


7-6-2018

# NMR Structural Investigation of Chlamydial Protein Complex and Lysine Glycomimetic for Drug Design

Abigael Chebichiy Songok

Louisiana State University and Agricultural and Mechanical College, asongo1@lsu.edu

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_dissertations](https://digitalcommons.lsu.edu/gradschool_dissertations)

 Part of the [Analytical Chemistry Commons](#), [Medicinal-Pharmaceutical Chemistry Commons](#), and the [Organic Chemistry Commons](#)

---

## Recommended Citation

Songok, Abigael Chebichiy, "NMR Structural Investigation of Chlamydial Protein Complex and Lysine Glycomimetic for Drug Design" (2018). *LSU Doctoral Dissertations*. 4677.

[https://digitalcommons.lsu.edu/gradschool\\_dissertations/4677](https://digitalcommons.lsu.edu/gradschool_dissertations/4677)

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

NMR STRUCTURAL INVESTIGATION OF CHLAMYDIAL PROTEIN COMPLEX AND LYSINE  
GLYCOMIMETIC FOR DRUG DESIGN

A Dissertation

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

in

The Department of Chemistry

by  
Abigael Chebichiy Songok  
B.S., University of Nairobi, 2007  
August 2018

*To my mother Leah and my sister Deborah, you allowed me to travel to a foreign land in pursuit of my dreams, your love was felt across the oceans, your great achievements in life are a great inspiration to me. I love you two*

*~Chichi*

## ACKNOWLEDGMENTS

To God for his unrelenting love towards me.

To my advisor Dr. Macnaughtan, for your patience despite my unending weaknesses. Thank you for nurturing me to be a successful scientist, for allowing me try out new procedures which were met with successes and failures, but both of which taught me how to handle life challenges. For giving me opportunities to attend scientific conferences and for your concern over my life after graduate school. During your sabbatical you still committed to helping me in preparation for defense, your input did not go unnoticed.

To Dr. Taylor, I can't imagine the degree of commitment you have for the students who are directly under your mentorship. Considering your level of dedication to my work, I'd say you treated me as your student. I am sure I did not measure up your standard, but you chose to help me anyway. You stepped in as my advisor in the absence of Dr. Macnaughtan. You had the kind of patience with me. I admire your perfection, organization, and thoroughness. You demonstrated all these not only in helping me through research and writing, but in your research lab and administrative duties.

To my committee members Dr. Murray and Dr. Ding, for your guidance and for setting aside time to attend my general exam and final exam.

To Macnaughtan research group members, current and past, who helped me transition smoothly in and out of graduate school. To Thilini, Justin, and Kendra for allowing me to be your 'boss'.

To Dr. Thomas and Dr. Zhang at the LSU NMR facility, Dr. Tonelli and Dr. Dashti at NMRFAM, and Connie David at LSU MS facility for your help in NMR and MS data recording and processing.

To Dr. Doerrler at LSU life science, Dr. Stephens and Dr. Richard at Pennington Biomedical Research Center, together with your students Jasmine and Pradip, for helping in the antimicrobial and anti-inflammatory tests.

To the LSU Chemistry department for accepting my application to graduate school and for providing financial support throughout my study.

To the Baton Rouge SDA church family and ACF@LSU for your mentorship and care. Thanks to Sang, Anne, Sitienei, Kikway, Andre, Mathaga, and Nerissa for all the rides and support outside school.

To my family, immediate and extended, thanks for checking on me frequently despite my limited communication

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
ABSTRACT .....	v
CHAPTER 1. INTRODUCTION: PRODUCTION OF PROTEINS AND PEPTIDES FOR DRUG DISCOVERY .....	1
CHAPTER 2. EXPRESSION AND NMR ANALYSIS OF <i>CHLAMYDIA</i> PROTEINS SCC1/SCC4.....	11
CHAPTER 3. STRUCTURAL MODIFICATION OF THE TRIPEPTIDE KPV BY REDUCTIVE GLYCOALKYLATION.....	50
CHAPTER 4. BIOLOGICAL ACTIVITY AND ENZYMATIC STABILITY TESTS OF KPV TRIPEPTIDE DERIVATIVES.....	73
CHAPTER 5. CONCLUSIONS .....	86
REFERENCES.....	87
APPENDIX A. SUPPLEMENTARY ON PROTEIN EXPRESSION AND NMR ANALYSIS .....	97
APPENDIX B. NMR SPECTRA FOR THE SYNTHESIZED COMPOUNDS .....	103
VITA .....	135

## ABSTRACT

Research and design of drugs for treatment against microbial infections require the study of pathogenic proteins involved during infection and replication. Drugs can be designed to interfere with the interaction network of these pathogenic proteins and inhibit the infection process. Determination of the structure and ligands of the target proteins in microbes is essential for designing mechanistic-based drugs. Two *Chlamydia* proteins involved in host cell invasion were investigated in this study. The goal is to identify amino acids involved at the binding interface of *Chlamydia* chaperon proteins Scc1 and Scc4 by solution NMR spectroscopy. This information will give a lead in designing drugs that will interfere with Scc1:Scc4 binding, thus preventing *Chlamydia* invasion. Two new purification methods for isolating untagged Scc4 and the formation of mixed heterodimer complexes for NMR studies are described. The experimental NMR spectroscopy data was recorded and used to analyze the structural conformations of Scc4 when transitioning from homodimer to the heterodimer complex. Research presented in the second section is a novel structural modification method for therapeutic peptides through glycoalkylation. The modification is intended to improve the pharmacokinetic properties of lysine-containing peptide drugs, through half-life elongation and increased oral bioavailability. Specifically, the therapeutic tripeptide, Ac-KPV-NH<sub>2</sub>, was modified with a sugar derivative. Although the effect of the modification could not be assessed through antimicrobial and anti-inflammatory tests, the glycoalkylated peptides are shown to be stable against pronase cleavage. The unmodified tripeptide was cleaved by pronase in less than 24 hours, indicating that the glycoalkylation confers proteolytic stability.

## CHAPTER 1

### INTRODUCTION: PRODUCTION OF PROTEINS AND PEPTIDES FOR DRUG DISCOVERY

#### 1.1 Proteins and peptides

Proteins and peptides are chains of amino acids linked by peptide bonds to form the primary structure.<sup>1</sup> Secondary structures are formed by local folding aided by hydrogen bonds, leading to the formation of  $\alpha$ -helices or  $\beta$ -sheets, depending on the amino acid sequence and side chain interactions.<sup>1</sup> Other sequences may form loops or coils. The regions of the secondary structure fold into tertiary structure giving the three-dimensional arrangement of the protein or peptide.<sup>2</sup>

Approximately 20,000 unmodified proteins are encoded in the human genome. Additionally, about 100 variant protein structures can be derived from a single gene through processes such as post translational modification, alternative splicing, and single amino acid polymorphism.<sup>3</sup> These proteins play a central role in biological systems through their interaction with each other or with molecules such as DNA, or RNA.<sup>4</sup> Improper folding and mutation of protein sequences interfere with their interaction network, leading to various diseases and disorders.<sup>4</sup>

Mechanism-based drugs that target proteins in biological systems are designed based on the molecular understanding of the protein structure and function. Protein targets are derived from the human genome<sup>5</sup> or from pathogens that cause infections.<sup>6</sup> Pathogenic proteins are transferred to the host cell by bacteria or viruses, through various mechanisms, to interfere with the host's cellular activities and allow the pathogen to thrive. To design drugs that target pathogenic proteins, there is a need for structural knowledge of the proteins involved and how they are transferred to the host cell. Bacterial systems, such as gram-negative bacteria, have complex microstructures used for host-pathogen interaction. These dedicated secretion systems form  $\beta$ -barrel channels classified into six types, called type I secretion system (T1SS) to type VI secretion system (T6SS), depending on the class of proteins being secreted.<sup>7</sup> Therapeutic approaches that interfere with these systems will inhibit microbial host cell invasion.

In addition to being used as drug targets, proteins and peptides have been utilized as therapeutic agents. A large number of bioactive peptides with varying activities have been isolated in animals.<sup>8</sup> These peptides are derived from pro-peptides (longer protein sequences), which are activated during post translational modification by enzymes such as prohormone convertases.<sup>9</sup> Protein or peptide drugs with

longer sequences are synthesized through recombinant methods,<sup>10</sup> while short peptides can be chemically synthesized. Biological activity of peptide motifs in large multifunctional proteins can be mimicked using short peptides bearing the same amino acid sequence as the peptide motif.<sup>11</sup> In most cases, specificity is enhanced and production costs are lower when short peptide sequences are used in place of the full-length proteins.

## 1.2 Therapeutic peptides

There are over 1000 known signaling peptides in animals that have been classified into different groups depending on their function. These classes include antimicrobial peptides, toxin and venom peptides, cytokines and growth factors, neuropeptides, peptide hormones, anti-inflammatory, and antifreeze peptides.<sup>8</sup> Studies on the therapeutic activities of both endogenous and exogenous peptides have shown that some peptides are multifunctional. Alpha melanocyte stimulating hormone ( $\alpha$ -MSH), for example, is an endogenous peptide with melanogenic effects, immunomodulatory effects, and antimicrobial activity.<sup>12</sup> Other examples of endogenous peptides are neurokinins, insulin, gastrin, glucagon, adrenocorticotrophic hormone (ACTH).<sup>13</sup> Exogenous bioactive peptides are largely derived from food.<sup>14</sup> Food proteins encrypt bioactive peptide sequences that are activated by proteolytic enzymes during fermentation, hydrolysis or digestion.<sup>15</sup> Food-derived peptides are essential for different physiological activities in the immune, nervous, gastrointestinal, and cardiovascular systems.<sup>15, 16</sup> Examples of short bioactive peptides are listed in Table 1.

Peptide-based therapeutics are produced by chemical synthesis, recombinant methods, transgenic systems, or isolated from natural sources. They have been used in the treatment of various diseases, including multiple sclerosis, cancer, Alzheimer's disease, congestive heart failure, deep vein thrombosis, lupus, ulcerative colitis, rheumatoid arthritis, and hepatitis C.<sup>17</sup> In terms of molecular size, peptide drugs fill the gap between small molecule drugs and protein drugs. Peptide drugs combine the positive features afforded by the small molecule drugs, such as homogeneity and ease in cellular uptake, with those of biologics, such as high target specificity and minimal side effects.<sup>18</sup> Short peptide sequences can be synthesized in solution or through solid phase peptide synthesis. Chemical synthesis enables structural manipulation to protect peptide drugs from enzymatic recognition and degradation. These modifications can be introduced during the stepwise synthesis with the use of modified building blocks. In contrast to the



structural modification of biologics in cellular systems, it is possible to control the extent and site of modifications using protecting groups.

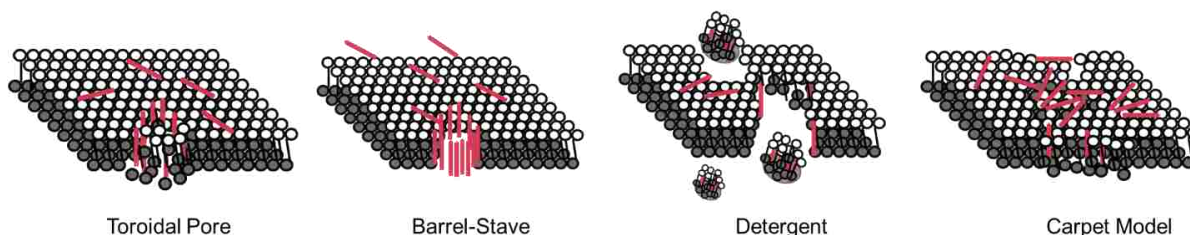
**Table 1.** Examples of bioactive peptides.<sup>15, 19</sup>

Peptide	Activity	Function
IP, MP, VP, LP	Anti-oxidants	Scavenge hydroxyl radicals
PTGADY	Immunomodulatory	Increases production of IL-2, IL-4, IL-6
ADVFNPR, VVLYK	Antihypertensive	Reduces endothelin-1 levels
LGTAVFK	Antimicrobial	Causes membrane destruction in bacteria and yeast
GPAE, GPGA	Anti-type 2 diabetes mellitus	Inhibits dipeptidyl peptidase-IV
FIMGPY	Anticancer	Antiproliferation activity by inducing apoptosis

### 1.3 Antimicrobial peptides

Cationic and amphiphilic properties of antimicrobial peptides (AMPs) allow their interaction with negatively charged bacterial membrane, distinguishable from positively charged mammalian membranes. The presence of lipopolysaccharide, anionic phospholipids, lipoteichoic acid, and teichoic acid in the bacterial membranes enable electrostatic interaction with cationic peptides.<sup>20</sup> At the water-membrane interface, antimicrobial peptides form an amphipathic structure segregating cationic (hydrophilic) residues from hydrophobic residues. This structure enables electrostatic interaction of the cationic residues with the negatively charged membrane.<sup>21</sup> The hydrophobic domain can interact with the fatty acyl chains in the bilayer membrane, thus creating pores that may lead to cell death through the loss of membrane potential or ions. Apart from disrupting microbial cytoplasmic membranes, other mechanisms have been identified that target cellular processes such as enzymatic activity, DNA and protein synthesis, protein folding, and synthesis of the cell wall.<sup>22</sup> Four main modes of actions have been proposed for both membrane-active and intracellular active AMPs. These are barrel-stave, toroidal, detergent, and carpet modes,<sup>23</sup> represented in Figure 1. In the barrel stave mode, the peptides form a transmembrane pore with the hydrophobic groups facing outward toward the membrane acyl chains while the hydrophilic groups face inward lining the pore.<sup>24</sup> The toroidal pore mode is like the barrel stave mode except that the peptides are intercalated with the membrane lipids to line the pore. In the carpet mode, the peptides accumulate on the membrane surface

through electrostatic interaction leading to membrane thinning. The detergent mode is characterized by initial accumulation of peptides at the membrane surface followed by lipid/peptide micellization.<sup>24, 25</sup> Examples of peptides utilizing these modes of action are cecropin (detergent), alamethicin (barrel-stave), magainin 2 (toroidal), and LL-37 (carpet).<sup>25</sup>

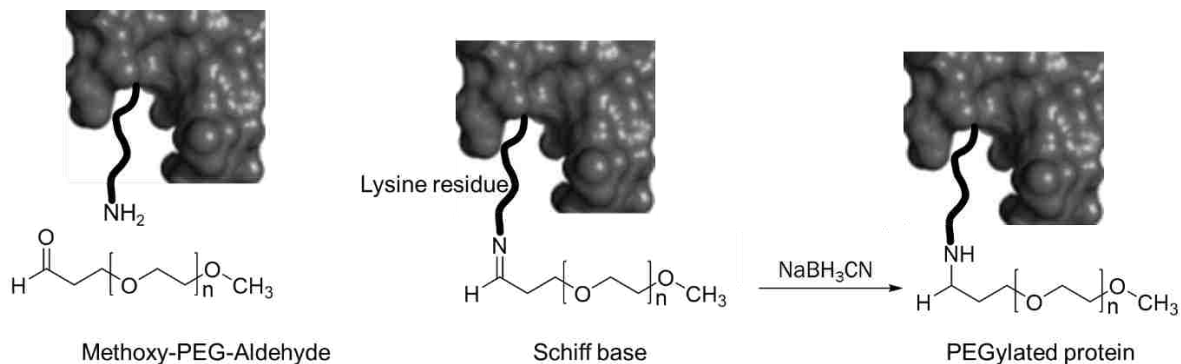


**Figure 1.** A pictorial representation of the different modes of action of AMPs adapted from Zhang *et al.*<sup>25</sup>

#### 1.4 Structural modification of therapeutic peptides

Some of the challenges that limit full exploitation of therapeutic peptides are low oral bioavailability and short half-life due to renal filtration or proteolytic degradation.<sup>26</sup> Proteases in bodily fluids, tissues, and cells easily degrade recognizable peptide sequences.<sup>27</sup> To overcome renal filtration and degradation, extensive modification of bioactive peptides, by incorporating unnatural amino acid(s), partial structural modifications, or covalent attachment with polymers, has been implemented. Peptide drugs have been coupled to polymers of sialic acid (polysialylation), ethylene glycol (PEGylation), and the carbohydrate polymer hydroxyethyl starch (HESylation).<sup>28</sup> The attachment of such polymers increases the hydrodynamic volume of the drug, which places it above the renal elimination cut-off, thereby allowing longer circulation times. In addition, some polymers are highly hydrophilic in nature, increasing the solubility of the drug. The bulky modifications sterically hinder proteolytic degradation.

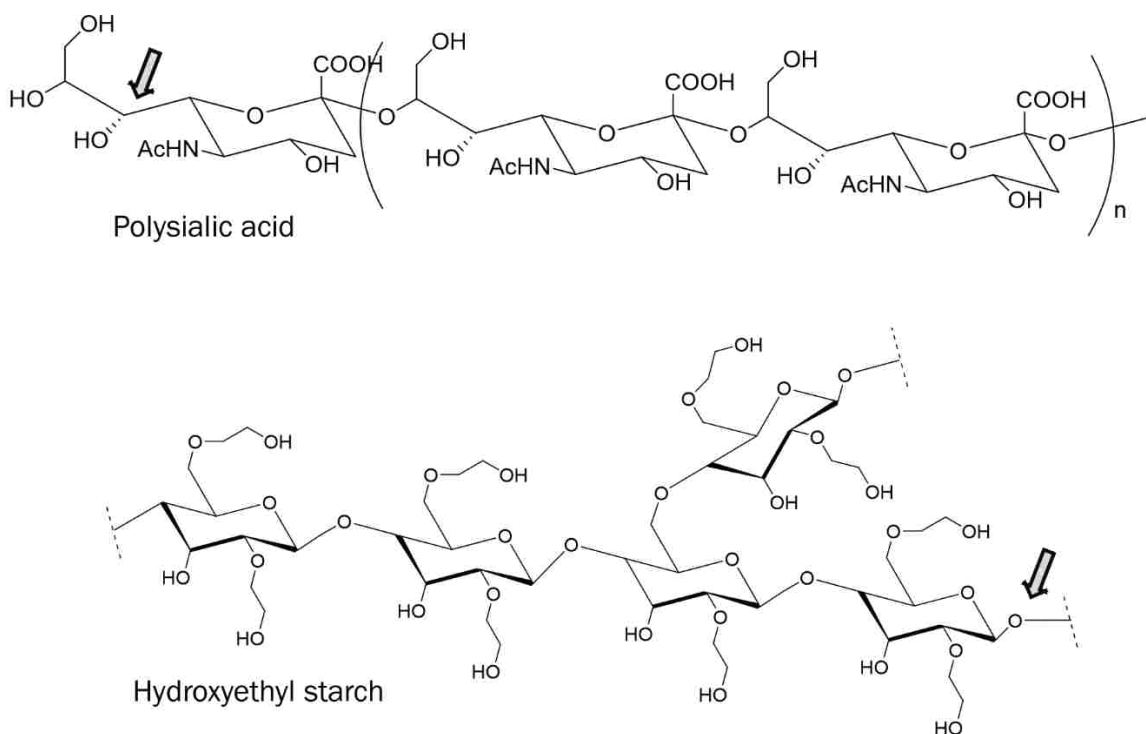
PEGylation is the covalent attachment of polyethylene glycol (PEG) to protein or peptide drugs. Conjugation processes depend on the PEG, which may be an activated ester that can be reacted with an amine in the protein or PEG aldehyde or glutaraldehyde coupled to amines through reductive alkylation. The aldehyde-amine reaction forms an intermediate Schiff base, which is reduced by  $\text{NaBH}_3\text{CN}$  (Figure 2).<sup>29, 30</sup>



**Figure 2.** Alkylating PEGylation adapted from Moosmann *et al.*<sup>30</sup>

Polysialylation is the coupling of polysialic acid to therapeutics. Unlike PEG polymers, polysialic acid is a naturally occurring biodegradable polymer that is as effective as PEG but shows minimal toxicity. The covalent attachment of this polymer to biopharmaceuticals is achieved through enzymatic conjugation methods or synthetic steps via carboxylate groups, hydroxyl groups, or cleavage of the 1,2-diols at the non-reducing end.<sup>28</sup> The 1,2 diols of polysialic acid can be cleaved by  $\text{NaIO}_4$  to give an aldehyde functional group at the position indicated by the arrow in Figure 3. The aldehyde group is then coupled to an amine group of the peptide drug. Polysialylation has been effectively used in half-life elongation and immunogenic reduction of insulin and asparaginase drugs.<sup>31, 32, 33</sup>

HESylation links the drug of interest to hydroxyethyl starch (HES) to increase solubility, hydrodynamic volume, and prevent enzymatic degradation. HES is a biodegradable polymer, a modified form of amylopectin, which may be used as a plasma volume expander (in place of albumin), since it is not susceptible to hydrolysis by plasma amylases as is the case with other starches.<sup>28, 34</sup> Conjugation to other molecules has been achieved using the enzyme, transglutaminase.<sup>33</sup> For site specific coupling, the acetal group at the reducing end, indicated by the arrow in Figure 3, is used to couple to the peptide drug. Other modifications include recombinant extension by the addition of a polypeptide sequence at the *N*- or *C*-terminus of a recombinant drug sequence.<sup>35</sup> PASylation is an example of an extension method that uses proline, alanine, and serine in the extended chain. Significance of some modifications are highlighted in Table 2.

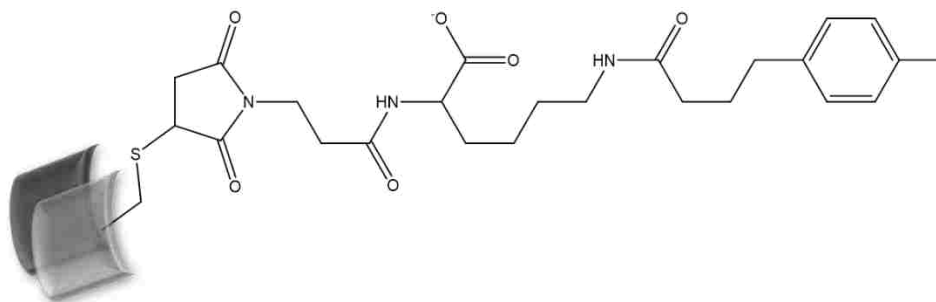


**Figure 3.** Polysialic acid and hydroxyethyl starch structures with arrows indicating the potential site of conjugation with peptide drugs.

In addition to polymers, several small organic molecules have been covalently bound to therapeutics peptides. The modifications include *N*-terminal acetylation or alkylation, glycosylation, phosphorylation, thiolation, and fatty acid conjugation.<sup>36</sup> Replacement of L-amino acids with D-amino acids has been successful in the case of vasopressin and desmopressin. Vasopressin is a naturally occurring antidiuretic hormone (Cys<sup>1</sup>-Tyr<sup>2</sup>-Phe<sup>3</sup>-Gln<sup>4</sup>-Asn<sup>5</sup>-Cys<sup>6</sup>-Pro<sup>7</sup>-L-Arg<sup>8</sup>-Gly<sup>9</sup>-NH<sub>2</sub>), the analogue desmopressin \*SH<sub>2</sub>C-H<sub>2</sub>C-C<sup>1</sup>-Tyr<sup>2</sup>-Phe<sup>3</sup>-Gln<sup>4</sup>-Asn<sup>5</sup>-Cys<sup>6</sup>\*-Pro<sup>7</sup>-D-Arg<sup>8</sup>-Gly<sup>9</sup>-NH<sub>2</sub>, is a non-peptide drug in which the L-arginine at position 8 has been replaced with a D-arginine and the cysteine 1 is removed (a disulfide bridge between the thiols is identified with asterisks).<sup>37,38</sup> Conjugation of 2-(3-maleimidopropanamido)-6-(4-(4-iodophenyl)butanamido) hexanoate (“Albu” tag, Figure 4) to an antibody fragment, scFv, reported by Trussel *et al.*, increased its serum-drug affinity, hence elongating the time in blood circulation.<sup>39</sup> Hydrocarbon stapling is used to hold together the alpha helical structures in peptide drugs and prevent unfolding upon exposure to harsh environmental conditions.<sup>40</sup> The amino acids within a helix structure are coupled by a hydrocarbon linker to make the structure more rigid.

**Table 2.** Effects of different structural modifications on drugs.<sup>36, 41</sup>

Method	Significance
<b>Post translational modifications</b> Cyclization, Phosphorylation, N-alkylation, Thiolation	Enhanced solubility
<b>N- or C- terminus modification</b> PEG conjugation, methylation, formylation, acetylation, esterification	Half-life elongation Increase functional diversity Enhanced solubility
<b>Conjugation with fatty acids</b> Reversible and irreversible lipidation	Improved lipophilicity, metabolic stability, and bioavailability
<b>Hydrophobic ion pairing</b>	Enhance permeation, increased bioavailability and solubility
<b>Complexation with cyclodextrin</b>	Protection against degradation and suppression of aggregation



**Figure 4.** “Albu” tag 2-(3-maleimidopropanamido)-6-(4-(4-iodophenyl)butanamido) hexanoate link to a cysteine residue of an antibody fragment adapted from Trussel *et al.*<sup>39</sup>

No modification method is universally applicable to all drugs. Some modifications are cause adverse side effects due to induced immunogenicity. Therefore, research is ongoing to define simple modification methods that will enhance the drug’s pharmacokinetics and pharmacodynamics without drawbacks.

## 1.5 Recombinant protein expression

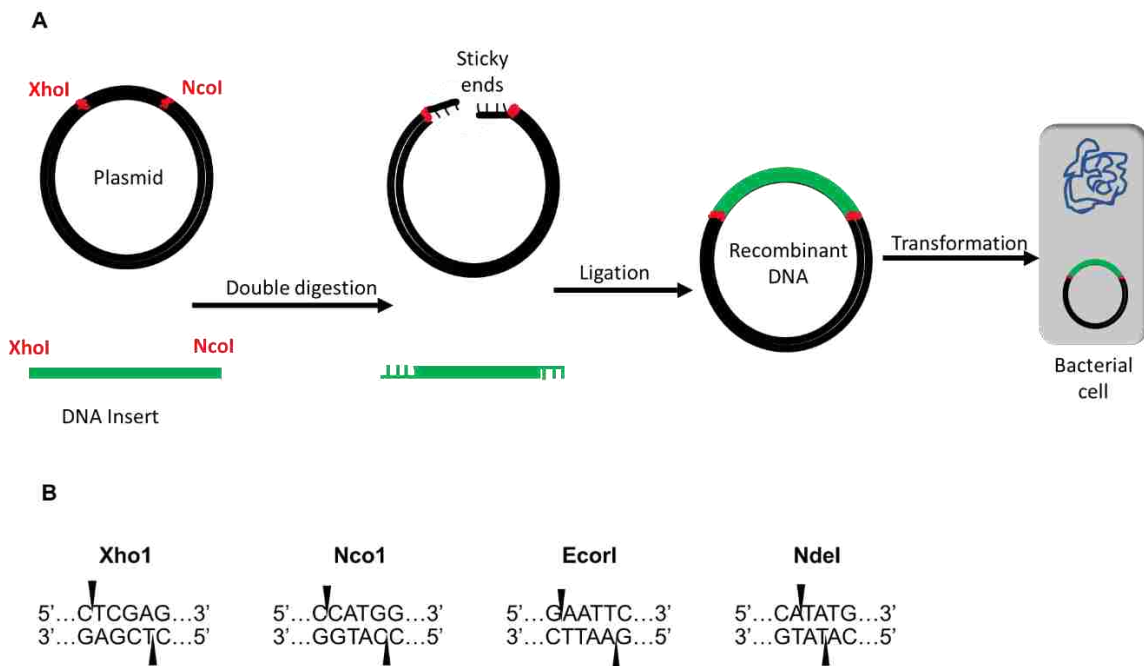
The recent advancement in recombinant DNA technology<sup>42</sup> has lowered the production costs and increased the annual sales of biopharmaceuticals. Biopharmaceuticals are therapeutic proteins and nucleic acids produced using biotechnology.<sup>43</sup> The market for biopharmaceuticals was valued at USD 176.9 billion globally in 2015 and is estimated to reach USD 291 billion in 2021, with a compound annual growth rate of 8.6%.<sup>44</sup> Progressively, there has been a notable increase in the annual approval rates for biologics and peptide drugs over small molecule drugs. Compared to small molecule drugs, biologics and peptides typically have high target specificity, thus causing minimal side effects and low immunogenicity. Additionally, their structural similarity to existing biological molecules shortens their FDA approval time, which makes them more attractive for production and marketing purposes.<sup>45</sup> The biological systems used in the production of protein therapeutics depend on the recombinant protein size and the desired post-translational modifications. These systems include bacteria, yeast, fungi, insects, mammals, and transgenic animals and plants.<sup>46, 47</sup>

Recombinant DNA technology<sup>10</sup> involves heterologous expression of proteins in non-native systems.<sup>48</sup> The cells are transfected with self-replicating DNA molecules encoding the protein of interest. The cloned vector (cyclic DNA) allows expression of proteins in high yields for different studies.<sup>49</sup> Besides expressing proteins from the human genome, proteins in other organisms such as fungi, bacteria, or viruses have been expressed and investigated as disease-causing agents or having therapeutic effects. Additionally, bacterial systems allow expression of isotopically labeled proteins that can be analyzed using NMR spectroscopy.

Bacterial plasmids used for making recombinant DNA carry a specific antibiotic resistance gene for selecting transformed cells. Gene fragments from a donor organism are inserted into these plasmids (vectors) to form a recombinant DNA.<sup>49</sup> Other vector systems used for gene transfer are vaccinia, baculovirus, adenovirus, and retroviral vectors.<sup>50</sup> Plasmids with specific cleavage sites are commercially available. The DNA inserts are constructed to have the same cleavage sites as the plasmid, which allows insertion of the gene of interest into the plasmid DNA sequence through double digestion (by restriction enzymes) and ligation. The restriction enzymes cut specific DNA sequences in both the plasmid and the

insert. Examples of commercially available restriction enzymes are XhoI, NcoI, EcoRI, and NdeI. The sticky ends (complementary bases) of the cut DNA are joined together by a ligase (cloning steps are illustrated in Figure 5) to form a cyclic DNA that is incorporated into biological protein expression systems.

Before selecting an expression system, one must consider several factors such as yield, post translational modification, solubility, cost, purification, toxicity, and function.<sup>50</sup> The biological systems used in the production of protein depend on the recombinant protein size and the desired post-translational modifications. These systems include bacteria, yeast, fungi, insects, mammals, transgenic animals, and plants.<sup>46, 47</sup> Bacterial systems (mainly *Escherichia coli*) dominate heterologous expression of recombinant proteins. However, post translational modification required for certain protein sequences may not be achieved in this system, thus limiting the nature of proteins that can be expressed.



**Figure 5.** Illustration of A) cloning steps and B) cleavage sites for some restriction enzymes.

Mammalian cells induce proper folding and glycosylation of eukaryotic proteins. Their major application is in the production of therapeutic proteins, vaccines, and diagnostics.<sup>50</sup> Commonly used cell lines are human embryonic kidney (HEK) and Chinese hamster ovary (CHO) cells. Many industrial enzymes are produced in yeast or fungi.<sup>51</sup> These systems produce high yields of recombinant proteins at low costs. Yeast has been engineered to produce glycoproteins for human use by eliminating the production of high-

mannose proteins (naturally produced by yeast), which induce immunogenicity and are easily degraded in humans.<sup>51</sup>

Baculovirus-insect cell systems are also efficient in producing eukaryotic proteins.<sup>52</sup> Essential glycosylation patterns such as the incorporation of terminal sialic acids in therapeutic glycoproteins have been achieved using transgenic insect cells.<sup>53</sup> The *Sf9* transgenic insect cells used in production of terminal sialylated *N*-glycans encode rat  $\alpha$ 2,6-sialyltransferase and bovine  $\beta$ 1,4-galactosyltransferase.<sup>53</sup> The use of plant cell expression systems is slowly rising in the pharmaceutical industry. The advantages of using this cell line are the notably high batch-to-batch reproducibility and the elimination of pathogen contamination that can incur when mammalian cells are used.<sup>54</sup>



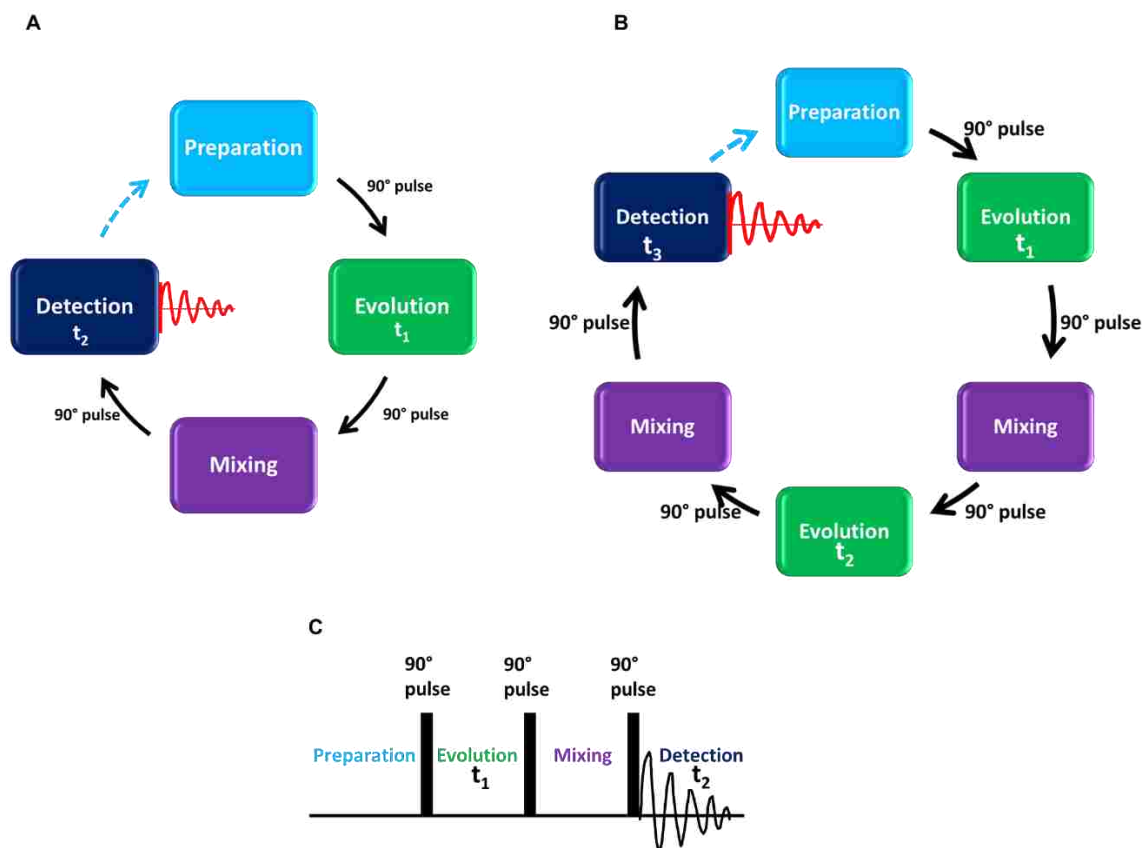
## CHAPTER 2 EXPRESSION AND NMR ANALYSIS OF *CHLAMYDIA* PROTEINS SCC1/SCC4

### 2.1 Protein NMR

Nuclear magnetic resonance occurs when nuclei are placed in a static magnetic field ( $B_0$ ), and a second oscillating magnetic field (radio frequency pulse perpendicular to  $B_0$ )<sup>55</sup> is applied to cause the nuclei to transition between energy levels. A fundamental property known as spin generates nuclear magnetic moments in some nuclei.<sup>56</sup> A spin can be zero, positive, or negative and exist in multiples of  $\frac{1}{2}$ .<sup>57</sup> Each unpaired electron, neutron, or proton has spin  $\frac{1}{2}$ . When placed in a magnetic field, the torque generated causes the magnetic moment to rotate (precess) around the magnetic field at a characteristic frequency of the nuclei known as the Larmor frequency.<sup>57</sup> The Larmor frequencies of nuclei detected in protein NMR, such as  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ , and  $^{31}\text{P}$ , differ because of an intrinsic property known as the gyromagnetic ratio,<sup>58</sup> which defines the sensitivity of different nuclei to the external magnetic field. Protein NMR experiments are used to determine protein structure and molecular dynamics by studying the interaction of nuclei.<sup>58</sup> When the spin  $\frac{1}{2}$  nuclei are placed in a strong magnetic field, two energy states are generated. The lower energy state spins (magnetic quantum number  $+\frac{1}{2}$ ) align in the direction of the external magnetic field ( $B_0$ ), while the higher energy state spins (magnetic quantum number  $-\frac{1}{2}$ ) align in the opposing direction of  $B_0$ .<sup>59</sup> The energy gap,  $\Delta E$ , between the two energy states is proportional to the strength of  $B_0$ . Irradiation of the nuclei with a radio frequency pulse (short, discrete energy equal to  $\Delta E$ ) causes the spins to align in one direction (coherence) and transition between the energy states.<sup>60</sup> Following excitation, two relaxation processes occur: Longitudinal relaxation, which is the realignment of the bulk magnetization (net magnetic moment) in the  $B_0$  direction, and transverse relaxation, which is caused by the loss of spin coherence.<sup>57</sup> The loss of coherence gradually causes reduction of the detected signal and is known as the free induction decay (FID).<sup>61</sup> The Fourier transform of the FID converts the signal from the time domain to the frequency domain. This transformation gives information about the population of the nuclei, chemical environment of the nuclei, and protein dynamics represented by the peak intensity, chemical shift, and linewidth in the frequency spectrum, respectively.<sup>58</sup> The chemical shifts are influenced by hydrogen bonding, the local electron cloud, torsion angles, and the secondary structure of proteins.<sup>58</sup>

Four basic steps are taken during NMR data acquisition: Preparation, evolution, mixing, and detection.<sup>62</sup> Preparation is a relaxation time delay that recovers the longitudinal magnetization of spins. A

selective  $90^\circ$  radio frequency pulse directed to a specific nucleus (determined by the Larmor frequency) creates a transverse bulk magnetization of spins, which are detected in 1D experiments.<sup>63</sup> After pulsing the nuclei with radio frequency energy, they can precess at their characteristic frequencies defined by their chemical environment in a process known as evolution.<sup>59</sup> The evolution time is regulated such that when another pulse is induced, it can cause transfer of coherence from one nuclei to another (mixing period).<sup>60</sup> The transferred coherence is then detected in 2D experiments.<sup>64</sup> Unique pulse sequences have been created for different experiments with examples shown in Figure 6.



**Figure 6.** Illustrative pulse sequences for A) 2D NMR B) 3D NMR experiments, and C) conventional representation of 2D NMR pulse sequences.

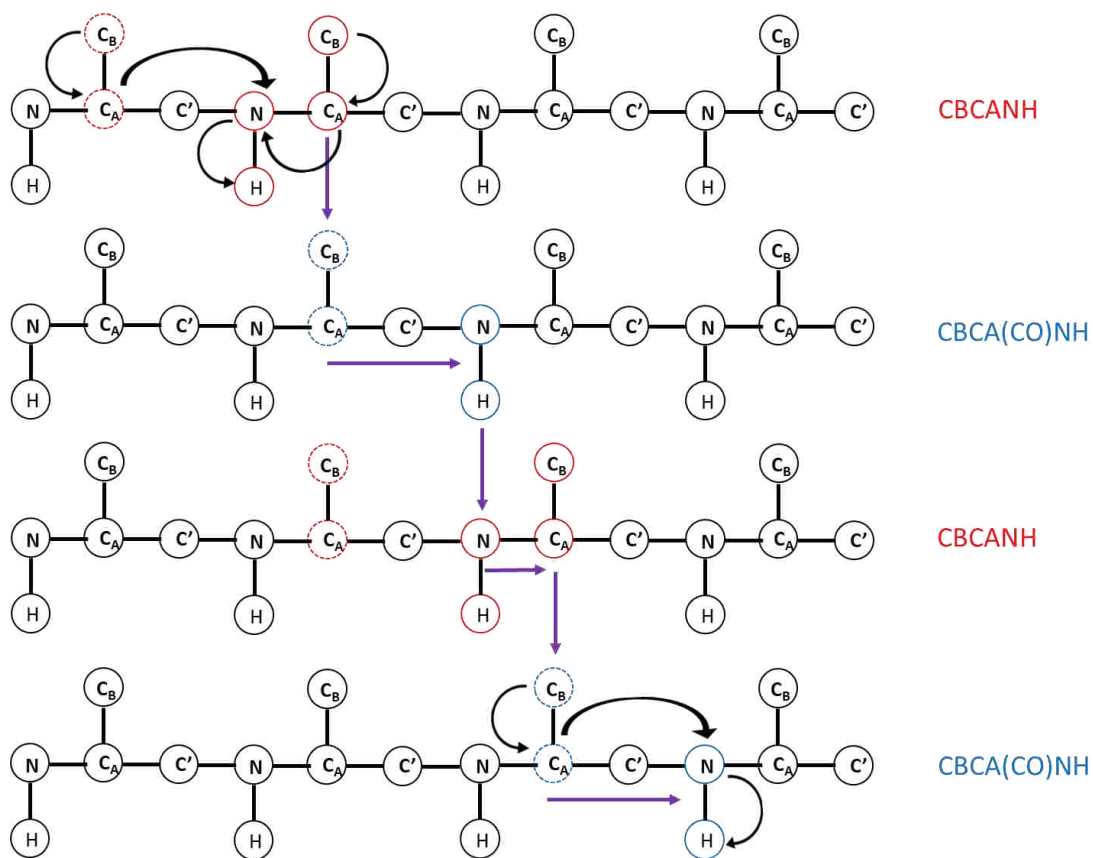
The pathway used for coherence transfer determines the kind of correlation observed. This transfer can occur along the covalent bond (through-bond) or between neighboring non-covalently attached nuclei (through-space).<sup>65</sup> Homonuclear through-bond or J-correlated spectroscopy includes COSY (correlation spectroscopy),<sup>66</sup> DQF-COSY (double quantum filtered COSY), and TOCSY (total correlation spectroscopy). In COSY experiments, magnetization is transferred to the nuclei within a 3-bond distance, while TOCSY

experiments allow successive transfer of magnetization throughout the amino acid spin system. The TOCSY spectrum is useful for protein side-chain assignments. Homonuclear through-space spectroscopy includes NOESY (nuclear Overhauser effect spectroscopy) and ROESY (rotating frame Overhauser effect spectroscopy). NOESY experiments correlate protons using the dipolar interaction of spins.<sup>67</sup> Correlation of spins within a 5 Å distance can be observed. The NOESY signal intensity is inversely proportional to the sixth power distance between the protons. This information is useful for the determination of protein tertiary structure. Heteronuclear J-correlated spectroscopy includes HMQC (heteronuclear multiple quantum coherence), HSQC (heteronuclear single quantum coherence), and refocus HSQC. These experiments correlate different nuclei such as <sup>13</sup>C-<sup>1</sup>H or <sup>1</sup>H-<sup>15</sup>N and are used in the protein backbone assignments.

