3.3 (12)	-1.4 (13)	-0.1 (12)
C5	13.1 (14)	8.6 (13)	12.6 (16)	0.3 (13)	3.8 (13)	0.7 (12)
C1	11.5 (14)	8.8 (13)	4.9 (14)	1.0 (11)	-0.6 (11)	0.3 (11)
C2	8.0 (13)	5.6 (13)	11.8 (16)	1.8 (11)	1.4 (12)	2.6 (12)
C4	18.0 (17)	8.4 (14)	15.8 (18)	-0.1 (13)	2.9 (14)	1.5 (13)
C6	10.9 (14)	15.3 (14)	13.7 (15)	-5.2 (12)	-3.5 (14)	2.2 (14)

Table 8.S10 Bond lengths for **2**.

atom	atom	length/Å	atom	atom	length/Å
Br1	C1	1.936 (3)	O3	C4	1.458 (5)
Br2	C1	1.932 (3)	O3	C6	1.336 (4)
Br3	C1	1.947 (3)	C3	C5	1.528 (5)
O1	C2	1.223 (4)	C3	C6	1.539 (5)
O2	C6	1.197 (5)	C5	C4	1.526 (5)
N1	C3	1.438 (4)	C1	C2	1.538 (4)
N1	C2	1.332 (5)			

Table 8.S11 Bond angles for **2**.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
C2	N1	C3	121.6 (3)	C2	C1	Br2	115.0 (2)
C6	O3	C4	111.2 (3)	C2	C1	Br3	105.7 (2)
C5	C3	N1	117.4 (3)	N1	C2	O1	123.8 (3)
C6	C3	N1	112.3 (3)	C1	C2	O1	119.6 (3)
C6	C3	C5	102.8 (3)	C1	C2	N1	116.5 (3)
C4	C5	C3	101.8 (3)	C5	C4	O3	105.7 (3)
Br2	C1	Br1	108.75 (16)	O3	C6	O2	123.0 (3)
Br3	C1	Br1	108.65 (16)	C3	C6	O2	127.8 (3)
Br3	C1	Br2	108.57 (16)	C3	C6	O3	109.1 (3)
C2	C1	Br1	110.0 (2)				

Table 8.S12 Hydrogen atom coordinates ($\text{\AA} \times 10^4$) and isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for **2**.

atom	x	y	z	U(eq)
H1	-2348 (3)	-3738 (3)	-5946 (3)	12.3 (7)
H3	-2469 (4)	-5936 (3)	-6075 (3)	11.8 (7)
H5a	-138 (4)	-6341 (3)	-6156 (4)	13.7 (7)
H5b	146 (4)	-5694 (3)	-4714 (4)	13.7 (7)
H4a	26 (4)	-7899 (3)	-4083 (4)	16.9 (8)
H4b	-997 (4)	-8290 (3)	-5289 (4)	16.9 (8)

Table 8.S13 Cartesian coordinates of the optimized geometry of **1**. SCF Energy = -632.406154 Hartree.

C	2.79423600	-1.14882200	-0.51359800
O	2.89150600	0.27287300	-0.76694000
C	1.38142100	0.06156300	1.04915900
C	2.03953700	-1.30904900	0.81607500
H	2.24967500	-1.58991700	-1.34622100
H	3.81187600	-1.53417600	-0.48382100
H	1.68747100	0.48143200	2.00746500
H	2.73176200	-1.53321900	1.62871700
H	1.30495200	-2.10911800	0.76306200
C	2.00200000	0.96798100	-0.03554900
O	1.78333200	2.13550500	-0.18948300
N	-0.07116000	0.10099100	1.03970900
H	-0.52222200	0.63291800	1.76403000
C	-0.78126300	-0.22349000	-0.07885300
O	-0.21585200	-0.69179800	-1.05756100
C	-2.30827900	-0.01345200	-0.06037200
C	-2.95621700	-1.40087900	-0.23860700
H	-2.73763100	-2.05383400	0.61066900
H	-4.04152100	-1.29843500	-0.31446500
H	-2.58701000	-1.88136400	-1.14444200
C	-2.65291700	0.88029300	-1.26815400
H	-3.73635000	1.00262000	-1.34199000
H	-2.20395000	1.87147900	-1.16873400
H	-2.28624500	0.43373100	-2.19181300
C	-2.83248300	0.63873200	1.22864600
H	-3.91684400	0.75165400	1.16524300
H	-2.62602800	0.02983400	2.11325800
H	-2.41791100	1.63955600	1.38046500

Table 8.S14 Cartesian coordinates of the optimized geometry of **2**. SCF Energy = 8235.128938 Hartree

C	-4.07961800	1.20396900	-0.52940700
O	-4.24591900	-0.21388400	-0.76574000
C	-2.70239000	-0.05585800	1.02546300
C	-3.29298900	1.34297800	0.78496000
H	-3.53505000	1.61339600	-1.37831300
H	-5.07746000	1.63584500	-0.48342600
H	-3.00122700	-0.44928700	1.99634900
H	-3.95736400	1.60992200	1.60709300
H	-2.52166400	2.10658800	0.71176000
C	-3.38754700	-0.94747200	-0.03558700
O	-3.23097500	-2.12522500	-0.17243000
N	-1.25230700	-0.16602400	0.98515100
H	-0.78562500	-0.66657100	1.72523500
C	-0.55841700	0.15330100	-0.12561400
O	-1.06452400	0.60228500	-1.13191900
C	0.99219000	0.00892500	-0.04486300
Br	1.68377000	1.80645200	0.40784500
Br	1.60972100	-1.27280000	1.33948900
Br	1.68218100	-0.56302900	-1.77486900

CHAPTER VIII – Addendum

Crystal Structure of *N*-(3-oxobutanoyl)-L-homoserine lactone

The structure and absolute configuration of the title compound, C₈H₁₁NO₄, which is a known quorum-sensing modulator, have been determined. The molecule exhibits signs of an intramolecular attractive carbonyl–carbonyl $n \rightarrow \pi^*$ interaction between the amide and lactone ester groups, specifically – a short contact of 2.709 (2) Å between the amide oxygen atom and ester carbon atom, approach of the amide oxygen atom to the ester carbonyl group along the Bürgi–Dunitz trajectory, at 99.1 (1), and pyramidalization of the ester carbonyl group by 1.1 (1). Moreover, a similar $n \rightarrow \pi^*$ interaction is observed for the amide carbonyl group approached by the ketone oxygen donor. These interactions apparently affect the conformation of the uncomplexed molecule, which adopts a different shape when bound to protein receptors. In the crystal, the molecules form translational chains along the *a* axis via N—HO hydrogen bonds.

This addendum has been published, in part, under the same title: Newberry, R.W.; Raines, R. T. *Acta Crystallogr.* **2016**, E72, 136-139.

R.W.N. synthesized and characterized the title compound and drafted the manuscript, which was edited by both authors.

N-Acyl homoserine lactones (AHLs) mediate quorum sensing in Gram-negative bacteria.^{436, 452} We have shown previously that AHLs engage in $n \rightarrow \pi^*$ interactions between the acyl and lactone ester carbonyl groups.¹⁷⁸ These interactions cause attraction through donation of oxygen lone pair (n) electron density into the π^* antibonding orbital of an acceptor carbonyl group.⁹⁰ This interaction is observed in the free molecule but not in structures of these compounds bound to their protein receptors, implicating these interactions in the potency of AHLs and their analogs. Our previous studies were restricted to AHLs with simple acyl appendages, but natural AHLs are also often oxidized at the 3-position to yield β -keto acyl groups, such as that reported here.

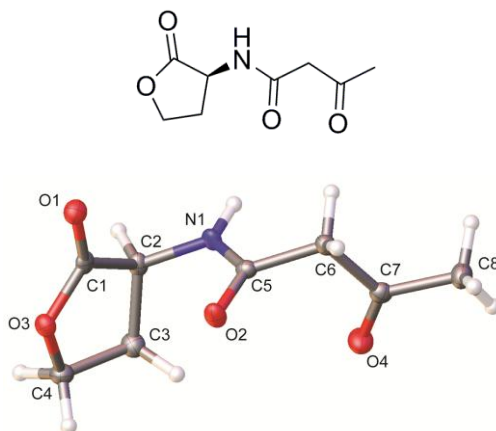


Figure 8.4 Molecular structure of *N*-(3-oxobutanoyl)-L-homoserine lactone. Thermal ellipsoids are drawn at 50% probability.

This is, to our knowledge, the first report of the structure of a free 3-oxo AHL (Figure 8.4). Individual molecules pack in linear arrays thanks to intermolecular hydrogen bonds between amide groups (Figure 8.5).

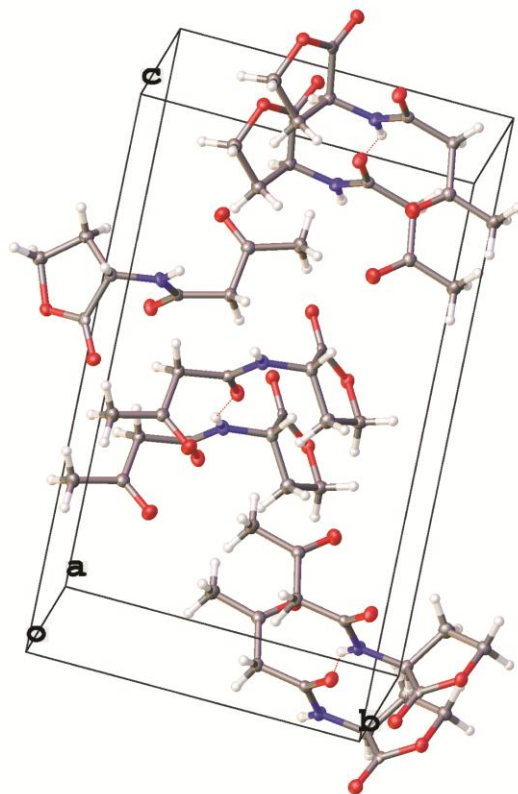


Figure 8.5 Packing of *N*-(3-oxobutanoyl)-L-homoserine lactone.

The molecule crystallizes as the keto tautomer, consistent with other β -keto amides.²⁵⁰ Like unoxidized AHLs, it displays the hallmark features of an attractive $n \rightarrow \pi^*$ interaction between the amide and ester carbonyl groups (Figure 8.6).

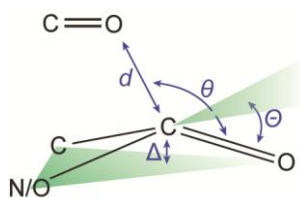


Figure 8.6 Structural parameters describing an $n \rightarrow \pi^*$ interaction.

Specifically, the donor oxygen makes a sub-van der Waals contact of 2.70926(16) Å with the acceptor carbonyl group, approaching with an angle of approach of 99.115(6)°, characteristic of

the Bürgi–Dunitz trajectory for nucleophilic addition.⁹³⁻⁹⁴ This geometry enables electron donation that in turn causes a characteristic pyramidalization of the acceptor carbonyl group. We observe that the carbonyl carbon rises 0.0157569(9) Å out of the plane of its substituents, creating a distortion angle θ of 1.10944(8)°. This signature has been used to diagnose the presence of these interactions in many molecules,^{98, 100, 102, 453} including polymers¹⁷⁹ and proteins.¹⁰³ Consistent with these observations, natural bond orbital (NBO) analysis of the crystal structure at the B3LYP/6-311+G(2d,p) level of theory predicts the release of 2.67 kcal/mol of energy due to the $n \rightarrow \pi^*$ interaction, indicating a significant contribution of this interaction to the conformation of this molecule (Figure 8.7).

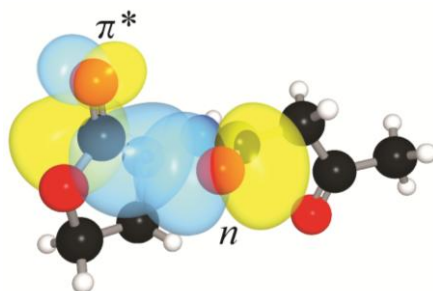


Figure 8.7 Overlap of oxo-AHL amide lone pair (n) and ester π^* orbitals.

Interestingly, a short contact was also observed between the ketone oxygen and amide carbonyl groups. In this case, the donor oxygen makes a 2.7461(2) Å contact at 107.4626(11)° to the amide carbonyl group. This contact causes the amide carbonyl group to distort 0.0076407(6) Å out of plane, corresponding to a distortion angle θ of 0.58657(6)°. The pyramidalization of the amide carbonyl group indicates a weaker $n \rightarrow \pi^*$ interaction from the ketone to the amide than from the amide to the ester, as would be expected for the enclosing of a four-membered ring relative to the enclosing of a five-membered ring, respectively. Indeed, NBO analysis predicts release of 1.42 kcal/mol of energy due to the $n \rightarrow \pi^*$ interaction between the ketone and amide

(Figure 8.8), which is nevertheless a significant contribution that likely biases the conformation of this molecule.

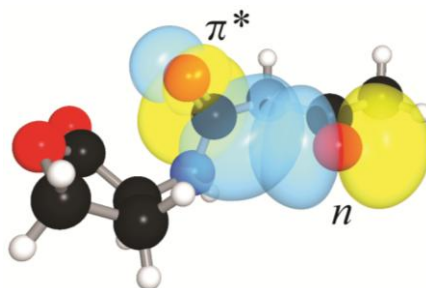


Figure 8.8 Overlap of oxo-AHL ketone lone pair (n) and amide π^* orbitals.

Based on the specific geometric parameters measured in this crystal structure, we conclude that the structure of unbound oxo-AHLs are influenced by $n \rightarrow \pi^*$ interactions, similarly to simple AHLs. Moreover, an additional $n \rightarrow \pi^*$ interaction specific to oxo-AHLs might bias their conformation further and thus affect their binding to protein receptors.

Acknowledgements. We thank I.A. Guzei and the Molecular Structure Laboratory at UW–Madison for assistance with data collection. This work was funded by grants CHE-1124944 (NSF) and R01 AR044276 (NIH). R.W.N. was supported by NIH Biotechnology Training Grant T32 GM008349 and by an ACS Division of Organic Chemistry Graduate Fellowship.

Synthesis and Crystallization. The title compound was prepared as reported previously.⁴⁵⁴ A small amount of solid product was dissolved in hexanes with a minimal amount of dichloromethane. Slow evaporation afforded high-quality crystals after ~4 days.

Table 8.S15 Crystal data and structure refinement for 3-oxo-AHL.

Identification code	3-oxo-AHL
Empirical formula	C ₈ H ₁₁ NO ₄
Formula weight	185.18
Temperature/K	99.97
Crystal system	orthorhombic
Space group	P2 ₁ 2 ₁ 2 ₁
a/Å	5.0215(4)
b/Å	9.8852(10)
c/Å	17.7668(14)
α/°	90
β/°	90
γ/°	90
Volume/Å ³	881.91(14)
Z	4
ρ _{calc} /g/cm ³	1.395
μ/mm ⁻¹	0.959
F(000)	392.0
Crystal size/mm ³	0.23 × 0.13 × 0.04
Radiation	CuKα (λ = 1.54178)
2θ range for data collection/°	9.956 to 146.58
Index ranges	-6 ≤ h ≤ 6, -12 ≤ k ≤ 11, -22 ≤ l ≤ 21
Reflections collected	11955
Independent reflections	1755 [R _{int} = 0.0275, R _{sigma} = 0.0152]
Data/restraints/parameters	1755/0/119
Goodness-of-fit on F ²	1.037
Final R indexes [I ≥ 2σ (I)]	R ₁ = 0.0265, wR ₂ = 0.0673
Final R indexes [all data]	R ₁ = 0.0275, wR ₂ = 0.0682
Largest diff. peak/hole / e Å ⁻³	0.22/-0.15
Flack parameter	-0.01(8)

Table 8.S16 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for 3-oxo-AHL. U_{eq} is defined as 1/3 of the trace of the orthogonalized U_{ij} tensor.

atom	x	y	z	$U(\text{eq})$
O1	1639 (3)	5285.1 (12)	5575.8 (7)	19.1 (3)
O2	-590 (2)	3797.1 (12)	4151.0 (7)	18.9 (3)
N1	3859 (3)	3996.8 (14)	4207.9 (8)	15.7 (3)
O3	157 (2)	6807.8 (12)	4755.7 (7)	16.5 (3)
O4	2366 (3)	2525.9 (13)	2628.4 (7)	25.4 (3)
C4	901 (4)	7334.0 (18)	4016.2 (10)	18.9 (4)
C7	2289 (3)	1584.4 (17)	3062.0 (9)	16.6 (3)
C1	1761 (3)	5789.9 (16)	4960.0 (9)	14.2 (3)
C8	2475 (4)	134.6 (18)	2812.8 (10)	23.0 (4)
C5	1638 (3)	3274.9 (17)	4102.1 (9)	14.3 (3)
C6	2005 (3)	1800.1 (16)	3906.5 (9)	16.1 (3)
C2	3720 (3)	5444.7 (16)	4328.4 (10)	15.8 (3)
C3	2703 (4)	6276.9 (18)	3659.1 (10)	20.0 (4)

Table 8.S17 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for 3-oxo-AHL. The anisotropic displacement factor exponent takes the form: $-2\pi^2 [h^2 a^{*2} U_{11} + 2hka^*b^* U_{12} + \dots]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
O1	20.1 (6)	18.2 (6)	18.9 (6)	0.5 (5)	2.2 (5)	-3.2 (5)
O2	10.5 (5)	18.2 (6)	27.9 (7)	-3.4 (5)	0.1 (5)	1.0 (5)
N1	9.7 (6)	16.1 (7)	21.2 (7)	-3.1 (6)	0.6 (5)	2.9 (5)
O3	14.0 (5)	16.1 (6)	19.2 (6)	-0.8 (5)	2.7 (5)	1.2 (5)
O4	35.9 (8)	20.1 (6)	20.1 (6)	3.1 (5)	0.6 (6)	0.5 (6)
C4	19.1 (8)	20.0 (8)	17.6 (8)	1.8 (7)	-1.3 (7)	0.8 (7)
C7	13.0 (7)	18.2 (8)	18.6 (8)	0.3 (6)	-1.0 (6)	-0.7 (7)
C1	10.9 (7)	12.6 (7)	19.0 (8)	-3.1 (6)	0.4 (6)	-4.8 (6)
C8	31.2 (10)	18.4 (8)	19.4 (8)	-2.4 (7)	-0.1 (7)	0.2 (8)
C5	12.5 (7)	17.2 (7)	13.1 (7)	0.9 (6)	0.0 (6)	1.6 (7)
C6	15.8 (8)	14.6 (7)	18.0 (8)	0.6 (6)	-0.2 (6)	0.9 (7)
C2	12.1 (8)	15.5 (8)	19.7 (8)	-2.8 (6)	2.0 (6)	-1.3 (6)
C3	20.6 (8)	20.2 (8)	19.3 (8)	1.3 (7)	3.8 (7)	-0.1 (8)

Table 8.S18 Bond lengths for 3-oxo-AHL.

atom atom length/Å			atom atom length/Å		
O1	C1	1.204 (2)	C4	C3	1.521 (2)
O2	C5	1.235 (2)	C7	C8	1.503 (2)
N1	C5	1.337 (2)	C7	C6	1.522 (2)
N1	C2	1.449 (2)	C1	C2	1.531 (2)
O3	C4	1.462 (2)	C5	C6	1.510 (2)
O3	C1	1.339 (2)	C2	C3	1.533 (2)
O4	C7	1.209 (2)			

Table 8.S19 Bond angles for 3-oxo-AHL.

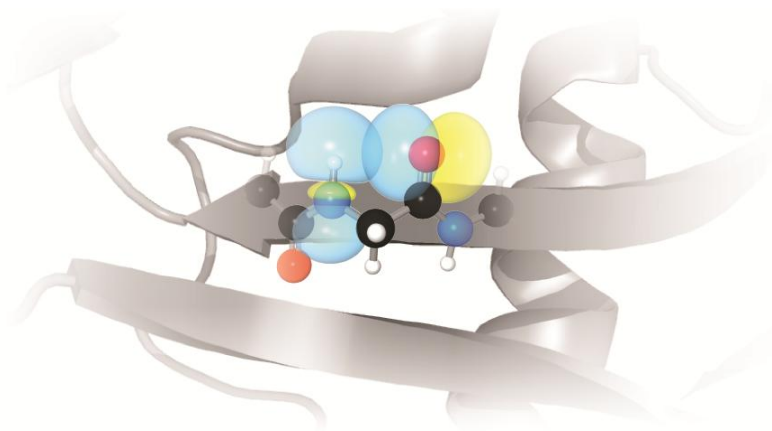
atom atom atom				angle/°	atom atom atom				angle/°
C5	N1	C2	120.52 (14)		O2	C5	N1	121.49 (15)	
C1	O3	C4	110.93 (13)		O2	C5	C6	122.03 (15)	
O3	C4	C3	106.42 (13)		N1	C5	C6	116.47 (14)	
O4	C7	C8	122.99 (15)		C5	C6	C7	111.94 (14)	
O4	C7	C6	121.55 (15)		N1	C2	C1	111.07 (13)	
C8	C7	C6	115.46 (14)		N1	C2	C3	115.56 (15)	
O1	C1	O3	121.80 (16)		C1	C2	C3	103.60 (14)	
O1	C1	C2	127.35 (16)		C4	C3	C2	104.07 (14)	
O3	C1	C2	110.81 (14)						

Table 8.S20 Hydrogen atom coordinates ($\text{\AA} \times 10^4$) and isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for 3-oxo-AHL.

atom	<i>x</i>	<i>y</i>	<i>z</i>	U(eq)
H1	5417	3589	4204	19
H4A	1855	8206	4068	23
H4B	-703	7480	3703	23
H8A	802	-332	2929	35
H8B	3944	-311	3078	35
H8C	2799	101	2269	35
H6A	455	1279	4092	19
H6B	3615	1452	4163	19
H2	5528	5781	4472	19
H3A	4199	6710	3388	24
H3B	1696	5701	3303	24

CHAPTER IX

Backbone C5 Hydrogen Bonds in Proteins



Current limitations in *de novo* protein structure prediction and design suggest an incomplete understanding of the interactions that govern protein folding. Herein, we demonstrate that previously unappreciated hydrogen bonds can occur within proteins between the amide proton and carbonyl oxygen of a single residue. Quantum calculations, infrared spectroscopy, and nuclear magnetic resonance spectroscopy show that these interactions share many hallmark features of canonical hydrogen bonds. Biophysical analyses of backbone-modified peptides demonstrates that selective attenuation or enhancement of the C5 hydrogen bond affects the stability of β -hairpins. These interactions are not only common, affecting at least 13% of β -sheet residues, but their cumulative impact provide several kcal/mol of stabilizing energy per every 100 residues in a typical protein. These results have broad implications for models of protein folding and stability.

Introduction

Predicting the structure of a protein from its sequence remains challenging,¹ motivating study of the noncovalent interactions that govern protein folding.⁶ Hydrogen bonding has received particular attention due to its unique ability to specify the geometry by which chemical groups interact.⁴⁴ In proteins, amide hydrogen-bond donors often approach carbonyl acceptors along the carbonyl bond axis, as in the α -helix and β -sheet.⁴⁶⁻⁴⁸ This observation is consistent with a modern description of the carbonyl lone pairs, featuring a predominantly s -type orbital (n_s) along the carbonyl bond with a p orbital (n_p) orthogonal to it (Figure 9.1A,B).⁴⁵⁵ Whereas n_s is poised to engage in common hydrogen-bond patterns in proteins, the role of n_p is less clear. In an α -helix, the π^* orbital of an adjacent carbonyl group can accept lone pair electron density from n_p , forming a so-called $n \rightarrow \pi^*$ interaction;^{92, 102} however, adjacent carbonyl groups in β -sheets are too distant to accept electron density from n_p . We therefore inquired as to what other electron acceptors could potentially engage with the carbonyl p -type lone pair in β -sheets.

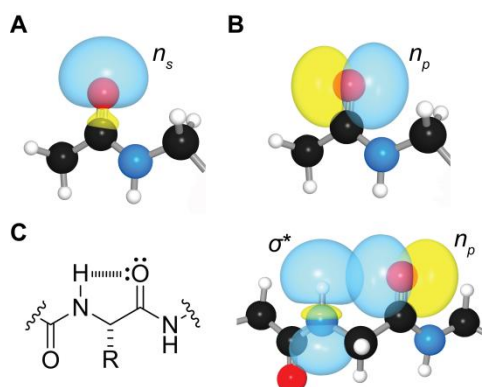


Figure 9.1 Carbonyl lone pairs in the peptide backbone of β -sheets, (A) s -Type and (B) p -type carbonyl lone pairs. (C) Putative C5 hydrogen bond, characterized by overlap of the p -type carbonyl lone pair and N–H σ^* orbital.

Upon inspection of the β -sheet, we noted close proximity of the carbonyl oxygen to the amide proton of the same residue, potentially creating a hydrogen bond through overlap of n_p

with the σ^* orbital of the N–H bond. (Figure 9.1C). This geometry has been termed “C5” for the size of the ring enclosed by the putative hydrogen bond.⁴⁵⁶⁻⁴⁵⁷ The C5 geometry has been observed for some amino acids in the gas phase,⁴⁵⁸⁻⁴⁵⁹ and calculations suggest it to be one of few minima on the potential energy surface.^{77, 460} The putative C5 hydrogen bond itself has, however, never been probed in the context of a peptide or protein in solution. If these interactions do indeed occur, their energies are likely to be weak, given that hydrogen-bond energies are maximized when the acceptor approaches along the donor bond axis;⁴⁴ indeed, algorithms for identifying hydrogen bonds in protein structure require this geometry for hydrogen-bond assignment.⁷⁹ It is unclear, therefore, if C5 interactions are truly hydrogen bonds or if they contribute to protein stability.

Computational Analysis of C5 Hydrogen Bonding

For an interaction to form, interpenetration of the donor and acceptor orbitals must occur. To evaluate the potential for orbital overlap in folded proteins, we measured the distance between putative donors and acceptors within β -sheet residues in sub-Å protein crystal structures, and found that over half featured oxygen–hydrogen distances <2.7 Å, which is the sum of the van der Waals radii for hydrogen and oxygen (Figure 9.2).

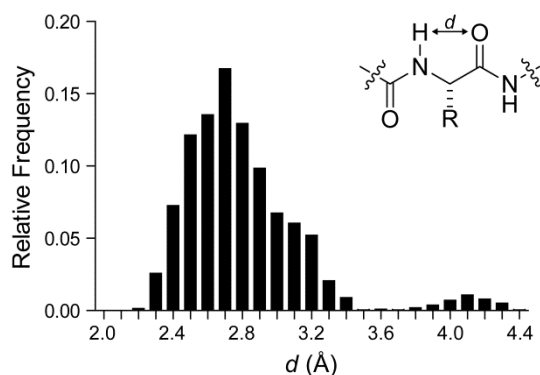


Figure 9.2 Frequencies of C5 hydrogen-bond donor–acceptor distances in β -sheets. NH \cdots O distances (d) from sub-Å protein crystal structures with assigned hydrogen coordinates.

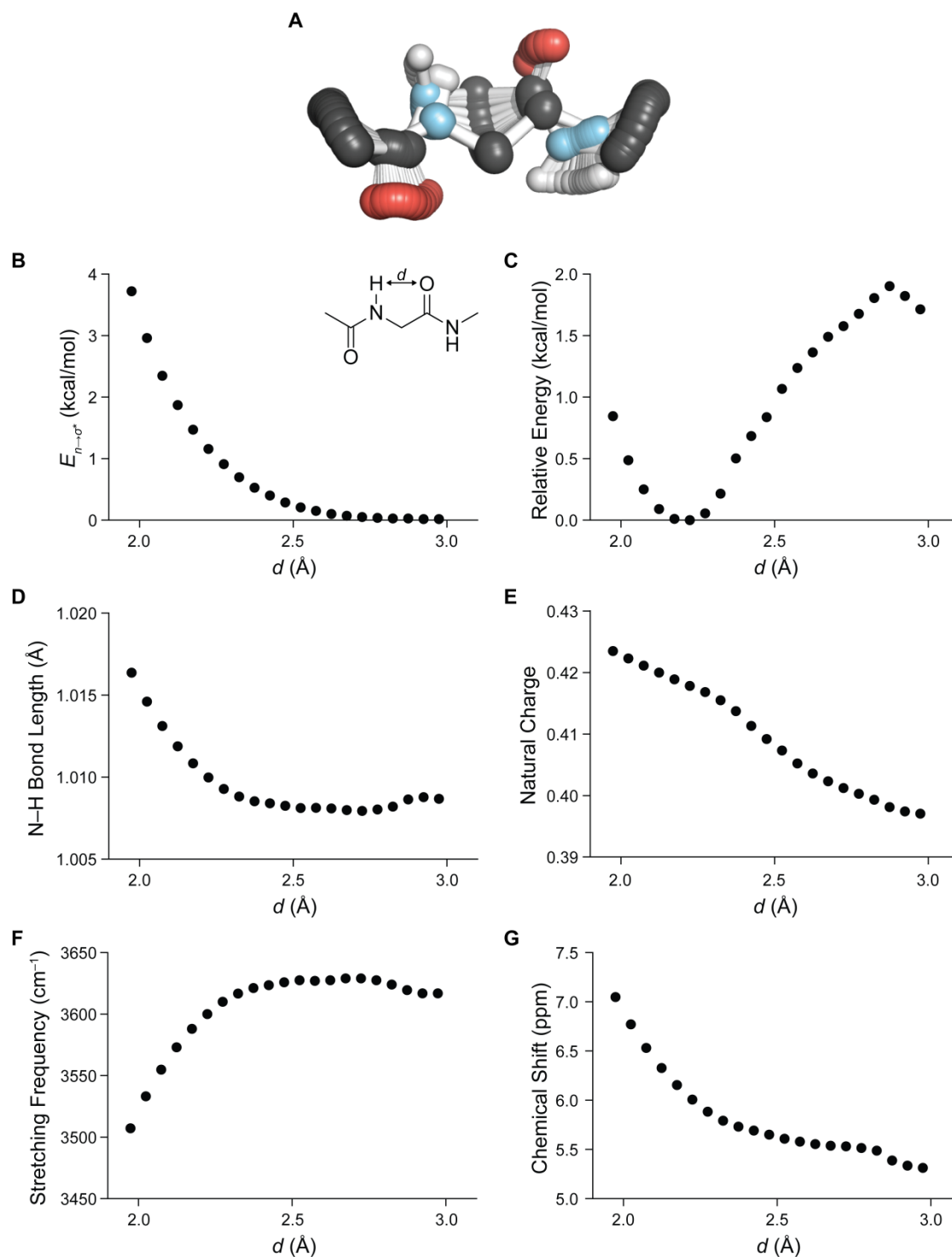


Figure 9.3 Computational characterization of AcGlyNHMe conformations.

(A) Overlay of AcGlyNHMe conformations generated by scanning the potential energy surface along the donor–acceptor distance. (B) C5 hydrogen bonding energy, (C) relative energy, (D) N–H bond length, (E) hydrogen charge, (F) N–H stretching frequency, and (G) donor ^1H chemical shift calculated for AcGlyNHMe conformations as a function of the C5 donor–acceptor distance.

To investigate whether these geometries enable C5 hydrogen bonding, we used density functional theory calculations to generate a set of AcGlyNHMe conformations that sample the donor–acceptor distances observed in proteins (Figure 9.3A). These conformers were then subjected to natural bond orbital (NBO) analysis to estimate the energy released by the mixing of the carbonyl *p*-type lone pair with the σ^* orbital of the hydrogen-bond acceptor.³⁹

Our calculations predict that C5 interactions can release significant energy, especially at donor–acceptor distances <2.5 Å. There, energy-release can exceed 0.25 kcal/mol (Figure 9.3B), which is similar to that expected from $n \rightarrow \pi^*$ interactions.⁹² In addition, the overall change in energy along this coordinate is similar to the hydrogen-bond energy estimated by NBO analysis (Figure 9.3C), suggesting that the interaction itself could make a significant contribution to the stability of these conformations. Moreover, as the donor–acceptor distance decreases, we predict lengthening of the N–H donor bond, increase in the partial positive charge on the donor hydrogen, a red shift of the donor stretching frequency, and a downfield chemical shift of the donor proton (Figure 9.3D–G), each suggesting that these interactions share the properties of typical hydrogen bonds.⁴⁴

Experimental Analysis of Individual C5 Hydrogen Bonds

Encouraged by these computational predictions, we set out to probe a single putative C5 hydrogen bond experimentally in a peptidic model system. To preorganize the putative donor and acceptor for interaction, we employed derivatives of diethylglycine (Deg), which has been shown by computation⁹² as well as NMR and vibrational spectroscopy⁴⁶¹ to populate the C5 geometry in solution. To probe the interaction, we compared diethylglycines bearing either an amide or an ester as the putative hydrogen-bond acceptor, as bona fide hydrogen bonds are

attenuated by replacing an amide acceptor with an ester.⁴²¹ Conformational analysis demonstrates that these compounds adopt the C5 geometry in solution (Figure 9.4), allowing us to isolate the effect of the putative C5 hydrogen bond.

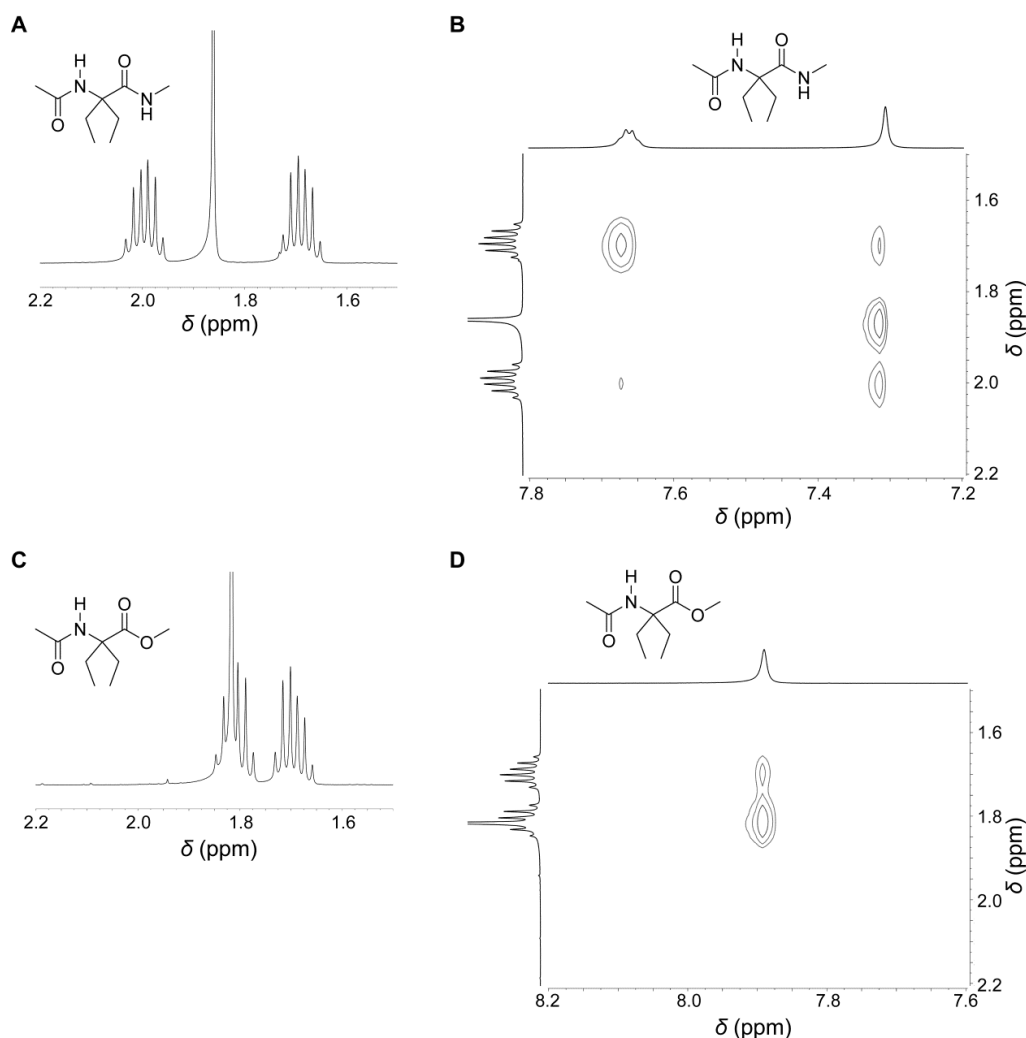


Figure 9.4 Conformational analysis of AcDegNHMe and AcDegOMe by NMR spectroscopy. (A) ¹H and (B) ¹H-¹H NOESY spectra of AcDegNHMe in DMSO-*d*₆. (C) ¹H and (D) ¹H-¹H NOESY spectra of AcDegOMe in DMSO-*d*₆. The one-dimensional ¹H NMR spectra shows two signals for the methylene groups with coupling constants consistent with gem splitting, indicating that the ethyl groups do not rotate freely on the NMR timescale. The presence of only two signals indicates an internal element of symmetry, which is consistent with the internal mirror plane of the C5 geometry. Moreover, the internal amide proton shows relatively equal NOEs to both methylene signals, indicating that it is oriented away from the ethyl groups; in contrast, the terminal amide proton has much stronger NOEs for one methylene over the other, indicating that it is oriented proximal to the ethyl groups. These correlations strongly suggest that the two carbonyl groups are oriented in opposite directions in solution. Similarity in the splitting patterns and coupling constants between AcDegNHMe and AcDegOMe indicates they adopt similar conformations.

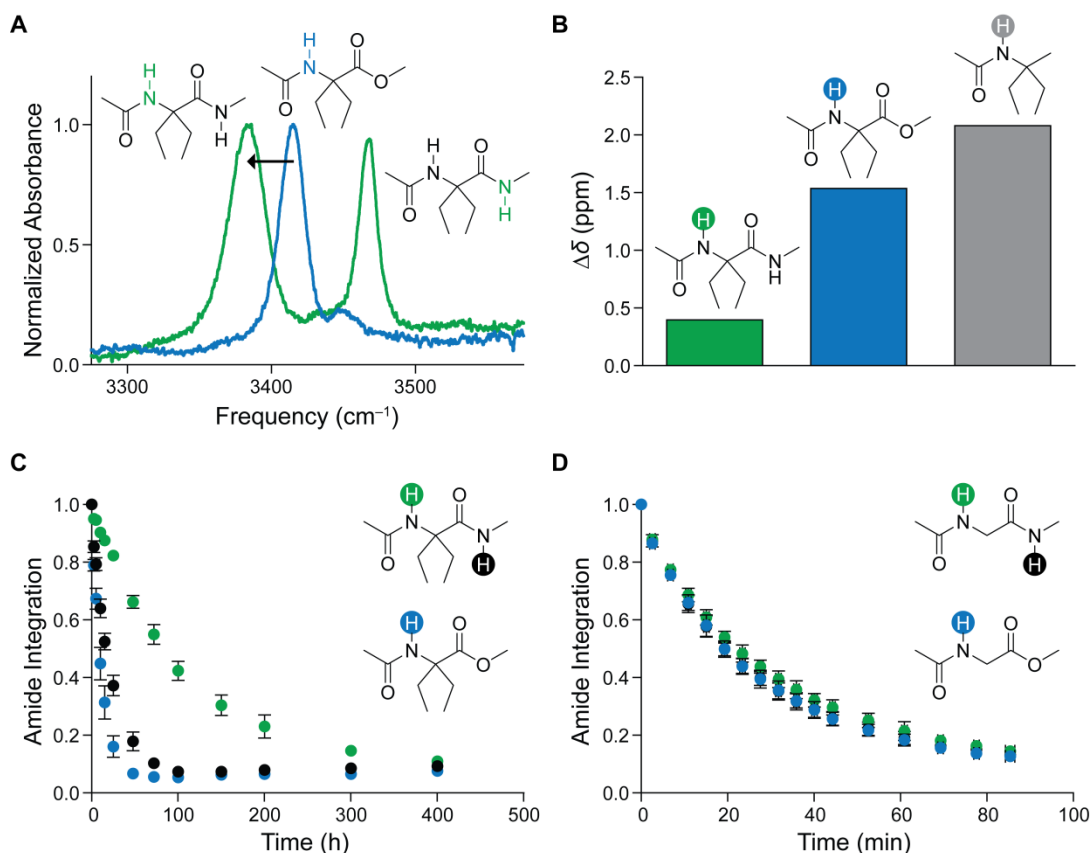


Figure 9.5 C5 Interactions display properties typical of hydrogen bonds.

(A) N–H stretching region of FTIR spectra of AcDegNHMe and AcDegOMe in CDCl_3 . (B) Change in chemical shift of amide protons of AcDegNHMe, AcDegOMe, and AcDegMe between CDCl_3 and $\text{DMSO}-d_6$ solutions. (C,D) Integration of donor amide ^1H NMR signals of AcDegNHMe and AcDegOMe (C), or AcGlyNHMe and AcGlyOMe (D) in $\text{DMSO}-d_6$ over time following the addition of D_2O .

Frequency analysis of the optimized C5 geometries of AcDegNHMe and AcDegOMe predicts that AcDegNHMe has a 30 cm^{-1} lower N–H stretching frequency than does AcDegOMe, consistent with the stronger C5 hydrogen bond predicted by NBO analysis (2.84 kcal/mol versus 1.73 kcal/mol). Infrared spectroscopy revealed that the N–H stretching mode of the putative C5 hydrogen-bond donor is indeed red-shifted—by 33 cm^{-1} —in AcDegNHMe relative to AcDegOMe (Figure 9.5A). The same mode of AcDegNHMe is also slightly red-shifted relative to that in AcDegNH₂ (Figure 9.6), consistent with a primary amide being a less effective hydrogen-bond acceptor than is a secondary amide.

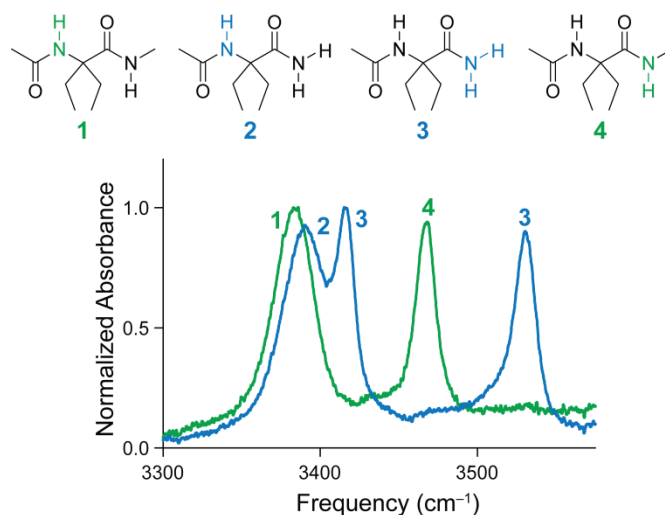


Figure 9.6 Fourier transform infrared spectra of AcDegNHMe and AcDegNH₂. Only the N–H stretching region is shown. Spectra were acquired in CDCl₃.

We next examined the effect of this interaction on the NMR properties of the donor proton. Replacing the acceptor amide with an ester creates an upfield chemical shift of the donor proton, despite the ester being more electron-withdrawing (Figure 9.7).

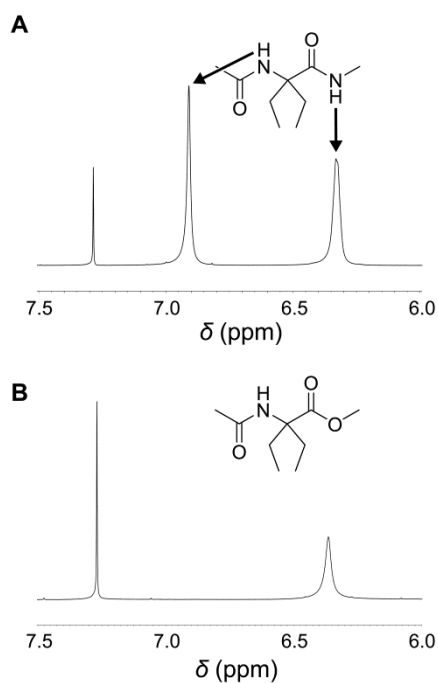


Figure 9.7 ¹H NMR spectra of AcDegNHMe and AcDegOMe. Only the amide region is shown. Spectra were acquired in CDCl₃. (A) AcDegNHMe and (B) AcDegOMe.

Switching from a solvent that does not engage in hydrogen bonding to one that does causes a downfield shift of the donor proton, and this effect is attenuated by the presence of an internal hydrogen bond. In accord with previous studies,⁴⁶¹ we found that the internal amide proton of AcDegNHMe experiences a much smaller chemical shift upon changing from CDCl₃ to DMSO-*d*₆ than do the amide protons of AcDegOMe or a diethylglycine derivative lacking any putative C5 hydrogen-bond acceptor (Figure 9.5B). We next assayed the ability of this interaction to slow the exchange of the donor proton. Following addition of D₂O to a DMSO-*d*₆ solution, the putative C5 hydrogen-bond donor in AcDegNHMe exchanges much more slowly than does the corresponding proton in AcDegOMe (Figure 9.5C), consistent with the stronger protective effect of the amide acceptor. This difference in exchange rate is contingent upon adoption of the C5 geometry, as the exchange rates of AcGlyNHMe and AcGlyOMe are indistinguishable (Figure 9.5D). These data establish that these interactions are consistent with hydrogen bonding.

C5 Hydrogen Bonding in β -Sheets

Confident that these interactions constitute meaningful hydrogen bonding, we next sought to characterize their contributions to the stability of proteins using tryptophan zipper (TrpZip2) β -hairpins as a model for the β -sheet (Table 9.1, Figure 9.8A).⁴⁶²

Table 9.1 Sequences and T_m values of TrpZip peptides.		
Peptide	Sequence	T_m (K)
TrpZip2-A	Ac-S-W-T-W-E-N-G-K-W-T-W-K-NH ₂	346.3 \pm 0.4
TrpZip2-B	Ac-S-W-T*W-E-N-G-K-W-T-W-K-NH ₂	330.4 \pm 0.5
TrpZip2-C	Ac-S-W-T-W-E-N-G-K-W*T-W-K-NH ₂	ND
TrpZip2-D	Ac-S-W-T*W-E-N-G-K-W*T-W-K-NH ₂	ND
TrpZip2-E	Ac-W-T-W-E-N-G-K-W-T-W-K-NH ₂	351.7 \pm 0.5
TrpZip2-F	Ac-W-T-W-E-N-G-K-W-T-W-K-OMe	349.5 \pm 0.4
TrpZip2-G	Ac-S-W-T-W-E-N-G-K-W-T-W-NH ₂	325.8 \pm 1.2
TrpZip2-H	F ₃ C-S-W-T-W-E-N-G-K-W-T-W-NH ₂	330.0 \pm 1.0
*Denotes ester linkage.		

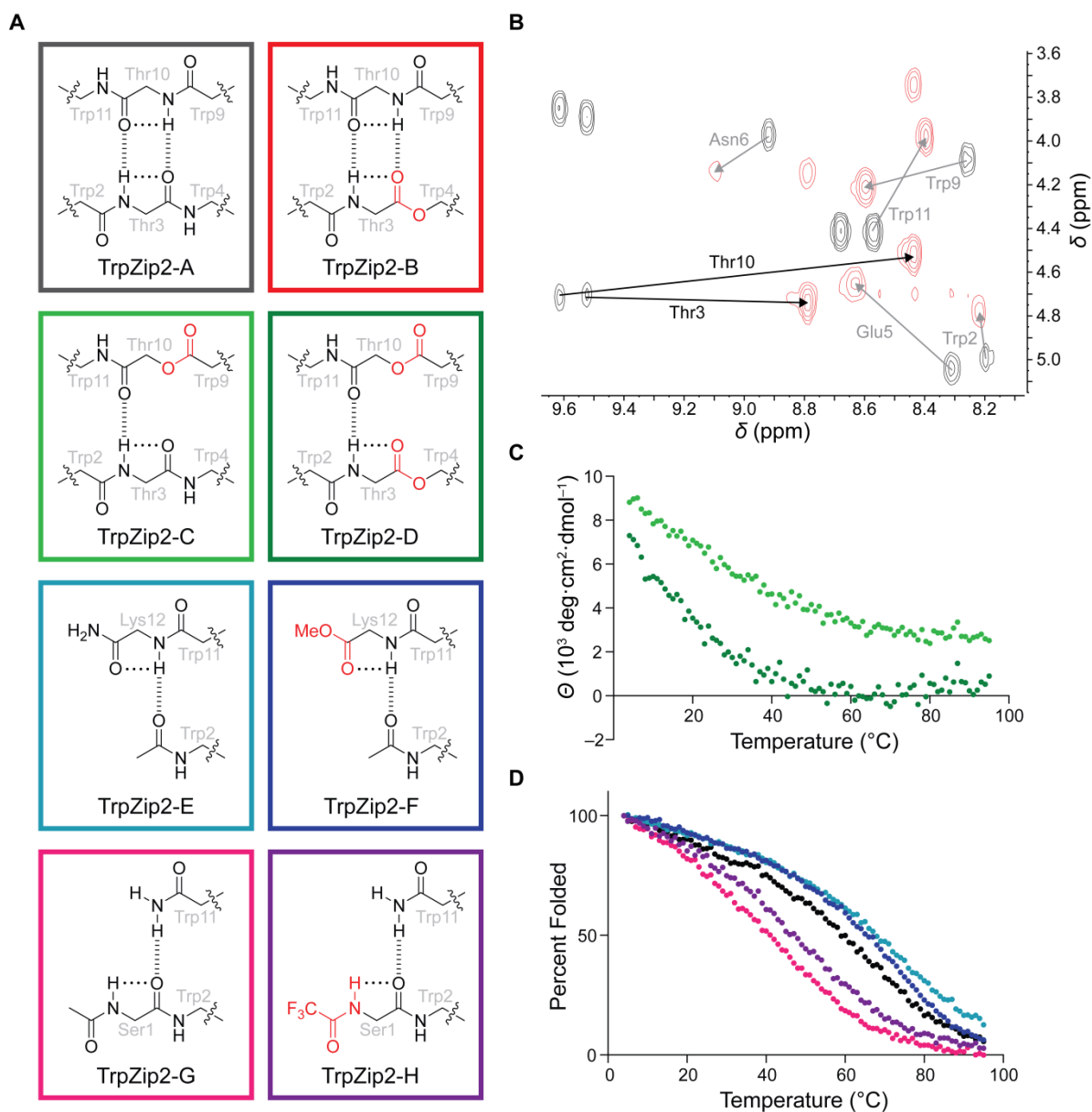


Figure 9.8 Perturbation of C5 hydrogen bonds impacts β -sheet stability.

(A) TrpZip2 peptides. (B) NH-H a region of ^1H - ^1H TOCSY spectra of TrpZip2-A and TrpZip2-B. (C,D) Thermal denaturation of TrpZip2 peptides.

To identify if C5 hydrogen bonding does occur in β -sheet structure, we compared the ^1H NMR chemical shifts of TrpZip2-A and TrpZip2-B (Figure 9.8B); the latter features an ester linkage between Thr3 and Trp4, where it can accept a canonical hydrogen bond from Thr10-NH and a C5 hydrogen bond from Thr3-NH. As expected, Thr10-NH experiences a large upfield shift in

the presence of a weaker hydrogen-bond acceptor. Importantly, Thr3-NH likewise experiences a large upfield shift, despite the greater inductive effect of the ester group, which, for comparison, leads to a downfield shift of Glu5-NH. The sensitivity of an amide proton to the identity of the C5 hydrogen-bond acceptor strongly suggests that C5 hydrogen bonds operate in folded proteins.

To evaluate the importance of the C5 interaction specifically, we eliminated the canonical hydrogen-bond donor by replacing it with an ester (TrpZip2-C). In the absence of the canonical donor, replacement of an amide acceptor with an ester attenuates the C5 hydrogen bond of interest selectively (TrpZip2-D). Circular dichroism spectra indicated this attenuation reduces structural content (Figure 9.9).

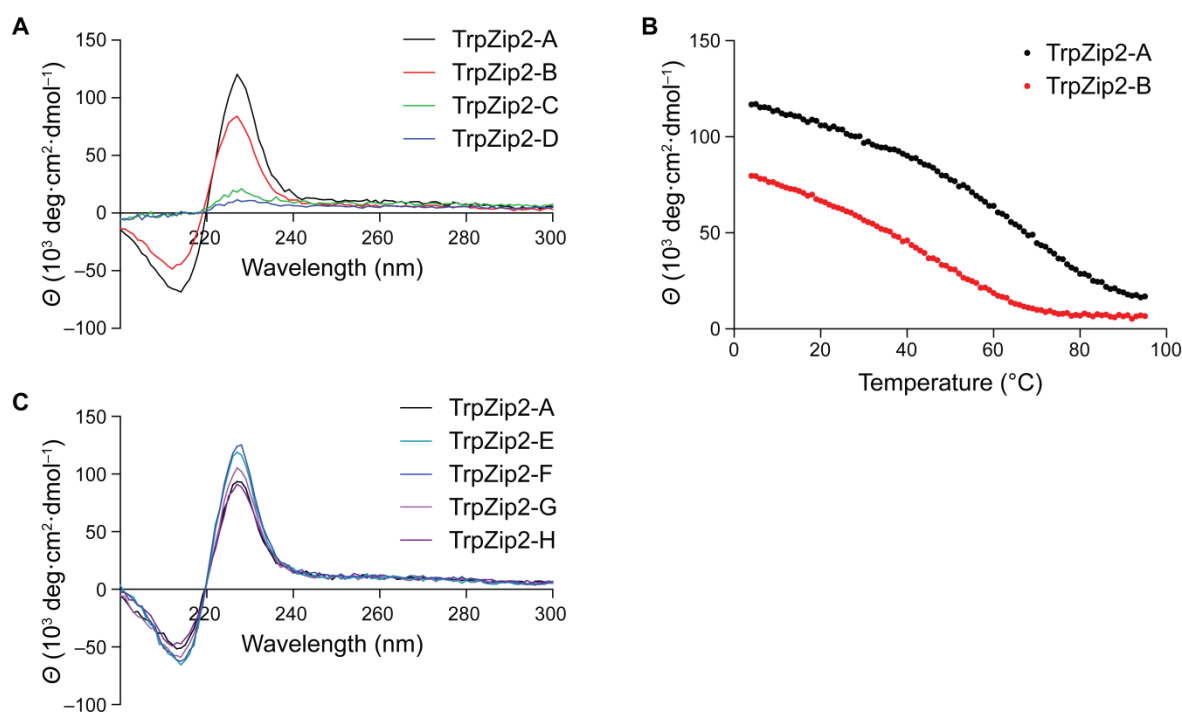


Figure 9.9 Far-UV circular dichroism spectra of TrpZip peptides. (A) Spectra of TrpZip2 peptides A–D. (B) Thermal denaturation of TrpZip2-A and TrpZip2-B. (C) Far-UV circular dichroism spectra of TrpZip2 peptides E–F.

Thermal denaturation demonstrated that TrpZip2-D also loses all measurable structure by 50 $^{\circ}\text{C}$, while TrpZip2-C retains some residual structure at temperatures up to ~75 $^{\circ}\text{C}$ (Figure 9.8C).

Because introduction of the second ester attenuates the interstrand oxygen–oxygen repulsion created by hydrogen bond deletion,⁴¹⁹ the observed decrease in the thermal stability of TrpZip2-D relative to TrpZip2-C might underestimate the contributions of C5 hydrogen bonding.

Though suggestive, results from analysis of TrpZip2-C and TrpZip2-D are clouded by the relative instability of these peptides. Moreover, attenuation of the C5 hydrogen-bond acceptor also necessarily perturbs the proposed interstrand $C^\alpha-H\cdots O=C$ hydrogen bond in β -sheets,⁸¹ potentially confounding our results. To improve the stability of these backbone-modified peptides and to isolate C5 interactions from other noncovalent interactions, we probed these interactions at the peptide termini, rather than in the center of the hairpin. Terminal modifications have been found to be less disruptive,⁴⁶³ and the C-terminal carbonyl group cannot accept $C^\alpha-H$ hydrogen bonds. Elimination of the N-terminal serine residue leaves the C-terminal carbonyl group without a canonical hydrogen-bond donor and isolates the C5 hydrogen bond of interest (TrpZip2-E). Thermal denaturation demonstrated that attenuation of the terminal C5 hydrogen bond with an ester acceptor (TrpZip2-F) lowers the value of T_m (which is the temperature at the midpoint of the thermal transition between the folded and unfolded states) by 2 °C (Figure 9.8D, Table 9.1), demonstrating the influence of a single C5 hydrogen bond. We then attempted to enhance the strength of the C5 hydrogen bond by replacing the N-terminal acetyl group with a trifluoroacetyl group, which increases the acidity of the N-terminal hydrogen-bond donor; the C5 hydrogen bond could then be isolated by removing the C-terminal lysine residue (TrpZip2-G and TrpZip2-H). We observed that the T_m value of the peptide capped with a trifluoroacetyl group is 4 °C higher than that of the peptide capped with an acetyl group (Figure 9.8D, Table 9.1), demonstrating that increasing the strength of only a single C5 hydrogen bond can increase the global stability of β -sheet structures.

C5 Hydrogen Bonding in Proteins

Given the prevalence of β -sheets in proteins, C5 hydrogen bonds en masse could contribute significantly to protein stability. To evaluate their potential contributions, we analyzed the energy of C5 hydrogen bonds by conducting NBO analysis on a set of AcGlyNHMe conformers that sample β -sheet geometry (Figure 9.10A).

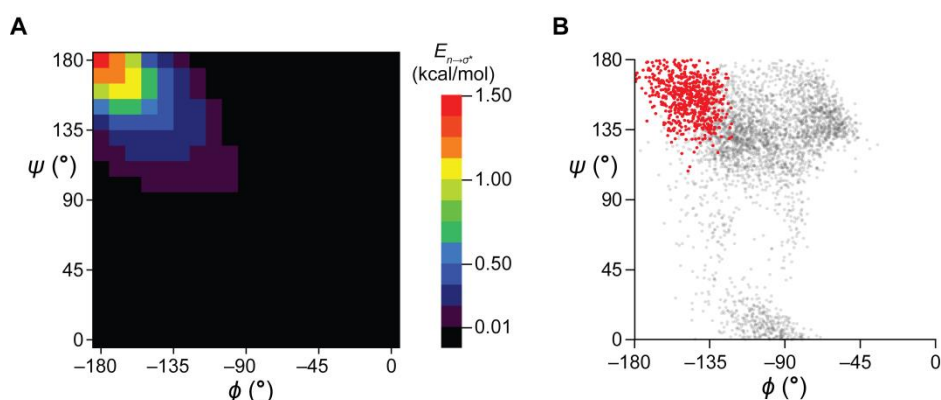


Figure 9.10 Energy and frequency of C5 hydrogen bonds in proteins.

(A) Ramachandran plot of the energy of C5 hydrogen bonds. (B) Ramachandran plot of residues in sub-Å protein crystal structures with reported hydrogen coordinates. Residues with donor-acceptor distances < 2.5 Å are shown in red.

We found a peak C5 hydrogen bond energy of 1.44 kcal/mol, with energies of at least 0.25 kcal/mol for residues with absolute backbone dihedral angles $> 140^\circ$; this region also corresponds to residues in sub-Å crystal structures with donor-acceptor distances < 2.5 Å (Figure 9.10B), which in turn correlated with several properties of hydrogen bonding in our computational analysis of AcGlyNHMe (Figure 9.3). Considering only residues with absolute backbone dihedral angles $> 140^\circ$, we found that $\sim 5\%$ of all residues engage in C5 hydrogen bonding, and 94% of the proteins we examined contain at least one C5 hydrogen bond. Most of these residues (62%) were assigned to β -sheet secondary structure, and 13% of all β -sheet residues engage in C5 hydrogen bonds. Antiparallel β -sheets have a higher frequency (14%) than

do parallel β -sheets (9%), consistent with differences in donor–acceptor distances (Figure 9.11), suggesting that C5 hydrogen bonds might contribute to differences in the stability of these two architectures.

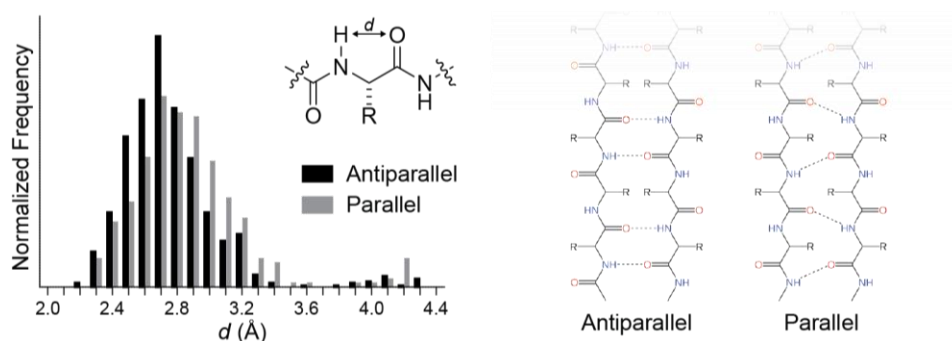


Figure 9.11 C5 hydrogen-bond donor–acceptor distances in parallel and antiparallel β -sheets. NH \cdots O distances (d) from sub-Å protein crystal structures with assigned hydrogen coordinates.

β -Bulges and bends each have C5 hydrogen-bond frequencies of $\sim 10\%$ and together make up 10% of residues with C5 interactions. The remaining interactions (27%) were not assigned to secondary structure, suggesting that these interactions impart stability to irregular loops or turns. Residues engaged in C5 hydrogen bonds are also less solvent-exposed than are other residues (Figure 9.12), suggesting that C5 hydrogen bonds compete effectively with water for the hydrogen-bond potential of the protein backbone.

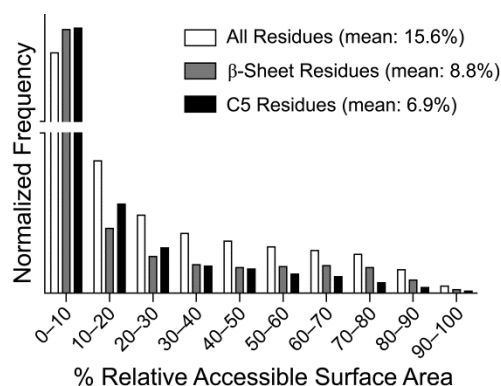


Figure 9.12 Solvent accessibility of residues engaged in C5 hydrogen bonds. Frequencies of relative backbone solvent-accessible surface areas for residues in high-resolution protein crystal structures.

To estimate the total contributions from C5 hydrogen bonds, we binned residues from a nonredundant set of high-resolution protein crystal structures by their backbone dihedral angles. We assigned an energy from our NBO calculations to each bin, and summed the energetic contributions, again only considering residues with absolute backbone dihedral angles $>140^\circ$. Notably, TrpZip peptides feature dihedral angles outside this range,⁴⁶² so our criteria likely provide a lower bound to the expected contributions of C5 hydrogen bonds. We found that C5 hydrogen bonds can contribute an average total stabilizing energy of ~ 4.5 kcal/mol per 100 residues, similar to contributions estimated for $n \rightarrow \pi^*$ ⁹² or cation- π interactions;¹¹⁶ moreover, β -rich proteins could benefit from significantly higher contributions (Figure 9.13). Although these estimates do not account for differences in individual microenvironments, they do suggest a role for C5 hydrogen bonds in stabilizing folded proteins.

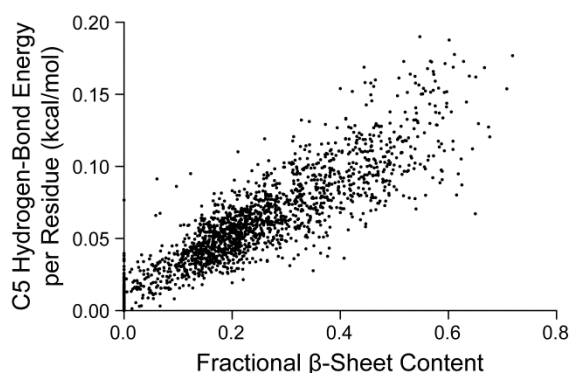


Figure 9.13 Relationship of C5 hydrogen-bonding energy with β -sheet content in proteins. Data are from high-resolution protein crystal structures.

We speculate that local C5 hydrogen bonds contribute to protein folding by directing the polyproline II secondary structure, which is believed to be common in the unfolded states of proteins,²⁰⁶ toward β -strand geometry. These local contacts could prepare strands for association into sheets, which are otherwise stabilized largely by nonlocal contacts. An overabundance of such interactions could, however, contribute to the formation of amyloids, since C5 hydrogen

bonds are stronger when backbone dihedral angles mirror one another (Figure 9.10A), which creates a flat conformation that proteins have evolved to avoid so as to prevent aggregation.⁴⁶⁴ Indeed, amyloid structures show highly symmetric backbone dihedral angles (Figure 9.14),⁴⁶⁵ suggesting a significant contribution from C5 hydrogen bonds for this general feature of protein folding landscapes.⁴⁶⁶

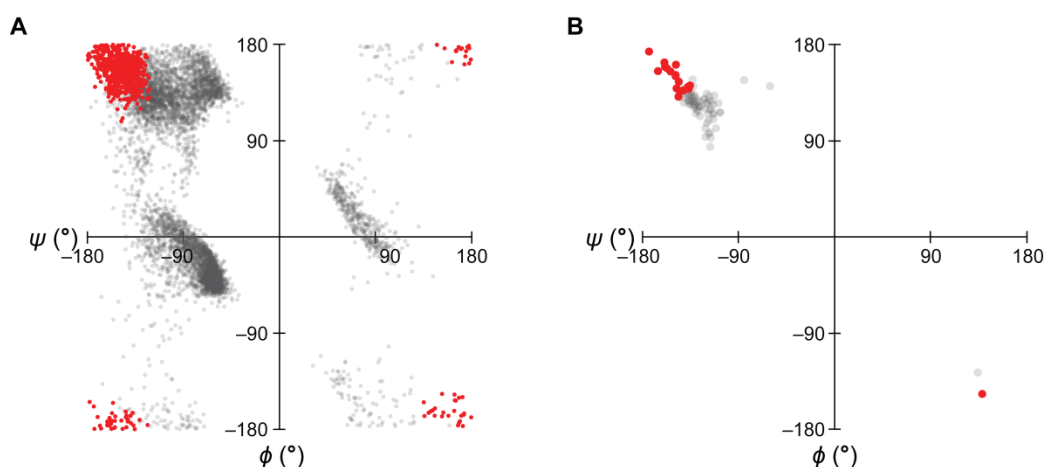


Figure 9.14 Ramachandran plots of residues engaged in C5 hydrogen bonds.

(A) Ramachandran plot of residues in sub-Å protein crystal structures with assigned hydrogen coordinates. Residues with hydrogen-bond donor-acceptor distances (d) < 2.5 Å are shown in red. (B) Ramachandran plot of residues in amyloid structures.^{465, 467} Residues with absolute backbone dihedral angles greater than 135° are shown in red.

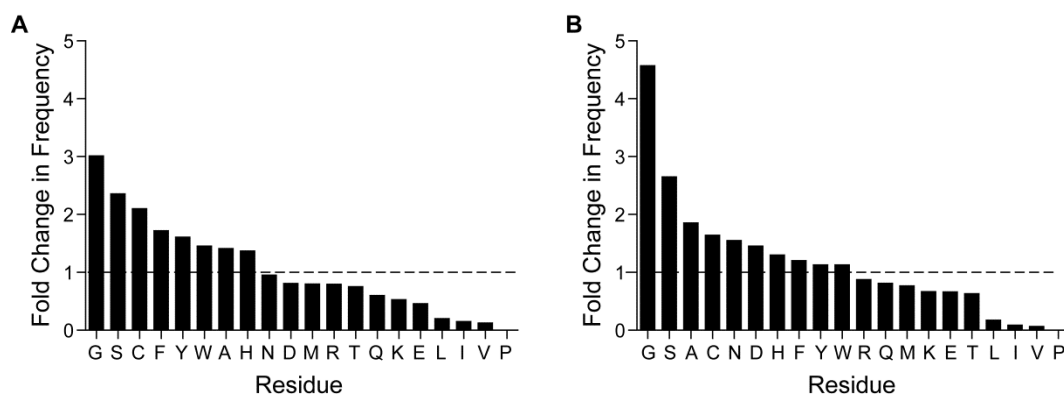


Figure 9.15 Relative frequency of amino acid residue-type engaged in C5 hydrogen bonds.

C5 Hydrogen bonds are identified in residues having $|\phi| > 150^\circ$ and $|\psi| > 150^\circ$, and are shown relative to their frequency in (A) the entire dataset, or (B) β -sheets.

Moreover, residues engaged in C5 hydrogen bonds are significantly enriched in small amino acids, most notably glycine (Figure 9.15), polymers of which have a high propensity for aggregation.⁴⁶⁸ Conversely, β -branched residues, which are abundant in β -sheets,⁴⁶⁹ are unlikely to engage in C5 hydrogen bonds. For example, whereas threonine is significantly more common in β -sheets than is serine, serine is much more likely to engage in C5 hydrogen bonds. Hence, evolution might have selected against the potentially deleterious effects of C5 hydrogen bonds in β -sheets by replacing small amino acids with bulky ones.

Conclusions

Our data indicate that C5 hydrogen bonds are common in folded proteins and can affect the stability and chemical properties of proteins. They likely compliment other noncanonical interactions in protein structure such as $C^\alpha-H\cdots O$ hydrogen bonds,⁸¹ cation- π ,¹¹⁶ and $n\rightarrow\pi^*$ interactions.⁹² They might also explain the anomalously slow exchange rates of solvent-exposed β -sheet residues in proteins (*e.g.*, Staphylococcal nuclease⁴⁷⁰) that are not engaged in canonical hydrogen bonds. Our results suggest that C5 hydrogen bonds merit inclusion in computational force fields, which do not explicitly treat such interactions,⁴⁷¹ despite their sharing many properties with traditional hydrogen bonds. Accurate accounting could improve models of protein structure, stability, and folding.

Acknowledgements

We thank G. J. Bartlett, D. N. Woolfson, I. C. Tanrikulu and L. L. Kiessling for contributive discussions, W. M. Westler for assistance with NMR spectroscopy, and T. Zhang and M. T. Zanni for assistance with IR spectroscopy. This work was supported by grants CHE-1124944

(NSF) and R01 AR044276 (NIH). The National Magnetic Resonance Facility at Madison is supported by grant P41 GM103399 (NIH). High-performance computing is supported by grant CHE-0840494 (NSF). The Biophysics Instrumentation Facility at UW–Madison was established with grants BIR-9512577 (NSF) and S10 RR013790 (NIH). R.W.N. was supported by NIH Biotechnology Training Grant T32 GM008349 and by an ACS Division of Organic Chemistry Graduate Fellowship.

Methods

Instrumentation. Nuclear magnetic resonance (NMR) spectra for compound characterization were acquired at ambient temperature with an Avance III 500 MHz spectrometer (^{13}C , 125 MHz), H/D exchange experiments were performed with a DMX 400 MHz spectrometer, and NMR spectra of peptides were acquired with an Avance III 600 MHz spectrometer equipped with a 1.7-mm cryogenic probe, all from Bruker (Billerica, MA) in the National Magnetic Resonance Facility at Madison (NMRFAM). ^{13}C spectra were proton-decoupled. Mass spectrometry was performed with a Micromass LCT (electrospray ionization, ESI) instrument from Waters (Milford, MA) in the Mass Spectrometry Facility of the Department of Chemistry at the University of Wisconsin–Madison. Fourier transfer infrared (FTIR) spectra were collected with a Nicolet iS10 spectrometer from Thermo Fisher Scientific (Waltham, MA). Circular dichroism (CD) spectra were collected with a Model 420 spectrometer from Aviv Biomedical (Lakewood, NJ) in the Biophysics Instrumentation Facility at the University of Wisconsin–Madison. Peptide synthesis was performed with a Prelude automated synthesizer from Protein Technologies (Tucson, AZ) in the University of Wisconsin–Madison Biotechnology Center.

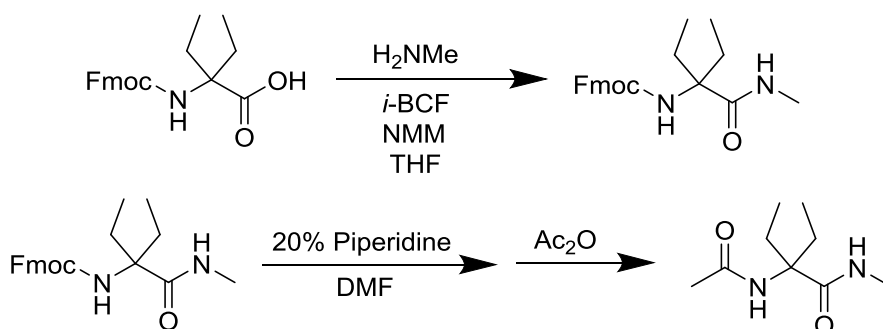
Peptide purification was accomplished with an LC-20 HPLC instrument from Shimadzu (Kyoto, Japan). Peptide characterization was performed with an LCMS 2020 instrument from Shimadzu.

Computational chemistry. All computations were performed using Gaussian 09³⁷⁰ and the B3LYP/6-311G+(2d,p) level of theory. A focused library of AcGlyNHMe conformations was constructed by optimizing the geometry of the compound for fixed values of the putative hydrogen-bond donor–acceptor distance (NH \cdots O), ranging from $d = 1.974$ to 2.974 Å. This range was selected to sample distances observed in folded proteins. Optimized geometries were subjected to frequency analysis, gauge-independent atomic orbital NMR calculations, and analysis by NBO 5.9 as implemented in Gaussian 09.³⁷¹ All energies were corrected by the zero-point vibrational energy. Similarly, optimized conformations of AcDegNHMe and AcDegOMe, where “Deg” refers to diethylglycine, were obtained at the same level of theory. Frequency calculations yielded no imaginary frequencies, indicating a true stationary point on the potential energy surface. Energies were corrected by the zero-point vibrational energy.

Bioinformatics. A nonredundant set (<25% pairwise sequence identity) of 192 protein crystal structures (>40 residues, $R < 20\%$) with a resolution of 1.0 Å or better was culled from the PDB of 28 November 2012 using the PISCES server,⁴³¹ as was a similar set of 1,884 crystal structures determined to a resolution of 1.6 Å or better. Secondary structure assignments were made using Kabsch and Sander criteria as implemented with the program PROMOTIF.⁴³³ Relative backbone solvent-accessible surface areas were calculated with the program NACCESS.⁴⁷² Residues with backbone atoms modeled in multiple conformations were excluded from analysis. Parallel and antiparallel β -sheets were identified as those forming only parallel or antiparallel contacts,

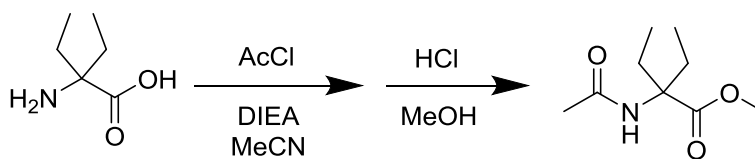
respectively. Strands engaged in mixed contacts were excluded from analysis for the purposes of comparing C5 frequencies between parallel and antiparallel β -sheets.

Synthesis of diethylglycine derivatives (General methods). Commercial chemicals were of reagent grade or better, and were used without further purification. Amino acid starting materials and peptide resins were obtained from Chem-Impex International. All other reagents were obtained from Sigma–Aldrich. Anhydrous THF and DMF were obtained from CYCLE-TAINER solvent-delivery systems from J. T. Baker (Phillipsburg, NJ). Reactions were monitored by thin-layer chromatography with visualization by UV light or staining with KMnO_4 or ninhydrin. Chromatography was performed with columns of silica gel 60, 230–400 mesh (Silicycle, Québec City, Canada). The removal of solvents and other volatile materials “under reduced pressure” refers to the use of a rotary evaporator at water-aspirator pressure (<20 torr) and a water bath of <45 °C.



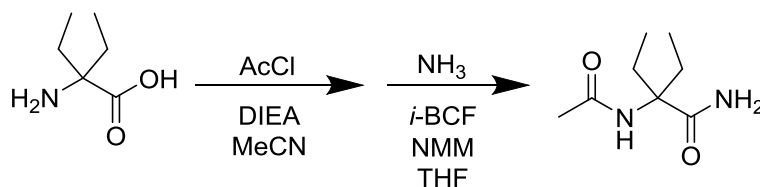
Synthesis of *N*-acetyldiethylglycine methyl amide (AcDegNHMe). Fmoc-Diethylglycine (1.0 g, 2.8 mmol) was dissolved in 50 mL of anhydrous THF containing NMM (0.62 mL, 5.6 mmol). The mixture was cooled to -20 °C and isobutyl chloroformate (0.37 mL, 2.8 mmol) was added dropwise. After 10 min, 3 mL of 2 M methylamine in THF was added, and the

reaction mixture was stirred overnight, allowing warmth to room temperature. The solution was filtered, and solvent was removed under reduced pressure. The residue was dissolved in EtOAc and washed with 1 M aqueous KH_2PO_4 , saturated aqueous NaHCO_3 , and brine. The organic portion was dried over anhydrous $\text{NaSO}_4(\text{s})$, and solvent was removed under reduced pressure. The residue was purified by chromatography on silica gel using an eluent of 2% v/v MeOH in DCM containing 1% v/v TEA, which afforded FmocDegNHMe as a white powder. The solid was dissolved in 20% v/v piperidine in DMF, and this solution was stirred for 1 h, after which the solvent was removed under reduced pressure. The residue was then dissolved in acetic anhydride and stirred overnight. After removing the solvent under reduced pressure, chromatography on silica gel using an eluent of 3% v/v MeOH in DCM afforded AcDegNHMe as a crystalline solid. ^1H NMR (CDCl_3) δ 6.91 (s, 1H), 6.33 (s, 1H), 2.88 (d, $J = 4.5$ Hz, 3H), 2.59 (dq, $J = 14.0, 7.4$ Hz, 2H), 2.03 (s, 3H), 1.55 (dq, $J = 14.0, 7.4$ Hz, 2H), 0.76 (t, $J = 7.4$ Hz, 6H); ^{13}C NMR δ 173.7, 169.2, 65.3, 28.8, 26.7, 24.3, 8.3; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 187.1442, observed 187.1439.



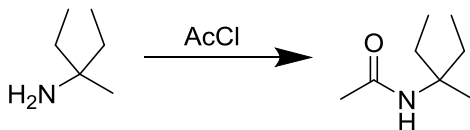
Synthesis of *N*-acetyldiethylglycine methyl ester (AcDegOMe). Diethylglycine (1.0 g, 7.6 mmol) was suspended in acetonitrile with TEA (2.1 mL, 15 mmol). Acetyl chloride (0.54 mL, 7.6 mmol) was added dropwise, and the reaction mixture was stirred at room temperature overnight. The solid was collected by filtration and purified by chromatography on silica gel in 12% v/v MeOH in DCM with 1% v/v TFA to remove residual starting material. The resulting solid was dissolved in MeOH, acidified with concentrated HCl, and heated at reflux for

1 h. After removing the solvent under reduced pressure, the residue was purified by chromatography on silica gel using an eluent of 4% v/v MeOH in DCM to afford AcDegOMe as a crystalline solid. ^1H NMR (CDCl_3) δ 6.36 (s, 1H), 3.78 (s, 3H), 2.48 (dq, $J = 14.0, 7.4$ Hz, 2H), 2.03 (s, 3H), 1.77 (dq, $J = 14.0, 7.4$ Hz, 2H), 0.74 (t, $J = 7.4$ Hz, 6H); ^{13}C NMR δ 174.8, 168.9, 66.3, 52.8, 27.9, 24.2, 8.6; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 188.1282, observed 188.1276.



Synthesis of *N*-acetyldiethylglycine amide (AcDegNH₂). Diethylglycine (1.0 g, 7.6 mmol) was suspended in acetonitrile with TEA (2.1 mL, 15 mmol). Acetyl chloride (0.54 mL, 7.6 mmol) was added dropwise, and the reaction mixture was stirred at room temperature overnight. The solid was collected by filtration and purified by chromatography of silica gel using an eluent of 12% v/v MeOH in DCM containing 1% v/v TFA. The resulting solid (0.15 g, 0.9 mmol) was dissolved in anhydrous THF containing NMM (0.19 mL, 1.8 mmol), and this solution was cooled to -20 °C. Isobutyl chloroformate (0.11 mL, 0.9 mmol) was added dropwise. After 10 min, 3 mL of 7 N ammonia in MeOH was added, and the mixture was stirred overnight, allowing warmth to room temperature. The solution was filtered, and solvent removed under reduced pressure. The residue was purified by reverse-phase HPLC (A = 0.1% v/v TFA in H₂O; B = 0.1% v/v TFA in MeCN) on a preparative C18 column using a linear gradient of B (5–95% v/v) over 30 min. Lyophilization yielded AcDegNH₂ as a white powder. ^1H NMR (CDCl_3) δ 6.77 (s, 1H), 6.05 (s, 1H), 5.79 (s, 1H), 2.60 (dq, $J = 14.0, 7.4$ Hz, 2H), 2.02 (s, 3H), 1.56 (dq, J

= 14.0, 7.4 Hz, 2H), 0.80 (t, J = 7.4 Hz, 6H); ^{13}C NMR δ 175.7, 169.5, 65.4, 28.8, 24.3, 8.3; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 173.1285, observed 173.1286.



Synthesis of 3-acetamido-3-methylpentane. 3-Amino-3-methylpentane (0.10 g, 1.0 mmol) was dissolved in acetyl chloride, and the reaction mixture was stirred for 1 h. The solution was diluted with DCM and washed with water, 1 M HCl, and saturated aqueous NaHCO_3 . The organic portion was dried over anhydrous $\text{NaSO}_4(\text{s})$. Removal of solvent under reduced pressure afforded AcDegMe as a colorless solid. ^1H NMR (CDCl_3) δ 5.01 (s, 1H), 1.93 (s, 3H), 1.77 (dq, J = 14.0, 7.4 Hz, 2H), 1.62 (dq, J = 14.0, 7.4 Hz, 2H), 1.20 (s, 3H), 0.82 (t, J = 7.4 Hz, 6H); ^{13}C NMR δ 169.4, 56.7, 30.3, 24.5, 23.3, 8.0; ESI-MS: $[\text{M} + \text{H}]^+$ calculated 144.1383, observed 144.1385.

Fourier transform infrared spectroscopy. FTIR spectra were acquired by dissolving diethylglycine derivatives in CDCl_3 to a final concentration of 10 mM. Following a purge with $\text{N}_2(\text{g})$, 256 scans between 400 and 4000 cm^{-1} were acquired and averaged, and the absorbance was calculated relative to background from neat CDCl_3 .

Hydrogen–deuterium exchange. Hydrogen–deuterium (H/D) exchange experiments were performed by co-dissolving AcDegNHMe and AcDegOMe, or AcGlyNHMe and AcGlyOMe in $\text{DMSO}-d_6$ to a final concentration of 50 mM each. Aliquots (0.50 mL) of the resulting solutions were transferred to NMR tubes. Exchange was initiated by the addition of 10 μL of D_2O at time

$t = 0$. Samples were mixed thoroughly by repeated inversion for 30 s before collection of the first spectrum. ^1H NMR spectra were collected by averaging 16 individual scans to provide adequate signal to noise. Integrations were determined from the area of the calculated fit of the amide region (7.2–8.0 ppm for diethylglycines or 7.6–8.4 ppm for glycines), as determined with the program Mestrenova 9.0 from MestreLab Research (Escondido, CA). Experiments were performed in triplicate.

Synthesis of TrpZip peptides. α -Hydroxy acids were prepared as described previously.⁴⁷³ Ester linkages were forged in solution by DIC/DMAP activation to yield dimers suitable for solid-phase synthesis. Solid-phase peptide synthesis was conducted with Fmoc chemistry on TentaGel S RAM resin, except for TrpZip2-F, which was synthesized on TentaGel HMBA resin. Amide-bond formation was accomplished by activation of 4 equiv of amino acid with 4 equiv of HCTU in the presence of 8 equiv of NMM for 1 h. N-Terminal acetylation was achieved by treatment with acetic anhydride or trifluoroacetic anhydride, as required. Release and global deprotection were performed by treatment of resin with a solution of 90% v/v TFA, 5% v/v phenol, 2.5% v/v TIPS, and 2.5% v/v H_2O for 3 h. For TrpZip2-F, release and global deprotection were performed according to procedures reported previously.⁴⁷⁴ Following precipitation in anhydrous diethyl ether, peptides were isolated by centrifugation and dissolved in 25% v/v MeCN in H_2O . Peptides were purified by reversed-phase HPLC on a preparative NucleoSil C18 column from Macherey–Nagel (Bethlehem, PA) using linear gradient of 25–50% v/v B over 45 min (A: 0.1% v/v TFA in H_2O ; B: 0.1% v/v TFA in MeCN) followed by lyophilization to yield white powders. Purified peptides were analyzed by LC–MS using an analytical Supelco Discovery BIO Wide Pore C5-5

column from Sigma–Aldrich with a linear gradient of 5–95% v/v B over 15 min (A: 0.1% v/v formic acid in H₂O; B: 0.1% v/v formic acid in MeCN).

NMR spectroscopy of peptides. As described previously,⁴⁶² TrpZip peptides were dissolved to a final concentration of ~1 mM in 40 μ L of 20 mM potassium phosphate buffer, pH 7.0, containing 10% v/v D₂O. Homonuclear ROESY, COSY, and TOCSY spectra were collected at 25 °C with water suppression, and resonances were assigned by standard methods. Specifically, sequential H ^{α} _{*i*}–NH_{*i*+1} ROESY correlations provided connectivities and confirmed the β -like backbone geometry.

Circular dichroism spectroscopy. As described previously,⁴⁶² far-UV CD spectra of TrpZip peptides were obtained at 20 μ M in 20 mM potassium phosphate buffer, pH 7.0. Ellipticity was recorded in 1-nm increments in the 200–300 nm range with 1-nm bandwidth and 3 s averaging. Thermal denaturation was performed at a concentration of 20 μ M, except for TrpZip2-C and TrpZip2-D, which were denatured at 100 μ M; for comparison, tryptophan zippers have been demonstrated previously to exist as monomers in solution at concentrations into the millimolar range.⁴⁶² Ellipticity at 228 nm was recorded by averaging for 15 s with a 1-nm bandwidth. Temperature was increased from 4 to 96 °C in 1-°C steps with a 5-min equilibration between steps. The overall rate of temperature increase was 6–10 °C/h. Data for peptides that were well folded at 4 °C were converted to percent folded. Values of T_m , which is the temperature at the midpoint of the thermal transition between the folded and unfolded states, were determined as described previously.⁴⁷⁵ All ellipticity data fitted well to a two-state model ($R^2 > 0.999$), and the ensuing T_m values were independent of the experimentally determined concentration.

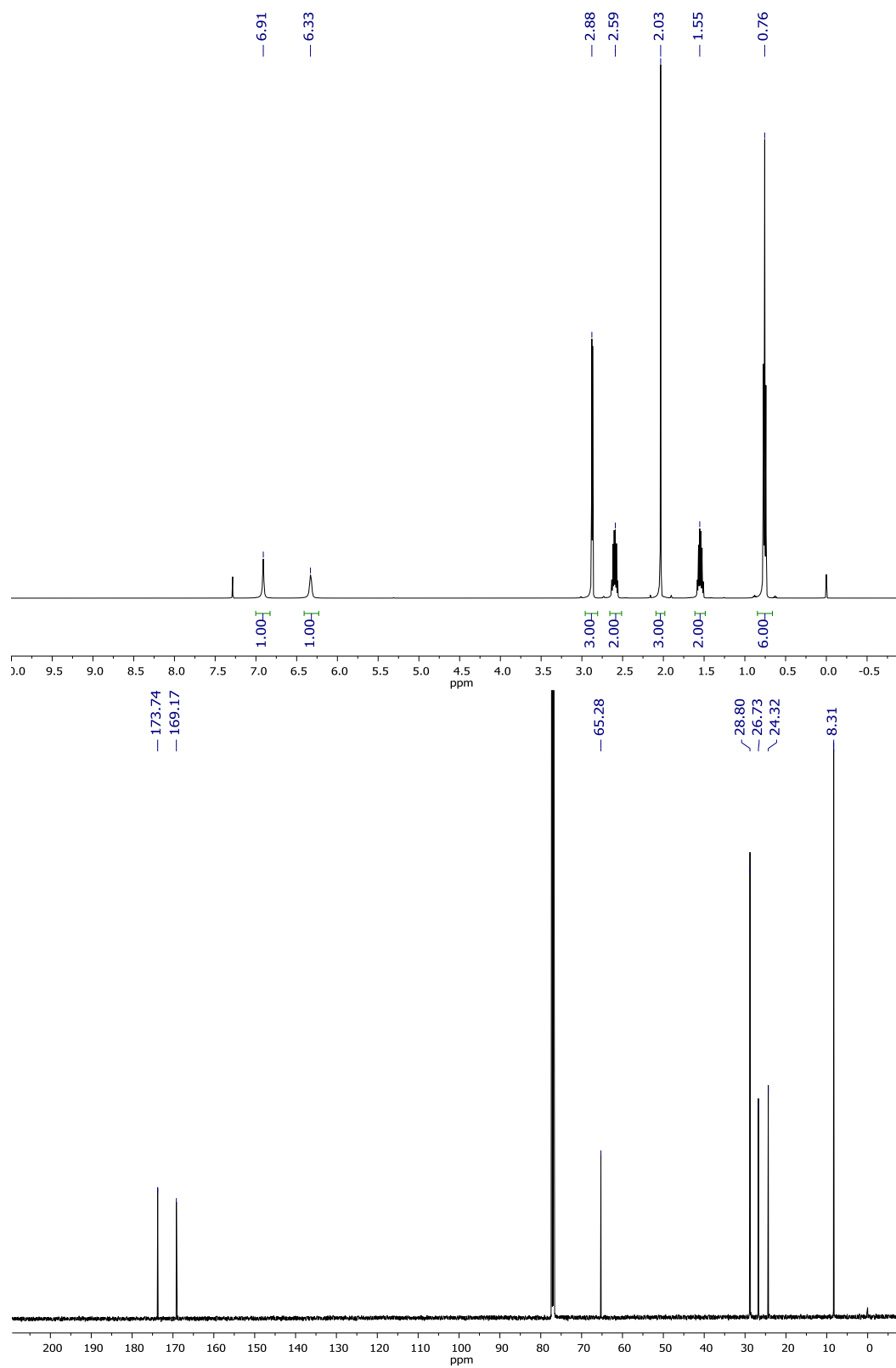


Figure 9.S1 ¹H and ¹³C NMR spectra of AcDegNHMe in CDCl₃.

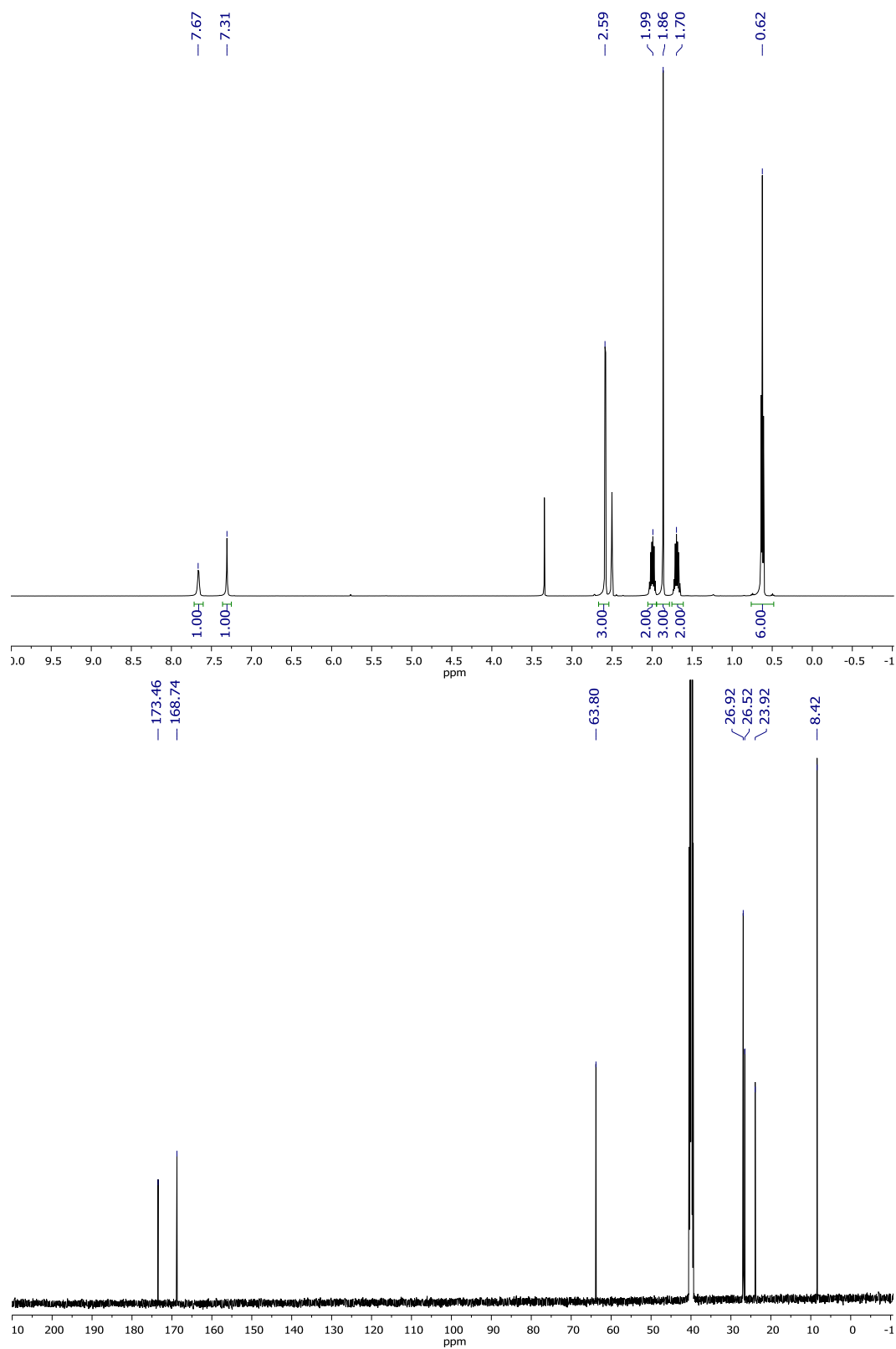


Figure 9.S2 ^1H and ^{13}C NMR spectra of AcDegNHMe in DMSO- d_6 .

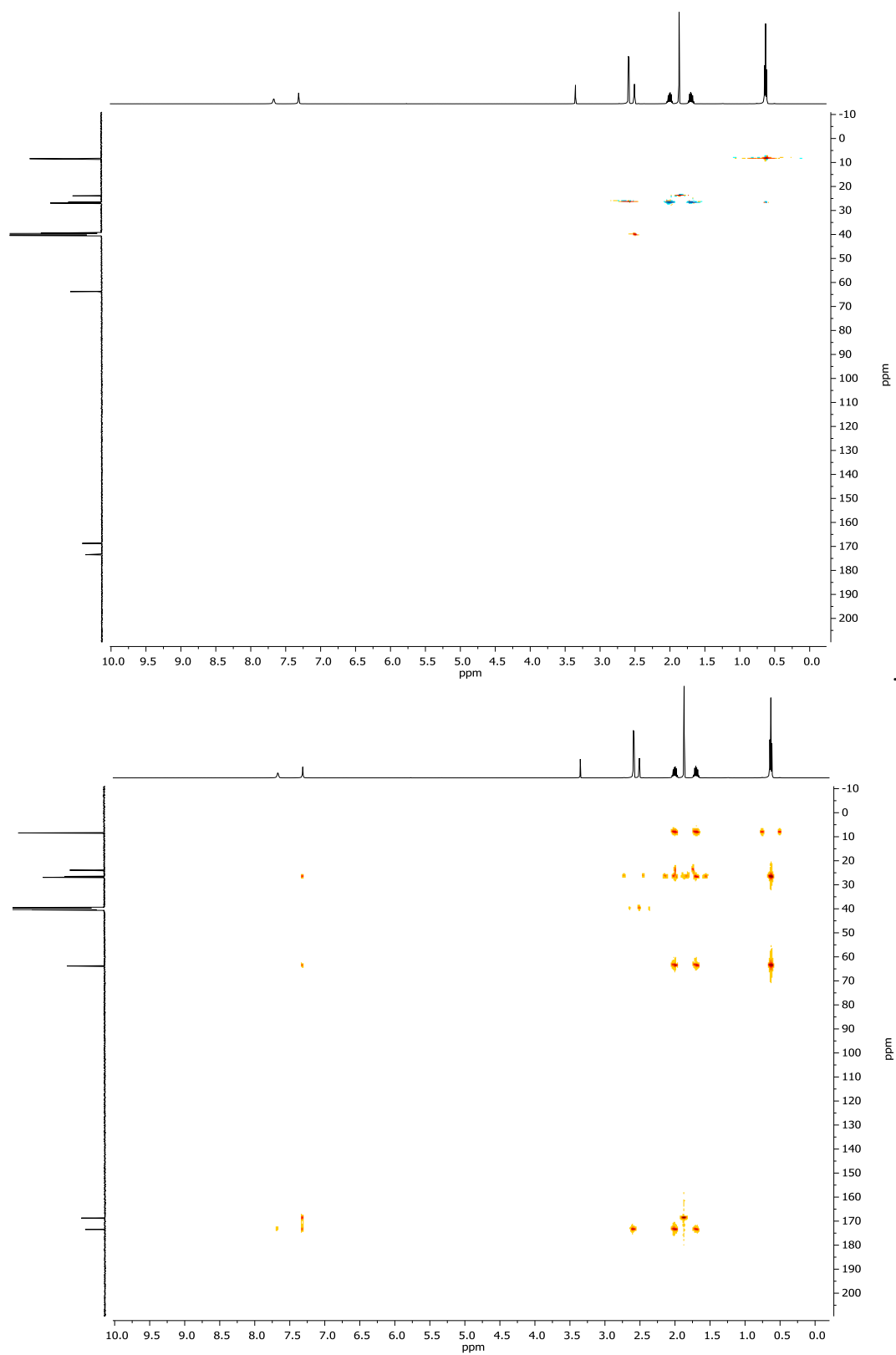


Figure 9.S3 ^1H - ^{13}C HSQC and HMBC spectra of AcDegNHMe in DMSO- d_6 .

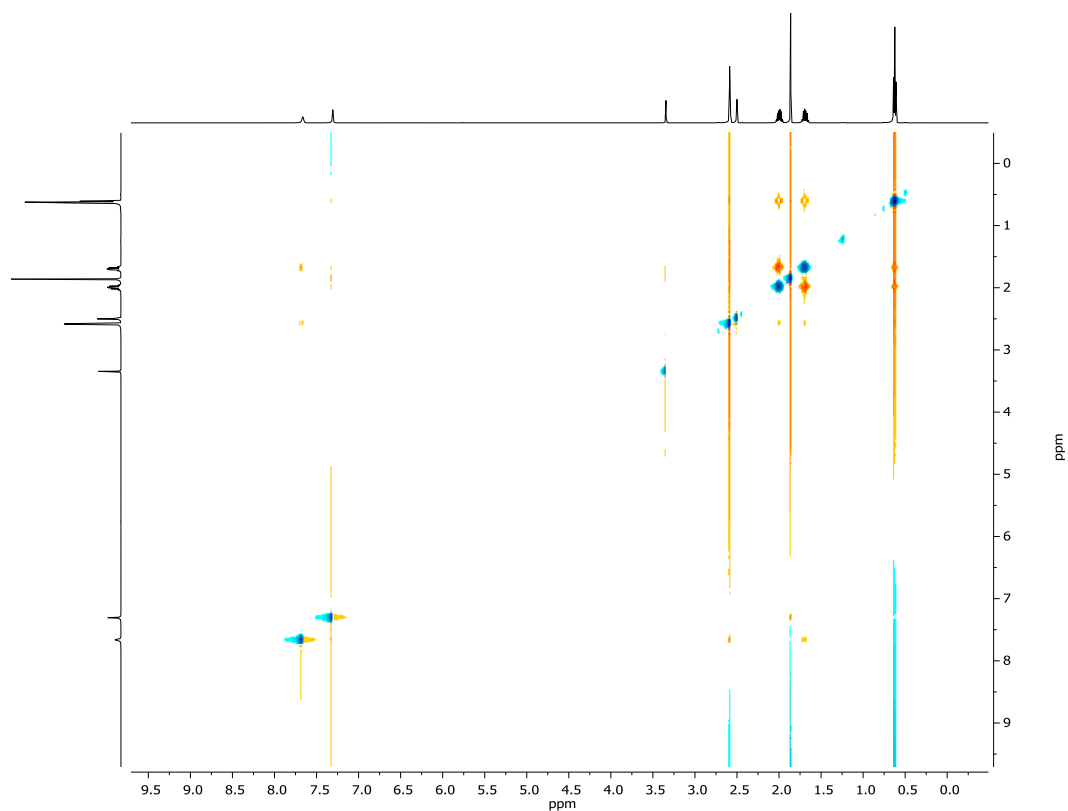


Figure 9.S4 ^1H - ^1H NOESY spectrum of AcDegNHMe in $\text{DMSO}-d_6$.

Table 9.S1 Cartesian coordinates of the optimized geometry of AcDegNHMe.

Corrected SCF Energy = -613.721928 Hartree.

C	-2.37544800	-0.00505700	0.25980000
O	-2.81007900	-0.01134900	-0.88575700
N	-1.05319900	-0.00027100	0.56445800
H	-0.75466900	0.00713000	1.53171700
C	0.05311900	0.00110400	-0.39137100
C	1.32649400	0.00652100	0.49660800
O	1.26463900	0.01587300	1.72048600
N	2.51152000	0.00799500	-0.15323100
H	2.52213800	-0.03697900	-1.15766600
C	3.77630700	-0.00622100	0.56602500
H	4.58634400	0.13950200	-0.14702500
H	3.79960200	0.79545000	1.30429200
H	3.92605800	-0.95321200	1.08942500
C	-3.31467300	-0.00272000	1.45394200
H	-2.80121900	0.00365800	2.41613600
H	-3.96048700	0.87409500	1.38610600
H	-3.95424200	-0.88472000	1.39450200
C	0.02054400	-1.26891000	-1.28361300
H	0.83455200	-1.20263200	-2.01285900
H	-0.90813300	-1.22696700	-1.85395700
C	0.10984800	-2.59565100	-0.53057300
H	-0.70894900	-2.70125500	0.18288500
H	0.04753000	-3.43119200	-1.23089200
H	1.04807800	-2.69502200	0.02142600
C	0.01433700	1.26764000	-1.28849900
H	-0.91299200	1.21786800	-1.86051100
H	0.83140500	1.20375400	-2.01415300
C	0.09409200	2.59748900	-0.54002600
H	0.02743200	3.42989900	-1.24368400
H	-0.72624900	2.70091300	0.17202600
H	1.03134200	2.70429200	0.01205900

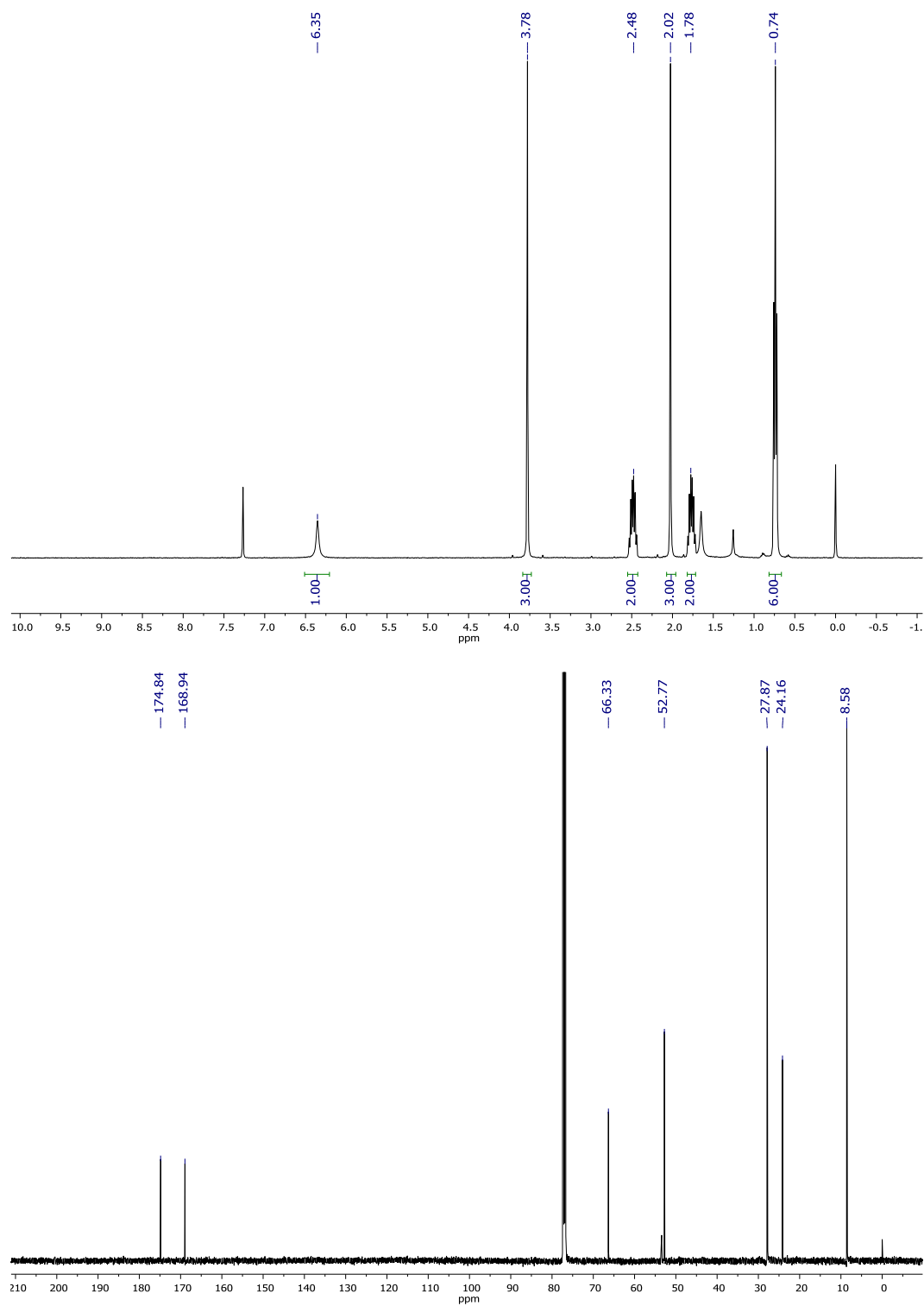


Figure 9.S5 ¹H and ¹³C NMR spectra of AcDegOMe in CDCl₃.

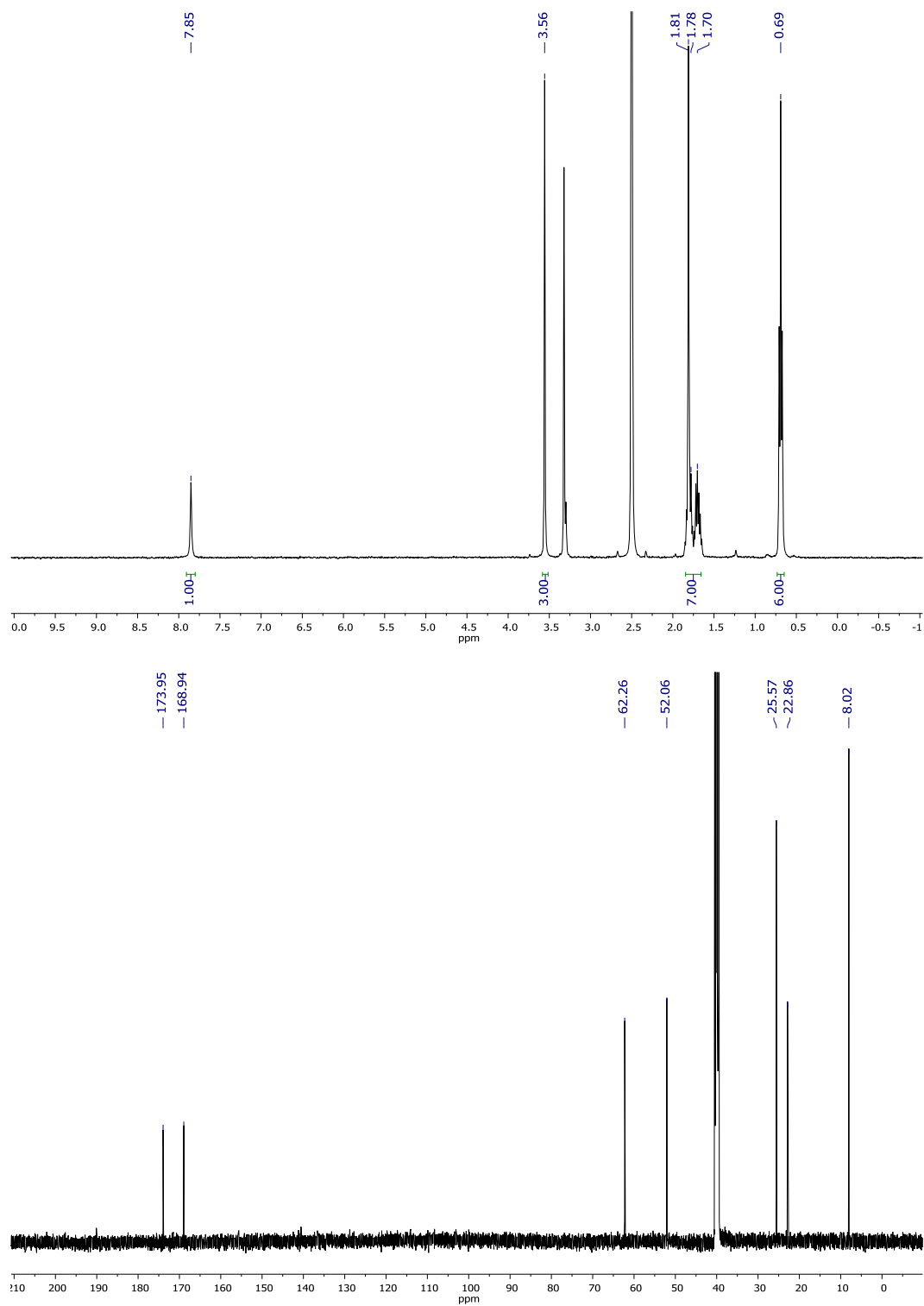


Figure 9.S6 ^1H and ^{13}C NMR spectra of AcDegOMe in DMSO- d_6 .

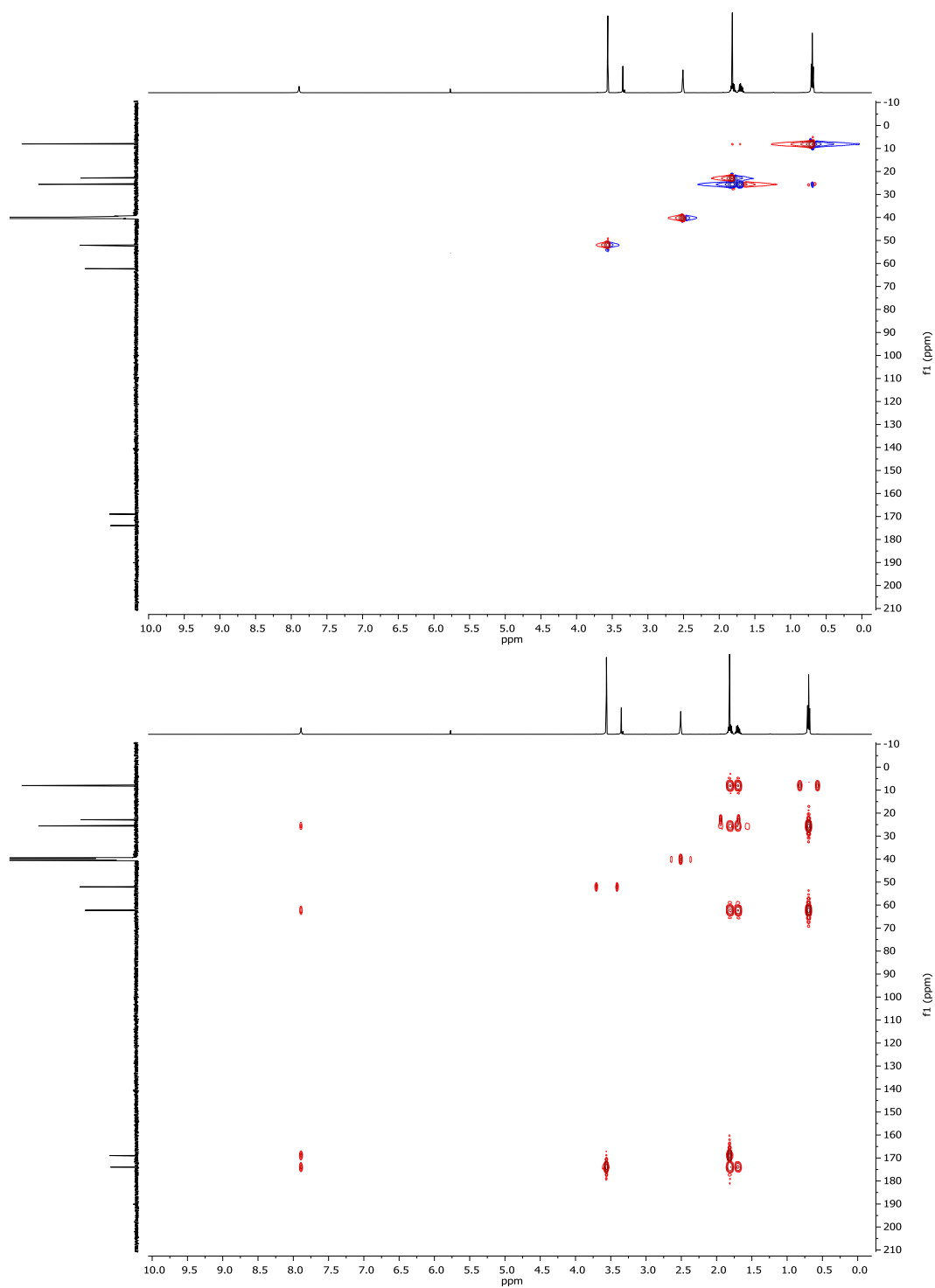


Figure 9.S7 ^1H - ^{13}C HSQC and HMBC spectra of AcDegOMe in DMSO- d_6 .