The pulse sequences for 3D NMR experiments are built from 2D NMR pulse sequences with an additional coherence transfer step. The 3D experiments are done by insertion of another evolution-mixing period before detection. Triple resonance experiments are used for protein backbone and side-chain assignments, 3D structure determination, and ligand binding studies. The three nuclei observed are <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N. A set of NMR data used for backbone and side chain assignment are selected from the following experiments: N15-HSQC, C13-HSQC, HNCO, HCCH-TOCSY, CBCA(CO)NH, HNCACB, HN(CO)CA, CCH-TOCSY, HNCA, HNCB, CB(CO)NH, HCCH-COSY, HN(CA)CO, HN(CO)CACB, C(CO)NH, CCH-COSY, H(CCO)NH, HBHA(CO)NH, HA(CO)NH.<sup>68,69</sup> Proteins that have less than 200 amino acid residues typically have sufficient signal for C $\alpha$  nuclei and therefore CBCANH and CBCA(CO)NH spectra are used for backbone assignment (Figure 7). In large proteins; however, the C $\alpha$  signal intensity is low and HNCA/HN(CA)CO and HN(CO)CA/HNCO spectra are used (Figure 8).<sup>70</sup>

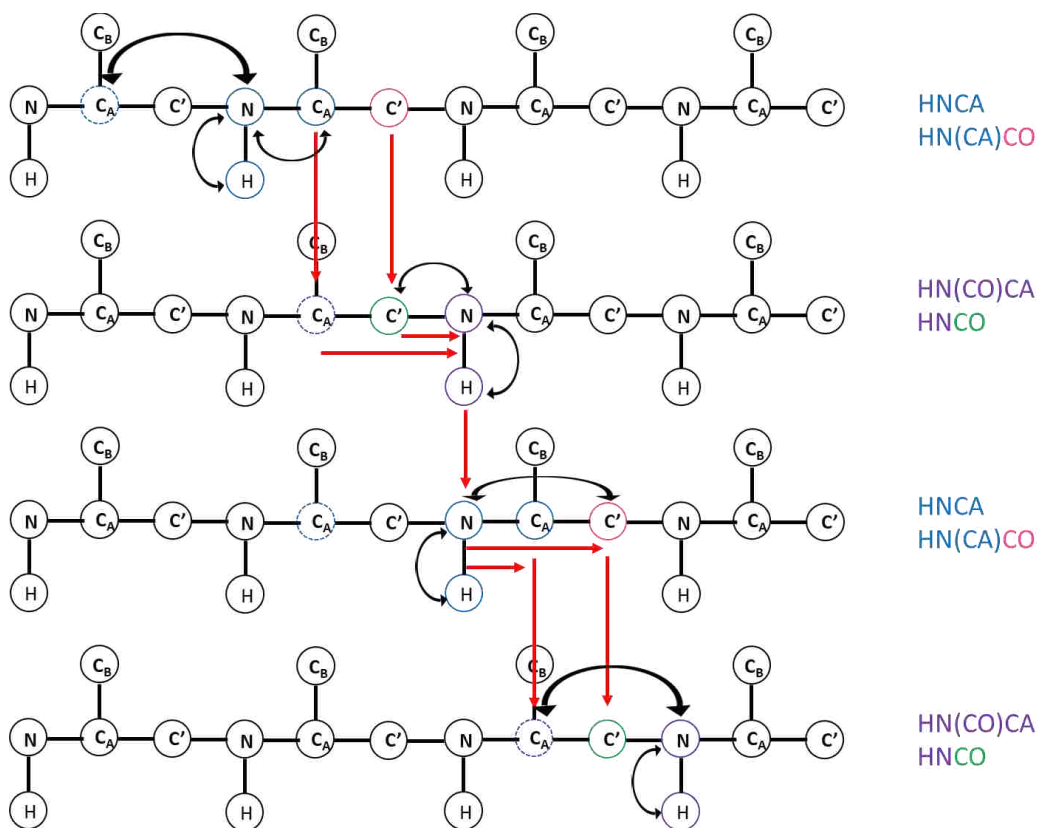
Protein-ligand interactions play a major role in signal transduction, as in the case of hormones, antigen-receptor, transcription-translation processes, and disease development. The ability of NMR to reveal these interactions at the atomic level is possible with experiments that monitor interactions by exchange of saturation transfer, changes in diffusion constants, changes in chemical shifts, relaxation time changes, and NOEs changes.<sup>71</sup> Two methods are mainly used to monitor these interactions. The first method monitors the chemical shift changes of the protein. The presence of a ligand causes environmental changes around the interface amino acids of the protein, which in turn alters their chemical shifts. Other amino acids further from the binding interface often do not experience any changes and their chemical-

shifts are unchanged. However, in cases where the binding process induces global structural changes, the chemical shift changes may occur in all signals, making it challenging to single out the changes at the binding surface. The second experimental approach uses transfer NOEs between the ligand and protein. This method gives information on the binding sites, conformation of ligands complexed with proteins, binding kinetics, and protein-ligand complex structure in solution.<sup>71, 72</sup>



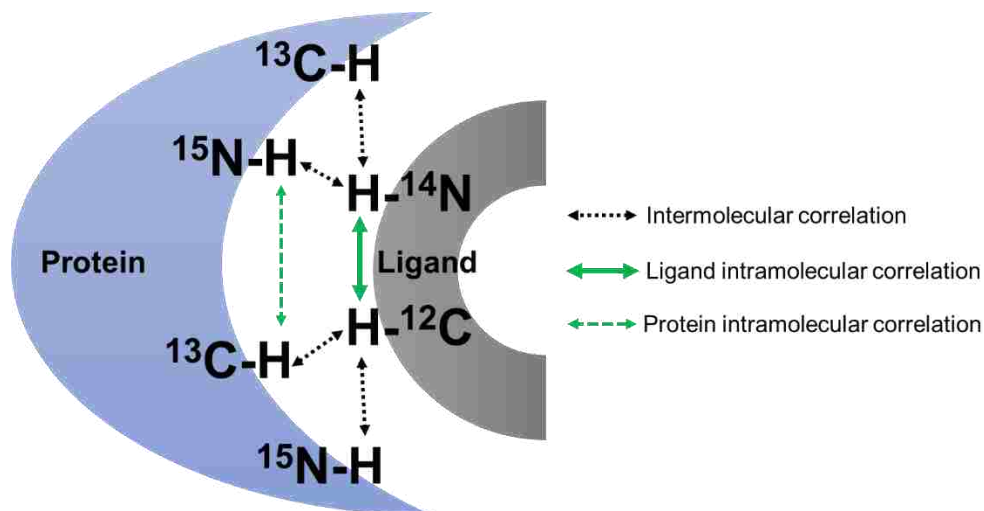
**Figure 7.** Backbone assignment using CBCANH and CBCA(CO)NH spectra.

It is possible to distinguish proton signals arising from different heavy atoms and their isotopes in protein-protein or protein-ligand complexes. To distinguish proton signals of ligands from those of protein, the protein is isotopically labeled with  $^{13}\text{C}/^{15}\text{N}$  (uniform or site specific). Intramolecular NOEs within the ligand are observed using isotope-filtered experiments while those of protein are observed using isotope-edited 2D experiments.<sup>72</sup> For intermolecular NOEs, one dimension of proton is isotope filtered while the other dimension is isotope edited, Figure 9 shows a schematic representation of protons in a protein-ligand system.

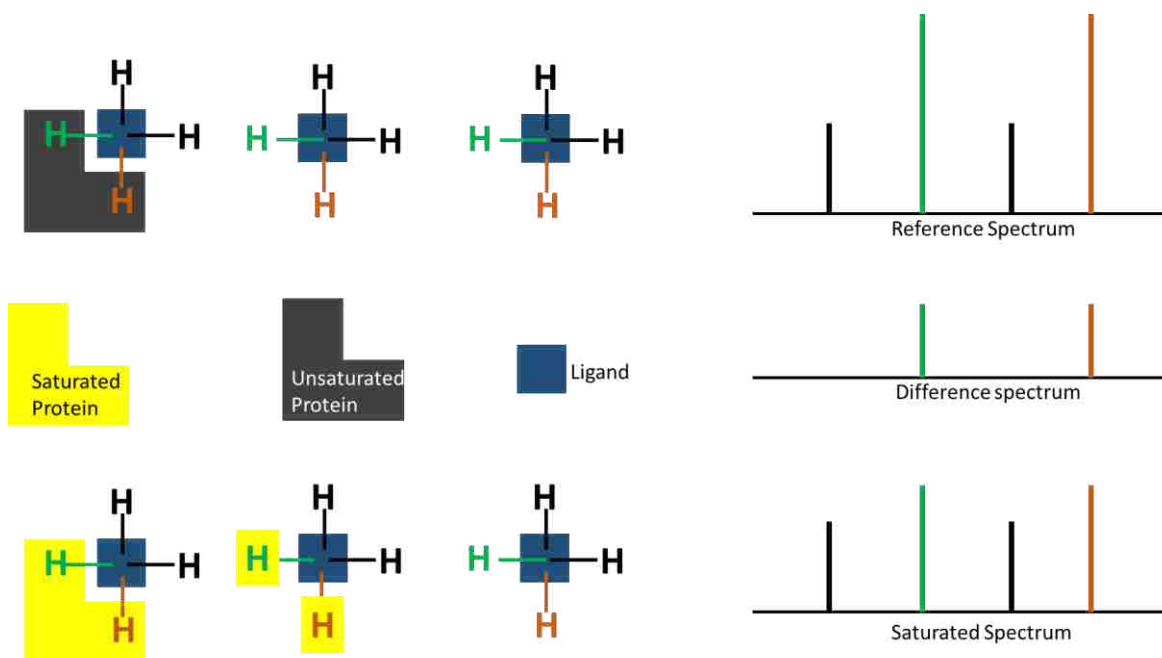


**Figure 8.** Backbone assignment using HNCA/HN(CA)CO and HN(CO)CA/HNCO.

Saturation transfer difference NMR has been used to screen for ligands (with medium to fast exchange rates), identify competing ligands, and determine the binding epitopes. Saturation transfer experiments selectively saturate the protein signals. This saturation is achieved by identifying and saturating protein resonance signals far from the ligand signal. The saturation will disseminate throughout the protein through spin diffusion and transfer to the ligand at the binding surface.<sup>73</sup> The signal intensity of the ligand, when the protein is saturated (saturated spectrum, Figure 10), will be lower than when it is not saturated (reference spectrum). The difference signal (difference spectrum) from the two conditions corresponds to the signal of the ligand protons responsible for binding. The STD NMR is applicable for high molecular weight complexes with a low molecular weight, medium-to-fast exchanging ligand. It has high sensitivity and requires minimal amounts of ligand; it can be coupled with other NMR pulse sequences giving rise to STD-HSQC, STD-HMQC, or STD-TOCSY.<sup>74</sup>



**Figure 9.** Illustration of intermolecular and intramolecular coherence transfer on a ligand-protein system adapted from Breeze *et al.*<sup>72</sup>



**Figure 10.** Schematic representation of signal observed in an STD experiment adapted from Yu *et al.*<sup>73</sup>

## 2.2 Chlamydia

*Chlamydia* is an intracellular bacterium that relies on the host cell for biosynthetic and metabolic pathways. Chlamydiae consists of three main species; namely, *Chlamydia trachomatis* (CT), *Chlamydia pneumoniae*, and *Chlamydia psittaci*.<sup>75</sup> *Chlamydia trachomatis* is the major cause of bacterial sexually transmitted infection worldwide. Untreated genital infections lead to proctitis, urethritis, and epididymitis in men, and tubal factor infertility, ectopic pregnancy, salpingitis, cervicitis, and pelvic inflammatory disease

in women.<sup>76,77</sup> Ocular infection causes inclusion conjunctivitis and trachoma.<sup>78</sup> *Chlamydia pneumoniae* and *Chlamydia psittaci* cause pneumonitis and psittacosis, respectively.<sup>79</sup> Serotypes A, B, and C of CT are the main causes of blindness, while serotype D-K cause both ocular and genital disease through sexual transmission.<sup>79</sup> The genetic manipulation of CT species has not been possible; hence, the underlying mechanism of CT infection and replication is not well known.

The developmental cycle of CT is divided into the early stage (elementary body (EB) to reticulate body (RB) differentiation), mid-stage (RB division and growth), and late stage (RB to EB differentiation).<sup>80</sup> When the RNA polymerase (RNAP) holoenzyme with the primary sigma factor, sigma 66 ( $\sigma^{66}$ ), interacts with the specific *Chlamydia* chaperon 4 (Scc4) in CT, the transcription process is inhibited.<sup>80</sup> Scc4 regulates transcription as a component of the RNAP holoenzyme, and the  $\sigma^{66}$  RNAP holoenzyme is reported to accumulate at the beginning of the RB to EB transition in the developmental cycle.<sup>80</sup> RNAP has four subunits viz.  $\alpha$ ,  $\alpha$ ,  $\beta$ , and  $\beta'$ . The  $\sigma^{66}$  subunit has four regions. Scc4 binds to the flap of the  $\beta$  subunit in RNAP and region four of  $\sigma^{66}$ .<sup>81</sup> The two subunits are essential for recognition of the specific -35 promoter regions of genes for transcription initiation. T3S proteins such as the *Chlamydia* outer membrane protein N (CopN) and specific *Chlamydia* chaperon 1 (Scc1) can interact with Scc4 and are hypothesized to hinder or reverse its anti  $\sigma^{66}$  activity.

*Chlamydia* infect a variety of host species including arthropods, vertebrates, and amoeba.<sup>75</sup> Typical hosts for different species are listed in Table 3. The species affecting humans are mainly *C. trachomatis*, *C. pneumoniae*, and *C. psittaci*. According to a 2012 report, there were 131 million new cases of CT infection globally, which translated to 4.2% among women and 2.7% among men aged 15-49; the regional prevalence was highest in the Americas and western pacific region.<sup>82</sup>

The Centers for Disease Control and Prevention report for 2017 indicates that CT infection is the most notifiable sexually transmitted disease in the United States. In 2016 alone, over 1.5 million cases were reported, which corresponded to 497.3 cases per 100,000 population, an increase of 4.7% from the previous year.<sup>83</sup> Between the years 2000 and 2016 there was a continuous rise in the *Chlamydia* infection rate. The infection was more prevalent in females than males. Both groups were highest at age 20-24 of sexually active individuals.<sup>83</sup>

**Table 3.** *Chlamydia* species and their respective hosts.<sup>75</sup>

Species	Typical host
<i>C. muridarum</i>	Mice, hamsters
<i>C. suis</i>	Swine
<i>C. pneumoniae</i>	Humans, horses, koalas
<i>C. psittaci</i>	Birds, poultry
<i>C. abortus</i>	Ruminants, swine
<i>C. caviae</i>	Guinea pigs
<i>C. felis</i>	Cats
<i>C. pecorum</i>	Ruminants, swine, koalas

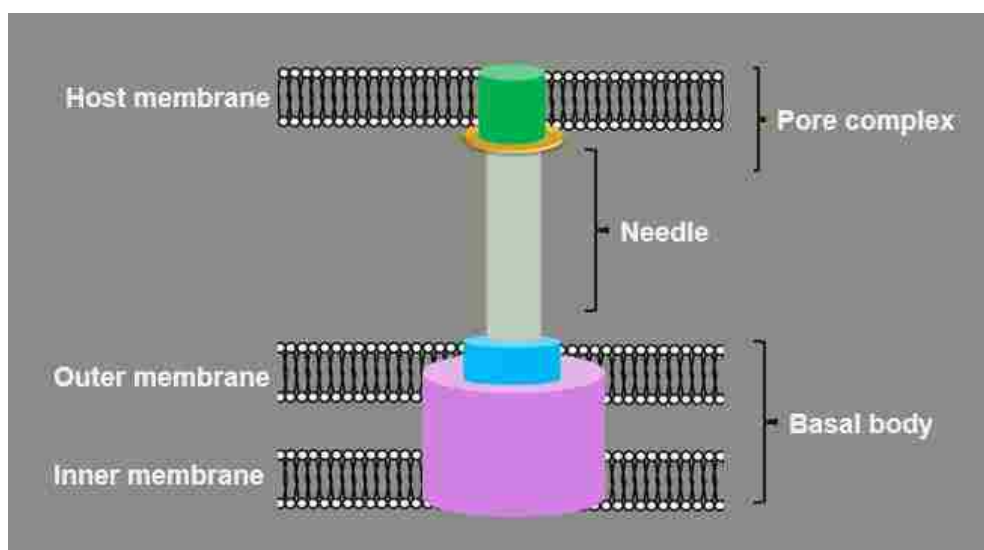
Simple chlamydial infections can be treated by primary antibiotics, such as azythromicins and tetracyclines. However, if the developmental cycle assumes the persistent stage, these drugs are ineffective. Additionally, co-infection with other STIs such as *gonorrhoea* can complicate the treatment process.<sup>84</sup> The  $\beta$ -lactam antibiotics used for the treatment of *gonorrhoea* induces persistence in *Chlamydia* infections. Therefore, successful treatment of co-infection can be achieved by broad spectrum antibiotics.<sup>84</sup> Other chlamydial antibiotics that have been used include fluoroquinolone, rifampin, lincomycin aminoglycosides, and trimethoprim.<sup>85, 86</sup>

Antimicrobial resistance by *Chlamydia* is attributed to a point mutation in the *gyrA*, foreign genomic island *Chlamydia* chromosome, a mutation in the peptidyl transferase region of 23S rRNA genes, and nucleotide substitution in the *rpoB* gene.<sup>86</sup> Some studies have shown that natural immunity to CT infections can be acquired with long term exposure.<sup>87</sup> Lower prevalence was also observed in mice with increasing age, indicating that individuals may acquire adaptive physiological processes to resist infection.<sup>88</sup> However, since the immunity acquired is serovar specific, the individuals who are immune to one serotype are susceptible to another. Therefore, there is a need to develop drugs to non-traditional antibiotic targets, such as virulence, the T3SS, and the unique developmental cycle of CT or develop broad-spectrum vaccines. Studies on different forms of vaccines have been conducted; these forms include the use of whole cell (live,



live attenuated, or UV-inactivated) or subunit vaccines. The subunit vaccines utilize antigens such as major outer membrane protein (MoMP), plasmid glycoprotein 3, chlamydial protease-like activity factor, polymorphic membrane proteins, and multivalent vaccines.<sup>88</sup>

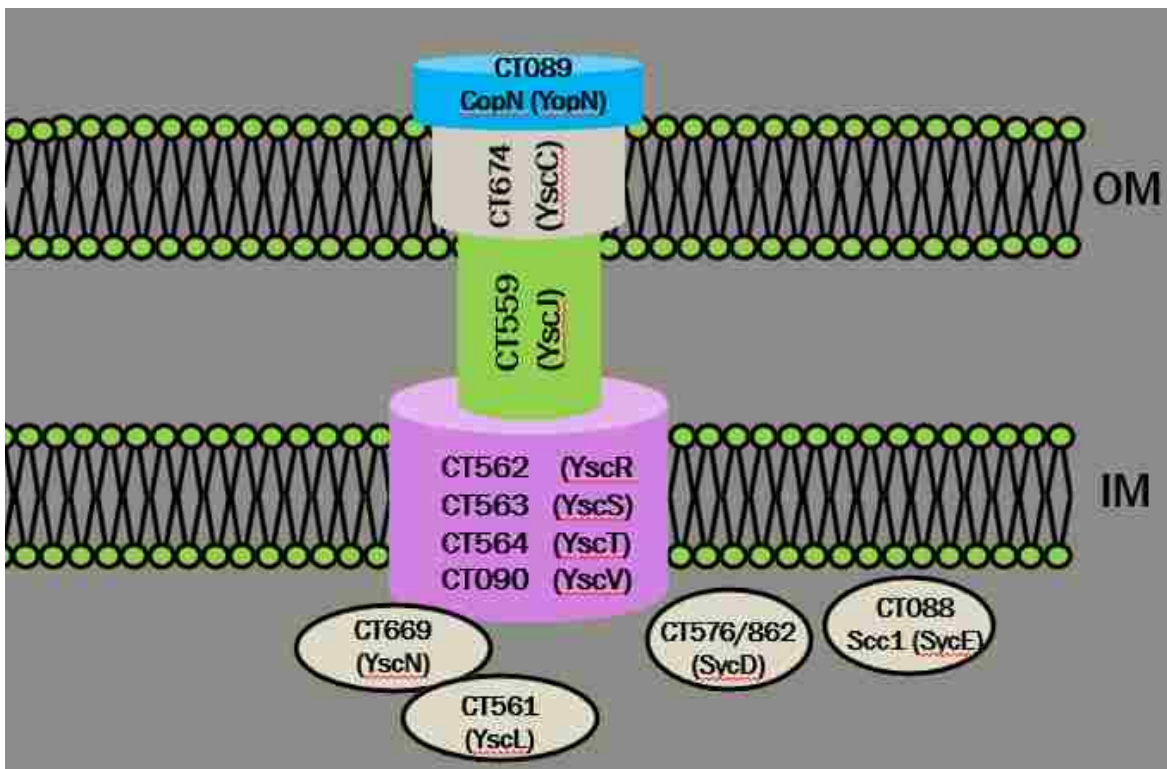
*Chlamydia* uses T3S apparatus to invade the host cell. The T3S injectisome serves as a channel to transfer effector proteins from the bacterial cell to the host cell. Effectors interfere with the host cell defense mechanism and facilitate endocytosis of the EB. The injectisome is made of multiple subunits of about 20 proteins that create a pathway for direct injection of effector proteins to the host cell, bypassing the extracellular matrix.<sup>89</sup> The injectisome comprises ATPase, effectors, translocators, a needle-like protein structure, and chaperons.<sup>90-92</sup> The structural arrangement is represented in Figure 11. A broad spectrum of diseases caused by T3SS-containing pathogens are outlined in Table 4.



**Figure 11.** Illustration of T3S apparatus spanning the bacterial membranes to the host membrane.

The commonly reported arrangement of T3S proteins in CT is illustrated in the model shown in Figure 12. The protein CT559 spans the periplasmic region connecting the inner membrane (IM) to the outer membrane (OM), while CT674 forms the outer membrane pore. The inner membrane proteins, CT562, CT563, CT564, CT091 (Cds1), and CT090 (Cds2), form the base.<sup>93</sup> The cytoplasmic protein, CT669, is presumed to be the ATPase that energizes the system in association with CT561. The other two proteins, CT576 and CT088 (Scc1), are chaperons.

Chlamydial infection slows down the host cell cytokinesis through two mechanisms. The first mechanism is the regulation of the G<sub>2</sub>/M transition by cleavage of the mitotic cyclin B<sub>1</sub>.<sup>94</sup> The second mechanism involves mitotic arrest effected by CopN, which binds directly to αβ-tubulin to hinder tubulin polymerization to microtubules.<sup>90</sup> Apart from modulating the host cell cycle, CopN functions as a plug for the T3SS, inhibiting protein secretion until activated by an external signal. It has been shown that CopN interacts with a class II chaperon Scc3 (specific *Chlamydia* protein 3) at its C-terminus and the class I chaperons, Scc1/Scc4 (in a heterodimer complex), at the N-terminus in *Chlamydia pneumoniae* (CP).<sup>91</sup> The three chaperons aid in the stabilization, secretion regulation, and translocation of CopN to the host cell. The CP protein complex Scc1/Scc4/CopN/Scc3 may function to regulate the secretion of translocators and effectors similarly to SycN/YscB/YopN/TyeA in *Yersinia*.<sup>92</sup>



**Figure 12.** A model of *Chlamydia* T3S apparatus, with *Yersinia* protein orthologs shown in parenthesis, adapted from Fields *et al.*<sup>93</sup>

**Table 4.** Disease causing pathogens that use a T3SS.<sup>7</sup>

<b>Pathogen</b>	<b>Disease</b>
<i>Pseudomonas aeruginosa</i>	Pneumonia, urinary tract infection, wound infection, septicemia, and endocarditis
<i>Chlamydia trachomatis</i>	Trachoma and conjunctivitis
<i>Chlamydia pneumonia</i>	Pneumonia, atherosclerotic disease
<i>Shigella</i> , <i>Salmonella</i> , and <i>Yersinia</i> species	Intestinal disease, plaque
<i>Salmonella</i> serovar Typhi	Enteric fever
<i>Bordetella</i>	Whooping cough
<i>Pseudomallei</i>	Bacteremia and pneumonia

This study investigated the binding interface of the Scc1:Scc4 heterodimer complex using solution NMR spectroscopy. The ultimate goals are to identify a therapeutic approach to interfere with this interface and inhibit CopN translocation, which will prevent chlamydial invasion. Recombinant expression of Scc1 alone is challenging due to its insolubility. However, co-expression of Scc1 with Scc4 removes the challenge by forming a soluble heterodimer complex.<sup>95</sup> Four protein sequences were used in this study, named Scc1His, HisScc4, Scc4, and HisScc4-22-end (amino acid sequence for Scc1His and HisScc4 provided in Appendix A-1). The first three proteins contain the full-length sequence of Scc1 or Scc4; the last protein has a shortened sequence of Scc4 beginning from amino acid number 22, which was expected to be a highly soluble protein with well resolved peaks in solution NMR spectra. The hexa-histidine tag (His) added to some protein sequences is used as a purification tag via Ni-affinity chromatography.

## **2.3 Cloning and protein expression**

### **2.3.1 Materials**

Cloned vectors, pACYC184-Scc1His, pET28-Scc4, and pCDF-HisScc4, were generously provided by Dr. Li Shen (LSU Health Sciences Center, New Orleans, LA).<sup>95</sup> pET28a vector was purchased from Novagen (Merck Sigma). The DNA insert (Scc4-22-end gene block fragment), pET 3' primer (CTA GTT ATT GCT CAG CGG), and pET upstream primer (ATG CGT CCG GCG TAG A) were purchased from IDT. The NcoI, XhoI, EcoR1, and T4 DNA ligase enzymes, NEB 10 Beta electrocompetent cells, BL21 (DE3)

Gold *E. coli* cells, SOC, NEB buffer 3.1, NEB buffer 2.1, and T4 ligase buffer were purchased from New England Biolabs (NEB). Polymerase chain reaction (PCR) purification kit and QIAprep spin miniprep kit were purchased from Invitrogen. Luria-Bertani (LB) broth, sodium phosphate dibasic, glycerol, methanol, glacial acetic acid, and 100X BME vitamin solution were purchased from VWR. Lennox LB agar was purchased from BD Biosciences. Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) was purchased from Anatrace. Spectinomycin (Spec) and sodium azide were purchased from Alfa Aesar. Kanamycin (Kan), hydrochloric acid, sodium hydroxide, acetonitrile, trifluoroacetic acid, iron (III) chloride, calcium chloride, zinc sulfate, copper (II) sulfate, sodium selenite, dithiothreitol (DTT),  $\beta$ -mercaptoethanol ( $\beta$ ME), nickel (II) chloride, manganese sulfate, ammonium bicarbonate, chloramphenicol (Cam), EDTA-free SIGMAFAST protease inhibitor tablets, tris (hydroxymethyl) aminomethane (Tris), sodium dodecyl sulfate (SDS), thiamine, cobalt chloride, sodium molybdate, and imidazole were purchased from Sigma-Aldrich (Merck). Sodium phosphate monobasic and glycine were purchased from Amresco. Nickel (II) chloride hexahydrate and ammonium chloride were purchased from Acros. Mini-PROTEAN TGX precast gels (4-20% Tris/glycine), 2x Laemmli sample buffer, Bio-gel<sup>®</sup> P4 gel, and Profinity IMAC uncharged resin were purchased from Bio-Rad. EDTA-free Pierce protease inhibitor mini tablets, Coomassie Plus protein assay reagent, bovine serum albumin standard (BSA), sodium chloride, boric acid, agarose, and EZ-Run Rec protein ladder were purchased from Thermo Fisher Scientific. Benzoylase nuclease (25-29 U/ $\mu$ L), sodium lauroyl sarcosinate (sarkosyl), ethidium bromide, and BugBuster 10x protein extraction reagent were purchased from Novagen (Millipore Sigma). Magnesium sulfate, molecular weight cut-off centrifugal filters, and polyethersulfone membrane (0.025  $\mu$ m) were purchased from EMD Millipore (Millipore Sigma). Sinapic acid and Coomassie brilliant blue R 250 were purchased from Fluka. <sup>13</sup>C-glucose (U-13C6, 99%), deuterium oxide (D<sub>2</sub>O, D, 99.9%), and <sup>15</sup>N-ammonium chloride (<sup>15</sup>NH<sub>4</sub>Cl, 15N, 99%) were purchased from Cambridge Isotope Laboratories. 4,4-Dimethyl-4-silapentane-1-sulfonic acid (DSS) was purchased from Millipore-Sigma. Phenylmethylsulphonyl fluoride (PMSF) was purchased from Calbio Chem.

All water was supplied by a Millipore Direct-Q3 ultrapure water system. Solid and liquid sterilization was accomplished using an AMSCO autoclave. Colonies were grown in an Imperial III incubator (Labline Instruments). Shigemi NMR tubes matched to D<sub>2</sub>O's magnetic susceptibility were purchased from Shigemi. Centrifugation was done on a Sorvall RC 6 centrifuge (Thermo Fisher Scientific). Protein purification was

done on a Biologic DuoFlow 10 chromatography system from Bio-Rad. Other apparatuses included a PowerWave XS2 microplate reader (BioTek), an Excella E24 incubator shaker (New Brunswick Scientific), a PCR cabinet (ESCO Technologies), an orbital shaker (Forma Scientific), a VCX 130 sonicator (SONICS), a compact UV lamp UVGL-25 (UVP), and a French pressure cell press (SLM Aminco). Lyophilization was done using a FreeZone Plus 6 liters freeze dry system (Labconco). Electroporation was done with an Eppendoff electroporator 2510. PCR was conducted using a Bio-Rad S1000 thermal cycler.

### **2.3.2 Cloning of pET28a-Scc4-22-end plasmid**

The Scc4-22-end DNA was cloned into the pET28a (+) vector using NcoI and XhoI restriction enzymes. EcoR1 restriction enzyme was utilized post ligation to linearize any empty pET28a vector. The vector, pET28a (+), and the DNA insert, Scc4-22-end, were digested separately in the digestion buffer (5  $\mu$ L NEB buffer 3.1, 1  $\mu$ L NcoI, 1  $\mu$ L XhoI, and 43  $\mu$ L sterile water) at 37 °C for 1 h. The enzymes were heat inactivated at 80 °C for 20 min. The digestion products were purified using the PCR purification kit, with B2 and B3 binding buffers for the insert and the vector, respectively.

The purified vector and insert (5  $\mu$ L each, 1:1 molar ratio and 2:1 molar ratio) were ligated in the ligation buffer (2  $\mu$ L 10X T4 ligase buffer, 2  $\mu$ L T4 DNA ligase, 11  $\mu$ L sterile water) at room temperature for 1 h. The ligase was heat inactivated at 65 °C for 20 min. Post ligation digestion was done by adding 0.2  $\mu$ L of EcoR1 enzyme and 5  $\mu$ L of 10X NEB buffer 2.1 to the ligation mixture and incubating at 37 °C for 2 h. EcoR1 was heat inactivated at 65 °C for 20 min. The mixtures were transferred to a polyethersulfone membrane suspended in sterile water in a petri dish to desalt for 15 min.

The ligation mixtures were transformed into NEB 10 Beta electrocompetent cells following the standard electroporation protocol. The transfected cells were cultured overnight on LB agar plates (kanamycin, 50  $\mu$ g/mL). Five colonies were selected and cultured in 10 mL of LB broth overnight. DNA was purified from 5 mL of the starter culture using the standard protocol for the QIAprep spin miniprep kit (using microcentrifugation). The purified DNA was analyzed by PCR amplification using the pET 3' (CTA GTT ATT GCT CAG CGG, Tm 53.7 °C) and pET upstream (ATG CGT CCG GCG TAG A, Tm 60.8 °C) primers, at an annealing temperature of 45 °C. The DNA sequence was confirmed by gene sequencing at LSU's GeneLab.

### 2.3.3 Expression and purification of HisScC4-22-end, ScC1His:ScC4, HisScC4, and untagged ScC4 proteins

The plasmids were individually transformed or co-transformed into BL21 (DE3) Gold *E. coli* expression cells using the heat shock protocol.<sup>96</sup> Single cell colonies of the transformed BL21 (DE3) Gold *E. coli* cells were grown on LB agar plates. The antibiotics in the agar and culture media were kanamycin, chloramphenicol, spectinomycin, and tetracycline. The respective concentrations of the antibiotics are listed in Table 5.

A starter culture of each transformed cell line was prepared by inoculating 100 mL of LB medium (containing the respective antibiotics) with cells from a single colony on the agar plate. The cells were incubated overnight in the incubator shaker at 250 rpm and 37°C. For the expression of unlabeled proteins, 50 mL starter cultures were transferred to 1 L culture flasks containing 950 mL of LB broth and the antibiotics. The cells were incubated in the shaker to an optical density (OD) of 0.6 at 37 °C, followed by induction with 0.5 mM IPTG. The cells were incubated for another 3 h and harvested by centrifugation at 3000 x g for 30 min at 4°C. The isotopically labeled proteins were expressed in minimal media<sup>97</sup> (Appendix A-2) supplemented with <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C-glucose. Triple labeled proteins were expressed in deuterated minimal media supplemented with <sup>13</sup>C-glucose and <sup>15</sup>NH<sub>4</sub>Cl. A 10 mL starter culture was used to inoculate 1 L of the minimal media. The cells were grown at 37 °C to 0.9 OD, and protein expression was induced with 0.5 mM IPTG. The cell culture was incubated at 16 °C for 20 h.

**Table 5.** Concentrations of different antibiotics used in the cell culture media.

Plasmid	Protein	Antibiotic resistance	Concentration (µg/mL)
pET28a(+)-ScC4-22-end	HisScC4-22-end	Kanamycin	Kan, 50
pACYCScC1-His & pET28-ScC4	ScC1His & ScC4	Chloramphenicol Kanamycin	Cam, 35 Kan, 50
pET28-ScC4	Untagged ScC4	Kanamycin	Kan, 50
pCDF-His-ScC4	HisScC4	Spectinomycin	Spec, 100

The cells were harvested by centrifugation at 3000 x g for 30 min at 4°C and re-suspended in 25 mL of the loading/wash buffer (20 mM Tris pH 8.0, 150 mM NaCl, 15 mM imidazole, and 5% glycerol)

containing 1X BugBuster protein extraction reagent, one-half of an EDTA-free protease inhibitor tablet, 1 mg phenylmethylsulphonyl fluoride (PMSF), and 1.25  $\mu$ L benzonase nuclease. Mechanical lysis was done using a French pressure cell or sonication. The lysed cells were centrifuged at 20,000 x g for 30 min at 4°C to obtain a cleared lysate. The hexahistidine (His) tagged proteins were purified using 4 mL of Ni-charged Profinity IMAC resin (Ni-column) packed on a column with a flow adaptor attached to a BioLogic DuoFlow FPLC. The cleared lysate was loaded into a Ni-column, the bound proteins were washed with 20 mL of loading/wash buffer, and the purified protein was eluted with 10 mL of 20 mM Tris pH 8.0, 150 mM NaCl, 250 mM imidazole, and 5% glycerol (elution buffer). A similar FPLC set up was used during protein desalting or to exchange the buffer, except that the column was packed with 15 mL Bio-gel P4 gel (P4-column). The purified proteins were exchanged into 10 mM ammonium bicarbonate buffer (ABC buffer, pH ~8.0, pH not adjusted) using the P4-column, lyophilized, and stored at -20 °C. All the buffers used in the purification steps contained 0.02% NaN<sub>3</sub> as a bacteriostat except for the ABC buffer.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to check the purity of the protein. The SDS-PAGE running buffer (800 mL) was 25 mM Tris, 192 mM glycine, and 0.1% SDS. The sample buffer was prepared by adding 5%  $\beta$ ME to 2x Laemmli buffer. Equal volumes of sample buffer and protein solution were mixed and heated at 95 °C for 5 min. The denatured proteins were loaded into 30  $\mu$ L-wells in Mini-PROTEAN TGX precast gels, 4-20% Tris/glycine, and separated at 150 V for 50 min. The protein bands were stained by placing the gel in Coomassie stain (0.25% w/v Coomassie Blue G-250, 10% acetic acid, and 40% methanol in water) for 1 h and transferred to de-stain solution (10% acetic acid and 25% methanol in water) for 1 h.

The Coomassie (Bradford) assay was used to quantify the purified proteins. Using a 96-well plate, 200  $\mu$ L of Coomassie Plus protein assay reagent was added to each well (for both samples and standards). Protein solution (2  $\mu$ L-15  $\mu$ L) was added to the sample wells and increasing volumes of 80  $\mu$ g/mL BSA solution (0, 10, 20, 30, and 40  $\mu$ L) were transferred to separate wells. All volumes were adjusted to 250  $\mu$ L with water. The absorbance at 595 nm was recorded on a microplate reader, and the proteins were quantified using a calculated BSA standard calibration curve.

### **2.3.4 Isolation of $^{15}\text{N}$ -Scc4 and $^{15}\text{N}$ -Scc1His from $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex**

Cleared lysate of the co-expressed  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex (25 mL) was loaded onto a Ni-column at a flow rate of 1 mL/min. The unbound proteins were washed (20 mL of wash buffer), and the complex was disrupted by elution of  $^{15}\text{N}$ -Scc4 with 0.5% sarkosyl in wash buffer (20 mL). The  $^{15}\text{N}$ -Scc1His trapped in the column was eluted with 10 mL of elution buffer or used to form on-column mixed heterodimer complex (discussed in the next section). The eluted  $^{15}\text{N}$ -Scc4 fractions were combined and diluted (3X) with Tris buffer (20 mM Tris, 150 mM NaCl, 0.02%  $\text{NaN}_3$  at pH 7.4) to lower the sarkosyl concentration to less than the critical micelle concentration (0.4%). The diluted  $^{15}\text{N}$ -Scc4 protein was concentrated using a 10 kDa molecular weight cut-off centrifugal filter. The buffer was exchanged to 20 mM Tris, 150 mM NaCl, 0.02%  $\text{NaN}_3$  at pH 7.4 using a P4 column. The collected fractions were then passed through a freshly equilibrated Ni-column to remove any residual  $^{15}\text{N}$ -Scc1His that may have eluted with  $^{15}\text{N}$ -Scc4. The flow through (pure  $^{15}\text{N}$ -Scc4) was collected, and the buffer was exchanged to ABC buffer. The sample was lyophilized and stored at  $-20\text{ }^\circ\text{C}$  or analyzed by NMR. Alternatively, pure  $^{15}\text{N}$ -Scc4 was isolated from the purified  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex. After loading the purified  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 onto a Ni-column (at a flow rate of 0.25 mL/min), the complex was disrupted, and the two proteins eluted following the same protocol describe above.

### **2.3.5 On-column formation of complexes with chain-selective isotopic labeling**

The  $^{15}\text{N}$ -labeled protein complex,  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 (purified and desalted) was loaded onto a Ni-column at a flow rate of 0.25 mL/min. The column was washed with 20 mL wash buffer, and  $^{15}\text{N}$ -Scc4 was eluted from the complex with 20 mL of 0.5% sarkosyl in wash buffer. Residual sarkosyl was removed from the column using additional wash buffer (15 mL). The unlabeled Scc4-lysate was loaded onto the column containing the trapped  $^{15}\text{N}$ -Scc1His at a flow rate of 0.5 mL/min. The unbound proteins were washed with 20 mL of wash buffer, and the newly formed mixed complex,  $^{15}\text{N}$ -Scc1His:Scc4, was eluted with 10 mL of elution buffer. Following a similar protocol, the mixed complex, Scc1His: $^{15}\text{N}$ -Scc4, was formed. In this case, Scc1His:Scc4 complex was not isotopically labeled, but Scc4-lysate was  $^{15}\text{N}$ -labeled. The mixed complexes were also formed on-column starting with a cleared lysate of  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex, in place of the purified complex following the same procedure.



### 2.3.6 NMR analysis of the purified proteins

The lyophilized proteins were reconstituted in 250  $\mu$ L of phosphate NMR buffer: 50 mM sodium phosphate, pH 7.4, 0.02%  $\text{NaN}_3$ , 50  $\mu$ M DSS, 10%  $\text{D}_2\text{O}$ . The proteins analyzed in Tris NMR buffer (20 mM Tris, 150 mM NaCl, pH 7.4, 0.02%  $\text{NaN}_3$ , 50  $\mu$ M DSS, 10%  $\text{D}_2\text{O}$ ) are indicated in their respective spectra. Dithiothreitol (DTT, 10 mM) was added to NMR samples containing  $^{15}\text{N}$ -Scc1His protein to prevent disulfide bond formation. The Scc4 sequence does not have a cysteine, so there was no need to add DTT to the Scc4 samples. The NMR analysis of  $^{15}\text{N}$ -Scc1His alone was done in elution buffer, with 10 mM DTT, 50  $\mu$ M DSS, and 10%  $\text{D}_2\text{O}$  added, due to its insolubility in phosphate buffer.

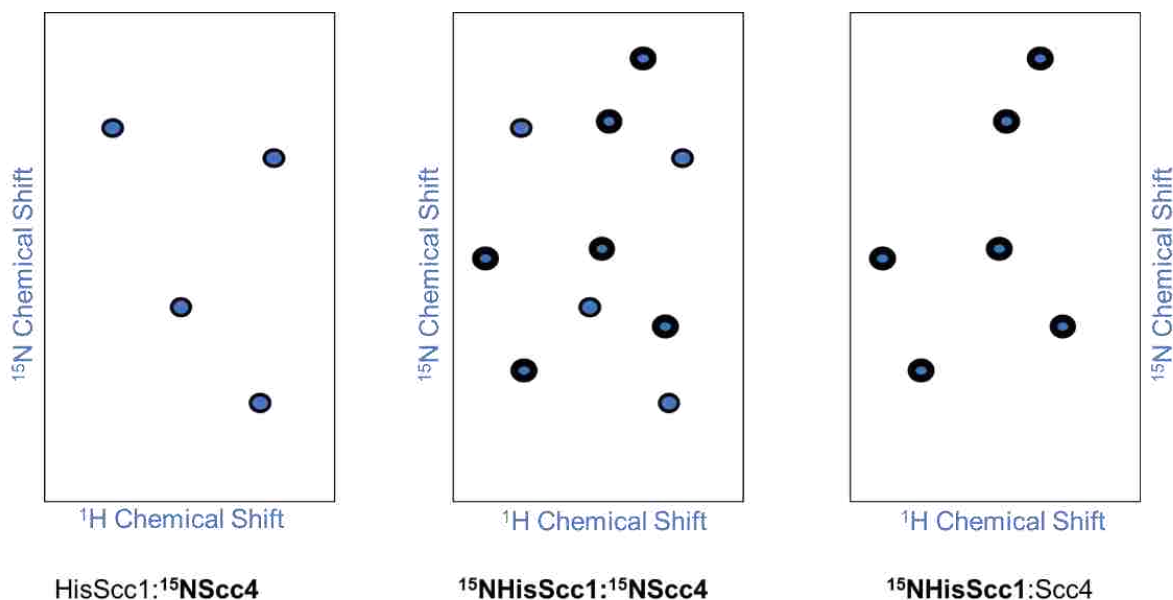
The reconstituted proteins were transferred to Shigemi NMR tubes for NMR analysis. NMR data were collected on a Bruker Avance III HD NMR spectrometer (500 MHz) and a Varian NMR spectrometer (700 MHz) at the LSU NMR facility, and a Bruker Avance III (600 MHz) and a Bruker Avance III HD (900 MHz) at NMRFAM (University of Wisconsin). Some protein samples were shipped to NMRFAM for 3D NMR analysis (in collaboration with Dr. Marco Tonelli). The recorded 3D NMR data included HNCA, HN(CO)CA, HNCO, HN(CA)CB, and HN(CO)(CA)CB experiments. All the 3D NMR spectra were TROSY-based and recorded at 25  $^\circ\text{C}$  on Bruker instruments equipped with cryogenic probes. The NMR data was acquired using a non-uniform sampling (NUS) rate of 40% to 50%. The data were processed using NMRPipe and the NUS data were reconstructed using SMILE (Bax group).<sup>98</sup> Detailed acquisition parameters for these experiments are provided in the Appendix A-3.

### 2.3.7 Triple Resonance NMR experiments for backbone assignment

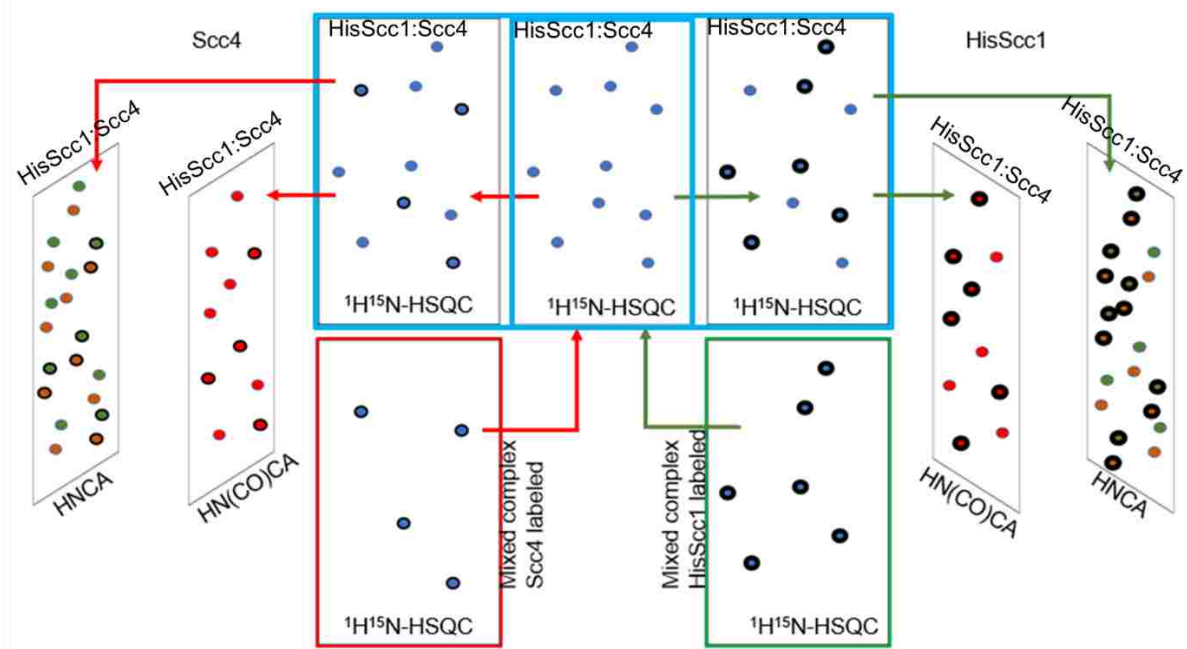
To identify the amino acids at the binding interface of the Scc1His:Scc4 complex, it is essential to do the backbone assignment. Due to the low yield and multiple preparation steps in the formation of the mixed complexes, it is more efficient to run 3D protein backbone experiments on the uniformly labeled complexes rather than the mixed labeled complexes. To use this 3D data for backbone assignment,  $^1\text{H}$ - $^{15}\text{N}$ -HSQC NMR spectra of the mixed labeled complexes are required to identify the individual protein peaks belonging to each protein. Once these peaks are identified, restricted peak picking in the 3D data sets can be used to assign the  $^1\text{H}$ ,  $^{15}\text{N}$ , and  $^{13}\text{C}$  nuclei in the backbone of each protein to the NMR peaks (illustrated in Figure 13). The peaks in one spectrum can be selected with respect to peaks in another spectrum. It is

therefore possible to pick peaks belonging to Scc4 in an  $^1\text{H}$ - $^{15}\text{N}$ -HSQC NMR spectrum of  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex using the restricted peak picking tool in the NMR SPARKY software.<sup>99</sup>

Triple resonance backbone assignment for Scc4 and Scc1His can be accomplished separately using the same set of 2D and 3D NMR data for the  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex. The pick peaking procedure for Scc4 backbone assignment is as follows: The peaks in the 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum of the mixed complex, Scc1His: $^{15}\text{N}$ -Scc4, are picked, followed by restricted peak picking of Scc4 peaks in the 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum of  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4. The peaks belonging to the Scc4 protein in the 3D spectra, such as HNCA, HN(CO)CA, HNCO, HNCACB, and HN(CO)CACB, can be picked using restricted peak picking with respect to the Scc4 peaks in the 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum of the  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex. Figure 14 is an illustration of how the same data set can be used to separately do backbone assignment for Scc1His and Scc4.