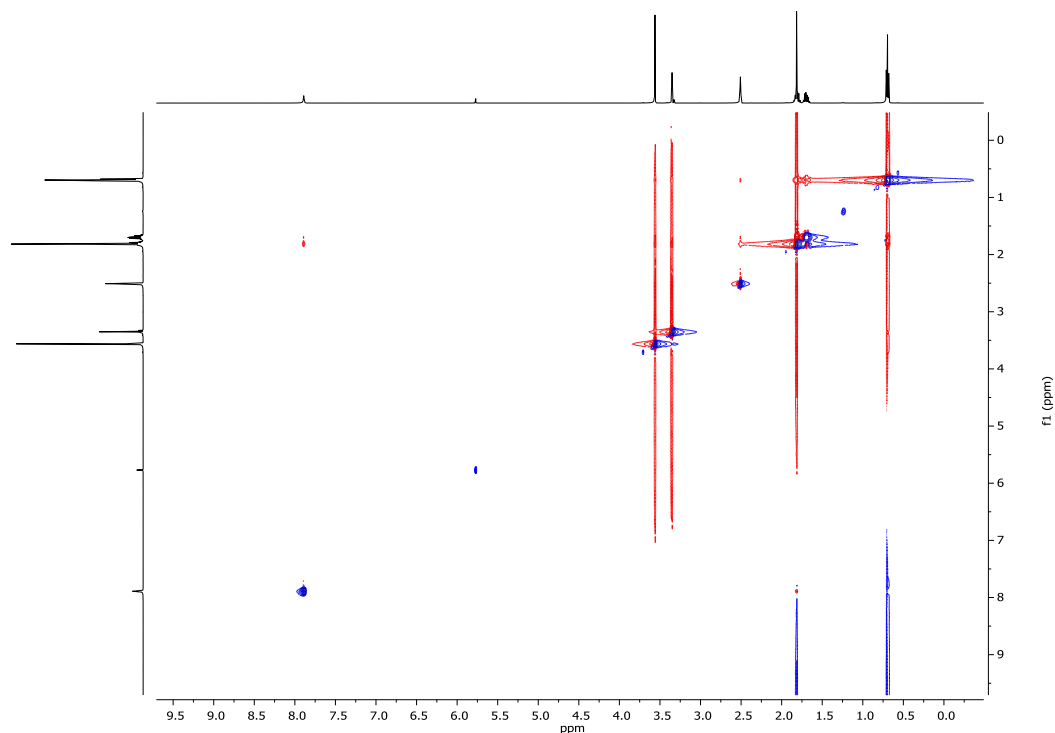


Figure 9.S8 ^1H - ^1H NOESY spectrum of AcDegOMe in $\text{DMSO-}d_6$.

Table 9.S2 Cartesian coordinates of the optimized geometry of AcDegOMe.
Corrected SCF Energy = -633.8604528 Hartree.

C	2.37371900	-0.00000500	0.23849700
O	2.78349200	-0.00002700	-0.91439500
N	1.05514700	-0.00000400	0.56915100
H	0.78976400	0.00001200	1.54362300
C	-0.06518500	-0.00001000	-0.37445000
C	-1.32210900	0.00000700	0.50696900
O	-1.31579400	0.00002100	1.71625000
C	-3.67891900	0.00002300	0.51797800
H	-4.46555200	0.00002100	-0.23154800
H	-3.74309700	-0.88846700	1.14491900
H	-3.74308400	0.88852700	1.14490000
C	3.33584500	0.00004000	1.41427900
H	2.84258300	-0.00007700	2.38711500
H	3.97711200	-0.87932100	1.33782500
H	3.97689300	0.87957300	1.33793500
C	-0.05969000	1.26748600	-1.27075900
H	-0.90636300	1.18951600	-1.95540700
H	0.85053300	1.22090000	-1.86921900
C	-0.12008900	2.59568800	-0.51737300
H	0.71952300	2.70487700	0.17139100
H	-0.07833900	3.42828900	-1.22255000
H	-1.04283500	2.70253200	0.05979700
C	-0.05970300	-1.26752700	-1.27072900
H	0.85051700	-1.22096000	-1.86919600
H	-0.90638000	-1.18956900	-1.95537400
C	-0.12010300	-2.59571200	-0.51731200
H	-0.07836400	-3.42832800	-1.22247100
H	0.71951500	-2.70488900	0.17144700
H	-1.04284500	-2.70253700	0.05986800
O	-2.44372200	0.00000600	-0.22431900

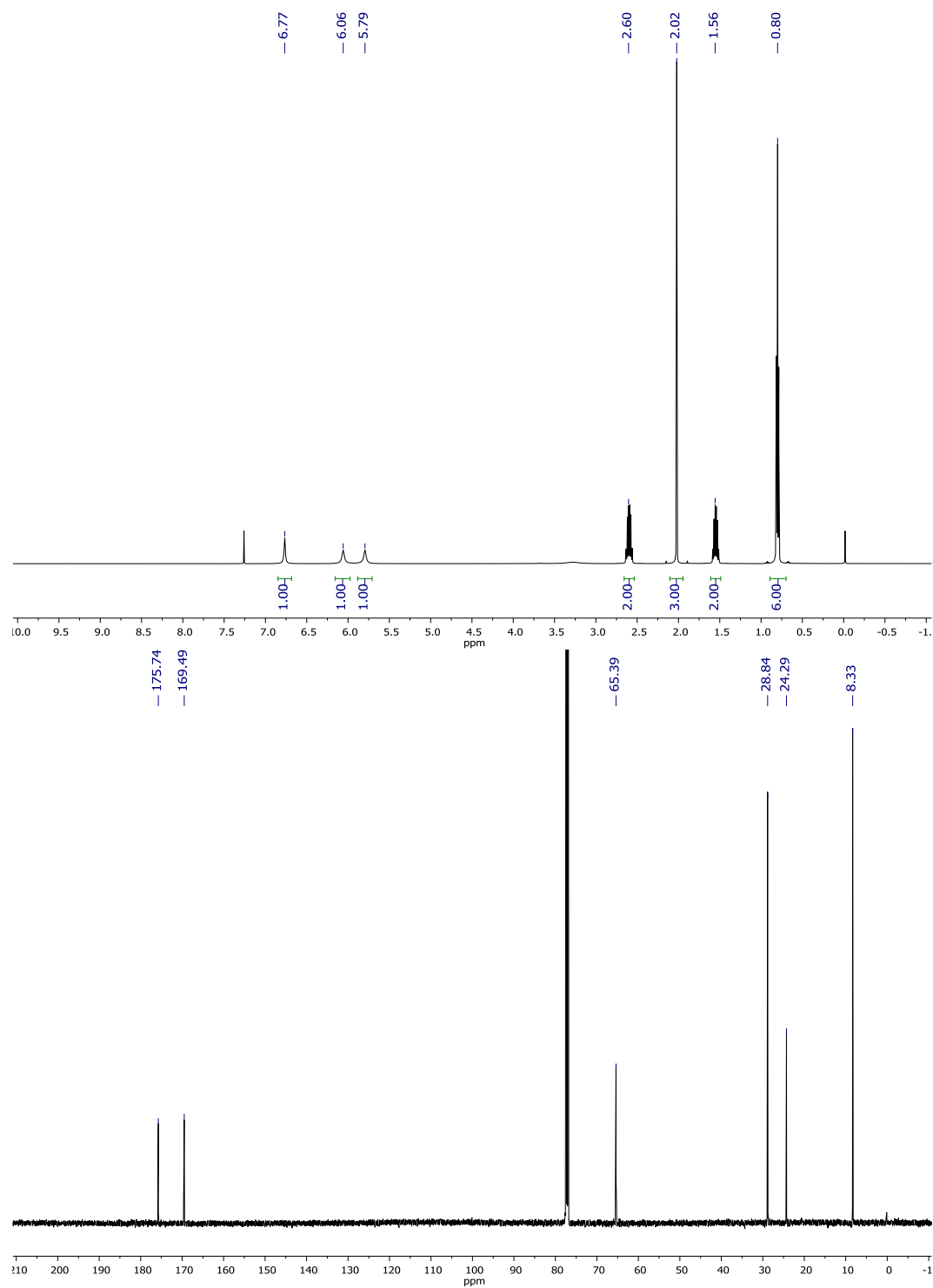


Figure 9.S9 ¹H and ¹³C NMR spectra of AcDegNH₂ in CDCl₃.

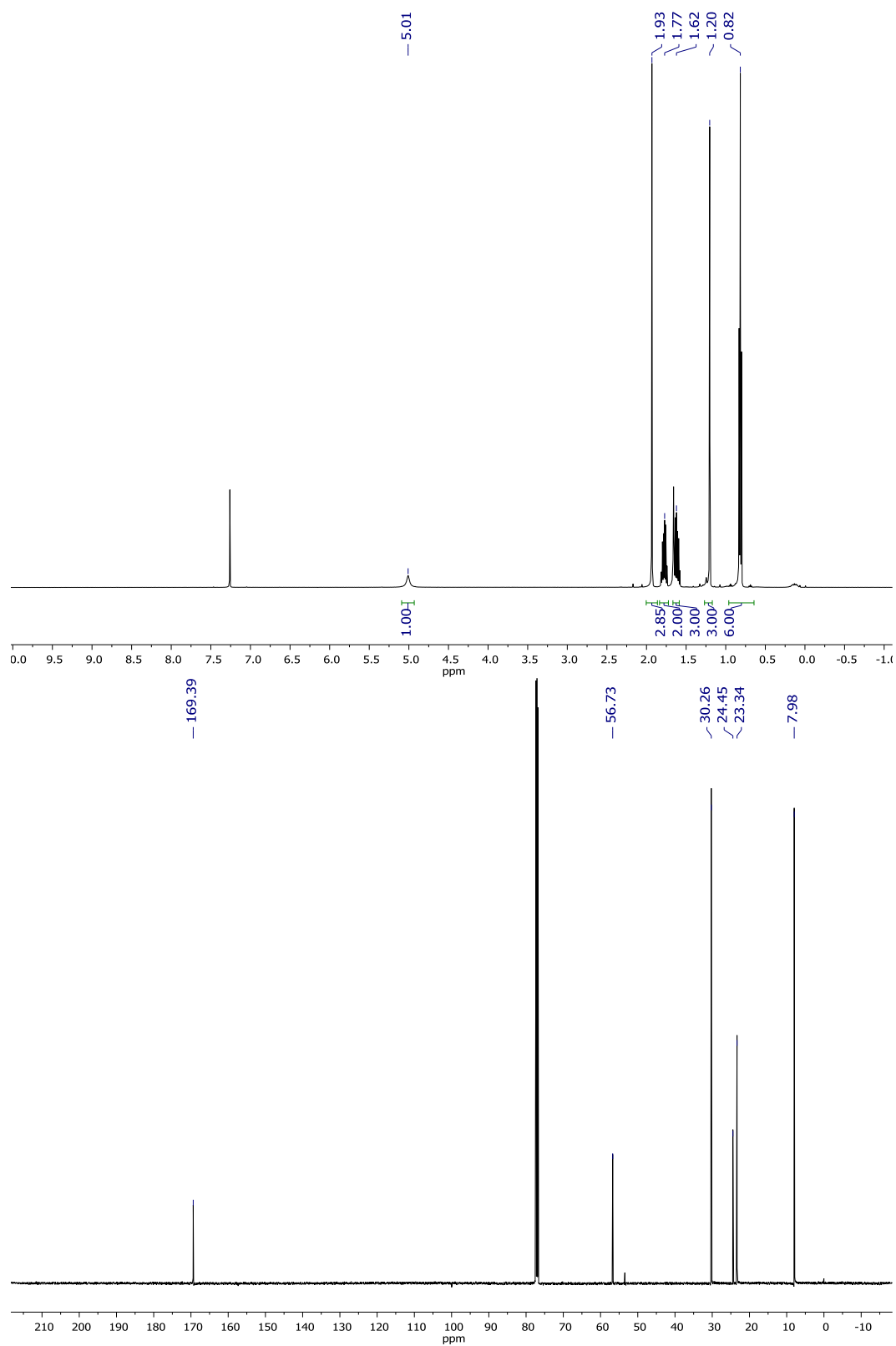


Figure 9.S10 ¹H and ¹³C NMR spectra of AcDegOMe in CDCl₃.

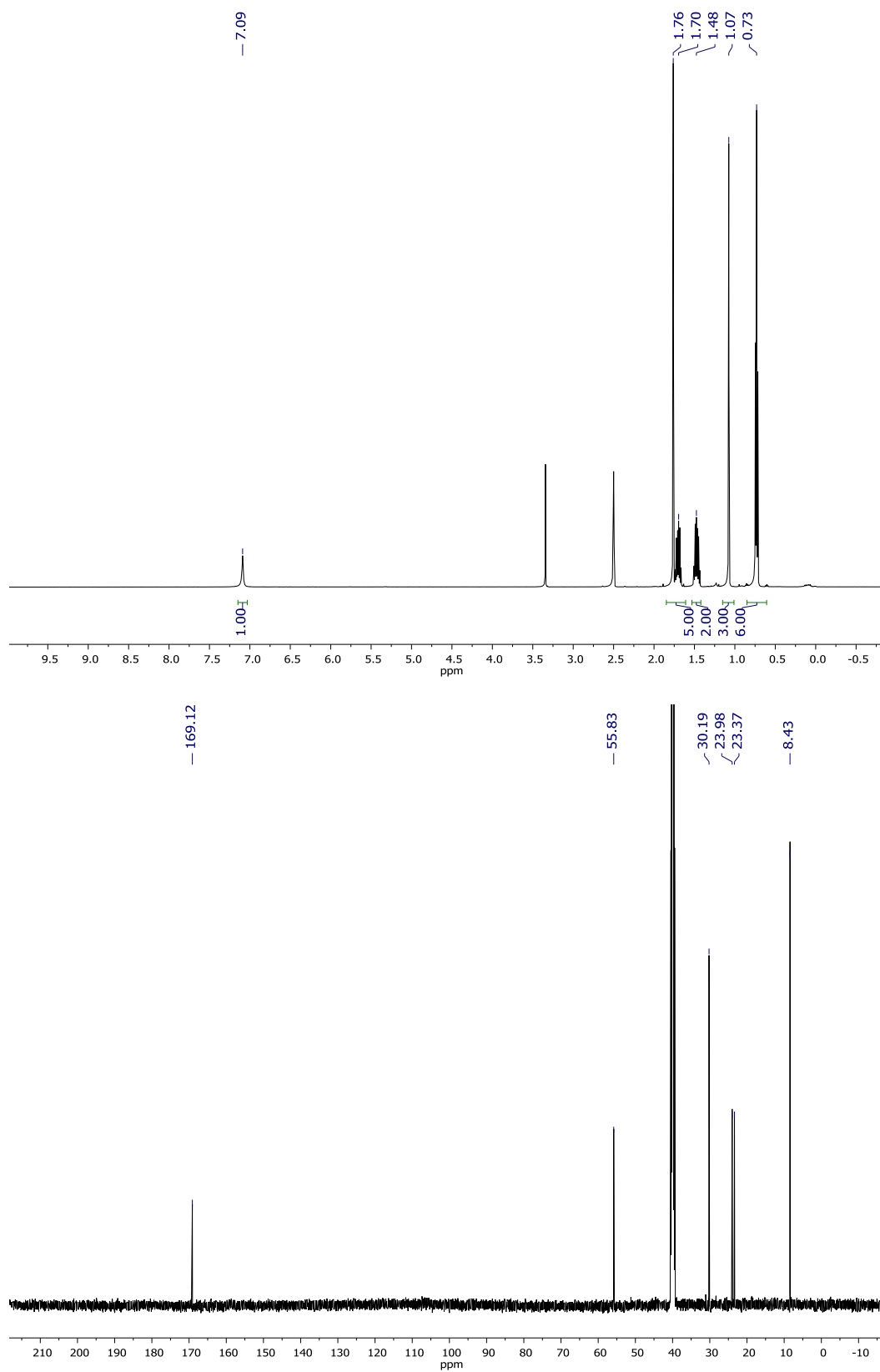


Figure 9.S11 ^1H and ^{13}C NMR spectra of AcDegOMe in DMSO- d_6 .

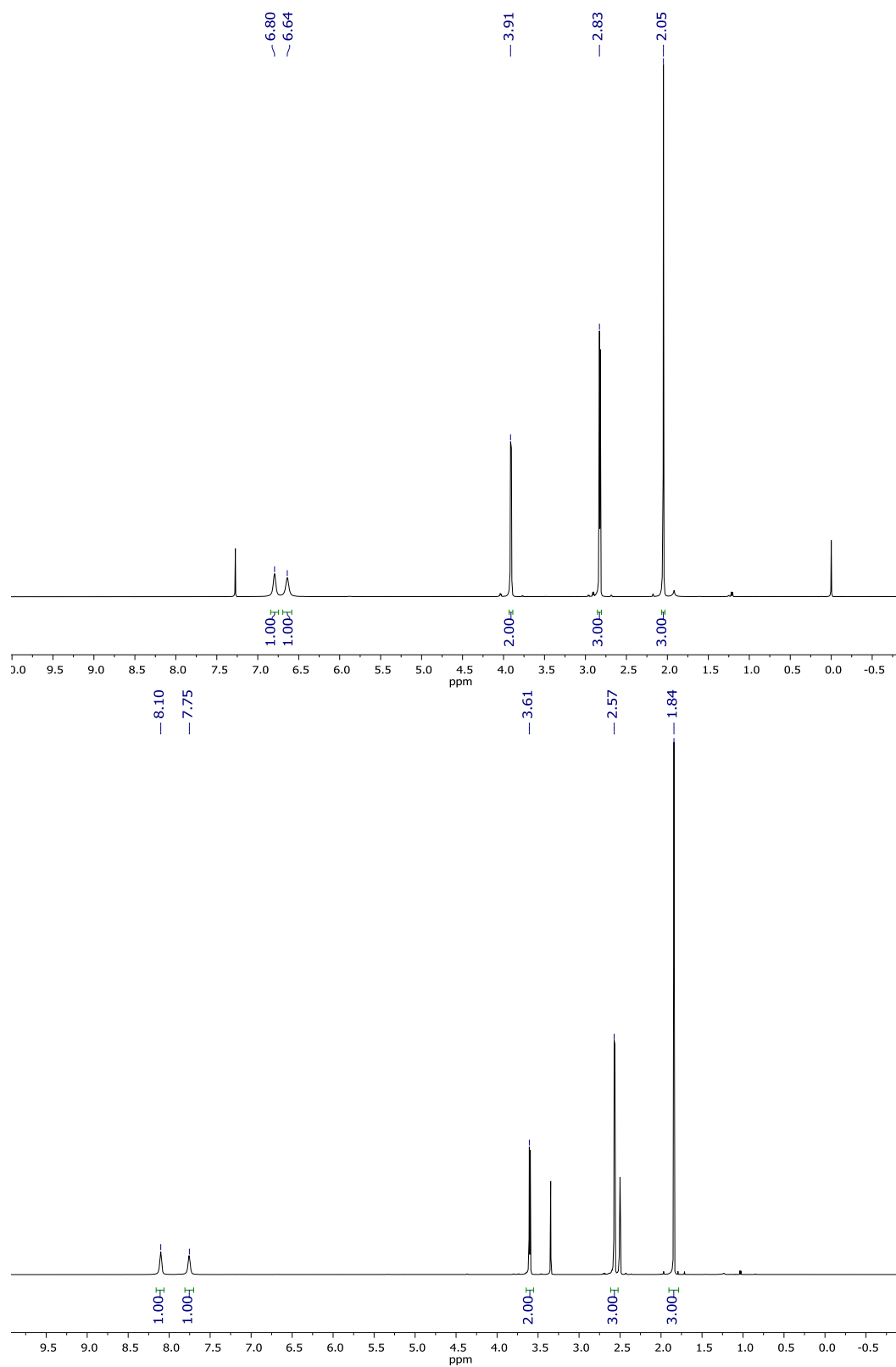


Figure 9.S12 ^1H NMR spectra of AcGlyNHMe in CDCl_3 and $\text{DMSO}-d_6$.

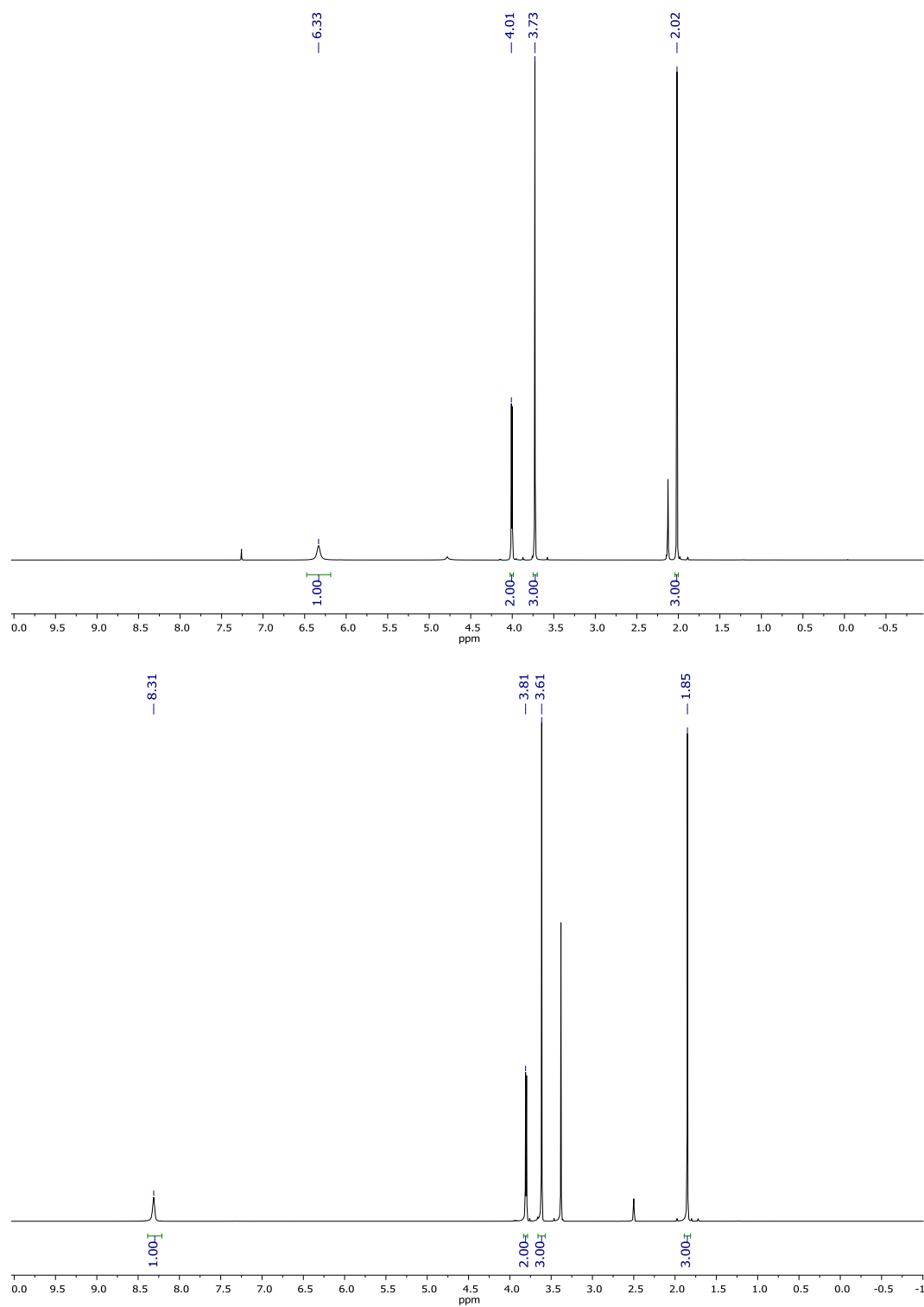


Figure 9.S13 ^1H NMR spectra of AcGlyOMe in CDCl_3 and $\text{DMSO}-d_6$.

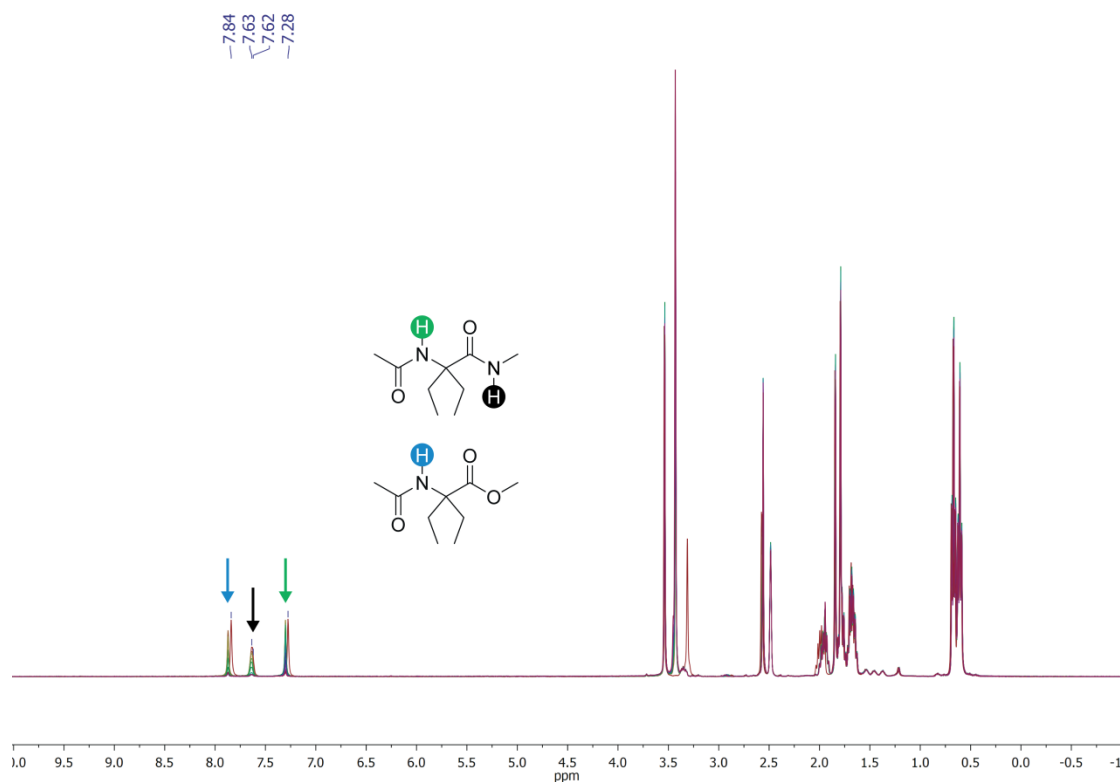


Figure 9.S14 Representative time-course for H/D exchange of AcDegNHMe and AcDegOMe. The overlaid ^1H NMR spectra were acquired during an H/D exchange experiment on an equimolar mixture of AcDegNHMe and AcDegOMe (50 mM) in $\text{DMSO-}d_6$ containing 2% v/v D_2O . This experiment was performed in triplicate, and the data were averaged to generate Figure 9.5C.

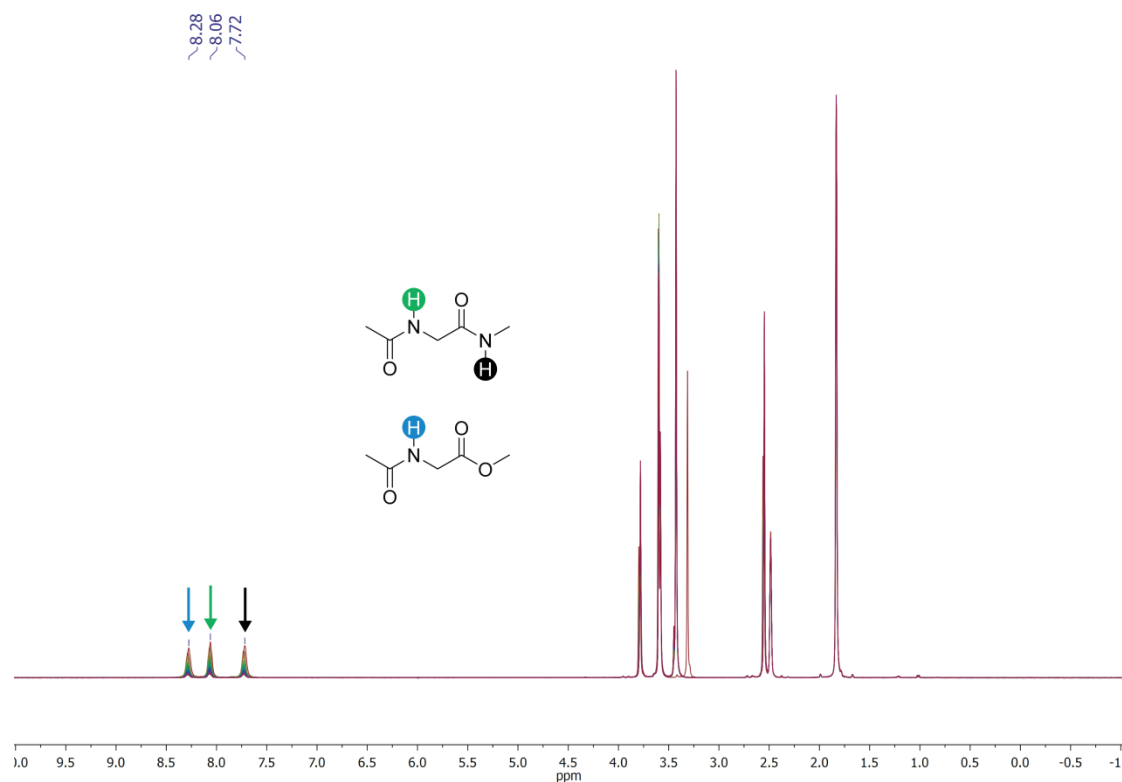


Figure 9.S15 Representative time-course for H/D exchange of AcGlyNHMe and AcGlyOMe. The overlaid ^1H NMR spectra were acquired during an H/D exchange experiment on an equimolar mixture of AcGlyNHMe and AcGlyOMe (50 mM) in $\text{DMSO-}d_6$ containing 2% v/v D_2O . This experiment was performed in triplicate, and the data were averaged to generate Figure 9.5D.

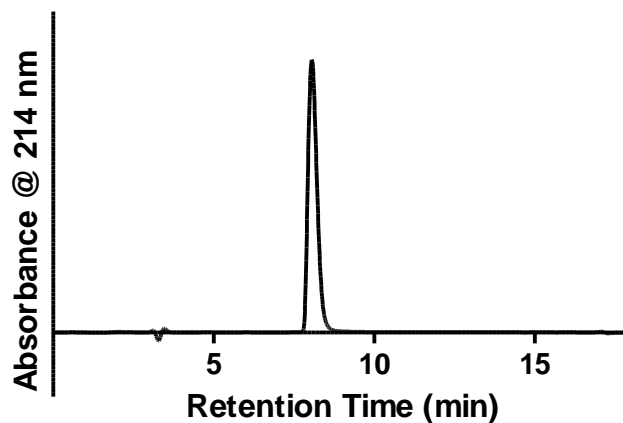


Figure 9.S16 Analytical LC–MS trace of purified TrpZip2-A. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{80}H_{105}N_{20}O_{19}$): 1649.78; observed 1649.55.

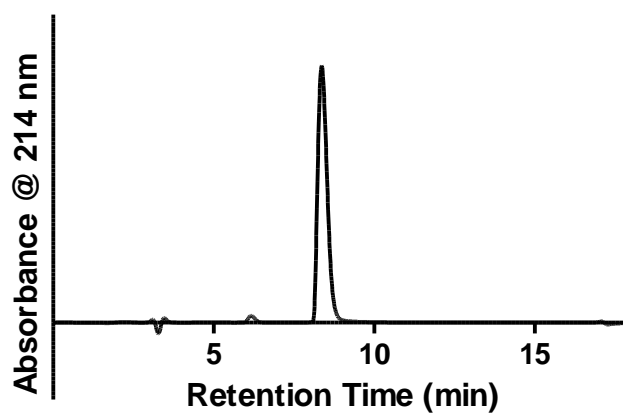


Figure 9.S17 Analytical LC–MS trace of purified TrpZip2-B. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{80}H_{105}N_{19}O_{20}$): 1650.76; observed 1650.55.

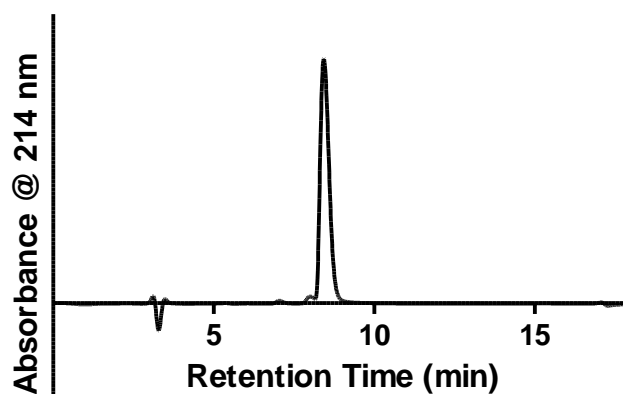


Figure 9.S18 Analytical LC–MS trace of purified TrpZip2-C. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{80}H_{105}N_{19}O_{20}$): 1650.76; observed 1650.50.

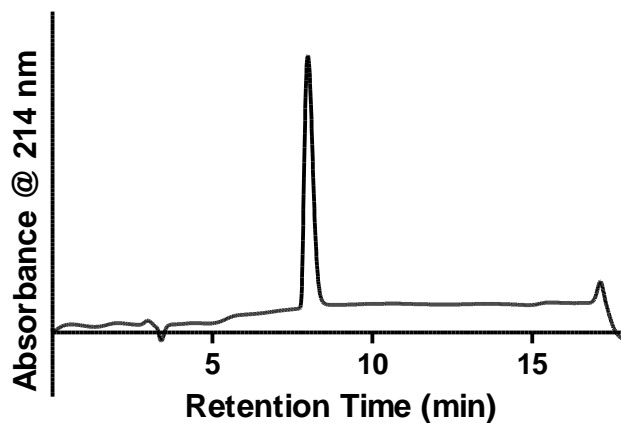


Figure 9.S19 Analytical LC-MS trace of purified TrpZip2-D. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{80}H_{105}N_{18}O_{21}$): 1651.75; observed 1651.50.

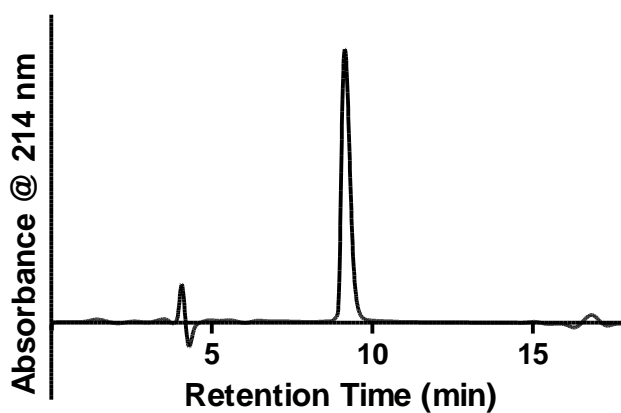


Figure 9.S20 Analytical LC-MS trace of purified TrpZip2-E. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{77}H_{100}N_{19}O_{17}$): 1562.75; observed 1562.55.

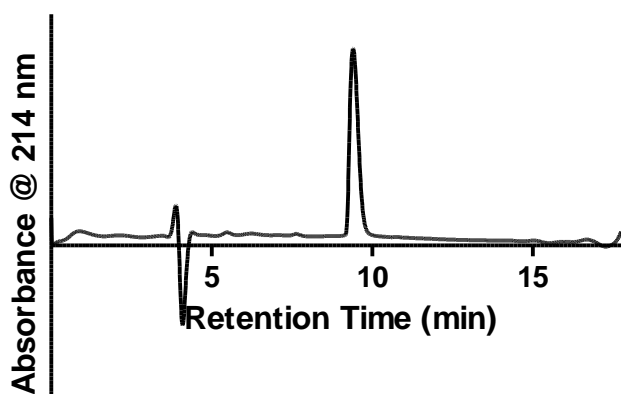


Figure 9.S21 Analytical LC-MS trace of purified TrpZip2-F. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{78}H_{101}N_{18}O_{18}$): 1577.75; observed 1577.60.

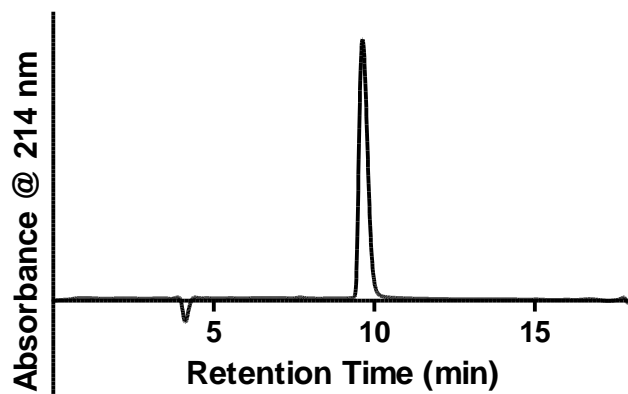


Figure 9.S22 Analytical LC-MS trace of purified TrpZip2-G. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{74}H_{93}N_{18}O_{18}$): 1521.68; observed 1521.60.

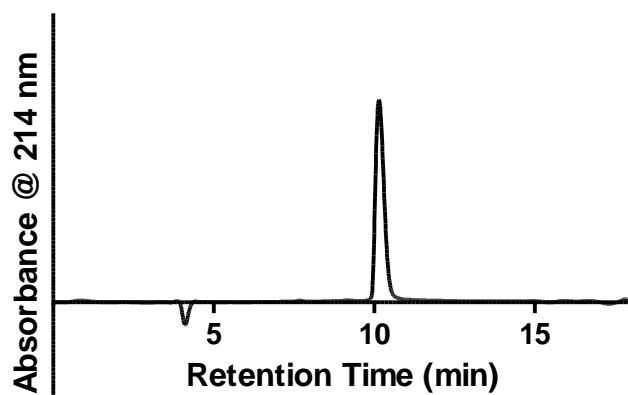


Figure 9.S23 Analytical LC-MS trace of purified TrpZip2-H. Calculated monoisotopic exact mass for $[M+H]^+$ ($C_{74}H_{90}N_{18}O_{18}F_3$): 1575.66; observed 1575.60.

Table 9.S3 ^1H Chemical shifts of TrpZip2-A.

Residue	Atom	δ
Ser1	H $^\alpha$	3.803
Ser1	H $^\beta$	3.429
Ser1	NH	7.343
Trp2	H $^\alpha$	4.986
Trp2	H $^{\beta 1}$	3.012
Trp2	H $^{\beta 2}$	2.818
Trp2	NH	8.190
Thr3	H $^\alpha$	4.703
Thr3	H $^\beta$	3.890
Thr3	H $^\gamma$	0.996
Thr3	NH	9.524
Trp4	H $^\alpha$	4.410
Trp4	H $^{\beta 1}$	2.671
Trp4	H $^{\beta 2}$	1.627
Trp4	NH	8.679
Glu5	H $^\alpha$	4.083
Glu5	H $^{\beta 1}$	1.683
Glu5	H $^{\beta 2}$	1.563
Glu5	H $^\gamma$	1.859
Glu5	NH	8.256
Asn6	H $^\alpha$	3.970
Asn6	H $^{\beta 1}$	2.726
Asn6	H $^{\beta 2}$	2.418
Asn6	NH	8.918
Gly7	H $^{\alpha 1}$	3.615
Gly7	H $^{\alpha 2}$	3.006
Gly7	NH	7.216
Lys8	H $^\alpha$	4.075
Lys8	H $^\delta$	1.479
Lys8	H $^\epsilon$	2.821
Lys8	H $^\gamma$	1.036
Lys8	NH	6.595
Trp9	H $^\alpha$	5.035
Trp9	H $^{\beta 1}$	3.030
Trp9	H $^{\beta 2}$	2.765
Trp9	NH	8.309
Thr10	H $^\alpha$	4.715
Thr10	H $^\beta$	3.850
Thr10	H $^\gamma$	1.011
Thr10	NH	9.614
Trp11	H $^\alpha$	4.415
Trp11	H $^{\beta 1}$	2.570
Trp11	H $^{\beta 2}$	1.542
Trp11	NH	8.570
Lys12	H $^\alpha$	3.871
Lys12	H $^\delta$	1.402
Lys12	H $^\gamma$	1.062
Lys12	NH	7.775

Table 9.S4 ^1H Chemical shifts of TrpZip2-B.

Residue	Atom	δ
Ser1	H $^\alpha$	4.102
Ser1	H $^\beta$	3.450
Ser1	NH	6.935
Trp2	H $^\alpha$	4.783
Trp2	H $^{\beta 1}$	2.696
Trp2	H $^{\beta 2}$	2.413
Trp2	NH	8.216
Thr3	H $^\alpha$	4.739
Thr3	H $^\beta$	4.143
Thr3	H $^\gamma$	0.899
Thr3	NH	8.790
Glu5	H $^\alpha$	4.210
Glu5	H $^{\beta 1}$	1.749
Glu5	H $^{\beta 2}$	1.587
Glu5	H $^\gamma$	1.873
Glu5	NH	8.596
Asn6	H $^\alpha$	4.138
Asn6	H $^{\beta 1}$	2.797
Asn6	H $^{\beta 2}$	2.525
Asn6	NH	9.090
Gly7	H $^{\alpha 1}$	3.016
Gly7	H $^{\alpha 2}$	2.577
Gly7	NH	5.511
Lys8	H $^\alpha$	4.340
Lys8	H $^\delta$	1.548
Lys8	H $^\epsilon$	2.842
Lys8	H $^\gamma$	1.144
Lys8	NH	7.216
Trp9	H $^\alpha$	4.651
Trp9	H $^{\beta 1}$	3.003
Trp9	H $^{\beta 2}$	2.573
Trp9	NH	8.630
Thr10	H $^\alpha$	4.519
Thr10	H $^\beta$	3.742
Thr10	H $^\gamma$	0.908
Thr10	NH	8.436
Trp11	H $^\alpha$	3.978
Trp11	H $^{\beta 1}$	2.514
Trp11	H $^{\beta 2}$	1.624
Trp11	NH	8.395
Lys12	H $^\alpha$	3.907
Lys12	H $^\delta$	1.253
Lys12	H $^\gamma$	0.972
Lys12	NH	7.793

Table 9.S5 Ranked changes in ^1H chemical shift from TrpZip2-A to TrpZip2-B

Residue	Atom	$\Delta\delta$ (ppm)
Gly7	NH	-1.705
Thr10	NH	-1.178
Thr3	NH	-0.734
Gly7	H $^{\alpha 1}$	-0.599
Trp11	H $^{\alpha}$	-0.437
Gly7	H $^{\alpha 2}$	-0.429
Ser1	NH	-0.408
Trp2	H $^{\beta 2}$	-0.405
Trp9	H $^{\alpha}$	-0.384
Trp2	H $^{\beta 1}$	-0.316
Trp2	H $^{\alpha}$	-0.203
Thr10	H $^{\alpha}$	-0.196
Trp9	H $^{\beta 2}$	-0.192
Trp11	NH	-0.175
Lys12	H $^{\delta}$	-0.149
Thr10	H $^{\beta}$	-0.108
Thr10	H $^{\gamma}$	-0.103
Thr3	H $^{\gamma}$	-0.097
Lys12	H $^{\gamma}$	-0.090
Trp11	H $^{\beta 1}$	-0.056
Trp9	H $^{\beta 1}$	-0.027
Glu5	H $^{\gamma}$	0.014
Lys12	NH	0.018
Lys8	H $^{\epsilon}$	0.021
Ser1	H $^{\beta}$	0.021
Glu5	H $^{\beta 2}$	0.024
Trp2	NH	0.026
Thr3	H $^{\alpha}$	0.036
Lys12	H $^{\alpha}$	0.036
Glu5	H $^{\beta 1}$	0.066
Lys8	H $^{\delta}$	0.069
Asn6	H $^{\beta 1}$	0.071
Trp11	H $^{\beta 2}$	0.082
Asn6	H $^{\beta 2}$	0.107
Lys8	H $^{\gamma}$	0.108
Glu5	H $^{\alpha}$	0.127
Asn6	H $^{\alpha}$	0.168
Asn6	NH	0.172
Thr3	H $^{\beta}$	0.253
Lys8	H $^{\alpha}$	0.265
Ser1	H $^{\alpha}$	0.299
Trp9	NH	0.321
Glu5	NH	0.340
Lys8	NH	0.621

Table 9.S6 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.974 \text{ \AA}$.

C	-1.82497500	-0.26278800	-0.06248700
O	-1.52832200	-1.21275700	-0.77410300
N	-1.16143000	0.93011400	-0.13304900
H	-1.24362800	1.57558200	0.63769100
C	0.04560600	1.05745800	-0.93668400
H	0.18323100	2.10360400	-1.20697800
H	-0.08441400	0.46259400	-1.83850200
C	1.25374100	0.62234900	-0.10779800
O	1.72189300	1.37129800	0.73717400
N	1.70493100	-0.63200500	-0.35573900
H	1.10359000	-1.22788800	-0.90705800
C	-2.97468900	-0.32611800	0.91989500
H	-3.09377400	0.58482600	1.50755400
H	-3.89570400	-0.51282500	0.36525700
H	-2.81736600	-1.17197800	1.59022200
C	2.74235900	-1.25104200	0.45270800
H	3.29800500	-1.96742300	-0.15261100
H	3.41771200	-0.47388700	0.80381200
H	2.32701200	-1.76684500	1.32376000

Table 9.S7 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.924 \text{ \AA}$.

C	-1.84461900	-0.27386000	-0.05635100
O	-1.55731200	-1.25346800	-0.72974900
N	-1.17071400	0.91021900	-0.17325000
H	-1.24893400	1.58595800	0.57166100
C	0.04050400	0.99332700	-0.97518000
H	0.17770600	2.02236800	-1.30622900
H	-0.08446200	0.34747400	-1.84185200
C	1.24614100	0.61142200	-0.11692400
O	1.66384800	1.38427800	0.73306600
N	1.76068400	-0.62117500	-0.34922800
H	1.20452900	-1.24725400	-0.91328800
C	-2.99418600	-0.28799200	0.92845900
H	-3.11632700	0.65250800	1.46686700
H	-3.91439600	-0.50793300	0.38497800
H	-2.83355700	-1.09598300	1.64344300
C	2.81340100	-1.18586700	0.47921200
H	3.41494700	-1.87487500	-0.11388900
H	3.44243300	-0.37479300	0.83948300
H	2.40854200	-1.71943500	1.34433200

Table 9.S8 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.874 \text{ \AA}$.

C	-1.88760800	-0.28680400	-0.04110300
O	-1.62296500	-1.32397000	-0.63134100
N	-1.18219500	0.86714500	-0.24411000
H	-1.26387700	1.60824300	0.43520300
C	0.03632700	0.86020000	-1.03669200
H	0.17543500	1.84401600	-1.48596900
H	-0.08293300	0.11933900	-1.82474800
C	1.24230900	0.58367700	-0.13797000
O	1.59179400	1.40905400	0.69321600
N	1.85217500	-0.61235700	-0.32485900
H	1.36458200	-1.29425500	-0.88604300
C	-3.04190100	-0.19311900	0.93452000
H	-3.16950700	0.80159500	1.36309100
H	-3.95854900	-0.47864800	0.41668400
H	-2.88023800	-0.91343100	1.73792400
C	2.93682500	-1.06756500	0.52978200
H	3.58460900	-1.74112500	-0.03130600
H	3.50990600	-0.20153600	0.85396200
H	2.56437300	-1.58672000	1.41775400

Table 9.S9 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.824 \text{ \AA}$.

C	-1.96325400	-0.29978100	0.00062600
O	-1.73327900	-1.44051600	-0.37143600
N	-1.19716800	0.76442100	-0.38686900
H	-1.29601100	1.63677900	0.10881400
C	0.03703000	0.56759300	-1.12334500
H	0.19028600	1.41075300	-1.80003100
H	-0.07278600	-0.34113700	-1.71207400
C	1.24112400	0.51456100	-0.17877900
O	1.48386300	1.44888300	0.57052600
N	2.00191500	-0.60501900	-0.25948300
H	1.62477300	-1.38993200	-0.76714500
C	-3.13972900	0.02765400	0.89673200
H	-3.22905100	1.08987700	1.12744600
H	-4.05452400	-0.31007100	0.40785500
H	-3.03678300	-0.53423700	1.82624600
C	3.14413800	-0.82875100	0.61193800
H	3.87952400	-1.44796900	0.09803200
H	3.58805100	0.13438400	0.85422300
H	2.85277000	-1.31886200	1.54533500

Table 9.S10 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.774 \text{ \AA}$.

C	-2.02555800	-0.29156100	0.03876600
O	-1.84879600	-1.48774000	-0.13415200
N	-1.19385800	0.65719500	-0.48785400
H	-1.27828400	1.60745600	-0.16226500
C	0.04309200	0.28995700	-1.14616800
H	0.20928000	0.94781500	-2.00320800
H	-0.07393100	-0.73006400	-1.50801200
C	1.25052700	0.44427600	-0.21550400
O	1.43317300	1.47768100	0.40993700
N	2.10006300	-0.61253600	-0.18387600
H	1.78853300	-1.47916800	-0.59273200
C	-3.20354000	0.23580700	0.83219700
H	-3.26685800	1.32461300	0.84687000
H	-4.12099500	-0.17136900	0.40582700
H	-3.12530800	-0.13127000	1.85687700
C	3.28276100	-0.63127300	0.66202700
H	4.04983800	-1.25470300	0.20243500
H	3.65467200	0.38634800	0.76030100
H	3.06091300	-1.01503700	1.66184000

Table 9.S11 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.724 \text{ \AA}$.

C	-2.07649100	-0.27686300	0.07022000
O	-1.95070300	-1.49164500	0.08021000
N	-1.18623500	0.55019300	-0.55610300
H	-1.24991500	1.54242600	-0.39066200
C	0.04998500	0.04536700	-1.11267100
H	0.22804000	0.50032500	-2.09171600
H	-0.07498100	-1.02773100	-1.24957300
C	1.25982800	0.37909500	-0.23290200
O	1.39179300	1.48326600	0.27207100
N	2.17995200	-0.61156400	-0.11858200
H	1.92518200	-1.53339100	-0.43522000
C	-3.25353200	0.40960300	0.73265500
H	-3.25006500	1.49377100	0.61294300
H	-4.17497900	0.00544300	0.31161300
H	-3.24341600	0.16722500	1.79639200
C	3.39325300	-0.45965600	0.66839600
H	4.19455100	-1.05037400	0.22383200
H	3.67537300	0.59096300	0.66813100
H	3.24722000	-0.77729500	1.70462000

Table 9.S12 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.674 \text{ \AA}$.

C	-2.11427500	-0.26128500	0.08343700
O	-2.04152000	-1.47672700	0.17883000
N	-1.17317600	0.48497100	-0.56893200
H	-1.20596200	1.48865600	-0.48180200
C	0.05412300	-0.10247500	-1.05788900
H	0.23185700	0.20666200	-2.09307700
H	-0.08209400	-1.18310600	-1.04298700
C	1.27293100	0.33724500	-0.23987700
O	1.38777900	1.48219600	0.16925900
N	2.22429400	-0.61567400	-0.07138000
H	1.99228100	-1.56625900	-0.31087700
C	-3.27800000	0.51619200	0.66404900
H	-3.22102400	1.58981800	0.47997200
H	-4.20399600	0.12822700	0.23768000
H	-3.31136600	0.33895500	1.74004700
C	3.45693600	-0.36515700	0.65834800
H	4.25973600	-0.97287900	0.24025300
H	3.71036500	0.68791900	0.55729300
H	3.35201600	-0.59395200	1.72255900

Table 9.S13 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.624 \text{ \AA}$.

C	-2.14634000	-0.24773700	0.08861200
O	-2.12281200	-1.46108500	0.22692900
N	-1.15973200	0.44058700	-0.55925300
H	-1.16058800	1.44774100	-0.51575100
C	0.05660800	-0.20610300	-0.99679300
H	0.22870300	-0.01210500	-2.06116900
H	-0.08872300	-1.27816700	-0.86737700
C	1.28597800	0.30627700	-0.24026200
O	1.39266500	1.47736200	0.08965500
N	2.25688200	-0.61914700	-0.03412400
H	2.03750400	-1.58631600	-0.21073700
C	-3.29498600	0.59209800	0.60958400
H	-3.19227800	1.65580300	0.39099600
H	-4.22326500	0.22403500	0.17074200
H	-3.36402200	0.45381200	1.68960200
C	3.50592000	-0.30213300	0.63969400
H	4.30760900	-0.92022400	0.23483300
H	3.73547700	0.74695500	0.46612300
H	3.43762500	-0.46624300	1.71869600

Table 9.S14 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.574 \text{ \AA}$.

C	-2.18641200	-0.23337500	0.08870400
O	-2.21772700	-1.44296200	0.25691600
N	-1.14389400	0.39936000	-0.52559100
H	-1.11327200	1.40702900	-0.52543400
C	0.05908500	-0.30517100	-0.90465100
H	0.21999700	-0.23906000	-1.98704200
H	-0.09338700	-1.35470400	-0.65311300
C	1.30301100	0.27674200	-0.22782500
O	1.40142200	1.46833800	0.02149300
N	2.29596600	-0.61957000	0.00176500
H	2.09091100	-1.59844300	-0.11879600
C	-3.32273700	0.66365300	0.53645400
H	-3.16734300	1.71675800	0.29843300
H	-4.24420600	0.32148400	0.06363500
H	-3.44639800	0.55516300	1.61499300
C	3.56936100	-0.24391800	0.59466900
H	4.37393700	-0.82398300	0.14136200
H	3.73588100	0.81420600	0.40558100
H	3.57596700	-0.40757600	1.67578500

Table 9.S15 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.524 \text{ \AA}$.

C	-2.23128200	-0.21878200	0.08294000
O	-2.32124300	-1.42548600	0.25289300
N	-1.12590500	0.36779700	-0.46024600
H	-1.06319200	1.37344600	-0.49238500
C	0.06032000	-0.38987700	-0.78305400
H	0.19942500	-0.45535000	-1.86937900
H	-0.09376100	-1.40300900	-0.41059500
C	1.32358300	0.25089100	-0.20444700
O	1.41745700	1.45655600	-0.03290500
N	2.33835600	-0.61914800	0.03179000
H	2.14590500	-1.60560400	-0.03666900
C	-3.35759000	0.72518500	0.45313900
H	-3.13634900	1.77256500	0.24346400
H	-4.25325000	0.43159700	-0.09606000
H	-3.57161100	0.61030500	1.51663700
C	3.63893800	-0.19663600	0.52549600
H	4.42952900	-0.74780000	0.01444600
H	3.75056000	0.86597300	0.32224900
H	3.73205700	-0.35590600	1.60314000

Table 9.S16 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.474 \text{ \AA}$.

C	-2.26106500	-0.20789500	0.07560500
O	-2.39414600	-1.41253500	0.23302200
N	-1.11219600	0.34710300	-0.40418600
H	-1.01899300	1.35019300	-0.44543500
C	0.06002200	-0.44793900	-0.68418400
H	0.17817800	-0.61069600	-1.76317900
H	-0.09009100	-1.42581700	-0.22488500
C	1.33651200	0.23075500	-0.18559300
O	1.42470200	1.44354700	-0.06944400
N	2.36914800	-0.61828400	0.04842500
H	2.18718100	-1.60882400	0.02546400
C	-3.37626900	0.76749200	0.39508000
H	-3.11219300	1.80920100	0.20787300
H	-4.24995500	0.50665300	-0.20385500
H	-3.65024400	0.65043300	1.44447200
C	3.68351600	-0.16134900	0.46998900
H	4.45814100	-0.74408700	-0.03004100
H	3.78491900	0.88575300	0.19387100
H	3.81365300	-0.24902900	1.55204400

Table 9.S17 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.424 \text{ \AA}$.