**Figure 13.** Illustration on how to apply restricted peak picking on  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of the  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 using mixed labeled samples. The peaks in one spectrum can be selected with respect to peaks in another spectrum using the 'kr' command in SPARKY.



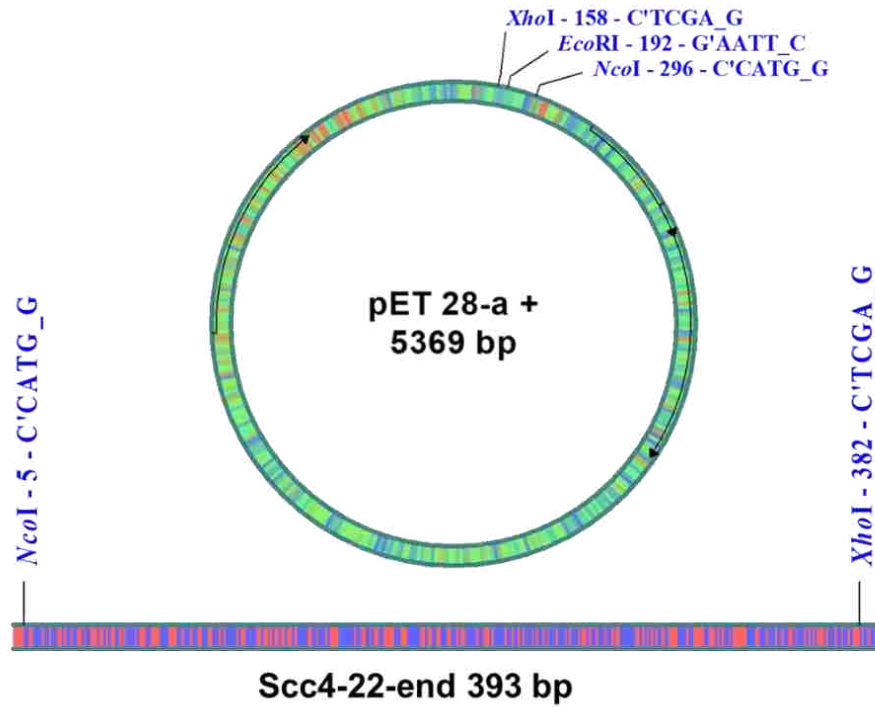
**Figure 14.** An illustration of how to use the same data set to separately conduct backbone assignment for each protein. The peaks in the 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum of the mixed complex, Sc4His: $^{15}\text{N}$ -Sc4 or  $^{15}\text{N}$ -Sc4His:Sc4, are picked, followed by restricted peak picking of  $^{15}\text{N}$ -Sc4 or  $^{15}\text{N}$ -Sc4His peaks in the 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum of  $^{15}\text{N}$ -Sc4His: $^{15}\text{N}$ -Sc4, respectively, and final peak picking in 3D spectra.

## 2.4 Results and discussion

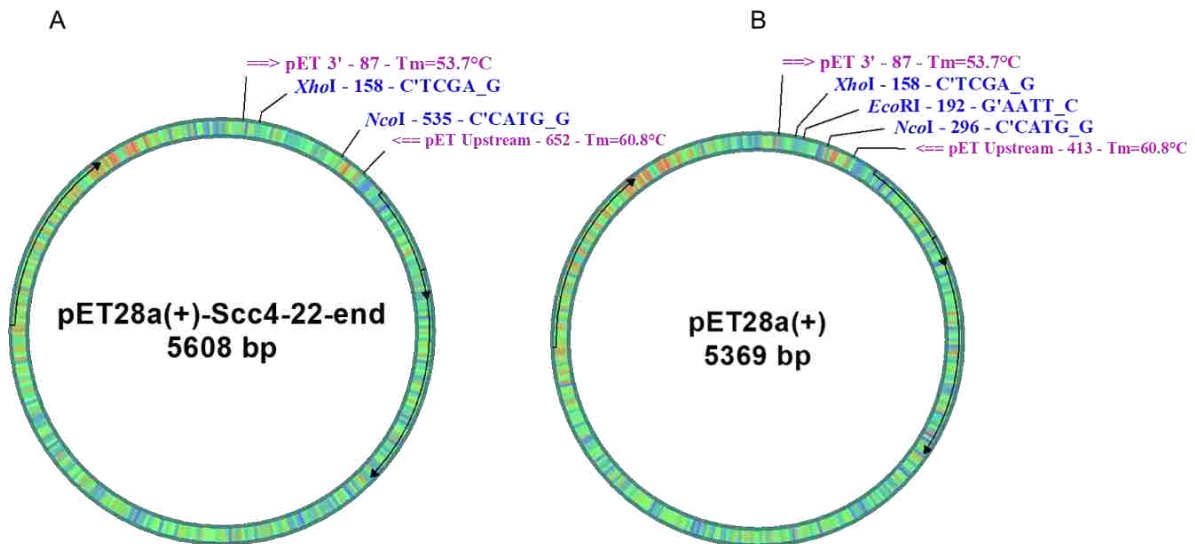
### 2.4.1 Cloning of pET28a-Sc4-22-end plasmid

A shorter sequence of the Sc4 protein (12.3 kDa) was predicted to be a better candidate for NMR studies compared to the full-length protein (14.7 kDa) due to its smaller size. The DNA sequence of the first 21 amino acids in the full-length protein is missing in the Sc4-22-end DNA construct. The Sc4-22-end DNA was cloned into the pET28a plasmid by double digestion and ligation. To verify that the cloning was successful, the purified plasmid was amplified using pET 3' and pET upstream primers. The amplification of the vector containing the Sc4-22-end DNA insert yields DNA strands of 565 bp, while the empty vectors yield DNA strands of 326 bp as illustrated in the pDRAW32 (AcaClone Software) images in Figures 15 and 16. The PCR amplification products were analyzed by agarose gel electrophoresis shown in Figure 17. The bands around 565 bp indicate that the cloning was successful (C5, C3, C2, C1 from the 2:1 ratio ligation and C5, C3, C2, C1 from the 1:1 ratio ligation). The purified plasmid from colony 2 (C2, analyzed in lane 4 of the agarose gel image in Figure 17) was submitted for DNA sequencing (LSU Genelab) and was

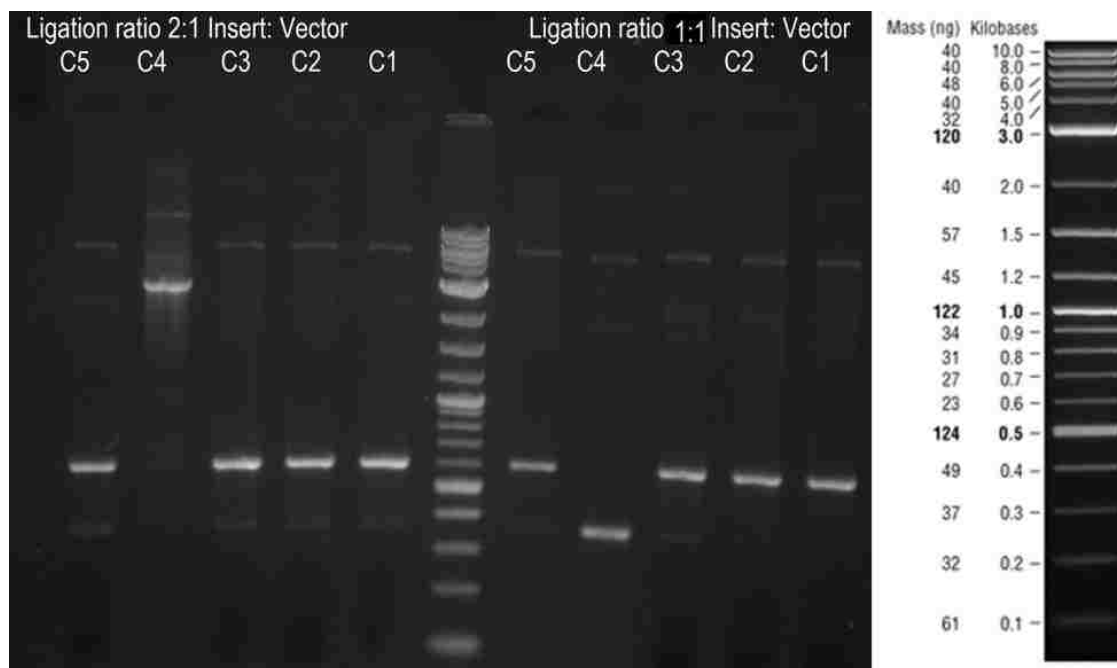
confirmed to be the correct sequence (Appendix A-4.1). The Scc4-22-end DNA sequence with the NcoI-XhoI restriction sites is provided in Appendix A-4.2.



**Figure 15.** pDRAW images of the Scc4-22-end linear DNA and pET28-a (+) circular DNA indicating the NcoI and XhoI restriction sites.



**Figure 16.** pDRAW32 primer analysis of the circular DNA of the A) empty and B) cloned vectors.



**Figure 17.** Agarose gel image of the PCR products from the pET28a(+) ligation with Scc4-22-end. Two ligation insert/vector ratios were used in the ligation reactions. Five colonies (C1, C2, C3, C4, C5) from cells transformed with each ligation reaction mixture are shown with the 2:1 insert/vector ratio (left) and 1:1 insert/vector ratio (right).

The translation of the Scc4-22-end DNA sequence in the pET28a-Scc4-22-end plasmid is shown below (HisScc4-22-end). The complete sequence of the Scc4 protein without the hexa-histidine tag is also shown. The full length Scc4 sequence includes the first 21 *N*-terminal amino acids, MLEKLIKNFVAYMGVASELEF, omitted in HisScc4-22-end sequence, which begins from amino acid number 22. After confirming the correct sequence, the purified plasmid from C2 was transformed into BL21 Gold (DE3) cells for Scc4-22-end protein expression.

MGSSHHHHHSSDADGSYVFPISLVRMRVRQNADEEIIISAFLEIPASMDIEKAYARMMEGNLFGQET  
GGAALGLDSDGHAVLVRVPGEVSQEDFASYIESVLNYAEAWLEDLGLSKTEQE

Theoretical pI/Mw: 4.18 / 12259.65; 124 amino acids

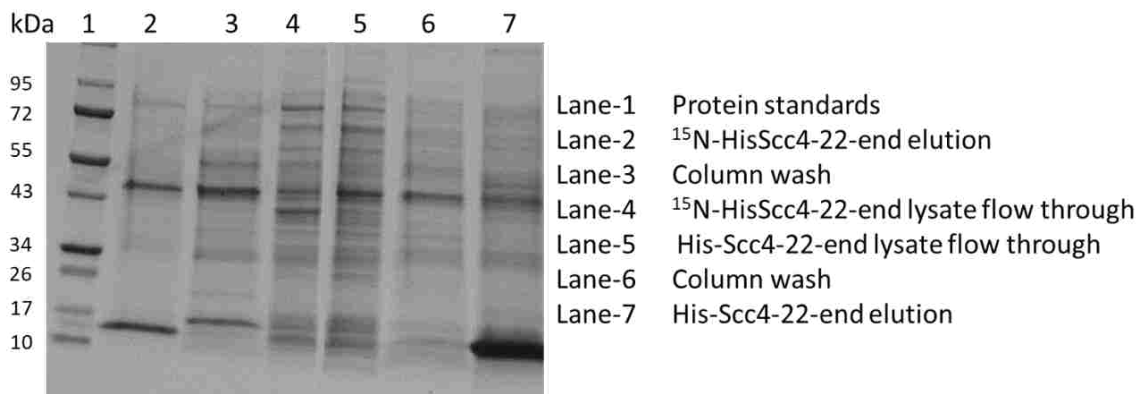
Scc4 amino acid sequence:

MLEKLIKNFVAYMGVASELEFDADGSYVFPISLVRMRVRQNADEEIIISAFLEIPASMDIEKAYARMMEG  
NLFGQETGGAALGLDSDGHAVLVRVPGEVSQEDFASYIESVLNYAEAWLEDLGLSKTEQE

Theoretical pI/Mw: 4.25 / 14674.55; 133 amino acids

## 2.4.2 Expression and purification of HisScC4-22-end, ScC1His:ScC4, HisScC4, and untagged ScC4 proteins

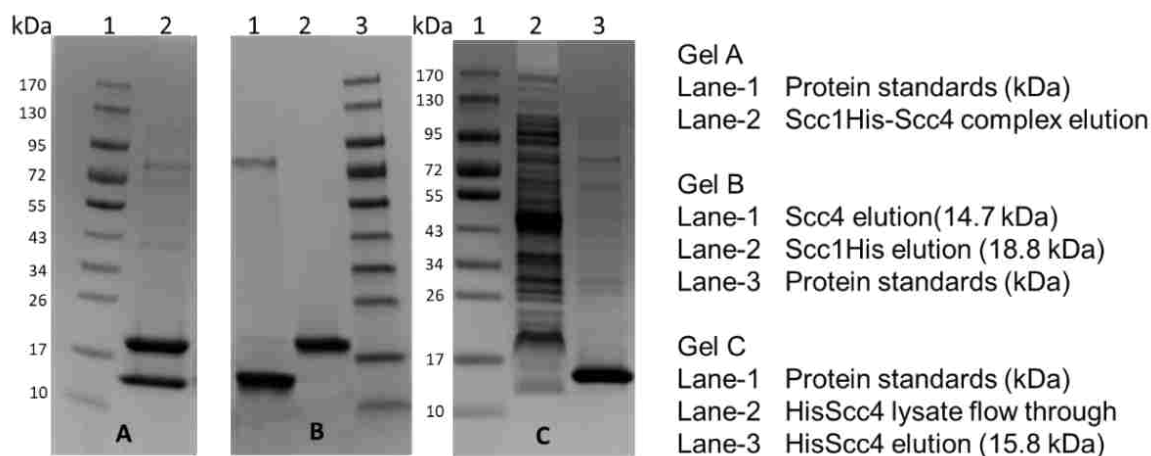
HisScC4-22-end was expressed in minimal media supplemented with  $^{15}\text{NHCl}_4$ . The elution fraction from Ni-IMAC purification of  $^{15}\text{N}$ -HisScC4-22-end protein was analyzed by SDS-PAGE (lane-2 Figure 18). Compared to the protein expressed in LB medium (lane-7), the band in lane-2 is faint and has almost the same intensity as the non-specific Ni-binding protein (the band between the 43 kDa and 55 kDa protein standards in lane-1) in the same lane. The yield was insufficient (about 0.5-1 mg per L culture) for 3D NMR studies, which require a minimum of 2 mg. Optimization is possible, but efforts were diverted to determine the best method for isolating full-length ScC4 (untagged) from the ScC1His:ScC4 protein complex.



**Figure 18.** SDS PAGE gel image of HisScC4-22-end protein expression in LB and minimal media. Lanes 2-4 are the analysis of minimal media expression and lanes 5-7 are the analysis of LB media expression. Lane-2 is  $^{15}\text{N}$ -HisScC4-22-end elution, Lane-3 is the column wash, Lane-4 is the  $^{15}\text{N}$ -HisScC4-22-end-lysate after flowing through the Ni-column, Lane-5 is the HisScC4-22-end-lysate after flowing through the Ni-column, Lane-6 is the column wash, and Lane-7 is HisScC4-22-end elution.

Based on the preceding NMR experiments of  $^{15}\text{N}$ -HisScC4 protein performed by Octavia Goodwin, it was determined that the histidine tag interfered with the stability of the protein. Expression of the full-length  $^{15}\text{N}$ -ScC4 protein without a histidine tag was achieved in our collaborator's, Dr. Li Shen's, laboratory; however, it required multiple purification steps that lead to the loss of protein. Non-denaturing conditions were established and used to isolate untagged ScC4 from the ScC1His:ScC4 protein complex.  $^{15}\text{N}$ -ScC1His and  $^{15}\text{N}$ -ScC4 proteins were co-expressed together. Because of their tight binding, the proteins co-purify in a 1:1 ratio using a Ni-column. The SDS-PAGE analysis of the purified  $^{15}\text{N}$ -ScC1His: $^{15}\text{N}$ -ScC4 complex, Figure 19A, shows the two bands corresponding to  $^{15}\text{N}$ -ScC1His (18.8 kDa) and  $^{15}\text{N}$ -ScC4 (14.7 kDa). To isolate untagged  $^{15}\text{N}$ -ScC4 protein, the complex was reloaded to the Ni-column, and  $^{15}\text{N}$ -ScC4 was eluted

with 0.5% sarkosyl in wash buffer. The remaining  $^{15}\text{N}$ -Scs1His protein was eluted with elution buffer. The two proteins were isolated with high purity as shown in the SDS-PAGE gel image, Figure 19B.  $^{15}\text{N}$ -Scs4 eluted with the non-specific, Ni-binding protein at around 72 kDa. After successful isolation of pure  $^{15}\text{N}$ -Scs1His and  $^{15}\text{N}$ -Scs4 protein, the impact of sarkosyl on the structure of Scs4 was assessed.

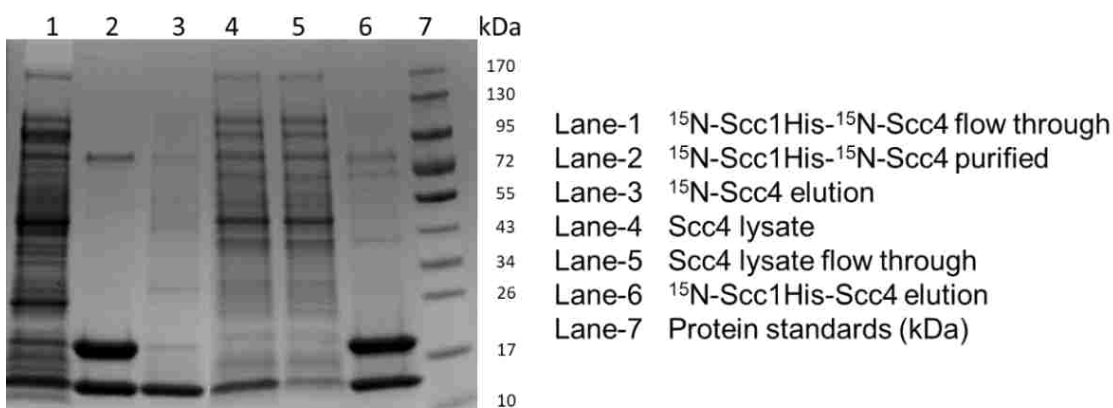


**Figure 19.** SDS-PAGE analysis of the purified proteins; A) Co-purification of the  $^{15}\text{N}$ -Scs1His: $^{15}\text{N}$ -Scs4 complex; B) purification of  $^{15}\text{N}$ -Scs4 (lane-1) and  $^{15}\text{N}$ -Scs1His (lane-2); C) purification of  $^{15}\text{N}$ -HisScs4 (lane-3).

For NMR spectral comparison and assessment of sarkosyl's effects on the Scs4 structure,  $^{15}\text{N}$ -HisScs4 protein was expressed and purified so its NMR spectrum could be compared to  $^{15}\text{N}$ -Scs4's spectrum. The SDS-PAGE analysis of purified  $^{15}\text{N}$ -HisScs4 (Figure 19C) shows a thick band corresponding to  $^{15}\text{N}$ -HisScs4 (15.8 kDa) in lane-3. Since the  $^{15}\text{N}$ -HisScs4 purification steps do not require sarkosyl, the NMR spectra of  $^{15}\text{N}$ -Scs4 and  $^{15}\text{N}$ -HisScs4 were used to compare the backbone structures of each protein and show possible changes to  $^{15}\text{N}$ -Scs4 due to sarkosyl (discussed in Section 3.3.3).

Interface NMR experiments require chain selective isotopic labeling of the proteins in the complex. Having established the conditions for isolating Scs1His and Scs4. The next step was to form mixed-heterodimer complexes of Scs1His:Scs4; in which only one protein was isotopically labeled. Two protein expressions, one in LB media and the other in labeled minimal media were performed to yield unlabeled Scs1His:Scs4 and labeled  $^{15}\text{N}$ -Scs1His: $^{15}\text{N}$ -Scs4 protein complexes. Isolation of each protein in the complex with 0.5% sarkosyl and the exchange of the proteins yield  $^{15}\text{N}$ -Scs1His:Scs4 and Scs1His: $^{15}\text{N}$ -Scs4 mixed labeled complexes, respectively. The complex was favorably reconstituted using

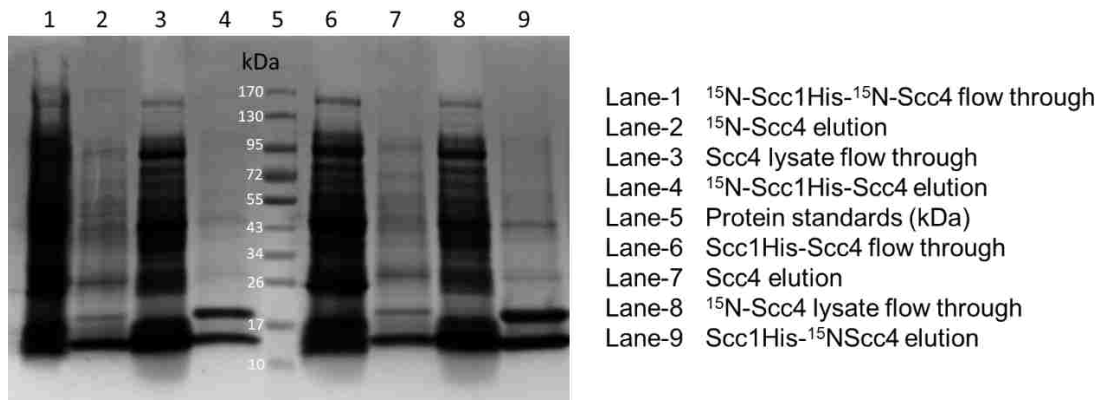
untagged Scc4 protein lysate that was separately expressed and loaded to Scc1His protein trapped in the Ni-column. The SDS-PAGE analysis (Figure 20) indicates successful on-column formation of the mixed heterodimer complex,  $^{15}\text{N}$ -Scc1His:Scc4. The flow through (lane 1) is the lysate of the complex after passing through the Ni-column to trap 1:1 bound  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4. The thick band around 14.7 kDa corresponds to the excess untagged-Scc4 not bound to  $^{15}\text{N}$ -Scc1His. The lysate of the separately expressed Scc4 is shown in lane 4 on the SDS-PAGE gel (Figure 20). After loading the Scc4 lysate onto the Ni-column containing  $^{15}\text{N}$ -Scc1His, Scc4 was successfully trapped, confirmed by the faded Scc4 band in the flow through (lane 5), indicating the formation of the on-column mixed complex (lane 6, the eluted complex).



**Figure 20.** SDS-PAGE analysis of the  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex disruption and mixed label complex formation (starting with the purified complex): Lane-1, the complex lysate flow through; Lane-2, the purified complex  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4; Lane-3,  $^{15}\text{N}$ -Scc4 eluted from the complex with 0.5% sarkosyl; Lane-4 lysate of the separately expressed Scc4; Lane-5 is Scc4 flow through; Lane-6, on-column formed mixed complex  $^{15}\text{N}$ -Scc1His:Scc4.

The mixed complexes were also formed without the initial purification of Scc1His:Scc4. In this procedure, the lysate of the complex was loaded onto the Ni-column, the unbound proteins washed, and Scc4 eluted with sarkosyl. The lysate of the separately expressed Scc4 was then loaded onto the column containing the trapped Scc1His, which formed the mixed labeled complex that was then eluted. The SDS-PAGE gel image (lane-4 and lane-9 Figure 21) depicts the newly formed mixed heterodimer complexes. It is important to note that the Scc4 eluted with 0.5% sarkosyl was not always pure (lane-2 and lane-7 Figure 21); a faint Scc1His protein band and other *E. coli* proteins were observed. Therefore, careful washing is necessary before elution of Scc4 to enhance its purity. The mixed complexes were analyzed by NMR to validate the purification method.

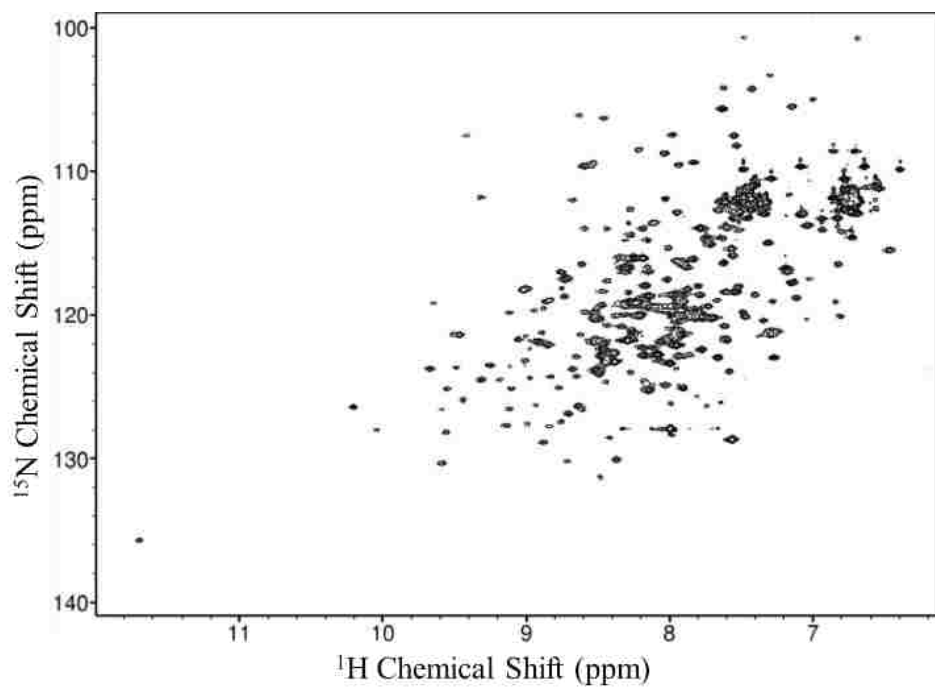




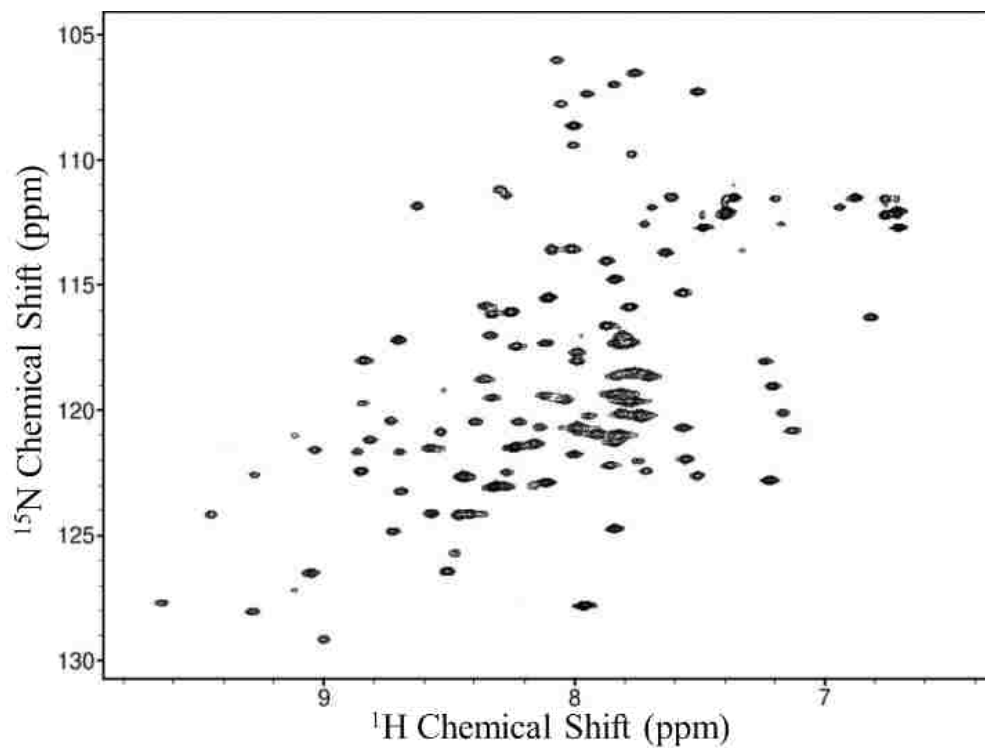
**Figure 21.** SDS PAGE analysis of the on-column mixed complex formation (starting with the lysate of the complex); Lane-1  $^{15}\text{N}$ -ScclHis:  $^{15}\text{N}$ -Sccl flow through, Lane-2  $^{15}\text{N}$ -Sccl elution with sarkosyl, Lane-3 Sccl lysate flow through, Lane-4  $^{15}\text{N}$ -ScclHis:Sccl mixed complex elution; and Lane-6 ScclHis: Sccl flow through, Lane-7 Sccl elution with sarkosyl, Lane-8  $^{15}\text{N}$ -Sccl lysate flow through, Lane-9 ScclHis:  $^{15}\text{N}$ -Sccl mixed complex elution.

### 2.4.3 NMR analysis of the purified proteins

$^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra were recorded for the isotopically labeled proteins. Although the  $^{15}\text{N}$ -ScclHis:  $^{15}\text{N}$ -Sccl heterodimer complex is large (33.5 kDa), it tumbles as a single unit in solution due to its strong binding. The spectrum displayed in Figure 22 was recorded in phosphate NMR buffer. Most peaks are well dispersed except for a few clusters. The variation in the peak intensities may be attributed to the presence of flexible regions in the protein structure, the presence of higher oligomeric states, or the presence of the un-complexed protein. The  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of  $^{15}\text{N}$ -Sccl has well dispersed peaks (Figure 23), indicating that the protein is properly folded and highly soluble in phosphate NMR buffer.

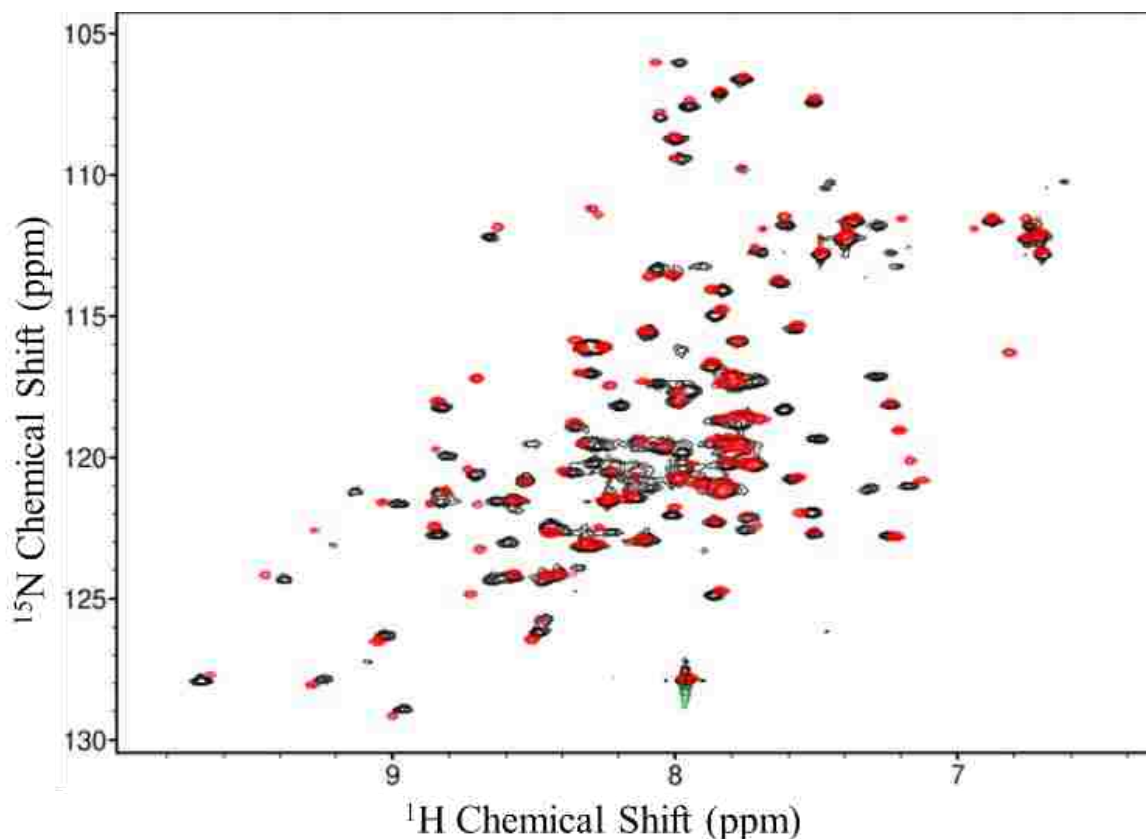


**Figure 22.**  $^1\text{H},^{15}\text{N}$ -HSQC spectrum of 0.75 mM  $^{15}\text{N}$ -ScclHis: $^{15}\text{N}$ -Sccl4 recorded at 700 MHz (Varian).



**Figure 23.**  $^1\text{H},^{15}\text{N}$ -HSQC spectrum for 0.50 mM  $^{15}\text{N}$ -Sccl4 recorded at 700 MHz (Varian).

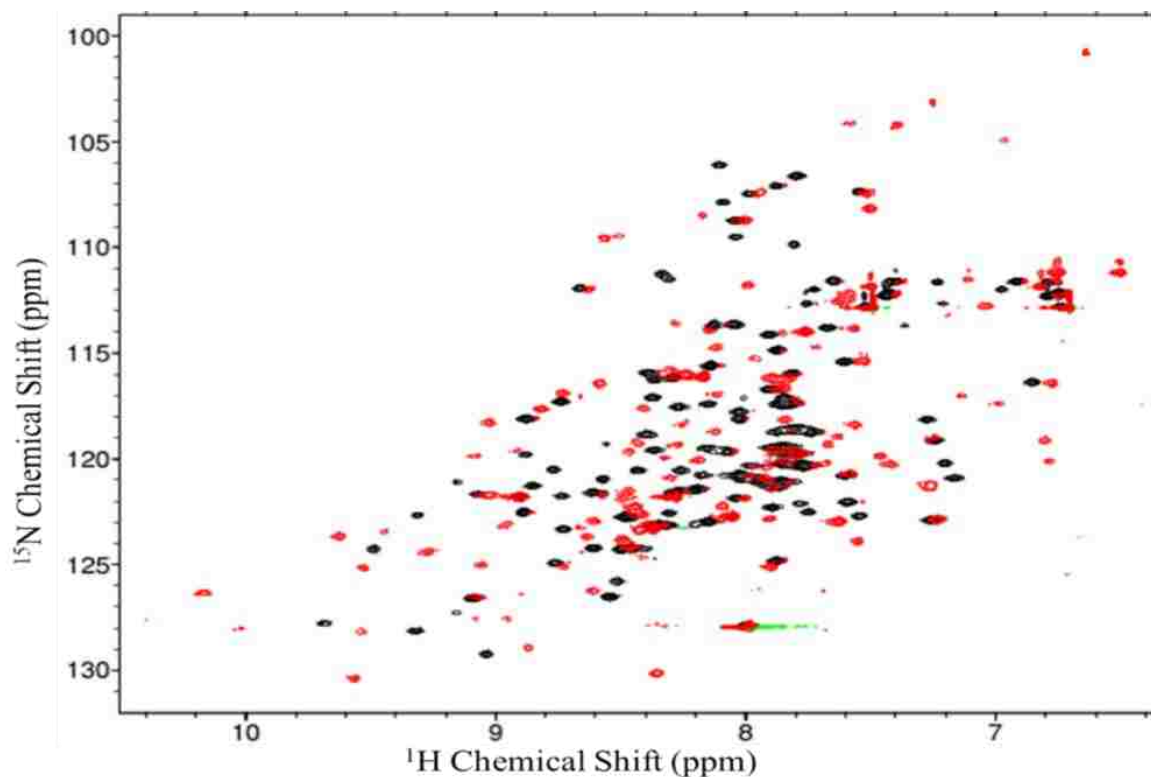
To determine if the use of sarkosyl affected the structure of Scc4,  $^{15}\text{N}$ -HisScc4 protein was expressed and purified (with no sarkosyl exposure) to compare its NMR spectrum with that of  $^{15}\text{N}$ -Scc4. The overlaid spectra of  $^{15}\text{N}$ -HisScc4 and  $^{15}\text{N}$ -Scc4 (Figure 24) show ~70% peak overlap, indicating that the use of sarkosyl in the isolation step did not change the backbone conformation of Scc4. The non-overlapped peaks show small shifts and are likely the effect of the hexahistidine tag on  $^{15}\text{N}$ -HisScc4.



**Figure 24.** Overlay of the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra for 0.50 mM  $^{15}\text{N}$ -HisScc4 (black/green for positive/negative signals, respectively) and  $^{15}\text{N}$ -Scc4 (red) recorded at 500 MHz (Bruker).

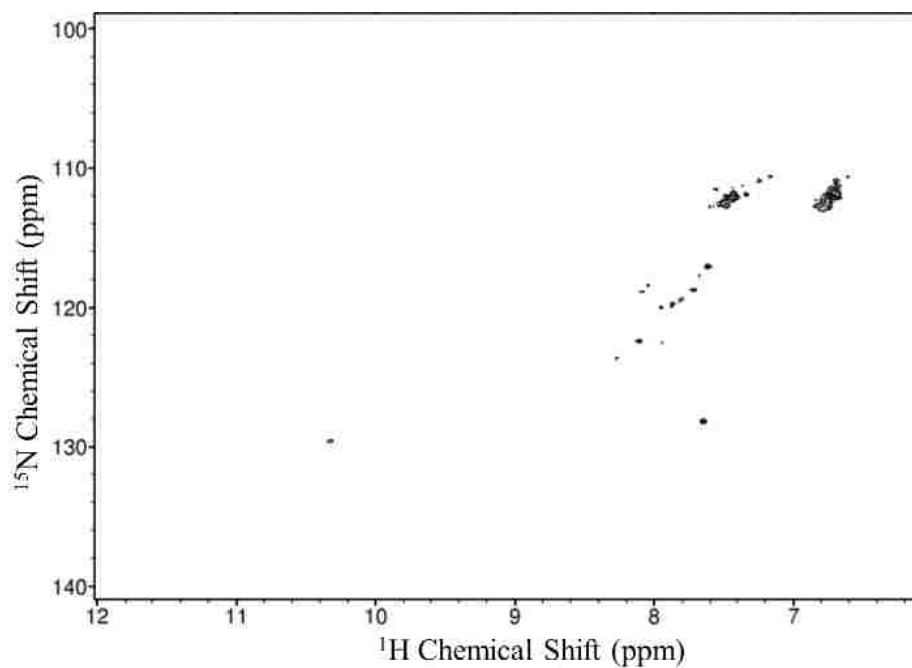
The spectra for the Scc1His: $^{15}\text{N}$ -Scc4 mixed complex and the untagged  $^{15}\text{N}$ -Scc4 were compared to determine the extent to which Scc4 changes when bound to Scc1. The minimal peak overlap (42%) observed in the Scc1His: $^{15}\text{N}$ -Scc4 and  $^{15}\text{N}$ -Scc4 spectra (Figure 25) implies that there is a conformational change when Scc4 binds to Scc1His. This change may be in accordance with the dual functionality of Scc4. The heterodimeric state is necessary for Scc4 to serve as a chaperon in the T3SS, while the homodimer Scc4 may be necessary for Scc4 to function as a transcriptional regulator of  $\sigma^{66}$ -mediated transcription.

Since the Scc4 homodimer and heterodimer assume different structures, it is necessary to study the Scc4 heterodimer and Scc4 homodimer structures separately.

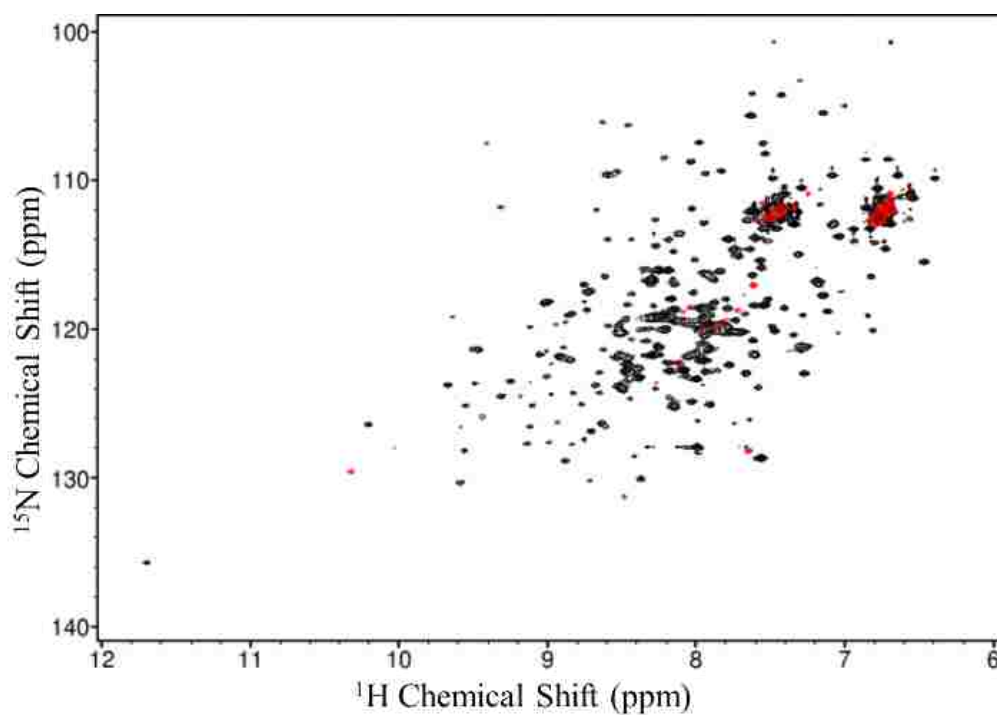


**Figure 25.** Overlay of the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra for Scc1His: $^{15}\text{N}$ -Scc4 (red) and  $^{15}\text{N}$ -Scc4 (black/green for positive/negative signals, respectively) recorded at 700 MHz (Varian).

Due to the limited solubility of Scc1His, it was not possible to acquire a meaningful  $^1\text{H}$ ,  $^{15}\text{N}$  -HSQC spectrum of  $^{15}\text{N}$ -Scc1His alone. The absence of many peaks for  $^{15}\text{N}$ -Scc1His in Figure 26 indicates that the protein is oligomerized or insoluble. The peaks observed are from flexible regions, particularly the amide-containing side chains. The NMR data was collected in the elution buffer because the protein precipitates in phosphate NMR buffer. Except for the clustered side chain peaks in Figure 27, there is no peak overlap between  $^{15}\text{N}$ -Scc1His (red) and  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 (black) spectra.