C	-2.29277800	-0.19612900	0.06446500
O	-2.46919300	-1.39933500	0.18946300
N	-1.09748300	0.33019600	-0.32076000
H	-0.97285700	1.32985500	-0.36576100
C	0.05817200	-0.50208700	-0.55639700
H	0.14380700	-0.77201100	-1.61713000
H	-0.07852600	-1.43256300	-0.00306100
C	1.35061700	0.21032800	-0.15893700
O	1.43471100	1.42768000	-0.09957600
N	2.40326100	-0.61800900	0.05777100
H	2.23344700	-1.61041300	0.08923200
C	-3.39928400	0.80638500	0.32523800
H	-3.08957100	1.84347900	0.18929900
H	-4.23100600	0.59068300	-0.34707900
H	-3.75832200	0.67032800	1.34622700
C	3.72987500	-0.12656300	0.39414800
H	4.48467100	-0.78651900	-0.03413200
H	3.84526500	0.87163700	-0.02270600
H	3.87889700	-0.06814400	1.47582700

Table 9.S18 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.374 \text{ \AA}$.

C	-2.32731700	-0.18735200	0.04244200
O	-2.54128000	-1.39030500	0.09103800
N	-1.08388700	0.31837600	-0.17651100
H	-0.93166300	1.31460800	-0.21502000
C	0.05705100	-0.54662000	-0.35807800
H	0.08995800	-0.94907700	-1.37895000
H	-0.04168800	-1.40434800	0.31056500
C	1.36597400	0.20116000	-0.10874500
O	1.43821000	1.42064600	-0.11837100
N	2.43889000	-0.60961200	0.07656400
H	2.28372300	-1.60434400	0.11733600
C	-3.43323400	0.83289400	0.22610300
H	-3.08792000	1.86592900	0.16621400
H	-4.19363100	0.66477300	-0.53795000
H	-3.90154300	0.66781300	1.19741800
C	3.79122500	-0.10846600	0.25764200
H	4.47277500	-0.57567000	-0.45592900
H	3.77581500	0.96537000	0.08693200
H	4.15151400	-0.29882600	1.27149300

Table 9.S19 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.324 \text{ \AA}$.

C	-2.34124400	-0.18174500	0.02339800
O	-2.57457400	-1.38221100	0.03704900
N	-1.07661400	0.30452100	-0.08356500
H	-0.89936500	1.29739500	-0.10584500
C	0.05643300	-0.58187300	-0.19066400
H	0.05389700	-1.10422400	-1.15518000
H	-0.00979700	-1.35597800	0.57866400
C	1.36932800	0.18921000	-0.06184000
O	1.42206900	1.40950700	-0.08724700
N	2.46484600	-0.60385400	0.05389300
H	2.32919600	-1.60154400	0.09587300
C	-3.44346000	0.85350900	0.12849200
H	-3.07919900	1.88147600	0.10360600
H	-4.14440600	0.70184200	-0.69357900
H	-3.98837300	0.68942800	1.05923100
C	3.81865600	-0.08209400	0.14110000
H	4.44908400	-0.51067700	-0.64068900
H	3.77032300	0.99614200	0.00881400
H	4.26278600	-0.29894500	1.11547800

Table 9.S20 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.274 \text{ \AA}$.

C	-2.34454600	-0.17858400	-0.00000700
O	-2.59191000	-1.37648300	-0.00002900
N	-1.07001800	0.29134600	0.00002500
H	-0.87110300	1.28082900	0.00003300
C	0.05714200	-0.60743500	0.00002700
H	0.02338700	-1.26324700	-0.87649400
H	0.02340000	-1.26323600	0.87655700
C	1.36757600	0.17865700	0.00001100
O	1.39998200	1.40006800	-0.00000600
N	2.47923900	-0.59950300	0.00003700
H	2.36035000	-1.60032100	-0.00002000
C	-3.43917600	0.86998000	-0.00000800
H	-3.06127600	1.89330100	-0.00002000
H	-4.06630100	0.71974000	-0.87991800
H	-4.06628400	0.71975900	0.87991800
C	3.82895900	-0.06011800	-0.00003500
H	4.37706200	-0.37872700	-0.88949200
H	3.75454500	1.02469400	0.00036000
H	4.37736800	-0.37937200	0.88899700

Table 9.S21 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.224 \text{ \AA}$.

C	-2.34145800	-0.17606200	-0.00002100
O	-2.60299700	-1.37108500	-0.00012700
N	-1.06199900	0.27864200	0.00011700
H	-0.84233600	1.26444800	0.00020800
C	0.05840400	-0.62725200	0.00012400
H	0.02502600	-1.28241900	-0.87691800
H	0.02511900	-1.28233800	0.87723400
C	1.36320900	0.16755900	0.00003800
O	1.37835100	1.38963600	-0.00003000
N	2.48365200	-0.59654500	0.00018100
H	2.37632800	-1.59869100	-0.00029900
C	-3.42351100	0.88546500	-0.00004200
H	-3.03304600	1.90401600	0.00007100
H	-4.05233400	0.74285200	-0.88002700
H	-4.05251700	0.74271400	0.87978900
C	3.82699400	-0.04118900	-0.00015800
H	4.37776700	-0.35119100	-0.89099200
H	3.73981700	1.04265800	0.00295200
H	4.37995000	-0.35625000	0.88750800

Table 9.S22 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.174 \text{ \AA}$.

C	-2.33825300	-0.17346800	-0.00004000
O	-2.61392200	-1.36546200	0.00001000
N	-1.05407800	0.26603100	0.00002100
H	-0.81330800	1.24778000	0.00005200
C	0.05960100	-0.64688100	0.00022500
H	0.02673800	-1.30151000	-0.87726100
H	0.02673400	-1.30115700	0.87797900
C	1.35889700	0.15652600	0.00011400
O	1.35693900	1.37905200	0.00023600
N	2.48784000	-0.59354500	0.00001900
H	2.39195700	-1.59687400	-0.00052300
C	-3.40765400	0.90079200	-0.00019900
H	-3.00479300	1.91446100	-0.00028800
H	-4.03822300	0.76553800	-0.88011400
H	-4.03830700	0.76571400	0.87968100
C	3.82469800	-0.02247900	-0.00026200
H	4.37894600	-0.32613600	-0.89109200
H	3.72492000	1.06027200	0.00285400
H	4.38114000	-0.33115100	0.88744000

Table 9.S23 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.124 \text{ \AA}$.

C	-2.33492400	-0.17079500	-0.00004800
O	-2.62465700	-1.35961200	-0.00008900
N	-1.04626300	0.25354700	0.00008800
H	-0.78396900	1.23083700	0.00016600
C	0.06071700	-0.66632800	0.00022200
H	0.02836200	-1.32021100	-0.87785900
H	0.02839500	-1.31997900	0.87847900
C	1.35462600	0.14554300	0.00012100
O	1.33579100	1.36831300	0.00015100
N	2.49178000	-0.59051400	0.00016900
H	2.40723500	-1.59488500	-0.00048000
C	-3.39162000	0.91595300	-0.00015500
H	-2.97653300	1.92463500	-0.00018400
H	-4.02378100	0.78820000	-0.88006600
H	-4.02388000	0.78828600	0.87969900
C	3.82204900	-0.00397500	-0.00030800
H	4.37926000	-0.30067700	-0.89160500
H	3.70988900	1.07755000	0.00362100
H	4.38224300	-0.30698100	0.88694400

Table 9.S24 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.074 \text{ \AA}$.

C	-2.33148200	-0.16806400	-0.00003900
O	-2.63521300	-1.35356000	-0.00004000
N	-1.03856400	0.24118900	0.00000100
H	-0.75433800	1.21362600	0.00008000
C	0.06175700	-0.68557500	0.00011600
H	0.03002600	-1.33873500	-0.87854600
H	0.02998400	-1.33859700	0.87887700
C	1.35041600	0.13464400	0.00009600
O	1.31488100	1.35747700	0.00013500
N	2.49549100	-0.58744100	0.00022900
H	2.42212300	-1.59271300	-0.00057000
C	-3.37542900	0.93093200	-0.00009800
H	-2.94828000	1.93452200	-0.00015500
H	-4.00914100	0.81058900	-0.87995900
H	-4.00913600	0.81069300	0.87978300
C	3.81909600	0.01427400	-0.00025400
H	4.37912000	-0.27525000	-0.89211900
H	3.69472500	1.09444700	0.00485900
H	4.38292500	-0.28341700	0.88645400

Table 9.S25 Cartesian coordinates of AcGlyNHMe optimized at $d = 2.024 \text{ \AA}$.

C	-2.32793800	-0.16529000	0.00004800
O	-2.64567400	-1.34729700	0.00037100
N	-1.03097800	0.22890400	-0.00081100
H	-0.72439900	1.19608100	-0.00047300
C	0.06271700	-0.70467600	-0.00086100
H	0.03196000	-1.35669400	-0.88043200
H	0.03131200	-1.35748000	0.87807800
C	1.34625600	0.12379600	-0.00014000
O	1.29421600	1.34652400	-0.00020300
N	2.49896400	-0.58433500	0.00084100
H	2.43666900	-1.59038000	-0.00068100
C	-3.35899900	0.94581900	0.00050400
H	-2.91983100	1.94416700	-0.00001800
H	-3.99472200	0.83276000	-0.87889300
H	-3.99355000	0.83316200	0.88080500
C	3.81582300	0.03231700	0.00028500
H	4.37742200	-0.24744600	-0.89369100
H	3.67951100	1.11100100	0.00987000
H	4.38425000	-0.26276900	0.88485500

Table 9.S26 Cartesian coordinates of AcGlyNHMe optimized at $d = 1.974 \text{ \AA}$.

C	-2.32427500	-0.16249700	0.00013600
O	-2.65595900	-1.34087500	0.00134800
N	-1.02351400	0.21677700	-0.00204400
H	-0.69424100	1.17833400	-0.00133800
C	0.06360400	-0.72354300	-0.00152800
H	0.03392900	-1.37505700	-0.88151800
H	0.03255900	-1.37550500	0.87807600
C	1.34217200	0.11306400	-0.00040100
O	1.27370200	1.33553700	0.00002400
N	2.50222900	-0.58114800	0.00002700
H	2.45080200	-1.58783500	-0.00026500
C	-3.34242800	0.96048200	0.00075000
H	-2.89139400	1.95349300	-0.00026200
H	-3.97989000	0.85460600	-0.87830700
H	-3.97785800	0.85547500	0.88138200
C	3.81231800	0.05004200	0.00119700
H	4.38112800	-0.23149100	-0.88754900
H	3.66420200	1.12719400	0.00089300
H	4.37946900	-0.23120400	0.89110300

Table 9.S27 Energies of AcGlyNHMe optimized at $d = 2.974$ – 1.974 Å. Values are corrected for the zero-point vibrational energy.

d (Å)	Energy (Hartree)
2.974	–456.533758
2.924	–456.533586
2.874	–456.533458
2.824	–456.533614
2.774	–456.533817
2.724	–456.533977
2.674	–456.534115
2.624	–456.534316
2.574	–456.534519
2.524	–456.534791
2.474	–456.535157
2.424	–456.535401
2.374	–456.535690
2.324	–456.536146
2.274	–456.536401
2.224	–456.536491
2.174	–456.536475
2.124	–456.536346
2.074	–456.536090
2.024	–456.535714
1.974	–456.535143

Table 9.S28 Donor–acceptor distances (Å) for various conformations of AcGlyNHMe.

ψ (°)	ϕ (°)								
	–180	–170	–160	–150	–140	–130	–120	–110	–100
180	2.182	2.197	2.241	2.311	2.403	2.513	2.633	2.761	2.891
170	2.197	2.199	2.230	2.290	2.373	2.475	2.591	2.716	2.845
160	2.241	2.230	2.249	2.297	2.369	2.463	2.572	2.693	2.820
150	2.311	2.289	2.296	2.332	2.393	2.477	2.579	2.694	2.817
140	2.402	2.371	2.368	2.392	2.443	2.518	2.611	2.719	2.837
130	2.510	2.472	2.460	2.475	2.517	2.582	2.666	2.767	2.879
120	2.630	2.587	2.568	2.576	2.609	2.665	2.742	2.835	2.940
110	2.756	2.710	2.687	2.689	2.715	2.764	2.833	2.919	3.018
100	2.884	2.838	2.813	2.811	2.832	2.874	2.937	3.016	3.108

Table 9.S29 Data used to generate Figure 9.10A.

ψ (°)	ϕ (°)								
	-180	-170	-160	-150	-140	-130	-120	-110	-100
180	1.44	1.28	0.89	0.47	0.17	0.03			
170	1.28	1.32	1.08	0.69	0.32	0.10	0.01		
160	0.90	1.08	1.03	0.76	0.44	0.18	0.04		
150	0.50	0.7	0.78	0.67	0.45	0.23	0.08	0.01	
140	0.21	0.36	0.47	0.47	0.37	0.22	0.10	0.03	
130	0.07	0.15	0.23	0.26	0.24	0.17	0.09	0.04	
120	0.01	0.04	0.09	0.12	0.13	0.11	0.07	0.03	0.01
110		0.01	0.03	0.04	0.06	0.05	0.04	0.03	0.01
100				0.01	0.02	0.02	0.02	0.02	0.01

CHAPTER IX – Addendum

Crystal Structures of Diethylglycine Derivatives

AcDegNHMe

The crystal lattice features a center of inversion. Solution of the structure shows a superposition of two arrangements that differ in the direction of the hydrogen bonding pattern. Specifically, the optimal refinement of the structure is achieved by dividing the occupancy of each amide group in two and fitting the observed density to a superposition of the two amide directionalities (for example, carbonyls pointed “up” versus pointed “down” in Figure 9.S24). Both orientations share a common conformation that differs qualitatively from that observed in solution.

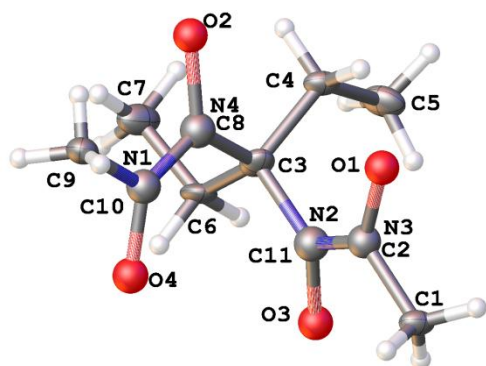


Figure 9.S24 Molecular structure of AcDegNHMe. Thermal ellipsoids are drawn at the 50% probability level.

Experimental

Single crystals of AcDegNHMe ($\text{C}_9\text{H}_{18}\text{N}_2\text{O}_2$) were grown by slow evaporation from hexane with minimal DCM. A suitable crystal was selected and mounted on a Bruker APEX-II CCD diffractometer. The crystal was kept at 99.98 K during data collection. Using Olex2,⁴⁷⁶ the

structure was solved with the ShelXS⁴⁵¹ structure solution program using direct methods and refined with the ShelXL³⁶⁹ refinement package using least squares minimization.

Table 9.S30 Crystal data and structure refinement for AcDegNHMe.

Identification code	AcDegNHMe
Empirical formula	C ₉ H ₁₈ N ₂ O ₂
Formula weight	186.25
Temperature/K	99.98
Crystal system	triclinic
Space group	P ⁻¹
a/Å	7.6768(6)
b/Å	8.5375(7)
c/Å	9.3954(14)
α/°	104.495(5)
β/°	95.056(9)
γ/°	115.922(5)
Volume/Å ³	522.21(10)
Z	2
ρ _{calc} /cm ³	1.184
μ/mm ⁻¹	0.680
F(000)	204.0
Crystal size/mm ³	1.2 × 0.21 × 0.16
Radiation	CuKα (λ = 1.54178)
2θ range for data collection/°	9.984 to 144.818
Index ranges	-9 ≤ h ≤ 9, -10 ≤ k ≤ 10, -10 ≤ l ≤ 11
Reflections collected	9035
Independent reflections	1982 [R _{int} = 0.0482, R _{sigma} = 0.0350]
Data/restraints/parameters	1982/0/122
Goodness-of-fit on F ²	1.121
Final R indexes [I ≥ 2σ (I)]	R ₁ = 0.1071, wR ₂ = 0.2537
Final R indexes [all data]	R ₁ = 0.1147, wR ₂ = 0.2579
Largest diff. peak/hole / e Å ⁻³	0.81/-0.64

Table 9.S31 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for AcDegNHMe. U_{eq} is defined as 1/3 of the trace of the orthogonalized U_{ij} tensor.

atom	x	y	z	U(eq)
C1	227 (6)	1518 (6)	2367 (5)	23.3 (8)
N1	1385 (5)	7323 (5)	3242 (4)	27.5 (8)
O1	178 (8)	3030 (7)	587 (6)	27.2 (12)
C2	-352 (5)	2775 (5)	1898 (4)	26.1 (8)
N2	-1366 (6)	3431 (5)	2672 (4)	28.9 (9)
O2	-315 (8)	7014 (7)	905 (6)	27.2 (12)
C3	-2012 (6)	4706 (5)	2242 (5)	25.8 (9)
N3	-352 (5)	2775 (5)	1898 (4)	26.1 (8)
O3	-2122 (8)	2718 (7)	3856 (6)	27.9 (12)
C4	-3435 (6)	3728 (5)	664 (4)	23.3 (8)
N4	-252 (6)	6336 (5)	2140 (4)	29.1 (9)
O4	1594 (8)	7069 (7)	4594 (6)	29.3 (12)
C5	-5395 (6)	2093 (6)	573 (5)	33.4 (10)
C6	-2996 (6)	5303 (6)	3471 (4)	28.3 (9)
C7	-3767 (7)	6604 (7)	3222 (5)	36.2 (11)
C8	-252 (6)	6336 (5)	2140 (4)	29.1 (9)
C9	3021 (7)	8979 (6)	3137 (5)	26.2 (9)
C10	1385 (5)	7323 (5)	3242 (4)	27.5 (8)
C11	-1366 (6)	3431 (5)	2672 (4)	28.9 (9)

Table 9.S32 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for AcDegNHMe. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11}+2hka^*b^*U_{12}+\dots]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C1	27 (2)	31 (2)	19 (2)	11.3 (16)	4.9 (16)	18.6 (18)
C3	23.2 (19)	27 (2)	29 (2)	11.1 (16)	-2.7 (16)	13.9 (16)
C4	24.5 (19)	31 (2)	20.8 (19)	12.8 (15)	3.6 (15)	16.8 (17)
C5	23 (2)	42 (2)	28 (2)	14.1 (19)	-2.7 (17)	9.7 (18)
C6	35 (2)	32 (2)	19.3 (19)	12.5 (16)	-4.0 (16)	16.3 (18)
C7	51 (3)	50 (3)	26 (2)	16 (2)	13 (2)	38 (2)
C9	33 (2)	26 (2)	20 (2)	7.1 (16)	-0.2 (17)	15.7 (18)

Table 9.S33 Bond lengths for AcDegNHMe.

atom	atom	length/Å	atom	atom	length/Å
C1	C2	1.471 (5)	N2	C3	1.501 (5)
N1	N4	1.330 (5)	C3	C4	1.544 (5)
N1	C9	1.458 (5)	C3	N4	1.489 (5)
O1	C2	1.369 (6)	C3	C6	1.536 (6)
O1	O2 ¹	1.408 (8)	C4	C5	1.517 (6)
C2	N2	1.313 (5)	C6	C7	1.522 (6)

¹-X,1-Y,-Z**Table 9.S34** Bond angles for AcDegNHMe.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
N4	N1	C9	120.3 (4)	N4	C3	N2	109.7 (3)
C2	O1	O2 ¹	167.9 (5)	N4	C3	C4	107.1 (3)
O1	C2	C1	112.4 (4)	N4	C3	C6	110.4 (3)
N2	C2	C1	120.0 (4)	C6	C3	C4	112.4 (3)
N2	C2	O1	127.5 (4)	C5	C4	C3	114.7 (3)
C2	N2	C3	122.4 (4)	N1	N4	C3	122.2 (4)
N2	C3	C4	109.8 (3)	C7	C6	C3	115.2 (3)
N2	C3	C6	107.5 (3)				

¹-X,1-Y,-Z

Table 9.S35 Hydrogen atom coordinates ($\text{\AA}\times 10^4$) and isotropic displacement parameters ($\text{\AA}^2\times 10^3$) for AcDegNHMe.

atom	<i>x</i>	<i>y</i>	<i>z</i>	U(eq)
H1A	150 (60)	550 (60)	1520 (50)	22 (10)
H1B	-530 (80)	1050 (70)	2960 (60)	38 (14)
H1C	1540 (90)	2180 (70)	2860 (60)	44 (14)
H1	1594 (8)	7069 (7)	4594 (6)	29.3 (12)
H8	-315 (8)	7014 (7)	905 (6)	27.2 (12)
H	178 (8)	3030 (7)	587 (6)	27.2 (12)
H2	-2122 (8)	2718 (7)	3856 (6)	27.9 (12)
H4A	-3716	4633	349	28
H4B	-2750	3305	-64	28
H9B	4090 (90)	8840 (70)	2950 (60)	46 (15)
H9C	2650 (80)	9280 (80)	2460 (70)	47 (16)
H5A	-6090	1412	-488	50
H5B	-6220	2528	1109	50
H5C	-5142	1283	1037	50
H6A	-4121	4188	3543	34
H6B	-2018	5910	4451	34
H7A	-4825	5982	2301	54
H7B	-2677	7701	3121	54
H7C	-4295	6967	4086	54
H9A	3230 (70)	10070 (70)	3890 (60)	39 (14)

Table 9.S36 Atomic occupancy for AcDegNHMe.

atom	occupancy	atom	occupancy	atom	occupancy
N1	0.5	H1	0.5	O1	0.5
H8	0.5	C2	0.5	H	0.5
N2	0.5	H2	0.5	O2	0.5
N3	0.5	O3	0.5	N4	0.5
O4	0.5	C8	0.5	C10	0.5
C11	0.5				

AcDegOMe

The compound crystallizes as a hydrogen-bonded array of alternating conformations that are pseudo-symmetric about a glide plane containing the hydrogen-bonded axis. The conformation differs qualitatively from that observed in solution.

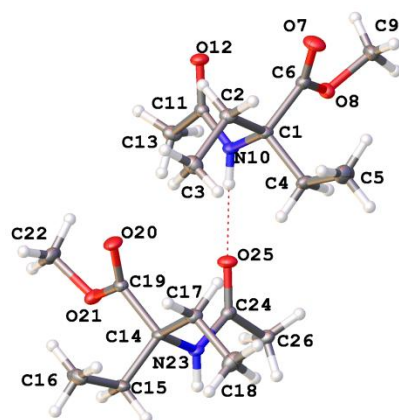


Figure 9.S25 Molecular structure of AcDegOMe. Thermal ellipsoids are drawn at the 50% probability level.

Experimental

Single crystals of AcDegOMe ($\text{C}_9\text{H}_{17}\text{NO}_3$) were grown by slow evaporation from hexane with minimal DCM. A suitable crystal was selected and mounted on a Bruker APEX-II CCD diffractometer. The crystal was kept at 100.05 K during data collection. Using Olex2,⁴⁷⁶ the structure was solved with the ShelXS⁴⁵¹ structure solution program using direct methods and refined with the ShelXL³⁶⁹ refinement package using least squares minimization.

Table 9.S37 Crystal data and structure refinement for AcDegOMe.

Identification code	AcDegOMe
Empirical formula	C ₉ H ₁₇ NO ₃
Formula weight	187.23
Temperature/K	100.05
Crystal system	triclinic
Space group	P-1
a/Å	7.434(2)
b/Å	9.770(3)
c/Å	14.488(4)
$\alpha/^\circ$	84.158(12)
$\beta/^\circ$	78.159(13)
$\gamma/^\circ$	81.529(14)
Volume/Å ³	1015.9(5)
Z	4
$\rho_{\text{calc}}/\text{g}/\text{cm}^3$	1.224
μ/mm^{-1}	0.091
F(000)	408.0
Crystal size/mm ³	0.314 × 0.193 × 0.106
Radiation	MoK α (λ = 0.71073)
2 Θ range for data collection/ $^\circ$	2.88 to 61.266
Index ranges	-10 ≤ h ≤ 10, -13 ≤ k ≤ 13, -20 ≤ l ≤ 20
Reflections collected	28342
Independent reflections	6229 [R_{int} = 0.0270, R_{sigma} = 0.0227]
Data/restraints/parameters	6229/0/243
Goodness-of-fit on F ²	1.078
Final R indexes [$I \geq 2\sigma(I)$]	R_1 = 0.0457, wR_2 = 0.1233
Final R indexes [all data]	R_1 = 0.0556, wR_2 = 0.1302
Largest diff. peak/hole / e Å ⁻³	0.50/-0.23

Table 9.S38 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for AcDegOMe. U_{eq} is defined as 1/3 of the trace of the orthogonalized U_{ij} tensor.

atom	<i>x</i>	<i>y</i>	<i>z</i>	U(eq)
C1	8768.3(16)	-677.9(11)	2959.6(8)	13.8(2)
C2	7008.5(16)	43.0(12)	3561.4(8)	16.5(2)
C3	5503.8(18)	-895.2(14)	3922.1(9)	22.0(2)
C4	9622.4(18)	-2006.5(12)	3467.1(9)	18.5(2)
C5	10103(2)	-1795.0(15)	4411.3(11)	28.1(3)
C6	10253.6(16)	304.4(12)	2726.6(8)	15.5(2)
O7	10265.7(14)	1294.7(10)	3146.7(7)	24.4(2)
O8	11629.9(12)	-171.2(9)	2034.1(7)	18.54(18)
C9	13277.4(17)	493.2(13)	1869.4(10)	20.8(2)
N10	8377.9(14)	-1086.2(10)	2083.1(7)	14.04(18)
C11	8049.9(15)	-163.8(11)	1366.7(8)	13.2(2)
O12	7957.0(12)	1103.4(8)	1410.6(6)	17.24(18)
C13	7838.0(18)	-777.6(13)	487.4(9)	18.8(2)
C14	5605.7(15)	-5224.7(11)	2466.2(8)	13.3(2)
C15	4171.3(16)	-6257.2(12)	2649.9(9)	17.7(2)
C16	2247.3(18)	-5692.1(14)	3167.1(11)	25.7(3)
C17	5964.8(16)	-4740.8(12)	3381.0(8)	15.8(2)
C18	6633(2)	-5918.7(14)	4054.6(9)	23.9(3)
C19	4850.1(15)	-3949.8(11)	1878.6(8)	14.5(2)
O20	4354.3(13)	-2812.0(9)	2161.5(7)	20.47(19)
O21	4686.1(13)	-4278.5(9)	1025.7(6)	19.35(18)
C22	3785(2)	-3176.3(13)	466.3(10)	23.2(3)
N23	7312.1(13)	-5925.2(10)	1919.9(7)	13.39(18)
C24	8762.5(16)	-5259.5(11)	1505.9(8)	13.8(2)
O25	8728.5(13)	-3992.5(9)	1516.7(6)	18.47(18)
C26	10423.8(17)	-6163.3(12)	1009.7(9)	18.7(2)

Table 9.S39 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for AcDegOMe. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11}+2hka^*b^*U_{12}+\dots]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C1	17.1 (5)	9.9 (5)	15.0 (5)	-2.5 (4)	-4.1 (4)	-1.2 (4)
C2	18.6 (5)	14.3 (5)	16.3 (5)	-3.3 (4)	-1.4 (4)	-2.7 (4)
C3	22.0 (6)	23.2 (6)	20.2 (6)	-3.4 (5)	0.2 (5)	-6.4 (5)
C4	23.5 (6)	12.0 (5)	21.1 (6)	0.9 (4)	-8.9 (4)	-1.3 (4)
C5	38.3 (8)	23.1 (6)	27.8 (7)	2.1 (5)	-18.6 (6)	-5.1 (6)
C6	16.1 (5)	12.6 (5)	17.8 (5)	-0.3 (4)	-4.4 (4)	-0.8 (4)
O7	28.7 (5)	17.1 (4)	29.0 (5)	-8.2 (4)	-2.5 (4)	-8.1 (4)
O8	14.2 (4)	15.3 (4)	26.2 (4)	-4.7 (3)	-2.1 (3)	-2.6 (3)
C9	15.6 (5)	16.6 (5)	31.2 (7)	-0.4 (5)	-5.7 (5)	-4.5 (4)
N10	18.0 (4)	9.0 (4)	15.8 (4)	-3.2 (3)	-4.2 (3)	-1.3 (3)
C11	12.5 (4)	11.5 (5)	15.4 (5)	-2.2 (4)	-1.4 (4)	-1.7 (4)
O12	22.2 (4)	9.2 (4)	20.4 (4)	-1.2 (3)	-4.2 (3)	-1.9 (3)
C13	25.5 (6)	16.1 (5)	16.0 (5)	-3.3 (4)	-4.2 (4)	-4.9 (4)
C14	14.5 (5)	7.6 (4)	17.3 (5)	-2.2 (4)	-2.1 (4)	-0.2 (4)
C15	16.7 (5)	9.6 (5)	26.5 (6)	-2.1 (4)	-2.4 (4)	-2.9 (4)
C16	18.7 (6)	18.2 (6)	38.5 (8)	-6.5 (5)	2.4 (5)	-5.4 (5)
C17	18.6 (5)	13.3 (5)	15.3 (5)	-3.3 (4)	-2.3 (4)	-0.8 (4)
C18	29.1 (6)	22.9 (6)	17.8 (6)	-1.9 (5)	-5.2 (5)	4.3 (5)
C19	14.1 (5)	10.5 (5)	19.0 (5)	-1.7 (4)	-3.6 (4)	-1.6 (4)
O20	25.7 (4)	9.2 (4)	27.1 (5)	-5.0 (3)	-8.5 (4)	3.1 (3)
O21	26.8 (4)	12.3 (4)	20.1 (4)	-2.7 (3)	-10.2 (3)	2.6 (3)
C22	29.3 (6)	16.7 (6)	25.2 (6)	1.1 (5)	-13.2 (5)	1.1 (5)
N23	15.2 (4)	7.2 (4)	17.2 (4)	-3.1 (3)	-2.1 (3)	0.5 (3)
C24	16.6 (5)	11.3 (5)	13.4 (5)	-1.7 (4)	-3.7 (4)	-0.2 (4)
O25	23.1 (4)	9.6 (4)	22.0 (4)	-2.6 (3)	-1.5 (3)	-2.6 (3)
C26	17.7 (5)	13.0 (5)	22.8 (6)	-2.1 (4)	0.2 (4)	1.4 (4)

Table 9.S40 Bond lengths for AcDegOMe.

atom	atom	length/Å	atom	atom	length/Å
C1	C2	1.5340 (16)	C14	C15	1.5399 (16)
C1	C4	1.5437 (16)	C14	C17	1.5367 (16)
C1	C6	1.5323 (16)	C14	C19	1.5382 (16)
C1	N10	1.4644 (15)	C14	N23	1.4651 (14)
C2	C3	1.5263 (17)	C15	C16	1.5244 (18)
C4	C5	1.5238 (19)	C17	C18	1.5251 (18)
C6	O7	1.1964 (15)	C19	O20	1.2040 (14)
C6	O8	1.3426 (15)	C19	O21	1.3412 (15)
O8	C9	1.4356 (15)	O21	C22	1.4483 (15)
N10	C11	1.3408 (15)	N23	C24	1.3443 (15)
C11	O12	1.2371 (14)	C24	O25	1.2360 (14)
C11	C13	1.5077 (16)	C24	C26	1.5051 (16)

Table 9.S41 Bond angles for AcDegOMe.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
C2	C1	C4	113.23 (10)	C17	C14	C15	112.92 (10)
C6	C1	C2	109.49 (9)	C17	C14	C19	108.54 (9)
C6	C1	C4	106.59 (9)	C19	C14	C15	107.69 (9)
N10	C1	C2	110.84 (9)	N23	C14	C15	106.72 (9)
N10	C1	C4	106.80 (9)	N23	C14	C17	110.84 (9)
N10	C1	C6	109.76 (9)	N23	C14	C19	110.07 (9)
C3	C2	C1	113.93 (10)	C16	C15	C14	114.93 (10)
C5	C4	C1	114.69 (10)	C18	C17	C14	113.96 (10)
O7	C6	C1	125.14 (11)	O20	C19	C14	124.76 (11)
O7	C6	O8	124.40 (11)	O20	C19	O21	123.49 (11)
O8	C6	C1	110.23 (10)	O21	C19	C14	111.60 (9)
C6	O8	C9	116.09 (10)	C19	O21	C22	115.32 (10)
C11	N10	C1	122.54 (9)	C24	N23	C14	122.85 (9)
N10	C11	C13	115.32 (10)	N23	C24	C26	115.21 (10)
O12	C11	N10	122.84 (11)	O25	C24	N23	122.58 (11)
O12	C11	C13	121.83 (10)	O25	C24	C26	122.20 (10)

Table 9.S42 Hydrogen atom coordinates ($\text{\AA}\times 10^4$) and isotropic displacement parameters ($\text{\AA}^2\times 10^3$) for AcDegOMe.

atom	<i>x</i>	<i>y</i>	<i>z</i>	U(eq)
H2A	7346	399	4110	20
H2B	6499	850	3180	20
H3A	5318	-1385	3398	33
H3B	4343	-331	4184	33
H3C	5884	-1572	4416	33
H4A	8739	-2698	3575	22
H4B	10764	-2398	3044	22
H5A	10934	-1081	4320	42
H5B	10720	-2668	4662	42
H5C	8965	-1500	4859	42
H9A	13118	1340	1454	31
H9B	14335	-138	1569	31
H9C	13503	731	2473	31
H10	8358	-1971	2024	17
H13A	7004	-129	157	28
H13B	7320	-1655	663	28
H13C	9053	-948	71	28
H15A	4058	-6561	2035	21
H15B	4644	-7086	3025	21
H16A	1750	-4881	2797	38
H16B	2330	-5421	3789	38
H16C	1425	-6410	3248	38
H17A	6905	-4091	3213	19
H17B	4805	-4225	3713	19
H18A	7684	-6511	3710	36
H18B	5623	-6471	4318	36
H18C	7022	-5531	4569	36
H22A	3669	-3525	-126	35
H22B	4527	-2404	326	35
H22C	2551	-2854	821	35
H23	7391	-6821	1860	16
H26A	10305	-6229	355	28
H26B	10497	-7092	1337	28
H26C	11551	-5757	1014	28

AcDegMe

The compound crystallizes as hydrogen-bonded arrays of alternating conformations.

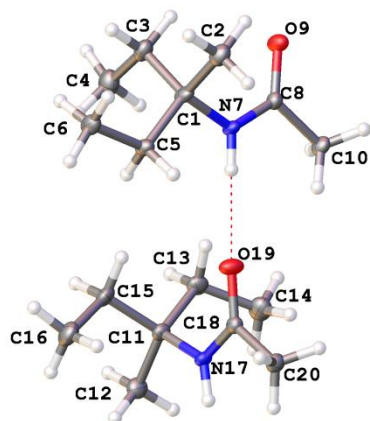


Figure 9.S26 Molecular structure of AcDegMe. Thermal ellipsoids are drawn at the 50% probability level.

Experimental

Single crystals of AcDegMe ($\text{C}_8\text{H}_{17}\text{NO}$) were obtained by slow evaporation from hexane with minimal DCM. A suitable crystal was selected and mounted on a Bruker APEX-II CCD diffractometer. The crystal was kept at 100.0 K during data collection. Using Olex2,⁴⁷⁶ the structure was solved with the ShelXS⁴⁵¹ structure solution program using direct methods and refined with the ShelXL³⁶⁹ refinement package using least squares minimization.

Table 9.S43 Crystal data and structure refinement for AcDegMe.

Identification code	AcDegMe
Empirical formula	C ₈ H ₁₇ NO
Formula weight	143.22
Temperature/K	100.0
Crystal system	triclinic
Space group	P-1
a/Å	7.2873(15)
b/Å	9.662(2)
c/Å	12.662(3)
α /°	91.534(5)
β /°	91.253(5)
γ /°	95.675(5)
Volume/Å ³	886.5(3)
Z	4
$\rho_{\text{calc}}/\text{g}/\text{cm}^3$	1.073
μ/mm^{-1}	0.070
F(000)	320.0
Crystal size/mm ³	? × ? × ?
Radiation	MoK α (λ = 0.71073)
2 Θ range for data collection/°	3.218 to 58.924
Index ranges	-10 ≤ h ≤ 10, -13 ≤ k ≤ 13, -17 ≤ l ≤ 17
Reflections collected	23945
Independent reflections	4896 [R_{int} = 0.1452, R_{sigma} = 0.0807]
Data/restraints/parameters	4896/0/189
Goodness-of-fit on F ²	1.060
Final R indexes [$I \geq 2\sigma(I)$]	R_1 = 0.0618, wR_2 = 0.1685
Final R indexes [all data]	R_1 = 0.0804, wR_2 = 0.1816
Largest diff. peak/hole / e Å ⁻³	0.96/-0.39

Table 9.S43 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for AcDegMe. U_{eq} is defined as 1/3 of the trace of the orthogonalized U_{ij} tensor.

atom	<i>x</i>	<i>y</i>	<i>z</i>	<i>U</i> (eq)
C1	-47 (2)	10948.5 (15)	7273.2 (13)	19.3 (3)
C2	-1210 (2)	9928.7 (17)	6473.5 (13)	24.2 (3)
C3	586 (2)	10136.5 (16)	8212.4 (13)	22.8 (3)
C4	1789 (2)	10998.1 (19)	9012.6 (14)	28.8 (4)
C5	-1176.3 (19)	12140.7 (15)	7630.0 (12)	17.7 (3)
C6	-2893 (2)	11696.1 (16)	8256.5 (13)	23.0 (3)
N7	1551.7 (16)	11636.1 (12)	6719.1 (10)	15.5 (3)
C8	2957.3 (19)	11005.1 (14)	6309.7 (11)	15.4 (3)
O9	3165.0 (15)	9757.9 (11)	6403.8 (9)	22.8 (3)
C10	4319 (2)	11936.1 (15)	5698.7 (12)	19.1 (3)
C11	4008 (2)	16400.1 (15)	8094.7 (11)	17.7 (3)
C12	4965 (2)	17737.6 (16)	8615.4 (13)	22.2 (3)
C13	5442 (2)	15364.5 (16)	7846.3 (13)	21.9 (3)
C14	6780 (2)	15785.4 (18)	6981.2 (14)	29.1 (4)
C15	2567 (2)	15697.7 (16)	8841.2 (13)	23.4 (3)
C16	935 (2)	16518.5 (19)	9070.3 (14)	29.8 (4)
N17	3106.1 (16)	16818.7 (12)	7108.5 (9)	15.2 (3)
C18	2130.9 (19)	15938.6 (14)	6414.5 (12)	16.4 (3)
O19	1874.9 (16)	14671.0 (11)	6527.2 (9)	23.8 (3)
C20	1356 (2)	16612.0 (15)	5458.9 (12)	19.2 (3)

Table 9.S45 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for AcDegMe. The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11}+2hka^*b^*U_{12}+\dots]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C1	18.4 (7)	15.8 (7)	24.7 (8)	6.8 (6)	8.9 (6)	2.4 (5)
C2	21.8 (7)	25.8 (8)	25.9 (8)	6.0 (6)	6.4 (6)	3.4 (6)
C3	23.6 (7)	25.5 (8)	20.7 (8)	7.5 (6)	5.8 (6)	7.1 (6)
C4	27.0 (8)	35.6 (9)	25.0 (9)	2.2 (7)	6.0 (6)	6.7 (7)
C5	16.6 (6)	15.0 (6)	22.4 (7)	3.8 (5)	5.5 (5)	3.6 (5)
C6	21.1 (7)	22.3 (7)	27.2 (8)	5.3 (6)	10.3 (6)	7.3 (6)
N7	16.3 (6)	12.0 (5)	19.0 (6)	5.3 (4)	6.2 (4)	2.5 (4)
C8	15.2 (6)	16.2 (6)	15.1 (7)	1.6 (5)	3.3 (5)	1.9 (5)
O9	24.0 (6)	16.6 (5)	29.0 (6)	4.8 (4)	10.2 (4)	5.1 (4)
C10	16.6 (7)	19.4 (7)	21.7 (8)	5.0 (6)	6.8 (5)	1.4 (5)
C11	22.8 (7)	16.2 (7)	14.5 (7)	3.6 (5)	-0.4 (5)	3.6 (5)
C12	24.8 (8)	22.3 (8)	19.2 (8)	-0.1 (6)	-2.1 (6)	1.2 (6)
C13	25.7 (8)	19.8 (7)	21.4 (8)	4.3 (6)	-1.6 (6)	7.8 (6)
C14	28.8 (8)	30.5 (9)	30.5 (9)	5.1 (7)	6.7 (7)	13.4 (7)
C15	31.3 (8)	22.3 (7)	17.3 (8)	8.3 (6)	4.9 (6)	2.3 (6)
C16	32.3 (9)	32.0 (9)	26.4 (9)	9.3 (7)	11.8 (7)	4.2 (7)
N17	20.1 (6)	11.6 (5)	14.3 (6)	3.6 (4)	1.3 (4)	2.3 (4)
C18	17.5 (6)	15.5 (7)	16.6 (7)	1.7 (5)	4.7 (5)	2.8 (5)
O19	30.5 (6)	13.6 (5)	26.8 (6)	3.6 (4)	-3.6 (5)	0.3 (4)
C20	22.6 (7)	19.1 (7)	16.1 (7)	2.7 (6)	0.1 (5)	1.8 (6)

Table 9.S46 Bond lengths for AcDegMe.

atom atom	length/ \AA	atom atom	length/ \AA
C1 C2	1.568 (2)	C11 C12	1.532 (2)
C1 C3	1.530 (2)	C11 C13	1.548 (2)
C1 C5	1.5441 (19)	C11 C15	1.547 (2)
C1 N7	1.4828 (18)	C11 N17	1.4828 (18)
C3 C4	1.503 (2)	C13 C14	1.519 (2)
C5 C6	1.528 (2)	C15 C16	1.522 (2)
N7 C8	1.3488 (17)	N17 C18	1.3467 (19)
C8 O9	1.2378 (17)	C18 O19	1.2330 (17)
C8 C10	1.512 (2)	C18 C20	1.514 (2)

Table 9.S47 Bond angles for AcDegMe.

atom	atom	atom	angle/°	atom	atom	atom	angle/°
C3	C1	C2	109.50 (12)	C12	C11	C13	110.18 (12)
C3	C1	C5	111.69 (12)	C12	C11	C15	110.37 (12)
C5	C1	C2	110.44 (12)	C15	C11	C13	108.47 (12)
N7	C1	C2	108.81 (12)	N17	C11	C12	106.32 (11)
N7	C1	C3	111.11 (12)	N17	C11	C13	110.71 (12)
N7	C1	C5	105.18 (11)	N17	C11	C15	110.79 (12)
C4	C3	C1	114.35 (14)	C14	C13	C11	115.61 (13)
C6	C5	C1	115.30 (12)	C16	C15	C11	115.23 (13)
C8	N7	C1	126.22 (12)	C18	N17	C11	124.92 (12)
N7	C8	C10	115.13 (12)	N17	C18	C20	115.20 (12)
O9	C8	N7	124.40 (13)	O19	C18	N17	123.63 (14)
O9	C8	C10	120.47 (12)	O19	C18	C20	121.16 (13)

Table 9.S48 Hydrogen atom coordinates ($\text{\AA} \times 10^4$) and isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for AcDegMe.

atom	x	y	z	U(eq)
H2A	-1556	10439	5853	36
H2B	-2326	9528	6815	36
H2C	-474	9181	6252	36
H3A	-517	9711	8570	27
H3B	1273	9371	7943	27
H4A	2855	11463	8662	43
H4B	2215	10394	9557	43
H4C	1080	11698	9341	43
H5A	-363	12820	8069	21
H5B	-1561	12625	6995	21
H6A	-3703	11009	7838	34
H6B	-3552	12510	8416	34
H6C	-2526	11284	8918	34
H7	1587	12542	6651	19
H10A	5480	12111	6107	29
H10B	3812	12821	5572	29
H10C	4547	11479	5020	29
H12A	5892	18160	8141	33
H12B	5566	17517	9281	33
H12C	4048	18392	8757	33
H13A	6164	15236	8502	26
H13B	4771	14453	7641	26
H14A	6104	15788	6304	44
H14B	7709	15121	6936	44
H14C	7389	16719	7146	44
H15A	2092	14780	8525	28
H15B	3196	15529	9520	28
H16A	273	16668	8408	45
H16B	1381	17420	9404	45
H16C	101	15997	9547	45
H17	3216	17709	6966	18
H20A	5	16466	5455	29
H20B	1810	16194	4814	29
H20C	1750	17612	5489	29

F₃CDegOH

The compound crystallizes as a complex with sodium ions and trifluoroacetate in 2:2:1 stoichiometry. Coordinates of acidic hydrogens are not reliably determined. In one case, modeling of a trifluoroacetyl group was improved by explicitly modeling additional, partially-occupied sites.

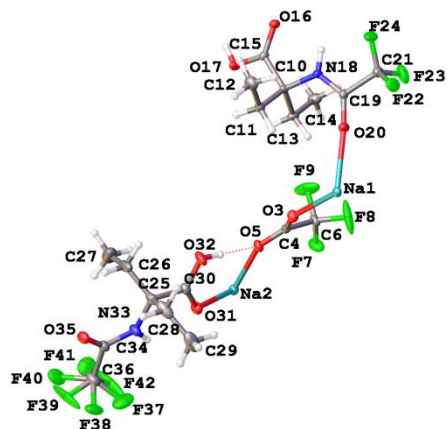


Figure 9.S27 Molecular structure of F₃CDeG-OH. Thermal ellipsoids are drawn at the 50% probability level.

Experimental

Single crystals of F₃CDeG-OH (C₁₈H₂₄F₉N₂NaO₈) were prepared by slow evaporation from hexane with minimal DCM. A suitable crystal was selected and mounted on a Bruker APEX-II CCD diffractometer. The crystal was kept at 100.04 K during data collection. Using Olex2,⁴⁷⁶ the structure was solved with the ShelXS⁴⁵¹ structure solution program using direct methods and refined with the ShelXL³⁶⁹ refinement package using least squares minimization.

Table 9.S49 Crystal data and structure refinement for F₃CDegOH.

Identification code	F ₃ CDegOH
Empirical formula	C ₁₈ H ₂₄ F ₉ N ₂ NaO ₈
Formula weight	590.38
Temperature/K	100.04
Crystal system	monoclinic
Space group	P2 ₁ /c
a/Å	14.470(4)
b/Å	13.372(5)
c/Å	12.713(4)
α/°	90
β/°	92.070(18)
γ/°	90
Volume/Å ³	2458.3(13)
Z	4
ρ _{calc} /g/cm ³	1.595
μ/mm ⁻¹	0.179
F(000)	1208.0
Crystal size/mm ³	0.244 × 0.113 × 0.088
Radiation	MoKα (λ = 0.71073)
2θ range for data collection/°	2.816 to 61.108
Index ranges	-20 ≤ h ≤ 20, -19 ≤ k ≤ 18, -18 ≤ l ≤ 18
Reflections collected	58109
Independent reflections	7512 [R _{int} = 0.0534, R _{sigma} = 0.0334]
Data/restraints/parameters	7512/0/380
Goodness-of-fit on F ²	1.025
Final R indexes [I ≥ 2σ (I)]	R ₁ = 0.0470, wR ₂ = 0.0980
Final R indexes [all data]	R ₁ = 0.0697, wR ₂ = 0.1088
Largest diff. peak/hole / e Å ⁻³	0.77/-0.64

Table 9.S50 Fractional atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for F_3CDegOH . U_{eq} is defined as 1/3 of the trace of the orthogonalized U_{ij} tensor.

atom	<i>x</i>	<i>y</i>	<i>z</i>	<i>U</i> (eq)
Na1	0	0	0	15.50 (18)
Na2	-5000	0	0	16.18 (18)
O3	-1715.3 (8)	329.9 (9)	-120.1 (9)	18.1 (2)
C4	-2514.1 (10)	103.3 (11)	-423.4 (11)	13.5 (3)
O5	-3272.6 (8)	389.9 (9)	-100.9 (9)	18.5 (2)
C6	-2581.2 (10)	-615.0 (12)	-1376.2 (13)	17.5 (3)
F7	-3331.2 (8)	-1165.0 (9)	-1402.4 (9)	30.3 (3)
F8	-1872.0 (9)	-1216.6 (13)	-1418.0 (14)	69.7 (6)
F9	-2616.4 (12)	-85.3 (11)	-2264.3 (9)	57.4 (5)
C10	-460.7 (10)	2313.7 (11)	-2709.6 (11)	13.5 (3)
C11	-462.6 (11)	2773.4 (12)	-1602.2 (12)	17.4 (3)
C12	314.3 (13)	3520.4 (14)	-1381.6 (14)	26.2 (4)
C13	-1226.7 (11)	1522.1 (12)	-2848.0 (12)	17.2 (3)
C14	-1326.7 (13)	1089.2 (13)	-3955.5 (13)	23.6 (3)
C15	-618.1 (10)	3138.1 (11)	-3531.8 (12)	13.7 (3)
O16	-38.9 (7)	3384.2 (8)	-4155.0 (8)	15.6 (2)
O17	-1444.2 (8)	3528.1 (9)	-3481.4 (9)	19.5 (2)
N18	446.2 (9)	1884.8 (10)	-2933 (1)	14.7 (2)
C19	827.6 (10)	1117.9 (11)	-2414.3 (12)	15.1 (3)
O20	520.6 (8)	664.0 (9)	-1670.4 (9)	18.8 (2)
C21	1737.2 (11)	755.6 (12)	-2885.9 (13)	19.7 (3)
F22	2329.3 (7)	428.8 (9)	-2152.6 (9)	28.9 (2)
F23	1551.3 (8)	-4.1 (10)	-3541.7 (10)	37.2 (3)
F24	2152.9 (8)	1456.8 (9)	-3436.8 (10)	37.6 (3)
C25	-4323.6 (11)	2327.2 (13)	2732.9 (13)	21.2 (3)
C26	-4248.4 (18)	3395.7 (15)	2314.6 (18)	41.2 (6)
C27	-4988 (2)	3669 (2)	1508 (2)	62.4 (9)
C28	-3583.5 (13)	2121.8 (18)	3608.9 (15)	35.2 (5)
C29	-3602.2 (15)	1071 (2)	4052.9 (16)	42.5 (6)
C30	-4193.8 (11)	1572.2 (13)	1844.0 (12)	18.3 (3)
O31	-4732.1 (8)	891.5 (9)	1652.9 (9)	20.0 (2)
O32	-3434.8 (8)	1741.8 (10)	1342.6 (10)	25.3 (3)
N33	-5251.6 (9)	2121.2 (11)	3113.5 (11)	19.3 (3)
C34	-5666.7 (11)	2655.1 (13)	3841.3 (13)	19.5 (3)
O35	-5369.1 (9)	3386.7 (9)	4311.8 (10)	24.6 (3)
C36	-6635.6 (13)	2255.7 (16)	4095.0 (16)	31.3 (4)
F37	-6624.8 (19)	1233 (2)	4199 (3)	51.2 (10)
F38	-6575 (2)	1943 (3)	5126 (2)	45.1 (10)
F39	-7017 (3)	2645 (4)	4830 (3)	88 (3)
F40	-7215 (2)	3052 (3)	4177 (3)	48.4 (10)
F41	-7216.1 (18)	2346 (3)	3210 (3)	45.9 (9)
F42	-6971 (3)	1620 (5)	3526 (4)	96 (3)

Table 9.S51 Anisotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for F_3CDegOH . The anisotropic displacement factor exponent takes the form: $-2\pi^2[h^2a^{*2}U_{11}+2hka^*b^*U_{12}+\dots]$.

atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
Na1	18.5 (4)	13.5 (4)	14.6 (4)	-1.6 (3)	2.9 (3)	1.4 (3)
Na2	19.3 (4)	15.0 (4)	14.2 (4)	-1.4 (3)	0.2 (3)	0.3 (3)
O3	16.5 (5)	19.7 (6)	17.8 (5)	-4.4 (4)	-1.7 (4)	-0.8 (4)
C4	16.5 (7)	12.2 (7)	12.0 (6)	-0.7 (5)	0.8 (5)	-0.1 (5)
O5	16.7 (5)	20.5 (6)	18.5 (5)	-4.4 (4)	3.6 (4)	2.8 (4)
C6	12.7 (7)	20.2 (8)	19.9 (7)	-6.3 (6)	3.1 (5)	-1.4 (6)
F7	28.0 (6)	32.8 (6)	30.6 (6)	-12.8 (5)	7.3 (4)	-17.0 (5)
F8	28.2 (7)	84.0 (11)	94.8 (12)	-74.3 (10)	-27.4 (7)	32.2 (7)
F9	112.2 (13)	45.5 (8)	14.6 (5)	-2.4 (5)	6.7 (7)	-38.6 (8)
C10	12.9 (6)	14.0 (7)	13.9 (6)	3.8 (5)	2.2 (5)	1.8 (5)
C11	20.2 (7)	18.2 (7)	14.0 (7)	1.3 (6)	3.7 (6)	3.5 (6)
C12	27.3 (9)	29.0 (9)	22.3 (8)	-6.2 (7)	0.0 (7)	-2.3 (7)
C13	15.4 (7)	17.6 (7)	18.8 (7)	5.0 (6)	1.6 (6)	-1.7 (6)
C14	26.8 (9)	22.8 (8)	21.0 (8)	1.6 (6)	-1.4 (7)	-6.7 (7)
C15	14.4 (6)	12.1 (7)	14.4 (6)	1.2 (5)	-0.6 (5)	-0.4 (5)
O16	16.3 (5)	15.3 (5)	15.2 (5)	3.6 (4)	2.1 (4)	-0.4 (4)
O17	15.0 (5)	21.2 (6)	22.4 (6)	9.4 (5)	3.0 (4)	4.9 (4)
N18	13.9 (6)	15.7 (6)	14.9 (6)	4.0 (5)	3.8 (5)	1.5 (5)
C19	14.6 (7)	13.5 (7)	17.1 (7)	0.9 (5)	1.0 (5)	0.9 (5)
O20	20.0 (5)	18.4 (6)	18.0 (5)	5.6 (4)	4.0 (4)	3.4 (4)
C21	18.5 (7)	17.9 (8)	22.9 (8)	3.6 (6)	3.7 (6)	2.8 (6)
F22	17.2 (5)	37.3 (6)	32.0 (6)	8.3 (5)	-0.2 (4)	9.1 (4)
F23	29.7 (6)	41.0 (7)	41.2 (7)	-19.9 (6)	6.1 (5)	6.4 (5)
F24	26.7 (6)	33.1 (6)	54.6 (8)	21.6 (6)	23.2 (5)	8.6 (5)
C25	18.8 (8)	25.3 (8)	20.1 (8)	-10.1 (6)	6.1 (6)	-5.8 (6)
C26	62.5 (15)	22 (1)	41.6 (12)	-11.2 (9)	34.7 (11)	-10.8 (9)
C27	102 (2)	41.1 (14)	46.8 (15)	18.2 (12)	35.0 (15)	27.9 (15)
C28	19.1 (8)	59.2 (14)	27.3 (9)	-21.6 (9)	-0.1 (7)	-6.4 (9)
C29	32.8 (11)	68.2 (16)	26 (1)	-8.4 (10)	-5.9 (8)	20.8 (11)
C30	17.0 (7)	21.5 (8)	16.4 (7)	-3.6 (6)	1.7 (6)	1.7 (6)
O31	19.8 (6)	22.8 (6)	17.5 (5)	-6.1 (4)	0.9 (4)	-1.7 (5)
O32	19.6 (6)	30.2 (7)	26.8 (6)	-12.8 (5)	8.8 (5)	-2.1 (5)
N33	17.1 (6)	22.4 (7)	18.5 (6)	-7.4 (5)	3.8 (5)	-4.3 (5)
C34	20.0 (7)	22.5 (8)	16.2 (7)	-2.4 (6)	3.0 (6)	-0.8 (6)
O35	30.1 (7)	22.4 (6)	21.8 (6)	-7.6 (5)	8.4 (5)	-3.5 (5)
C36	23.9 (9)	37.9 (11)	32.7 (10)	-8.0 (8)	10.2 (7)	-4.4 (8)
F37	31.4 (14)	42.4 (16)	79 (3)	21.6 (16)	-4.2 (14)	-15.7 (11)
F38	30.2 (15)	57 (2)	49.3 (19)	26.7 (16)	20.8 (12)	6.3 (14)
F39	66 (3)	124 (5)	78 (3)	-85 (4)	60 (3)	-63 (3)
F40	24.2 (14)	65 (2)	57 (2)	17.9 (18)	11.8 (14)	16.2 (13)
F41	23.0 (12)	49.7 (19)	64.1 (19)	2.8 (15)	-10.5 (11)	-2.9 (12)
F42	58 (3)	138 (6)	97 (4)	-105 (5)	60 (3)	-72 (4)

Table 9.S52 Bond lengths for F₃CDegOH.

atom	atom	length/Å	atom	atom	length/Å
Na1	O3	2.5201 (13)	O16	Na1 ⁷	2.4145 (13)
Na1	O3 ¹	2.5201 (13)	N18	C19	1.3284 (19)
Na1	O16 ²	2.4146 (13)	C19	O20	1.2210 (19)
Na1	O16 ³	2.4146 (13)	C19	C21	1.544 (2)
Na1	O20 ¹	2.4456 (12)	C21	F22	1.3177 (19)
Na1	O20	2.4456 (12)	C21	F23	1.335 (2)
Na2	O5	2.5611 (13)	C21	F24	1.3275 (19)
Na2	O5 ⁴	2.5611 (13)	C25	C26	1.530 (3)
Na2	O31 ⁴	2.4348 (12)	C25	C28	1.542 (3)
Na2	O31	2.4348 (12)	C25	C30	1.532 (2)
Na2	O35 ⁵	2.3814 (14)	C25	N33	1.470 (2)
Na2	O35 ⁶	2.3814 (14)	C26	C27	1.501 (4)
O3	C4	1.2428 (18)	C28	C29	1.515 (4)
C4	O5	1.2457 (19)	C30	O31	1.217 (2)
C4	C6	1.546 (2)	C30	O32	1.309 (2)
C6	F7	1.3103 (18)	N33	C34	1.329 (2)
C6	F8	1.307 (2)	C34	O35	1.217 (2)
C6	F9	1.332 (2)	C34	C36	1.545 (3)
C10	C11	1.536 (2)	O35	Na2 ⁸	2.3814 (14)
C10	C13	1.538 (2)	C36	F37	1.374 (4)
C10	C15	1.531 (2)	C36	F38	1.375 (3)
C10	N18	1.4692 (19)	C36	F39	1.219 (3)
C11	C12	1.522 (2)	C36	F40	1.361 (4)
C13	C14	1.524 (2)	C36	F41	1.385 (3)
C15	O16	1.2191 (19)	C36	F42	1.206 (3)
C15	O17	1.3078 (18)			

¹-X,-Y,-Z; ²-X,-1/2+Y,-1/2-Z; ³+X,1/2-Y,1/2+Z; ⁴-1-X,-Y,-Z; ⁵-1-X,-1/2+Y,1/2-Z; ⁶+X,1/2-Y,-1/2+Z; ⁷-X,1/2+Y,-1/2-Z; ⁸-1-X,1/2+Y,1/2-Z

Table 9.S53 Bond angles for F₃CDegOH.

atom atom atom	angle/°	atom atom atom	angle/°
O3 Na1 O3 ¹	180.0	C12 C11 C10	113.93 (13)
O16 ² Na1 O3	80.30 (4)	C14 C13 C10	114.26 (13)
O16 ² Na1 O3 ¹	99.70 (4)	O16 C15 C10	123.32 (13)
O16 ³ Na1 O3 ¹	80.30 (4)	O16 C15 O17	125.12 (14)
O16 ³ Na1 O3	99.70 (4)	O17 C15 C10	111.56 (13)
O16 ² Na1 O16 ³	180.00 (5)	C15 O16 Na1 ⁷	123.94 (10)
O16 ² Na1 O20 ¹	85.73 (4)	C19 N18 C10	124.29 (13)
O16 ³ Na1 O20 ¹	94.27 (4)	N18 C19 C21	113.22 (13)
O16 ² Na1 O20	94.26 (4)	O20 C19 N18	127.73 (15)
O16 ³ Na1 O20	85.74 (4)	O20 C19 C21	118.94 (14)
O20 ¹ Na1 O3 ¹	102.56 (4)	C19 O20 Na1	170.15 (11)
O20 Na1 O3	102.56 (4)	F22 C21 C19	111.77 (14)
O20 Na1 O3 ¹	77.44 (4)	F22 C21 F23	107.50 (14)
O20 ¹ Na1 O3	77.44 (4)	F22 C21 F24	108.19 (14)
O20 ¹ Na1 O20	180.0	F23 C21 C19	109.02 (13)
O5 ⁴ Na2 O5	180.0	F24 C21 C19	113.09 (13)
O31 ⁴ Na2 O5 ⁴	79.53 (4)	F24 C21 F23	107.02 (14)
O31 ⁴ Na2 O5	100.47 (4)	C26 C25 C28	111.17 (17)
O31 Na2 O5	79.53 (4)	C26 C25 C30	110.30 (15)
O31 Na2 O5 ⁴	100.47 (4)	C30 C25 C28	108.37 (15)
O31 Na2 O31 ⁴	180.0	N33 C25 C26	111.52 (16)
O35 ⁵ Na2 O5 ⁴	90.20 (4)	N33 C25 C28	110.36 (15)
O35 ⁶ Na2 O5	90.20 (4)	N33 C25 C30	104.90 (13)
O35 ⁶ Na2 O5 ⁴	89.80 (4)	C27 C26 C25	114.0 (2)
O35 ⁵ Na2 O5	89.80 (4)	C29 C28 C25	114.40 (16)
O35 ⁵ Na2 O31 ⁴	84.27 (5)	O31 C30 C25	123.07 (15)
O35 ⁶ Na2 O31	84.27 (5)	O31 C30 O32	125.06 (15)
O35 ⁵ Na2 O31	95.73 (5)	O32 C30 C25	111.86 (14)
O35 ⁶ Na2 O31 ⁴	95.73 (5)	C30 O31 Na2	128.27 (11)
O35 ⁶ Na2 O35 ⁵	180.00 (9)	C34 N33 C25	124.74 (14)
C4 O3 Na1	151.54 (10)	N33 C34 C36	113.32 (15)
O3 C4 O5	130.05 (14)	O35 C34 N33	127.80 (16)
O3 C4 C6	115.20 (13)	O35 C34 C36	118.87 (15)
O5 C4 C6	114.70 (13)	C34 O35 Na2 ⁸	168.32 (12)
C4 O5 Na2	145.77 (10)	F37 C36 C34	110.91 (19)
F7 C6 C4	113.39 (13)	F37 C36 F41	99.7 (3)
F7 C6 F9	105.70 (14)	F38 C36 C34	106.09 (19)
F8 C6 C4	112.78 (13)	F39 C36 C34	116.8 (2)
F8 C6 F7	107.73 (15)	F39 C36 F37	110.8 (3)
F8 C6 F9	107.39 (16)	F39 C36 F41	107.9 (3)
F9 C6 C4	109.45 (14)	F40 C36 C34	108.1 (2)
C11 C10 C13	110.80 (13)	F40 C36 F38	100.6 (3)
C15 C10 C11	109.42 (13)	F41 C36 C34	109.19 (19)

C15	C10	C13	109.18 (12)	F42	C36	C34	117.8 (2)
N18	C10	C11	111.38 (12)	F42	C36	F38	111.5 (4)
N18	C10	C13	110.75 (13)	F42	C36	F40	111.2 (4)
N18	C10	C15	105.14 (12)				

¹-X,-Y,-Z; ²+X,1/2-Y,1/2+Z; ³-X,-1/2+Y,-1/2-Z; ⁴-1-X,-Y,-Z; ⁵-1-X,-1/2+Y,1/2-Z; ⁶+X,1/2-Y,-1/2+Z; ⁷-X,1/2+Y,-1/2-Z; ⁸-1-X,1/2+Y,1/2-Z

Table 9.S54 Hydrogen atom coordinates ($\text{\AA} \times 10^4$) and isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for F_3CDegOH .

atom	<i>x</i>	<i>y</i>	<i>z</i>	U(eq)
H11A	-1062	3113	-1511	21
H11B	-411	2228	-1076	21
H12A	279	4055	-1908	39
H12B	912	3179	-1416	39
H12C	252	3807	-678	39
H13A	-1096	969	-2347	21
H13B	-1823	1827	-2664	21
H14A	-1507	1621	-4452	35
H14B	-1802	567	-3972	35
H14C	-735	801	-4155	35
H17	-1542	3910	-3998	29
H18	757	2153	-3444	18
H26A	-4278	3867	2913	49
H26B	-3637	3479	2000	49
H27A	-4986	3187	927	94
H27B	-4869	4341	1237	94
H27C	-5592	3657	1832	94
H28A	-2966	2246	3325	42
H28B	-3670	2602	4189	42
H29A	-3490	589	3491	64
H29B	-4208	941	4345	64
H29C	-3120	1003	4610	64
H32	-3367	1295	887	38
H33	-5554	1607	2840	23

Table 9.S55 Atomic occupancy for F_3CDegOH .

atom	occupancy	atom	occupancy	atom	occupancy
F37	0.533 (5)	F38	0.467 (5)	F39	0.533 (5)
F40	0.467 (5)	F41	0.533 (5)	F42	0.467 (5)

CHAPTER X

Conclusions and Future Directions

Conclusions

The results herein provide key insights into the role of $n \rightarrow \pi^*$ interactions in biomolecule structure and function. Specifically, I have shown that $n \rightarrow \pi^*$ interactions are distinct from dipolar interactions due to their important charge contributions. I have also estimated a lower bound of 0.27 kcal/mol for the energy of a typical $n \rightarrow \pi^*$ interaction. My data also demonstrate the utility of thioamides for probing $n \rightarrow \pi^*$ interactions in folded peptides while emphasizing the importance of isolating the study of individual interactions. The discovery of $n \rightarrow \pi^*$ interactions in polylactide provided strong evidence that $n \rightarrow \pi^*$ interactions can play a key organizing role for polymers, including peptides, given the highly local nature of these interactions. Earlier studies had predicted the existence of $n \rightarrow \pi^*$ interactions in protein structure based on the intimacy of carbonyl contacts; I have further reported the effects of those contacts on protein structure and electronics, providing more direct evidence for $n \rightarrow \pi^*$ interactions in proteins. I also provide the first evidence of these effects on the $n \rightarrow \pi^*$ interaction on protein–ligand interactions, specifically concerning quorum sensing modulators. Finally, I have articulated, for the first time, a previously unrecognized force in protein folding: the C5 hydrogen bond. Despite an atypical geometry, this interaction does qualify as a hydrogen bond, and moreover, it provides important stability and organization to the β -sheet.

Future Directions

Detection of carbonyl interactions continues to be a significant challenge. The only reliable tool for unequivocal assignment of $n\rightarrow\pi^*$ interactions, for example, is analysis of high-resolution crystal structures. Complimenting this approach with solution-phase analyses would prove immensely helpful in characterizing the role of carbonyl interactions in a variety of processes but especially protein folding. This goal is a challenging one for several reasons. First, examining individual chemical moieties in complex systems is a general challenge in analytical chemistry and biochemistry; therefore, success in this endeavor will likely require specific labeling strategies or sophisticated NMR approaches, each of which can provide atomic resolution. Second, and perhaps more importantly, because carbonyl interactions tend to be very weak, they are likely populated only transiently at ambient temperatures. Moreover, single interactions impart only very subtle changes to the involved moieties, so detection of the effects of these interactions requires exceptionally high-quality data sets to assign these effects accurately above background noise. NMR spectroscopy seems an excellent candidate for developing these strategies; however, initial attempts to correlate the presence of an $n\rightarrow\pi^*$ interaction with changes in common NMR observables were unsuccessful (Appendix A). Distance measurements by NMR spectroscopy might present a somewhat cruder alternative, but determining distances in interactions involving oxygen might require ^{17}O -labeling,⁴⁷⁷ which is challenging and can be cost-prohibitive.

Alternatively, one could envision observing carbonyl interactions by two-dimensional infrared (2DIR) spectroscopy; however, at least in the case of the $n\rightarrow\pi^*$ interaction, this approach likely suffers from fatal convolutions. Specifically, vibrational coupling is highly dependent upon the angular alignment of the transition dipole moments of the oscillators,⁴⁷⁸⁻⁴⁷⁹

which does not necessarily correlate with the presence of a carbonyl interaction. For example, 2DIR studies,⁴⁸⁰ along with Raman spectroscopy⁴⁸¹⁻⁴⁸² and associated theoretical treatments,⁴⁸³ have observed strong vibrational coupling in α -helices but only very weak, if any, coupling in PPII helices, despite the presence of attractive $n \rightarrow \pi^*$ interactions in both. Indeed, the 2DIR spectrum of AcProNMe₂ shows no discernible coupling (Figure 10.1).

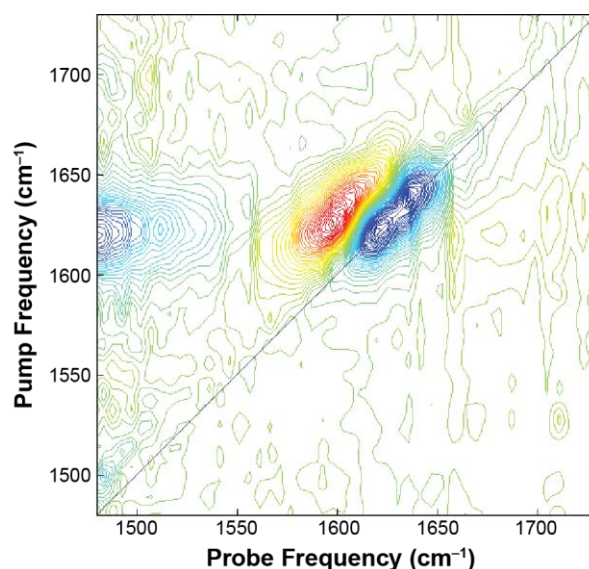


Figure 10.1 Two-dimensional infrared spectrum of AcProNMe₂ in CDCl₃ at 25 °C. Data collected by T. Zhang and M. T. Zanni.

Moreover, in β -sheets, coupling of adjacent carbonyl vibrations results in band splitting of the in-phase and out-of-phase oscillations,⁴⁷⁹ complicating the study of specific interactions such as the C5 hydrogen bond. These results highlight the exquisite dependence of vibrational coupling on the relative orientations of the oscillators, which almost certainly prohibits the use of correlated vibrational spectroscopy for studying carbonyl interactions.

Fundamental questions remain regarding the $n \rightarrow \pi^*$ interaction. Perhaps most important is how other interactions affect the $n \rightarrow \pi^*$ interaction and vice versa. For example, we showed in collaboration with the Woolfson lab that asparagine side chains that form stronger hydrogen

bonds are less likely to form $n \rightarrow \pi^*$ interactions and vice versa,¹⁰⁶ suggesting that $n \rightarrow \pi^*$ interactions could stabilize sub-optimal hydrogen bond geometries. Thus far, however, we have only examined the interplay of hydrogen bonding and $n \rightarrow \pi^*$ interactions in the case where they share an electron pair donor; there are, of course, other means by which an $n \rightarrow \pi^*$ interaction and a hydrogen bond would interact. For example, one would also predict that donation of a hydrogen bond by an amide group would attenuate its ability to act as an acceptor of the $n \rightarrow \pi^*$ interaction. Conversely, if the $n \rightarrow \pi^*$ acceptor were to accept a hydrogen bond, the polarization induced by the hydrogen bond would likely enhance the $n \rightarrow \pi^*$ interaction. Alternatively, if an amide serving as an electron-pair donor donates a hydrogen bond, it will likely be more nucleophilic than if no hydrogen bond were donated, which should enhance the $n \rightarrow \pi^*$ interaction. These hypotheses can likely be assayed by examining solvent effects, though the molecules on which we have focused heretofore are generally devoid of hydrogen bond donors.

We have observed that the formation of an $n \rightarrow \pi^*$ interaction polarizes the acceptor (Chapter 7),¹⁰³ which should improve its ability to serve as an electron pair donor; this polarization should create positive cooperativity between $n \rightarrow \pi^*$ interactions that could further stabilize repetitive secondary structures such as the α -helix. Assaying this effect, though, poses additional experimental challenges. First, one must devise a model system containing two distinct $n \rightarrow \pi^*$ interactions. Second, one must be able to observe changes in the strength of one of these interactions specifically, and observing the effects of $n \rightarrow \pi^*$ interactions remains challenging (see above). Finally, and most importantly, the effects of cooperativity are likely to be subtle, meaning that one must search for subtle changes in a weak interaction, which might not appear above the uncertainty in the chosen measurement. Indeed, our initial attempts to probe the cooperativity of $n \rightarrow \pi^*$ interactions have proven unsuccessful, largely due to synthetic

challenges (see below). In particular, we have found the incorporation of thioamides at and between proline residues to be quite challenging,⁴⁸⁴ and unfortunately, most of our analyses are predicated on examination of proline derivatives, for multiple reasons. First, the proline ring preorganizes the adjacent carbonyl groups for interaction, thereby surmounting the entropic barrier to interaction. Second, modification of the pyrrolidine ring allows for exquisite control over the strength of an $n \rightarrow \pi^*$ interaction (see Chapter 3).⁴⁸⁵ Third, the ability to assay the strength of $n \rightarrow \pi^*$ interactions by examining the equilibrium between isomers of the peptidyl-proline bond using NMR spectroscopy has proven invaluable (see Chapter 4).^{98, 100} Therefore, an alternative choice of model systems is not obvious.

Though we have collected significant evidence that $n \rightarrow \pi^*$ interactions can and do occur in protein structure,^{92, 103} our estimates of their thermodynamic contributions rely on analysis of small-molecule models, often with significant emphasis on computation.¹⁰⁰ In order to integrate $n \rightarrow \pi^*$ interactions into modern molecular mechanical approaches, it will be necessary to improve experimental measurements of the energy of individual $n \rightarrow \pi^*$ interactions in peptide and protein models. Substitution of backbone amides with thioamides provides a promising avenue for the direct perturbation of $n \rightarrow \pi^*$ interactions in peptide and proteins, allowing for experimental determination of the thermodynamic contributions of $n \rightarrow \pi^*$ interactions to protein structure and folding.¹⁰⁰ Indeed, our work on collagen (Chapter 5) has shown that this strategy can be used modulate the stability of large molecules.⁴⁸⁴ Those results also emphasized the necessity of isolating individual interactions in order to draw meaningful conclusions regarding their contributions. In the case of collagen, such isolation is afforded by evolution, which leaves particular backbone groups unburdened by canonical hydrogen bonding and allows for selective perturbation of the $n \rightarrow \pi^*$ interaction with thioamides. In the case of the α -helix, however, the

most common element of secondary structure in proteins and the site of most $n \rightarrow \pi^*$ interactions in proteins, all backbone carbonyl groups participate in canonical $i, i+4$ hydrogen bonds, which are necessarily perturbed upon introduction of backbone thioamides.

To circumvent this convolution, it will be necessary to isolate particular $n \rightarrow \pi^*$ interactions from nearby hydrogen bonds, which can be accomplished by studying the $n \rightarrow \pi^*$ interaction in exceptionally small α -helices that lack some canonical hydrogen bonds. One such model is afforded by the work of Fairlie and coworkers.⁴⁸⁶

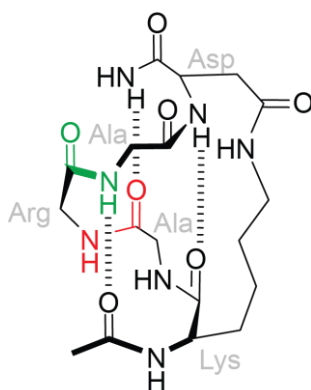


Figure 10.2 Stapled α -helical turn developed by Fairlie and coworkers. Carbonyl groups serving only as hydrogen bond donors (green) or acceptors (red) are highlighted.

Though small, these molecules present significant synthetic challenges. Successful stapling requires selective deprotection of the side chain amine and carboxyl groups in order to introduce the staple on resin. The original synthesis achieved stapling by reductive cleavage of allyl carbamates and esters, respectively, using a palladium catalyst. In my hands, palladium reactions are poisoned by the presence of a thioamide. To circumvent this challenge, one could employ amine and carboxyl side chain protecting groups that cleave in the presence of hydrazine.⁴⁸⁷ Unfortunately, this protection scheme reduces their coupling efficiency and increases cost. Moreover, these protecting groups are incompatible with selective thionation of dipeptide

fragments, necessitating the use of thioacylation for thioamide incorporation. Importantly, even if these synthetic challenges are overcome, it remains difficult to isolate the effect of a single $n \rightarrow \pi^*$ interaction in this system; removal of any hydrogen bond donors or acceptors eliminates helicity, so perturbing any $n \rightarrow \pi^*$ interaction in this molecule will necessarily also perturb a hydrogen bond, confounding interpretation.

Alternatively, larger helices, such as those stabilized by appropriate salt bridges,²⁵ can be modified to isolate the $n \rightarrow \pi^*$ interaction through use of amide-to-ester mutations that remove the relevant hydrogen-bond donors. The technique was exploited to allow selective perturbation of the C5 hydrogen bond in Chapter 9. One could therefore envision the following design:

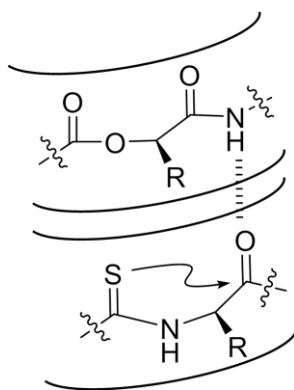


Figure 10.3 Helical peptide scaffold for assaying contributions from $n \rightarrow \pi^*$ interactions.

This design isolates the position of interest by incorporating a backbone ester group four residues C-terminal to the site of thioamide incorporation so as to eliminate the canonical hydrogen bond donor. This is a distinct advantage over the stapled design because one can more effectively isolate the $n \rightarrow \pi^*$ interaction without fatally compromising thermal stability. Additionally, in this latter scheme, only the most N-terminal carbonyl of the helix is probed, which offers two main advantages. First, the N-terminal residues do not donate canonical hydrogen bonds to subsequent turns of the helix, improving the selectivity of the thioamide perturbation. Second, the N-

terminal residue does not receive an $n \rightarrow \pi^*$ interaction. Given the thioamide is simultaneously a stronger donor and weaker acceptor of the $n \rightarrow \pi^*$ interaction than the oxoamide,^{171, 378} it is necessary to segregate these two effects. Placing the thioamide at the N terminus also provides the added advantage of synthetic accessibility. Since the N terminus is the final section of a peptide to be synthesized on solid phase, one can generate a large pool of the peptide lacking the N-terminal groups and then divide that pool for a variety of modifications. N-terminal substitutions are also less likely to complicate later coupling steps.

Designed mini-proteins may offer additional opportunities for studying these interactions. For example, the Trp-cage consists of a PPII helix folded onto an α -helix thanks to a structured turn.²⁹² The length of this “protein” makes it amenable to solid-phase generation of backbone-modified variants. Moreover, this peptide is significantly more stable than other systems of focus, and is amenable to characterization by NMR spectroscopy. This particular choice of model, however, does create its own challenges. In particular, the N-terminal residue of the α -helix is asparagine, which bears multiple thionation-competent groups, complicating synthesis. Moreover, the proline-rich PPII helix is likely to be recalcitrant to thionation based on our previous results.

Despite the advantages of the experimental designs described above, the synthetic challenges to incorporating thioamides into the peptide backbone have heretofore precluded further study. There are two basic approaches to thiopeptide synthesis: thioacylation and segment condensation of fragments into which thioamides are introduced by thionation; unfortunately, both have failed to yield the appropriate products in my hands. In particular, the synthesis of common thioacylation reagents is not robust and cannot be generalized to key amino acids.^{366, 381} Moreover, those that can be generated are not usually competent for solid-phase synthesis

directly. With regard to thionation, I have achieved significant success in introducing thioamides into amino-acid dimers and trimers with excellent regioselectivity provided largely by sterics.⁴⁸⁴ Unfortunately, some of these species are not competent for segment condensation, and it is not immediately clear why. Successful research involving synthetic peptides relies on robust methods for introducing chemical functionalities site-specifically, in high yield, and with excellent fidelity for a wide range of amino acids. In contrast, the generation of thiopeptides is idiosyncratic at best. Improved synthetic methods to access thioamide-substituted peptides would therefore be welcomed.

Alternatively, one could take a cruder, more global approach to investigating the role of $n \rightarrow \pi^*$ interactions in protein structure. Knowing that thioamides generally enhance the $n \rightarrow \pi^*$ interaction, one could introduce them broadly across a peptide or protein by global thionation. One could hypothesize that such a treatment would increase the helicity of the protein due to the correlation of $n \rightarrow \pi^*$ interactions with helical secondary structure. The effects of thioamide incorporation of helical stability so far have been mixed, with both computational⁴⁸⁸⁻⁴⁹⁰ and experimental studies⁴⁹¹⁻⁴⁹³ offering contradictory results. Perhaps thioamide-substituted peptides are less likely to form β -rich amyloid fibers as a result of increased helicity, though the effects of the thioamide on hydrogen bonding interactions are likely dominant in this case. Global thionation could be of particular interest for the study of intrinsically disordered proteins (IDPs). Indeed, as many IDPs adopt structure upon binding to a ligand,⁴⁹⁴⁻⁴⁹⁵ binding affinity could provide an additional readout as to the effect of thionation of the structure of these proteins. Ultimately, however, the $n \rightarrow \pi^*$ interaction is a highly specific, highly local interaction, so conclusions about individual interactions from such a crude experiment are likely to be limited.

To probe the thermodynamic contributions of $n \rightarrow \pi^*$ interactions in peptides or proteins, alternative perturbation strategies could be employed. Indeed, a wide variety of peptide bond isosteres exist that could modulate the $n \rightarrow \pi^*$ interaction.¹⁷¹ Fluoroalkenes have proven especially useful for probing $n \rightarrow \pi^*$ interactions,¹⁸⁰ and may provide a method for effectively deleting $n \rightarrow \pi^*$ interactions while maintaining structural rigidity and relative polarity. The syntheses of these groups is challenging, though advances in olefin metathesis could improve access.⁴⁹⁶ Alternatively, depsipeptides are a more tractable target, and the enhanced electrophilicity of the ester relative to the amide makes it a more potent $n \rightarrow \pi^*$ acceptor.¹⁰⁰ However, the ability of esters to donate $n \rightarrow \pi^*$ interactions is less clear, especially given that they are less basic than amides; nevertheless my calculations in Chapter 6 suggest that esters are competent donors. One could perhaps employ vibrational spectroscopy of α -aminoisobutyric acid (Aib) derivatives to address this issue; despite the lack of a constraining ring, Aib monomers and polymers have a strong tendency toward helical conformations,⁴⁹⁷ especially the 3_{10} -helix,⁴⁹⁸⁻⁴⁹⁹ and would therefore preorganize adjacent carbonyls for an $n \rightarrow \pi^*$ interaction that could be probed by ester substitution. Unfortunately, like all isosteres of the peptide bond, these modifications necessary perturb backbone hydrogen bonding, complicating analysis. In fact, thioamide substitution is probably the most conservative peptide-bond modification available, at least with respect to hydrogen bonding; esters and fluoroalkenes are much more disruptive to hydrogen bonds, perhaps limiting their utility.

Another approach one could employ for investigating $n \rightarrow \pi^*$ interactions would be to examine systems that undergo changes in the number of carbonyl groups, and therefore, the number of potential $n \rightarrow \pi^*$ interactions. For example, acetylation is a common post-translational modification, particularly affecting histones and transcriptional regulation. Changes in protein

structure or function upon acetylation may therefore be dependent upon $n \rightarrow \pi^*$ formation. Given that it normally affects lysine residues, however, acetylation also causes changes in protein electrostatics. To discriminate between these two effects, one could potentially study the structure and activity of proteins incorporating nonnatural amino acids; for example, one could compare the effect of replacing a lysine with either acetyl lysine or a linear hydrocarbon, which would isolate the effect of the $n \rightarrow \pi^*$ interaction.

It is likely that additional roles of $n \rightarrow \pi^*$ interactions remain to be identified. In particular, though these interactions are typically features of secondary structure, where they make key contributions, they may mediate other important contacts at the levels of tertiary or quaternary structure. Moreover, these interactions may also mediate protein–ligand interactions, some of which I highlight in Appendix D. My initial screen of these contacts examined them quite broadly, but additional insights might be gleaned from focusing on enzyme active sites, high-throughput identification of which is possible through bioinformatic approaches.⁵⁰⁰⁻⁵⁰¹ Identification of such interactions may also suggest a role for them in enzymatic catalysis. In addition, it has been observed that $n \rightarrow \pi^*$ interactions reduce the electrophilicity of the acceptor carbonyl, which should reduce its reactivity with nucleophiles such as water. Indeed, the Kent group has determined that the $n \rightarrow \pi^*$ interaction is likely to blame for the slow reactivity of prolyl thioesters in native chemical ligation (NCL) reactions.²²³ A quantitative treatment of this effect could be useful for planning a variety of chemical transformations.

It has also been speculated that $n \rightarrow \pi^*$ interactions might improve electron transfer through the peptide backbone.⁹² The peptide backbone has been predicted to serve as a conduit for tunneling electrons,⁵⁰²⁻⁵⁰⁵ and computations implicate helical geometries as particularly permissive.⁵⁰⁶⁻⁵⁰⁷ Fortuitously, appropriate peptide model systems have been developed for

studying electron transfer in peptides.⁵⁰⁸⁻⁵⁰⁹ Substitution of backbone amides with thioamides or other appropriate peptide bond isosteres (see above) could therefore provide a method to probe the role of the $n \rightarrow \pi^*$ interaction in this dynamic process.

Systems outside biology undoubtedly also exploit the $n \rightarrow \pi^*$ interaction. For example, we showed that polylactides are stabilized by an extensive network of $n \rightarrow \pi^*$ interactions (Chapter 6).¹⁷⁹ It is likely that other synthetic polymers, polyesters in particular, depend on these types of interactions. Identification of such contributions is reliant upon high-resolution structural data, which is often lacking for polymers, as their polydispersity is generally much greater than that of biological samples. As more high-resolution structural data is collected for synthetic polymers, additional intermolecular interactions may be discovered.

The studies I have described herein demonstrate that proteins have evolved to capitalize upon and exploit the quantum mechanical features of their constituent chemical functionalities in order to establish structural order in these complex biopolymers. It is likely that there are still other interactions contributing to the structure of proteins that we have yet to identify, some of which might have important quantum-mechanical character. For example, in β -sheets, backbone carbonyl oxygens can form short contacts with α -protons of adjacent residues, resulting in a C–H \cdots C5 hydrogen bond. One would expect these interactions to be still weaker than either C–H \cdots hydrogen bonding or C5 hydrogen bonding, and indeed, my initial calculations indicate that they are likely at least one-third weaker than typical examples of the former two interactions; they might also be less widely distributed across protein structure.

Recently, an additional contribution, termed “chalcogen bonding,” has been identified as a contributor to the binding of methyl transfer cofactor SAM.⁵¹⁰ Chalcogen bonds were originally articulated by Dunitz, who observed geometric preferences for contacts between sp^3 chalcogens

in crystals.⁵¹¹⁻⁵¹² Based on the geometries observed in the solid state, as well as from quantum mechanical calculations, these contacts afford stabilization through donation of sulfur lone-pair electron density into an antibonding orbital involving the partner chalcogen. The importance for SAM binding can thus be rationalized as the interaction of a nucleophilic cysteine with the potent electron pair acceptor of the S–CH₃ σ^* antibonding orbital. Though proteins generally do not feature such electrophilic chalcogens, the number of potential interactions suggest that these donor-acceptor interactions might have the potential to contribute to protein folding, though no systematic analyses have been performed to date. Such an inventory is likely challenging because proteins predominantly feature unalkylated chalcogens; the location of the relevant lone pairs can only be inferred from the location of the two substituents, and hydrogen locations in proteins are rarely determined with sufficient accuracy to allow unequivocal conclusions regarding chalcogen contacts involving hydroxyls and sulfhydryls. Moreover, hydrogen bonding is likely to be dominant in these cases, reducing contributions from chalcogen bonding. Although an inventory might be possible for interactions between protein thioethers, the relative infrequency of methionine residues in proteins suggests that contributions of such interactions would be minimal.

APPENDIX A

Correlation of $n \rightarrow \pi^*$ Interactions with NMR Observables

The detection of $n \rightarrow \pi^*$ interactions is largely limited to high-resolution crystal structures, preventing characterization of their role in dynamic processes such as protein folding. To address this challenge, NMR spectroscopy has been evaluated for its potential to identify $n \rightarrow \pi^*$ interactions in solution. Specifically, a variety of NMR measurements have been collected on a series of proline model compounds to identify parameters that correlate with the presence of the $n \rightarrow \pi^*$ interaction.

$n \rightarrow \pi^*$ Interactions have been implicated in a variety of molecules, where we believe they contribute to conformation and stability. It is also likely that they guide dynamic processes. For example, we have found a similarity between unfolded polyalanine and its depsipeptide counterpart, polylactide, which is incapable of hydrogen bonding; specifically, they both adopt the PPII geometry that we believe is stabilized by pervasive $n \rightarrow \pi^*$ interactions. This result suggests that $n \rightarrow \pi^*$ interactions may direct early events in the folding process, before the formation of the native hydrogen bonding pattern. The highly local nature of these interactions suggests that they may determine early conformational biases in protein structure that eventually lead to folding.

In order to test these types of hypotheses, it is necessary to observe the presence of $n \rightarrow \pi^*$ interactions in solution with reasonable time and atomic resolution. Unfortunately, our current methods for identifying $n \rightarrow \pi^*$ interactions rely on analysis of high-resolution crystal structures, which do not provide sufficient information regarding dynamic processes in solution. As an alternative, NMR spectroscopy offers excellent atomic resolution on a variety of time scales, and thus would be a promising approach to detection of $n \rightarrow \pi^*$ interactions in dynamic processes. There are two main challenges for studying these interactions by NMR. First, one partner in the $n \rightarrow \pi^*$ interaction is usually oxygen, which is generally invisible to NMR spectroscopy. Though ^{17}O is magnetic, it has a large quadrupole moment that diminishes the quality of the spectra; moreover, this is a very expensive label to introduce, and it often cannot be introduced beyond 20% abundance. The second main challenge to studying $n \rightarrow \pi^*$ interactions by NMR is the fact that the interaction involves the π^* orbital of the acceptor, which has a node at both atomic nuclei. Therefore, even if an electronic interaction occurs, the effects of that interaction may not be transduced to the nuclei that are probed by NMR, or at least not directly. It is apparent,

therefore, that in order to apply NMR spectroscopy to the observation of $n \rightarrow \pi^*$ interactions, one must consider a wide variety of the observables generated by NMR.

To begin evaluating the potential for NMR spectroscopy to observe $n \rightarrow \pi^*$ interactions in solution, a variety of chemical shifts and coupling constants were recorded for AcProOMe bearing ^{13}C labels at both carbonyls, focusing on parameters directly related to the carbonyl moieties. The parameters were obtained by solution-state ^1H -, ^{13}C -, and ^1H -undecoupled ^{13}C -spectra. The results are summarized in Table A.1.

Table A.1 NMR properties of AcProOMe.

Parameter	<i>trans</i>	<i>cis</i>
$\delta_{\text{C}'\text{-donor}}$ (ppm)	169.52	169.66
$\delta_{\text{C}'\text{-acceptor}}$ (ppm)	172.96	172.72
$^1J_{\text{C}'\text{-C}\alpha}$ (Hz)	63.50	61.75
$^1J_{\text{C}\alpha\text{-H}\alpha}$ (Hz)	150.03	147.46
$^2J_{\text{C}'\text{-H}\alpha}$ (Hz)	4.44	2.58
^{13}C Linewidth (Hz)	1.697 (125 MHz)	1.452 (125 MHz)
	2.130 (188 MHz)	2.082 (188 MHz)

Data collected in D_2O at 25 °C.

Though some parameters differ measurably between the *trans* and *cis* conformations, the differences are subtle, as expected. The largest change is for coupling of the acceptor carbonyl to the α -proton. This parameter could therefore find utility as a probe of these interactions, though the α -proton region is typically congested for full-length proteins, which could complicate analysis.

As an alternative metric, I compared the longitudinal relaxation rates (T_1) of carbonyl groups involved in $n \rightarrow \pi^*$ interactions, this time focusing on an unlabeled sample of AcProNMe₂ (Figure A.1). I estimated the relaxation rates using an inversion-recovery experiment, as has been performed previously. No significant changes in the rate of relaxation were observed.

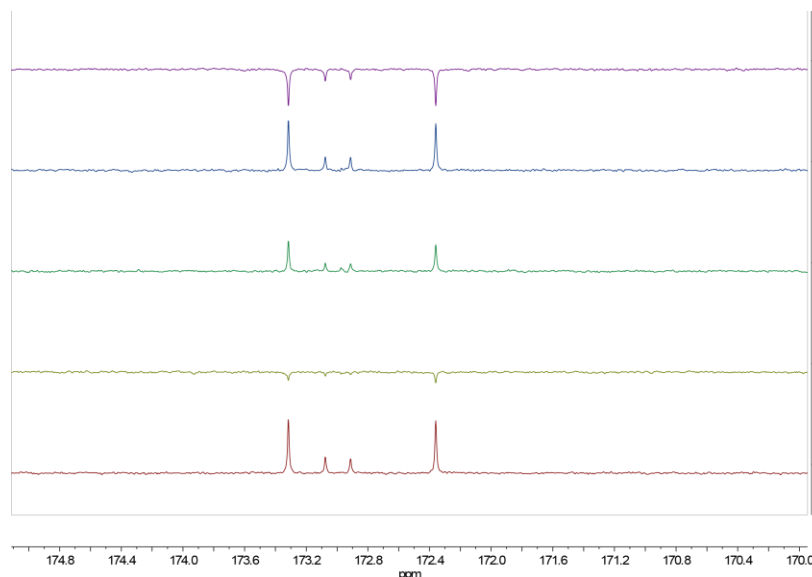


Figure A.1 Measurement of carbonyl longitudinal relaxation by inversion recovery. ^{13}C NMR spectra of AcProNMe₂ in D₂O at 25 °C acquired at various delay intervals.

As a preliminary measure of the effect of $n \rightarrow \pi^*$ interactions on other relaxation phenomena, in particular chemical shift anisotropy (CSA), I compared the linewidths of carbonyl signals in the presence and absence of the $n \rightarrow \pi^*$ interaction. Though there is some difference in the presence of an $n \rightarrow \pi^*$ interaction (Table A.1), this difference is smaller at higher field strengths. A more direct measure of the CSA of the carbonyl group can be achieved using solid-state NMR, though this sacrifices the applications available to a solution-state method.

Finally, though not a direct NMR measurement *per se*, I measured the rate of amide isomerization in the presence and absence of $n \rightarrow \pi^*$ interactions using an NMR experiment. Specifically, I selectively excited one of two methyl groups on a thioamide-containing proline derivative and observed chemical exchange as the amide bond rotated, equilibrating the magnetization (Figure A.2).

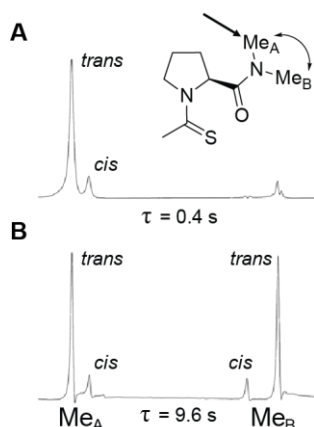


Figure A.2 Magnetization transfer by chemical exchange due to amide isomerization. ^1H NMR spectra in the 3–4 ppm region acquired in D_2O at 45 $^\circ\text{C}$.

The use of a thioamide $n \rightarrow \pi^*$ donor not only increases the strength of the interaction, but isolates the bond rotation of the acceptor; with an oxoamide donor, there would be two concurrent bond rotations that could be confounding. Upon analysis, there was not significant change in the isomerization rate based on the conformation of the thioacetyl moiety.

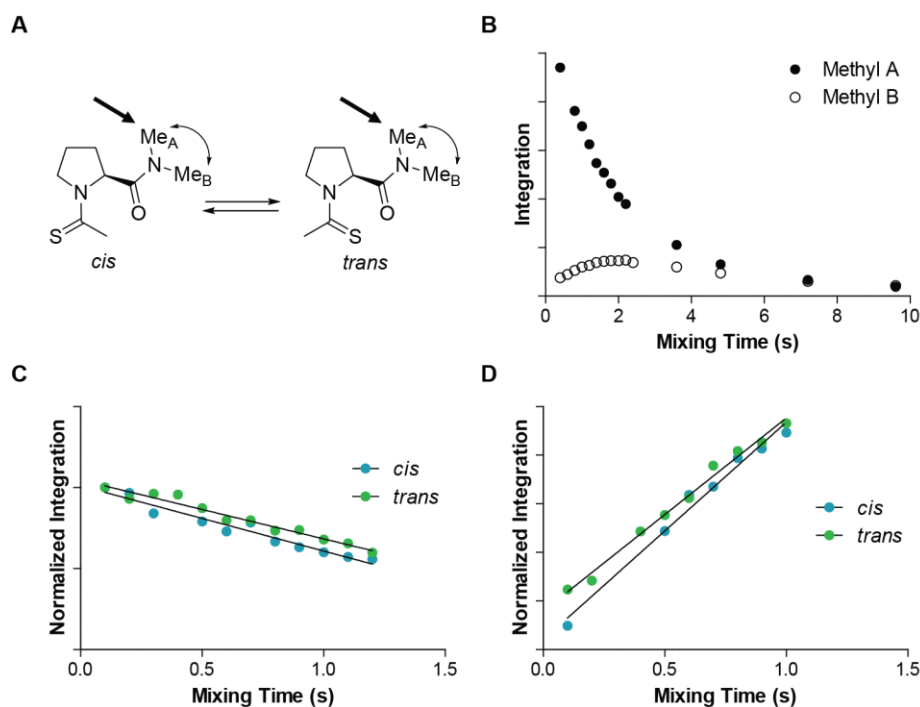


Figure A.3 Amide isomerization kinetics obtained from NMR spectroscopy. (A) Conformational equilibrium of proline derivatives. (B) Integration of amide methyl groups over time following excitation. (C) Chemical exchange over time as observed from the excited methyl group. (D) Chemical exchange over time as observed from the non-excited methyl group.

My preliminary results do not identify any NMR observables that are closely related to the presence or strength of an $n \rightarrow \pi^*$ interaction. Other approaches may be available, such as distance measurements involving ^{17}O probes or CSA measurements by solid-state approaches.⁵¹³⁻⁵¹⁸

APPENDIX B

Solvent Effects on Intermolecular Interactions in Proline Residues

Noncovalent interactions are particularly sensitive to environmental variables such as phase, temperature, pressure, and solvation. Indeed, the affect of these parameters has been used to delineate the nature of several intermolecular interactions. To shed light on the nature of the $n \rightarrow \pi^*$ interaction, including its electrostatic character, a series of proline model compounds were subjected to conformational analysis in different solvents. NMR spectroscopy has previously shown that these compounds engages in attractive $n \rightarrow \pi^*$ interactions and that those interactions are not well described by electrostatics. The solvent effects described herein support this notion. In addition, these data suggest an interplay of $n \rightarrow \pi^*$ interactions with hydrogen bonds.

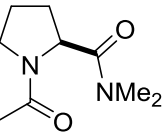
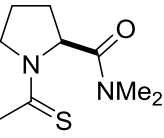
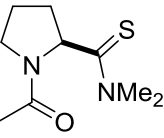
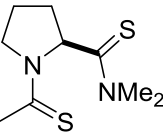
The nature of interactions occurring between carbonyl groups has recently been the subject of some debate. Specifically, some have proposed that carbonyl groups interact primarily through classic dipolar mechanisms, despite a large body of evidence indicating charge transfer character of these interactions. We have shown previously that substituting the carbonyl oxygen with sulfur is an effective strategy for modulating carbonyl interactions and for probing their nature. Relative to carboxylic amides, thioamides are more nucleophilic and have larger dipole moments. Therefore, if classical electrostatics dominates the interaction between carbonyl groups, substituting either amide of an interacting pair with a thioamide should cause a similar effect. Conversely, if interactions between carbonyl groups have significant charge transfer character, thioamide substitution should have a much larger effect at the carbonyl that orients the chalcogen toward the carbon of its partner, owing to the enhanced nucleophilicity of thioamides.

To evaluate these predictions experimentally, a series of *N*-acetyl proline dimethyl amides have been prepared and characterized. The *N*-acetyl peptide bond populates both the *cis* and *trans* conformations, but an attractive carbonyl-carbonyl interaction exists only in the *trans* conformation. Therefore, perturbations that increase the strength of the interaction, regardless of its nature, should increase population of the *trans* conformer. The two peptide bonds isomers equilibrate slowly on the NMR timescale, so their equilibrium populations can be determined by integrating ^1H NMR signals for each isomer.

It is evident from measurements in a wide variety of solvents that thioamide substitution has very different effects at the two different positions within these molecules. Specifically, substitution of the *N*-acetyl amide, which approaches the prolyl amide through its chalcogen substituent, uniformly increases the preference of the molecule for the *trans* conformation, indicating a stronger interactions, whereas substitution of the prolyl amide causes little change in

the conformational preferences of these molecules, and in some cases actually decreases the *trans* preference. It is clear from these data that interactions between carbonyl groups are not accurately described by considerations from classical electrostatics.

Table B.1 $K_{trans/cis}$ values of proline derivatives in various solvents.

Solvent	Dielectric Constant				
		1	2	3	4
1,4-Dioxane	2.3	6.7	4.4	8.0	8.8
Chloroform	4.8	8.2	15.2	8.5	Large
Acetic Acid	6.2	6.4	15.6	3.7	5.8
Tetrahydrofuran	7.5	3.1	4.3	3.2	4.1
Methanol	33.0	4.2	7.0	2.1	3.5
Acetonitrile	37.5	2.6	3.1	2.0	2.8
Water	80.0	3.4	4.1	2.2	6.3

Moreover, despite relatively large *trans* preferences for these molecules in chloroform, there is no systematic change in conformational preferences as the dielectric constant of the solvent changes. For example, despite similar dielectric constants, tetrahydrofuran and chloroform support very different conformational populations. These solvents also differ in their ability to participate in hydrogen bonding, though these propensities also do not have a uniform effect on the conformational preferences of these molecules. Because thioamides are weaker hydrogen bond acceptors than are carboxylic amides, they can be used to probe the effect of hydrogen bonds to solvent on carbonyl-carbonyl interactions. For example, consider the results in acetic acid, which is a potent hydrogen bond donor. In acetic acid, there is a large increase in *trans* preference upon thionation of the *N*-acetyl peptide bond, especially relative to the same perturbation observed in other solvents. This is likely due to polarization of the acceptor carbonyl by formation of a hydrogen bond to solvent. In contrast, acetic acid promotes a very

low *trans* preference for **3**, where hydrogen bonding to the *N*-acetyl peptide bond likely distracts the donor carbonyl from forming an attractive carbonyl-carbonyl interaction.

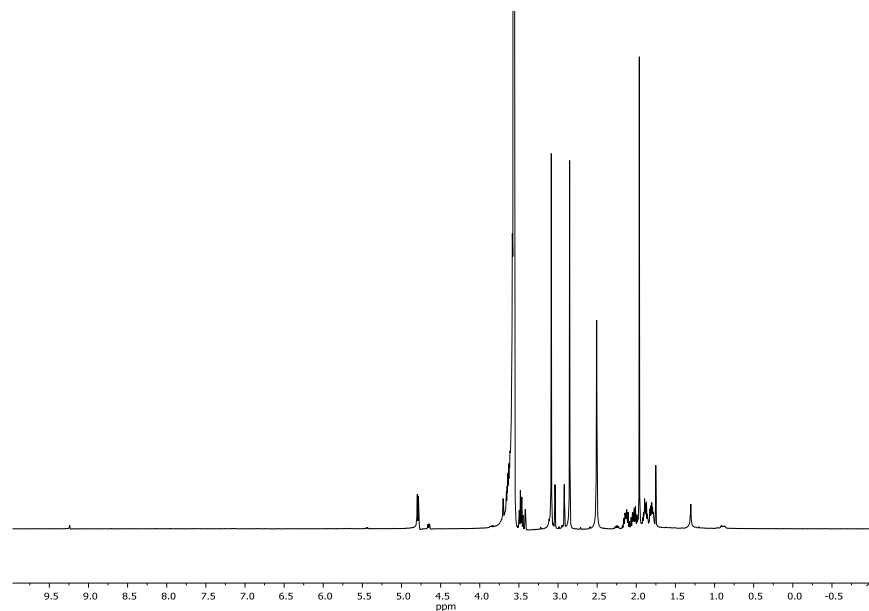


Figure B.1 ^1H NMR spectrum of **1** in 1,4-dioxane- d_8 .

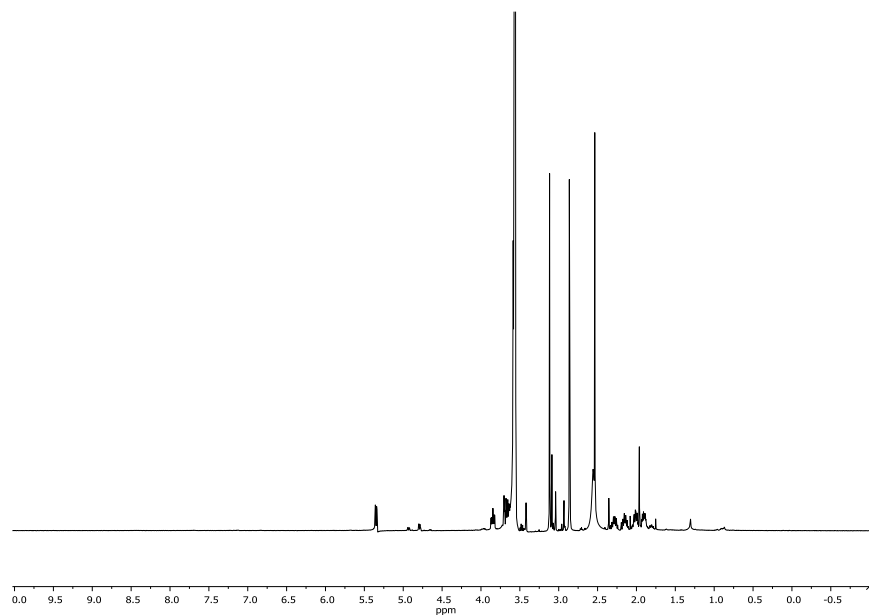


Figure B.2 ^1H NMR spectrum of **2** in 1,4-dioxane- d_8 .

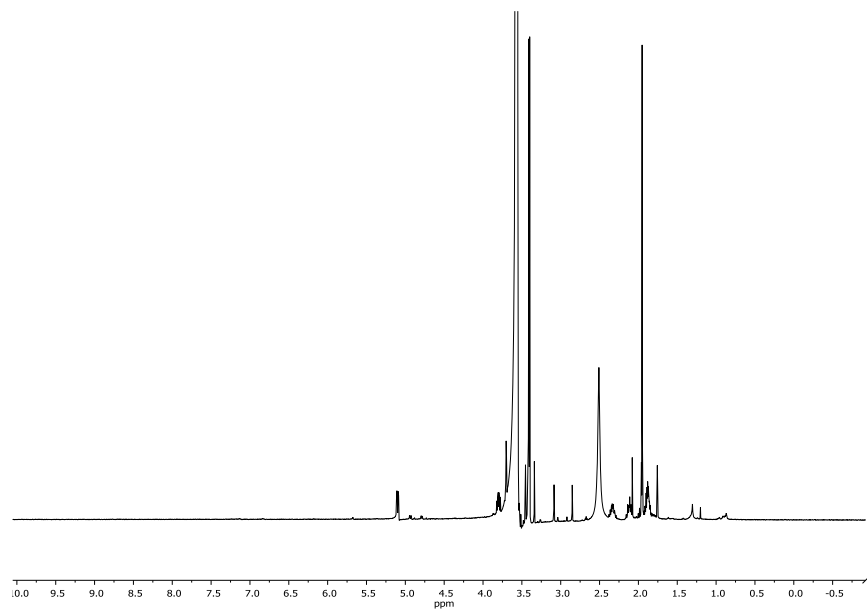


Figure B.3 ^1H NMR spectrum of **3** in 1,4-dioxane- d_8 .

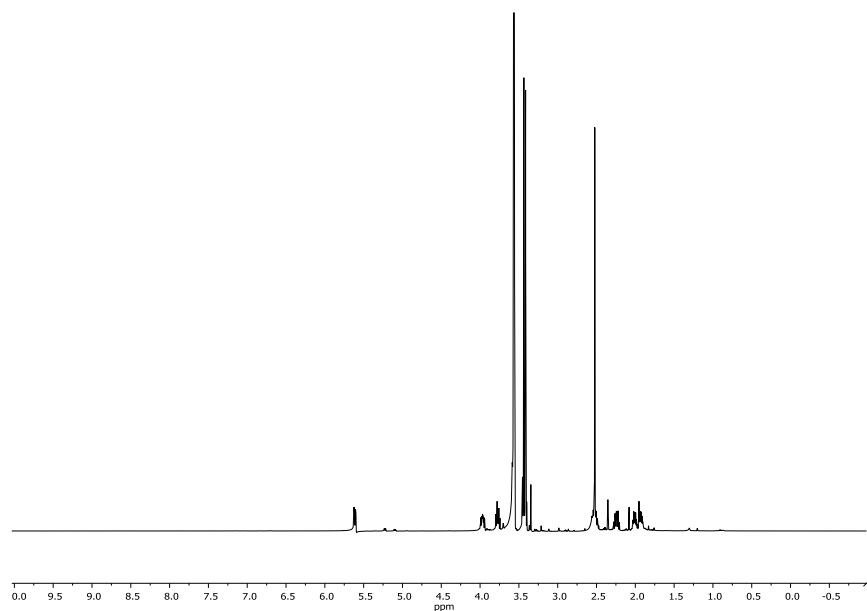


Figure B.4 ^1H NMR spectrum of **4** in 1,4-dioxane- d_8 .

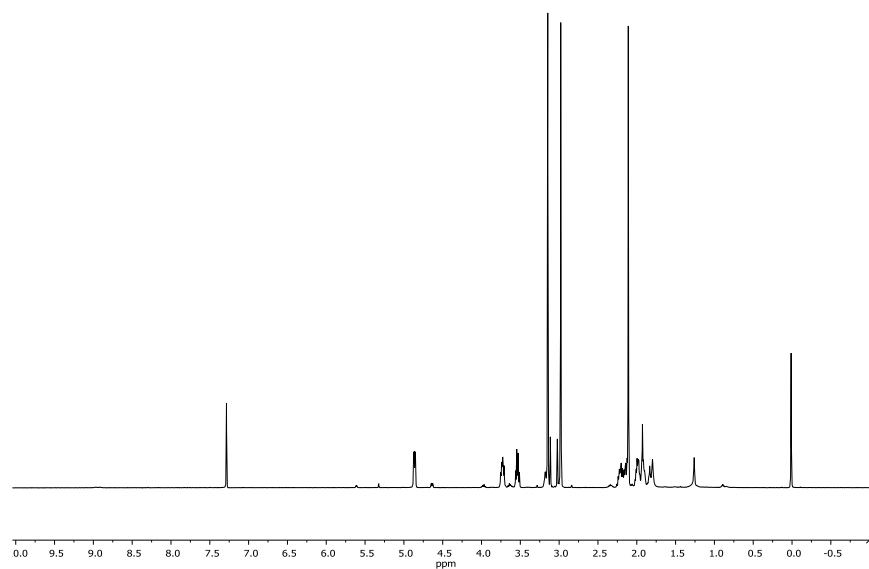


Figure B.5 ^1H NMR spectrum of **1** in chloroform-*d*.

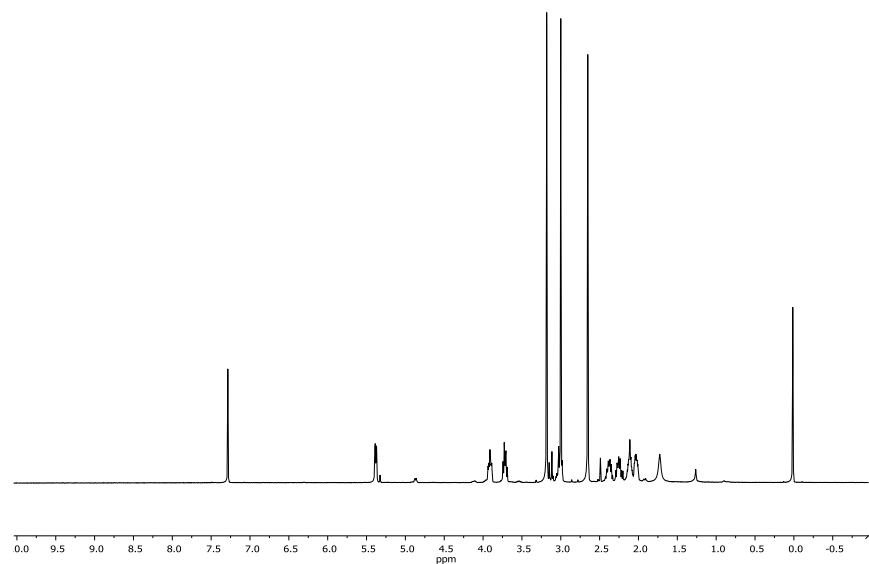


Figure B.6 ^1H NMR spectrum of **2** in chloroform-*d*.