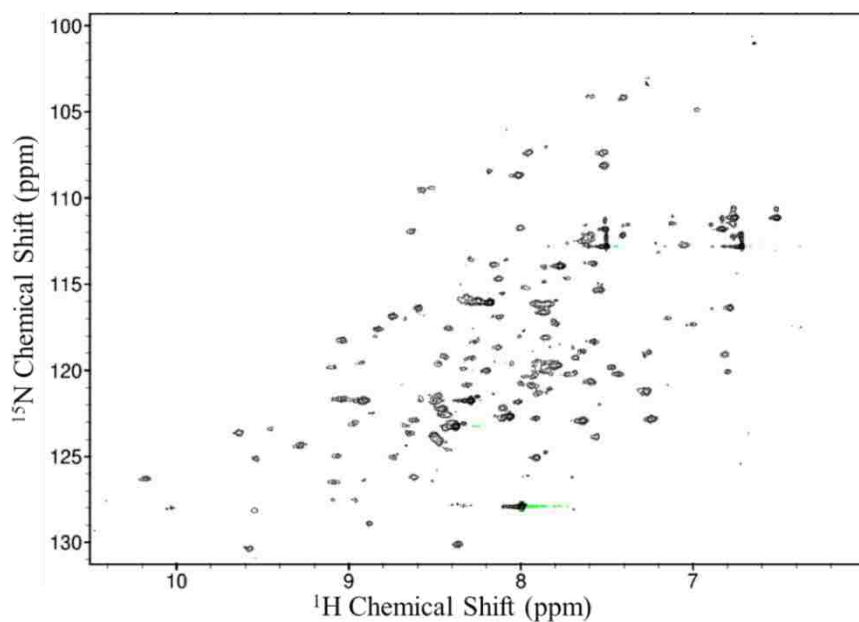
**Figure 26.**  $^1\text{H},^{15}\text{N}$ -HSQC spectrum for  $^{15}\text{N}$ -ScclHis (in elution buffer, with 10 mM DTT, 50  $\mu\text{M}$  DSS, and 10%  $\text{D}_2\text{O}$  recorded at 700 MHz (Varian).



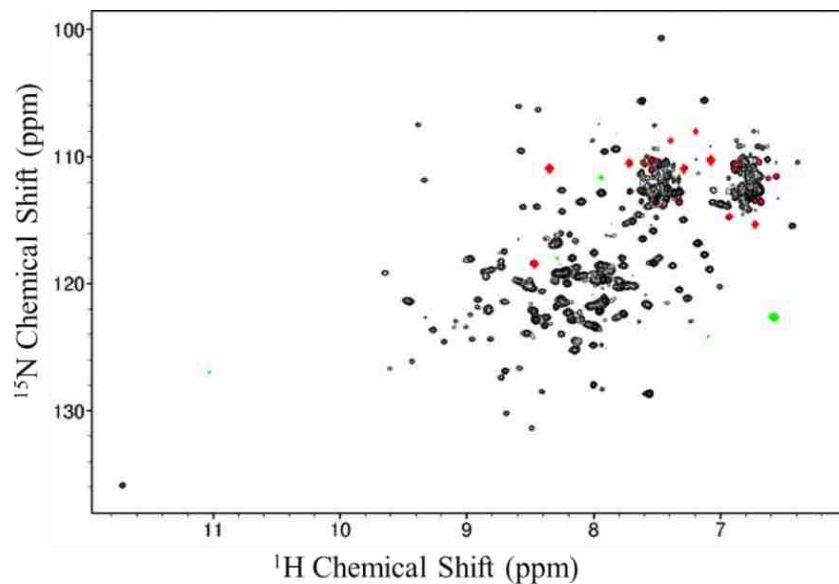
**Figure 27.**  $^1\text{H},^{15}\text{N}$ -HSQC overlay spectra of  $^{15}\text{N}$ -ScclHis (red, in 20 mM Tris pH 8.0, 150 mM NaCl, 250 mM imidazole, and 5% glycerol) and  $^{15}\text{N}$ -ScclHis: $^{15}\text{N}$ -Sccl4 (black) recorded at 700 MHz (Varian).

#### 2.4.4 NMR analysis of complexes with chain-selective isotopic labeling

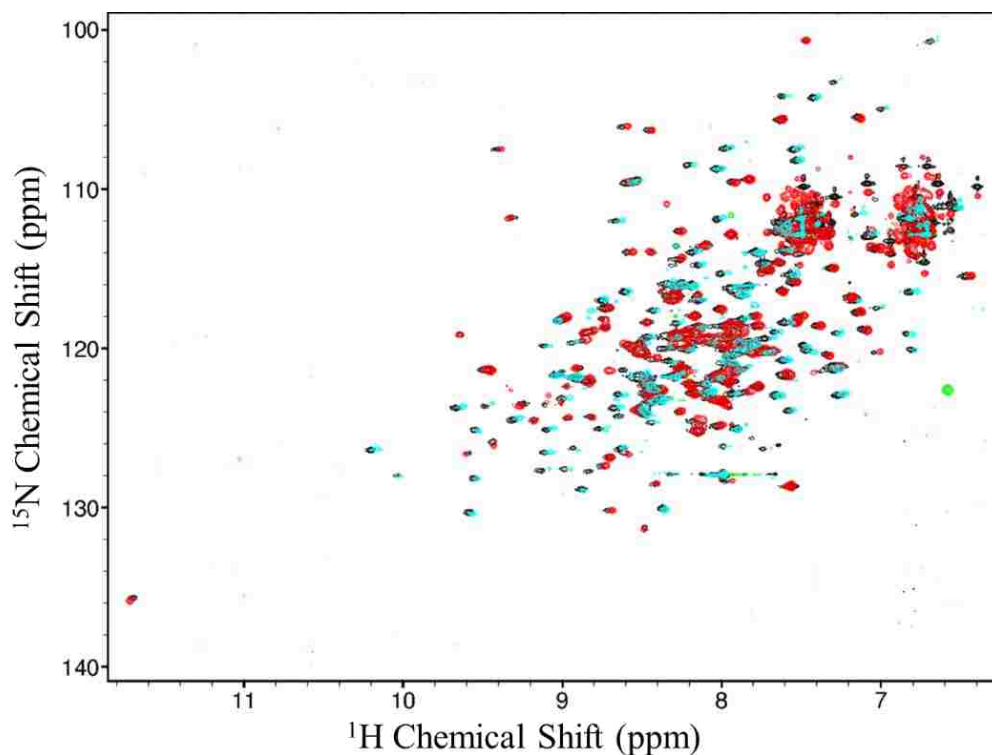
The successful formation of mixed heterodimer complexes was evaluated by comparing the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectra of  $\text{ScclHis}:\text{Sccl}^{15}\text{N}$ -Sccl4 (Figure 28) and  $^{15}\text{N}$ -Sccl1His:Sccl4 (Figure 29) with that of the  $^{15}\text{N}$ -Sccl1His: $^{15}\text{N}$ -Sccl4 (Figure 22) complex spectrum. The spectrum of  $^{15}\text{N}$ -Sccl1His:Sccl4 (Figure 29) shows many more peaks from Sccl1His when it is in complex with Sccl4 than the number of peaks observed in the spectrum of  $^{15}\text{N}$ -Sccl1His alone (Figure 26), which reflects the enhanced solubility of Sccl1His when complexed with Sccl4. Comparing the spectrum of  $^{15}\text{N}$ -Sccl1His:Sccl4 (Figure 29) with the spectrum of  $^{15}\text{N}$ -Sccl1His: $^{15}\text{N}$ -Sccl4 (Figure 22), there are additional peaks (colored red in Figure 29) that do not correspond with either free  $^{15}\text{N}$ -Sccl1His (Figure 26) or free  $^{15}\text{N}$ -Sccl4 (Figure 23). These peaks may originate from a contaminating protein or peptide. In Figure 30, the spectrum of the complex ( $^{15}\text{N}$ -Sccl1His: $^{15}\text{N}$ -Sccl4, black) is overlapped with the spectra of the two on-column formed mixed complex proteins ( $\text{Sccl1His}:\text{Sccl}^{15}\text{N}$ -Sccl4 and  $^{15}\text{N}$ -Sccl1His:Sccl4). The overlap of the spectra of the three protein samples confirms that the heterodimer complex and the mixed complexes have similar structural conformations. The spectra for the mixed labeled samples were used to uniquely assign peaks corresponding to  $^{15}\text{N}$ -Sccl1His and  $^{15}\text{N}$ -Sccl4 in the 2D  $^{15}\text{N}$ -HSQC spectrum of the  $^{15}\text{N}$ -Sccl1His: $^{15}\text{N}$ -Sccl4 complex.



**Figure 28.**  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum for 0.20 mM  $\text{Sccl1His}:\text{Sccl}^{15}\text{N}$ -Sccl4 recorded at 700 MHz (Varian). The black/green peaks correspond to positive/negative signals, respectively.



**Figure 29.**  $^1\text{H},^{15}\text{N}$ -HSQC spectrum for 0.2 mM  $^{15}\text{N}$ -ScclHis:Sccl4 recorded at 900 MHz (Bruker). The red colored peaks indicate peaks not observed in the  $^{15}\text{N}$ -ScclHis: $^{15}\text{N}$ -Sccl4 spectrum. The black/green peaks correspond to positive/negative signals, respectively.



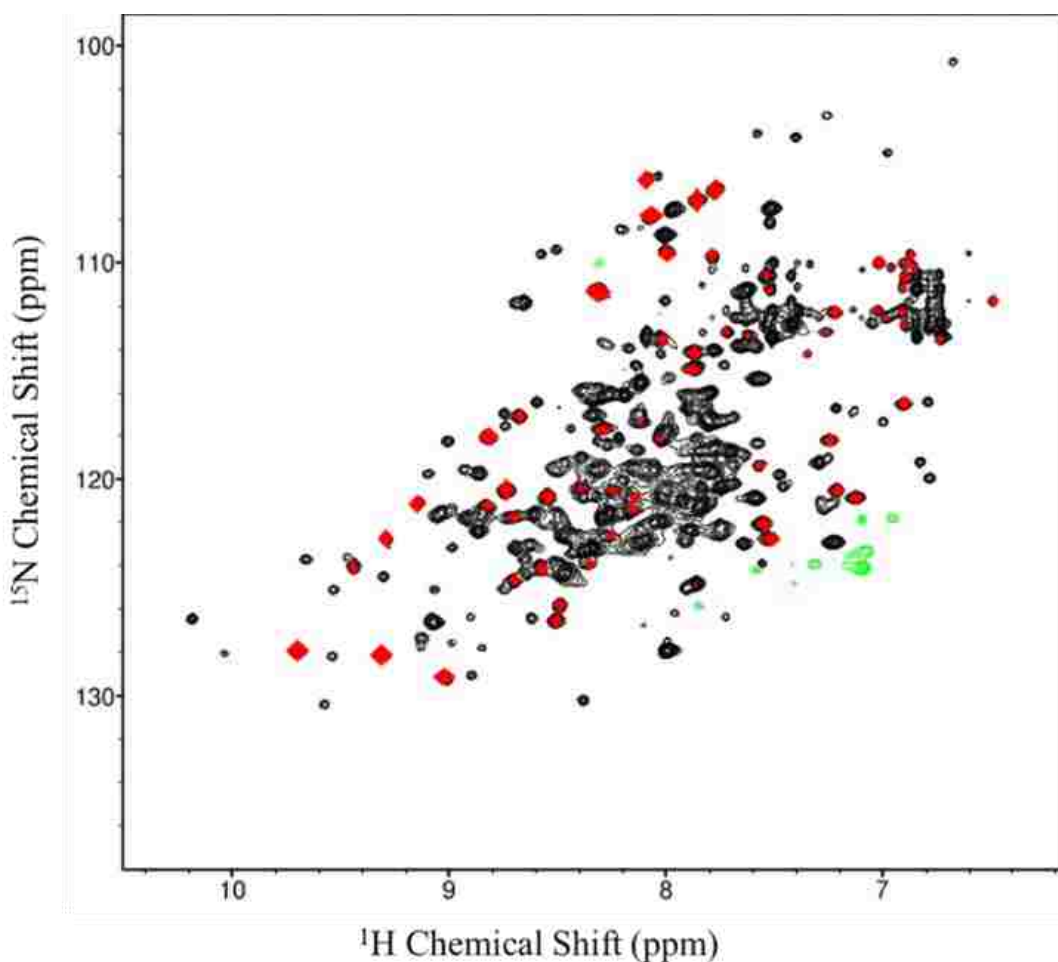
**Figure 30.**  $^1\text{H},^{15}\text{N}$ -HSQC spectra for  $^{15}\text{N}$ -ScclHis: $^{15}\text{N}$ -Sccl4 (black/green for positive/negative peaks, respectively) and ScclHis: $^{15}\text{N}$ -Sccl4 (cyan) in phosphate buffer recorded at 700 MHz (Varian), and  $^{15}\text{N}$ -ScclHis:Sccl4 (red) in Tris NMR buffer recorded at 900 MHz (Bruker). Tris buffer was used to prevent the precipitation of ScclHis, which was thought to have led to the release of  $^{15}\text{N}$ -Sccl4 (ScclHis is slightly soluble in Tris buffer).

It is expected that the on-column formed mixed complexes have a 1:1 ratio of Scc1His and Scc4 proteins. Early samples, such as those in Figure 30, only showed a few extra peaks in the  $^{15}\text{N}$ -Scc1His:Scc4 sample. However, free  $^{15}\text{N}$ -Scc4 peaks were observed in  $^1\text{H},^{15}\text{N}$ -HSQC spectrum of recent Scc1His: $^{15}\text{N}$ -Scc4 samples analyzed at NMRFAM. (Figure 34). To check the purity of the mixed complex samples (after elution from the Ni-column), Scc1His: $^{15}\text{N}$ -Scc4 and  $^{15}\text{N}$ -Scc1His:Scc4 elution fractions were separately loaded onto a size exclusion chromatography column (SEC 3-70), which separate proteins based on their molecular weights. The impurity present in the mixed complex samples is free  $^{15}\text{N}$ -Scc4 (not bound to Scc1His). Scc1His:Scc4 (33.5 kDa, heterodimer) and Scc4 (29.4 kDa, homodimer) have different elution volume/time from the SEC column because of their different molecular sizes. For comparison, the complex,  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4, was also purified by SEC. The chromatograms (Figure 32) show two peaks for the  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex and three peaks for each of the on-column formed mixed complexes. The SDS-PAGE analysis of the fractions obtained from the SEC column for a mixed complex sample (Figure 33), indicates that peak 1 is the higher oligomeric form of the complex (which eluted with some impurities), peak 2 is the heterodimer Scc1His: $^{15}\text{N}$ -Scc4 mixed complex, and the additional peak (peak 3 not observed in the  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex) is free  $^{15}\text{N}$ -Scc4. This confirmed that the extra peaks (red peaks Figure 31) observed in the Scc1His: $^{15}\text{N}$ -Scc4 NMR spectrum belonged to  $^{15}\text{N}$ -Scc4 homodimer. Tris buffer was used to prevent the precipitation of Scc1His, which was thought to have led to the release of  $^{15}\text{N}$ -Scc4 (Scc1His is slightly soluble in Tris buffer); however, extra  $^{15}\text{N}$ -Scc4 peaks are still visible. Additional purification by SEC is required for the mixed complexes after the initial Ni-column purification.

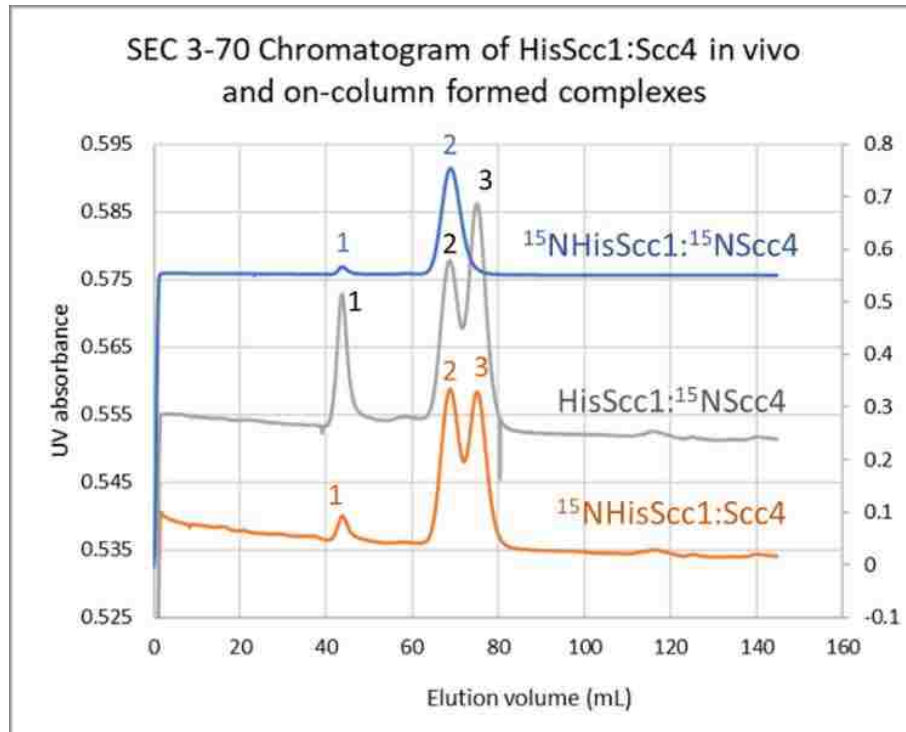
To further confirm that the extra peaks in Scc1His: $^{15}\text{N}$ -Scc4 sample were for free  $^{15}\text{N}$ -Scc4, the NMR spectra of  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 (black), Scc1His: $^{15}\text{N}$ -Scc4 (red), and  $^{15}\text{N}$ -Scc4 (lime green) were overlaid. Figure 34 shows the two forms of Scc4 in the Scc1His: $^{15}\text{N}$ -Scc4 sample. The homodimer Scc4 (not in complex) is shown by the red intense peaks overlapped by the lime green peaks, and the heterodimer form (in complex with Scc1His) is shown by the red less intense peaks overlapping the black peaks.



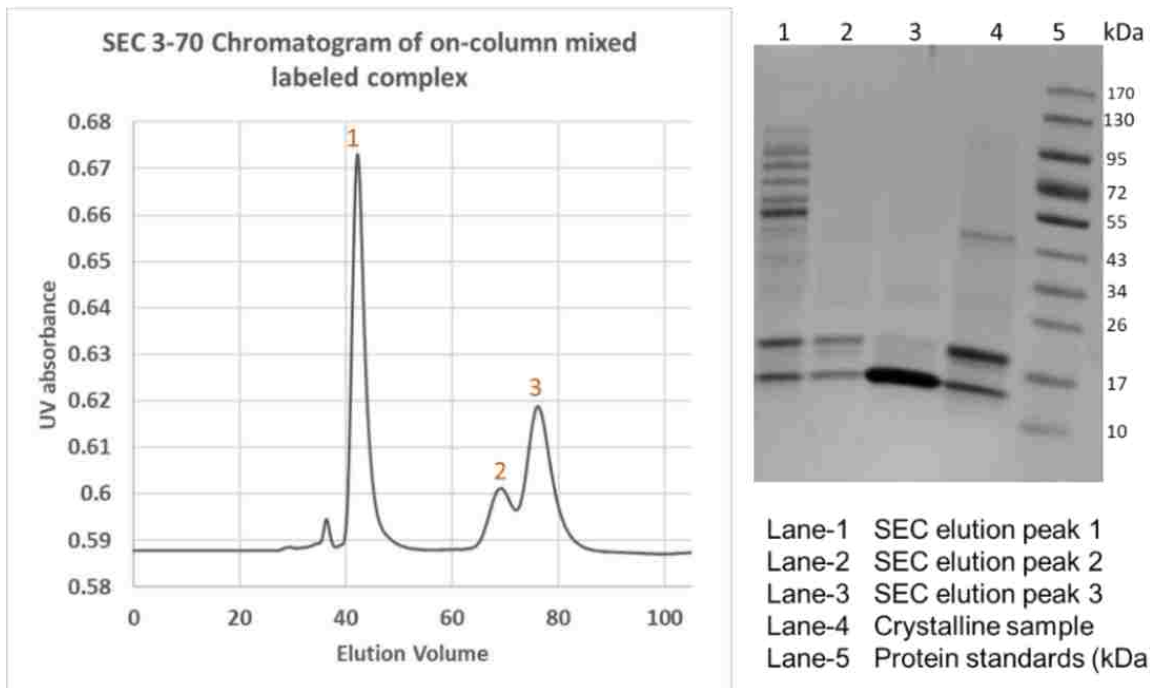
The presence of Scc4 homodimer in the eluted Scc1His:<sup>15</sup>N-Scc4 mixed complex was unexpected. The binding ratio of Scc1His and Scc4 is 1:1, and it is unusual for Scc1His to trap excess Scc4 from the lysate during the formation of the mixed complex. There is a possibility that the on-column complex is Scc1His:Scc4 mixed complex secondarily bound to Scc4 homodimer. The secondary binding may fail during centrifugation or when the sample buffer is changed leading to the presence of free Scc4 in solution. A second possibility is non-specific binding of Scc4 to the Ni-column, which elutes together with the mixed complex. The third possibility is that some Scc1His dissociates from the complex and precipitates leaving free Scc4, which is soluble and visible in both NMR and SEC.



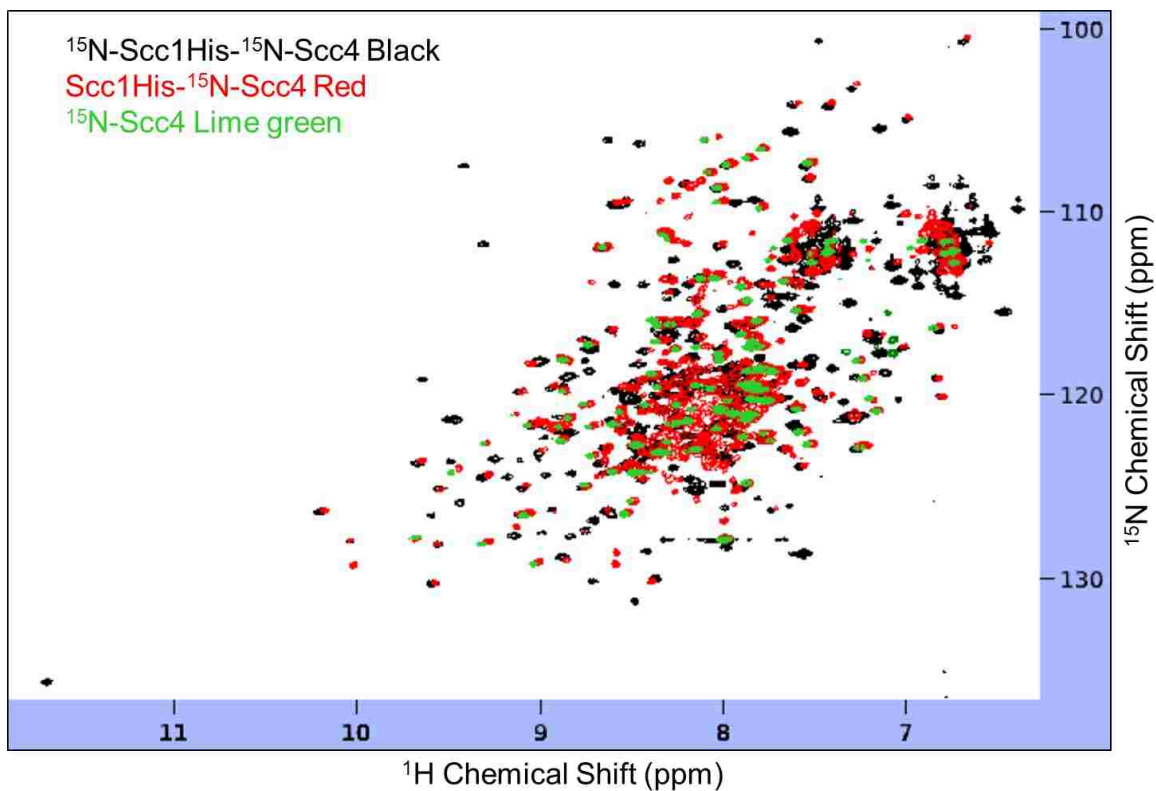
**Figure 31.** <sup>1</sup>H,<sup>15</sup>N-HSQC spectrum of 0.2 mM Scc1His:<sup>15</sup>N-Scc4 (black/green for positive/negative peaks, respectively) in Tris buffer (20 mM Tris, 150 mM NaCl, 0.02% NaN<sub>3</sub> pH 7.4) recorded on 800 MHz NMR spectrometer (Agilent). The extra peaks colored red are for the free Scc4.



**Figure 32.** Size exclusion chromatograms for the  $^{15}\text{N}$ -Sc1His: $^{15}\text{N}$ -Sc4 complex and the on-column formed complexes (Sc1His: $^{15}\text{N}$ -Sc4 and  $^{15}\text{N}$ -Sc1His: Sc4) in 50 mM phosphate buffer, pH 7.4.



**Figure 33.** SEC chromatogram of the on-column mixed labeled complex, Sc1His: $^{15}\text{N}$ -Sc4 (after isolating the crystalline portion), and the SDS-PAGE analysis of the SEC fractions.

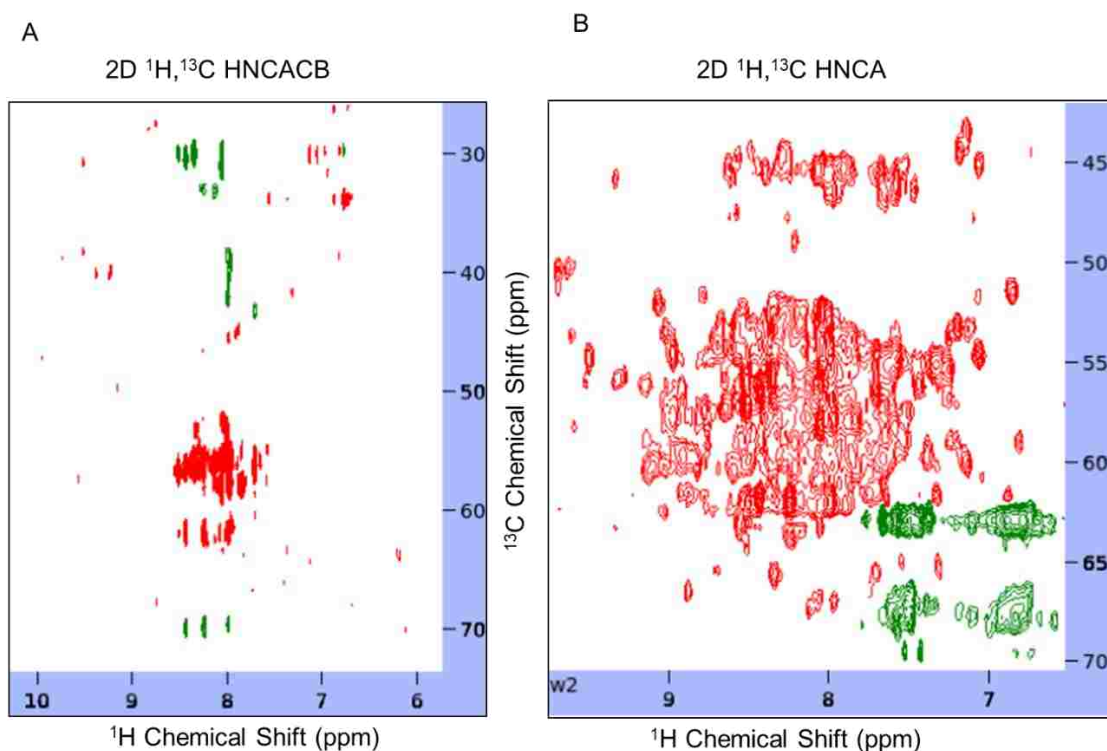


**Figure 34.**  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC overlay spectra of  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 (black), Scc1His: $^{15}\text{N}$ -Scc4 (red), and  $^{15}\text{N}$ -Scc4 (green).

After eluting the on-column mixed complex samples from the Ni-column, the fractions are combined and concentrated using molecular weight cut-off centrifugal filters. During centrifugation, tiny crystals were observed in the residual sample. The crystalline portion was separated and analyzed by SDS-PAGE. The gel image, lane-4 Figure 33, indicates that the crystals contain the Scc1His:Scc4 complex. These crystals will be tested for diffraction in collaboration with Dr. Marcia Newcomer, Department of Biological Sciences, LSU.

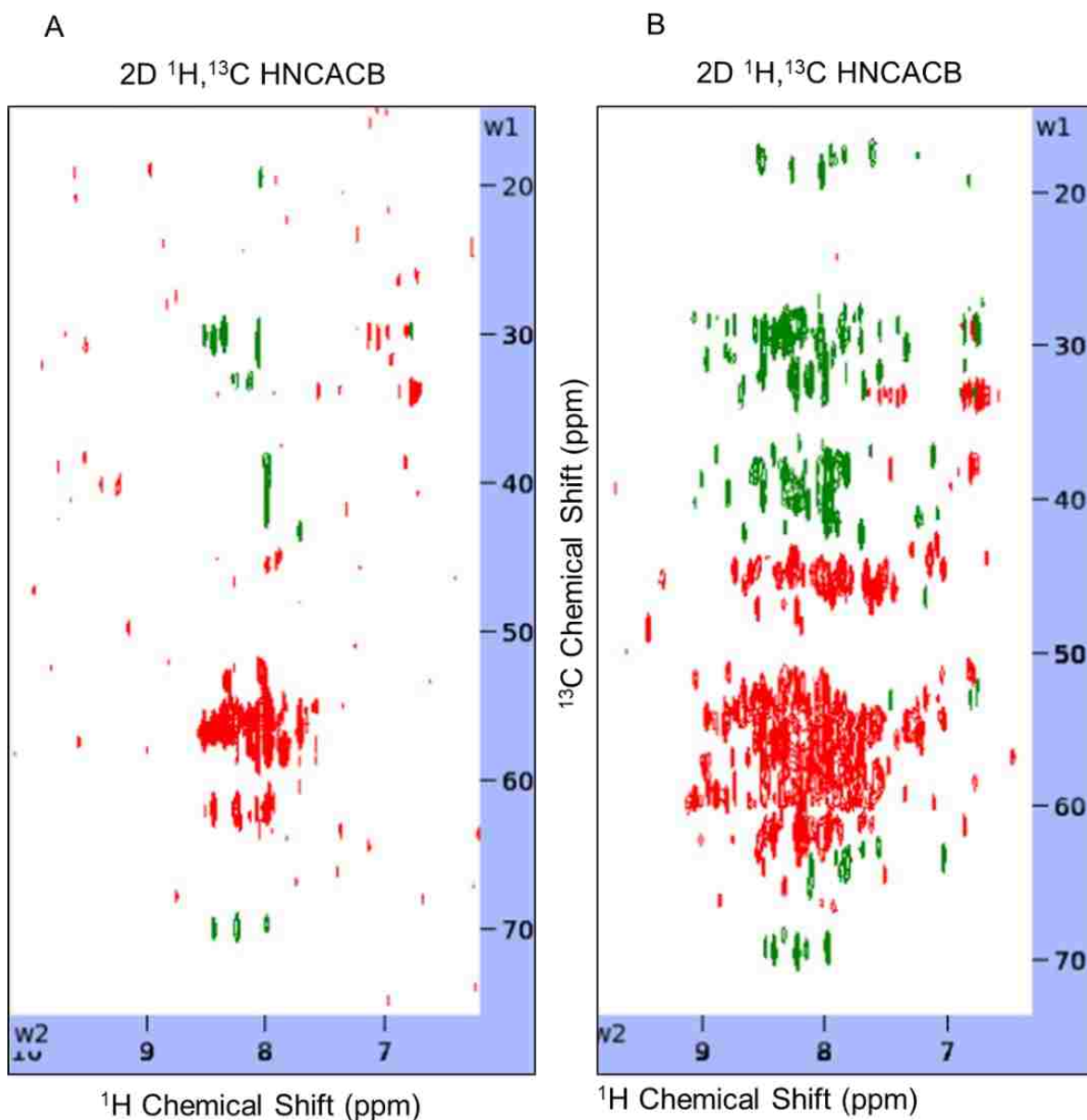
Before carrying out protein backbone NMR experiments, it was necessary to check the  $^{13}\text{C}$  signal intensity in the protein samples. Due to the size of the Scc1His:Scc4 complex (about 33 kDa), the double labeled  $^{15}\text{N}$ ,  $^{13}\text{C}$ -Scc1His: $^{15}\text{N}$ ,  $^{13}\text{C}$ -Scc4 did not give strong signals for  $\text{C}\beta$  (green peaks) nuclei in the 2D HNCACB spectrum, shown in Figure 35A. However, there was a relatively good signal for  $\text{C}\alpha$  (red peaks) nuclei in the HNCA 2D spectrum (Figure 35B). It was determined that a deuterium-labeled complex was necessary to enhance the  $^{13}\text{C}$  signal (Figure 36). The spectrum for the double labeled complex (Figure 36A) has a weak  $^{13}\text{C}$  signal compared to the triple labeled complex (Figure 36B). Large proteins with slow

tumbling rates have short transverse relaxation times ( $T_2$ ), which produces broad signals with low intensity. Deuteration of proteins minimizes cross-relaxation, leading to increased  $^{13}\text{C}$  signal compared to protonated proteins.



**Figure.35.**  $^1\text{H},^{13}\text{C}$ -HSQC spectra for comparison of the  $^{13}\text{C}$  signal intensity in A) 2D HNCACB and B) 2D HNCA spectra of the double labeled  $^{15}\text{N},^{13}\text{C}$ -ScclHis: $^{15}\text{N},^{13}\text{C}$ -Sccl4 protein complex.

The  $^1\text{H},^{15}\text{N}$ -HSQC spectrum for the triple labeled protein sample was not well resolved; however, the 2D  $^1\text{H},^{15}\text{N}$  HNCA spectrum (Figure 37) was used in place of the  $^1\text{H},^{15}\text{N}$ -HSQC spectrum for automated backbone assignment. The selected peaks (overlaid with black spots) in the 2D  $^1\text{H},^{15}\text{N}$  HNCA spectrum belong to Sccl4 in the complex, identified through restricted peak picking with respect to the  $^1\text{H},^{15}\text{N}$ -HSQC spectrum of ScclHis: $^{15}\text{N}$ -Sccl4. Restricted peak picking on 3D spectra with respect to the Sccl4 peaks in 2D  $^1\text{H},^{15}\text{N}$  HNCA spectrum (Figure 37) was performed using the automated PINE<sup>68</sup> sparky assignment (PINE-SPARKY2) software to give the  $^1\text{H},^{15}\text{N}$ -HSQC plot in Figure 38. The assignments may not be accurate because a small data set (2D  $^1\text{H},^{15}\text{N}$ -HNCA, HNCA, HN(CO)CA, HNCOS spectra) was used. More 3D data is needed to improve the assignments, and assignments will be checked manually. The backbone assignment output (Figure 38) corresponds to the Sccl4 peaks selected in the 2D  $^1\text{H},^{15}\text{N}$ -HNCA spectrum. A similar procedure can be followed to conduct backbone assignments for ScclHis.

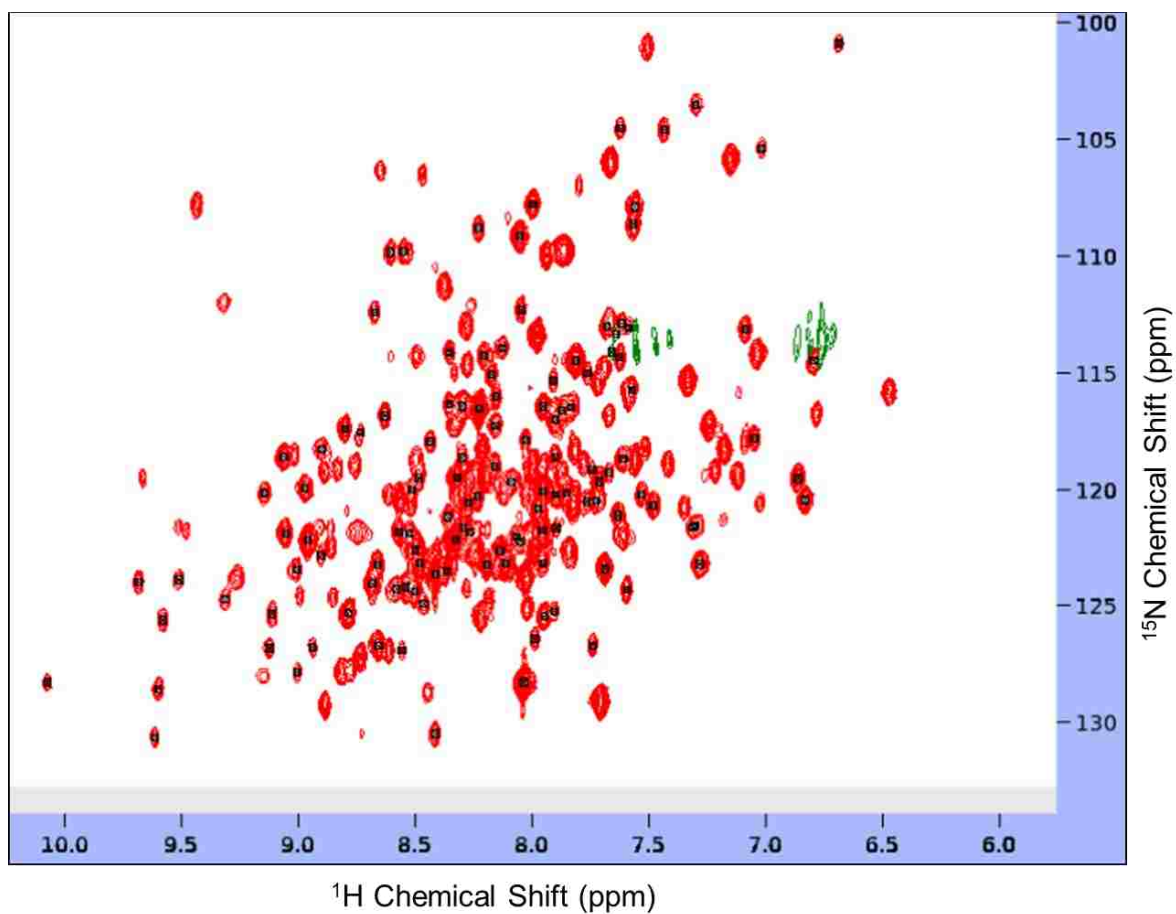


**Figure 36.**  $^1\text{H},^{13}\text{C}$ -HSQC spectra for comparison of  $^{13}\text{C}$  signal intensity in the 2D HNCACB spectra of the (A) double labelled ( $^{15}\text{N},^{13}\text{C}$ ) and (B) triple labelled ( $^2\text{H},^{15}\text{N},^{13}\text{C}$ ) Scc1His:Scc4 protein complex

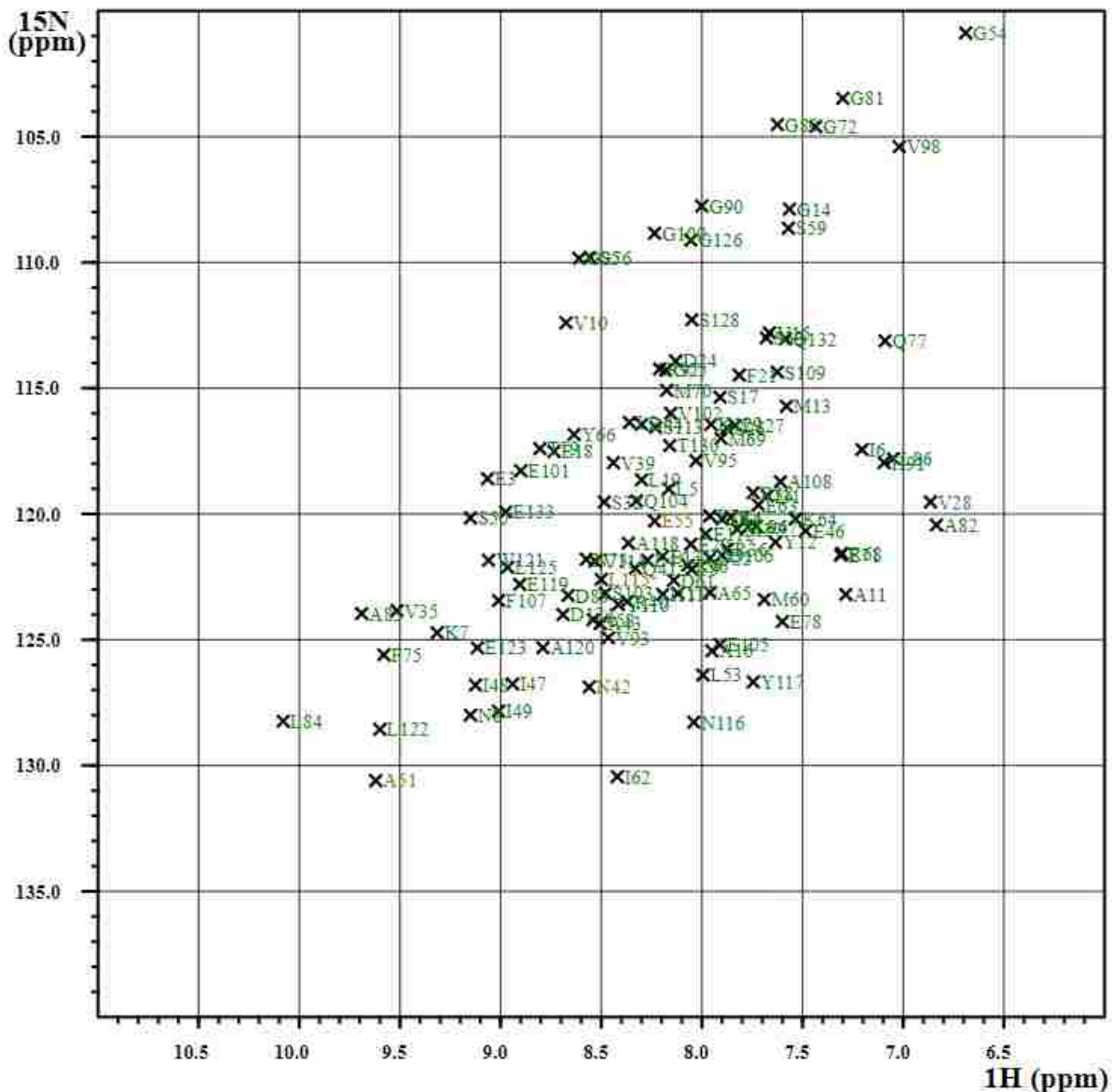
## 2.5 Conclusion

A method to express and isolate Scc4 from the Scc1His:Scc4 complex has been developed. An on-column protocol for the formation of the mixed labeled protein complexes has been presented. Two-dimensional NMR spectra for Scc4, HisScc4, Scc1His:Scc4, and the mixed labeled complexes have given structural insights on the proteins. The structure for Scc4 in the heterodimer complex with Scc1His differs from the homodimer Scc4. Sarkosyl does not change the backbone structure of Scc4, and there is a slight

change in Scc4's structure when a hexahistidine tag is present at the *N*-terminus of the protein. The on-column mixed complexes have the same structure as the *in vivo* formed complex. The 2D  $^1\text{H},^{15}\text{N}$ -HSQC spectra of the mixed labeled samples were used to identify peaks for each protein in the 2D  $^1\text{H},^{15}\text{N}$ -HSQC of the  $^{15}\text{N}$ -Scc1His: $^{15}\text{N}$ -Scc4 complex, where both proteins were labeled. These initial assignments will make triple resonance backbone assignments feasible.



**Figure 37.** 2D HNCA spectrum for deuterated  $^{15}\text{N},^{13}\text{C}$ -Scc1His: $^{15}\text{N},^{13}\text{C}$ -Scc4 recorded on a 900 MHz spectrometer (Bruker). The selected peaks are for Scc4 identified through restricted peak picking with respect to  $^1\text{H},^{15}\text{N}$ -HSQC spectrum of the Scc1His: $^{15}\text{N}$ -Scc4 mixed complex.



**Figure 38.** Illustration of how restricted peak picking can be applied to Scc1His:Scc4 using the  $^1\text{H},^{15}\text{N}$ -HSQC spectrum to assign peaks belonging to Scc4 alone. Featured is the I-PINE HSQC plot for sparky automated assignment using 2D  $^1\text{H},^{15}\text{N}$ -HNCA, HNCA, HN(CO)CA, HNC(O) spectra for  $^{15}\text{N},^{13}\text{C}$ -Scc1His: $^{15}\text{N},^{13}\text{C}$ -Scc4 (the assignments may not be accurate, as more data and manual verification are needed).