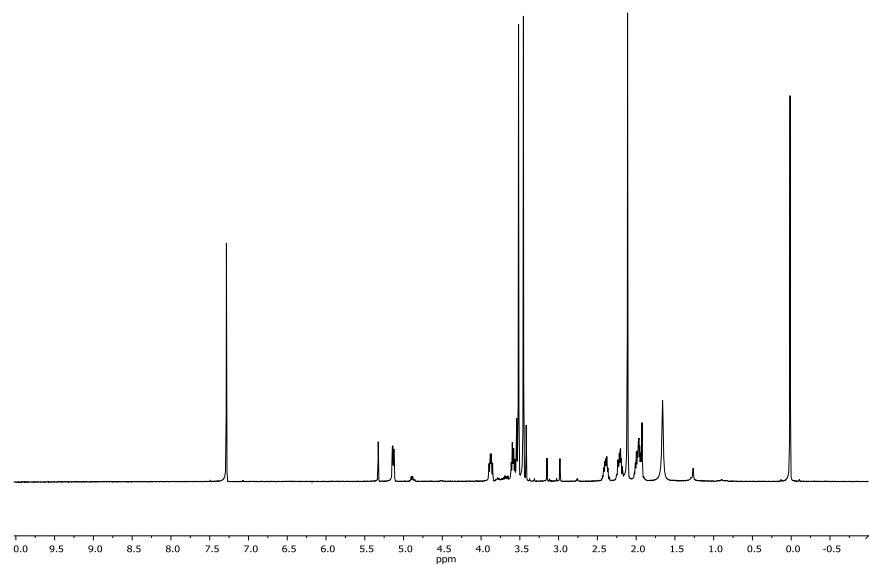


Figure B.7 ^1H NMR spectrum of **3** in chloroform-*d*.

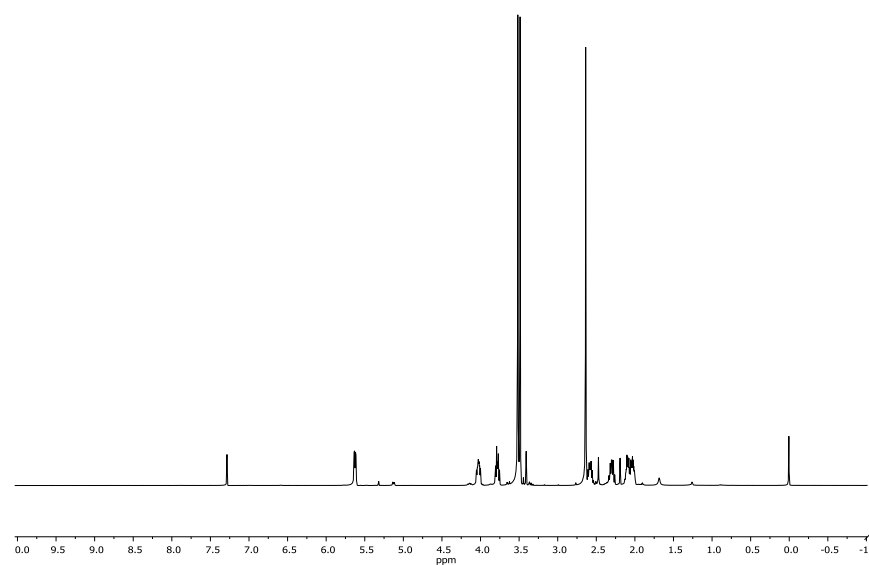


Figure B.8 ^1H NMR spectrum of **4** in chloroform-*d*.

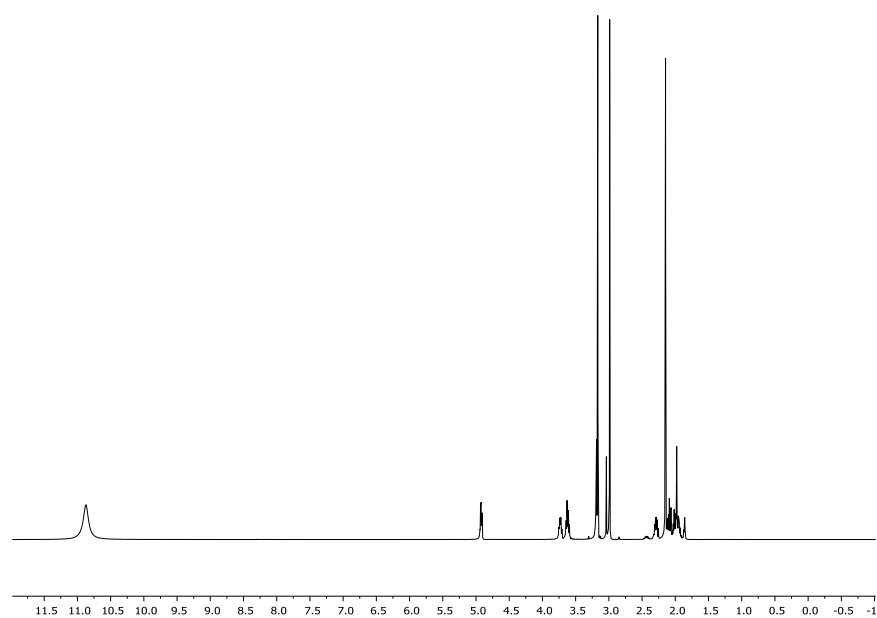


Figure B.9 ^1H NMR spectrum of **1** in acetic acid- d_4 .

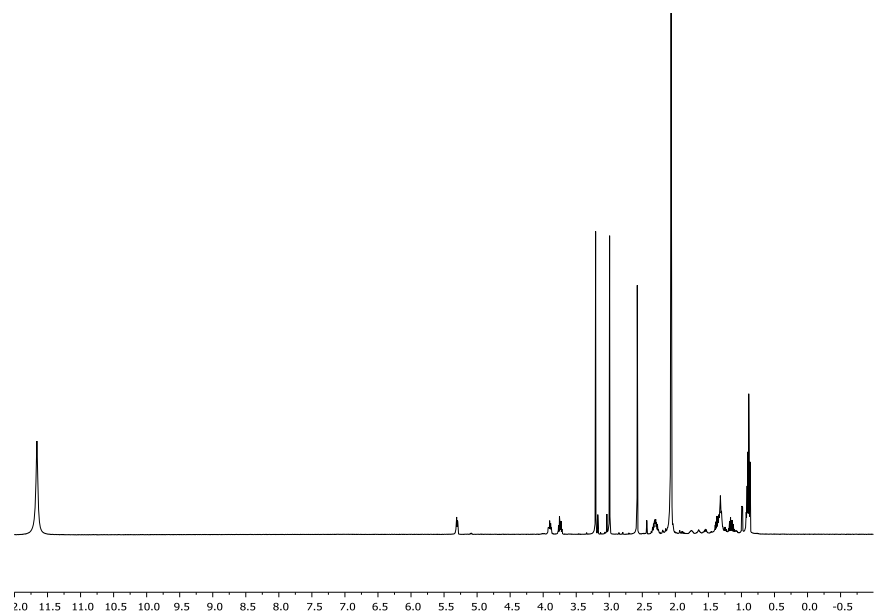


Figure B.10 ^1H NMR spectrum of **2** in acetic acid- d_4 .

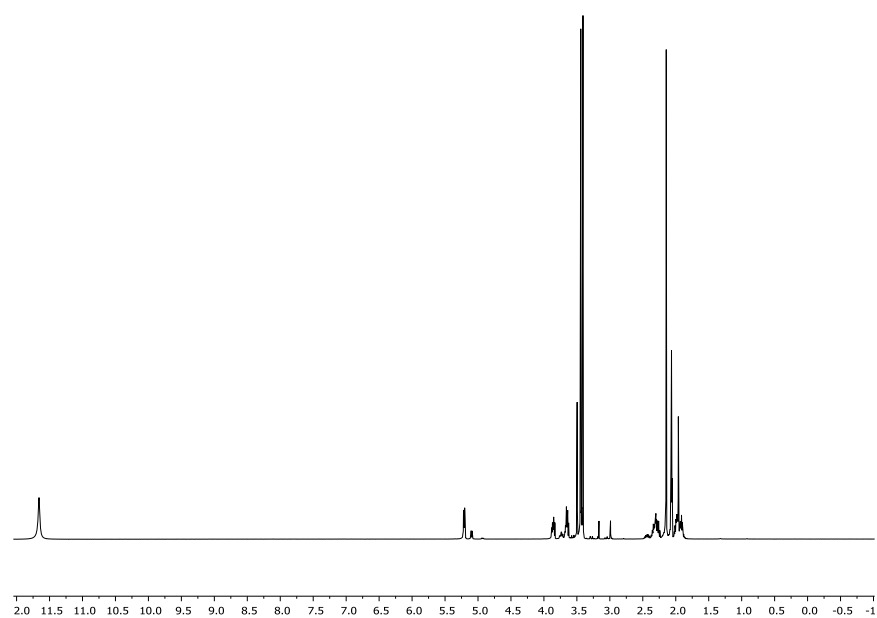


Figure B.11 ^1H NMR spectrum of **3** in acetic acid- d_4 .

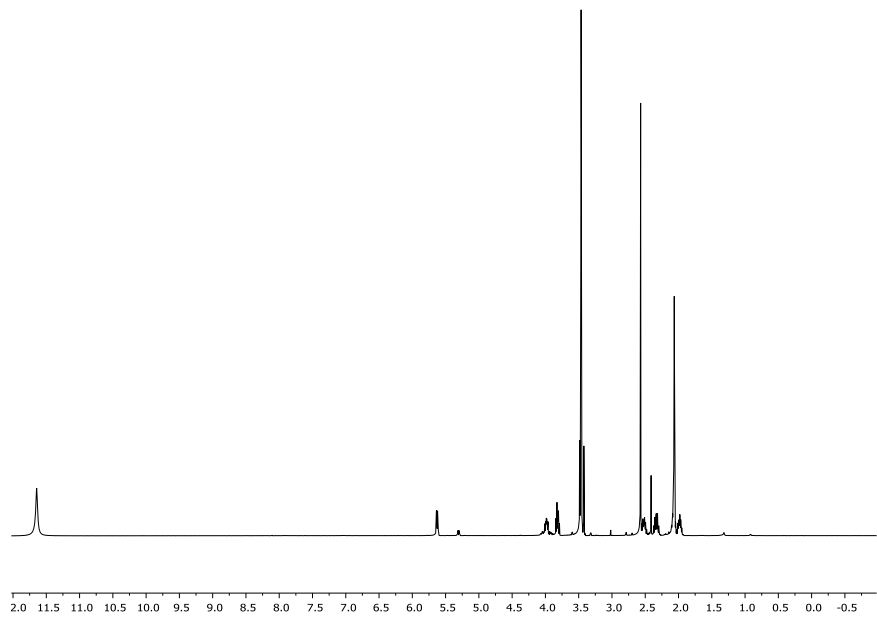


Figure B.12 ^1H NMR spectrum of **4** in acetic acid- d_4 .

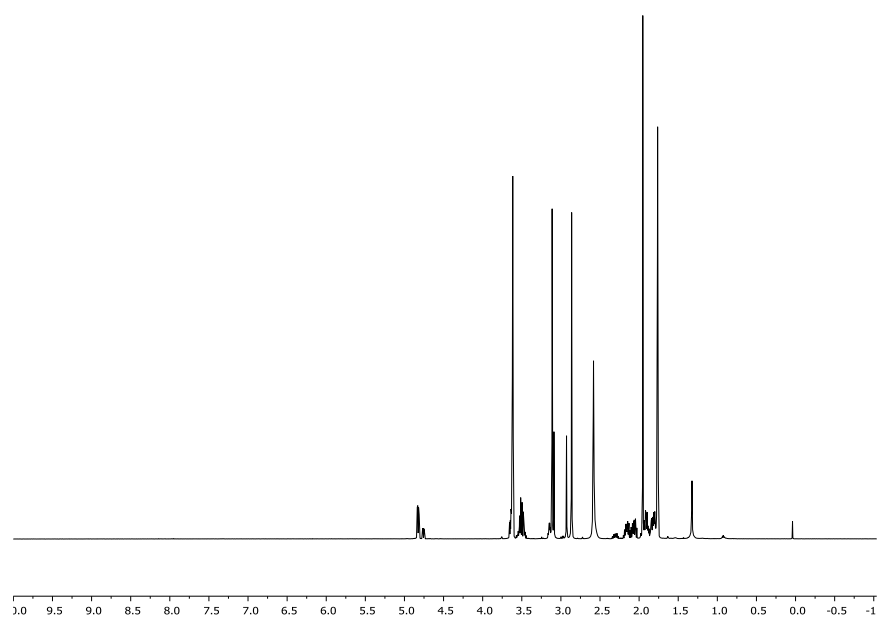


Figure B.13 ^1H NMR spectrum of **1** in $\text{THF-}d_8$.

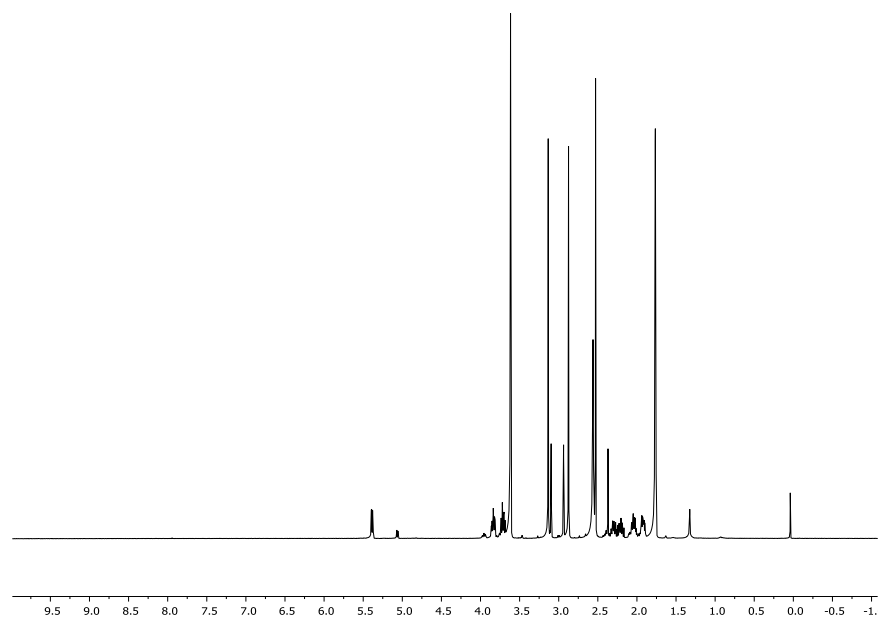


Figure B.14 ^1H NMR spectrum of **2** in $\text{THF-}d_8$.

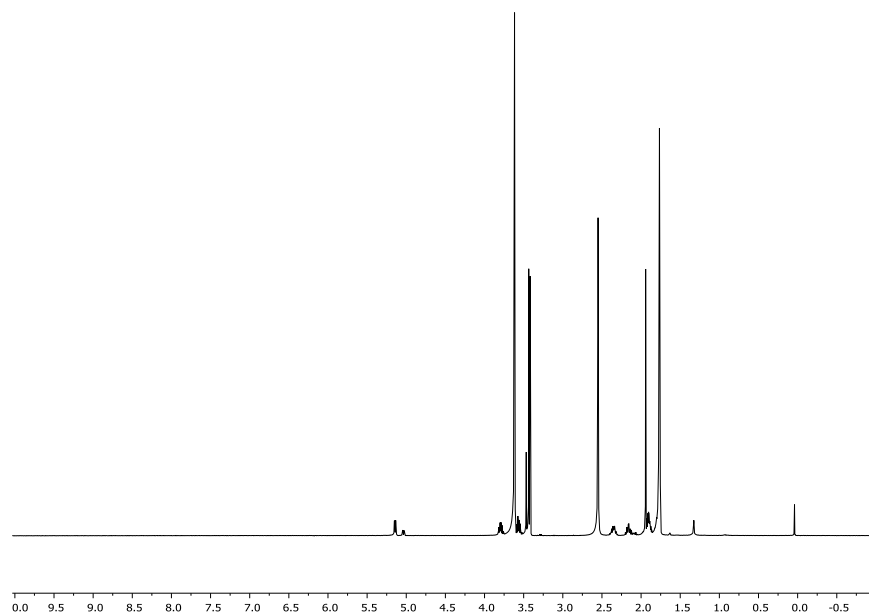


Figure B.15 ^1H NMR spectrum of **3** in $\text{THF-}d_8$.

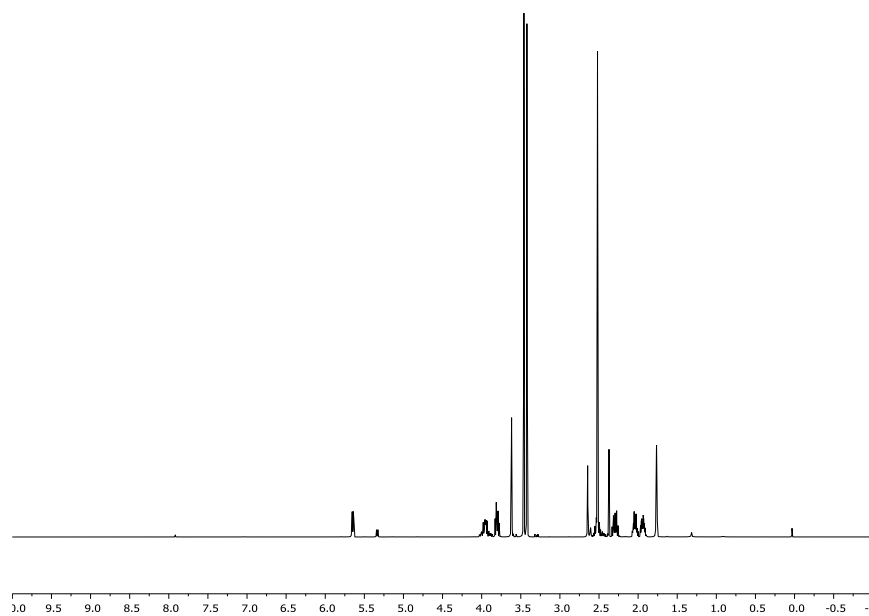


Figure B.16 ^1H NMR spectrum of **4** in $\text{THF-}d_8$.

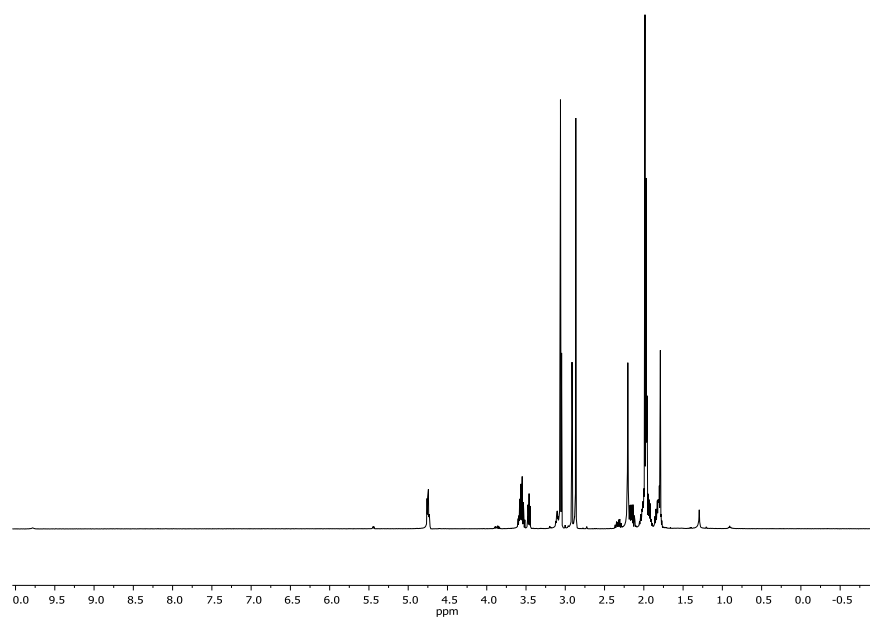


Figure B.17 ^1H NMR spectrum of **1** in acetonitrile- d_3 .

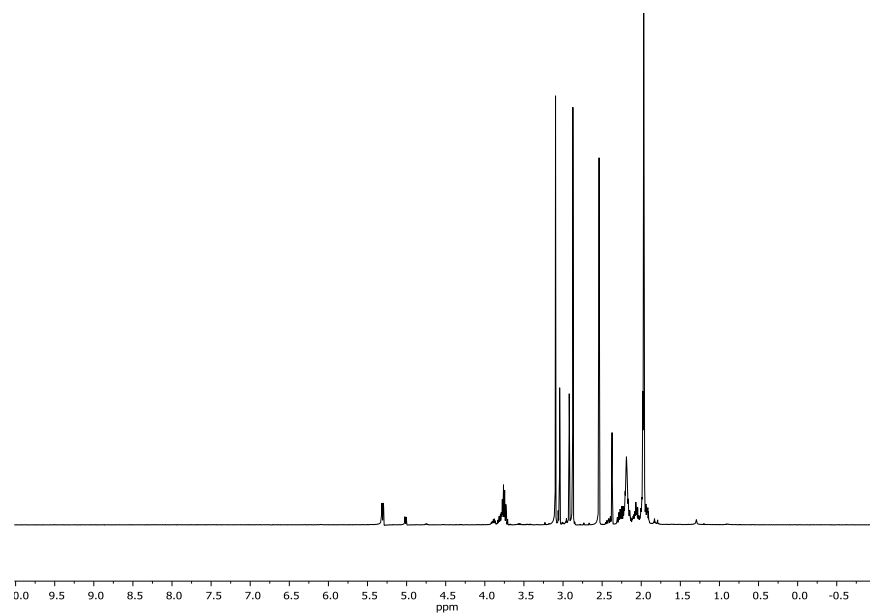


Figure B.18 ^1H NMR spectrum of **2** in acetonitrile- d_3 .

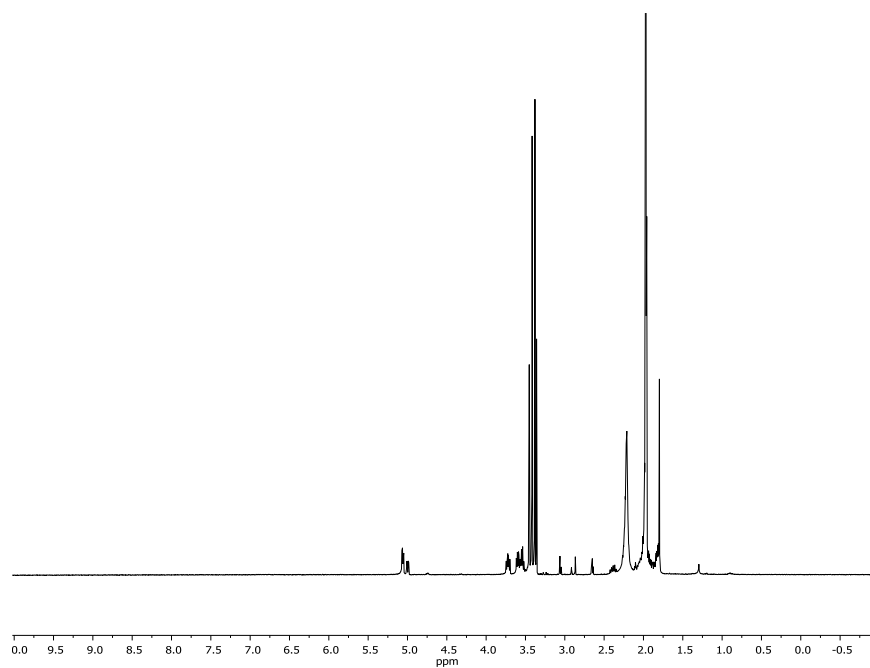


Figure B.19 ^1H NMR spectrum of **3** in acetonitrile- d_3 .

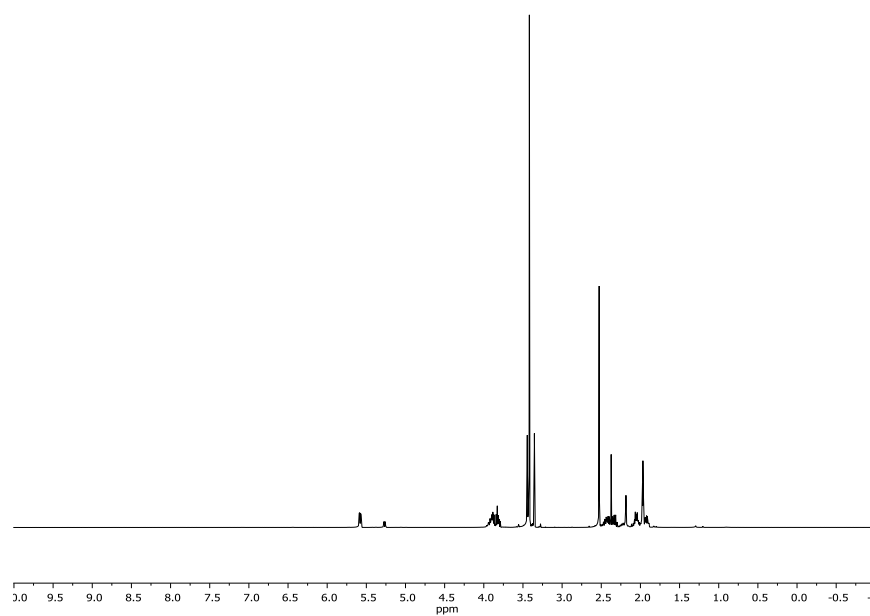


Figure B.20 ^1H NMR spectrum of **4** in acetonitrile- d_3 .

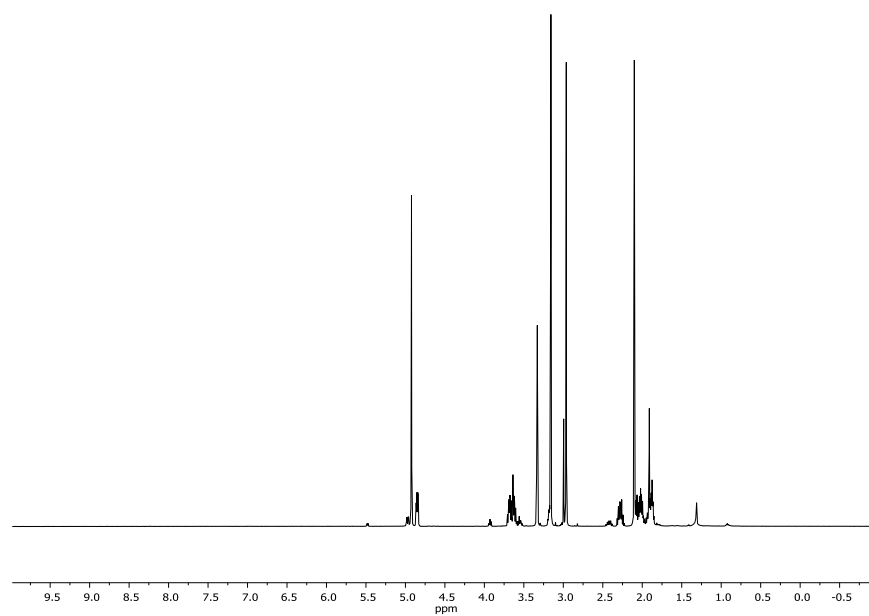


Figure B.21 ^1H NMR spectrum of **1** in methanol- d_4 .

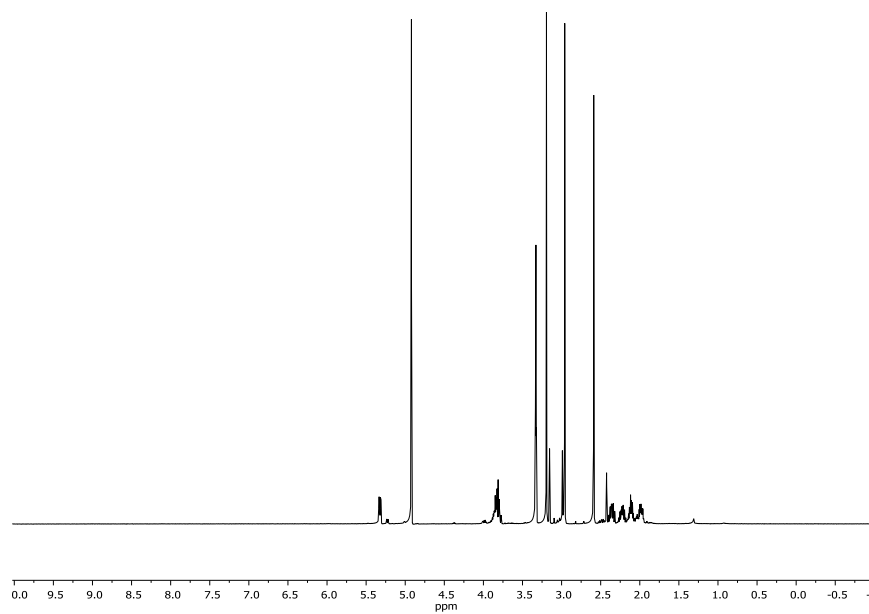


Figure B.22 ^1H NMR spectrum of **2** in methanol- d_4 .

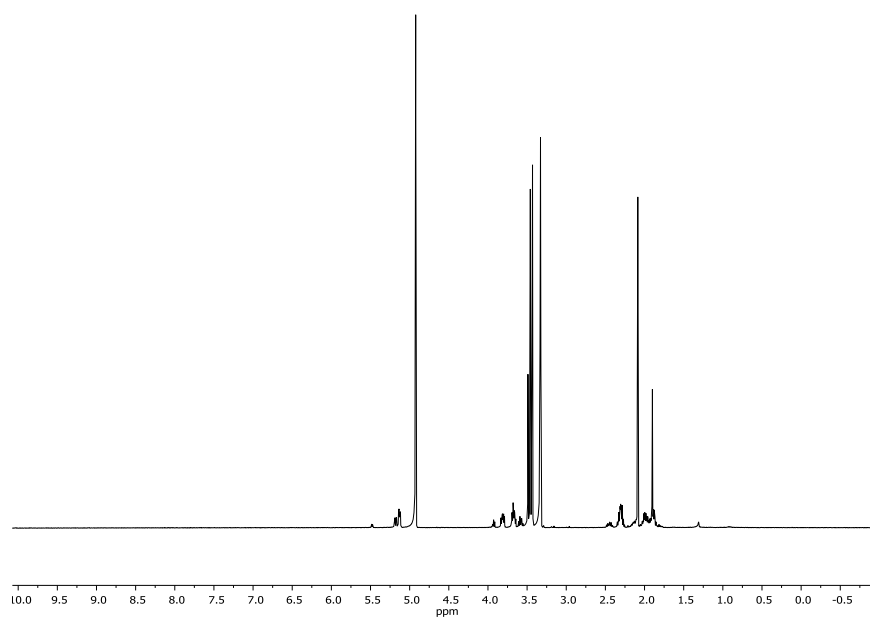


Figure B.23 ^1H NMR spectrum of **3** in methanol- d_4 .

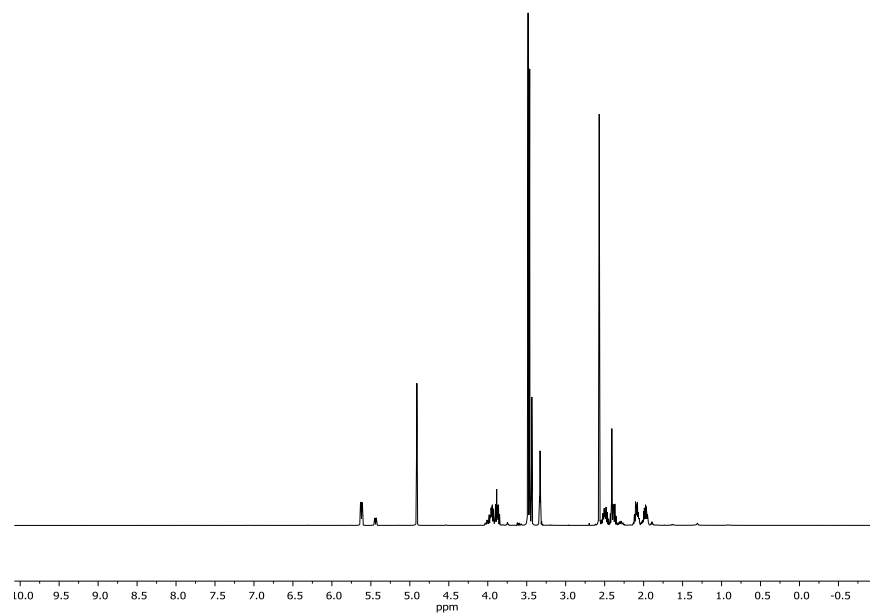


Figure B.24 ^1H NMR spectrum of **4** in methanol- d_4 .

APPENDIX C

Thioesters: Alternative $n \rightarrow \pi^*$ Acceptors

The chemical tools available for modulating $n \rightarrow \pi^*$ interactions remain limited. To expand this toolset, thioesters have been evaluated as acceptors of the $n \rightarrow \pi^*$ interaction. Thioesters are important reactive groups for chemical biology that allow for the construction of complex molecules, and the reactivity of these groups has previously been shown to depend on $n \rightarrow \pi^*$ interactions. Here, it is demonstrated that typical thioesters, as well as thionoesters and dithioesters, engage in significant $n \rightarrow \pi^*$ interactions, thereby providing additional opportunities to perturb $n \rightarrow \pi^*$ interactions in peptides or proteins.

Previous attempts to characterize the $n \rightarrow \pi^*$ interaction have focused on perturbations to the donor or the donor-acceptor distance. In order to evaluate the sensitivity of this interaction to perturbations of the acceptor, a series of *N*-acetyl proline thioesters have been prepared and characterized.

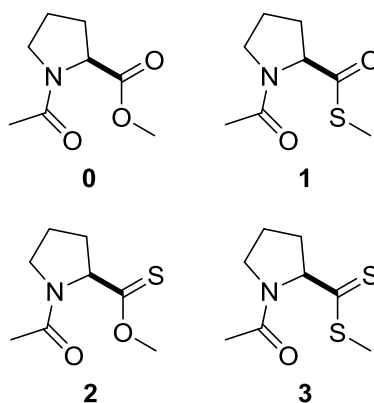
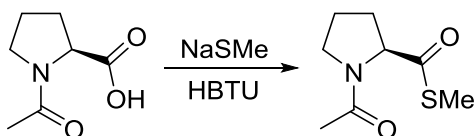


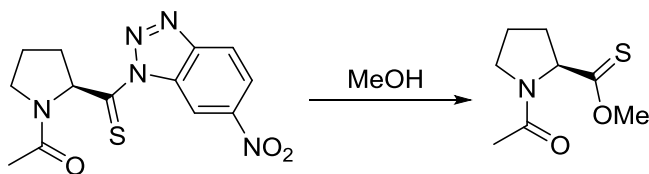
Table C.1 Conformational parameters of proline thioesters

Compound	$K_{trans/cis}$	$E_{n \rightarrow \pi^*}$ (kcal/mol)	
		<i>endo</i>	<i>exo</i>
0	4.60	0.42	1.29
1	2.79	0.33	0.99
2	2.19	0.33	0.99
3	1.91	0.30	0.73

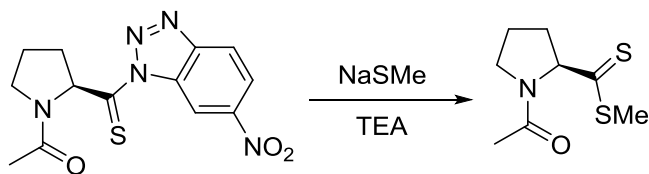
Computational results predicted that the dithioester would be a somewhat weaker electronic acceptor than would be the thioester or thionoesters, as was borne out experimentally. However, computation predicted identical interaction energies for the monothio isomers, but measurement of the conformation preferences of these molecules experimentally indicated a somewhat stronger interaction in the thioester. All three thioester variants attenuated the $n \rightarrow \pi^*$ interaction relative to ester acceptor, though less than does replacement with a dimethyl amide. Strategic thioester substitution may therefore serve as a strategy to make subtle perturbations of the $n \rightarrow \pi^*$ interaction.



***N*-acetyl-(2*S*)-proline methyl thioester (1).** 1.0 g *N*-acetyl proline (6.4 mmol), 2.4 g HBTU (6.4 mmol), and 0.45 g sodium thiomethoxide were dissolved in MeCN and allowed to stir overnight at room temperature. Purification was achieved by reverse-phase HPLC with a 5-95% B gradient over 30 min. using 0.1% TFA in H₂O (A) and 0.1% TFA in MeCN (B) as eluents. Lyophilization afforded a yellow oil. ¹H NMR (D₂O), mixture of two rotamers, δ 4.72 (dd, *J* = 8.8, 2.2 Hz, 0.25H), 4.58 (dd, *J* = 8.8, 3.1 Hz, 0.75H), 3.72-3.40 (m, 2H), 2.30-1.80 (m, 10H), 2.28 (s), 2.24 (s), 2.08 (s), 1.99 (s); ¹³C NMR (D₂O), mixture of two rotamers, δ 208.3, 207.9, 176.4, 176.1, 70.5, 68.6, 51.2, 49.4, 34.4, 32.9, 26.5, 24.6, 24.2, 24.0, 13.8, 13.5. ESI-MS: [M+H]⁺ calculated 188.0740, observed 188.0739.



***N*-acetyl-(2*S*)-proline methyl thionoester (2).** 1-(*N*-acetyl-(2*S*)-thiopropyl)-6-nitrobenzotriazole was refluxed in MeOH for 3 h. Purification was achieved by reverse-phase HPLC with a 5-95% B gradient over 30 min. using 0.1% TFA in H₂O (A) and 0.1% TFA in MeCN (B) as eluents. Lyophilization afforded a yellow oil. ¹H NMR (D₂O), mixture of two rotamers, δ 4.87 (dd, *J* = 8.8, 2.4 Hz, 0.35H), 4.75 (dd, *J* = 8.8, 3.1 Hz, 0.65H), 4.12 (s, 0.94H), 4.08 (s, 2.06H), 3.75-3.42 (m, 2H), 2.40-1.85 (m, 7H), 2.07 (s), 1.93 (s); ¹³C NMR (D₂O), mixture of two rotamers, δ 225.6, 224.7, 176.2, 175.4, 72.1, 70.6, 62.8, 62.5, 51.3, 49.4, 35.6, 34.5, 26.2, 24.3, 24.1. ESI-MS: [M+H]⁺ calculated 188.0740, observed 188.0739.



***N*-acetyl-(2*S*)-proline methyl dithioester (3).** 0.2 g 1-(*N*-acetyl-(2*S*)-thiopropyl)-6-nitrobenzotriazole (0.6 mmol), 0.04 g sodium thiomethoxide (0.6 mmol), and 0.09 mL triethylamine (0.6 mmol) were dissolved in MeCN and allowed to stir overnight at room temperature. Purification was achieved by reverse-phase HPLC with a 5-95% B gradient over 30 min. using 0.1% TFA in H₂O (A) and 0.1% TFA in MeCN (B) as eluents. Lyophilization afforded a yellow oil. ¹H NMR (D₂O), mixture of two rotamers, δ 5.25 (dd, *J* = 8.8, 2.4 Hz, 0.33H), 5.15 (dd, *J* = 8.9, 3.1 Hz, 0.66H), 3.82-3.50 (m, 2H), 2.64 (s, 1.15H), 2.60 (s, 2.16H), 2.50-1.85 (m, 7H), 2.10 (s), 1.94 (s); ¹³C NMR (D₂O), mixture of two rotamers, δ 244.4, 243.8, 176.6, 175.9, 77.2, 75.4, 51.7, 50.0, 38.5, 37.1, 26.1, 24.4, 24.2, .4.1, 21.7, 21.4. ESI-MS: [M+H]⁺ calculated 204.0511, observed 204.0510.

Computational methodology. Geometries of **1**, **2**, and **3** were optimized for the *trans* conformation of the *N*-acetyl peptide bond and both the *endo* and *exo* conformations of the pyrrolidine ring at the B3LYP/6-311+G(2d,p) level of theory using Gaussian 09. Optimized geometries yielded no imaginary frequencies and were subject to NBO analysis at the same level of theory using NBO 5.9 as implemented in Gaussian.

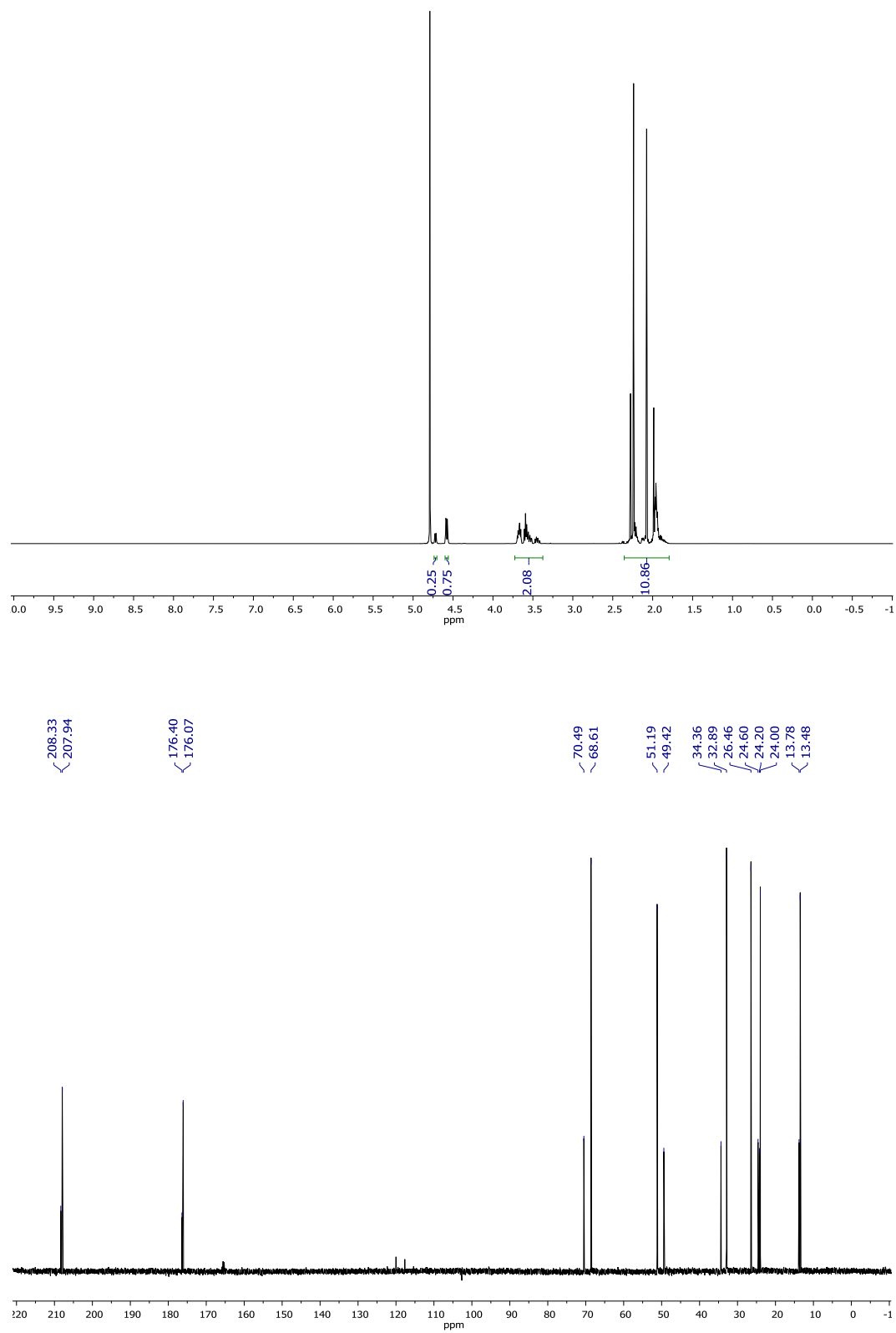


Figure C.1 1H and ^{13}C NMR spectra of **1** in D_2O .

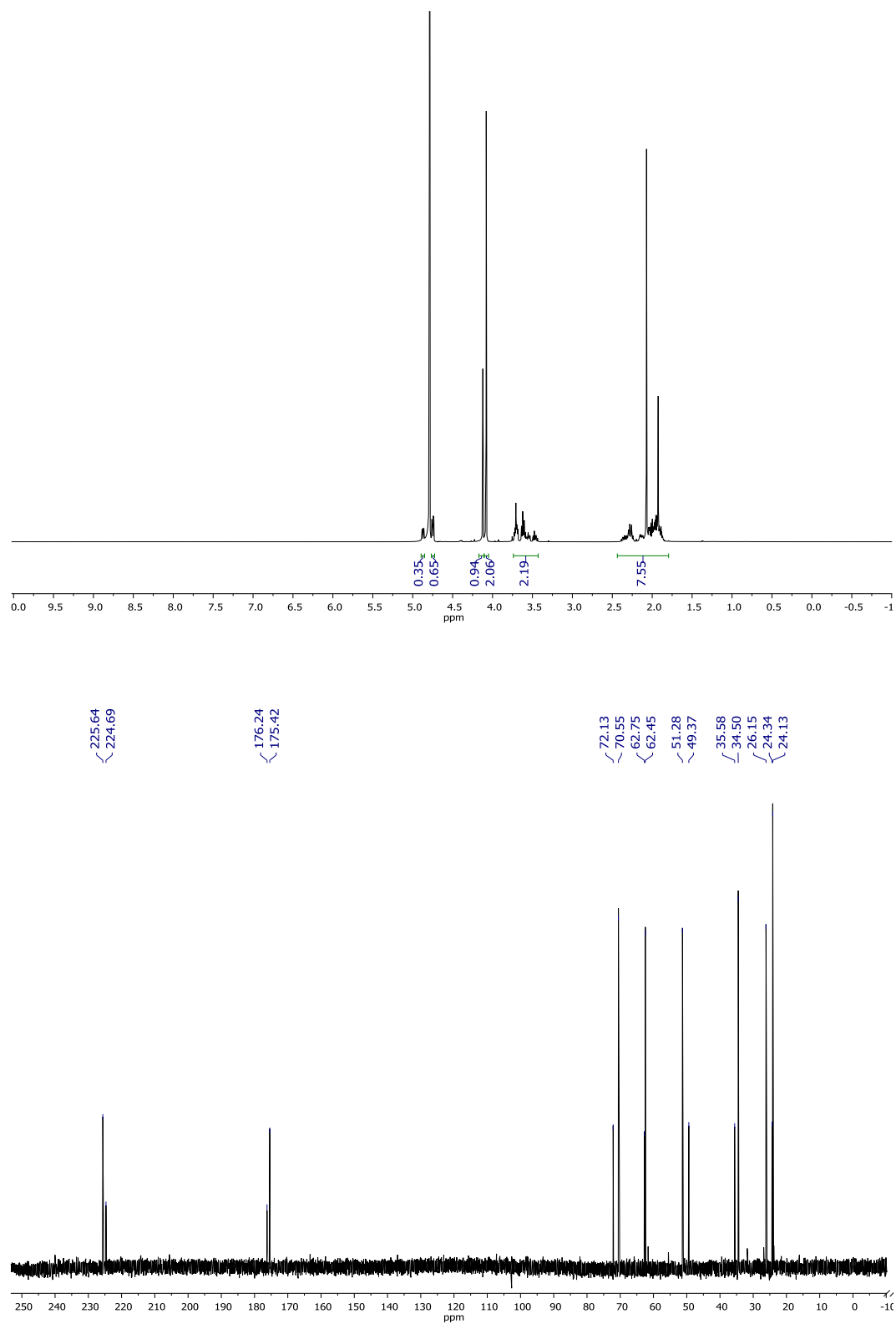


Figure C.2 ¹H and ¹³C NMR spectra of 2 in D₂O.

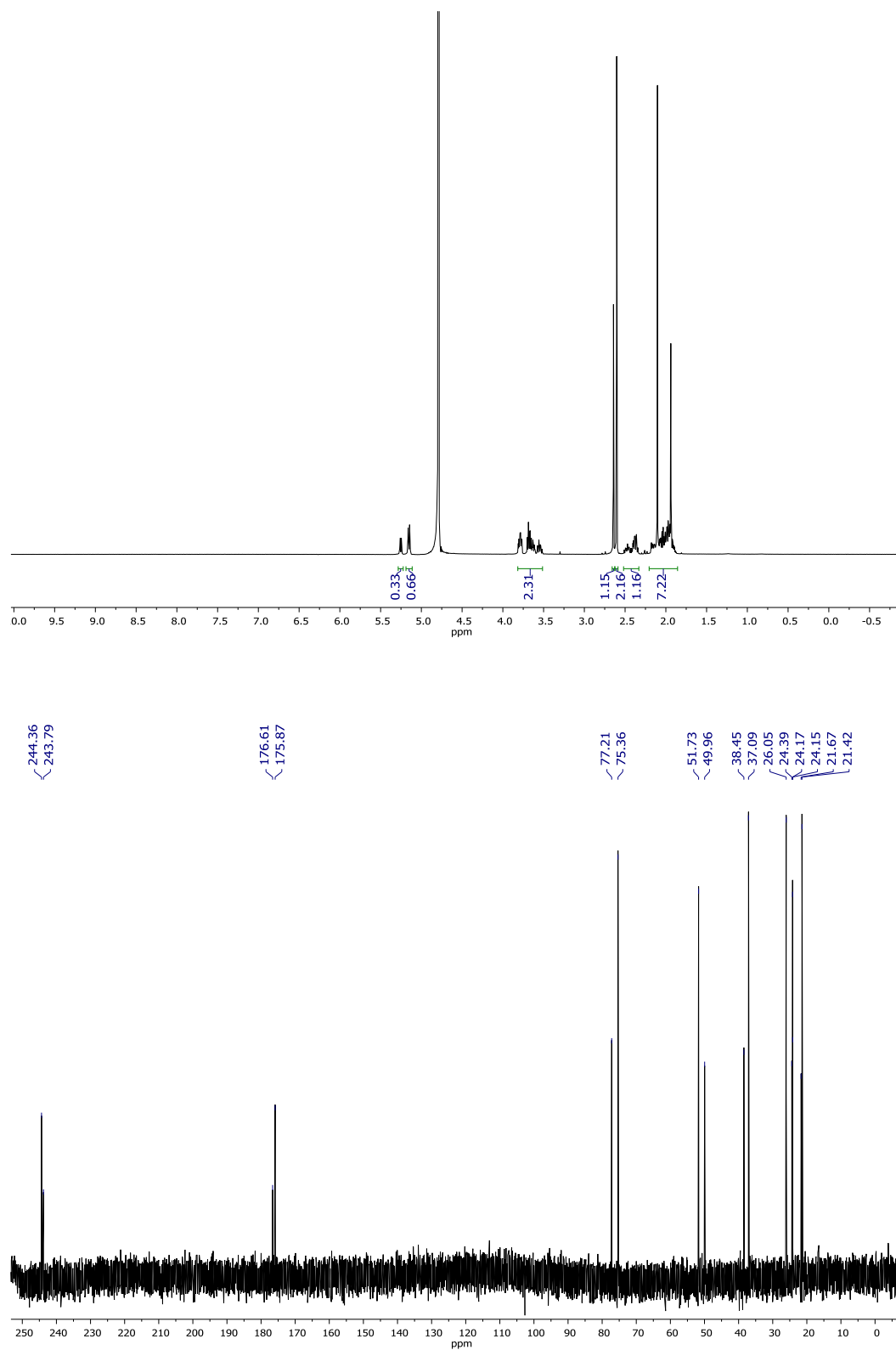


Figure C.3 ¹H and ¹³C NMR spectra of **3** in D₂O.

Table C.2 Cartesian coordinates of the optimized *trans/endo* geometry of **1**.

C	-0.12274300	0.45232300	-0.71312200
H	0.04683900	-0.00504800	-1.68630600
C	-2.18887600	1.15596500	0.38269800
H	-2.53231500	0.93039100	1.39104600
H	-3.06607700	1.36198600	-0.23717500
C	-1.20306400	2.33062200	0.35984500
H	-0.65108900	2.36656400	1.29770600
H	-1.70547100	3.28650200	0.22416000
C	-0.25438700	1.98813000	-0.79728600
H	-0.70405900	2.24870300	-1.75608600
H	0.70641200	2.49368600	-0.73195700
N	-1.41317300	0.04044200	-0.17698300
C	-1.76870900	-1.27684800	-0.18969800
C	-3.12247600	-1.62180300	0.39665200
H	-3.13073000	-1.44036900	1.47262400
H	-3.92025200	-1.02430200	-0.04588700
H	-3.31274200	-2.67433600	0.21401600
C	1.03153400	0.05146400	0.21865700
C	3.66596000	-0.55358200	0.75632800
H	3.29323600	-1.38739200	1.34298200
H	4.64063100	-0.79364500	0.33970200
H	3.73231500	0.33540300	1.37607000
S	2.56035100	-0.26018200	-0.66007700
O	0.93646500	0.00645500	1.41882900
O	-1.02315700	-2.12844900	-0.65448200

Table C.3 Cartesian coordinates of the optimized *trans/exo* geometry of **1**.

C	-0.14857600	0.61567700	-0.49514500
C	-0.38747400	2.11538500	-0.19016700
C	-1.91552200	2.24949800	-0.13402000
C	-2.35944100	0.91999200	0.47910000
H	0.01820400	0.43836200	-1.55874100
H	0.04264100	2.36177500	0.78053200
H	0.07578100	2.75614800	-0.93614200
H	-2.24110100	3.10858800	0.44980800
H	-2.32717100	2.34775300	-1.13943400
H	-2.29755300	0.94995300	1.57105400
H	-3.37522600	0.64827100	0.20178900
C	-1.54445200	-1.37483700	-0.23523900
C	1.06442300	0.09460600	0.28486500
O	-0.66090300	-2.05044900	-0.74544200
O	1.03255900	-0.16736000	1.45986200
C	-2.85031500	-1.98117600	0.23796000
H	-3.69147200	-1.61232300	-0.35130800
H	-2.78267400	-3.05774100	0.12060100
H	-3.04538400	-1.74110000	1.28354100
C	3.74183800	-0.46989800	0.58167400
H	3.77444500	0.27275700	1.37287000
H	3.45529500	-1.43407600	0.99006100
H	4.71010200	-0.54201600	0.09337900
N	-1.39018100	-0.02970100	-0.07332200
S	2.55245300	0.02740900	-0.70339300

Table C.4 Cartesian coordinates of the optimized *trans/endo* geometry of **2**.

C	0.03897800	0.43019000	-0.83615600
H	0.12539800	-0.02687800	-1.81996900
C	-1.93756500	1.24139000	0.34876600
H	-2.23027800	1.03940400	1.37817200
H	-2.84007800	1.47667000	-0.22191300
C	-0.90907400	2.37592200	0.25912900
H	-0.32454700	2.41642200	1.17614000
H	-1.38093700	3.34607700	0.11458500
C	-0.01343400	1.97293900	-0.92100200
H	-0.47554100	2.25152200	-1.86866400
H	0.97423900	2.42694800	-0.88612000
N	-1.24234300	0.08900200	-0.24022500
C	-1.70333100	-1.19422500	-0.30153400
C	-3.05552300	-1.45936500	0.32781100
H	-3.01608600	-1.29352100	1.40551800
H	-3.82620800	-0.80368600	-0.07943900
H	-3.32209900	-2.49360600	0.13636900
C	1.24245900	-0.05257100	-0.02892500
O	2.25858300	-0.24825500	-0.86322900
C	3.52643900	-0.68257100	-0.34273300
H	3.40789300	-1.62951700	0.17868100
H	4.16510100	-0.79301500	-1.21302900
H	3.92556000	0.05982100	0.34523200
S	1.29406900	-0.23568700	1.60022200
O	-1.04918500	-2.07911000	-0.83673200

Table C.5 Cartesian coordinates of the optimized *trans/exo* geometry of **2**.

C	-0.01653200	0.62587400	-0.64486100
C	-0.24118900	2.13344800	-0.36445700
C	-1.76192100	2.26433100	-0.20177200
C	-2.15270500	0.94831900	0.47438300
H	0.06915000	0.42759100	-1.71347000
H	0.25661600	2.40723400	0.56473200
H	0.16320900	2.75171900	-1.16239400
H	-2.04772200	3.13648500	0.38363000
H	-2.24679200	2.33633000	-1.17626900
H	-2.00620800	1.00239500	1.55696300
H	-3.18513100	0.66758200	0.28204400
C	-1.42265600	-1.34536100	-0.32936100
C	1.26572800	0.11372200	0.00171500
C	-2.70589100	-1.94296000	0.21220900
H	-3.57604100	-1.55059000	-0.31683400
H	-2.66233100	-3.01759100	0.06820800
H	-2.83241100	-1.72239500	1.27236100
C	3.53296300	-0.39905300	-0.53963900
H	3.95273300	0.26489400	0.21293800
H	3.47993500	-1.40936500	-0.14071500
H	4.11431300	-0.37176100	-1.45546600
N	-1.22440600	-0.01171800	-0.12738100
S	1.44289600	-0.21360100	1.59964000
O	-0.59210600	-2.02163500	-0.92229700
O	2.21940600	0.04503300	-0.91865400

Table C.6 Cartesian coordinates of the optimized *trans/endo* geometry of **3**.