**CHAPTER 3**  
**STRUCTURAL MODIFICATION OF THE TRIPEPTIDE KPV BY REDUCTIVE GLYCOALKYLATION**

**3.1 Alpha Melanocyte Stimulating hormone ( $\alpha$ -MSH 1-13) and the KPV ( $\alpha$ -MSH 11-13) tripeptide motif**

Short linear peptide motifs of less than 10 amino acid residues have the ability to modulate biological processes such as localization of proteins, post-translational modification, and ligand-receptor interactions.<sup>100</sup> These peptide motifs form the basis for small molecule peptidomimetic drugs. The smallest peptide motifs reported to have biological activities have three contiguous amino acids, a minimum of 25 heavy atoms.<sup>101</sup> Ung *et al.* have reviewed endogenous tripeptides and tripeptide motifs embedded in large protein sequences,<sup>11</sup> with varying biological activities summarized in Table 6. The activities reported for these tripeptides, indicate that a complete protein sequence may not be necessary for some signaling or regulatory activities to occur. The tripeptide KPV which was selected for this study, is derived from the C-terminal amino acids of alpha melanocyte stimulating hormone ( $\alpha$ -MSH), whose sequence is contained in proopiomelanocortin hormone.<sup>12</sup>

**Table 6.** Endogenous tripeptides and tripeptide motifs.<sup>11</sup>

Endogenous tripeptides	Tripeptide motifs
ECG, (glutathione), antioxidant, cofactor	DLF/SLF, inhibition of $\beta$ protein of bacterial replisome, antimicrobial
EHP, stimulates pituitary gland controlling thyroid-stimulating hormone secretion	ELR, chemokine, growth factor binding motif GGQ, release factor, stop codon recognition
FEG, inhibition of anaphylaxis, anti-inflammatory, modulates cardiac leukocyte adhesion	GPE, neuroprotection HAV, cadherin motif, cell-cell interactions, and adhesion
GHK, tissue remodeling and wound healing	HGK, vitronectin inhibition
PLG, modulator of the dopamine D2receptor	HPQ, streptavidin binding motif KPV, anti-inflammatory properties
	LDV, vascular cell adhesion molecule 1/ fibronectin adhesion motif
	RGD, cell adhesion signal and modulation of thrombosis
	SKL, peroxisomal targeting

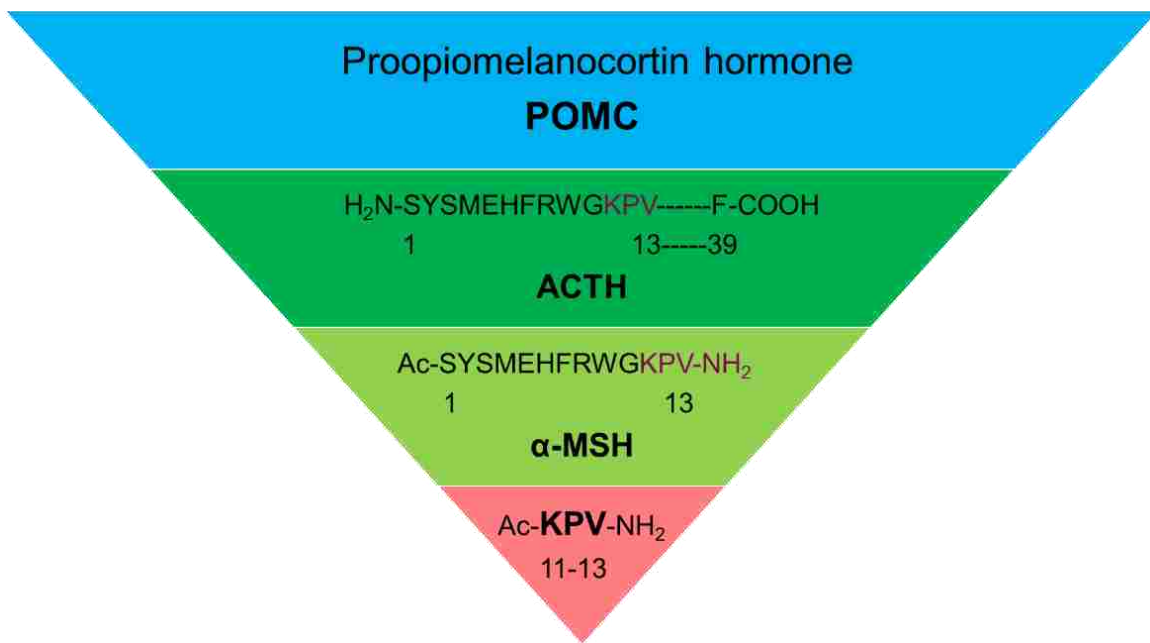
---

Portions of this chapter previously appeared in PLOS ONE, Songok, *et al.* PLoS ONE 2018, 13, e0199686.



Proopiomelanocortin hormone is composed of 267 amino acids.<sup>102</sup> Proopiomelanocortin (POMC) derived peptides are produced in a number of peripheral tissues including reproductive organs, anterior and intermittent lobes of the pituitary, arcuate nucleus of the hypothalamus neurons, and neurons in the commissural nucleus of the solitary tract.<sup>103</sup> The genes for *pomc* consist of 2 introns and 3 exons. Exon 2 and 3 encode POMC which is processed to hormonal peptides including adrenocorticotrophic releasing hormone (ACTH),  $\beta$ -lipotropin ( $\beta$ -LPH), melanocyte stimulating hormone (MSH), and  $\beta$ -endorphin ( $\beta$ -EP). These peptides influence various physiological processes including food uptake, pigmentation, stress response, glucose balance, behavior, and energy homeostasis.<sup>104</sup>

Proopiomelanocortin preprotein undergoes multiple types of post translational processing such as glycosylation and cleavage by the enzymes prohormone convertase 1, 2, or 3 (PC1, PC2, and PC3) to give  $\alpha$ -,  $\beta$ -, and  $\gamma$ -MSH functional peptides.<sup>104</sup> Alpha melanocyte stimulating hormone was first discovered as a pigment-inducing peptide.<sup>12</sup> Subsequent studies revealed multiple functions of this hormone including anti-inflammatory and antimicrobial effects.<sup>105, 106</sup> Cleavage of POMC by PC1 to ACTH followed by cleavage of ACTH by PC2/7B2 (cofactor) leads to the formation of  $\alpha$ -MSH. Further processing of  $\alpha$ -MSH through C-terminal amidation and N-terminal acetylation is needed for full activity. Alpha-MSH production from POMC and the terminal acetylated/amidated sequence is shown in Figure 39.



**Figure 39.** C-terminal amidated and N-terminal acetylated  $\alpha$ -MSH.

Melanocortin peptides signal through the seven transmembrane domain G-protein-coupled receptors (GPCRs); classified into five subtypes commonly referred to as melanocortin receptor 1 through 5 (MC1R, MC2R, MC3R, MC4R, and MC5R).<sup>107</sup> Melanocortin receptors are expressed in different cell types of the central nervous system or peripheral tissues. Stimulation of these receptors accomplish different functions depending on the cell types as indicated in Figure 40.

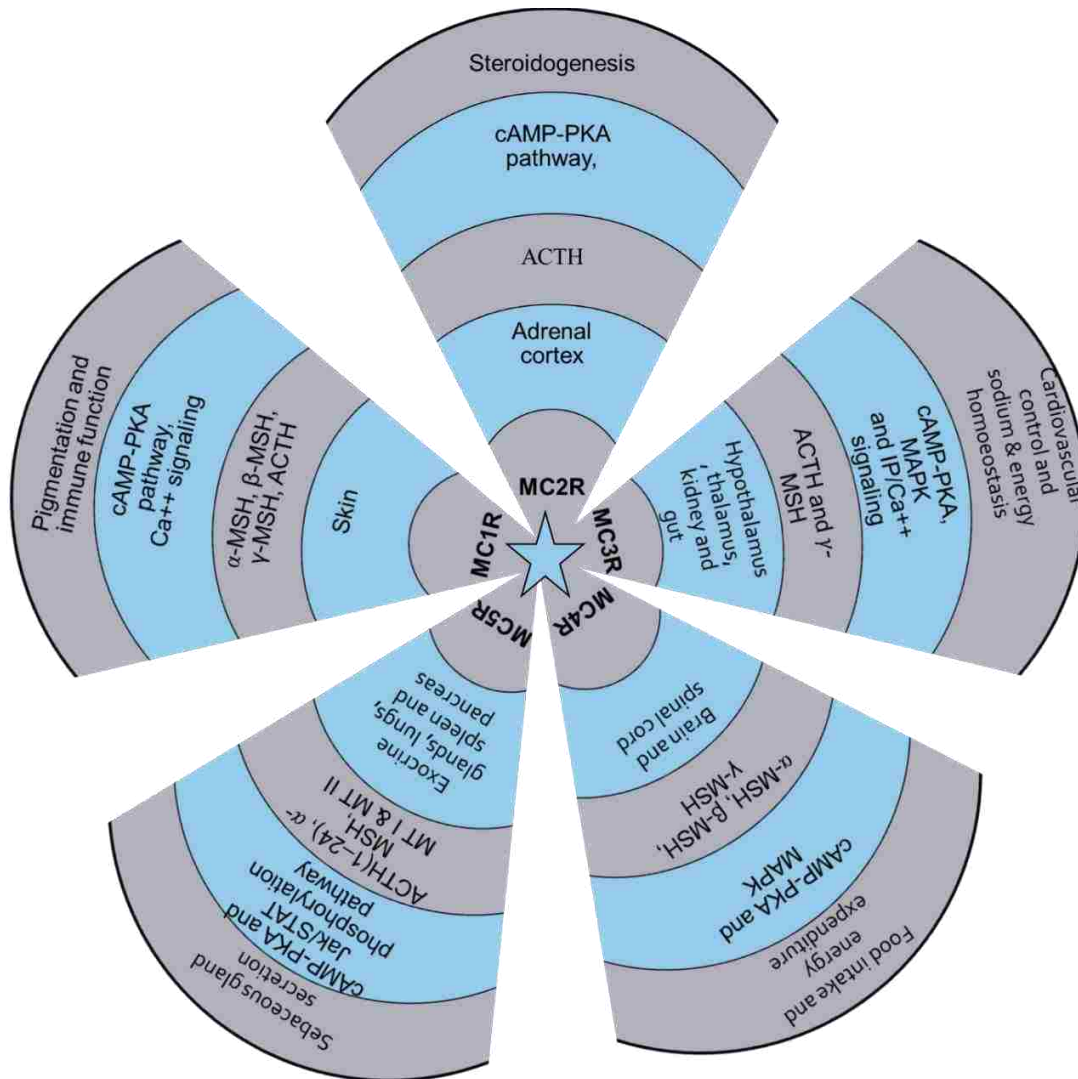
The core sequence His-Phe-Arg-Trp (HFRW) present in  $\alpha$ -,  $\beta$ -,  $\gamma$ -MSH, and ACTH stimulates the MCRs for signal transduction through different signaling pathways shown in Figure 40. Melanocortin receptors are unique in that they have both endogenous agonist and endogenous antagonists. Agouti signaling protein (ASIP) and Agouti related protein (AgRP) are the two endogenous antagonists of MC1R, MC3R, and MC4R. Agouti signaling protein is expressed at the hair follicle and it inhibits the stimulation of MC1R by  $\alpha$ -MSH at the surface of melanocytes leading to a shift of pigmentation color from eumelanin (black/brown) to phaeomelanin (red/yellow).<sup>108-110</sup> Agouti signaling protein has high affinity for MC1R and therefore inhibits adenylyl cyclase stimulation which in turn prevent the synthesis of eumelanin. AgRP, on the other hand, binds to MC3R and MC4R for metabolism and body weight regulation.<sup>111</sup>

### **3.2 Anti-inflammatory and antimicrobial activity of $\alpha$ -MSH**

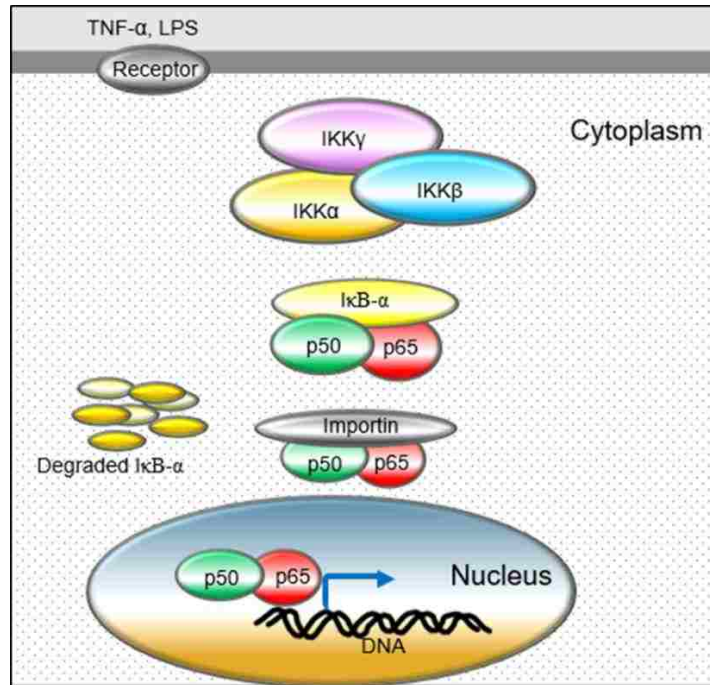
Alpha melanocyte stimulating hormone inhibits the NF $\kappa$ B inflammatory pathway.<sup>105</sup> Nuclear factor kappa B is a heterodimer complex of the Rel family of proteins (p65, cRel, RelB, p50, and p52) that binds to the DNA sequence,  $\kappa$ B.<sup>112</sup> The heterodimer p65/p50 is activated in the canonical pathway, shown in Figure 41 and translocate to the nucleus leading to expression of molecules such as cell adhesion molecules, cytokines, and acute phase response proteins.<sup>112</sup> Inflammatory agents such as TNF- $\alpha$ , LPS, IL1, or CD40L activate this pathway through their respective receptors. The inflammation signal leads to phosphorylation of inhibitor  $\kappa$ B- $\alpha$  (I $\kappa$ B- $\alpha$ ) in the inactive cytoplasmic complex I $\kappa$ B $\alpha$ / p65/p50 by I $\kappa$ B- $\alpha$  kinase (IKK $\alpha$ ); followed by ubiquitination and degradation.<sup>105</sup> This process exposes the nuclear translocation signal in p65 leading to the translocation of the heterodimer complex to the nucleus and the onset of the transcription process.

In addition to the pigmentary effects,  $\alpha$ -MSH has anti-inflammatory activity. It signals through the seven-transmembrane G-protein coupled receptors MC1R, MC3R-MC5R. Upon binding to the MC1R,

adenylyl cyclase is activated leading to amplification of cytoplasmic cyclic adenosine monophosphate (cAMP). The cAMP activates protein kinase A (PKA) which leads to upregulation of cAMP responsive binding element (CREB) and microphthalmia (MITF). In melanocytes, this process leads to increased production of the enzyme tyrosinase which initiates melanogenesis by oxygenation of the amino acid tyrosine to produce dopaquinone.<sup>110</sup>



**Figure 40.** Distribution and function of melanocortin receptors.<sup>20, 108</sup> Mitogen activated protein kinase (MAPK), Protein kinase A (PKA), inositol phosphate (IP), cyclic AMP (cAMP), and Janus kinase/signal transducers and activators of transcription (Jak/STAT).



**Figure 41.** Canonical NFκB activation pathway

Increased concentration of cAMP leads to different responses in different cells. In the inflamed cells, α-MSH inhibits the expression of proinflammatory cytokines, through induction of cAMP.<sup>105</sup> Increased cAMP prevents phosphorylation of inhibitor kappa B alpha (IκB-α) in the IκBα/ p65/p50 complex by IκBα kinase (IKK).<sup>20</sup> The downstream process involves the inhibition of the transcription regulator nuclear factor kappa B (NFκB, aka p65/p50 complex) from translocating to the nucleus; hence preventing the expression on NFκB target genes with respect to the inflammatory agents. α-MSH down regulates the production of interleukin-8 (IL-8), nitric oxide, interferon-γ (IFN-γ), interleukin-1β (IL-1β), monocyte chemoattractant protein 1 (MCP-1), tumor necrosis factor α (TNF-α), and limits the inflammatory cells trafficking into sites of injury.<sup>113</sup>

Several immunomodulatory peptides such as neuropeptide Y, neurokinin 1, and proenkephalin A have been shown to additionally have antimicrobial properties (AMPs).<sup>20</sup> On the other hand, AMPs including human cathelicidin LL-37 and human β defensin 2 (HBD2) exhibit immunomodulatory properties.<sup>114</sup> Besides its antipyretic and anti-inflammatory properties, α-MSH has potent antimicrobial activity, depicted by the α-helical structure, cationic nature, and occurrence in defense cells (keratinocytes, melanocytes, and fibroblasts). The occurrence of α-MSH and its receptor in both defense cells and immune cells (monocytes,

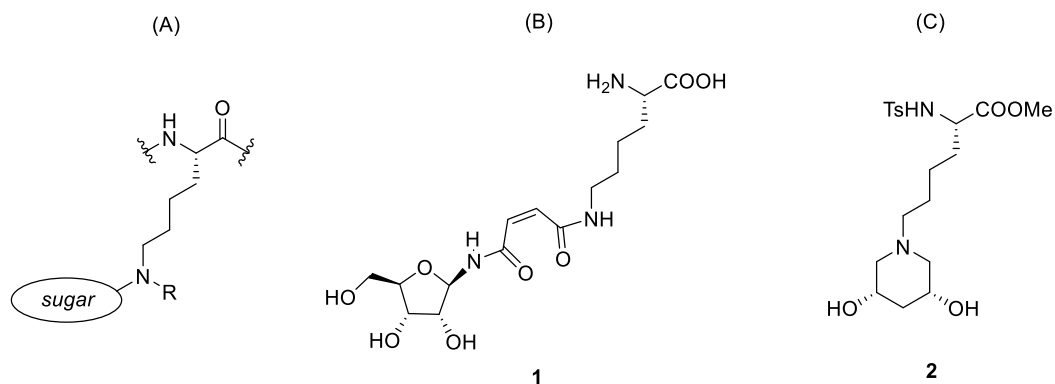
neutrophils, and macrophages)<sup>106</sup> supports the dual functionality as an immunomodulatory and antimicrobial peptide. Alpha MSH is known to have a broad spectrum of antimicrobial activity against *Escherichia coli* (*E. coli*), *Staphylococcus aureus* (*S. aureus*), and *Candida albicans* (*C. albicans*).<sup>11, 115</sup>

Madhuri *et al.* identified membrane permeation as the mechanism by which  $\alpha$ -MSH exerts its antimicrobial activity against *S. aureus*, which paralleled cell permeation by gramicidin.<sup>106</sup> On the other hand, Cutuli *et al.* demonstrated that the killing of *C. albicans* by  $\alpha$ -MSH is mediated through induction of cAMP, comparable to forskolin,<sup>116</sup> which confirmed an earlier report that cAMP enhancement inhibits protein and mRNA synthesis in *C. albicans*.<sup>117</sup> Adenylyl cyclase inhibitor dideoxyadenosine attenuated the candidicidal activity of  $\alpha$ -MSH.

The antimicrobial activity of  $\alpha$ -MSH 1-13 lies in the C-terminal tripeptide KPV ( $\alpha$ -MSH 11-13), which has been shown to have similar potency to full length  $\alpha$ -MSH (1-13).<sup>116</sup> Since KPV lacks the pigmentary effect induced by  $\alpha$ -MSH,<sup>118</sup> it is the preferred candidate for antimicrobial and anti-inflammatory drug development. A novel structural modification of therapeutic peptides by reductive “glycoalkylation” was devised in this study and applied to the tripeptide Ac-KPV-NH<sub>2</sub>. The tripeptide was synthesized in solution and a D-glucose derived aldehyde used to modify the lysine residue at the  $\alpha$ - and  $\epsilon$ -amine. The modification is expected to stabilize the tripeptide against enzymatic degradation and to enhance solubility.

### 3.3 Introduction to “glycoalkylation” and aminocyclization reactions

The term “glycoalkylation” has been used to imply the modification of lysine with a generic sugar, Figure 42A. It was earlier used by Schlimme *et al.* when describing the *N*-glycosylation of mono- and bicyclic dicarboxylic acid imides using ribose.<sup>119</sup> Lysine was glycoalkylated at *N* $\epsilon$  through a ring opening reaction with *N*-ribosylated maleimide, to afford compound **1** (Figure 42B). The term “glycoalkylation” is redefined here to mean the capping of *N* $\alpha$  or *N* $\epsilon$  amine groups of lysine in H-KPV-NH<sub>2</sub> with a piperidine diol derived from D-glucose, *viz.* compound **2** (Figure 42C). This modification is expected to enhance the pharmacokinetic properties of peptide drugs by conferring stability towards proteolytic and enhancing solubility in physiologically relevant solvents.



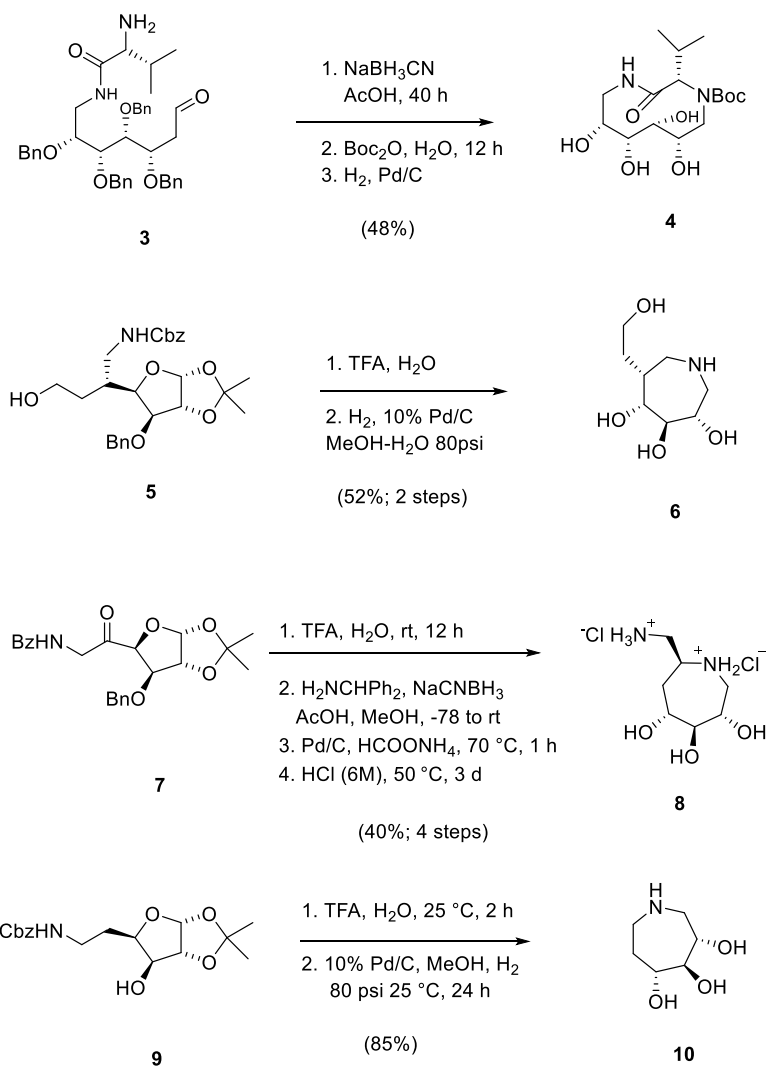
**Figure 42.** (A) Modification of lysine at the  $\epsilon$ -amine with a generic sugar molecule. (B) Glycoalkylated lysine, as described by Schlimme *et al.* (C) Glycoalkylated Ts-Lys-OMe, as described herein.

Aminocyclization is the formation of a heterocyclic ring wherein an amine serves as a nucleophile. Roy *et al.* utilized this approach to synthesize 10-membered macrocycle **4** (Scheme 1) containing a conformationally constrained hydroxyethylamine (HEA) derived from D-glucose.<sup>120</sup> Hydrolysis of the acetal in compound **5** and aminocyclization led to the seven-membered ring iminosugar **6**.<sup>121</sup> Iminosugars such as nojirimycin and 1-deoxynojirimycin are known inhibitors of glycosyl transferases and glycosidases.<sup>121</sup> 6-Amino-1-dideoxynojirimycin (**8**) was synthesized in an analogous fashion by Begoña *et al.* from D-glucose derivative **7**.<sup>122</sup> Trihydroxyazepane **10**, a glycosidase inhibitor, was synthesized from a D-glucose derivative **9** by Dhavale *et al.* Hydrolysis of the acetal in **9** and hydrogenolysis of the Cbz group led to the formation of the 7-membered trihydroxyazepane **10**.<sup>123</sup>

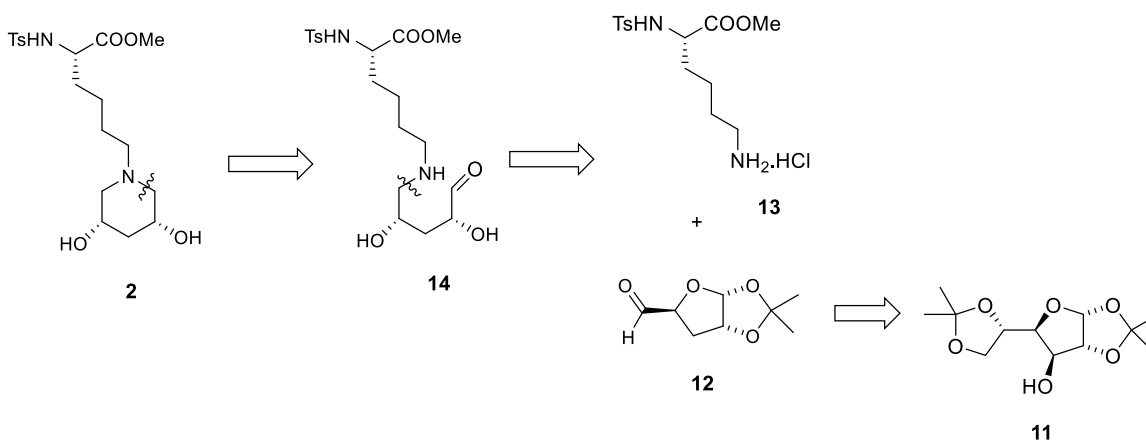
### 3.4. Retrosynthetic analysis of glycoalkylation reaction

“Glycoalkylation” of lysine can be accomplished following reductive alkylations similar to those discussed above. With the ultimate goal of modifying lysine residues in therapeutic peptides, a lysine derivative Ts-Lys-OMe (**13**, Scheme 2) was chosen as a model system for glycoalkylation. The N- and C-protecting groups were to mimic the lysine residue in a peptide chain. The retrosynthetic analysis of a model system, *viz.* single lysine residue, is shown in Scheme 2. The aldehyde **12**, derived from commercially available **11**, was envisaged to couple to the lysine derivative **13** via intermolecular reductive alkylation. Acid hydrolysis of the acetal functionality in the initially-formed adduct is expected to yield **14**, in equilibrium with the imine arising from condensation of the aldehyde and the secondary amine. Reduction of the imine by cyanoborohydride will afford compound **2**.

**Scheme 1.** Synthesis of cyclic amines derived from D-glucose



**Scheme 2.** Retrosynthetic analysis of lysine glycoalkylation with the D-glucose derived aldehyde **12**



### 3.5 Materials and Methods

The reagents were obtained from commercial sources and used without further purification except for methanol (MeOH), triethylamine (Et<sub>3</sub>N), piperidine, collidine, and toluene which were distilled after overnight stirring in CaH<sub>2</sub> with exception of MeOH which was stirred overnight with magnesium turnings. Dry tetrahydrofuran (THF) was obtained from a solvent purification system. Deuterated solvents were purchased from Cambridge Isotope Laboratories. Deuterated methanol (CD<sub>3</sub>OD), methanol (MeOH), ethyl acetate (EtOAc), dichloromethane (CH<sub>2</sub>Cl<sub>2</sub>), dimethylformamide (DMF), anhydrous magnesium sulfate (MgSO<sub>4</sub>), and hexanes were purchased from EMD Millipore. TLC silica gel 60G F<sub>254</sub> glass plates were purchased from Millipore. Tetramethylsilane (TMS), triethylamine (Et<sub>3</sub>N), diisopropylethylamine (iPr<sub>2</sub>NEt), 2,4,6-trimethylpyridine (collidine), piperidine, acetic anhydride (Ac<sub>2</sub>O), pyridine, imidazole, petroleum ether, sulfuric acid (H<sub>2</sub>SO<sub>4</sub>), 4 Å molecular sieves, acetonitrile, sodium periodate (NaIO<sub>4</sub>) were purchased from Sigma Aldrich. N,N'-Dicyclohexylcarbodiimide (DCC) was purchased from Aldrich. Silica and Nα-*p*-Tosyl-L-lysine methyl ester hydrochloride were purchased from Sigma. Sodium borohydride (NaBH<sub>4</sub>), *N*-Hydroxysuccinimide (NHS), L-proline, sodium cyanoborohydride (NaBH<sub>3</sub>CN), sodium hydride (NaH), iodomethane (MeI), and tri-*n*-butyltinhydride (Bu<sub>3</sub>SnH) were purchased from Acros Chemicals. Hydrochloric acid and sodium bicarbonate (NaHCO<sub>3</sub>) were purchased from Anachemia. Trifluoroacetic acid (TFA), Fmoc-Lys(Boc)-OH, Boc-Lys(Fmoc)-OH, and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate (HATU) were purchased from Chem-impex Int'l Inc. Toluene, tetrahydrofuran (THF), valine amide hydrochloride, 1,2:5,6-di-*O*-isopropylidene- $\alpha$ -*D*-glucofuranose, glacial acetic acid, and ethanol were purchased from VWR. Carbon disulfide (CS<sub>2</sub>) was purchased from Fisher Chemical and D-proline purchased from Alfa Aesar. <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on a Bruker AVIII-400-Nanobay spectrometer, AV500-Prodigy or Bruker AVIII-400-3. Chemical shifts are expressed in ppm downfield of TMS, in deuterated solvents as specified. Optical rotations were measured on a JASCO-DIP-370 polarimeter. HRMS was carried out using an ESI TOF 6210 (Electrospray Time-of-Flight) mass spectrometer (Agilent Technologies). HPLC (Agilent technologies) purification was performed on a Sorbent Purity C18 300Å 5µm column (250 × 10.0 mm), 1.0 mL/min, 20-90% acetonitrile + 0.1% formic acid, 20 min, monitoring UV-absorbance at λ = 218 nm and λ = 254 nm. Crystal data was collected on Bruker Kappa APEX-II DUO diffractometer.



### 3.6 Synthesis of the aldehyde from D-glucose

**3.6.1 1,2:5,6-Di-O-isopropylidene-3-O-(S-methyldithiocarbonate)- $\alpha$ -D-glucofuranose (15).**<sup>124</sup> A solution of 1,2:5,6-di-O-isopropylidene- $\alpha$ -D-glucofuranose **11** (4.006 g, 15.39 mmol, 1.00 equiv) and imidazole (4 mg, 0.58 mmol, 0.03 equiv) in THF (100 mL) was stirred at 0 °C under a nitrogen atmosphere. 60% dispersion of NaH (1.26 g, 31.50 mmol, 2.05 equiv) was added in portions over 10 min. The reaction was warmed to rt and stirred for 30 min. To the reaction mixture was added CS<sub>2</sub> (3.00 mL, 3.78 g, 49.64 mmol 3.23 equiv) resulting in a bright yellow solution. After 45 min, MeI (1.92 mL, 4.39 g, 30.84 mmol, 2.00 equiv) was added and stirring continued for 50 min under nitrogen at rt. The reaction was quenched by addition of glacial acetic acid ( $\approx$  2 mL), the mixture was filtered, and the filtrate concentrated. The residue was taken up in ethyl acetate (150 mL), washed with NaHCO<sub>3</sub> (3 x 80 mL) and dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated. The residue was purified by flash column chromatography eluting with 10:1 hexanes-EtOAc to afford **15** as a yellow crystalline solid (4.13 g, 76%).  $R_f$  0.38 (8:1 hexanes-EtOAc).  $[\alpha]_D^{25} = -27.8$  ( $c$  2.3, CHCl<sub>3</sub>). Lit.<sup>125</sup>  $[\alpha]_D^{25} = -23.6$  ( $c$  8.1, CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  1.25 (s, 6H), 1.34 (s, 3H), 1.46 (s, 3H), 2.53 (s, 3H), 4.03-4.12 (m, 2H), 4.27-4.19 (m, 2H), 4.67(d,  $J = 3.8$  Hz, 1H), 5.85-5.83 (m, 2H); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  19.3, 25.3, 26.3, 26.7, 26.8, 67.0, 72.4, 79.7, 82.8, 84.2, 105.0, 109.4, 112.4, 214.8. HRMS (ESI) calcd for C<sub>14</sub>H<sub>23</sub>O<sub>6</sub>S<sub>2</sub> (M+H)<sup>+</sup> 351.0931, obsd 351.0938.

**3.6.2 3-Deoxy-1,2:5,6-di-O-isopropylidene- $\alpha$ -D-glucofuranose (16).**<sup>126</sup> A solution of 1,2:5,6-Di-O-isopropylidene-3-O-(S-methyldithiocarbonate)- $\alpha$ -D-glucofuranose **15** (2.81 g, 8.02 mmol, 1.0 equiv) in toluene (60 mL) was added as a whole to a solution of Bu<sub>3</sub>SnH (5 mL, 5.4 g 18.58 mmol, 2.3 equiv) in toluene (60 mL). The mixture was stirred on reflux (128 °C) under N<sub>2</sub> for 7 h. The mixture turned from bright yellow to pale yellow as the reaction progressed. The solvent was removed under low pressure, the residue dissolve in acetonitrile (200 mL) and washed with petroleum ether (3 x 65 mL). The acetonitrile layer was concentrated, and the residue purified by flash column chromatography, eluting with 3:1 hexanes-EtOAc to give **16**, as a colorless viscous liquid (1.14 g, 58%).  $R_f$  0.28 (3:1 hexanes-EtOAc).  $[\alpha]_D^{25} = -7.93$  ( $c$  5.19, CHCl<sub>3</sub>). Lit.<sup>125</sup>  $[\alpha]_D^{22} = -7.60$  ( $c$  1.8, CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  1.28 (s, 3H), 1.32 (s, 3H), 1.39 (s, 3H), 1.47 (s, 3H), 1.7 (ddd,  $J = 13.5, 10.0, 4.8$  Hz, 1H), 2.15 (dd,  $J = 13.6, 4.0$  Hz, 1H), 3.7 (ddd,  $J = 8.8, 6.7, 4.2$  Hz, 1H), 4.04-4.15 (m, 3H), 4.72 (t,  $J = 4.4$  Hz, 1H), 5.70 (d,  $J = 3.6$  Hz, 1H); <sup>13</sup>C NMR (100 MHz,

CDCl<sub>3</sub>) δ 25.1, 25.6, 26.4, 26.7, 35.2, 67.1, 76.7, 78.5, 80.3, 105.5, 109.5, 111.2. HRMS (ESI) calcd for C<sub>12</sub>H<sub>21</sub>O<sub>5</sub> (M+H)<sup>+</sup> 245.1384, obsd 245.1387.

**3.6.3 3-Deoxy-1,2-O-isopropylidene- $\alpha$ -D-glucofuranose (17).**<sup>127</sup> A solution of 3-deoxy-1,2:5,6-di-O-isopropylidene- $\alpha$ -D-glucofuranose (1.025 g, 4.19 mmol) in methanol (20 mL) was stirred at rt, followed by dropwise addition of 0.9 % H<sub>2</sub>SO<sub>4</sub> (5 mL) over 5 min. The mixture was stirred for 5 h, neutralized with saturated NaHCO<sub>3</sub> (3 mL) and extracted with EtOAc (3 x 30 mL). The combined organic layers were dried over anhydrous MgSO<sub>4</sub>, filtered, concentrated and purified by flash column chromatography eluting with 9:1 EtOAc-Hexanes, to afford compound **17** as a colorless crystalline solid (450.1 mg, 53 %) *R*<sub>f</sub> 0.24 (9:1 EtOAc-Hexanes).  $[\alpha]_D^{25} = -17.2$  (*c* 2.80, CHCl<sub>3</sub>). Lit.<sup>128</sup>  $[\alpha]_D = -14$  (*c* 1.2, EtOH) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 1.33 (s, 3H), 1.51 (s, 3H), 1.85 (ddd, *J* = 13.5, 10.7, 4.8 Hz, 1H), 2.07 (dd, *J* = 13.5, 4.5 Hz, 1H), 2.59 (t, *J* = 5.4 Hz, 1H), 2.87 (d, *J* = 3.9 Hz, 1H), 3.56-3.62 (m, 1H), 3.65-3.74 (m, 1H), 3.91 (dt, *J* = 3.7, 7.5 Hz, 1H), 4.22 (td, *J* = 10.7, 4.4 Hz, 1H) 4.76 (t, *J* = 4.2 Hz, 1H), 5.81 (d, *J* = 3.7 Hz, 1H); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 26.0, 26.7, 33.6, 63.5, 72.1, 78.5, 80.5, 105.1, 111.3; HRMS (ESI) calcd for C<sub>9</sub>H<sub>17</sub>O<sub>5</sub> (M+H)<sup>+</sup> 205.1071, obsd 205.1073.

**3.6.4 1,2-O-Isopropylidene- $\alpha$ -D-glucofuranose-5-carbaldehyde (12).**<sup>129</sup> A solution of 3-Deoxy-1,2-O-isopropylidene- $\alpha$ -D-glucofuranose (405 mg, 1.98 mol, 1 equiv) in ethanol (2 mL) was added to a stirred solution of NaIO<sub>4</sub> (1.048 g, 4.90 mol, ~ 2.5 equiv) in water (10 mL). The reaction mixture was stirred for 30 min at rt, diluted with ethanol (60 mL), and filtered. The filtrate was evaporated, taken up in the EtOAc and filtered through anhydrous MgSO<sub>4</sub>. The filtrate was evaporated, and the residue purified by flash column chromatography, using a short bed (3 in) of silica gel, eluting with 9:1 EtOAc-hexanes to afford compound **12** as a colorless syrup (314 mg, 92%). *R*<sub>f</sub> 0.40 (9:1 EtOAc-hexanes).  $[\alpha]_D^{25} = -33.6$  (*c* 1.0, CHCl<sub>3</sub>). Lit.<sup>129</sup>  $[\alpha]_D^{22} = +32$  (*c* 1.4, CHCl<sub>3</sub>)\*. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 1.33 (s, 3H), 1.51 (s, 3H), 1.82 (ddd, *J* = 13.5, 11.1, 4.6 Hz, 1H), 2.33 (dd, *J* = 13.5, 5.2 Hz, 1H), 4.56 (ddd, *J* = 7.1, 5.2, 1.8 Hz, 1H), 4.78 (t, *J* = 4.1 Hz, 1H), 5.94 (d, *J* = 3.5 Hz, 1H), 9.68 (d, *J* = 1.9 Hz, 1H); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 26.2, 26.8, 34.5, 79.9, 81.5, 106.3, 112.2, 199.9; HRMS (ESI) calcd for C<sub>8</sub>H<sub>13</sub>O<sub>4</sub> (M+H)<sup>+</sup> 173.0808, obsd 173.0807.

### 3.7 Glycoalkylation of Ts-Lys-OMe

**3.7.1 *N*α-Tosyl-*N*ε-1,2-*O*-isopropylidene-α-*D*-glucofuranose-*L*-lysine methyl ester (**18**).** Flame-dried, 4 Å molecular sieves (143.0 mg) were added to a solution of 1,2-*O*-isopropylidene-α-*D*-glucofuranose-5-carbaldehyde **12**<sup>129, 130</sup> (132.9 mg, 0.77 mmol, 1.0 equiv) and triethylamine (300 μL, 216.6 mg, 2.15 mmol, 2.8 equiv) in dry methanol (20 mL). *N*α-*p*-Tosyl-*L*-lysine methyl ester hydrochloride **13** (270.9 mg, 0.77 mmol, 1.0 equiv) was added as a solid in a single portion. The mixture was stirred at rt under N<sub>2</sub> for 18 h. The molecular sieves were removed by filtration, washing well with methanol. The filtrate was concentrated to give imine. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.59 -N=CH-, d, *J* = 4.5 Hz.

Sodium borohydride (40.0 mg, 1.06 mmol, 1.0 equiv) was added to a stirred solution of imine in dry methanol (15 mL) at 0 °C and stirred under N<sub>2</sub> for 4 h. The reaction was quenched by dropwise addition of 2M HCl (600 μL), the mixture concentrated, and the residue partitioned between EtOAc (40 mL) and water (10 mL). The aqueous layer was further extracted with EtOAc (2 x 20 mL) and the combined organic extracts were concentrated. The residue was purified by flash chromatography on silica gel, eluting with 95:5 CH<sub>2</sub>Cl<sub>2</sub>-MeOH to afford **18** as a brownish solid (240.4 mg, 66 %). *R*<sub>f</sub> 0.56 (9:1 CH<sub>2</sub>Cl<sub>2</sub>-MeOH). [α]<sub>D</sub><sup>26</sup>+9.53 (*c* 2.5, CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 1.31 (s, 3H), 1.47-1.32 (m, 4H), 1.49 (s, 3H), 1.76-1.57 (m, 3H), 2.07 (dd, *J* = 13.2, 4.3 Hz, 1H), 2.40 (s, 3H), 2.62 (t, *J* = 6.6 Hz, 2H), 2.71 (dd, *J* = 12.4, 7.0 Hz, 1H), 2.87 (dd, *J* = 12.4, 3.3 Hz, 1H), 3.47 (s, 3H), 3.89 (dd, *J* = 7.3, 5.2 Hz, 1H), 4.37 (ddd, *J* = 14.2, 3.8, 3.5 Hz, 1H), 4.72 (t, *J* = 4.2 Hz, 1H), 5.80 (d, *J* = 3.7 Hz, 1H), 7.28 (d, *J* = 8.0 Hz, 2H), 7.70 (d, *J* = 8.3 Hz, 2H); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 21.5, 22.5, 26.1, 26.7, 28.3, 32.8, 36.5, 49.2, 52.2, 52.4, 55.6, 76.5, 80.4, 105.5, 111.0, 127.2(2C), 129.5(2C), 136.7, 143.6, 172.1. HRMS (ESI) calcd for C<sub>22</sub>H<sub>35</sub>N<sub>2</sub>O<sub>7</sub>S (M+H)<sup>+</sup> 471.2159, obsd 471.2149.

**3.7.2 *N*α-Tosyl-*N*ε-(2*S*,4*R*)-dihydropiperidine-*L*-lysine methyl ester (**2**).** A solution of *N*α-tosyl-*N*ε-1,2-*O*-isopropylidene-α-*D*-glucofuranose-*L*-lysine methyl ester **18** (105.0 mg, 0.24 mmol, 1.0 equiv) in TFA-water<sup>120, 122</sup> (2:1 v/v) solution was stirred for 3 h at rt. The TFA was co-evaporated with toluene, and the residue was diluted with water and lyophilized. The dried sample was dissolved in dry MeOH (3 mL) and cooled to 0 °C. Sodium borohydride (30.6 mg, 0.49 mmol, 2.0 equiv) was added and stirring continued for 4 h under N<sub>2</sub>. The reaction was quenched by the dropwise addition of 2M HCl (0.5 mL). The mixture was

concentrated, and the residue purified by flash column chromatography on silica gel eluting with 9:1 CH<sub>2</sub>Cl<sub>2</sub>-MeOH to afford compound **2** (42.0 mg, 42%). *R<sub>f</sub>* 0.37 (9:1 CH<sub>2</sub>Cl<sub>2</sub>-MeOH).  $[\alpha]_D^{25} +7.3$  (c 1.1, CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 1.21 (dt, *J* = 11.0 Hz, 1H), 1.28-1.41 (m, 2H), 1.41-1.55 (m, 2H), 1.55-1.75 (m, 2H), 1.81 (t, *J* = 9.7 Hz, 2H), 2.19-2.30 (m, 1H), 2.38 (t, *J* = 7.3 Hz, 2H) 2.45 (s, 3H), 2.94 (dd, *J* = 10.5, 3.4 Hz, 2H), 3.44 (s, 3H), 3.66-3.71 (m, 2H), 3.86 (dd, *J* = 8.6, 5.5 Hz, 1H), 7.38 (d, *J* = 8.1 Hz, 2H), 7.72 (d, *J* = 8.2 Hz, 2H); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ 20.1, 22.9, 25.0, 32.1, 41.3, 51.1, 55.6, 57.4, 59.4, 59.5, 64.9(2C), 126.8 (2C), 129.2(2C), 137.8, 143.4, 172.2. HRMS (ESI) calcd for C<sub>19</sub>H<sub>31</sub>N<sub>2</sub>O<sub>6</sub>S (M+H)<sup>+</sup> 415.1903, obsd 415.1904.

### 3.8 Synthesis and glycoalkylation of KPV-NH<sub>2</sub>

**3.8.1 Fmoc-K(Boc)-PV-NH<sub>2</sub> (22).** *N*-Hydroxysuccinimide (143.3 mg, 1.28 mmol, 1.0 equiv) and DCC (264.1 mg, 1.28 mmol, 1.0 equiv) were added to a solution of Fmoc-Lys(Boc)-OH **21** (600.0 mg, 1.28 mmol, 1.0 equiv) in CH<sub>2</sub>Cl<sub>2</sub> (20 mL) at 0 °C. The mixture was stirred for 20 min, warmed to rt, stirred for 4 h and filtered through a plug of cotton in a Pasteur pipette. The filtrate was concentrated, placed in the freezer for 2 h, filtered a second time and the filtrate concentrated. The residue was dissolved in DMF (6 mL) and cooled in an ice bath. To the stirred mixture was added L-proline (147.4 mg, 1.28 mmol, 1.0 equiv) and diisopropylethylamine (268 μL, 199.0 mg, 1.54 mmol, 1.2 equiv). The mixture was stirred at 0 °C for 10 min, warmed to rt and stirred for 14 h. Dimethylformamide was removed by a stream of air. The residue was taken up in EtOAc (100 mL) and washed with 2M HCl (80 mL). The layers were separated, and the aqueous layer was further extracted with EtOAc (3 x 20 mL). The organic fractions were combined, filtered through anhydrous MgSO<sub>4</sub> and concentrated to afford the dipeptide acid that was used directly without purification *R<sub>f</sub>* 0.32 (9:1 CH<sub>2</sub>Cl<sub>2</sub>-MeOH).

Valine amide hydrochloride (195.4 mg, 1.28 mmol, 1.0 equiv), HATU (535.5 mg, 1.41 mmol, 1.1 equiv), and 2,4,6-collidine (340 μL, 312.8 mg, 2.58 mmol, 2.0 equiv) were added to a stirred solution of Boc-Lys(Fmoc)-Pro-OH in CH<sub>2</sub>Cl<sub>2</sub> (6 mL) at 0 °C. After 10 min, the reaction was warmed to rt and stirred for 18 h under N<sub>2</sub>. The mixture was concentrated and the tripeptide **22** was isolated by flash column chromatography, eluting with 20:1 CH<sub>2</sub>Cl<sub>2</sub>-MeOH, as a colorless solid (134 mg, 44%) *R<sub>f</sub>* 0.55 (9:1 CH<sub>2</sub>Cl<sub>2</sub>-MeOH).  $[\alpha]_D^{25} +56.7$  (c 1.4, DMSO). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 0.94 (d, *J* = 6.1 Hz, 3H), 0.96 (d, *J* = 6.6

Hz, 3H), 1.42 (s, 9H), 1.32-1.58 (m, 4H), 1.58-1.70 (m, 1H), 1.71-1.76 (m, 1H), 2.01-2.11 (m, 5H), 2.93-3.18 (m, 2H), 3.57-3.76 (m, 1H), 3.76-3.82 (m, 1H), 4.17-4.22 (m, 2H), 4.29-4.45 (m, 3H), 4.50 (dd,  $J = 7.6, 3.9$  Hz, 1H), 6.70 (d,  $J = 7.5$  Hz, NH\*), 7.07 (t,  $J = 5.5$  Hz, NH\*) 7.30 (t,  $J = 7.4$  Hz, 2H), 7.39 (t,  $J = 7.4$  Hz, 2H), 7.61 (d,  $J = 7.2$  Hz, 2H), 7.76 (d,  $J = 7.5$  Hz, 2H);  $^{13}\text{C}$  NMR (100 MHz, DMSO- $d_6$ )  $\delta$  18.2, 19.8, 23.0, 25.1, 28.7, 29.0, 29.7, 31.0 (2C), 39.9, 47.1 (2C), 52.9, 57.7, 59.8, 66.1, 77.8, 120.5, 125.8, 127.5, 128.0, 141.2, 144.3, 156.0, 156.6, 171.4, 171.5, 173.3; HRMS (ESI) calcd for  $\text{C}_{36}\text{H}_{50}\text{N}_5\text{O}_7$  (M+H) $^+$  664.3705, found 664.3688.

\*Does not integrate to a full proton due to proton exchange with  $\text{CD}_3\text{OD}$

**3.8.2  $\alpha\text{G}'\text{-K(Boc)PV-NH}_2$  (**25**).** Piperidine (552  $\mu\text{L}$ , 475.8 mg, 5.59 mmol, 5.6 equiv) was added to a solution of tripeptide **22** (797.8 mg, 1.20 mmol, 1.0 equiv) in dry DMF (20 mL). The reaction was stirred at rt for 30 min. The solvent was evaporated by a stream of air, the residue partitioned between  $\text{CH}_2\text{Cl}_2$  (20 mL) and  $\text{H}_2\text{O}$  (10 mL), and the layers separated. The aqueous layer was further washed with  $\text{CH}_2\text{Cl}_2$  (3 x 10 mL) and lyophilized to afford the free amine that was used in the next reaction without further purification ( $R_f$  0.59, 6:4:1  $\text{CHCl}_3\text{-MeOH-H}_2\text{O}$ ).

Triethylamine (250  $\mu\text{L}$ , 181.4 mg, 1.79 mmol, 3.0 equiv) and flame dried 4 $\text{\AA}$  powdered molecular sieves (75.0 mg) were added to a solution of tripeptide amine (263.4 mg, 0.60 mmol, 1.0 equiv) in dry MeOH (3 mL). The mixture was stirred at rt and a solution of the aldehyde (328.3 mg, 1.91 mmol, 3.2 equiv) in dry MeOH (3 mL) was added. The mixture was left to stir at rt for 24 h. The reaction was filtered through a pad of Celite<sup>®</sup> that was washed well with MeOH. The filtrate was cooled to 0  $^\circ\text{C}$ ,  $\text{NaBH}_4$  (73.2 mg, 1.93 mmol, 3.2 equiv) was added, and the mixture was stirred for 4 h under  $\text{N}_2$ . The reaction was quenched by dropwise addition of 2M HCl (250  $\mu\text{L}$ ). The mixture was concentrated, and the residue purified by flash column chromatography, eluting with 9:1  $\text{CH}_2\text{Cl}_2\text{-MeOH}$  to afford the tripeptide **25** (194 mg, 27%)  $R_f$  0.54 (9:1  $\text{CH}_2\text{Cl}_2\text{-MeOH}$ ).  $[\alpha]_{\text{D}}^{25}$  -68.7 (c 1.9,  $\text{CHCl}_3$ ).  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  1.00 (d,  $J = 6.6$  Hz, 3H), 1.01 (d,  $J = 6.8$  Hz, 3H), 1.31 (s, 3H), 1.45 (s, 9H), 1.48 (s, 3H), 1.40-1.53 (m, 3H), 1.54-1.62 (m, 2H), 1.63-1.73 (m, 2H), 1.98-2.07 (m, 3H), 2.08-2.14 (m, 2H), 2.14-2.20 (m, 1H), 2.58 (dd,  $J = 12.7, 6.6$  Hz, 1H), 2.79 (dd,  $J = 12.7, 3.5$  Hz, 1H) 2.89-3.12 (app. t,  $J = 5.1$  Hz, 3H), 3.62-3.74 (m, 2H), 3.77- 3.82 (m, 1H), 4.21 (d,  $J = 6.5$  Hz, 1H), 4.25-4.30 (m, 1H), 4.59 (dd,  $J = 8.2, 4.3$  Hz, 1H), 4.76 (app. t,  $J = 4.2$  Hz, 1H), 5.78 (d,  $J = 3.7$

Hz, 1H); <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD)\* δ 17.1, 18.5, 22.4, 24.6, 25.0, 25.6, 27.4, 28.7, 29.4, 30.7, 32.4, 36.1, 39.7, 47.1, 50.0, 58.3, 58.8, 59.9, 77.9, 80.3, 105.7, 110.8, 157.1, 172.8 (2C), 174.2, 174.7. HRMS (ESI) calcd for C<sub>29</sub>H<sub>52</sub>N<sub>5</sub>O<sub>8</sub> (M+H)<sup>+</sup> 598.3810, obsd 598.3809.

\* Reported for the major conformation only two species were observed that were presumed to be rotamers about the prolyl amide bond.

**3.8.3 αG\*-KPV-NH<sub>2</sub> (26).** A solution of compound **25** (78.0 mg, 0.13 mmol, 1.0 equiv) in TFA-H<sub>2</sub>O (2:1 v/v, 4.5 mL) was stirred for 3.5 h. The mixture was diluted with toluene (20 mL) and concentrated. The residue was dissolved in MeOH and stirred at 0 °C. Solid NaHCO<sub>3</sub> (35.3 mg) was added to neutralize the solution. NaBH<sub>3</sub>CN (16.3 mg, 0.26 mmol, 2.0 equiv) was added and the mixture stirred for 15 h. The reaction was quenched by the dropwise addition of 2M HCl (~600 μL), concentrated, and the residue subjected to HPLC to afford compound **26** (17.8 mg, 31 %). *t<sub>R</sub>* 15.5 min. *R<sub>f</sub>* 0.13 (6:4:1 CHCl<sub>3</sub>-MeOH-H<sub>2</sub>O). [α]<sub>D</sub><sup>25</sup> -28.9 (c 0.1, MeOH); <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 1.01 (d, *J* = 6.8 Hz, 3H), 1.02 (d, *J* = 6.7 Hz, 3H), 1.24 (app. q, *J* = 10.7 Hz, 1H), 1.29-1.52 (m, 2H), 1.62-1.78 (m, 3H), 1.79-1.92 (m, 1H), 1.94-2.13 (m, 5H), 2.15-2.28 (m, 2H), 2.31 (t, *J* = 10.0 Hz, 1H), 2.87-3.02 (m, 4H), 3.56 (dd, *J* = 10.1, 3.9 Hz, 1H), 3.60-3.78 (m, 3H), 3.88-3.92 (m, 1H), 4.18 (d, *J* = 6.8 Hz, 1H), 4.53 (dd, *J* = 8.4, 4.0 Hz, 1H); <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD) δ 17.2, 18.4, 22.7, 24.4, 24.5, 27.0, 29.4, 30.7, 39.2, 41.6, 47.4, 55.3, 56.9, 58.5, 60.3, 65.2, 65.7 (2C), 171.0, 173.2, 174.7; HRMS (ESI) calcd for C<sub>21</sub>H<sub>40</sub>N<sub>5</sub>O<sub>5</sub> (M+H)<sup>+</sup> 442.3024, obsd 442.3029.

**3.8.4 Ac-KPV-NH<sub>2</sub> (27).** A solution of tripeptide **22** (362.2 mg, 0.55 mmol, 1.0 equiv) in piperidine (544 μL, 470.0 mg, 5.50 mmol, 10.0 equiv) and DMF (5 mL) was stirred for 30 min. The solvent was evaporated by a stream of air, and the residue partitioned between CH<sub>2</sub>Cl<sub>2</sub> (50 mL) and H<sub>2</sub>O (30 mL). The aqueous layer was further washed with CH<sub>2</sub>Cl<sub>2</sub> (2 x 20 mL) and lyophilized to afford the free amine (232 mg, 96%). *R<sub>f</sub>* 0.28 (9:1 CH<sub>2</sub>Cl<sub>2</sub>-MeOH).

A portion of the free amine (94 mg, 0.213 mmol) was dissolved in a mixture of Ac<sub>2</sub>O-pyridine (1:1 v/v, 6 mL) and stirred for 15 h, concentrated and purified by flash column chromatography, eluting with 100:7 CH<sub>2</sub>Cl<sub>2</sub>-MeOH to give the acetylated tripeptide, Ac-K(Boc)-PV-NH<sub>2</sub> (72 mg, 70 %). *R<sub>f</sub>* 0.50 (20:3 CH<sub>2</sub>Cl<sub>2</sub>-MeOH). [α]<sub>D</sub><sup>25</sup> -65.4 (c 1.5, CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 0.99 (d, *J* = 3.4 Hz, 3H), 1.00 (d, *J* = 3.4 Hz, 3H), 1.43-1.56 (m, 4H), 1.45 (s, 9H), 1.59-1.72 (m, 1H), 1.77-1.87 (m, 1H), 1.96-2.19 (m, 5H),

1.98 (s, 3H), 3.06 (app. t,  $J = 6.0$  Hz, 2H), 3.66-3.73 (m, 1H), 3.73-3.92 (m, 1H), 4.22 (app. t,  $J = 6.3$  Hz, 1H), 4.54 (dd,  $J = 8.0, 3.9$  Hz, 1H), 4.54-4.59 (m, 1H), 6.59 (br s, NH\*), 7.91 (d,  $J = 8.2$  Hz, NH\*), 8.18 (d,  $J = 7.0$  Hz, NH\*);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  17.1, 18.5, 20.9, 22.6, 24.7, 27.4(3C), 28.8, 29.3, 30.7 (2C), 39.6, 47.3, 51.3, 58.3, 60.1, 78.4, 157.1, 171.8, 171.9, 172.8, 174.7. HRMS (ESI) calcd for  $\text{C}_{23}\text{H}_{42}\text{N}_5\text{O}_6$  (M+H)<sup>+</sup> 484.3135, obsd 484.3130.

\*Does not integrate for a full proton due to deuterium exchange

The acetylated tripeptide, Ac-K(Boc)PV-NH<sub>2</sub> (72.0 mg, 0.19 mmol) was dissolved in a mixture  $\text{CH}_2\text{Cl}_2$ -TFA (1:1 v/v, 4 mL) and stirred at rt for 30 min. The mixture was concentrated, and the residue dissolved in toluene and concentrated again. The residue was purified by HPLC to afford the free amine **27** (40.8 mg, 71 %).  $t_R$  16.2 min.  $R_f$  0.36 (20:13:3:1  $\text{CHCl}_3$ -MeOH- $\text{H}_2\text{O}$ - $\text{NH}_3$ ).  $[\alpha]_D^{25}$  -85.5 ( $c$  0.6, MeOH).  $^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  1.00 (d,  $J = 2.5$  Hz, 3H), 1.01 (d,  $J = 2.5$  Hz, 3H), 1.42-1.55 (m, 2 H), 1.62-1.73 (m, 3 H), 1.75-1.84 (m, 1 H), 1.96-2.15 (m, 4H), 1.99 (s, 3H), 2.18-2.24 (m, 1H), 2.78 (t,  $J = 7.0$  Hz, 2H), 3.66-3.71 (m, 1H), 3.86-3.92 (m, 1H), 4.18 (d,  $J = 6.9$  Hz, 1H), 4.53 (dd,  $J = 8.4, 4.6$  Hz, 1H), 4.60 (dd,  $J = 8.1, 5.9$  Hz, 1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  17.1, 18.5, 20.9, 22.2, 24.7, 29.0 (2C), 30.6, 30.7, 39.7, 47.4, 51.1, 58.5, 60.1, 171.6, 171.8, 172.9, 174.7; HRMS (ESI) calcd for  $\text{C}_{18}\text{H}_{34}\text{N}_5\text{O}_4$  (M+H)<sup>+</sup> 384.2611, obsd 384.2606.

**3.8.5 Boc-K(Fmoc)-PV-NH<sub>2</sub> (24).** Boc-Lys(Fmoc)-OH **23** (600.0 mg, 1.28 mmol) was treated, by analogy to the procedure described for the conversion of **21** to **22**, to afford **24** (360.0 mg, 42%)  $R_f$  0.43 (9:1  $\text{CH}_2\text{Cl}_2$ -MeOH).  $[\alpha]_D^{25}$  -55.7 ( $c$  0.8, MeOH).  $^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  0.98 (d,  $J = 2.1$  Hz, 3H), 0.99 (d,  $J = 2.1$  Hz, 3H), 1.43 (s, 9H), 1.43-1.64 (m, 5H), 1.72-1.81 (m, 1H), 1.95-2.17 (m, 5H), 3.13 (app. t,  $J = 6.2$  Hz, 2H), 3.63 (dd,  $J = 16.0$  Hz, 9.6 Hz, 1H), 3.79 (dd,  $J = 16.0$  Hz, 6.8 Hz, 1H), 4.17-4.22 (m 1H), 4.21 (d,  $J = 6.6$  Hz, 1H), 4.27-4.34 (m, 1H), 4.35 (d,  $J = 6.9$  Hz, 2H), 4.55 (dd,  $J = 8.0$  Hz, 3.8 Hz, 1H), 7.31 (t,  $J = 7.4$  Hz, 2H), 7.39 (t,  $J = 7.4$  Hz, 2H), 7.65 (d,  $J = 7.4$  Hz, 2H), 7.79 (d,  $J = 7.4$  Hz, 2H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  17.2, 18.5, 22.5, 24.7, 27.4, 28.7, 29.1, 30.6, 30.9, 39.9, 47.1, 52.4, 58.3, 60.1, 66.2, 79.2, 119.6, 124.8, 126.8, 127.4, 141.2, 144.0, 157.4, 157.5, 172.6, 172.8, 174.7; HRMS (ESI) calcd for  $\text{C}_{36}\text{H}_{50}\text{N}_5\text{O}_7$  (M+H)<sup>+</sup> 664.3705, obsd 664.3710.

**3.8.6 Boc-K( $\epsilon$ G')PV-NH<sub>2</sub> (28).** Following the same series of reactions in the conversion of **22** to **25** above, compound **24** (125.0 mg, 0.19 mmol) was converted to **28** (41 mg, 60%). *R<sub>f</sub>* 0.78 (6:4:1 CHCl<sub>3</sub>-MeOH-H<sub>2</sub>O).  $[\alpha]_D^{25}$  -59.3 (*c* 1.1, MeOH). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$  1.00 (d, *J* = 2.1 Hz, 3H), 1.01 (d, *J* = 2.1 Hz, 3H), 1.33 (s, 3H), 1.45 (s, 9H), 1.49 (s, 3H), 1.40-1.55 (m, 1H), 1.61-1.84 (m, 5H), 1.99-2.27 (m, 6H), 2.95-3.03 (m, 3H), 3.25 (dd, *J* = 12.8, 2.4 Hz, 1H), 3.66-3.27 (m, 1H), 3.83-3.88 (m, 1H), 4.18 (d, *J* = 6.6 Hz, 1H), 4.37 (t, *J* = 6.7 Hz, 1H), 4.42-4.48 (m, 1H), 4.55 (dd, *J* = 8.2, 4.2 Hz, 1H), 4.84 (t, *J* = 4.1 Hz, 1H), 5.89 (d, *J* = 3.5 Hz, 1H); <sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD,)  $\delta$  17.2, 18.5, 22.3, 24.7, 25.0, 25.7, 25.9, 27.3, 29.0, 30.7, 30.8, 36.2, 47.4, 48.0, 50.7, 52.0, 58.5, 60.3, 74.0, 79.2, 80.4, 105.9, 111.3, 156.5, 172.2, 172.9, 174.7. HRMS (ESI) calcd for C<sub>29</sub>H<sub>51</sub>N<sub>5</sub>O<sub>8</sub> (M+H)<sup>+</sup> 598.3810, obsd 598.3817.

**3.8.7 H-K( $\epsilon$ G\*)PV-NH<sub>2</sub> (29).** By analogy to the procedure described for conversion of **25** to **26**, compound **28** (139.0 mg, 0.23 mmol) was converted to **29**. The crude product was purified by HPLC to afford  $\epsilon$ -glycoalkylated tripeptide **x** (24 mg, 23 %). *t<sub>R</sub>* 15.6 min). *R<sub>f</sub>* 0.20 (6:4:1 CHCl<sub>3</sub>-MeOH-H<sub>2</sub>O).  $[\alpha]_D^{25}$  -35.5 (*c* 0.4, MeOH). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$  1.00 (d, *J* = 6.8 Hz, 3H), 1.01 (d, *J* = 6.7 Hz, 3H), 1.19-1.35 (m, 1H), 1.41-1.54 (m, 2H), 1.42-1.64 (m, 2H), 1.66-1.76 (m, 1H), 1.78-1.92 (m, 3H), 1.96-2.16 (m, 4H), 2.20-2.29 (m, 2H), 2.50 (app. t, *J* = 7.7 Hz, 2H), 2.98 (dd, *J* = 10.6, 3.7 Hz, 2H), 3.63-3.66 (m, 1H), 3.68-3.75 (m, 3H), 3.91 (t, *J* = 6.2 Hz, 1H), 4.20 (d, *J* = 6.7 Hz, 1H), 4.58 (dd, *J* = 8.1, 4.6 Hz, 1H); <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD)  $\delta$  17.1, 18.4, 22.5, 24.6, 25.8, 28.8, 30.7, 32.4, 41.7, 47.2, 51.9, 57.5, 58.4, 59.6 (2C), 60.1, 65.0 (2C), 171.4, 172.6, 174.7. HRMS (ESI) calcd for C<sub>21</sub>H<sub>40</sub>N<sub>5</sub>O<sub>5</sub> (M+H)<sup>+</sup> 442.3024, obsd 442.3032.

## 3.9 Discussion

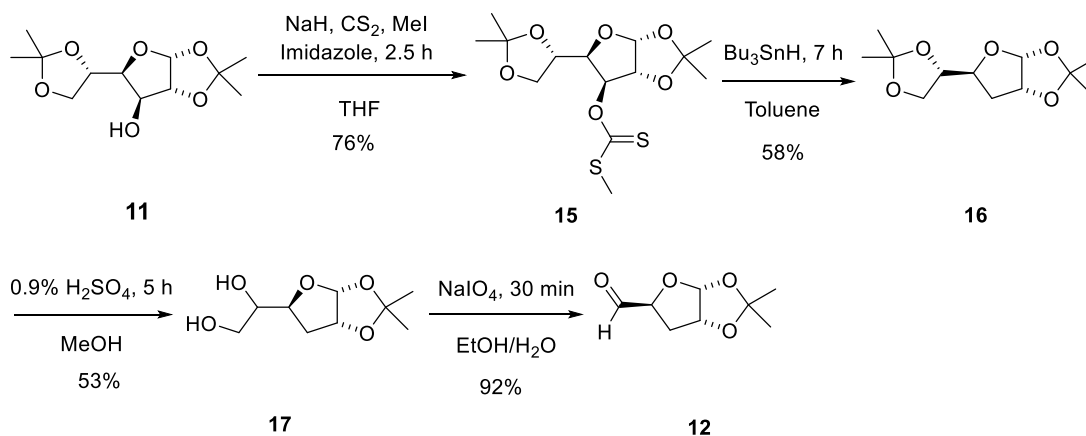
### 3.9.1 Synthesis of the aldehyde **12**

The aldehyde **12** (Scheme 3) was synthesized from commercially available diacetone-D-glucose **11** through Barton-McCombie deoxygenation at C3.<sup>126</sup> Compound **11** was reacted with excess NaH and catalytic quantities of imidazole. Addition of carbon disulfide to the reaction mixture led to the formation of dithiocarbonate salt which was alkylated with iodomethane to afford the *S*-methyl xanthate ester **15**. Reductive homolytic cleavage of C-O bond in **15** with tributyltin hydride, driven by the formation of a strong Sn-S bond, afforded **16**. Subjection of compound **16** to mild acid hydrolysis led to deprotection of C5,6-diol



and oxidative cleavage of the C5,6-diol by  $\text{NaIO}_4$ , afforded **12**. The NMR spectra are provided in Appendix B-1.

### Scheme 3. Synthesis of the aldehyde (**12**)

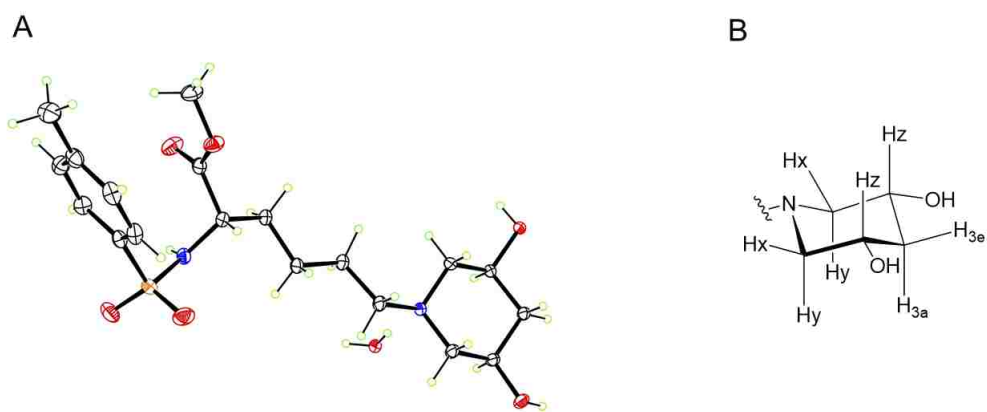
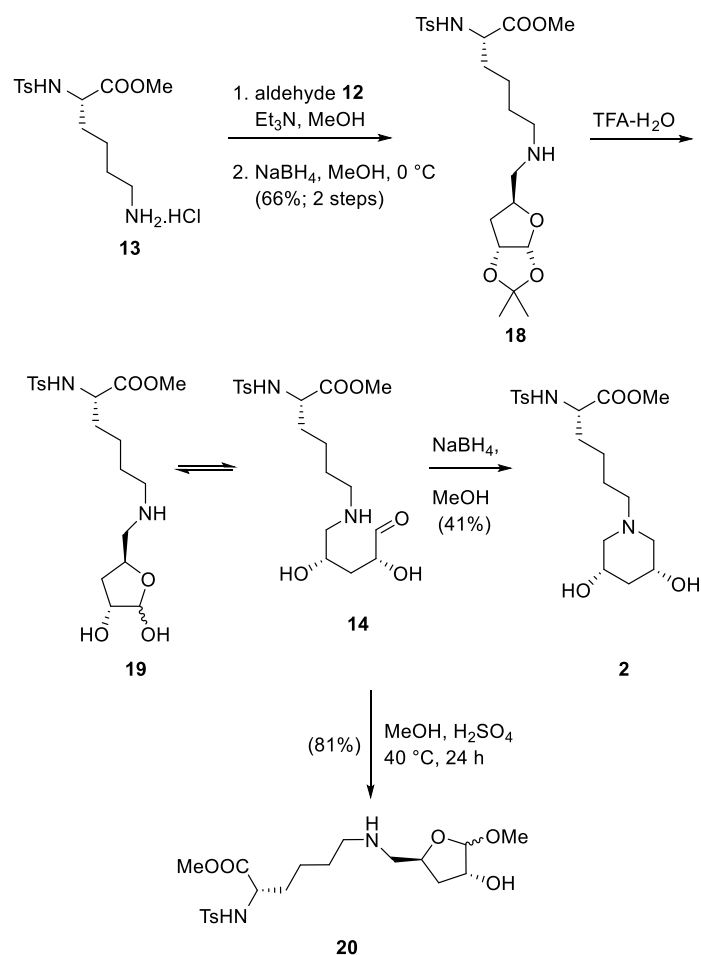


### 3.9.2 Synthesis of the glycoalkylated lysine derivative **2**

Reductive alkylation of **13** with the aldehyde **12** afforded compound **18** which was subjected to hydrolysis and intramolecular reductive alkylation to form the piperidine diol. Acid hydrolysis of **18** exposes the hemiacetal **19**, which in equilibrium with the aldehyde **14**. The aldehyde condenses with the  $\epsilon$ -NH leading to compound **2**. Multiple mineral acids were investigated for the hydrolysis of the acetal compound **18**. Treatment of compound **18** with a 1:1 10%  $\text{H}_2\text{SO}_4$ -MeOH (v/v) solution led to ring opening and trapping of the aldehyde **14** as cyclic methylated acetal **20**.

An extensive search of the literature for hydrolysis conditions that preceded successful intramolecular ring formation led to the use of a TFA- $\text{H}_2\text{O}$  acid system. Different TFA- $\text{H}_2\text{O}$  ratios: 2:1,<sup>131, 132</sup> 4:1,<sup>120</sup> and 3:2<sup>121, 123</sup> have been used by different scientists to perform the aminocyclization reactions outlined in Section 3.3. Treatment of compound **18** with 2:1 TFA/ $\text{H}_2\text{O}$ , followed by reduction with sodium borohydride, afforded compound **2**. Formation of the product was confirmed by NMR, ESI-MS, and X-ray crystallography. A solution of purified **2** in methanol formed crystals following storage at 4 °C. The crystal structure is shown in Figure 43A and the CCDC deposition number is 1825648. The NMR spectra are provided in Appendix B-2.

**Scheme 4.** Synthesis of the glycoalkylated lysine derivative (**2**)



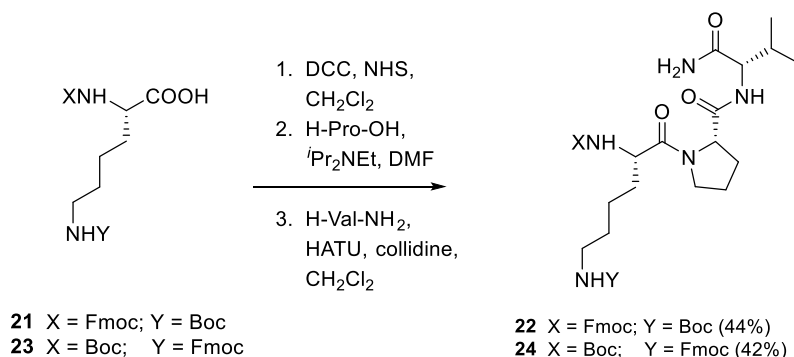
**Figure 43.** A) ORTEP drawing of compound **2**. B) The piperidine ring of compound **2** showing the three pairs of equivalent hydrogens Hx, Hy, and Hz.

From the crystal structure, it can be seen that the piperidine-2,4-diol ring is symmetric along the ring plane passing through N and C3. Each hydroxyl group of the diol adopts an equatorial orientation. <sup>1</sup>H NMR analysis of compound **2** confirms the symmetry of the piperidine, showing three pairs of equivalent protons (Figure 43A): H<sub>x</sub> (H1<sub>e</sub> and H5<sub>e</sub>); H<sub>y</sub> (H1<sub>a</sub> and H5<sub>a</sub>); and H<sub>z</sub> (H2 and H4). A doublet of doublet peak was observed at δ 2.94 corresponding to H1<sub>e</sub>, H5<sub>e</sub> with a large geminal coupling constant ( $J_{1e,1a}$  and  $J_{5e,5a} = 10.4$  Hz) and a small vicinal coupling constant ( $J_{1e,2a}$  and  $J_{5e,4a} = 3.3$  Hz). This small vicinal coupling constant places H2 and H4 in axial positions, consistent with the equatorial orientation of the hydroxyl groups in the crystal structure.

### 3.9.3 Synthesis of Ac-KPV-NH<sub>2</sub> and the glycoalkylated analogs

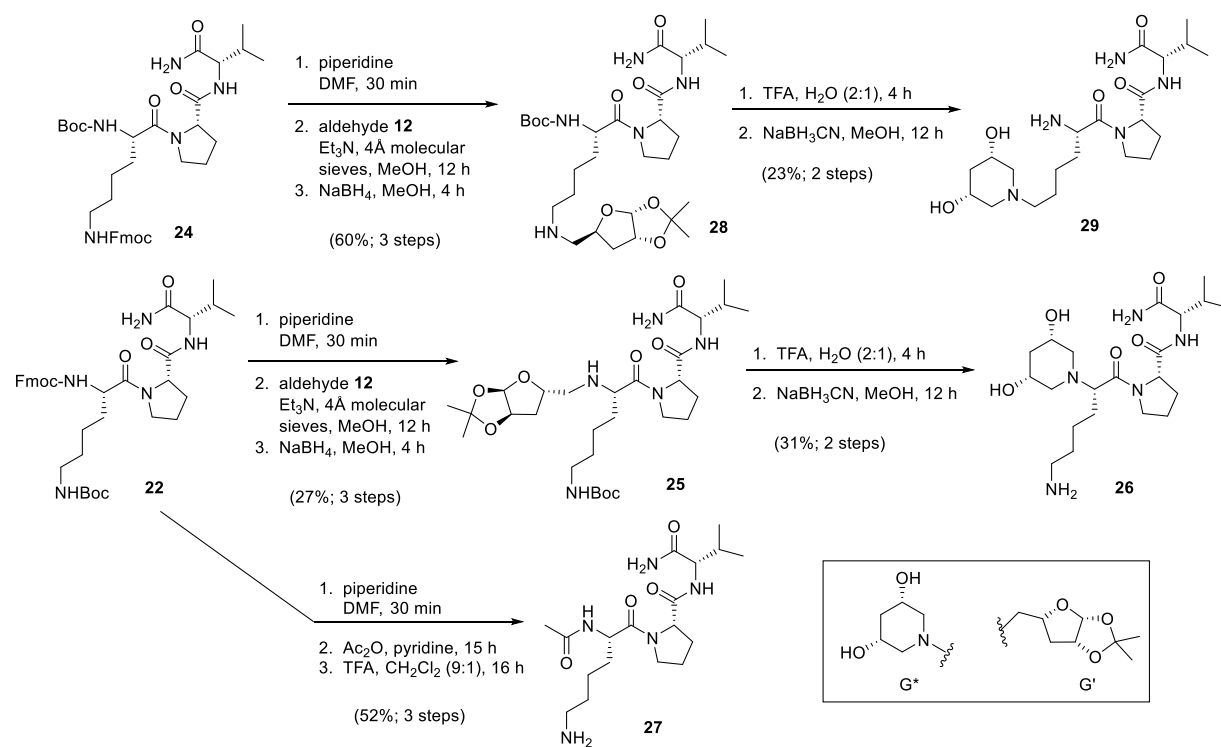
Following the success of our model system, the possibility of applying the same modification to a short peptide sequence containing a lysine residue was explored. From the list of the biologically relevant tripeptides and tripeptide motifs provided by Ung and Winkler,<sup>11</sup> the tripeptide Ac-KPV-NH<sub>2</sub> was selected. Its dual functionality as an antimicrobial and anti-inflammatory peptide would be used to assess the effect of glycoalkylation on the biological activity. Three derivatives of H-KPV-NH<sub>2</sub> were synthesized: αG\*-KPV-NH<sub>2</sub> **26**, H-K(εG\*)PV-NH<sub>2</sub> **29**, and the parent peptide Ac-KPV-NH<sub>2</sub> **27** were synthesized (with G\* defined in Scheme 6). The core tripeptides **22** and **24** were synthesized and used for α- and ε-glycoalkylation respectively (Scheme 5). The activated acid of **21** was coupled with proline to give the dipeptide acid which was coupled with valine amide to afford **22**. Compound **24** was synthesized from **23** by analogy.

**Scheme 5.** Synthesis of the α- and ε-protected tripeptides **22** and **24**



Fmoc deprotection of the tripeptide **22** followed by reductive alkylation of the resulting primary amine with compound **12**, afforded **25** (scheme 6). Acid hydrolysis of the acetal in **25** unmasked an aldehyde which condensed with the secondary  $\alpha$ -NH of lysine, through intramolecular reductive alkylation to form the piperidine diol in compound **26** ( $\alpha$ -glycoalkylation). The  $\epsilon$ -glycoalkylation was accomplished using compound **24**, by analogy, to afford **29** via the intermediate compound **28**. The control peptide Ac-KPV-NH<sub>2</sub> (**27**) was synthesized from tripeptide **22**. Fmoc deprotection of **22** gave the primary  $\alpha$ -amine which was acetylated. Cleavage of the Boc group from sidechain amine afforded **27**. The NMR spectra are provided in Appendix B-3 and B-4.

**Scheme 6.** Synthesis of compounds **26**, **27**, and **29** from compounds **22**, and **24**.



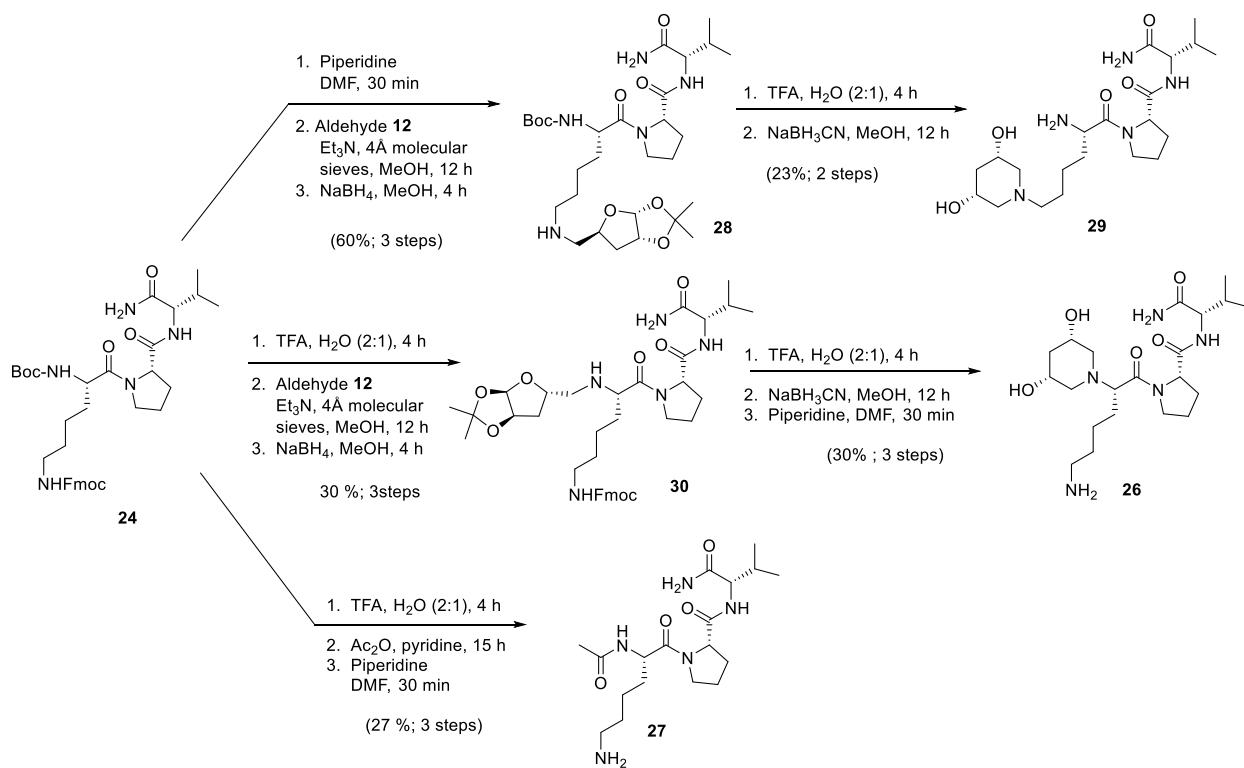
During the intramolecular cyclization to form the piperidine diol ring, reduction of the imine by NaBH<sub>3</sub>CN, under acidic conditions, led to low yields of **26** and **29**. The reduction was therefore done under basic conditions<sup>133</sup> by addition of NaHCO<sub>3</sub> to the reaction mixture. The polar diol and amine functionality in the final products made it impossible to purify the compounds by silica gel column chromatography. Reverse phase HPLC purification was used, unfortunately, some fractions of the glycopeptides eluted with

residual borohydride salts. The reported low yields are mainly attributed to losses incurred during purification.

During subsequent efforts to prepare more of **26**, **27**, and **29**, for biological studies (*vide infra*), it was found that all three tripeptides could be synthesized from Boc-K(Fmoc)PV-NH<sub>2</sub> following the reaction steps summarized in Scheme 7. While additional reaction and purification steps are incurred in the synthesis of the alpha-glycoalkylated tripeptide; this approach has the advantage of a common intermediate that could be prepared on large scale (hundreds of milligrams) and some practical advantages. Retaining the Fmoc group, and its inherent UV-activity, during the intramolecular cyclization made it easier to monitor the reaction by TLC. It was also possible to purify compound **30** by silica gel chromatography, which removed borohydride salts and thus they are not carried forward to complicate the HPLC purification of **26**. The final purification of compound **26**, after Fmoc deprotection, is simplified to a phase separation between, CH<sub>2</sub>Cl<sub>2</sub> and H<sub>2</sub>O, followed by lyophilization of the aqueous layer. Chromatographic behavior (TLC and HPLC) and <sup>1</sup>H NMR spectra of **26** and **27** were identical to those obtained for batches synthesized via Scheme 6.

Following the successful synthesis of the three compounds αG<sup>+</sup>-KPV-NH<sub>2</sub> **26**, H-K(εG<sup>+</sup>)PV-NH<sub>2</sub> **29**, and the parent peptide Ac-KPV-NH<sub>2</sub> **27**. The compounds were used for antimicrobial and anti-inflammatory tests to check the effects of glycoalkylation on the biological activity Ac-KPV-NH<sub>2</sub> of (discussed in the next chapter). Stability of the tripeptides towards proteolytic cleavage was also tested.

Scheme 7. Alternative synthetic route for compounds **26** and **27**.



## CHAPTER 4

### BIOLOGICAL ACTIVITY AND ENZYMATIC STABILITY TESTS OF KPV TRIPEPTIDE DERIVATIVES

#### 4.1 Antimicrobial activity of Ac-KPV-NH<sub>2</sub>

The tripeptide Ac-KPV-NH<sub>2</sub> has been shown to have similar antimicrobial activity to the tridecapeptide  $\alpha$ -MSH. Cutuli *et al.* demonstrated the antimicrobial effects of Ac-KPV-NH<sub>2</sub> against *Staphylococcus aureus* and *Candida albicans*, with significant activity at concentrations ranging from picomolar (quantity of  $\alpha$ -MSH found in human plasma) to micromolar.<sup>116</sup> Charnley *et al.* showed that altering the stereochemistry of Ac-KPV-NH<sub>2</sub> to Ac-KP<sub>D</sub>V-NH<sub>2</sub> (L-valine substituted with D-valine) did not increase its antimicrobial activity,<sup>134</sup> different from what was observed for its anti-inflammatory activity. The anti-inflammatory effect was shown to increase with <sub>D</sub>Val substitution.<sup>135</sup> Charnley *et al.* further reported that antimicrobial effects of Ac-KP<sub>D</sub>V-NH<sub>2</sub> were similar in both gram negative and Gram positive bacteria (*Escherichia coli* and *Staphylococcus aureus*).<sup>134</sup> They demonstrated that replacing lysine with alanine did not have any significant effect on the antimicrobial activity of Ac-KPV-NH<sub>2</sub>, implying that the cationic charge in lysine is not necessary. They therefore proposed an Ac-XP<sub>(D/L)</sub>V-NH<sub>2</sub> sequence for antimicrobial activity.<sup>134</sup>

Singh *et al.* studied the antimicrobial effect of Ac-KPV-NH<sub>2</sub> ( $\alpha$ -MSH 11-13) against MRSA and MSSA and compared to that of the full-length peptide,  $\alpha$ -MSH 1-13. Their findings showed that the C-terminal tripeptide retained similar activity as the  $\alpha$ -MSH 1-13 and the presence of physiological salt concentration did not affect the potency of the peptide.<sup>136</sup> Their study also indicated that the full-length and the C-terminal fragment of  $\alpha$ -MSH disrupted the bacterial membrane.<sup>136</sup>

In the current study, the antimicrobial activity of Ac-KPV-NH<sub>2</sub> and the glycopeptide analogs were investigated against *Staphylococcus aureus* USA300, by agar diffusion test, colony forming units and liquid culture.

---

Portions of this chapter previously appeared in PLOS ONE, Songok, *et al.* *PLoS ONE* **2018**, 13, e0199686..