C	-0.20528000	0.40858900	-0.81144100
H	-0.06945900	-0.10024700	-1.76566500
C	-2.28336700	1.14482000	0.24790200
H	-2.59536400	0.99239000	1.28026900
H	-3.18048300	1.26698600	-0.36512400
C	-1.34264000	2.34792300	0.10881500
H	-0.80120200	2.50056700	1.04000400
H	-1.88144500	3.26420500	-0.12577600
C	-0.37049600	1.93478100	-1.00460300
H	-0.81052600	2.11198900	-1.98673900
H	0.57862800	2.46346200	-0.96355000
N	-1.48289900	0.01162200	-0.23552600
C	-1.85128000	-1.30277100	-0.22841600
C	-3.20377600	-1.62365000	0.37436800
H	-3.21556900	-1.38541900	1.43905400
H	-4.00302400	-1.05325600	-0.10052200
H	-3.38990600	-2.68470800	0.24461900
C	1.00026100	0.06939600	0.07543700
C	3.72790300	-0.53708400	0.27528700
H	3.48920200	-1.35454000	0.94953700
H	4.60875500	-0.77656000	-0.31621900
H	3.88677500	0.37415700	0.84476500
S	0.98526500	0.12333600	1.71298800
O	-1.11897400	-2.17034100	-0.68426900
S	2.38271900	-0.31019100	-0.91773300

Table C.7 Cartesian coordinates of the optimized *trans/exo* geometry of **3**.

C	-0.23761400	0.61135700	-0.59608200
C	-0.49518200	2.12334100	-0.36234400
C	-2.02136500	2.23723900	-0.24826800
C	-2.41622700	0.92880100	0.43972200
H	-0.11974700	0.39549100	-1.66015100
H	-0.02754700	2.42799200	0.57237300
H	-0.07530700	2.72595600	-1.16420500
H	-2.33482200	3.11614800	0.31248100
H	-2.47827100	2.28577000	-1.23765800
H	-2.29749600	1.00219100	1.52444800
H	-3.43997500	0.63333500	0.22511900
C	-1.65505600	-1.36528100	-0.33055700
C	1.02412600	0.13672600	0.13002200
C	-2.95451300	-1.96265300	0.17270200
H	-3.80739700	-1.57705800	-0.38846600
H	-2.90280700	-3.03801800	0.03701200
H	-3.11787000	-1.73562300	1.22636100
C	3.76358500	-0.42107000	0.06930400
H	3.92555500	0.35711000	0.80978400
H	3.58287100	-1.36955400	0.56686000
H	4.61904600	-0.49860400	-0.59800400
N	-1.46262200	-0.03126700	-0.12119900
S	2.35207900	-0.00265400	-0.98685000
S	1.09422000	-0.11605700	1.74765700
O	-0.80940000	-2.04220700	-0.89968300

APPENDIX D

$n \rightarrow \pi^*$ Interactions between Proteins and their Ligands

The role of $n \rightarrow \pi^*$ interactions in intermolecular associations remains undercharacterized. To address this challenge, contacts between proteins and their ligands were identified from Relibase using geometric criteria. Three main groups of hits were identified: peptides, flavins, and pendant glycans. These results may illuminate specific functional roles of the $n \rightarrow \pi^*$ interaction.

The $n \rightarrow \pi^*$ interaction mediates molecular recognition events in a wide variety of molecules. Most of these events involve intramolecular contacts, given the relatively low energy of individual $n \rightarrow \pi^*$ interactions. The role of $n \rightarrow \pi^*$ interactions in intermolecular associations is less clear. While we have posited that intramolecular $n \rightarrow \pi^*$ interactions can affect intermolecular association indirectly, as in the case of *N*-acyl homoserine lactones,¹⁷⁸ examples of true intermolecular contacts mediated by $n \rightarrow \pi^*$ interactions are scarce.²⁰⁵ It is unclear, therefore, if $n \rightarrow \pi^*$ interactions can contribute to ligand binding by proteins.

To address this issue, $n \rightarrow \pi^*$ interactions between proteins and ligands were identified from the Relibase dataset.⁵¹⁹ Specifically, the entire database was queried for instances in which a carbonyl oxygen on a protein donates an $n \rightarrow \pi^*$ interaction to a carbonyl group on a ligand, based on geometric criteria.⁹² Analysis focused on donation from the protein to the ligand since organic molecules often feature more potent electrophiles than proteins.

This query produced 606 hits. Among them, three major subgroups were identified. The first were peptide ligands. While some of these ligands appeared to bind to protein active sites, others appeared to be artifacts of crystallization. The next major group are the flavins (Table D.1), which often feature contact from an amide side chain (Figure D.1). These contacts do not appear to involve hydrogen bonding, suggesting a significant contribution from the $n \rightarrow \pi^*$ interaction itself. Finally, a very common $n \rightarrow \pi^*$ interaction was observed at asparagine residues that undergo N-glycosylation (Figure and Table D.2). While this interaction is not intermolecular, it may help to organize the glycan relative to the protein.

These results provide the first insights into the prevalence of $n \rightarrow \pi^*$ interactions in protein-ligand or protein-appendage interactions. This approach suffers from several shortcomings. First, the entire Relibase dataset was considered without regard to sequence redundancy or

resolution. Confident assignment of $n \rightarrow \pi^*$ interactions between proteins and ligands, as well as their frequency, will require improvements in both areas, which are fortunately straightforward. More targeted results could also be obtained by constraining the search to enzyme active sites, the identification of which can be achieved with bioinformatics methods.⁵⁰⁰⁻⁵⁰¹

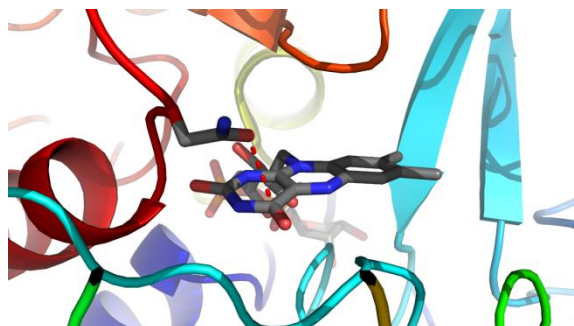


Figure D.1 Representative $n \rightarrow \pi^*$ interaction (red) between an asparagine side chain and a bound molecule of flavin adenine dinucleotide (PDB ID: 3cox).

Table D.1 Flavins receiving $n \rightarrow \pi^*$ interactions from proteins

PDB ID	Flavin	PDB ID	Flavin
3u33	FAD	3hf3	FMN
2dvl	FAD	2wqf	FMN
2r0p	FAD	2jbt	FMN
2r0c	FAD	3ek3	FMN
2jif	FAD	3gfq	FMN
3nf4	FAD	3s2y	FMN
3mdd	FAD	4h6p	FMN
1egd	FAD	1llw	FMN
1ivh	FAD	1ea0	FMN
1jqi	FAD	1ofd	FMN
3mpj	FAD	1ofe	FMN
1ukw	FAD	3mhu	FMN
4hr3	FAD	3gr7	FMN
3cox	FAD	3gr8	FMN
3djl	FAD	2q3r	FMN
2vig	FAD	1gvr	FMN
4eip	FAD	3kru	FMN
2gmh	FAD	3krz	FMN
2gmj	FAD	1v5z	FMN
3awi	FAD	2isj	FMN
		2ifa	FMN
		2oz0	FMN
		1ldc	FMN
		2q3o	FMN

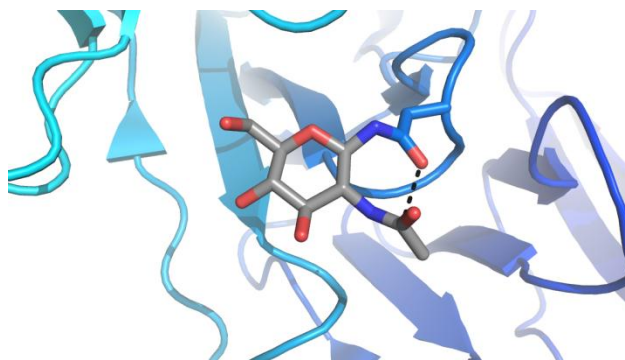


Figure D.2 Representative $n \rightarrow \pi^*$ interaction (black) between an asparagine side chain and a pendant *N*-acetylglucosamine (PDB ID: 3nox).

Table D.2 Glycans receiving $n \rightarrow \pi^*$ interactions from proteins

PDB ID	Glycan	PDB ID	Glycan
1bhg	MAN-MAN-MAN-MAN-MAN-NAG	4e36	NAG
1zpu	MAN-MAN-NAG-NAG-BMA-MAN	2cgy	NAG
3kq4	MAN-NAG-NAG-BMA-MAN	2va9	NAG
3iai	MAN-NAG-NAG-BMA-MAN	2v98	NAG
4dl1	MAN-NAG-NAG-BMA-MAN	1yym	NAG
3ltf	MAN-NAG-NAG-MAN-MAN-MAN	3kls	NAG
3ala	NAG-FUC	1kcd	NAG
4i78	NAG-FUC	3l4v	NAG
3l5h	NAG-NAG-BMA-BMA	4dda	NAG
1ivg	NAG-NAG-BMA-MAN-MAN-MAN	3d5o	NAG
3ula	NAG-NAG-BMA	3nox	NAG
1zpu	NAG-NAG-BMA	3sx4	NAG
4exn	NAG-NAG-BMA	2wg1	NAG
3csy	NAG-NAG-BMA	3o95	NAG
1ypz	NAG-NAG-MAN	1zag	NAG
2e9e	NAG-NAG-MAN	1v3d	NAG
2g41	NAG-NAG-NAG	3ije	NAG
2dp8	NAG-NAG-NAG	2vc2	NAG
2xwb	NAG-NAG	3t3m	NAG
1bcs	NAG-NAG	2vdm	NAG
1whs	NAG-NAG	3sm5	NAG
3lrl	NAG-NAG	1uhg	NAG
1n1m	NAG-NAG	4kbp	NAG
2rft	NAG-NAG	3h6g	NAG
3csy	NAG-NAG	1ieb	NAG
2r9j	NAG-NAG	2rl8	NAG
3usa	NAG-NAG	1som	NAG

1myp	NAG-NAG	4dte	NAG
2zmb	NAG-NAG	1s5r	NAG
3qf1	NAG-NAG	3vi4	NAG
4eby	NAG-NAG	4dvs	NAG
1oyh	NAG-NAG	2ajl	NAG
1gpk	NAG-NDG	3d67	NAG
1jv2	NAG-NDG	3al4	NAG
1m1x	NAG-NDG	3vtr	NAG
1d0m	NAG-NDG	3d66	NAG
1flc	NDG-NAG-BMA	3d67	NAG
2dqz	SIA	4cox	NAG
3gvk	SLB-SIA-SIA	3qyt	NAG
3ulv	NAG	3fvc	NAG
3g0g	NAG	3m3d	NAG
1dbn	NAG	3pps	NAG
2ft3	NAG	1nu6	NAG
3jwo	NAG	3qbj	NAG
1jjb	NAG	3ccc	NAG
1cfj	NAG		

REFERENCES

1. Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012**, 338, 1042-1046.
2. Anfinsen, C. B.; Haber, E.; Sela, M.; F. H. White, J. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **1961**, 47, 1309-1314.
3. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **1973**, 181, 223-230.
4. Venter, J. C., et al. The sequence of the human genome. *Science* **2001**, 291, 1304-1351.
5. Chiti, F.; Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **2006**, 75, 333-366.
6. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, 29, 7133-7155.
7. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, 14, 1-63.
8. Tanford, C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* **1962**, 84, 4240-4247.
9. Pace, C. N.; Scholtz, J. M.; Grimsley, G. R. Forces stabilizing proteins. *FEBS Lett.* **2014**, 588, 2177-2184.
10. Ponder, J. W.; Richards, F. M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **1987**, 193, 775-791.
11. Woolfson, D. N. The design of coiled-coil structures and assemblies. *Adv. Protein Chem.* **2005**, 70, 79-112.
12. Crick, F. H. C. The packing of α -helices: Simple coiled-coils. *Acta Crystallogr.* **1953**, 6, 689-697.
13. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, 437, 640-647.
14. Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, 157, 105-132.
15. London, F. The general theory of molecular forces. *Trans. Faraday Soc.* **1937**, 33, 8-26.

16. Lennard-Jones, J. E. Cohesion. *Proc. Phys. Soc.* **1931**, *43*, 461-482.
17. Wagner, J. P.; Schreiner, P. R. London dispersion in molecular chemistry—reconsidering steric effects. *Angew. Chem. Int. Ed.* **2015**, *54*, 12274-12296.
18. Horowitz, A.; Serrano, L.; Avron, B.; Bycroft, M.; Fersht, A. R. Strength and cooperativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **1990**, *216*, 1031-1044.
19. Serrano, L.; Horowitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R. Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **1990**, *29*, 9343-9352.
20. Paulini, R.; Muller, K.; Diederich, F. Orthogonal multipolar interactions in structural chemistry and biology. *Angew. Chem. Int. Ed.* **2005**, *44*, 1788-1805.
21. Yang, A.-S.; Honig, B. Free energy determinants of secondary structure formation: I. α -helices. *J. Mol. Biol.* **1995**, *252*, 351-365.
22. Yang, A.-S.; Honig, B. Free energy determinants of secondary structure formation: II. Antiparallel β -sheets. *J. Mol. Biol.* **1995**, *252*, 366-376.
23. Yang, A.-S.; Hitz, B.; Honig, B. Free energy determinants of secondary structures: III. β -Turns and their role in protein folding. *J. Mol. Biol.* **1996**, *259*, 873-882.
24. Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. The α -helix dipole and the properties of proteins. *Nature* **1978**, *273*, 443-446.
25. Baker, E. G.; Bartlett, G. J.; Crump, M. P.; Sessions, R. B.; Linden, N.; Faul, C. F.; Woolfson, D. N. Local and macroscopic electrostatic interactions in single α -helices. *Nat. Chem. Biol.* **2015**, *11*, 221-228.
26. Perutz, M. F. Electrostatic effects in proteins. *Science* **1978**, *201*, 1187-1191.
27. Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301-332.
28. Ensign, D. L.; Webb, L. J. Factors determining electrostatic fields in molecular dynamics simulations of the Ras/effector interface. *Proteins* **2011**, *79*, 3511-3524.
29. Ragain, C. M.; Newberry, R. W.; Ritchie, A. W.; Webb, L. J. Role of electrostatics in differential binding of RalGDS to Rap mutations E30D and K31E investigated by vibrational spectroscopy of thiocyanate probes. *J. Phys. Chem. B* **2012**, *116*, 9326-9336.
30. Schutz, C. N.; Warshel, A. What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins* **2001**, *44*, 400-417.

31. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio Jr., R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549-2564.
32. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. The polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comp.* **2013**, *9*, 4046-4063.
33. Klotz, I. M.; Franzen, J. S. Hydrogen bonds between model peptide groups in solution. *J. Am. Chem. Soc.* **1962**, *84*, 3461-3466.
34. Baldwin, R. L. In search of the energetic role of peptide hydrogen bonds. *J. Biol. Chem.* **2003**, *278*, 17581-17588.
35. Marqusee, S.; Robbins, V. H.; Baldwin, R. L. Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5286-5290.
36. Rohl, C. A.; Chakrabarty, A.; Baldwin, R. L. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* **1996**, *5*, 2623-2637.
37. Fleming, P. J.; Rose, G. D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* **2005**, *14*, 1911-1917.
38. Hong, H. Toward understanding driving forces in membrane protein folding. *Arch. Biochem. Biophys.* **2014**, *564*, 297-313.
39. Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899-926.
40. Weinhold, F.; Klein, R. A. What is a hydrogen bond? Resonance covalency in the supramolecular domain. *Chem. Educ. Res. Pract.* **2014**, *15*, 276-285.
41. Reed, A. E.; Weinhold, F. Natural bond orbital analysis of near-Hartree-Fock water dimer. *J. Chem. Phys.* **1983**, *78*, 4066-4073.
42. Cordier, F.; Grzesiek, S. Direct observation of hydrogen bonds in proteins by interresidue $^3\text{J}_{\text{NC}'}$ scalar couplings. *J. Am. Chem. Soc.* **1999**, *121*, 1601-1602.
43. Isaacs, E. D.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A. Covalency of the hydrogen bond in ice: A direct X-ray measurement. *Phys. Rev. Lett.* **1999**, *82*, 600-603.
44. Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.* **2002**, *41*, 48-76.

45. O'Meara, M. J.; Leaver-Fay, A.; Tyka, M. D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; Kuhlman, B. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J Chem Theory Comput* **2015**, *11*, 609-622.
46. 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. USA* **1951**, *37*, 205-211.
47. Pauling, L.; Corey, R. B. Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 729-740.
48. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.
49. Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97-179.
50. Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. Hydrogen bonding in globular proteins. *J. Mol. Biol.* **1992**, *226*, 1143-1159.
51. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **1982**, *4*, 187-217.
52. Mayo, S. L.; Olafson, B. D.; Goddard III, W. A. DREIDING: A generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897-8909.
53. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Kenneth M. Merz, J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
54. Das, R.; Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363-382.
55. Dahiyat, B. I.; Mayo, S. L. De novo protein design: Fully automated sequence selection. *Science* **1997**, *278*, 82-87.
56. Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003**, *302*, 1364-1368.
57. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas III, C. F.; Hilvert, D.; Houk, K. N.;

- Stoddard, B. L.; Baker, D. De novo computational design of retro-aldol enzymes. *Science* **2008**, *319*, 1387-91.
58. Moult, J. A decade of CASP: Progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **2005**, *15*, 285-289.
 59. Kryshchuk, A.; Fidelis, K.; Moult, J. CASP9 results compared to those of previous CASP experiments. *Proteins* **2011**, *79*, 196-207.
 60. Fleishman, S. J., et al. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J. Mol. Biol.* **2011**, *414*, 289-302.
 61. Stranges, P. B.; Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **2013**, *22*, 74-82.
 62. Moult, J.; Fidelis, K.; Kryshchuk, A.; Schwede, T.; Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)—round X. *Proteins* **2014**, *82*, 1-6.
 63. Koehl, P.; Levitt, M. A brighter future for protein structure prediction. *Nat. Struct. Biol.* **1999**, *6*, 108-111.
 64. Bardwell, D. A., et al. Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. *Acta Crystallogr.* **2011**, *B67*, 535-551.
 65. Lommerse, J. P. M.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Mooji, W. T. M.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; Eijck, B. P. v.; Verwer, P.; Williams, D. E. A test of crystal structure prediction of small organic molecules. *Acta Crystallogr.* **2000**, *B56*, 697-714.
 66. Derewenda, Z. S.; Derewenda, U.; Kobos, P. M. (His)C^ε–H···O=C< hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol.* **1994**, *241*, 83-93.
 67. Ramachandran, G. N.; Sasisekharan, V. Refinement of the structure of collagen. *Biochim. Biophys. Acta* **1965**, *109*, 314-316.
 68. Ramachandran, G. N.; Venkatachalam, C. M. The stability of the two-bonded collagen triple helix. *Biochim. Biophys. Acta* **1966**, *120*, 457-458.
 69. Taylor, R.; Kennard, O. Crystallographic evidence for the existence of C–H···O, C–H···N, and C–H···Cl hydrogen bonds. *J. Am. Chem. Soc.* **1982**, *104*, 5063-5070.
 70. Gu, Y.; Kar, T.; Scheiner, S. Fundamental properties of the CH···O interaction: Is it a true hydrogen bond? *J. Am. Chem. Soc.* **1999**, *121*, 9411-6422.

71. Steiner, T. Unrolling the hydrogen bond properties of C–H \cdots O interactions. *Chem. Commun.* **1997**, 8, 727-723.
72. Qian, W.; Krimm, S. Vibrational spectroscopy of hydrogen bonding: Origin of the different behaviour of the C–H \cdots O hydrogen bond. *J. Phys. Chem. A* **2002**, 106, 6628-6636.
73. Hobza, P.; Havlas, Z. Blue-shifting hydrogen bonds. *Chem. Rev.* **2000**, 100, 4253-4264.
74. Yohannan, S.; Faham, S.; Yang, D.; Grosfeld, D.; Chamberlain, A. K.; Bowie, J. U. A C α –H \cdots O hydrogen bond in a membrane protein is not stabilizing. *J. Am. Chem. Soc.* **2004**, 126, 2284-2285.
75. Arbely, E.; Arkin, I. T. Experimental measurement of the strength of a C α –H \cdots O bond in a lipid bilayer. *J. Am. Chem. Soc.* **2004**, 126, 5362-5363.
76. Scheiner, S.; Kar, T.; Gu, Y. Strength of the C $^{\alpha}$ H \cdots O hydrogen bond of amino acid residues. *J. Biol. Chem.* **2001**, 276, 9832-9837.
77. Scheiner, S. Relative strengths of NH \cdots O and CH \cdots O hydrogen bonds between polypeptide chain segments. *J. Phys. Chem. B* **2005**, 109, 16132-16141.
78. Vargas, R.; Garza, J.; Dixon, D. A.; Hay, B. P. How strong is the C $^{\alpha}$ –H \cdots O=C hydrogen bond? *J. Am. Chem. Soc.* **2000**, 122, 4750-4755.
79. McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, 238, 777-793.
80. Bartlett, G. J.; Woolfson, D. N. On the satisfaction of backbone-carbonyl lone pairs of electrons in protein structures. *Protein Sci.* **2016**, DOI: 10.1002/pro.2896.
81. Derewenda, Z. S.; Lee, L.; Derewenda, U. The occurrence of C–H \cdots O hydrogen bonds in proteins. *J. Mol. Biol.* **1995**, 252, 248-262.
82. Scheiner, S. Contributions of NH \cdots O and CH \cdots O hydrogen bonds to the stability of β -sheets in proteins. *J. Phys. Chem. B* **2006**, 110, 18670-18679.
83. Manikandan, K.; Ramakumar, S. The occurrence of C–H \cdots O hydrogen bonds in α -helices and helix termini in globular proteins. *Proteins* **2004**, 56, 768-781.
84. Chakrabarti, P.; Chakrabarti, S. C–H \cdots O hydrogen bond involving proline residues in α -helices. *J. Mol. Biol.* **1998**, 284, 867-873.
85. Bella, J.; Berman, H. M. Crystallographic evidence for C $^{\alpha}$ –H \cdots O=C hydrogen bonds in a collagen triple helix. *J. Mol. Biol.* **1996**, 264, 734-742.

86. Jiang, L.; Lai, L. CH \cdots O hydrogen bonds at protein-protein interfaces. *J. Biol. Chem.* **2002**, *277*, 37732-37740.
87. Senes, A.; Gerstein, M.; Engelman, D. M. Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with β -branched residues at neighboring positions. *J. Mol. Biol.* **2000**, *296*, 921-936.
88. Senes, A.; Ubarretxena-Belandia, I.; Engelman, D. M. The Ca-H \cdots O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9056-9061.
89. Mueller, B. K.; Subramaniam, S.; Senes, A. A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Ca-H hydrogen bonds. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E888-E895.
90. Hinderaker, M. P.; Raines, R. T. An electronic effect on protein structure. *Protein Sci.* **2003**, *12*, 1188-1194.
91. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **2001**, *123*, 777-778.
92. Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. $n\rightarrow\pi^*$ Interactions in proteins. *Nat. Chem. Biol.* **2010**, *6*, 615-620.
93. Bürgi, H. B.; Dunitz, J. D.; Shefter, E. Geometric reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Am. Chem. Soc.* **1973**, *95*, 5065-5067.
94. Bürgi, H. B.; Dunitz, J. D.; Shefter, E. Chemical reaction paths. IV. Aspects of O \cdots C=O interactions in crystals. *Acta Crystallogr.* **1974**, *B30*, 1517-1527.
95. Fischer, F. R.; Wood, P. A.; Allen, F. H.; Diederich, F. Orthogonal dipolar interactions between amide carbonyl groups. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17290-17294.
96. Fah, C.; Hardegger, L. A.; Ebert, M. O.; Schweizer, W. B.; Diederich, F. Self-association based on orthogonal C=O \cdots C=O interactions in the solid and liquid state. *Chem. Commun.* **2010**, *46*, 67-69.
97. Worley, B.; Richard, G.; Harbinson, G. S.; Powers, R. ^{13}C NMR reveals no evidence of $n\rightarrow\pi^*$ interactions in proteins. *PLoS One* **2012**, *7*, e42075.
98. Choudhary, A.; Gandla, D.; Krow, G. R.; Raines, R. T. Nature of amide carbonyl-carbonyl interactions in proteins. *J. Am. Chem. Soc.* **2009**, *131*, 7244-7246.
99. Kamer, K. J.; Choudhary, A.; Raines, R. T. Intimate interactions with carbonyl groups: Dipole-dipole or $n\rightarrow\pi^*$? *J. Org. Chem.* **2013**, *78*, 2099-2103.

100. Newberry, R. W.; VanVeller, B.; Guzei, I. A.; Raines, R. T. $n \rightarrow \pi^*$ Interactions of amides and thioamides: Implications for protein stability. *J. Am. Chem. Soc.* **2013**, *135*, 7843-7846.
101. Persch, E.; Dumele, O.; Diederich, F. Molecular recognition in chemical and biological systems. *Angew. Chem. Int. Ed.* **2015**, *54*, 3290-3327.
102. Choudhary, A.; Raines, R. T. Signature of $n \rightarrow \pi^*$ interactions in α -helices. *Protein Sci.* **2011**, *20*, 1077-1081.
103. Newberry, R. W.; Bartlett, G. J.; Vanveller, B.; Woolfson, D. N.; Raines, R. T. Signatures of $n \rightarrow \pi^*$ interactions in proteins. *Protein Sci.* **2014**, *23*, 284-288.
104. Horng, J. C.; Raines, R. T. Stereoelectronic effects on polyproline conformation. *Protein Sci.* **2006**, *15*, 74-83.
105. Pal, T. K.; Sankararamakrishnan, R. Quantum chemical investigations on intraresidue carbonyl-carbonyl contacts in aspartates of high-resolution protein structures. *J. Phys. Chem. B* **2010**, *114*, 1038-1049.
106. Bartlett, G. J.; Newberry, R. W.; Vanveller, B.; Raines, R. T.; Woolfson, D. N. Interplay of hydrogen bonds and $n \rightarrow \pi^*$ interactions in proteins. *J. Am. Chem. Soc.* **2013**, *135*, 18682-18688.
107. Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210-1250.
108. Dougherty, D. A. Cation- π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **1996**, *271*, 163-168.
109. Dougherty, D. A. The cation- π interaction. *Acc. Chem. Res.* **2013**, *46*, 885-893.
110. Ma, J. C.; Dougherty, D. A. The cation- π interaction. *Chem. Rev.* **1997**, *97*, 1303-1324.
111. Mecozzi, S.; West Jr., A. P.; Dougherty, D. A. Cation- π interactions in simple aromatics: Electrostatics provide a predictive tool. *J. Am. Chem. Soc.* **1996**, *118*, 2307-2308.
112. Mecozzi, S.; West Jr., A. P.; Dougherty, D. A. Cation- π interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10566-10571.
113. Kumpf, R. A.; Dougherty, D. A. A mechanism for ion selectivity in potassium channels: Computational studies of cation- π interactions. *Science* **1993**, *261*, 1708-1710.

114. Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025-1040.
115. Loewenthal, R.; Sancho, J.; Fersht, A. R. Histidine–aromatic interactions in barnase: Elevation of histidine pK_a and contribution to protein stability. *J. Mol. Biol.* **1992**, *224*, 759-770.
116. Gallivan, J. P.; Dougherty, D. A. Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. USA* **1999**, *98*, 9459-9464.
117. Tsou, L. K.; Tatko, C. D.; Waters, M. L. Simple cation- π interaction between a phenyl ring and a protonated amine stabilizes an alpha-helix in water. *J. Am. Chem. Soc.* **2002**, *124*, 14917-14921.
118. Waters, M. L. Aromatic interactions in peptides: Impact on structure and function. *Biopolymers* **2004**, *76*, 435-445.
119. Crowley, P. B.; Golovin, A. Cation- π interactions in protein-protein interfaces. *Proteins* **2005**, *59*, 231-239.
120. Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic rings in chemical and biological recognition: Energetics and structures. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808-4842.
121. Levitt, M.; Perutz, M. F. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **1988**, *201*, 751-754.
122. Steiner, T.; Koellner, G. Hydrogen bonds with π -acceptors in proteins: Frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **2001**, *305*, 535-557.
123. Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. C-H $\cdots\pi$ interactions in proteins. *J. Mol. Biol.* **2001**, *307*, 357-377.
124. Plevin, M. J.; Bryce, D. L.; Boisbouvier, J. Direct detection of CH/ π interactions in proteins. *Nat. Chem.* **2010**, *2*, 466-471.
125. Chen, W.; Enck, S.; Price, J. L.; Powers, D. L.; Powers, E. T.; Wong, C. H.; Dyson, H. J.; Kelly, J. W. Structural and energetic basis of carbohydrate–aromatic packing interactions in proteins. *J. Am. Chem. Soc.* **2013**, *135*, 9877-9884.
126. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. Carbohydrate–aromatic interactions in proteins. *J. Am. Chem. Soc.* **2015**, *137*, 15152-15160.
127. Vyas, N. K. Atomistic features of protein–carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1991**, *1*, 732-740.

128. Burley, S. K.; Petsko, G. A. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. *Science* **1985**, *229*, 23-28.
129. Hunter, C. A.; Sanders, J. K. M. The nature of π - π interactions. *J. Am. Chem. Soc.* **1990**, *112*, 5525-5534.
130. Hunter, C. A.; Singh, J.; Thornton, J. M. π - π Interactions: The geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **1991**, *218*, 837-846.
131. McGaughey, G. B.; Gagné, M.; Rappé, A. K. π -Stacking interactions: Alive and well in proteins. *J. Biol. Chem.* **1998**, *273*, 15458-15463.
132. Kannan, N.; Wishveshwara, S. Aromatic clusters: A determinant of thermal stability of thermophilic proteins. *Protein Eng.* **2000**, *13*, 753-761.
133. Tatko, C. D.; Waters, M. L. Selective aromatic interactions in β -hairpin peptides. *J. Am. Chem. Soc.* **2002**, *124*, 9372-9373.
134. Butterfield, S. M.; Patel, P. R.; Waters, M. L. Contribution of aromatic interactions to α -helix stability. *J. Am. Chem. Soc.* **2002**, *124*, 9751-9755.
135. Serrano, L.; Bycroft, M.; Fersht, A. R. Aromatic-aromatic interactions and protein stability: Investigation by double-mutant cycles. *J. Mol. Biol.* **1991**, *218*, 465-475.
136. Burley, S. K.; Petsko, G. A. Dimerization energetics of benzene and aromatic amino acid side chains. *J. Am. Chem. Soc.* **1986**, *108*, 7995-8001.
137. Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *J. Am. Chem. Soc.* **2001**, *124*, 6133-6143.
138. Frontera, A.; Gamez, P.; Mascal, M.; Mooibroek, T. J.; Reedijk, J. Putting anion- π interactions into perspective. *Angew. Chem. Int. Ed.* **2011**, *50*, 9564-9583.
139. Jackson, M. R.; Beahm, R.; Duvvuru, S.; Narasimhan, C.; Wu, J.; Wang, H.-N.; Philip, V. M.; Hinde, R. J.; Howell, E. E. A preference for edgewise interactions between aromatic rings and carboxylate anions: The biological relevance of anion-quadrupole interactions. *J. Phys. Chem. B* **2007**, *111*, 8242-8249.
140. Shi, Z.; Olson, C. A.; Bell Jr., A. J.; Kallenbach, N. R. Non-classical helix-stabilizing interactions: C-H \cdots O H-bonding between Phe and Glu side chains in α -helical peptides. *Biophys. Chem.* **2002**, *101*, 267-279.
141. Philip, V.; Harris, J.; Adams, R.; Nguyen, D.; Spiers, J.; Baudry, J.; Howell, E. E.; Hinde, R. J. A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the Protein Data Bank: Searching for anion- π pairs. *Biochemistry* **2011**, *50*, 2939-2950.

142. Egli, M.; Sarkhel, S. Lone pair–aromatic interactions: To stabilize or not to stabilize. *Acc. Chem. Res.* **2007**, *40*, 197-205.
143. Reid, K. S. C.; Lindley, P. F.; Thornton, J. M. Sulphur-aromatic interactions in proteins. *FEBS J.* **1985**, *190*, 209-213.
144. Viguera, A. R.; Serrano, L. Side-chain interactions between sulfur-containing amino acids and phenylalanine in α -helices. *Biochemistry* **1995**, *34*, 8771-8779.
145. Tatko, C. D.; Waters, M. L. Investigation of the nature of the methionine– π interaction in β -hairpin peptide model systems. *Protein Sci.* **2004**, *13*, 2515-2522.
146. Alber, F.; Kuonen, O.; Scapozza, L.; Folkers, G.; Carloni, P. Density functional studies on herpes simplex virus type 1 thymidine kinase-substrate interactions: The role of Tyr-172 and Met-128 in thymine fixation. *Proteins* **1998**, *31*, 453-459.
147. Spencer, D. S.; Stiles, W. E. The M32L substitution of staphylococcal nuclease: Disagreement between theoretical prediction and experimental protein stability. *J. Mol. Biol.* **1996**, *257*, 497-499.
148. Zauhar, R. J.; Colbert, C. L.; Morgan, R. S.; Welsh, W. J. Evidence for a strong sulfur–aromatic interactions derived from crystallographic data. *Biopolymers* **2000**, *53*, 233-248.
149. Minton, A. P.; Wilf, J. Effect of macromolecular crowding upon the structure and function of an enzyme: Glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* **1981**, *20*, 4821-4826.
150. Ellis, R. J. Macromolecular crowding: Obvious but underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597-604.
151. Homouz, D.; Perham, M.; Samiotakis, A.; Cheung, M. S.; Wittung-Stafshede, P. Crowded, cell-like environment induces shape changes in aspherical protein. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11754-11759.
152. Zhou, H. X.; Rivas, G.; Minton, A. P. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375-397.
153. Fedyukina, D. V.; Cavagnero, S. Protein folding at the exit tunnel. *Annu. Rev. Biophys.* **2011**, *40*, 337-359.
154. Pechmann, S.; Willmund, F.; Frydman, J. The ribosome as a hub for protein quality control. *Mol. Cell* **2013**, *49*, 411-421.
155. Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **1996**, *381*, 571-580.

156. Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **2011**, *475*, 324-332.
157. White, S. H.; Wimley, W. C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319-365.
158. Dowhan, W.; Bogdanov, M. Lipid-dependent membrane protein topogenesis. *Annu. Rev. Biochem.* **2009**, *78*, 515-540.
159. Parodi, A. J. Protein glucosylation and its role in protein folding. *Annu. Rev. Biochem.* **2000**, *69*, 69-93.
160. Ron, D.; Walter, P. Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell. Biol.* **2007**, *8*, 519-529.
161. Walter, P.; Ron, D. The unfolded protein response: From stress pathway to homeostatic regulation. *Science* **2011**, *334*, 1081-1086.
162. Glickman, M. H.; Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiol. Rev.* **2002**, *82*, 373-428.
163. Powers, E. T.; Morimoto, R. I.; Dillin, A.; Kelly, J. W.; Balch, W. E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **2009**, *78*, 959-991.
164. Powers, E. T.; Balch, W. E. Diversity in the origins of proteostasis networks—a driver for protein function in evolution. *Nat. Rev. Mol. Cell. Biol.* **2013**, *14*, 237-248.
165. Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16789-16794.
166. Metrangolo, P.; Neukirch, H.; Pilati, T.; Resnati, G. Halogen bonding based recognition processes: A world parallel to hydrogen bonding. *Acc. Chem. Res.* **2005**, *38*, 386-395.
167. Voth, A. R.; Khuu, P.; Oishi, K.; Ho, P. S. Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nat. Chem.* **2009**, *1*, 74-79.
168. Dahiyat, B. I.; Mayo, S. L. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10172-10177.
169. Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **1953**, *171*, 737-738.
170. Allen, F. H.; Baalham, C. A.; Lommerse, J. P. M.; Raithby, P. R. Carbonyl-carbonyl interactions can be competitive with hydrogen bonds. *Acta Crystallogr.* **1998**, *B54*, 320-329.

171. Choudhary, A.; Raines, R. T. An evaluation of peptide-bond isosteres. *ChemBioChem* **2011**, *12*, 1801-1807.
172. Sonntag, L.-S.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H. The “azido gauche effect”—implications for the conformation of azidoproline. *J. Am. Chem. Soc.* **2006**, *128*, 14697-14703.
173. Kuemin, M.; Nagel, Y. A.; Schweizer, S.; Monnard, F. W.; Ochsenfeld, C.; Wennemers, H. Tuning the cis/trans conformer ratio of Xaa-Pro amide bonds by intramolecular hydrogen bonds: The effect on PPII helix stability. *Angew. Chem. Int. Ed.* **2010**, *49*, 6324-6327.
174. Shoulders, M. D.; Kotch, F. W.; Choudhary, A.; Guzei, I. A.; Raines, R. T. The aberrance of the 4*S* diastereomer of 4-hydroxyproline. *J. Am. Chem. Soc.* **2010**, *132*, 10857-10865.
175. Choudhary, A.; Kamer, K. J.; Powner, M. W.; Sutherland, J. D.; Raines, R. T. A stereoelectronic effect in prebiotic nucleotide synthesis. *ACS Chem. Biol.* **2010**, *5*, 655-657.
176. Choudhary, A.; Pua, K. H.; Raines, R. T. Quantum mechanical origin of the conformational preferences of 4-thiaproline and its *S*-oxides. *Amino Acids* **2011**, *41*, 181-186.
177. Choudhary, A.; Kamer, K. J.; Raines, R. T. An $n \rightarrow \pi^*$ interaction in aspirin: Implications for structure and reactivity. *J. Org. Chem.* **2011**, *76*, 7933-7937.
178. Newberry, R. W.; Raines, R. T. A key $n \rightarrow \pi^*$ interaction in *N*-acyl homoserine lactones. *ACS Chem. Biol.* **2014**, *9*, 880-883.
179. Newberry, R. W.; Raines, R. T. $n \rightarrow \pi^*$ Interactions in poly(lactic acid) suggest a role in protein folding. *Chem. Commun.* **2013**, *49*, 7699-7701.
180. Jakobsche, C. E.; Choudhary, A.; Miller, S. J.; Raines, R. T. $n \rightarrow \pi^*$ Interaction and $n(\pi)$ Pauli repulsion are antagonistic for protein stability. *J. Am. Chem. Soc.* **2010**, *132*, 6651-6653.
181. Hodges, J. A.; Raines, R. T. Energetics of an $n \rightarrow \pi^*$ interaction that impacts protein structure. *Org. Lett.* **2006**, *8*, 4695-4697.
182. Adhikari, U.; Scheiner, S. Preferred configurations of peptide-peptide interactions. *J. Phys. Chem. A* **2013**, *117*, 489-496.
183. Shoulders, M. D.; Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.

184. Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. Code for collagen's stability deciphered. *Nature* **1998**, *392*, 666-667.
185. DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. Collagen stability: Insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. *J. Am. Chem. Soc.* **2002**, *124*, 2497-2505.
186. Lesarri, A.; Cocinero, E. J.; Lopez, J. C.; Alonso, J. L. Shape of 4(*S*)- and 4(*R*)-hydroxyproline in gas phase. *J. Am. Chem. Soc.* **2005**, *127*, 2572-2579.
187. Jenkins, C. L.; Lin, G.; Duo, J.; Rapolu, D.; Guzei, I. A.; Raines, R. T.; Krow, G. R. Substituted 2-azabicyclo[2.1.1]hexanes as constrained proline analogues: Implications for collagen stability. *J. Org. Chem.* **2004**, *69*, 8565-8573.
188. Krow, G. R.; Edupuganti, R.; Gandla, D.; Yu, F.; Sender, M.; Sonnet, P. E.; Zdilla, M. J.; DeBrosse, C.; Cannon, K. C.; Ross, C. W.; Choudhary, A.; Shoulders, M. D.; Raines, R. T. Synthesis of conformationally constrained 5-fluoro- and 5-hydroxymethano pyrrolidines. Ring-puckered mimics of gauche- and anti-3-fluoro- and 3-hydroxy pyrrolidines. *J. Org. Chem.* **2011**, *76*, 3626-3634.
189. Jenkins, C. L.; Vasbinder, M. M.; Miller, S. J.; Raines, R. T. Peptide bond isosteres: Ester or (*E*)-alkene in the backbone of the collagen triple helix. *Org. Lett.* **2005**, *7*, 2619-2622.
190. Shoulders, M. D.; Hodges, J. A.; Raines, R. T. Reciprocity of steric and stereoelectronic effects in the collagen triple helix. *J. Am. Chem. Soc.* **2006**, *128*, 8112-8113.
191. Kotch, F. W.; Guzei, I. A.; Raines, R. T. Stabilization of the collagen triple helix by *O*-methylation of hydroxyproline residues. *J. Am. Chem. Soc.* **2008**, *130*, 2952-2953.
192. Dai, N.; Wang, X. J.; Etzkorn, F. A. The effect of a trans-locked Gly-Pro alkene isostere on collagen triple helix stability. *J. Am. Chem. Soc.* **2008**, *130*, 5396-5397.
193. Dai, N.; Etzkorn, F. A. Cis-trans proline isomerization effects on collagen triple-helix stability are limited. *J. Am. Chem. Soc.* **2009**, *131*, 13728-13732.
194. Shoulders, M. D.; Satyshur, K. A.; Forest, K. T.; Raines, R. T. Stereoelectronic and steric effects in side chains preorganize a protein main chain. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 559-564.
195. Kumin, M.; Sonntag, L.-S.; Wennemers, H. Azidoproline containing helices: Stabilization of the polyproline II structure by a functionalizable group. *J. Am. Chem. Soc.* **2007**, *129*, 466-467.
196. Chiang, Y. C.; Lin, Y. J.; Horng, J. C. Stereoelectronic effects on the transition barrier of polyproline conformational interconversion. *Protein Sci.* **2009**, *18*, 1967-1977.

197. Bielska, A. A.; Zondlo, N. J. Hyperphosphorylation of tau induces local polyproline II helix. *Biochemistry* **2006**, *45*, 5527-5537.
198. Brown, A. M.; Zondlo, N. J. A propensity scale for type II polyproline helices (PPII): Aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline–aromatic interactions. *Biochemistry* **2012**, *51*, 5041-5051.
199. Elbaum, M. B.; Zondlo, N. J. OGlcNAcylation and phosphorylation have similar structural effects in α -helices: Post-translational modifications as inducible start and stop signals in α -helices, with greater structural effects on threonine modification. *Biochemistry* **2014**, *53*, 2242-2260.
200. Wilhelm, P.; Lewandowski, B.; Trapp, N.; Wennemers, H. A crystal structure of an oligoproline PPII-helix, at last. *J. Am. Chem. Soc.* **2014**, *136*, 15829-15832.
201. Fufezan, C. The role of Buergi-Dunitz interactions in the structural stability of proteins. *Proteins* **2010**, *78*, 2831-2838.
202. Creighton, T. E., *Proteins: Structures and molecular properties*. 2nd ed.; Freeman: New York, 1993.
203. Gao, J.; Kelly, J. W. Toward quantification of protein backbone–backbone hydrogen bonding energies: An energetic analysis of an amide-to-ester mutation in an α -helix within a protein. *Protein Sci.* **2008**, *17*, 1096-1011.
204. Reddy, D. N.; George, G.; Prabhakaran, E. N. Crystal-structure analysis of *cis*-X-pro-containing peptidomimetics: Understanding the steric interactions at *cis* X-pro amide bonds. *Angew Chem Int Ed Engl* **2013**.
205. Choudhary, A.; Kamer, K. J.; Raines, R. T. A conserved interaction with the chromophore of fluorescent proteins. *Protein Sci.* **2012**, *21*, 171-177.
206. Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. Conformation of the backbone in unfolded proteins. *Chem. Rev.* **2006**, *106*, 1877-1897.
207. Butterfoss, G. L.; Renfrew, P. D.; Kuhlman, B.; Kirshenbaum, K.; Bonneau, R. A preliminary survey of the peptoid folding landscape. *J. Am. Chem. Soc.* **2009**, *131*, 16798-167807.
208. Gorske, B. C.; Bastian, B. L.; Geske, G. D.; Blackwell, H. E. Local and tunable $n \rightarrow \pi^*$ interactions regulate amide isomerism in the peptoid backbone. *J. Am. Chem. Soc.* **2007**, *129*, 8928-8929.
209. Gorske, B. C.; Stringer, J. R.; Bastian, B. L.; Fowler, S. A.; Blackwell, H. E. New strategies for the design of folded peptoids revealed by a survey of noncovalent interactions in model systems. *J. Am. Chem. Soc.* **2009**, *131*, 16555-16567.

210. Caumes, C.; Roy, O.; Faure, S.; TAILLEFUMIER, C. The click triazolium peptoid side chain: A strong *cis*-amide inducer enabling chemical diversity. *J. Am. Chem. Soc.* **2012**, *134*, 9553-6.
211. Laursen, J. S.; Engel-Andreasen, J.; Fristrup, P.; Harris, P.; Olsen, C. A. *Cis-trans* amide bond rotamers in β -peptoids and peptoids: Evaluation of stereoelectronic effects in backbone and side chains. *J. Am. Chem. Soc.* **2013**, *135*, 2835-2844.
212. Gorske, B. C.; Nelson, R. C.; Bowden, Z. S.; Kufe, T. A.; Childs, A. M. "Bridged" $n \rightarrow \pi^*$ interactions can stabilize peptoid helices. *J. Org. Chem.* **2013**, *78*, 11172-11183.
213. Adhikari, U.; Scheiner, S. Contributions of various noncovalent bonds to the interaction between an amide and S-containing molecules. *ChemPhysChem* **2012**, *13*, 3535-3541.
214. Reddy, D. N.; Thirupathi, R.; Tumminakatti, S.; Prabhakaran, E. N. A method for stabilizing the *cis* prolyl peptide bond: Influence of an unusual $n \rightarrow \pi^*$ interaction in 1,3-oxazine and 1,3-thiazine containing peptidomimetics. *Tetrahedron Lett.* **2012**, *53*, 4413-4417.
215. Singh, S. K.; Kumar, S.; Das, A. Competition between $n \rightarrow \pi_{Ar}^*$ and conventional hydrogen bonding (N-H \cdots N) interactions: An *ab initio* study of the complexes of 7-azaindole and fluorosubstituted pyridines. *Phys. Chem. Chem. Phys.* **2013**, *16*, 8819-8827.
216. Azofra, L. M.; Scheiner, S. Complexes containing CO₂ and SO₂. Mixed dimers, trimers and tetramers. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5142-5149.
217. Azofra, L. M.; Scheiner, S. Complexation of *n* SO₂ molecules (*n* = 1, 2, 3) with formaldehyde and thioformaldehyde. *J. Chem. Phys.* **2014**, *140*, 034302.
218. Blanco, S.; Lopez, J. C.; Mata, S.; Alonso, J. L. Conformations of γ -aminobutyric acid (GABA): The role of the $n \rightarrow \pi^*$ interaction. *Angew. Chem. Int. Ed.* **2010**, *49*, 9187-9192.
219. Cabezas, C.; Alonso, J. L.; Lopez, J. C.; Mata, S. Unveiling the shape of aspirin in the gas phase. *Angew. Chem. Int. Ed.* **2012**, *51*, 1375-1378.
220. Wang, H.; Kohler, P.; Overman, L. E.; Houk, K. N. Origins of stereoselectivities of dihydroxylations of *cis*-bicyclo[3.3.0]octenes. *J. Am. Chem. Soc.* **2012**, *134*, 16054-8.
221. Liu, P.; Yang, X.; Birman, V. B.; Houk, K. N. Origin of enantioselectivity in benztetramisole-catalyzed dynamic kinetic resolution of azlactones. *Org. Lett.* **2012**, *14*, 3288-3291.
222. Lajkiewicz, N. J.; Roche, S. P.; Gerard, B.; Porco, J., J. A. Enantioselective photocycloaddition of 3-hydroxyflavones: Total syntheses and absolute configuration

- assignments of (+)-ponapensin and (+)-elliptifoline. *J. Am. Chem. Soc.* **2012**, *134*, 13108-13113.
223. Pollock, S. B.; Kent, S. B. H. An investigation into the origin of the dramatically reduced reactivity of peptide-prolyl-thioesters in native chemical ligation. *Chem. Commun.* **2011**, *47*, 2342-2344.
 224. MacArthur, M. W.; Thornton, J. M. Influence of proline residues on protein conformation. *J. Mol. Biol.* **1991**, *218*, 397-412.
 225. Brandts, J. F.; Halvorson, H. R.; Brennan, M. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **1975**, *14*, 4953-4963.
 226. Mauger, A. B.; Witkop, B. Analogs and homologs of proline and hydroxyproline. *Chem. Rev.* **1966**, *66*, 47-86.
 227. Gottlieb, A. A.; Fujita, Y.; Udenfriend, S.; Witkop, B. Incorporation of *cis*- and *trans*-4-fluoro-L-prolines into proteins and hydroxylation of the *trans* isomer during collagen biosynthesis. *Biochemistry* **1965**, *4*, 2507-2513.
 228. Bakerman, S.; Martin, R. L.; Burgstahler, A. W.; Hayden, J. W. *In vivo* studies with fluoroprolines. *Nature* **1966**, *212*, 849-850.
 229. Takeuchi, T.; Prockop, D. J. Biosynthesis of abnormal collagens with amino acid analogues I. Incorporation of L-azetidine-2-carboxylic acid and *cis*-4-fluoro-L-proline into protocollagen and collagen. *Biochim. Biophys. Acta* **1969**, *175*, 142-155.
 230. Takeuchi, T.; Rosenbloom, J.; Prockop, D. J. Biosynthesis of abnormal collagens with amino acid analogues II. Inability of cartilage cells to extrude collagen polypeptides containing L-azetidine-2-carboxylic acid or *cis*-4-fluoro-L-proline. *Biochim. Biophys. Acta* **1969**, *175*, 156-164.
 231. Uitto, J.; Prockop, D. J. Incorporation of proline analogues into collagen polypeptides. Effects on the production of extracellular procollagen and on the stability of the triple-helical structure of the molecule. *Biochim. Biophys. Acta* **1974**, *336*, 234-251.
 232. Hutton, J., J. J.; Marglin, A.; Witkop, B.; Kurtz, J.; Berger, A.; Udenfriend, S. Synthetic polypeptides as substrates and inhibitors of collagen proline hydroxylase. *Arch. Biochem. Biophys.* **1968**, *125*, 779-785.
 233. Diegmann, R. F.; Ondrejickova, O.; Katz, E. Oxygen-18 and fluoroproline studies on the synthesis of hydroxyproline and oxoproline in actinomycin. *Arch. Biochem. Biophys.* **1969**, *131*, 276-287.

234. Yoder, N. C.; Kumar, K. Fluorinated amino acids in protein design and engineering. *Chem. Soc. Rev.* **2002**, *31*, 335-341.
235. Salwiczek, M.; Nyakatura, E. K.; Gerling, U. I.; Ye, S.; Koksche, B. Fluorinated amino acids: Compatibility with native protein structures and effects on protein-protein interactions. *Chem. Soc. Rev.* **2012**, *41*, 2135-2171.
236. Merkel, L.; Budisa, N. Organic fluorine as a polypeptide building element: *in vivo* expression of fluorinated peptides, proteins and proteomes. *Org. Biomol. Chem.* **2012**, *10*, 7241-7261.
237. Demange, L.; Menez, A.; Dugave, C. Practical synthesis of Boc and Fmoc protected 4-fluoro and 4-difluoroproline from *trans*-4-hydroxyproline. *Tetrahedron Lett.* **1998**, *39*, 1169-1172.
238. Hudlicky, M.; Merola, J. S. New stereospecific syntheses and X-ray diffraction structures of (-)-D-erythro- and (+)-L-threo-4-fluoroglutamic acid. *Tetrahedron Lett.* **1990**, *31*, 7403-7406.
239. Avent, A. G.; Bowler, A. N.; Doyle, P. M.; Marchand, C. M.; Young, D. W. Stereospecific synthesis of 4-fluoroglutamic acids. *Tetrahedron Lett.* **1992**, *33*, 1509-1512.
240. Hudlicky, M. Stereospecific syntheses of all four stereoisomers of 4-fluoroglutamic acid. *J. Fluorine Chem.* **1993**, *60*, 193-210.
241. Kronenthal, D. R.; Mueller, R. H.; Kuester, P. L.; Kissick, T. P.; Johnson, E. J. Stereospecific Friedel-Crafts alkylation of benzene with 4-mesyloxy-L-prolines. A new synthesis of 4-phenylprolines. *Tetrahedron Lett.* **1990**, *31*, 1241-1244.
242. Doi, M.; Nishi, Y.; Kiritoshi, N.; Iwata, T.; Nago, M.; Nakano, H.; Uchiyama, S.; Nakazawa, T.; Wakamiya, T.; Kobayashi, Y. Simple and efficient syntheses of Boc- and Fmoc-protected 4(*R*)- and 4(*S*)-fluoroproline solely from 4(*R*)-hydroxyproline. *Tetrahedron* **2002**, *58*, 8453-8459.
243. Chorghade, M. S.; Mohapatra, D. K.; Sahoo, G.; Gurjar, M. K.; Mandlecha, M. V.; Bhoite, N.; Moghe, S.; Raines, R. T. Practical syntheses of 4-fluoroproline. *J. Fluorine Chem.* **2008**, *129*, 781-784.
244. Thomas, K. M.; Naduthambi, D.; Tririyia, G.; Zondlo, N. J. Proline editing: A divergent strategy for the synthesis of conformationally diverse peptides. *Org. Lett.* **2005**, *7*, 2397-2400.
245. Pandey, A. K.; Naduthambi, D.; Thomas, K. M.; Zondlo, N. J. Proline editing: A general and practical approach to the synthesis of functionally and structurally diverse peptides.

- Analysis of steric versus stereoelectronic effects of 4-substituted prolines on conformation within peptides. *J. Am. Chem. Soc.* **2013**, *135*, 4333-4363.
246. Donohue, J.; Trueblood, K. N. The crystal structure of hydroxy-L-proline. II. Determination and description of the structure. *Acta Crystallogr.* **1952**, *5*, 419-431.
 247. Matheison, A. M.; Welsh, H. K. The crystal structure of copper proline dihydrate. *Acta Crystallogr.* **1952**, *5*, 599-604.
 248. Mitsui, Y.; Tsuboi, M.; Iitaka, Y. The crystal structure of DL-proline hydrochloride. *Acta Crystallogr.* **1969**, *B25*, 2182-2192.
 249. Higashijima, T.; Tasumi, M.; Miyazawa, T. ¹H Nuclear magnetic resonance studies of *N*-acetyl-L-proline *N*-methylamide. Molecular conformations, hydrogen bonding, and thermodynamic quantities in various solvents. *Biopolymers* **1977**, *16*, 1259-1270.
 250. Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallogr.* **2002**, *B58*, 380-388.
 251. Improta, R.; Benzi, C.; Barone, V. Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* **2001**, *123*, 12568-12577.
 252. Grieg, J. T.; McLeod, R. S. Conformations of *cis*- and *trans*-4-fluoro-L-proline in aqueous solution. *J. Am. Chem. Soc.* **1973**, *95*, 5725-5729.
 253. Panasik Jr., N.; Eberhardt, E. S.; Edison, A. S.; Powell, D. R.; Raines, R. T. Inductive effects on the structure of proline residues. *Int. J. Pept. Prot. Res.* **1994**, *44*, 262-269.
 254. Eberhardt, E. S.; Panasik Jr., N.; Raines, R. T. Inductive effects on the energetics of prolyl peptide bond isomerization: Implications for collagen folding and stability. *J. Am. Chem. Soc.* **1996**, *118*, 12261-12266.
 255. Rich, A.; Crick, F. H. C. The molecular structure of collagen. *J. Mol. Biol.* **1961**, *3*, 483-506.
 256. Ramshaw, J. A. M.; Shah, N. K.; Brodsky, B. Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J. Struct. Biol.* **1998**, *122*, 86-91.
 257. Sakakibara, S.; Inouye, K.; Shudo, K.; Kishida, T.; Kobayashi, Y.; Prockop, D. J. Synthesis of (Pro-Hyp-Gly)_n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **1973**, *303*, 198-202.