#### 4.2 Anti-inflammatory activity of Ac-KPV-NH<sub>2</sub>

The tripeptide Ac-KPV-NH<sub>2</sub> has been shown to have similar anti-inflammatory properties to  $\alpha$ -MSH.<sup>113, 135</sup> However, there is no evidence that Ac-KPV-NH<sub>2</sub> modulates this process through melanocortin receptors. There was dose-dependent response on anti-inflammatory effect of Ac-KPV-NH<sub>2</sub> in activated microglia,<sup>113</sup> which is not characteristic of a receptor signaling pathway. On the other hand, KdPT (IL-1 $\beta$ <sup>193-195</sup>) an analog of Ac-KPV-NH<sub>2</sub> (containing D-proline) and an antagonist of IL-1 $\beta$  has been shown to compete with IL-1 $\beta$  for the same receptor site. The analgesic effect of IL-1 $\beta$ , given intraperitoneally, was attenuated by subcutaneous pretreatment with KdPT (85  $\mu$ g/150 g) in rat.<sup>137</sup> This inhibition was overcome by larger doses of IL-1 $\beta$  which confirmed competitive antagonism.<sup>137</sup>

Although the actual mechanism of KPV's action is not fully determined, it has been shown to have anti-inflammatory activity both *in vitro* (summarized in Table 7) and *in vivo* as observed in different cell lines.<sup>12</sup> KPV and its derivatives suppressed the production of proinflammatory cytokines (TNF- $\alpha$ , IL-6<sup>113</sup> and IL-8), inhibited the activation of NF- $\kappa$ B<sup>138</sup> for nuclear translocation, and reduced production of NO during the inflammation process.<sup>139</sup> The anti-inflammatory effects of Ac-KPV-NH<sub>2</sub> and the glycopeptides have been assessed in the current study, for inhibition of NF- $\kappa$ B nuclear translocation in 3T3-L1 adipocyte cells, with the inflammation induced by TNF $\alpha$ .

#### 4.3 Enzymatic stability of therapeutic peptides

The high target specificity and good tolerance achieved in the use of therapeutic peptides has enabled their broad range of application in treatment of various diseases, including cancer and metabolic disorders.<sup>140</sup> The use of natural peptide drugs is faced with the challenge of low oral bioavailability, due to proteolytic degradation and minimal absorption across the epithelial barrier. To minimize enzymatic degradation of the peptides, structural modifications such as C-terminal amidation, N-terminal acetylation, site specific modifications, and the use of unnatural amino have been implemented.<sup>36</sup>

Structural modification of Ac-KPV-NH<sub>2</sub> by reductive glycoalkylation is expected to confer proteolytic stability to the tripeptide. To evaluate the effect of this modification on proteolytic stability, the glycoalkylated KPV analogs were exposed to pronase and the reactivity compared with that of the parent peptide Ac-KPV-



NH<sub>2</sub>. Pronase comprises of a mixture of aminopeptidases, endopeptidases, serine-type proteases, and a carboxypeptidase.<sup>141</sup> It is extracted from *Streptomyces griseus* and often used for complete degradation of proteins to their constituent amino acids.<sup>141</sup>

**Table 7.** Anti-inflammatory effects of KPV analogs *in vitro*.<sup>12</sup>

Target molecule /process	Peptide	Effect of $\alpha$ -MSH	Cell type
<b>Proinflammatory cytokines</b>	KPV	LPS/PMA-induced TNF- $\alpha$ expression ↓	Murine connective tissue fibroblasts cell line L292
	KPV	LPS/IFN- $\gamma$ -induced TNF- $\alpha$ and IL-6 expression ↓	Murine microglial cell line (N9 clone)
	KPV	$\beta$ A/IFN- $\gamma$ -induced NO <sub>2</sub> <sup>-</sup> production ↓ iNOS expression ↓	Murine microglial cell line (N9 clone)
	KPV	IL-1 $\beta$ -induced IL-8 expression ↓	Human colon, Jurkat cells
	KdPT	IL-1 $\beta$ -induced IL-6 and IL-8 expression ↓	SZ95 human sebocytes
<b>Transcription factor NF-B</b>	KPV	TNF- $\alpha$ -induced NF- $\kappa$ B activation ↓	HIV-1-infected promonocytic U1 cell line
	KdPV	LPS-induced NF- $\kappa$ B activation ↓	Rat alveolar type II epithelial cells
	KPV	LPS/IFN- $\gamma$ -induced NF- $\kappa$ B activation ↓	RAW 264.7 murine macrophage cell line
	KPV, KPdV	TNF- $\alpha$ -induced NF- $\kappa$ B activation ↓	HaCaT keratinocytes, human keratinocytes
	KPV	IL-1 $\beta$ -induced NF- $\kappa$ B activation ↓	Human colon, Jurkat cells
	KdPT	IL-1 $\beta$ -induced NF- $\kappa$ B activation ↓	SZ95 human sebocytes
<b>Non-cytokine pro-inflammatory mediators</b>	KPV	IFN- $\gamma$ /LPS-induced NO <sub>2</sub> <sup>-</sup> production ↓	Murine microglial cell line (N9 clone)
	KPV	LPS/IFN- $\gamma$ -induced NO production ↓	RAW 264.7 murine macrophage cell line
	Cyclic KPv, KPdV Derivatives	LPS/IFN- $\gamma$ -induced NO production ↓	RAW 264.7 murine macrophage cell line
<b>Cytokine suppressors</b>	KPV	IL-10 production ↑	Human peripheral blood monocytes

$\beta$ A,  $\beta$ -Amyloid protein; PMA, phorbol-12-myristate-13-acetate; ↓, decreased; ↑, increased.

#### 4.4 Antimicrobial tests

##### 4.4.1 Disk diffusion method

Agar diffusion testing (Kirby-Bauer testing<sup>142</sup>) was performed to test the sensitivity of various bacteria to our synthetic products: Ac-KPV-NH<sub>2</sub>, H-K( $\epsilon$ G\*)PV-NH<sub>2</sub> and  $\alpha$ G\*-KPv-NH<sub>2</sub>. The bacteria tested were W3110: *E. coli* K12; WBB06: *E. coli*  $\Delta$ (*rfaC-rfaF*) (an *E. coli* strain missing two genes involved in

lipopolysaccharide synthesis and sensitive to numerous antibiotics);<sup>143</sup> *Salmonella enterica*; *Vibrio cholera*, C6706; and *Staphylococcus aureus* USA300. Ampicillin and water were used as positive and negative controls, respectively. Disks of filter paper were spotted with 5  $\mu$ L of water or 10 mg/mL solutions of ampicillin, Ac-KPV-NH<sub>2</sub>, H-K( $\epsilon$ G\*)PV-NH<sub>2</sub> or  $\alpha$ G\*-KPV-NH<sub>2</sub>. Each bacterium and compound was tested by placing the disks on a lawn of bacterial cells and incubating overnight at 37 °C. Visual inspection of the plates for zones of inhibited cell growth around each disc was used to determine the sensitivity of the bacteria for each compound.

#### 4.4.2 Colony forming units

An overnight culture of *Staphylococcus aureus* JE2<sup>144</sup> was grown in Luria-Bertani (LB) medium. 1 mL was centrifuged and washed in MOPS minimal medium (Teknova Inc., Hollister CA) and resuspended in the same medium. Cells were diluted to approximately 10<sup>7</sup> cells/mL in MOPS medium, MOPS medium + 10 mM CaCl<sub>2</sub>, MOPS medium + 10% laked horse blood (Thermo-Fisher Scientific), or MOPS medium + 10 mM CaCl<sub>2</sub> + 10% laked horse blood both with and without 0.1 mM Ac-KPV-NH<sub>2</sub> (Bachem, Bubendorf, Switzerland). Cells were incubated for 2 hours at 37 °C at which time viability was determined by colony counting of serially diluted cells.

#### 4.4.3 Liquid culture

*Staphylococcus aureus* cells were cultured overnight in 2 mL Luria-Bertani broth medium in presence of Ac-KPV-NH<sub>2</sub> (50  $\mu$ g/mL, 100  $\mu$ g/mL, 200  $\mu$ g/mL, 500  $\mu$ g/mL, 1000  $\mu$ g/mL),  $\alpha$ G\*-KPV-NH<sub>2</sub> (50  $\mu$ g/mL), K( $\epsilon$ G\*)PV-NH<sub>2</sub> (50  $\mu$ g/mL), and ampicillin (10  $\mu$ g/mL). Visual inspection of the cell density in the LB medium was used to assess the inhibition.

### 4.5. Anti-inflammatory tests

#### 4.5.1 Adipocytes cell culture, treatment with KPV tripeptides, and TNF $\alpha$ induction

A monolayer of murine 3T3-L1 adipocyte cells<sup>145</sup> were grown in 4 mL Dulbecco's Modified Eagle's Medium (DMEM) containing 10% bovine serum. The cells were serum deprived by changing the media to DMEM containing 0.3% bovine serum albumin (BSA) 20 h before treatment with the peptides. The cells were treated with the KPV tripeptide derivatives, **26**, **27**, or **29** (25  $\mu$ M), at 37 °C for 2h, and inflammation

was induced by addition of TNF $\alpha$  (0.5 nM) with incubation at 37 °C for 20 min. The cells were fractionated to separate the cytosolic and nucleic proteins.

#### **4.5.2 Cell fractionation and SDS PAGE analysis**

Culture plates were aspirated and 2 mL nuclear homogenization buffer (NHB) containing 20 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, with protease inhibitors (1  $\mu$ M phenylmethylsulfonyl fluoride, 1  $\mu$ M 1,10-phenanthroline, 10  $\mu$ M leupeptin, 50 mU trypsin inhibitory aprotinin) and phosphatase inhibitors (0.1  $\mu$ M NaF, 1 mM Na<sub>3</sub>VO<sub>4</sub> and 1 mM Na<sub>2</sub>MoO<sub>4</sub>) was added. The cells were scrapped off and transferred to 15 mL tubes. Nonidet P-40 was added to a 0.5% final concentration. The mixture was homogenized in a Dounce homogenizer with 16 strokes and centrifuged at 3500 rpm, 4 °C for 5 min. 1 mL of the supernatant (cytosol) was transferred to a clean tube and the rest discarded. The pellet was washed by resuspending in NHB buffer (1 mL), centrifuged and the supernatant discarded. The washed pellet was then suspended using a P1000 pipetor in 600  $\mu$ L NEB buffer (20 mM HEPES pH 7.9, 420 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 25% glycerol) to yield the nuclear proteins. The cytosolic and nucleic extracts were frozen at -80 °C for 15 min. The extracts were thawed and 1.5  $\mu$ L of DNase added to the nuclear extracts. The mixture was passed through a 20 gauge needle 5x to homogenize and centrifuged at 10,000 rpm, 4 °C for 10 min. The supernatant (nuclear extract) and the cytosol were quantified by BCA assay.

#### **4.5.3 Gel electrophoresis, Western blot and imaging**

Extracted proteins (100  $\mu$ g) and the loading controls (PDC-E<sub>2</sub> and ERK) were loaded and separated on an SDS-polyacrylamide gel (7.5%). The protein bands were transferred to a nitrocellulose membrane with 192 mM glycine, 25 mM Tris, and 20% methanol.<sup>145</sup> The nitrocellulose was removed and dried on a Whatman filter paper, and the molecular weight of the bands marked. Strips containing proteins of interest were cut from the nitrocellulose and blocked with 4% nonfat milk (overnight, 4 °C) or 1 h at room temperature. Strips were washed two times for 5 min with 1X Tris buffered saline (TBS-T) and incubated with 15 mL of the primary antibody solution (in a Kapak “seal a meal” bag) for 90 min at rt in the orbital shaker (150 rpm). Rinsing was done 3x using 1X TBST-T for 15 min each, and the strips incubated with the secondary antibody, antibody-HRP (horseradish peroxidase) conjugate, for 90 min, rt, 150 rpm. The blots were rinsed 3x for 10 min each in 1X TBS-T solution and visualized by enhanced chemiluminescence.

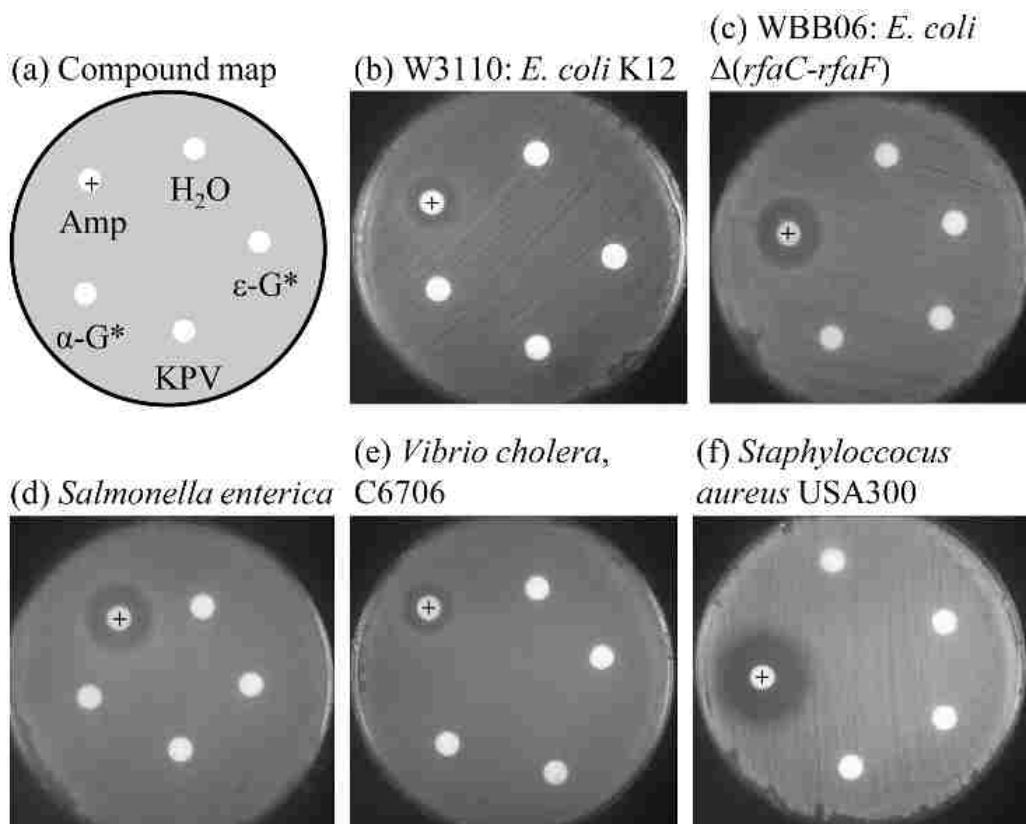
The imaging process was done in the dark room by incubating the blot for 2 min with the luminol- peroxidase mixture. The strips were wrapped in a plastic bag and glow-in-the-dark labels placed on the top right corner. The film was opened in the dark and placed in the developer cartridge for one to five minutes after which the film was developed.

#### 4.6 Enzymatic stability tests

Pronase incubation solutions were prepared in D<sub>2</sub>O by adding 20  $\mu$ L of 1M NH<sub>4</sub>HCO<sub>3</sub> and 40  $\mu$ L of 50 mM CaCl<sub>2</sub> to a solution of the tripeptide (200  $\mu$ g Ac-KPV-NH<sub>2</sub> and K(G\*)PV-NH<sub>2</sub> or 100  $\mu$ g G\*-KPV-NH<sub>2</sub> in 300  $\mu$ L D<sub>2</sub>O). The pH of the resulting mixture was adjusted to 7.0 with 3.7% HCl (10-12  $\mu$ L). The volume was adjusted to 395  $\mu$ L and <sup>1</sup>H-NMR spectra were recorded on a 500 MHz Bruker instrument. Pronase (VWR, *Streptomyces griseus*,  $\geq$ 45,000 proteolytic units/g dry weight) solution (5  $\mu$ L of 2 mg/mL stock solution) was added to the incubation solution and <sup>1</sup>H NMR spectra were recorded at 15 min intervals for 3 h. The first hour of incubation was at room temperature, after which the reaction was warmed to 37 °C and spectra were recorded at this temperature for the next 2 h. The reaction was then incubated at 37 °C and monitored by recording the spectra at additional times up to 48 h, 96 h, and 24 h for **26**, **29**, and **27**, respectively.

#### 4.7 Discussion

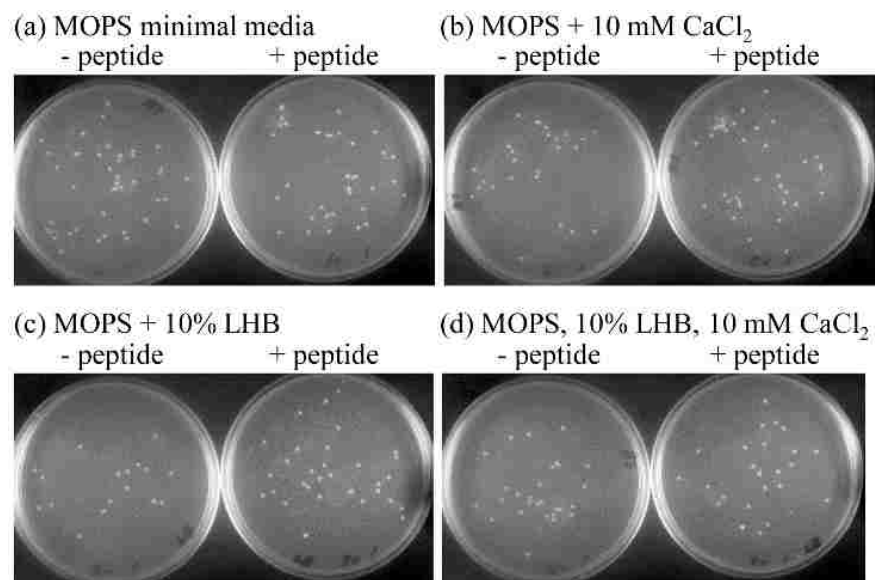
The antimicrobial activity of the three compounds, **26**, **27**, and **29**, were tested by Pradip Panta, a graduate student in Dr. William Doerrler's laboratory (Department of Biological Sciences, LSU), using the agar diffusion method against bacterial strains *E. coli* K12; WBB06: *E. coli*  $\Delta$ (*rfaC-rfaF*); *Salmonella enterica*; *Vibrio cholera*, C6706; and *Staphylococcus aureus* USA300. Ampicillin was used as a positive control. None of the H-KPV-NH<sub>2</sub> derivatives inhibited growth; however, an inhibition zone was observed for ampicillin (Figure 44). The discs with the positive control of ampicillin (indicated with a + on the disc) are surrounded by zones where cell growth was inhibited. No inhibition zones are visible for the negative control, water, or any of the tripeptides. These results were surprising because Ac-KPV-NH<sub>2</sub> has been reported as an anti-microbial agent<sup>116, 134, 136</sup> To verify the activity of Ac-KPV-NH<sub>2</sub>, the peptide was purchased from Bachem (Bubendorf, Switzerland), the same supplier as was used in Charnley *et al.*<sup>134</sup> The same protocols were followed as reported by Cutuli *et al.*<sup>116, 134</sup> but negative results were observed again.



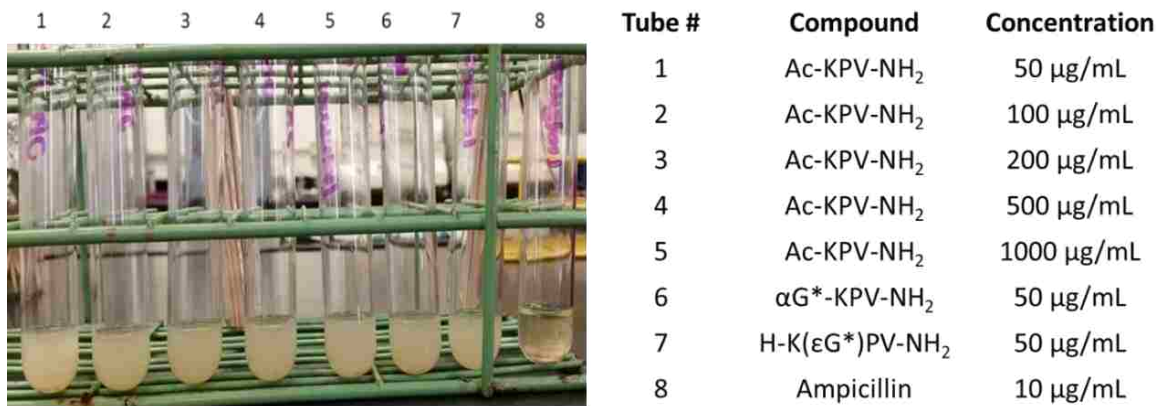
**Figure 44.** (a) Compounds, spotted on filter disks and placed on cultures, include ampicillin (designated by +), water, H-K( $\epsilon$ G\*)PV-NH<sub>2</sub>, Ac-KPV-NH<sub>2</sub>, and  $\alpha$ G\*-KPV-NH<sub>2</sub> (starting with ampicillin and moving clockwise). Several strains of bacteria were tested: (b) W3110: *E. coli* K12; (c) WBB06: *E. coli*  $\Delta(rfaC-rfaF)$ ; (d) *Salmonella enterica*; (e) *Vibrio cholera*, C6706; and (f) *Staphylococcus aureus* USA300.

Additional tests were conducted to compare the colony forming units of the cells exposed or not exposed to the Ac-KPV-NH<sub>2</sub>. *Staphylococcus aureus* cells grown overnight and suspended in MOPS minimal media were exposed to the Ac-KPV-NH<sub>2</sub> for 2 h and cultured in LB agar plates. Visual counts of the colony forming units showed no difference between the cells exposed (+ peptide) or not exposed (- peptide) to the peptide. The images of plates used for colony counting are shown in Figure 45. Exposure of the cells to the commercial Ac-KPV-NH<sub>2</sub> at 0.1 mM did not affect the viability of the cells. Media with calcium and laked horse blood were also tested, but these additives did not alter the negative results.

Lastly, the cells were grown in LB broth with varying concentration of the peptide and ampicillin as positive control. The respective culture tubes are shown in Figure 46. None of the KPV derivatives hindered growth.



**Figure 45.** *Staphylococcus aureus* JE2 cells were incubated in the absence (- peptide) and presence (+ peptide) of 0.1 mM Ac-KPV-NH<sub>2</sub> in various buffers: (a) MOPS minimal medium, (b) MOPS medium + 10 mM CaCl<sub>2</sub>, (c) MOPS medium + 10% laked horse blood, and (d) MOPS medium + 10 mM CaCl<sub>2</sub> + 10% laked horse blood. Viability was determined by colony counting of serially diluted cells (10 cells/plate).



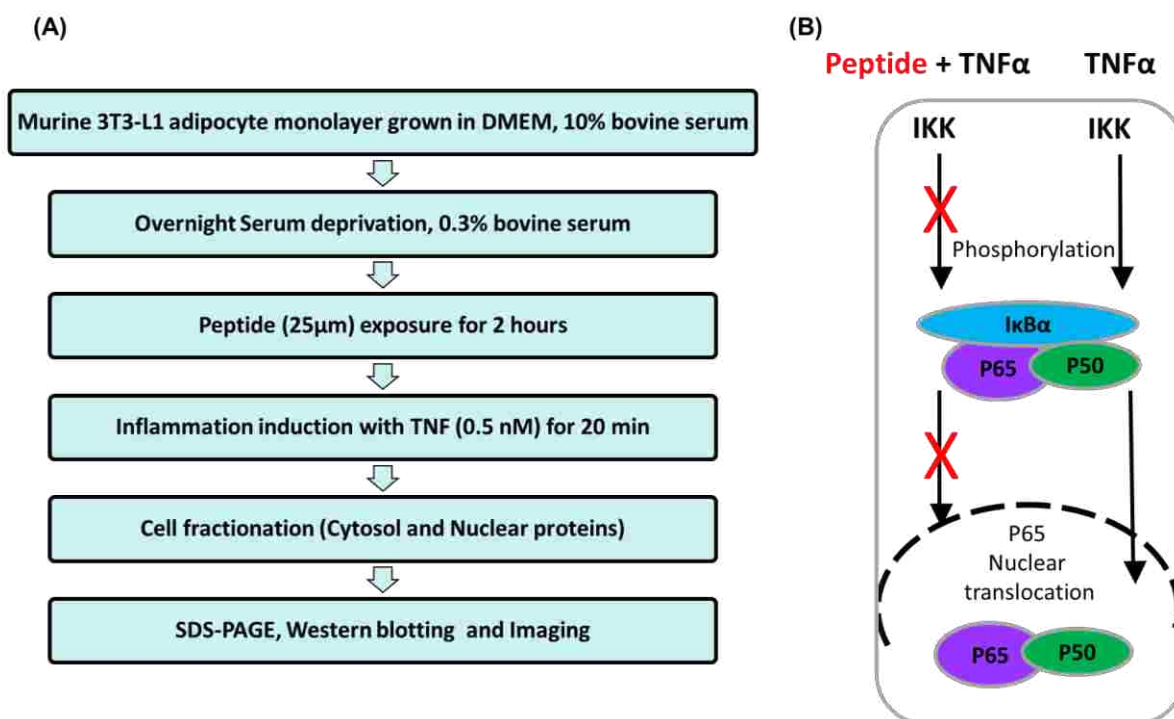
**Figure 46.** Overnight cell culture of *Staphylococcus aureus* in LB media containing KPV tripeptides at varying concentrations as indicated.

The three antimicrobial tests discussed above were negative, contrary to earlier reports by Cutuli *et al.*<sup>116</sup>, Charnley *et al.*<sup>134</sup>, and Singh *et al.*<sup>136</sup>. Our findings paralleled the later report by Lau *et al.*<sup>146</sup>, which showed no antimicrobial activity. Additionally, previous tests by Rauch *et al.* failed to show the antimicrobial activity of the Ac-KPV-NH<sub>2</sub> against *Candida albicans*. Conflicting reports in the literature are summarized in Table 8.

**Table 8.** Conflicting reports on the antimicrobial activity of Ac-KPV-NH<sub>2</sub>

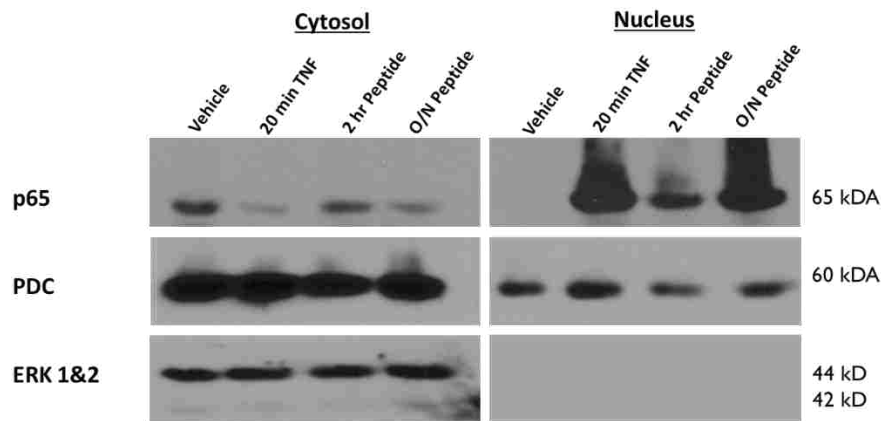
Pathogen	Activity	Reference
MRSA USA 100, MRSA USA 300	No activity (>100 μM)	Lau <i>et al.</i> 2015 <sup>146</sup>
<i>S. aureus</i>	Activity (nM- μM)	Charnley <i>et al.</i> 2008 <sup>134</sup>
<i>S. aureus</i> , <i>Candida albicans</i>	Activity (fM- μM)	Cutuli <i>et al.</i> 2000 <sup>116</sup>
MSSA ATCC 29213, MRSA ATCC 33591	Activity (nM- μM)	Singh <i>et al.</i> 2011 <sup>136</sup>

After failing to demonstrate antimicrobial activity for Ac-KPV-NH<sub>2</sub>, its anti-inflammatory activity was investigated. The anti-inflammatory experiments were conducted in Dr. Jacqueline Stephens' laboratory at the Pennington Biomedical Research Center, with the help of Jasmine Burrell (graduate student) and Dr. Allison Richard (postdoctoral associate). The inhibition of nuclear factor kappa B (NFκB) and the translocation of p65 to the nucleus by the KPV peptides was investigated. Steps in the procedure and the peptide inhibition mechanism are illustrated in Figure 47.

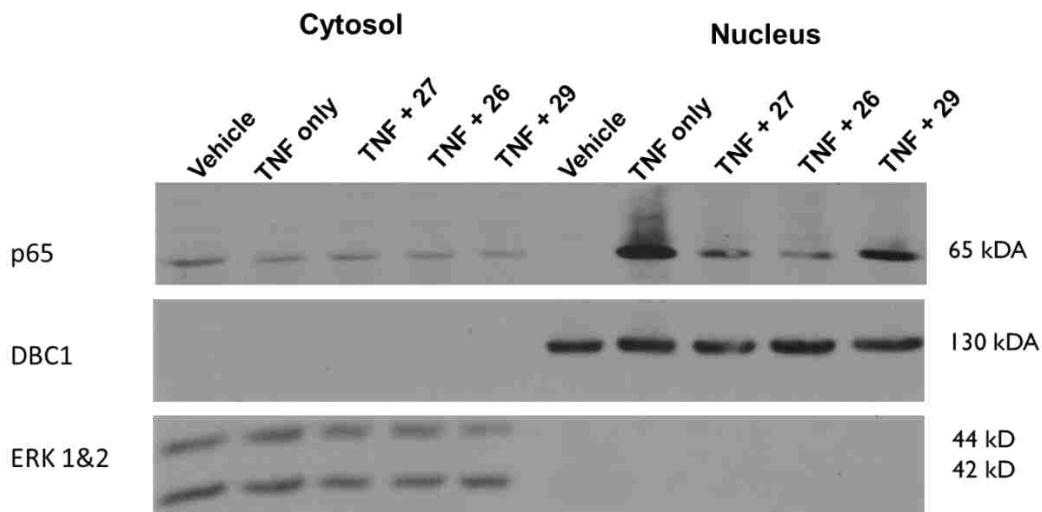


**Figure 47.** A) Steps in the procedure for determination of p65 localization; B) expected inhibition mechanism of p65 translocation to the nucleus (right). IKK, inhibitor kappa B kinase; TNF, tumor necrosis factor.

The gel images (Figures 48 and 49) indicate some degree of inhibition of p65 translocation by Ac-KPV-NH<sub>2</sub> and the glycopeptides. However, these results could not be reproduced in subsequent experiments, including positive results for the Ac-KPV-NH<sub>2</sub> control. There is a need to optimize parameters, such as the peptide concentration and exposure time, in order to make accurate conclusions. There is also a possibility that the adipocytes have a different response pathway to TNF $\alpha$  induction; such as the inhibition of nitric oxide production.



**Figure 48.** Western blot images showing the effect of Ac-KPV-NH<sub>2</sub> on the translocation of p65 to the nucleus. ERK 1&2, extracellular signal regulated kinase (cytosol protein); PDC, phosducin.



**Figure 49.** Western blot images showing the effects of the tripeptides, **26**, **27**, and **29**, on p65 translocation to the nucleus. N, Nuclear proteins; C, Cytosol proteins; ERK 1&2, extracellular signal regulated kinase (cytosol protein); DBC1, Deleted brain cancer 1 (nuclear protein). ERK and DBC1 were used as gel loading standards.



Since the antimicrobial and anti-inflammatory tests were negative and not reproducible, respectively, we investigated the proteolytic stability of the modified tripeptides, **26** and **29**, compared to the parent tripeptide, **27**. The peptides were exposed to pronase, and the peptide cleavage reactions were monitored by NMR spectroscopy. Pronase is a mixture of proteases and aminopeptidases that cleave proteins to their constituent amino acids. The peptide bonds in Ac-KPV-NH<sub>2</sub> (20:1 peptide-pronase w/w) were cleaved in less than 24 h. This result is indicated by the upfield shifts of proline (about 0.2 ppm) and valine (about 0.5 ppm) H $\alpha$  peaks in the <sup>1</sup>H NMR spectrum, Figure 50C. The shift is attributed to the removal of the electron-withdrawing N-acyl group. Additionally, the change in proline's H $\alpha$  peak from the apparent triplet to a doublet of doublet is an indicator of a conformational change in the pyrrolidine ring. The pronase reaction with the alpha modified tripeptide **26** (10:1 peptide-pronase w/w) was monitored for 35 h and no change in the spectrum was observed, Figure 50A. To further test this stability, the ratio was adjusted to 5:1 peptide-pronase w/w, and the reaction was monitored for seven days. There was no change in the spectrum, confirming that the peptide is highly resistant to pronase cleavage.

The reaction of pronase with the  $\epsilon$ -glycoalkylated peptide **29** (20:1 peptide-pronase w/w) was less clear. Within 50 hours of incubation, the lysine and the piperidine diol proton peaks were shifted upfield. The broad peaks attributed to piperidine and the lysine  $\epsilon$ -CH protons may be due to incomplete reaction. The broadening can also be caused by the interaction of lysine side-chain with the autoproteolytic products of pronase. Prolonged incubation for 96 hours revealed an additional valine H $\alpha$  peak (Figure 50B) at the same chemical shift of valine H $\alpha$  in the cleaved Ac-KPV-NH<sub>2</sub>. Based on the ratio of peak intensities between the two valine H $\alpha$  peaks (original and cleaved), the pronase cleaved one-third of the peptide at the P-V bond under this condition. Since the proline peaks were unchanged, it suggests that the K-P peptide bond is stable and the P-V bond is starting to cleave (additional NMR spectra provided in Appendix B-5).

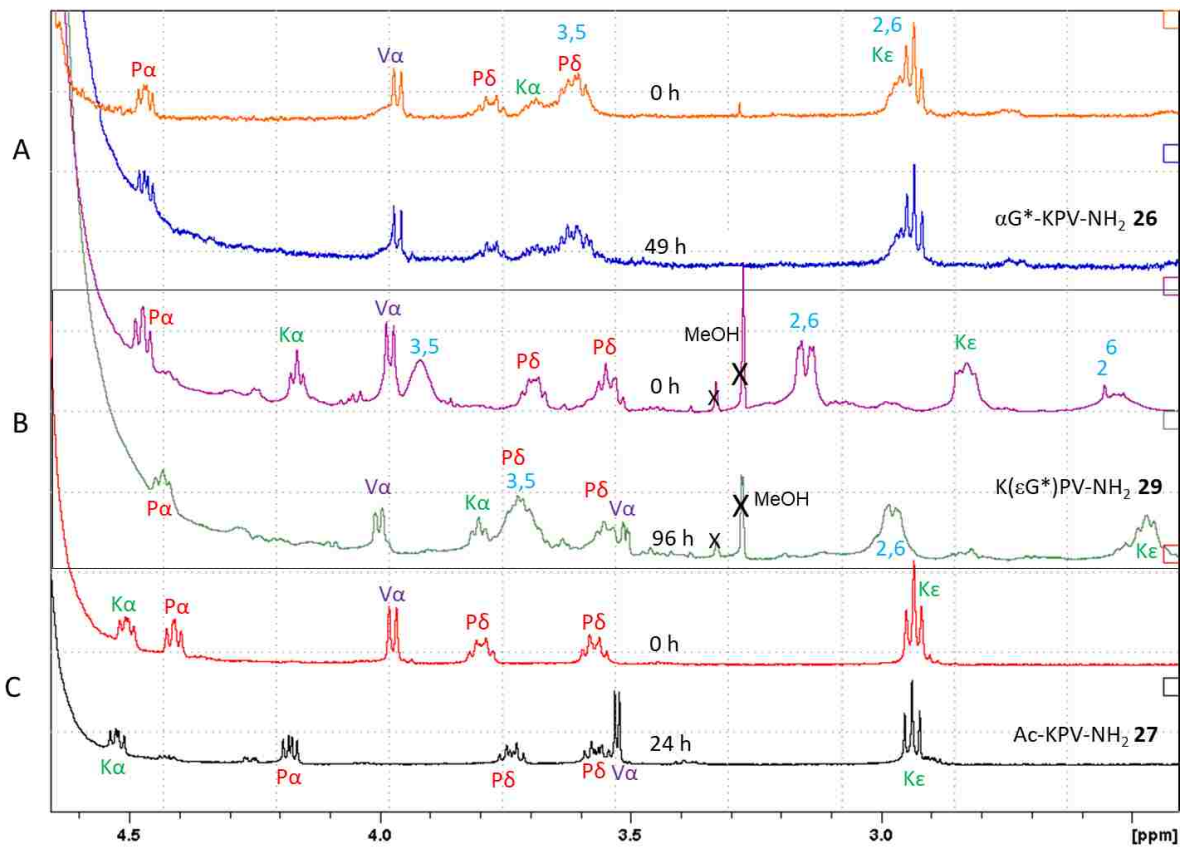
#### 4.8. Conclusion

The lysine derivative, Ts-Lys-OMe, was successfully modified by double reductive alkylation with a sugar derivative at the epsilon amine. The modification was confirmed by a crystal structure showing the location and conformation of the dihydroxy piperidine ring. The same modification was applied to the tripeptide sequence, KPV, the C-terminal sequence of the peptide  $\alpha$ -MSH. A regioselective glycoalkylation

was accomplished independently at each of the alpha and epsilon amines of lysine. The two glycoalkylated peptides and the parent peptide did not show any antimicrobial activity against *Staphylococcus aureus* under the conditions tested; even though the same protocol and conditions were followed to test the antimicrobial activity as reported in literature.<sup>116, 134</sup> The unexpected results were corroborated by a 2015 report by Lau *et al*<sup>146</sup>, that was published during our studies. They also observed no antimicrobial activity for Ac-KPV-NH<sub>2</sub> against *Staphylococcus aureus*.

The effect of the “glycoalkylation” modification was alternatively assessed through a proteolytic stability assay. Glycoalkylation induced resistance of αG-KPV-NH<sub>2</sub> against proteolytic cleavage by pronase. The parent peptide was cleaved in less than 24 hours while the epsilon glycoalkylated peptide H-K(εG\*)PV-NH<sub>2</sub> was resistant to cleavage for several days.

Considering the amount of effort directed to the synthesis of the glycopeptides, it was very disappointing to learn of the inactivity of the parent peptide. It is now apparent that it would have been prudent to test the antimicrobial activity of Ac-KPV-NH<sub>2</sub> before proceeding to synthesize the glycoalkylated analogs. This strategy should be applied to other selected peptide sequences used in the future for glycoalkylation. While there were several literature reports indicating the antimicrobial activity of Ac-KPV-NH<sub>2</sub>, most primary reports were given by the same group of authors and the others were review articles.



**Figure 50.** Stability of the tripeptides to pronase cleavage, monitored by  $^1\text{H}$  NMR at 500 MHz in  $\text{D}_2\text{O}$ . (A) Compound **26**; (B) Compound **29**; and (C) Compound **27**. Time of incubation is shown in hours at the center of each spectrum.

## CHAPTER 5 CONCLUSION

Based on the NMR analysis of the Scc4 protein, it was determined that the backbone structure for the Scc4 homodimer protein was different from that of the Scc4 heterodimer (in complex with Scc1His). While there is a need to solve the 3D structure of the Scc4 homodimer for transcription regulation studies, determination of the binding interface should be studied using the Scc1His-Scc4 complex and the mixed complex samples. Two protocols were developed for purification of untagged Scc4 and for selective isotopic labeling of Scc4 or Scc1His in Scc1His:Scc4 complex. A methodology was developed for triple resonance backbone assignment for individual proteins from the NMR data of the complex. The backbone assignment task was initialized; however, more NMR data is needed to complete backbone assignment and to determine the binding interface.

The *in vitro* disruption of the binding interface in Scc1His:Scc4 protein complex using sarkosyl, indicates that there is possibility that the complex can serve as a pathogenic protein target. Identification of the interface epitopes will help in designing drugs for *in vivo* complex disruption. These drugs will inhibit types III secretion system processes thus limiting chlamydial invasion and replication.

The structural modification of the tripeptide KPV by reductive glycoalkylation was accomplished. The modification was shown to confer proteolytic stability to the glycopeptides against pronase cleavage. The effect of the modification on the biological activity of the peptides could not be determined. The initial positive anti-inflammatory results indicated that the peptides may have anti-inflammatory effects. Optimization of the conditions, such as incubation time and peptide concentration, may enhance the reproducibility. The choice of the cell line may also play a major role in determining the inflammation pathway.

Glycoalkylation should be investigated for broad application in the structural modification of lysine containing peptide drugs. Considering that pronase is a mixture of proteases and amino peptidases, the stability to pronase cleavage by the modified peptides imply that glycoalkylation could be used to enhance oral bioavailability of peptide drugs.

## REFERENCES

1. Lara, J. M. Intracellular management of information: From DNA to proteins. *Triple C: Communication, Capitalism & Critique* **2009**, *7*, 376-385.
2. Guarracino, D. A.; Gentile, K.; Grossman, A.; Li, E.; Refai, N.; Mohnot, J.; King, D. Salt-bridging effects on short amphiphilic helical structure and introducing sequence-based short beta-turn motifs. *Journal of Biomolecular Structure & Dynamics* **2018**, *36*, 475-485.
3. Ponomarenko, E. A.; Poverennaya, E. V.; Ilgisonis, E. V.; Pyatnitskiy, M. A.; Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V.; Archakov, A. I. The size of the human proteome: The width and depth. *International Journal of Analytical Chemistry* **2016**, *2016*, 7436849-6p.
4. Gonzalez, M. W.; Kann, M. G. Chapter 4: Protein interactions and disease. *Plos Computational Biology* **2012**, *8*, e1002819
5. Bull, S. C.; Doig, A. J. Properties of protein drug target classes. *PLoS ONE* **2015**, *10*, e0117955.
6. Sosa, E. J.; Burguener, G.; Lanzarotti, E.; Defelipe, L.; Radusky, L.; Pardo, A. M.; Marti, M.; Turjanski, A. G.; Fernández DoPorto, D. Target-pathogen: A structural bioinformatic approach to prioritize drug targets in pathogens. *Nucleic Acids Research* **2018**, *46*, D413-D418.
7. Green, E. R.; Meccas, J. Bacterial secretion systems – An overview. *Microbiology spectrum* **2016**, *4*, 10.1128/microbiolspec.VMBF-0012-2015.
8. Russo, A. F. Overview of neuropeptides: Awakening the senses? *Headache* **2017**, *57 Suppl 2*, 37-46.
9. Zanin, J. P.; Unsain, N.; Anastasia, A. Growth factors and hormones pro-peptides: the unexpected adventures of the BDNF prodomain. *Journal of Neurochemistry* **2017**, *141*, 330-340.
10. Andersen, D. C.; Krummen, L. Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology* **2002**, *13*, 117-123.
11. Ung, P.; Winkler, D. A. Tripeptide motifs in biology: targets for peptidomimetic design. *Journal of Medicinal Chemistry* **2011**, *54*, 1111-1125.
12. Brzoska, T.; Luger, T. A.; Maaser, C.; Abels, C.; Bohm, M. Alpha-melanocyte-stimulating hormone and related tripeptides: Biochemistry, antiinflammatory and protective effects *in vitro* and *in vivo*, and future perspectives for the treatment of immune-mediated inflammatory diseases. *Endocrine Reviews* **2008**, *29*, 581-602.
13. Myers, R. D. Neuroactive peptides: Unique phases in research on mammalian brain over three decades. *Peptides* **1994**, *15*, 367-381.
14. Hartmann, R.; Meisel, H. Food-derived peptides with biological activity: from research to food applications. *Current Opinion in Biotechnology* **2007**, *18*, 163-169.
15. Korhonen, H.; Pihlanto, A. Bioactive peptides: Production and functionality. *International Dairy Journal* **2006**, *16*, 945-960.

16. Mains, R. E.; Eipper, B. A. Chapter 20 - Peptides A2 - Brady, Scott T. In: *Basic Neurochemistry* (8<sup>th</sup> edition). Edited by Siegel, G. J.; Albers, R. W.; Price, D. L. New York: Academic Press; **2012**, 390-407.
17. Latham, P. W. Therapeutic peptides revisited. *Nature. Biotechnology*. **1999**, *17*, 755-757.
18. Marqus, S.; Pirogova, E.; Piva, T. J. Evaluation of the use of therapeutic peptides for cancer treatment. *Journal of Biomedical Science* **2017**, *24*, 21.
19. Daliri, E. B.-M.; Oh, D. H.; Lee, B. H. Bioactive peptides. *Foods* **2017**, *6*, 32.
20. Singh, M.; Mukhopadhyay, K. Alpha-melanocyte stimulating hormone: an emerging anti-inflammatory antimicrobial peptide. *BioMed Research International* **2014**, *2014*, 10.
21. Matsuzaki, K. Why and how are peptide–lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. *Biochimica et Biophysica Acta - Biomembranes* **1999**, *1462*, 1-10.
22. Nguyen, L. T.; Haney, E. F.; Vogel, H. J. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends in Biotechnology*, *29*, 464-472.
23. Pálffy, R.; Gardlík, R.; Behuliak, M.; Kadasi, L.; Turna, J.; Celec, P. On the physiology and pathophysiology of antimicrobial peptides. *Molecular Medicine* **2009**, *15*, 51-59.
24. Yeaman, M. R.; Yount, N. Y. Mechanisms of antimicrobial peptide action and resistance. *Pharmacological Reviews* **2003**, *55*, 27-55.
25. Zhang, M.; Zhao, J.; Zheng, J. Molecular understanding of a potential functional link between antimicrobial and amyloid peptides. *Soft Matter* **2014**, *10*, 7425-7451.
26. Otvos, L.; Wade, J. D. Current challenges in peptide-based drug discovery. *Frontiers in Chemistry* **2014**, *2*, 62.
27. Böttger, R.; Hoffmann, R.; Knappe, D. Differential stability of therapeutic peptides with different proteolytic cleavage sites in blood, plasma and serum. *PLoS ONE* **2017**, *12*, e0178943.
28. Ahmed, B.; Robert, L.; Martin, M.; Gerhard, W. Challenges for PEGylated proteins and alternative half-life extension technologies based on biodegradable polymers. In: *Tailored Polymer Architectures for Pharmaceutical and Biomedical Applications*. vol. 1135: American Chemical Society; **2013**, 215-233.
29. Ikeda, Y.; Katamachi, J.; Kawasaki, H.; Nagasaki, Y. Novel protein PEGylation chemistry via glutaldehyde-functionalized PEG. *Bioconjugate Chemistry* **2013**, *24*, 1824-1827.
30. Moosmann, A.; Blath, J.; Lindner, R.; Müller, E.; Böttinger, H. Aldehyde PEGylation kinetics: a standard protein versus a pharmaceutically relevant single chain variable fragment. *Bioconjugate Chemistry* **2011**, *22*, 1545-1558.
31. Fernandes, A. I.; Gregoriadis, G. The effect of polysialylation on the immunogenicity and antigenicity of asparaginase: Implication in its pharmacokinetics. *International Journal of pharmaceutics* **2001**, *217*, 215-224.
32. Jain, S.; Hreczuk-Hirst, D. H.; McCormack, B.; Mital, M.; Epenetos, A.; Laing, P.; Gregoriadis, G. Polysialylated insulin: Synthesis, characterization and biological activity *in vivo*. *Biochimica et Biophysica Acta - General Subjects* **2003**, *1622*, 42-49.