258. Berg, R. A.; Prockop, D. J. The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen. *Biochem. Biophys. Res. Commun.* **1973**, *52*, 115-120.
259. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **1994**, *266*, 75-81.
260. Bella, J.; Brodsky, B.; Berman, H. M. Hydration structure of a collagen peptide. *Structure* **1995**, *3*, 893-906.
261. Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. A hyperstable collagen mimic. *Chem. Biol.* **1999**, *6*, 63-70.
262. Mooney, S. D.; Kollman, P. A.; Klein, T. E. Conformational preferences of substituted prolines in the collagen triple helix. *Biopolymers* **2002**, *64*, 63-71.
263. Hodges, J. A.; Raines, R. T. Stereoelectronic and steric effects in the collagen triple helix: Toward a code for strand association. *J. Am. Chem. Soc.* **2005**, *127*, 15923-15932.
264. Hodges, J. A.; Raines, R. T. Stereoelectronic effects on collagen stability: The dichotomy of 4-fluoroproline diastereomers. *J. Am. Chem. Soc.* **2003**, *125*, 9262-9263.
265. Doi, M.; Nishi, Y.; Uchiyama, S.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Characterization of collagen model peptides containing 4-fluoroproline; (4(*S*)-fluoroproline-Pro-Gly)₁₀ forms a triple helix, but (4(*R*)-fluoroproline-Pro-Gly)₁₀ does not. *J. Am. Chem. Soc.* **2003**, *125*, 9922-9923.
266. Doi, M.; Nishi, Y.; Uchiyama, S.; Nishiuchi, Y.; Nishio, H.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Collagen-like triple helix formation of synthetic (Pro-Pro-Gly)₁₀ analogues: (4(*S*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-Gly)₁₀, (4(*R*)-hydroxyprolyl-4(*R*)-hydroxyprolyl-Gly)₁₀ and (4(*S*)-fluoroprolyl-4(*R*)-fluoroprolyl-Gly)₁₀. *J. Pept. Sci.* **2005**, *11*, 609-616.
267. Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. Triple-helix propensity of hydroxyproline and fluoroproline: Comparison of host-guest and repeating tripeptide collagen models. *J. Am. Chem. Soc.* **2003**, *125*, 11500-11501.
268. Malkar, N. B.; Lauer-Fields, J. L.; Borgia, J. A.; Fields, G. B. Modulation of triple-helical stability and subsequent melanoma cellular responses by single-site substitution of fluoroproline derivatives. *Biochemistry* **2002**, *41*, 6054-6064.
269. Chattopadhyay, S.; Murphy, C. J.; McAnulty, J. F.; Raines, R. T. Peptides that anneal to natural collagen *in vitro* and *ex vivo*. *Org. Biomol. Chem.* **2012**, *10*, 5892-5897.
270. Chattopadhyay, S.; Raines, R. T. Review collagen-based biomaterials for wound healing. *Biopolymers* **2014**, *101*, 821-833.

271. Nishi, Y.; Uchiyama, S.; Doi, M.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Different effects of 4-hydroxyproline and 4-fluoroproline on the stability of collagen triple helix. *Biochemistry* **2005**, *44*, 6034-6042.
272. Dunitz, J. D.; Taylor, R. Organic fluorine hardly ever accepts hydrogen bonds. *Chem. Eur. J.* **1997**, *3*, 89-98.
273. Shoulders, M. D.; Kamer, K. J.; Raines, R. T. Origin of the stability conferred upon collagen by fluorination. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3859-3862.
274. Marsh, E. N.; Suzuki, Y. Using ^{19}F NMR to probe biological interactions of proteins and peptides. *ACS Chem. Biol.* **2014**, *9*, 1242-1250.
275. Kawahara, K.; Nemoto, N.; Motooka, D.; Nishi, Y.; Doi, M.; Uchiyama, S.; Nakazawa, T.; Nishiuchi, Y.; Yoshida, T.; Ohkubo, T.; Kobayashi, Y. Polymorphism of collagen triple helix revealed by ^{19}F NMR of model peptide [Pro-4(R)-hydroxyprolyl-Gly] $_3$ -[Pro-4(R)-fluoroprolyl-Gly]-[Pro-4(R)-hydroxyprolyl -Gly] $_3$. *J. Phys. Chem. B* **2012**, *116*, 6908-6915.
276. Friedman, L.; Higgin, J. J.; Moulder, G.; Barstead, R.; Raines, R. T.; Kimble, J. Prolyl 4-hydroxylase is required for viability and morphogenesis in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4736-4741.
277. Holster, T.; Pakkanen, O.; Soininen, R.; Sormunen, R.; Nokelainen, M.; Kivirikko, K. I.; Myllyharju, J. Loss of assembly of the main basement membrane collagen, type IV, but not fibril-forming collagens and embryonic death in collagen prolyl 4-hydroxylase I null mice. *J. Biol. Chem.* **2007**, *282*, 2512-2519.
278. Gorres, K. L.; Edupuganti, R.; Krow, G. R.; Raines, R. T. Conformational preferences of substrates for human prolyl 4-hydroxylase. *Biochemistry* **2008**, *47*, 9447-9455.
279. Gorres, K. L.; Raines, R. T. Direct and continuous assay for prolyl 4-hydroxylase. *Anal. Biochem.* **2009**, *386*, 181-185.
280. Adzhubei, A. A.; Sternberg, M. J. E. Left-handed polyproline II helices commonly occur in globular proteins. *J. Mol. Biol.* **1993**, *229*, 472-493.
281. Kuemin, M.; Engel, J.; Wennemers, H. Temperature-induced transition between polyproline I and II helices: Quantitative fitting of hysteresis effects. *J. Pept. Sci.* **2010**, *16*, 596-600.
282. Lin, Y. J.; Horng, J. C. Impacts of terminal (4R)-fluoroproline and (4S)-fluoroproline residues on polyproline conformation. *Amino Acids* **2014**, *46*, 2317-2324.

283. Lin, L.-N.; Brandts, J. F. Kinetic mechanism for conformational transitions between poly-L-prolines I and II: A study utilizing the cis-trans specificity of a proline-specific protease. *Biochemistry* **1980**, *19*, 3055-3059.
284. Berisio, R.; Vitagliano, L. Polyproline and triple-helix motifs in host-pathogen recognition. *Curr. Protein Pept. Sci.* **2012**, *13*, 855-865.
285. Feng, S.; Chen, J. K.; Yu, H.; Simon, J. A.; Schreiber, S. L. Two binding orientations for peptides to the Src SH3 domain: Development of a general model for SH3-ligand interactions. *Science* **1994**, *266*, 1241-1247.
286. Ruzza, P.; Siligardi, G.; Donella-Deana, A.; Calderan, A.; Hussain, R.; Rubini, C.; Cesaro, L.; Osler, A.; Guiotto, A.; Pinna, L. A.; Borin, G. 4-Fluoroproline derivative peptides: Effect on PPII conformation and SH3 affinity. *J. Pept. Sci.* **2006**, *12*, 462-471.
287. Zheng, T. Y.; Lin, Y. J.; Horng, J. C. Thermodynamic consequences of incorporating 4-substituted proline derivatives into a small helical protein. *Biochemistry* **2010**, *49*, 4255-4263.
288. Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. High-resolution X-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7517-7522.
289. Tang, H. C.; Lin, Y. J.; Horng, J. C. Modulating the folding stability and ligand binding affinity of Pin1 WW domain by proline ring puckering. *Proteins* **2014**, *82*, 67-76.
290. Wintjens, R.; Wieruszeski, J. M.; Drobecq, H.; Rousselot-Pailley, P.; Buee, L.; Lippens, G.; Landrieu, I. ¹H NMR study on the binding of Pin1 Trp-Trp domain with phosphothreonine peptides. *J. Biol. Chem.* **2001**, *276*, 25150-25156.
291. Naduthambi, D.; Zondlo, N. J. Stereoelectronic tuning of the structure and stability of the Trp cage miniprotein. *J. Am. Chem. Soc.* **2006**, *128*, 12430-12431.
292. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425-430.
293. Boulegue, C.; Milbradt, A. G.; Renner, C.; Moroder, L. Single proline residues can dictate the oxidative folding pathways of cysteine-rich peptides. *J. Mol. Biol.* **2006**, *358*, 846-856.
294. Pokidysheva, E.; Milbradt, A. G.; Meier, S.; Renner, C.; Haussinger, D.; Bachinger, H. P.; Moroder, L.; Grzesiek, S.; Holstein, T. W.; Ozbek, S.; Engel, J. The structure of the Cys-rich terminal domain of hydra minicollagen, which is involved in disulfide networks of the nematocyst wall. *J. Biol. Chem.* **2004**, *279*, 30395-30401.

295. Barth, D.; Musiol, H. J.; Schutt, M.; Fiori, S.; Milbradt, A. G.; Renner, C.; Moroder, L. The role of cystine knots in collagen folding and stability, part I. Conformational properties of (Pro-Hyp-Gly)₅ and (Pro-(4S)-FPro-Gly)₅ model trimers with an artificial cystine knot. *Chem. Biol.* **2003**, *9*, 3692-3702.
296. Scholz, S.; Liebler, E. K.; Eickmann, B.; Fritz, H. J.; Diederichsen, U. Variation of the intercalating proline in artificial peptides mimicking the DNA binding and bending ihf protein. *Amino Acids* **2012**, *43*, 289-298.
297. Rice, P. A.; Yang, S.-w.; Mizuuchi, K.; Nash, H. A. Crystal structure of an IHF–DNA complex: A protein-induced DNA U-turn. *Cell* **1996**, *87*, 1295-1306.
298. Lepthien, S.; Merkel, L.; Budisa, N. In vivo double and triple labeling of proteins using synthetic amino acids. *Angew. Chem. Int. Ed.* **2010**, *49*, 5446-5450.
299. Merkel, L.; Schauer, M.; Antranikian, G.; Budisa, N. Parallel incorporation of different fluorinated amino acids: On the way to “Teflon” proteins. *ChemBioChem* **2010**, *11*, 1505-1507.
300. Hoesl, M. G.; Budisa, N. Expanding and engineering the genetic code in a single expression experiment. *ChemBioChem* **2011**, *12*, 552-555.
301. Oldach, F.; Al Toma, R.; Kuthning, A.; Caetano, T.; Mendo, S.; Budisa, N.; Sussmuth, R. D. Congeneric lantibiotics from ribosomal in vivo peptide synthesis with noncanonical amino acids. *Angew. Chem. Int. Ed.* **2012**, *51*, 415-418.
302. Larregola, M.; Moore, S.; Budisa, N. Congeneric bio-adhesive mussel foot proteins designed by modified prolines revealed a chiral bias in unnatural translation. *Biochem. Biophys. Res. Commun.* **2012**, *421*, 646-650.
303. Hoesl, M. G.; Acevedo-Rocha, C. G.; Nehring, S.; Royter, M.; Wolschner, C.; Wiltshi, B.; Budisa, N.; Antranikian, G. Lipase congeners designed by genetic code engineering. *ChemCatChem* **2011**, *3*, 213-221.
304. Nilsson, B. L.; Soellner, M. B.; Raines, R. T. Chemical synthesis of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91-118.
305. Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. Fluoroprolines as tools for protein design and engineering. *Angew. Chem. Int. Ed.* **2001**, *40*, 923-925.
306. Golbik, R.; Yu, C.; Weyher-Stingl, E.; Huber, R.; Moroder, L.; Budisa, N.; Schiene-Fischer, C. Peptidyl prolyl *cis/trans*-isomerases: Comparative reactivities of cyclophilins, FK506-binding proteins, and parvulins with fluorinated oligopeptide and protein substrates. *Biochemistry* **2005**, *44*, 16026-16034.

307. Kim, W.; George, A.; Evans, M.; Conticello, V. P. Cotranslational incorporation of a structurally diverse series of proline analogues in an *Escherichia coli* expression system. *ChemBioChem* **2004**, *5*, 928-936.
308. Debelle, L.; Tamburro, A. M. Elastin: Molecular description and function. *Int. J. Biochem. Cell Biol.* **1999**, *31*, 261-272.
309. Kim, W.; McMillian, R. A.; Snyder, J. P.; Conticello, V. P. A stereoelectronic effect on turn formation due to proline substitution in elastin-mimetic polypeptides. *J. Am. Chem. Soc.* **2005**, *127*, 18121-18132.
310. Crespo, M. D.; Rubini, M. Rational design of protein stability: Effect of (2*S*,4*R*)-4-fluoroproline on the stability and folding pathway of ubiquitin. *PLoS One* **2011**, *6*, e19425.
311. Komander, D.; Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **2012**, *81*, 203-229.
312. Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **1987**, *194*, 531-544.
313. Edwardraja, S.; Sriram, S.; Govindan, R.; Budisa, N.; Lee, S. G. Enhancing the thermal stability of a single-chain Fv fragment by *in vivo* global fluorination of the proline residues. *Mol. Biosyst.* **2011**, *7*, 258-265.
314. Steiner, T.; Hess, P.; Bae, J. H.; Wiltshi, B.; Moroder, L.; Budisa, N. Synthetic biology of proteins: Tuning GFPs folding and stability with fluoroproline. *PLoS One* **2008**, *3*, e1680.
315. Deepankumar, K.; Nadarajan, S. P.; Ayyadurai, N.; Yun, H. Enhancing the biophysical properties of mRFP1 through incorporation of fluoroproline. *Biochem. Biophys. Res. Commun.* **2013**, *440*, 509-514.
316. Holzberger, B.; Marx, A. Replacing 32 proline residues by a noncanonical amino acid results in a highly active DNA polymerase. *J. Am. Chem. Soc.* **2010**, *132*, 15708-15713.
317. Holzberger, B.; Obeid, S.; Welte, W.; Diederichs, K.; Marx, A. Structural insights into the potential of 4-fluoroproline to modulate biophysical properties of proteins. *Chem. Sci.* **2012**, *3*, 2924-2931.
318. Rubini, M.; Scharer, M. A.; Capitani, G.; Glockshuber, R. (4*R*)- and (4*S*)-fluoroproline in the conserved *cis*-prolyl peptide bond of the thioredoxin fold: Tertiary structure context dictates ring puckering. *ChemBioChem* **2013**, *14*, 1053-1057.
319. Torbeev, V. Y.; Hilvert, D. Both the *cis-trans* equilibrium and isomerization dynamics of a single proline amide modulate β 2-microglobulin amyloid assembly. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20051-20056.

320. Trinh, C. H.; Smith, D. P.; Kalverda, A. P.; Phillips, S. E.; Radford, S. E. Crystal structure of monomeric human β -2-microglobulin reveals clues to its amyloidogenic properties. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9771-9776.
321. Jahn, T. R.; Parker, M. J.; Homans, S. W.; Radford, S. E. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* **2006**, *13*, 195-201.
322. Eichner, T.; Radford, S. E. Understanding the complex mechanisms of β_2 -microglobulin amyloid assembly. *FEBS J.* **2011**, *278*, 3868-3883.
323. Park, B. K.; Kitteringham, N. R.; O'Neill, P. M. Metabolism of fluorine-containing drugs. *Annu. Rev. Pharmacol. Toxicol.* **2001**, *41*, 443-470.
324. Tran, T. T.; Patino, N.; Condom, R.; Frogier, T.; Guedj, R. Fluorinated peptides incorporating a 4-fluoroproline residue as potential inhibitors of HIV protease. *J. Fluorine Chem.* **1997**, *82*, 125-130.
325. Staas, D. D., et al. Discovery of potent, selective 4-fluoroproline-based thrombin inhibitors with improved metabolic stability. *Bioorg. Med. Chem.* **2006**, *14*, 6900-6916.
326. Tucker, T. J., et al. Design and synthesis of a series of potent and orally bioavailable noncovalent thrombin inhibitors that utilize nonbasic groups in the P1 position. *J. Med. Chem.* **1998**, *41*, 3210-3219.
327. Mandal, P. K.; Ren, Z.; Chen, X.; Kalurachchi, K.; Liao, W. S.-L.; McMurray, J. S. Structure-activity studies of phosphopeptidomimetic prodrugs targeting the Src homology 2 (SH2) domain of signal transducer and activator of transcription 3 (Stat3). *Int. J. Pept. Prot. Res.* **2013**, *19*, 3-12.
328. White, J. F.; Noinaj, N.; Shibata, Y.; Love, J.; Kloss, B.; Xu, F.; Gvozdenovic-Jeremic, J.; Shah, P.; Shiloach, J.; Tate, C. G.; Grisshammer, R. Structure of the agonist-bound neurotensin receptor. *Nature* **2012**, *490*, 508-513.
329. Held, C.; Hubner, H.; Kling, R.; Nagel, Y. A.; Wennemers, H.; Gmeiner, P. Impact of the proline residue on ligand binding of neurotensin receptor 2 (NTS2)-selective peptide-peptoid hybrids. *ChemMedChem* **2013**, *8*, 772-778.
330. Zhuang, W.; Zhao, X.; Zhao, G.; Guo, L.; Lian, Y.; Zhou, J.; Fang, D. Synthesis and biological evaluation of 4-fluoroproline and 4-fluoropyrrolidine-2-acetic acid derivatives as new GABA uptake inhibitors. *Bioorg. Med. Chem.* **2009**, *17*, 6540-6546.
331. van der Ley, M. [^{18}F]fluorine labeled aliphatic amino acids. *J. Labelled Cpd. Radiopharm.* **1983**, *20*, 453-461.

332. Hamacher, K. Synthesis of N.C.A. *cis*- and *trans*-4-[¹⁸F]fluoro-L-proline, radiotracers for PET-investigation of disordered matrix protein synthesis. *J. Labelled Cpd. Radiopharm.* **1999**, *42*, 1135-1144.
333. Mazza, S. M. Stereospecific, semi-automated, N.C.A. syntheses of *cis*-4-[¹⁸F]fluoro-L-proline and *trans*-4-[¹⁸F]fluoro-L-proline. *J. Labelled Cpd. Radiopharm.* **2000**, *43*, 1047-1058.
334. Azad, B. B.; Ashique, R.; Labiris, N. R.; Chirakal, R. Temperature effects on the stereospecificity of nucleophilic fluorination: Formation of *trans*-4-[¹⁸F]fluoro-L-proline during the synthesis of *cis*-4-[¹⁸F]fluoro-L-proline. *J. Labelled Cpd. Radiopharm.* **2012**, *55*, 23-28.
335. Wester, H.-J.; Herz, M.; Senekowitsch-Schmidtke, R.; Schwaiger, M.; Stocklin, G.; Hamacher, K. Preclinical evaluation of 4-[¹⁸F]fluoroprolines: Diastereomeric effect on metabolism and uptake in mice. *Nucl. Med. Biol.* **1999**, *26*, 259-265.
336. Borner, A. R.; Langen, K.-J.; Herzog, H.; Hamacher, K.; Muller-Mattheis, V.; Schmitz, T.; Ackermann, R.; Coenen, H. H. Whole-body kinetics and dosimetry of *cis*-4-[¹⁸F]fluoro-L-proline. *Nucl. Med. Biol.* **2001**, *28*, 287-292.
337. Langen, K.-J.; Borner, A. R.; Muller-Mattheis, V.; Hamacher, K.; Herzog, H.; Ackermann, R.; Coenen, H. H. Uptake of *cis*-4-[¹⁸F]fluoro-L-proline in urologic tumors. *J. Nucl. Med.* **2001**, *42*, 752-754.
338. Langen, K.-J.; Muhlensiepen, H.; Schmieder, S.; Hamacher, K.; Broer, S.; Borner, A. R.; Schneeweiss, F. H. A.; Coenen, H. H. Transport of *cis*- and *trans*-4-[¹⁸F]fluoro-L-proline in F98 glioma cells. *Nucl. Med. Biol.* **2002**, *29*, 685-692.
339. Langen, K.-J.; Jarosch, M.; Hamacher, K.; Muhlensiepen, H.; Weber, F.; Floeth, F.; Pauleit, D.; Herzog, H.; Coenen, H. H. Imaging of gliomas with *cis*-4-[¹⁸F]fluoro-L-proline. *Nucl. Med. Biol.* **2004**, *31*, 67-75.
340. Langen, K. J.; Hamacher, K.; Bauer, D.; Broer, S.; Pauleit, D.; Herzog, H.; Floeth, F.; Zilles, K.; Coenen, H. H. Preferred stereoselective transport of the D-isomer of *cis*-4-[¹⁸F]fluoro-proline at the blood–brain barrier. *J. Cereb. Blood Flow Metab.* **2005**, *25*, 607-616.
341. Zimny, M.; Klosterhalfen, B.; Conze, J.; Hamacher, K.; Fehler, S.; Schumpelick, V.; Coenen, H. H.; Buell, U. Uptake of *cis*-4-[¹⁸F]fluoro-L-proline in scar formation: A marker of collagen synthesis? *Nucl. Med. Commun.* **2002**, *23*, 695-698.
342. Skovgaard, D.; Kjaer, A.; Heinemeier, K. M.; Brandt-Larsen, M.; Madsen, J.; Kjaer, M. Use of *cis*-[¹⁸F]fluoro-proline for assessment of exercise-related collagen synthesis in musculoskeletal connective tissue. *PLoS One* **2011**, *6*, e16678.

343. Shoulders, M. D.; Guzei, I. A.; Raines, R. T. 4-Chloroprolines: Synthesis, conformational analysis, and effect on the collagen triple helix. *Biopolymers* **2008**, *89*, 443-54.
344. Kuemin, M.; Schweizer, S.; Ochsenfeld, C.; Wennemers, H. Effects of terminal functional groups on the stability of the polyproline II structure: A combined experimental and theoretical study. *J. Am. Chem. Soc.* **2009**, *131*, 15474-15482.
345. Erdmann, R. S.; Wennemers, H. Effect of sterically demanding substituents on the conformational stability of the collagen triple helix. *J. Am. Chem. Soc.* **2012**, *134*, 17117-17124.
346. Cadamuro, S. A.; Reichold, R.; Kusebauch, U.; Musiol, H. J.; Renner, C.; Tavan, P.; Moroder, L. Conformational properties of 4-mercaptoproline and related derivatives. *Angew. Chem. Int. Ed.* **2008**, *47*, 2143-2146.
347. Cipolla, L.; Airoidi, C.; Bini, D.; Gregori, M.; Marcelo, F.; Jiménez-Barbero, J.; Nicotra, F. Fructose-based proline analogues: Exploring the prolyl *trans/cis*-amide rotamer population in model peptides. *Eur. J. Org. Chem.* **2011**, *2011*, 128-136.
348. DeTar, D. F.; Luthra, N. P. Conformations of proline. *J. Am. Chem. Soc.* **1977**, *99*, 1232-1244.
349. Erdmann, R. S.; Wennemers, H. Conformational stability of collagen triple helices functionalized in the Yaa position by click chemistry. *Org. Biomol. Chem.* **2012**, *10*, 1982-1986.
350. Ivanova, G.; Yakimova, B.; Angelova, S.; Stoineva, I.; Enchev, V. Influence of pH on the *cis-trans* isomerization of valine-proline dipeptide: An integrated NMR and theoretical investigation. *J. Mol. Struct.* **2010**, *975*, 330-334.
351. Lee, K. K.; Park, K. H.; Joo, C.; Kwon, H. J.; Jeon, J.; Jung, H. I.; Park, S.; Han, H.; Cho, M. Infrared probing of 4-azidoproline conformations modulated by azido configurations. *J. Phys. Chem. B* **2012**, *116*, 5097-5110.
352. Thomas, K. M.; Naduthambi, D.; Zondlo, N. J. Electronic control of amide *cis-trans* isomerism via the aromatic-prolyl interaction. *J. Am. Chem. Soc.* **2006**, *128*, 2216-2217.
353. Taylor, C. M.; Handre, R.; Edwards, P. J. B. The impact of pyrrolidine hydroxylation on the conformation of proline-containing peptides. *J. Org. Chem.* **2005**, *70*, 1306-1315.
354. Taylor, C. M.; Hardre, R.; Edwards, P. J. B.; Park, J. H. Factors affecting conformation in proline-containing peptides. *Org. Lett.* **2004**, *5*, 4413-4416.
355. Torbeev, V. Y.; Fumi, E.; Ebert, M.-O.; Schweizer, W. B.; Hilvert, D. *cis-trans* Peptide-bond isomerization in α -methylproline derivatives. *Helv. Chim. Acta* **2012**, *95*, 2411-2420.

356. Williams, K. R.; Adhyaru, B.; German, I.; Alvarez, E. The cis-trans equilibrium of *N*-acetyl-L-proline. *J. Chem. Ed.* **2002**, *79*, 372-373.
357. Zhang, K.; Teklebrhan, R. B.; Schreckenbach, G.; Wetmore, S.; Schweizer, F. Intramolecular hydrogen bond-controlled prolyl amide isomerization in glucosyl 3'(*S*)-hydroxy-5'-hydroxymethylproline hybrids: Influence of a C-5'-hydroxymethyl substituent on the thermodynamics and kinetics of prolyl amide *cis/trans* isomerization. *J. Org. Chem.* **2009**, *74*, 3735-3743.
358. Liang, G.-B.; Rito, C. J.; Gellman, S. H. Variations in the turn-forming characteristics of *N*-acyl proline units. *Biopolymers* **1992**, *32*, 293-301.
359. Benzi, C.; Improta, R.; Scalmani, G.; Barone, V. Quantum mechanical study of the conformational behavior of proline and 4*R*-hydroxyproline dipeptide analogues in vacuum and in aqueous solution. *J. Comp. Chem.* **2002**, *23*, 341-350.
360. Guzei, I. A.; Choudhary, A.; Raines, R. T. Pyramidalization of a carbonyl carbon atom in *N*-selenoacetyl-(2*S*)-proline methyl ester. *Acta Crystallogr.* **2013**, *E69*, o805-o806.
361. Lide, D. R., Dipole moments. In *CRC Handbook of Chemistry & Physics*, 93 ed.; CRC Press: Boca Raton, FL, 2013.
362. Wiberg, K. B.; Rablen, P. R. Why does thioformamide have a larger rotational barrier than formamide? *J. Am. Chem. Soc.* **1995**, *117*, 2201-2209.
363. Wiberg, K. B.; Rush, D. J. Solvent effects on the thioamide rotational barrier: An experimental and theoretical study. *J. Am. Chem. Soc.* **2001**, *123*, 2038-2046.
364. Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **2003**, *424*, 805-808.
365. Eisenberg, D.; Jucker, M. The amyloid state of proteins in human diseases. *Cell* **2012**, *148*, 1188-1203.
366. Shalaby, M. A.; Grote, C. W.; Rapoport, H. Thiopeptide synthesis. α -Amino thionoacid derivatives of nitrobenzotriazole as thioacylating agents. *J. Org. Chem.* **1996**, *61*, 9045-9048.
367. *Mestrenova*, 9.0; MestreLab Research: Escondido, CA.
368. Bruker *SADABS*, Bruker AXS, Inc.: Madison, WI, 2015.
369. Sheldrick, G. M. ShelXL. *Acta Crystallogr.* **2015**, *C71*, 3-8.
370. Frisch, M. J., et al. *Gaussian 09, Revision A.1*, Gaussian, Inc.: Wallingford, CT, 2009.

371. Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. *NBO 5.9*, Theoretical Chemistry Institute, University of Wisconsin–Madison: Madison, WI, 2012.
372. Okuyama, K.; Hongo, C.; Kukushima, R.; Wu, G.; Narita, H.; Noguchi, K.; Tanaka, Y.; Nishino, N. Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: Implications for proline ring puckering. *Biopolymers* **2004**, *76*, 367-377.
373. Shoulders, M. D.; Raines, R. T. Interstrand dipole-dipole interactions can stabilize the collagen triple helix. *J. Biol. Chem.* **2011**, *286*, 22905-22912.
374. Erdmann, R. S.; Wennemers, H. Importance of ring puckering versus interstrand hydrogen bonds for the conformational stability of collagen. *Angew. Chem. Int. Ed.* **2011**, *50*, 6835-6838.
375. Siebler, C.; Erdmann, R. S.; Wennemers, H. Switchable proline derivatives: Tuning the conformational stability of the collagen triple helix by pH changes. *Angew. Chem. Int. Ed.* **2014**, *53*, 10340-10344.
376. Wiberg, K. B.; Rush, D. J. Methyl rotational barriers in amides and thioamides. *J. Org. Chem.* **2002**, *67*, 826-830.
377. Frank, R.; Jakob, M.; Thunecke, F.; Fischer, G.; Schutkowski, M. Thioxylation as one-atom-substitution generates a photoswitchable element within the peptide backbone. *Angew. Chem. Int. Ed.* **2000**, *39*, 1120-1122.
378. Lee, H.-J.; Choi, Y.-S.; Lee, K.-B.; Park, J.; Yoon, C.-J. Hydrogen bonding abilities of thioamide. *J. Phys. Chem. A* **2002**, *106*, 7010-7017.
379. Zacharie, B.; Sauve, G.; Penney, C. Thioacylating agents. Use of thiobenzimidazolone derivatives for the preparation of thiotuftsins analogs. *Tetrahedron* **1993**, *49*, 10489-10500.
380. Hoeg-Jensen, T.; Olsen, C. E.; Holm, A. Thioacylation achieved by activation of a monothiocarboxylic acid with phosphorus reagents. *J. Org. Chem.* **1994**, *59*, 1257-1263.
381. Brain, C. T.; Hallett, A.; Ko, S. Y. Thioamide synthesis: Thioacyl-*N*-phthalimides as thioacylating agents. *J. Org. Chem.* **1997**, *62*, 3808-3809.
382. Katritzky, A. R.; Witek, R. M.; Rodriguez-Garcia, V.; Mohapatra, P. P.; Rogers, J. W.; Cusido, J.; Abdel-Fattah, A. A. A.; Steel, P. J. Benzotriazole-assisted thioacylation. *J. Org. Chem.* **2005**, *70*, 7866-7881.
383. Brown, D. W.; Campbell, M. M.; Walker, C. V. Endothiopeptides. *Tetrahedron* **1983**, *39*, 1075-1083.

384. Jurayj, J.; Cushman, M. Approaches to the synthesis of endothiopeptides: Synthesis of a thioamide-containing C-terminal bombesin nonapeptide. *Tetrahedron* **1992**, *48*, 8601-8614.
385. Sifferlen, T.; Rueping, M.; Gademann, K.; Jaun, B.; Seebach, D. β -Thiopeptides: Synthesis, NMR solution structure, CD spectra, and photochemistry. *Helv. Chim. Acta* **1999**, *82*, 2067-2093.
386. Formaggio, F.; Crisma, M.; Toniolo, C.; Peggion, C. All-thioamidated homo- α -peptides: Synthesis and conformation. *Eur. J. Org. Chem.* **2013**, *2013*, 3455-3463.
387. Batjargal, S.; Wang, Y. J.; Goldberg, J. M.; Wissner, R. F.; Petersson, E. J. Native chemical ligation of thioamide-containing peptides: Development and application to the synthesis of labeled α -synuclein for misfolding studies. *J. Am. Chem. Soc.* **2012**, *134*, 9172-9182.
388. Wissner, R. F.; Batjargal, S.; Fadzen, C. M.; Petersson, E. J. Labeling proteins with fluorophore/thioamide Förster resonant energy transfer pairs by combining unnatural amino acid mutagenesis and native chemical ligation. *J. Am. Chem. Soc.* **2013**, *135*, 6529-40.
389. Batjargal, S.; Huang, Y.; Wang, Y. J.; Petersson, E. J. Synthesis of thioester peptides for the incorporation of thioamides into proteins by native chemical ligation. *J. Pept. Sci.* **2014**, *20*, 87-91.
390. We speculate that stacking interactions between the electron-rich Fmoc group and electron-deficient triazole slowed acylation, allowing for loss of sulfur by hydrolysis.
391. Ozturk, T.; Ertas, R.; Mert, O. Use of Lawesson's reagent in organic syntheses. *Chem. Rev.* **2007**, *107*, 5210-5278.
392. Thorsen, M.; Yde, B.; Pederson, U.; Clausen, K.; Lawesson, S.-O. Studies on amino acids and peptides—V. Syntheses of endothionated melanostatin analogs. *Tetrahedron* **1983**, *39*, 3429-3435.
393. Jensen, O. E.; Lawesson, S.-O. Studies on amino acids and peptides VIII: Synthesis and crystal structure of two monothiated analogies of Boc-Gly-S-Ala-Aib-OMe. *Tetrahedron* **1985**, *41*, 5595-5606.
394. Hollosi, M.; Kollat, E.; Kajtar, J.; Kajtar, M.; Fasman, G. D. Chiroptical labeling of folded polypeptide conformations: The thioamide probe. *Biopolymers* **1990**, *30*, 1061-1072.
395. De Poli, M.; Clayden, J. Thionoglycine as a multifunctional spectroscopic reporter of screw-sense preference in helical foldamers. *Org. Biomol. Chem.* **2014**, *12*, 836-843.

396. Chen, Y. S.; Chen, C. C.; Horng, J. C. Thermodynamic and kinetic consequences of substituting glycine at different positions in a Pro-Hyp-Gly repeat collagen model peptide. *Biopolymers* **2011**, *96*, 60-68.
397. Stultz, C. M. The folding mechanism of collagen-like model peptides explored through detailed molecular simulations. *Protein Sci.* **2006**, *15*, 2166-2177.
398. Huang, Y.; Cong, Z.; Yang, L.; Dong, S. A photoswitchable thioxopeptide bond facilitates the conformation-activity correlation study of insect kinin. *J. Pept. Sci.* **2008**, *14*, 1062-1068.
399. Bregy, H.; Heimgartner, H.; Helbing, J. A time-resolved spectroscopic comparison of the photoisomerization of small β -turn-forming thioxopeptides. *J. Phys. Chem. B* **2009**, *113*, 1756-1762.
400. Goldberg, J. M.; Batjargal, S.; Petersson, E. J. Thioamides as fluorescence quenching probes: Minimalist chromophores to monitor protein dynamics. *J. Am. Chem. Soc.* **2010**, *132*, 14718-14720.
401. Yao, S.; Zutshi, R.; Chmielewski, J. Endothiopeptide inhibitors of HIV-1 protease. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 699-704.
402. Siebler, C.; Erdmann, R. S.; Wennemers, H. From azidoproline to functionalizable collagen. *Chimia* **2013**, *67*, 891-895.
403. Erdmann, R. S.; Wennemers, H. Functionalizable collagen model peptides. *J. Am. Chem. Soc.* **2010**, *132*, 13957-13959.
404. Scheirs, J.; Long, T. E., *Modern Polyesters: Chemistry and Technology of Polyesters and Copolyesters*. John Wiley & Sons: West Sussex, England, 2003.
405. Piemonte, V., *Polylactic Acid: Synthesis, Properties and Applications*. Nova Science Publishers: Hauppauge, NY, 2011.
406. Sin, L. T.; Rahmat, A. R.; Rahman, W. A. W. A., *Polylactic Acid: PLA Biopolymer Technology and Applications*. Elsevier: Oxford, UK, 2012.
407. Garlotta, D. A literature review of poly(lactic acid). *J. Polym. Environ.* **2001**, *9*, 63-84.
408. Ikada, Y.; Tsuji, H. Biodegradable polyesters for medical and ecological applications. *Macromol. Rapid Commun.* **1999**, *21*, 117-132.
409. Puiggali, J.; Ikada, Y.; Tsuji, H.; Cartier, L.; Okihara, T.; Lotz, B. The frustrated structure of poly-L-lactide). *Polymer* **2000**, *41*, 8921-8930.

410. Cartier, L.; Okihara, T.; Ikada, Y.; Tsuji, H.; Puiggali, J.; Lotz, B. Epitaxial crystallization and crystalline polymorphism of polylactides. *Polymer* **2000**, *41*, 8909-8919.
411. Santis, P. D.; Kovacs, A. J. Molecular conformation of poly(*S*-lactic acid). *Biopolymers* **1968**, *6*, 299-306.
412. Hoogsteen, W.; Postema, A. R.; Pennings, A. J.; Brinke, G. t. Crystal structure, conformation, and morphology of solution-spun poly(L-lactide) fibers. *Macromolecules* **1990**, *23*, 634-642.
413. Kobayashi, J.; Asahi, T.; Ichiki, M.; Oikawa, A.; Suzuki, H.; Watanabe, T.; Fukada, E.; Shikinami, Y. Structural and optical properties of poly lactic acids. *J. App. Phys.* **1995**, *77*, 2957-2973.
414. Aleman, C.; Lotz, B.; Puiggali, J. Crystal structure of the alpha-form of poly(L-lactide). *Macromolecules* **2001**, *34*, 4795-4801.
415. Sasaki, S.; Asakura, T. Helix distortion and crystal structure of the α -form of poly(L-lactide). *Macromolecules* **2003**, *36*, 8385-8390.
416. Wasanasuk, K.; Tashiro, K.; Hanesaka, M.; Ohhara, T.; Kurihara, K.; Kuroki, R.; Tamada, T.; Ozeki, T.; Kanamoto, T. Crystal structure analysis of poly(L-lactic acid) α form on the basis of the 2-dimensional wide-angle synchrotron X-ray and neutron diffraction measurements. *Macromolecules* **2011**, *44*, 6441-6452.
417. Pawlak, T.; Jaworska, M.; Potrzebowski, M. J. NMR crystallography of α -poly(L-lactide). *Phys. Chem. Chem. Phys.* **2013**, *15*, 3137-3145.
418. Deechongkit, S.; Dawson, P. E.; Kelly, J. W. Toward assessing the position-dependent contributions of backbone hydrogen bonding to β -sheet folding thermodynamics employing amide-to-ester perturbations. *J. Am. Chem. Soc.* **2004**, *126*, 16762-16771.
419. Fu, Y.; Gao, J.; Bieschke, J.; Dendle, M. A.; Kelly, J. W. Amide-to-*E*-olefin versus amide-to-ester backbone H-bond perturbations: Evaluating the O–O repulsion for extracting H-bond energies. *J. Am. Chem. Soc.* **2006**, *128*, 15948-15949.
420. Koh, J. T.; Cornish, V. W.; Schultz, P. G. An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **1997**, *36*, 11314-11322.
421. Deechongkit, S.; Ngyuen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. Context-dependent contributions of backbone hydrogen bonding to β -sheet folding energetics. *Nature* **2004**, *430*, 101-105.

- 422. Shi, Z.; Woody, R. W.; Kallenbach, N. R. Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem.* **2002**, *62*, 163-240.
- 423. Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9190-9195.
- 424. Hamburger, J. B.; Ferreón, J. C.; Whitten, S. T.; Hilser, V. J. Thermodynamic mechanism and consequences of the polyproline II (PPII) structural bias in the denatured states of proteins. *Biochemistry* **2004**, *43*, 9790-9799.
- 425. Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. Urea promotes polyproline II helix formation: Implications for protein denatured states. *Biochemistry* **2005**, *44*, 6269-6275.
- 426. Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins* **2004**, *55*, 502-507.
- 427. Grdadolnik, J.; Mohacek-Grosev, V.; Baldwin, R. L.; Avbelj, F. Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1794-1798.
- 428. Shi, Z.; Kallenbach, N. R. Ramachandran redux. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3-4.
- 429. Esposito, L.; Vitagliano, L.; Zagari, A.; Mazzarella, L. Pyramidalization of backbone carbonyl carbon atoms in proteins. *Protein Sci.* **2001**, *9*, 2038-2042.
- 430. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- 431. Wang, G.; Dunbrack, R. L. PISCES: A protein sequence culling server. *Bioinformatics* **2003**, *19*, 1589-1591.
- 432. Improtà, R.; Vitagliano, L.; Esposito, L. Peptide bond distortions from planarity: New insights from quantum mechanical calculations and peptide/protein crystal structures. *PLoS One* **2011**, *6*, 1-10.
- 433. Hutchinson, E. G.; Thornton, J. M. PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci.* **1996**, *5*, 212-220.
- 434. Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735-746.

435. Lario, P. I.; Vrielink, A. Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution. *J. Am. Chem. Soc.* **2003**, *125*, 12787-12794.
436. Waters, C. M.; Bassler, B. L. Quorum sensing: Cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 319-346.
437. Hastings, J. W.; Greenberg, E. P. Quorum sensing: The explanation of a curious phenomenon reveals a common characteristic of bacteria. *J. Bacteriol.* **1999**, *181*, 2667-2668.
438. Eberhard, A.; Burlingame, A. L.; Eberhard, C.; Kenyon, G. L.; Nealson, K. H.; Oppenheimer, N. J. Structural identification of autoinducer of *Photobacterium fischeri* luciferase. *Biochemistry* **1981**, *20*, 2444-2449.
439. Fuqua, C.; Parsek, M. R.; Greenberg, E. P. Regulation of gene expression by cell-to-cell communication: Acyl-homoserine lactone quorum sensing. *Annu. Rev. Gen.* **2001**, *35*, 439-468.
440. Churchill, M. E. A.; Chen, L. Structural basis of acyl-homoserine lactone-dependent signaling. *Chem. Rev.* **2011**, *111*, 68-85.
441. Ng, W.-L.; Bassler, B. L. Bacterial quorum-sensing network architectures. *Annu. Rev. Genet.* **2009**, *43*, 197-222.
442. Rasmussen, T. B.; Givskov, M. Quorum-sensing inhibitors as anti-pathogenic drugs. *Int. J. Med. Microbiol.* **2006**, *296*, 149-161.
443. Stringer, J. R.; Crapster, J. A.; Guzei, I. A.; Blackwell, H. E. Construction of peptoids with all *trans*-amide backbones and peptoid reverse turns via the tactical incorporation of *N*-aryl side chains capable of hydrogen bonding. *J. Org. Chem.* **2010**, *75*, 6068-6078.
444. Pauling, L. The nature of the chemical bond. III. The transition from one extreme bond type to another. *J. Am. Chem. Soc.* **1932**, *54*, 988-1003.
445. Pauling, L., In *The Nature of the Chemical Bond*, Cornell University Press: Ithaca, NY, 1939; pp 186-193.
446. Choudhary, A.; Fry, C. G.; Raines, R. T. Modulation of an $n \rightarrow \pi^*$ interaction with α -fluoro groups. *ARKIVOC* **2010**, 251-262.
447. Rackham, D. M.; Chakrabarti, J. K.; Davies, G. L. O. Study of hydrolysis of amido-substituted γ -butyrolactones by pH-stat titration and ultraviolet spectroscopic analyses. *Talanta* **1981**, *28*, 329-331.
448. Yates, E. A.; Philipp, B.; Buckley, C.; Atkinson, S.; Chhabra, S. R.; Sockett, R. E.; Goldner, M.; Dessaux, Y.; Cámara, M.; Smith, H.; Williams, P. *N*-Acylhomoserine

- lactones undergo lactonolysis in a pH-, temperature-, and acyl chain length-dependent manner during growth of *Yersinia pseudotuberculosis* and *Pseudomonas aeruginos*. *Infect. Immun.* **2002**, *70*, 5635-5646.
449. Choudhary, A.; Fry, C. G.; Kamer, K. J.; Raines, R. T. An $n \rightarrow \pi^*$ interaction reduces the electrophilicity of the acceptor carbonyl group. *Chem. Commun.* **2013**, *49*, 8166-8168.
 450. Geske, G. D.; O'Neill, J. C.; Miller, D. M.; Mattmann, M. E.; Blackwell, H. E. Modulation of bacterial quorum sensing with synthetic ligands: Systematic evaluation of *N*-acylated homoserine lactones in multiple species and new insights into their mechanisms of action. *J. Am. Chem. Soc.* **2007**, *129*, 13613-13625.
 451. Sheldrick, G. M. ShelXS. *Acta Crystallogr.* **2008**, *A64*, 112-122.
 452. Miller, M. B.; Bassler, B. L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **2001**, *55*, 165-199.
 453. Choudhary, A.; Newberry, R. W.; Raines, R. T. $n \rightarrow \pi^*$ Interactions engender chirality in carbonyl groups. *Org. Lett.* **2014**, *16*, 3421-3423.
 454. Eberhard, A.; Schineller, J. B. Chemical synthesis of bacterial autoinducers and analogs. *Meth. Enz.* **2000**, *305*, 301-315.
 455. Clauss, A. D.; Nelsen, S. F.; Ayoub, M.; Moore, J. W.; Landis, C. R.; Weinhold, F. Rabbit-ears hybrids, VSEPR sterics, and other orbital anachronisms. *Chem. Educ. Res. Pract.* **2014**, *15*, 417-434.
 456. Burgess, A. W.; Scheraga, H. A. Stable conformations of dipeptides. *Biopolymers* **1973**, *12*, 2177-2183.
 457. Toniolo, C. Intramolecularly hydrogen-bonded peptide conformations. *Crit. Rev. Biochem.* **1980**, *9*, 1-44.
 458. Blanco, S.; Lesarri, A.; Lopez, J. C.; Alonso, J. L. The gas-phase structure of alanine. *J. Am. Chem. Soc.* **2004**, *126*, 11675-11683.
 459. Dian, B. C.; Longarte, A.; Mercier, S.; Evans, D. A.; Wales, D. J.; Zwier, T. S. The infrared and ultraviolet spectra of single conformations of methyl-capped dipeptides: *N*-acetyl tryptophan amide and *N*-acetyl tryptophan methyl amide. *J. Chem. Phys.* **2002**, *117*, 10688-10702.
 460. Zimmerman, S. S.; Pottle, M. S.; Nemethy, G.; Scheraga, H. S. Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. *Macromolecules* **1977**, *10*, 1-9.
 461. Toniolo, C.; Bonora, G. M.; Bavoso, A.; Benedetti, E.; Blasio, B. D.; Pavone, V.; Pedone, C.; Barone, V.; Lelj, F.; Leplawy, M. T.; Kaczmarek, K.; Redlinski, A.

- Structural versatility of peptides from C ^{α,α} -dialkylated glycines. II. An IR absorption and ¹H-NMR study of homo-oligopeptides from C ^{α,α} -diethylglycine. *Biopolymers* **1988**, *27*, 373-379.
462. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β -hairpins. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5578-5583.
 463. Culik, R. M.; Jo, H.; DeGrado, W. F.; Gai, F. Using thioamides to site-specifically interrogate the dynamics of hydrogen bond formation in β -sheet folding. *J. Am. Chem. Soc.* **2012**, *134*, 8026-8029.
 464. Richardson, J. S.; Richardson, D. C. Natural β -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754-2759.
 465. Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J.; McFarlane, H. T.; Madsen, A. O.; Riek, C.; Eisenberg, D. Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **2007**, *447*, 453-457.
 466. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **1999**, *24*, 329-332.
 467. Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riek, C.; Grothe, R.; Eisenberg, D. Structure of the cross- β spine of amyloid-like fibrils. *Nature* **2005**, *435*, 773-778.
 468. Ohnishi, S.; Kamikubo, H.; Onitsuka, M.; Kataoka, M.; Shortle, D. Conformational preference of polglycine in solution to elongated structure. *J. Am. Chem. Soc.* **2006**, *128*, 16338-16344.
 469. Minor Jr., D. L.; Kim, P. S. Measurement of the β -sheet-forming propensities of amino acids. *Nature* **1994**, *367*, 660-663.
 470. Skinner, J. J.; Lim, W. K.; Bedard, S.; Black, B. E.; Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **2012**, *21*, 996-1005.
 471. Kortemme, T.; Morozov, A. V.; Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **2003**, *326*, 1239-1259.
 472. Hubbard, S. J.; Thornton, J. M. *NACCESS, 2.1.1*, Department of Biochemistry and Molecular Biology, University College: London, 1993.
 473. Deechongkit, S.; You, S.-L.; Kelly, J. W. Synthesis of all nineteen appropriately protected chiral α -hydroxy acid equivalent of the α -amino acids for Boc solid-phase depsi-peptide synthesis. *Org. Lett.* **2004**, *6*, 497-500.

474. Abd-Elgalil, W. R.; Gallazzi, F.; Lever, S. Z. Total solid-phase synthesis of bombesin analogs with different functional groups at the C-terminus. *J. Pept. Sci.* **2007**, *13*, 487-492.
475. Huang, R.; Wu, L.; McElheny, D.; Bour, P.; Roy, A.; Keiderling, T. A. Cross-strand coupling and site-specific unfolding thermodynamics of a TrpZip β -hairpin peptide using ^{13}C isotope labeling and IR spectroscopy. *J. Phys. Chem. B* **2009**, *113*, 5661-5674.
476. Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H. Olex2. *J. Appl. Cryst.* **2009**, *42*, 339-341.
477. Yamada, S. Effects of C(O)–N bond rotation on the ^{13}C , ^{15}N , and ^{17}O NMR chemical shifts, and infrared carbonyl absorption in a series of twisted amides. *J. Org. Chem.* **1996**, *61*, 941-946.
478. Asplund, M. C.; Zanni, M. T.; Hochstrasser, R. M. Two-dimensional infrared spectroscopy of peptides by phase-controlled femtosecond vibrational photon echoes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8219-8224.
479. Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A. Amide I two-dimensional infrared spectroscopy of proteins. *Acc. Chem. Res.* **2008**, *41*, 432-441.
480. Kim, Y. S.; Wang, J.; Hochstrasser, R. M. Two-dimensional infrared spectroscopy of the alanine dipeptide in aqueous solution. *J. Phys. Chem. B* **2005**, *109*, 7511-7521.
481. Chi, Z.; Chen, X. G.; Holtz, J. S. W.; Asher, S. A. UV resonance Raman-selective amide vibrational enhancement: Quantitative methodology for determining protein secondary structure. *Biochemistry* **1998**, *37*, 2854-2864.
482. Mikhonin, A. V.; Asher, S. A. Uncoupled peptide bond vibrations in α -helical and polyproline II conformations of polyalanine peptides. *J. Phys. Chem. B* **2005**, *109*, 3047-3052.
483. Myshakina, N. S.; Asher, S. A. Peptide bond vibrational coupling. *J. Phys. Chem. B* **2007**, *111*, 4271-4279.
484. Newberry, R. W.; VanVeller, B.; Raines, R. T. Thioamides in the collagen triple helix. *Chem. Commun.* **2015**, *51*, 9624-9627.
485. Newberry, R. W.; Raines, R. T. 4-fluoroproline: Conformational analysis and effects on the stability and folding of peptides and proteins. *Top. Heterocycl. Chem.* **2016**.
486. Shepherd, N. E.; Hoang, H. N.; Abbenante, G.; Fairlie, D. P. Single turn peptide α -helices with exceptional stability in water. *J. Am. Chem. Soc.* **2005**, *127*, 2974-2983.

487. Isidro-Llobet, A.; Alvarez, M.; Albericio, F. Amino acid-protecting groups. *Chem. Rev.* **2009**, *109*, 2455-2504.
488. Tran, T. T.; Treutlein, H.; Burgess, A. W. Conformational analysis of thiopeptides: (ϕ, ψ) maps of thio-substituted dipeptides. *J. Comp. Chem.* **2001**, *22*, 1026-1037.
489. Tran, T. T.; Zheng, J.; Treutlein, H.; Burgess, A. W. Effects of thioamide substitutions on the conformation and stability of α - and 3_{10} -helices. *J. Am. Chem. Soc.* **2002**, *124*, 5222-5230.
490. Huang, W.; Eichenberger, A. P.; van Gunsteren, W. F. Molecular dynamics simulation of thionated hen egg white lysozyme. *Protein Sci.* **2012**, *21*, 1153-1161.
491. Miwa, J. H.; Pallivathucal, L.; Gowda, S.; Lee, K. E. Conformational stability of helical peptides containing a thioamide linkage. *Org. Lett.* **2002**, *4*, 4655-4657.
492. Reiner, A.; Wildemann, D.; Fischer, G.; Kiefhaber, T. Effect of thioxopeptide bonds on α -helix structure and stability. *J. Am. Chem. Soc.* **2008**, *130*, 8079-8084.
493. Bachmann, A.; Wildemann, D.; Praetorius, F.; Fischer, G.; Kiefhaber, T. Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3952-3957.
494. Wright, P. E.; Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31-38.
495. Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **2013**, *1834*, 932-951.
496. Koh, M. J.; Nguyen, T. T.; Zhang, H.; Schrock, R. R.; Hoveyda, A. H. Direct synthesis of Z-alkenyl halides through catalytic cross-metathesis. *Nature* **2016**, *531*, 459-465.
497. Toniolo, C.; Borona, G. M.; Barone, V.; Bavoso, A.; Benedetti, E.; Blasio, B. D.; Grimaldi, P.; Lelj, F.; Pavone, V.; Pedone, C. Conformation of pleionomers of α -aminoisobutyric acid. *Macromolecules* **1985**, *18*, 895-902.
498. Toniolo, C.; Borona, G. M.; Bavoso, A.; Benedetti, E.; Blasio, B. D.; Pavone, V.; Pedone, C. A long, regular polypeptide 3_{10} -helix. *Macromolecules* **1986**, *18*, 472-479.
499. Bavoso, A.; Benedetti, E.; Blasio, B. D.; Pavone, V.; Pedone, C.; Toniolo, C.; Bonora, G. M. Long polypeptide 3_{10} -helices at atomic resolution. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1988-1992.
500. Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. M. Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **2002**, *324*, 105-121.

501. Porter, C. T.; Bartlett, G. J.; Thornton, J. M. The Catalytic Site Atlas: A resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.* **2004**, *32*, D129-D133.
502. Isied, S. S.; Ogawa, M. Y.; Wishart, J. F. Peptide-mediated intramolecular electron transfer: Long-range distance dependence. *Chem. Rev.* **1992**, *92*, 381-394.
503. Gray, H. B.; Winkler, J. R. Electron transfer in proteins. *Annu. Rev. Biochem.* **1996**, *65*, 537-561.
504. Gray, H. B.; Winkler, J. R. Electron tunneling through proteins. *Quart. Rev. Biophys.* **2003**, *36*, 341-372.
505. Long, Y. T.; Abu-Irhayem, E.; Kraatz, H. B. Peptide electron transfer: More questions than answers. *Chemistry* **2005**, *11*, 5186-194.
506. Shin, Y.-g. K.; Newton, M. D.; Isied, S. S. Distance dependence of electron transfer across peptides with different secondary structures: The role of peptide energetics and electronic coupling. *J. Am. Chem. Soc.* **2003**, *125*, 3722-3732.
507. Issa, J. B.; Krogh-Jespersen, K.; Isied, S. S. Conformational dependence of electronic coupling across peptide bonds: A ramachandran map. *J. Phys. Chem. C* **2010**, *114*, 20809-20812.
508. Antonello, S.; Formaggio, F.; Moretto, A.; Toniolo, C.; Maran, F. Anomalous distance dependence of electron transfer across peptide bridges. *J. Am. Chem. Soc.* **2003**, *125*, 2874-2875.
509. Polo, F.; Antonello, S.; Formaggio, F.; Toniolo, C.; Maran, F. Evidence against the hopping mechanism as an important electron transfer pathway for conformationally constrained oligopeptides. *J. Am. Chem. Soc.* **2005**, *127*, 492-493.
510. Fick, R. J.; Kroner, G. M.; Nepal, B.; Magnani, R.; Horowitz, S.; Houtz, R. L.; Scheiner, S.; Trievel, R. C. Sulfur-oxygen chalcogen bonding mediates AdoMet recognition in the lysine methyltransferase SET7/9. *ACS Chem Biol* **2016**, DOI: 10.1021/acscchembio.5b00852
511. Rosenfield Jr., R. E.; Parthasarathy, R.; Dunitz, J. D. Directional preferences of nonbonded atomic contacts with divalent sulfur. 1. Electrophiles and nucleophiles. *J. Am. Chem. Soc.* **1977**, *99*, 4860-4862.
512. Row, T. N. G.; Parthasarathy, R. Directional preferences of nonbonded atomic contacts with divalent sulfur in terms of its orbital orientations. 2. S-S interactions and nonspherical shape of sulfur in crystals. *J. Am. Chem. Soc.* **1981**, *103*, 477-479.

513. Yang, D.; Konrat, R.; Kay, L. E. A multidimensional NMR experiment for measurement of the protein dihedral angle ψ based on cross-correlated relaxation between $^1\text{H}^\alpha$ – $^{13}\text{C}^\alpha$ dipolar and $^{13}\text{C}'$ (carbonyl) chemical shift anisotropy mechanisms. *J. Am. Chem. Soc.* **1997**, *119*, 11938-11940.
514. Tjandra, N.; Bax, A. Large variations in $^{13}\text{C}^\alpha$ chemical shift anisotropy in proteins correlate with secondary structure. *J. Am. Chem. Soc.* **1997**, *119*, 9576-9577.
515. Skrynnikov, N.; Konrat, R.; Muhundiram, D. R.; Kay, L. E. Relative orientation of peptide planes in proteins is selected in carbonyl–carbonyl chemical shift anisotropy cross-correlation spin relaxation. *J. Am. Chem. Soc.* **2000**, *122*, 7059-7071.
516. Wei, Y.; Lee, D.-K.; Ramamoorthy, A. Solid-state ^{13}C NMR chemical shift anisotropy tensors of polypeptides. *J. Am. Chem. Soc.* **2001**, *123*, 6118-6126.
517. Blanco, F. J.; Tycko, R. Determination of polypeptide backbone dihedral angles in solid state NMR by double quantum ^{13}C chemical shift anisotropy measurements. *J. Magn. Reson.* **2001**, *149*, 131-138.
518. Loth, K.; Pelupessy, P.; Bodenhausen, G. Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlation relaxation in NMR spectroscopy. *J. Am. Chem. Soc.* **2005**, *127*, 6062-6068.
519. Hendlich, M.; Bergner, A.; Günther, J.; Klebe, G. Relibase: Design and development of a database for comprehensive analysis of protein–ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607-620.