33. Pasut, G. Polymers for protein conjugation. *Polymers* **2014**, *6*, 160-178.
34. Treib, J.; Baron, J. F.; Grauer, M. T.; Strauss, R. G. An international view of hydroxyethyl starches. *Intensive Care Medicine* **1999**, *25*, 258-268.
35. Schellenberger, V.; Wang, C.-w.; Geething, N. C.; Spink, B. J.; Campbell, A.; To, W.; Scholle, M. D.; Yin, Y.; Yao, Y.; Bogin, O.; Cleland, J. L.; Silverman, J.; Stemmer, Willem, P. C. A recombinant polypeptide extends the *in vivo* half-life of peptides and proteins in a tunable manner. *Nature Biotechnology* **2009**, *27*, 1186-1190.
36. Mahajan, A.; Rawat, A. S.; Bhatt, N.; Chauhan, M. K. Structural modification of proteins and peptides. *Indian Journal of Pharmaceutical Education and Research* **2014**, *48*, 34-47.
37. Wang, J.; Breslow, E.; Sykes, B. D. Differential binding of desmopressin and vasopressin to neurophysin-II. *Journal of Biological Chemistry* **1996**, *271*, 31354-31359.
38. Qureshi, S.; Galiveeti, S.; Bichet, D. G.; Roth, J. Diabetes insipidus: Celebrating a century of vasopressin therapy. *Endocrinology* **2014**, *155*, 4605-4621.
39. Trüssel, S.; Dumelin, C.; Frey, K.; Villa, A.; Buller, F.; Neri, D. New strategy for the extension of the serum half-life of antibody fragments. *Bioconjugate Chemistry* **2009**, *20*, 2286-2292.
40. Verdine, G. L.; Hilinski, G. J. Stapled peptides for intracellular drug targets. *Methods in Enzymology* **2012**, *503*, 3-33.
41. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. The future of peptide-based drugs. *Chemical Biology & Drug Design* **2013**, *81*, 136-147.
42. Pavlou, A. K.; Reichert, J. M. Recombinant protein therapeutics--success rates, market trends and values to 2010. *Nature Biotechnology* **2004**, *22*, 1513-1519.
43. Jozala, A. F.; Geraldles, D. C.; Tundisi, L. L.; Feitosa, V. d. A.; Breyer, C. A.; Cardoso, S. L.; Mazzola, P. G.; Oliveira-Nascimento, L. d.; Rangel-Yagui, C. d. O.; Magalhães, P. d. O.; Oliveira, M. A. d.; Pessoa, Adalberto. Biopharmaceuticals from microorganisms: From production to purification. *Brazilian Journal of Microbiology* **2016**, *47*, 51-63.
44. Biopharmaceuticals Market-Growth, Trends & Forecasts (2017-2022) [<https://www.mordorintelligence.com/industry-reports/global-biopharmaceuticals-market-industry>]
45. Leader, B.; Baca, Q. J.; Golan, D. E. Protein therapeutics: A summary and pharmacological classification. *Nature Reviews Drug Discovery* **2008**, *7*, 21-39.
46. Demain, A. L.; Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances* **2009**, *27*, 297-306.
47. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discovery Today* **2010**, *15*, 40-56.
48. Farrokhi, N.; Hrmova, M.; Burton, R. A.; Fincher, G. B. Heterologous and cell free protein expression systems. *Methods in Molecular Biology* **2009**, *513*, 175-198.
49. Bennett, P. M. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology* **2008**, *153*, S347-S357

50. Khan, K. H. Gene expression in mammalian cells and its applications. *Advanced Pharmaceutical Bulletin* **2013**, 3, 257-263.
51. Gerngross, T. U. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nature Biotechnology* **2004**, 22, 1409-1414.
52. Jarvis, D. L. Baculovirus-insect cell expression systems. *Methods in Enzymology* **2009**, 463, 191-222.
53. Kost, T. A.; Condreay, J. P.; Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nature Biotechnology* **2005**, 23, 567-575.
54. Tekoah, Y.; Shulman, A.; Kizhner, T.; Ruderfer, I.; Fux, L.; Nataf, Y.; Bartfeld, D.; Ariel, T.; Gingis-Velitski, S.; Hanania, U.; Shaaltiel, Y. Large-scale production of pharmaceutical proteins in plant cell culture—the protalix experience. *Plant Biotechnology Journal* **2015**, 13, 1199-1208.
55. Bloch, F. Nuclear induction. *Physical Review* **1946**, 70, 460-474.
56. Alvarez, L. W.; Bloch, F. A quantitative determination of the neutron moment in absolute nuclear magnetons. *Physical Review* **1940**, 57, 111-122.
57. Pake, G. E. Fundamentals of nuclear magnetic resonance absorption .1. *American Journal of Physics* **1950**, 18, 438-452.
58. Kleckner, I. R.; Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta* **2011**, 1814, 942-968.
59. Atta ur, R.; Choudhary, M. I.; Atia tul, W. Chapter 2 - Creating NMR Signals. In: *Solving problems with NMR spectroscopy (2<sup>nd</sup> Edition)*. Edited by Atta ur, R.; Choudhary, M. I.; Atia tul, W. Boston: Academic Press; **2016**, 35-98.
60. Bodenhausen, G.; Kogler, H.; Ernst, R. R. Selection of coherence-transfer pathways in NMR pulse experiments. *Journal of Magnetic Resonance (1969)* **1984**, 58, 370-388.
61. Shapiro, L. M.; Matzat, S. J.; Gold, G. E. 41 - Functional magnetic resonance imaging. In: *Rheumatology (6<sup>th</sup> Edition)*. Edited by Hochberg, M. C.; Silman, A. J.; Smolen, J. S.; Weinblatt, M. E.; Weisman, M. H. Philadelphia. **2015**, 322-330.
62. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of exchange processes by 2-dimensional NMR-spectroscopy. *Journal of Chemical Physics* **1979**, 71, 4546-4553.
63. Morris, G. A.; Freeman, R. Selective excitation in Fourier transform nuclear magnetic resonance. *Journal of Magnetic Resonance* **2011**, 213, 214-243.
64. Rinaldi, P. L. Three-dimensional solution NMR spectroscopy of complex structures and mixtures. *Analyst* **2004**, 129, 687-699.
65. Atta ur, R.; Choudhary, M. I.; Atia tul, W. Chapter 8 - Playing with dimensions in NMR spectroscopy. In: *Solving Problems with NMR Spectroscopy (2<sup>nd</sup> edition)*. Edited by Atta ur, R.; Choudhary, M. I.; Atia tul, W. Boston: Academic Press; **2016**, 387-413.
66. Marion, D.; Wüthrich, K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins. *Biochemical and Biophysical Research Communications* **1983**, 113, 967-974.



67. Grzesiek, S.; Wingfield, P.; Stahl, S.; Kaufman, J. D.; Bax, A. Four-dimensional <sup>15</sup>N-separated NOESY of slowly tumbling perdeuterated <sup>15</sup>N-enriched proteins. Application to HIV-1 Nef. *Journal of the American Chemical Society* **1995**, *117*, 9594-9595.
68. Lee, W.; Tonelli, M.; Dashti, H.; Eghbalnia, H. R.; Markley, J. L. I-PINE, an integrative probabilistic NMR assignment system. *In preparation* (cited as requested by the authors).
69. Grzesiek, S. B. A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance* **1992**, *99*, 201-207.
70. Venters, R. A.; Thompson, R.; Cavanagh, J. Current approaches for the study of large proteins by NMR. *Journal of Molecular Structure* **2002**, *602-603*, 275-292.
71. Meyer, B.; Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition* **2003**, *42*, 864-890.
72. Breeze, A. L. Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2000**, *36*, 323-372.
73. Yu, F.; Roy, S.; Arevalo, E.; Schaeck, J.; Wang, J.; Holte, K.; Duffner, J.; Gunay, N. S.; Capila, I.; Kaundinya, G. V. Characterization of heparin-protein interaction by saturation transfer difference (STD) NMR. *Analytical and Bioanalytical Chemistry* **2014**, *406*, 3079-3089.
74. Acton, T. B.; Xiao, R.; Anderson, S.; Aramini, J.; Buchwald, W. A.; Ciccocanti, C.; Conover, K.; Everett, J.; Hamilton, K.; Huang, Y. J.; Janjua, H.; Kornhaber, G.; Lu, J.; Lee, D. Y.; Liu, G.; Maglagui, M.; Ma, L.; Patel, D.; Rossi, P.; Sahdev, S.; Shastry, R.; Swapna, G. V. T.; Tang, Y.; Tong, S.; Wang, D.; Wang, H.; Zhao, L.; Montelione, G. T. Preparation of protein samples for NMR structure, function, and small molecule screening studies. *Methods in Enzymology* **2011**, *493*, 21-60.
75. Wheelhouse, N.; Longbottom, D. Endemic and emerging chlamydial infections of animals and their zoonotic implications. *Transboundary and Emerging Diseases* **2012**, *59*, 283-291.
76. Malhotra, M.; Sood, S.; Mukherjee, A.; Muralidhar, S.; Bala, M. Genital *Chlamydia trachomatis*: An update. *The Indian Journal of Medical Research* **2013**, *138*, 303-316.
77. Paavonen, J.; Eggert-Kruse, W. *Chlamydia trachomatis*: Impact on human reproduction. *Human Reproduction Update* **1999**, *5*, 433-447.
78. Cai, S.; He, F.; Samra, H. S.; de la Maza, L. M.; Bottazzi, M. E.; Joshi, S. B.; Middaugh, C. R. Biophysical and stabilization studies of the *Chlamydia trachomatis* mouse pneumonitis major outer membrane protein. *Molecular Pharmacology* **2009**, *6*, 1553-1561.
79. Beagley, K. W.; Timms, P. *Chlamydia trachomatis* infection: incidence, health costs and prospects for vaccine development. *Journal of Reproductive Immunology* **2000**, *48*, 47-68.
80. Rao, X.; Deighan, P.; Hua, Z.; Hu, X.; Wang, J.; Luo, M.; Wang, J.; Liang, Y.; Zhong, G.; Hochschild, A.; Shen, L. A regulator from *Chlamydia trachomatis* modulates the activity of RNA polymerase through direct interaction with the beta subunit and the primary sigma subunit. *Genes & Development* **2009**, *23*, 1818-1829.
81. Hanson, B. R.; Slepkin, A.; Peterson, E. M.; Tan, M. *Chlamydia trachomatis* Type III secretion proteins regulate transcription. *Journal of Bacteriology* **2015**, *197*, 3238-3244.

82. Newman, L.; Rowley, J.; Vander Hoorn, S.; Wijesooriya, N. S.; Unemo, M.; Low, N.; Stevens, G.; Gottlieb, S.; Kiarie, J.; Temmerman, M. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. *PLoS ONE* **2015**, *10*, e0143304.
83. 2016 Sexually transmitted disease surveillance [https://www.cdc.gov/std/stats16/chlamydia.htm]
84. Sandoz, K. M.; Rockey, D. D. Antibiotic resistance in *Chlamydiae*. *Future Microbiology* **2010**, *5*, 1427-1442.
85. Somani, J.; Bhullar, V. B.; Workowski, K. A.; Farshy, C. E.; Black, C. M. Multiple drug-resistant *Chlamydia trachomatis* associated with clinical treatment failure. *The Journal of Infectious Diseases* **2000**, *181*, 1421-1427.
86. Mestrovic, T.; Ljubin-Sternak, S. Molecular mechanisms of *Chlamydia trachomatis* resistance to antimicrobial drugs. *Frontier in Bioscience (Landmark edition)* **2018**, *23*, 656-670.
87. Brunham, R. C.; Kimani, J.; Bwayo, J.; Maitha, G.; Maclean, I.; Yang, C.; Shen, C.; Roman, S.; Nagelkerke, N. J.; Cheang, M.; Plummer, F. A. The epidemiology of *Chlamydia trachomatis* within a sexually transmitted diseases core group. *Journal of Infectious Disease* **1996**, *173*, 950-956.
88. de la Maza, L. M.; Zhong, G.; Brunham, R. C. Update on *Chlamydia trachomatis* vaccinology. *Clinical and Vaccine Immunology* **2017**, *24*.
89. Coburn, B.; Sekirov, I.; Finlay, B. B. Type III secretion systems and disease. *Clinical Microbiology Reviews* **2007**, *20*, 535-549.
90. Archuleta, T. L.; Du, Y.; English, C. A.; Lory, S.; Lesser, C.; Ohi, M. D.; Ohi, R.; Spiller, B. W. The *Chlamydia* effector chlamydial outer protein N (CopN) sequesters tubulin and prevents microtubule assembly. *Journal of Biological Chemistry* **2011**, *286*, 33992-33998.
91. Silva-Herzog, E.; Joseph, S. S.; Avery, A. K.; Coba, J. A.; Wolf, K.; Fields, K. A.; Plano, G. V. Scc1 (CP0432) and Scc4 (CP0033) function as a type III secretion chaperone for CopN of *Chlamydia pneumoniae*. *Journal of Bacteriology* **2011**, *193*, 3490-3496.
92. Hsia, R.-c.; Pannekoek, Y.; Ingerowski, E.; Bavoil, P. M. Type III secretion genes identify a putative virulence locus of *Chlamydia*. *Molecular Microbiology* **1997**, *25*, 351-359.
93. Fields, K. A.; Hackstadt, T. Evidence for the secretion of *Chlamydia trachomatis* CopN by a type III secretion mechanism. *Molecular Microbiology* **2000**, *38*, 1048-1060.
94. Balsara, Z. R.; Misaghi, S.; Lafave, J. N.; Starnbach, M. N. *Chlamydia trachomatis* infection induces cleavage of the mitotic cyclin B1. *Infection and Immunity* **2006**, *74*, 5602-5608
95. Shen, L.; Macnaughtan, M. A.; Frohlich, K. M.; Cong, Y.; Goodwin, O. Y.; Chou, C.-w.; LeCour, L.; Krup, K.; Luo, M.; Worthylake, D. K. Multipart chaperone-effector recognition in the type III secretion system of *Chlamydia trachomatis*. *Journal of Biological Chemistry* **2015**, *290*, 28141-28155.
96. Froger, A.; Hall, J. E. Transformation of plasmid DNA into *E. coli* using the heat shock method. *Journal of Visualized Experiments : JoVE* **2007**, 253.

97. Marley, J.; Lu, M.; Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* **2001**, *20*, 71-75.
98. Ying, J.; Delaglio, F.; Torchia, D. A.; Bax, A. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *Journal of Biomolecular NMR* **2017**, *68*, 101-118.
99. Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015**, *31*, 1325-1327.
100. Neduva, V.; Russell, R. B. DILIMOT: Discovery of linear motifs in proteins. *Nucleic Acids Research* **2006**, *34*, W350-W355.
101. Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *Journal of Chemical Information and Computer Sciences* **2001**, *41*, 856-864.
102. Zamyatnin, A. A. Fragmentomics of natural peptide structures. *Biochemistry-Moscow* **2009**, *74*, 1575-1585.
103. Mountjoy, K. G. Functions for pro-opiomelanocortin-derived peptides in obesity and diabetes. *Biochemical Journal* **2010**, *428*, 305-324.
104. Harris, R. M.; Dijkstra, P. D.; Hofmann, H. A. Complex structural and regulatory evolution of the pro-opiomelanocortin gene family. *General and Comparative Endocrinology* **2014**, *195*, 107-115.
105. Manna, S. K.; Aggarwal, B. B. Alpha-melanocyte-stimulating hormone inhibits the nuclear transcription factor NF-kappa B activation induced by various inflammatory agents. *Journal of Immunology* **1998**, *161*, 2873-2880.
106. Madhuri; Shireen, T.; Venugopal, S. K.; Ghosh, D.; Gadepalli, R.; Dhawan, B.; Mukhopadhyay, K. *In vitro* antimicrobial activity of alpha-melanocyte stimulating hormone against major human pathogen *Staphylococcus aureus*. *Peptides* **2009**, *30*, 1627-1635.
107. MacNeil, D. J.; Howard, A. D.; Guan, X. M.; Fong, T. M.; Nargund, R. P.; Bednarek, M. A.; Goulet, M. T.; Weinberg, D. H.; Strack, A. M.; Marsh, D. J. *et al* The role of melanocortins in body weight regulation: opportunities for the treatment of obesity. *European Journal of Pharmacology* **2002**, *440*, 141-157.
108. Coll, A. P. Effects of pro-opiomelanocortin (POMC) on food intake and body weight: mechanisms and therapeutic potential? *Clinical Science* **2007**, *113*, 171.
109. Lu, D. S.; Willard, D.; Patel, I. R.; Kadwell, S.; Overton, L.; Kost, T.; Luther, M.; Chen, W. B.; Woychik, R. P.; Wilkison, W. O.; Cone, R. D. Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. *Nature* **1994**, *371*, 799-802.
110. Galván, I.; Alonso-Alvarez, C. The expression of melanin-based plumage is separately modulated by exogenous oxidative stress and a melanocortin. *Proceedings of the Royal Society B: Biological Sciences* **2009**, *276*, 3089.
111. McNulty, J. C.; Jackson, P. J.; Thompson, D. A.; Chai, B.; Gantz, I.; Barsh, G. S.; Dawson, P. E.; Millhauser, G. L. Structures of the agouti signaling protein. *Journal of Molecular Biology* **2005**, *346*, 1059-1070.

112. Sankar, G. May, M. J.; Kopp, E. B. NF- $\kappa$ B and Rel proteins: Evolutionarily conserved mediators of immune responses. *Annual Review of Immunology* **1998**, *16*, 225-260.
113. Delgado, R.; Carlin, A.; Airaghi, L.; Demitri, M. T.; Meda, L.; Galimberti, D.; Baron, P.; Lipton, J. M.; Catania, A. Melanocortin peptides inhibit production of proinflammatory cytokines and nitric oxide by activated microglia. *Journal of Leukocyte Biology* **1998**, *63*, 740-745.
114. Brogden, N. K.; Mehalick, L.; Fischer, C. L.; Wertz, P. W.; Brogden, K. A. The emerging role of peptides and lipids as antimicrobial epidermal barriers and modulators of local inflammation. *Skin Pharmacology and Physiology* **2012**, *25*, 167-181.
115. Catania, A.; Colombo, G.; Rossi, C.; Carlin, A.; Sordi, A.; Lonati, C.; Turcatti, F.; Leonardi, P.; Grieco, P.; Gatti, S. Antimicrobial properties of alpha-MSH and related synthetic melanocortins. *Scientific World Journal* **2006**, *6*, 1241-1246.
116. Cutuli, M.; Cristiani, S.; Lipton, J. M.; Catania, A. Antimicrobial effects of alpha-MSH peptides. *Journal of Leukocyte Biology* **2000**, *67*, 233-239.
117. Bhattacharya, A.; Datta, A. Effect of cyclic AMP on RNA and protein synthesis in *Candida albicans*. *Biochemical and Biophysical Research Communications* **1977**, *77*, 1438-1444.
118. Sawyer, T. K.; Staples, D. J.; Castrucci, A. M. L.; Hadley, M. E.; Al-Obeidi, F. A.; Cody, W. L.; Hruby, V. J.  $\alpha$ -Melanocyte stimulating hormone message and inhibitory sequences: Comparative structure-activity studies on melanocytes. *Peptides* **1990**, *11*, 351-357.
119. Schlimme, E.; Frister, H.; Raezke, K. P. Glycosylation of mono- and bicyclic dicarboxylic acid amides. *Nucleoside & Nucleotide* **1988**, *7*, 577-580.
120. Roy, A.; Sanjayan, G. J. Sugar-amino acid cyclic conjugates as novel conformationally constrained hydroxyethylamine transition-state isosteres. *Tetrahedron Letters* **2012**, *53*, 3361-3363.
121. Jabgunde, A. M.; Kalamkar, N. B.; Chavan, S. T.; Sabharwal, S. G.; Dhavale, D. D. Synthesis of new six- and seven-membered 1-*N*-minosugars as promising glycosidase inhibitors. *Bioorganic & Medicinal Chemistry* **2011**, *19*, 5912-5915.
122. Begoña, P. M.; Fernández, F.; Estévez, J. C.; Estévez, R. J. A nitro sugar mediated synthesis of 6-amino-1,5,6-trideoxy-1,5-imino-d-glucitol (6-amino-1,6-dideoxyojirimycin). *Tetrahedron: Asymmetry* **2009**, *20*, 503-507.
123. Dhavale, D. D.; Chaudhari, V. D.; Tilekar, J. N. An expeditious synthesis of a (3*S*,4*S*,5*R*)-trihydroxyazepane. *Tetrahedron Letters* **2003**, *44*, 7321-7323.
124. Iacono, S.; Rasmussen, J. R. Deoxygenation of secondary alcohols: 3-Deoxy-1,2:5,6-Di-*O*-isopropylidene- $\alpha$ -D-ribo-hexofuranose. In: *Organic Syntheses. vol. 64*: John Wiley & Sons, Inc.; **1986**, 57-62.
125. Cui, L.; Ling, C. C.; Sadowska, J.; Bundle, D. R. Synthesis of modified *Trichinella spiralis* disaccharide epitopes and a comparison of their recognition by chemical mapping and saturation transfer difference NMR. *Carbohydrate Research* **2014**, *383*, 1-13.
126. Barton, D. H. R.; McCombie, S. W. A new method for the deoxygenation of secondary alcohols. *Journal of the Chemical Society, Perkin Transactions 1* **1975**, 1574-1585.

127. Manna, S.; Viala, J.; Yadagiri, P.; Falck, J. R. Synthesis of 12(*S*),20-, 12(*S*),19(*R*)-, and 12(*S*),19(*S*)-dihydroxyeicosa-cis-5,8,14-trans-10-tetraenoic acids, metabolites of 12(*S*)-hete. *Tetrahedron Letters* **1986**, *27*, 2679-2682.
128. Gurjar, M. K.; Nagaprasad, R.; Ramana, C. V.; Karmakar, S.; Mohapatra, D. K. Ring-closing metathesis mediated total synthesis of microcarpalide and herbarumin III. *Arkivoc* **2005**, 237-257.
129. Rauter, A. P.; Figueiredo, J.; Ismael, M.; Canda, T.; Font, J.; Figueiredo, M. Efficient synthesis of  $\alpha,\beta$ -unsaturated  $\gamma$ -lactones linked to sugars. *Tetrahedron: Asymmetry* **2001**, *12*, 1131-1146.
130. Le Bouc, G.; Thomassigny, C.; Greck, C. Diastereoselective syntheses of 1-deoxyhomonojirimycin and two new 1,5,6-trideoxy-1,5-iminoheptitols with d-allo- and l-talo-configuration. *Tetrahedron: Asymmetry* **2006**, *17*, 2006-2014.
131. Chakraborty, C.; Dhavale, D. D. Short and efficient synthesis of (2*S*,3*R*,4*R*,5*R*) and (2*S*,3*R*,4*R*,5*S*)-tetrahydroxyazepanes via the Henry reaction. *Carbohydrate Research* **2006**, *341*, 912-917.
132. Dhavale, D. D.; Markad, S. D.; Karanjule, N. S.; PrakashaReddy, J. Asymmetric dihydroxylation of D-glucose derived  $\alpha,\beta$ -unsaturated ester: Synthesis of azepane and nojirimycin analogues. *The Journal of Organic Chemistry* **2004**, *69*, 4760-4766.
133. Borch, R. F.; Bernstein, M. D.; Durst, H. D. Cyanohydridoborate anion as a selective reducing agent. *Journal of the American Chemical Society* **1971**, *93*, 2897-2904.
134. Charnley, M.; Moir, A. J. G.; Douglas, C. W. I.; Haycock, J. W. Anti-microbial action of melanocortin peptides and identification of a novel X-Pro-d/l-Val sequence in Gram-positive and Gram-negative bacteria. *Peptides* **2008**, *29*, 1004-1009.
135. Hiltz, M. E.; Catania, A.; Lipton, J. M. Anti-inflammatory activity of  $\alpha$ -MSH(11–13) analogs: Influences of alteration in stereochemistry. *Peptides* **1991**, *12*, 767-771.
136. Singh, M.; Mukhopadhyay, K. C-terminal amino acids of alpha-melanocyte-stimulating hormone are requisite for its antibacterial activity against *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* **2011**, *55*, 1920-1929.
137. Ferreira, S. H.; Lorenzetti, B. B.; Bristow, A. F.; Poole, S. Interleukin-1[ $\beta$ ] as a potent hyperalgesic agent antagonized by a tripeptide analogue. *Nature* **1988**, *334*, 698-700.
138. Moustafa, M.; Helen Kemp, E.; MacNeil, S.; Szabo, M.; Haycock, J. W.; Ghanem, G. E.; Morandini, R. Inhibition of tumor necrosis factor- $\alpha$  stimulated NF $\kappa$ B/p65 in human keratinocytes by  $\alpha$ -melanocyte stimulating hormone and adrenocorticotrophic hormone peptides. *Journal of Investigative Dermatology* **2002**, *119*, 1244-1253.
139. Galimberti, D.; Baron, P.; Meda, L.; Prat, E.; Scarpini, E.; Delgado, R.; Catania, A.; Lipton, J. M.; Scarlato, G.  $\alpha$ -MSH peptides inhibit production of nitric oxide and tumor necrosis factor- $\alpha$  by microglial cells activated with  $\beta$ -amyloid and interferon  $\gamma$ . *Biochemical and Biophysical Research Communications* **1999**, *263*, 251-256.
140. Fosgerau, K.; Hoffmann, T. Peptide therapeutics: current status and future directions. *Drug Discovery Today* **2015**, *20*, 122-128.
141. Sweeney, P. J.; Walker, J. M. Pronase (EC 3.4.24.4). In: *Enzymes of Molecular Biology*. Edited by Burrell, M. M. Totowa, NJ: Humana Press; **1993**, 271-276.

142. Bauer, A. W.; Perry, D. M.; Kirby, W. M. Single-disk antibiotic-sensitivity testing of *Staphylococci*; an analysis of technique and results. *AMA Arch Intern Med* **1959**, *104*, 208-216.
143. Brabetz, W.; Muller-Loennies, S.; Holst, O.; Brade, H. Deletion of the heptosyltransferase genes *rfaC* and *rfaF* in *Escherichia coli* K-12 results in an Re-type lipopolysaccharide with a high degree of 2-aminoethanol phosphate substitution. *European Journal of Biochemistry* **1997**, *247*, 716-724.
144. Fey, P. D.; Endres, J. L.; Yajjala, V. K.; Widhelm, T. J.; Boissy, R. J.; Bose, J. L.; Bayles, K. W. A genetic resource for rapid and comprehensive phenotype screening of nonessential *Staphylococcus aureus* genes. *MBio* **2013**, *4*, e00537-00512.
145. Zhao, P.; Stephens, J. M. STAT1, NF- $\kappa$ B and ERKs play a role in the induction of lipocalin-2 expression in adipocytes. *Molecular Metabolism* **2013**, *2*, 161-170.
146. Lau, Q. Y.; Choo, X. Y.; Lim, Z. X.; Kong, X. N.; Ng, F. M.; Ang, M. J. Y.; Hill, J.; Brian Chia, C. S. A head-to-head comparison of the antimicrobial activities of 30 ultra-short antimicrobial peptides against *Staphylococcus aureus*, *Pseudomonas aeruginosa* and *Candida albicans*. *International Journal of Peptide Research and Therapeutics* **2015**, *21*, 21-28.

**APPENDIX A  
SUPPLEMENTARY ON PROTEIN EXPRESSION AND NMR ANALYSIS**

**A-1 Protein sequences**

**A-1.1 Scc1His**

MSGHHHHHSSSSGTENLYFQGAMQNQFEQLLTELGTQINSPLTPDSNNACIVRFGYNNVAVQIEEDGN  
SGFLVAGVMLGKLPENTFRQKIFKAALSINGSPQSNIKGTLGYGEISNQLYLCDRNLNMTYLNKEKLARYLV  
LFSQHANIWMQSISKGELPDLHALGMYHL

Theoretical pI/Mw: 6.28 / 18783.19; 169 amino acids

**A-1.2 HisScc4**

MASHHHHHHMLEKLIKNFVAYMGVASELEFDADGSYVFPISLVRMRVRQNADEEIIISAFLEIPASMDIE  
KAYARMMEGNLFQGETGGAALGLDSDGHAVLVRVPGEVSQEDFASYIESVLNYAEAWLEDLGLSKTE  
QE

Theoretical pI/Mw: 4.69 / 15786.75; 142 amino acids

**A-2. Minimal media recipe (1 L)**

100 mL of 10X M9 salts (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 20 mM KH<sub>2</sub>PO<sub>4</sub>, 10 mM NaCl) without NH<sub>4</sub>Cl, 1 mL of 1M MgSO<sub>4</sub>, 100µL of 1M CaCl<sub>2</sub>, 1mL of trace metals (5 µM FeCl<sub>3</sub>, 20 µM CaCl<sub>2</sub>, 10 µM MnSO<sub>4</sub>, ZnSO<sub>4</sub>, 2 µM CoCl<sub>2</sub>, CuSO<sub>4</sub>, NiCl<sub>2</sub>, H<sub>3</sub>BO<sub>3</sub>, Na<sub>2</sub>MoO<sub>4</sub>, Na<sub>2</sub>SeO<sub>3</sub>), 10 mL BME vitamins, 1 mg thiamine, 0.4% glucose, 1 g of <sup>15</sup>N ammonium chloride, and the respective antibiotics.

### A-3 Acquisition parameters for 3D NMR experiments

**Table 9** Parameters used for triple resonance backbone NMR experiments

	d1	# scans	# complex points			# spectral window (ppm)		
			H	N	C	H	N	C
HNCO	1.5	8	1024	32	36	15.88	33.23	14.73
HNCA	1.5	32	1024	32	52	15.88	33.23	29.08
HN(CA)CB	1.5	64	1024	32	64	15.88	33.23	63.14
HN(CO)CA	1.5	32	1024	36	55	16	36.56	30.14
HN(COCA)CB	1.5	64	1024	32	64	16	36.56	66.3
		Offset (ppm)						
	H	N	C	NUS rate (%)	d Strength (MHz)			
HNCO	4.7	116	172.9	47.5	900			
HNCA	4.7	116	52	45	900			
HN(CA)CB	4.7	116	41	48	900			
HN(CO)CA	4.7	117.5	51	48	600			
HN(COCA)CB	4.7	117.5	39	43	600			

### A-4 DNA sequences

#### A-4.1 His-Scc4-22-end DNA sequence

TAT ACC ATG GGCAGCAGCCATCACCACCATCACCATTCTAGT GAT GCA GAC GGC TCT TAC GTT  
TTC CCG ATT TCC AGC CTG GTT CGC ATG CGT GTC CGC CAG AAC GCC GAT GAA GAG ATC ATC  
ATC TCC GCG TTC CTG GGT GAA ATC CCG GCG AGC ATG GAC ATC GAA AAA GCG TAT GCG CGC  
ATG ATG GAA GGC AAC CTG TTC GGC CAG GAA ACT GGC GGT GCT GCT CTG GGT CTG GAC  
TCC GAC GGT CAC GCT GTA CTG GTC CGT CGT GTT CCG GGC GAA GTG TCT CAG GAG GAT TTC  
GCC TCC TAT ATC GAA TCT GTG CTG AAT TAT GCG GAA GCC TGG CTG GAG GAT CTG GGT CTG  
TCT AAG ACC GAG CAG GAG TAA CTC GAG CAC CAC

#### A-4.2 pET28(+)-His-Scc4-22-end DNA sequences

1 ATCCGGATAT AGTTCCTCCT TTCAGCAAAA AACCCCTCAA GACCCGTTTA  
51 GAGGCCCAA GGGGTTATGC TAGTTATTGC TCAGCGGTGG CAGCAGCCAA  
101 CTCAGCTTCC TTTCGGGCTT TGTTAGCAGC CGGATCTCAG TGGTGGTGGT  
151 GGTGGTGCTC GAGTTACTCC TGCTCGGTCT TAGACAGACC CAGATCCTCC  
201 AGCCAGGCTT CCGCATAATT CAGCACAGAT TCGATATAGG AGGCGAAATC  
251 CTCCTGAGAC ACTTCGCCCG GAACACGACG GACCAGTACA GCGTGACCGT



301 CGGAGTCCAG ACCCAGAGCA GCACCGCCAG TTTCTGGCC GAACAGGTTG  
351 CCTTCCATCA TGC GCGCATA CGCTTTTTTCG ATGTCCATGC TCGCCGGGAT  
401 TTCACCCAGG AACGCGGAGA TGATGATCTC TTCATCGGCG TTCTGGCGGA  
451 CACGCATGCG AACCAGGCTG GAAATCGGGA AAACGTAAGA GCCGTCTGCA  
501 TCACTAGAAT GGTGATGGTG GTGATGGCTG CTGCCCATGG TATATCTCCT  
551 TCTTAAAGTT AAACAAAATT ATTTCTAGAG GGAATTGTT ATCCGCTCAC  
601 AATTCCCCTA TAGTGAGTCG TATTAATTC GCGGGATCGA GATCTCGATC  
651 CTCTACGCCG GACGCATCGT GGCCGGCATC ACCGGCGCCA CAGGTGCGGT  
701 TGCTGGCGCC TATATCGCCG ACATCACCGA TGGGGAAGAT CGGGCTCGCC  
751 ACTTCGGGCT CATGAGCGCT TGTTTCGGCG TGGGTATGGT GGCAGGCCCC  
801 GTGGCCGGGG GACTGTTGGG CGCCATCTCC TTGCATGCAC CATTCTTGC  
851 GCGGCGGGT CTCAACGGCC TCAACCTACT ACTGGGCTGC TTCCTAATGC  
901 AGGAGTCGCA TAAGGGAGAG CGTCGAGATC CCGGACACCA TCGAATGGCG  
951 CAAAACCTTT CGCGGTATGG CATGATAGCG CCCGGAAGAG AGTCAATTCA  
1001 GGGTGGTAA TGTGAAACCA GTAACGTTAT ACGATGTCGC AGAGTATGCC  
1051 GGTGTCTCTT ATCAGACCGT TTCCGCGTG GTGAACCAGG CCAGCCACGT  
1101 TTCTGCGAAA ACGCGGAAA AAGTGAAGC GCGGATGGCG GAGCTGAATT  
1151 ACATTCCCAA CCGCGTGGCA CAACAACCTGG CGGGCAAACA GTCGTTGCTG  
1201 ATTGGCGTTG CCACCTCCAG TCTGGCCCTG CACGCGCCGT CGCAAATTGT  
1251 CGCGGCGATT AAATCTCGCG CCGATCAACT GGGTGCCAGC GTGGTGGTGT  
1301 CGATGGTAGA ACGAAGCGGC GTCGAAGCCT GTAAAGCGGC GGTGCACAAT  
1351 CTTCTCGCGC AACGCGTCAG TGGGCTGATC ATTAACCTATC CGCTGGATGA  
1401 CCAGGATGCC ATTGCTGTGG AAGCTGCCTG CACTAATGTT CCGGCGTTAT  
1451 TTCTTGATGT CTCTGACCAG ACACCCATCA ACAGTATTAT TTTCTCCCAT  
1501 GAAGACGGTA CGCGACTGGG CGTGGAGCAT CTGGTCGCAT TGGGTCACCA  
1551 GCAAATCGCG CTGTTAGCGG GCCCATTAAG TTCTGTCTCG GCGCGTCTGC  
1601 GTCTGGCTGG CTGGCATAAA TATCTCACTC GCAATCAAAT TCAGCCGATA  
1651 GCGGAACGGG AAGGCGACTG GAGTGCCATG TCCGGTTTTT ACAAACCAT

1701 GCAAATGCTG AATGAGGGCA TCGTTCCCAC TCGGATGCTG GTTGCCAACG  
1751 ATCAGATGGC GCTGGGCGCA ATGCGCGCCA TTACCGAGTC CGGGCTGCGC  
1801 GTTGGTGC GG ATATCTCGGT AGTGGGATAC GACGATACCG AAGACAGCTC  
1851 ATGTTATATC CCGCCGTTAA CCACCATCAA ACAGGATTTT CGCCTGCTGG  
1901 GGCAAACCAG CGTGGACCGC TTGCTGCAAC TCTCTCAGGG CCAGGCGGTG  
1951 AAGGGCAATC AGCTGTTGCC CGTCTCACTG GTGAAAAGAA AAACCACCCT  
2001 GGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA  
2051 TGCAGCTGGC ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA  
2101 CGCAATTAAT GTAAGTTAGC TCACTCATT A GGCACCGGGA TCTCGACCGA  
2151 TGCCCTTGAG AGCCTTCAAC CCAGTCAGCT CCTTCCGGTG GCGCGGGGGC  
2201 ATGACTATCG TCGCCGCACT TATGACTGTC TTCTTTATCA TGCAACTCGT  
2251 AGGACAGGTG CCGGCAGCGC TCTGGGTCAT TTTCGGCGAG GACCGCTTTC  
2301 GCTGGAGCGC GACGATGATC GGCCTGTCGC TTGCGGTATT CGGAATCTTG  
2351 CACGCCCTCG CTCAAGCCTT CGTCACTGGT CCCGCCACCA AACGTTTCGG  
2401 CGAGAAGCAG GCCATTATCG CCGGCATGGC GGCCCCACGG GTGCGCATGA  
2451 TCGTGCTCCT GTCGTTGAGG ACCCGGCTAG GCTGGCGGGG TTGCCTTACT  
2501 GGTTAGCAGA ATGAATCACC GATACGCGAG CGAACGTGAA GCGACTGCTG  
2551 CTGCAAACG TCTGCGACCT GAGCAACAAC ATGAATGGTC TTCGGTTTCC  
2601 GTGTTTCGTA AAGTCTGGAA ACGCGGAAGT CAGCGCCCTG CACCATTATG  
2651 TTCCGGATCT GCATCGCAGG ATGCTGCTGG CTACCCTGTG GAACACCTAC  
2701 ATCTGTATTA ACGAAGCGCT GGCATTGACC CTGAGTGATT TTTCTCTGGT  
2751 CCCGCCGCAT CCATACCGCC AGTTGTTTAC CCTCACAACG TTCCAGTAAC  
2801 CGGGCATGTT CATCATCAGT AACCCGTATC GTGAGCATCC TCTCTCGTTT  
2851 CATCGGTATC ATTACCCCCA TGAACAGAAA TCCCCCTTAC ACGGAGGCAT  
2901 CAGTGACCAA ACAGGAAAAA ACCGCCCTTA ACATGGCCCG CTTTATCAGA  
2951 AGCCAGACAT TAACGCTTCT GGAGAAACTC AACGAGCTGG ACGCGGATGA  
3001 ACAGGCAGAC ATCTGTGAAT CGCTTACGA CCACGCTGAT GAGCTTTACC  
3051 GCAGCTGCCT CGCGCGTTTC GGTGATGACG GTGAAAACCT CTGACACATG

3101 CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG CCGGGAGCAG  
3151 ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CGGGGCGCAG  
3201 CCATGACCCA GTCACGTAGC GATAGCGGAG TGTATACTGG CTTAACTATG  
3251 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATATG CGGTGTGAAA  
3301 TACCGCACAG ATGCGTAAGG AGAAAATACC GCATCAGGCG CTCTTCCGCT  
3351 TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT  
3401 ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA  
3451 ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT  
3501 AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC CCCCTGACGA  
3551 GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC  
3601 TATAAAGATA CCAGGCGTTT CCCCTGGAA GCTCCCTCGT GCGCTCTCCT  
3651 GTTCCGACCC TGCCGTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG  
3701 AAGCGTGGCG CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT  
3751 AGGTCGTTTCG CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC  
3801 GACCGCTGCG CTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG  
3851 ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG  
3901 CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC  
3951 GGCTACACTA GAAGGACAGT ATTTGGTATC TGCCTCTGC TGAAGCCAGT  
4001 TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG  
4051 CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC AGCAGATTAC GCGCAGAAAA  
4101 AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA  
4151 GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCATGAAC AATAAACTG  
4201 TCTGCTTACA TAAACAGTAA TACAAGGGGT GTTATGAGCC ATATTCAACG  
4251 GGAAACGTCT TGCTCTAGGC CGCGATTAAT TTCCAACATG GATGCTGATT  
4301 TATATGGGTA TAAATGGGCT CGCGATAATG TCGGGCAATC AGGTGCGACA  
4351 ATCTATCGAT TGTATGGGAA GCCCGATGCG CCAGAGTTGT TTCTGAAACA  
4401 TGGCAAAGGT AGCGTTGCCA ATGATGTTAC AGATGAGATG GTCAGACTAA  
4451 ACTGGCTGAC GGAATTTATG CCTCTTCCGA CCATCAAGCA TTTTATCCGT

4501 ACTCCTGATG ATGCATGGTT ACTCACCCT GCGATCCCCG GGAAAACAGC  
4551 ATTCCAGGTA TTAGAAGAAT ATCCTGATTC AGGTGAAAAT ATTGTTGATG  
4601 CGCTGGCAGT GTTCCTGCGC CGGTTGCATT CGATTCCTGT TTGTAATTGT  
4651 CCTTTAACA GCGATCGCGT ATTCGTCTC GCTCAGGCGC AATCACGAAT  
4701 GAATAACGGT TTGGTTGATG CGAGTGATTT TGATGACGAG CGTAATGGCT  
4751 GGCCTGTTGA ACAAGTCTGG AAAGAAATGC ATAACTTTT GCCATTCTCA  
4801 CCGGATTCAG TCGTCACTCA TGGTGATTTT TCACTTGATA ACCTTATTTT  
4851 TGACGAGGGG AAATTAATAG GTTGTATTGA TGTTGGACGA GTCGGAATCG  
4901 CAGACCGATA CCAGGATCTT GCCATCCTAT GGAAGTGCCT CGGTGAGTTT  
4951 TCTCCTTCAT TACAGAAACG GCTTTTTCAA AAATATGGTA TTGATAATCC  
5001 TGATATGAAT AAATTGCAGT TTCATTTGAT GCTCGATGAG TTTTCTAAG  
5051 AATTAATTCA TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAACA  
5101 AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT GAAATTGTAA  
5151 ACGTTAATAT TTTGTTAAAA TTCGCGTTAA ATTTTTGTTA AATCAGCTCA  
5201 TTTTTTAACC AATAGGCCGA AATCGGCAA ATCCCTTATA AATCAAAGA  
5251 ATAGACCGAG ATAGGGTTGA GTGTTGTTCC AGTTTGGAAC AAGAGTCCAC  
5301 TATTAAAGAA CGTGGACTCC AACGTCAAAG GCGAAAAAC CGTCTATCAG  
5351 GCGATGGCC CACTACGTGA ACCATCACCC TAATCAAGTT TTTTGGGGTC  
5401 GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC CCCCATTTA  
5451 GAGCTTGACG GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA  
5501 GCGAAAGGAG CGGGCGCTAG GGCGCTGGCA AGTGTAGCGG TCACGCTGCG  
5551 CGTAACCACC ACACCCGCCG CGCTTAATGC GCCGCTACAG GGCGCGTCCC  
5601 ATTCGCCA

APPENDIX B  
NMR SPECTRA FOR THE SYNTHESIZED COMPOUNDS

B-1  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of compounds in Scheme 3

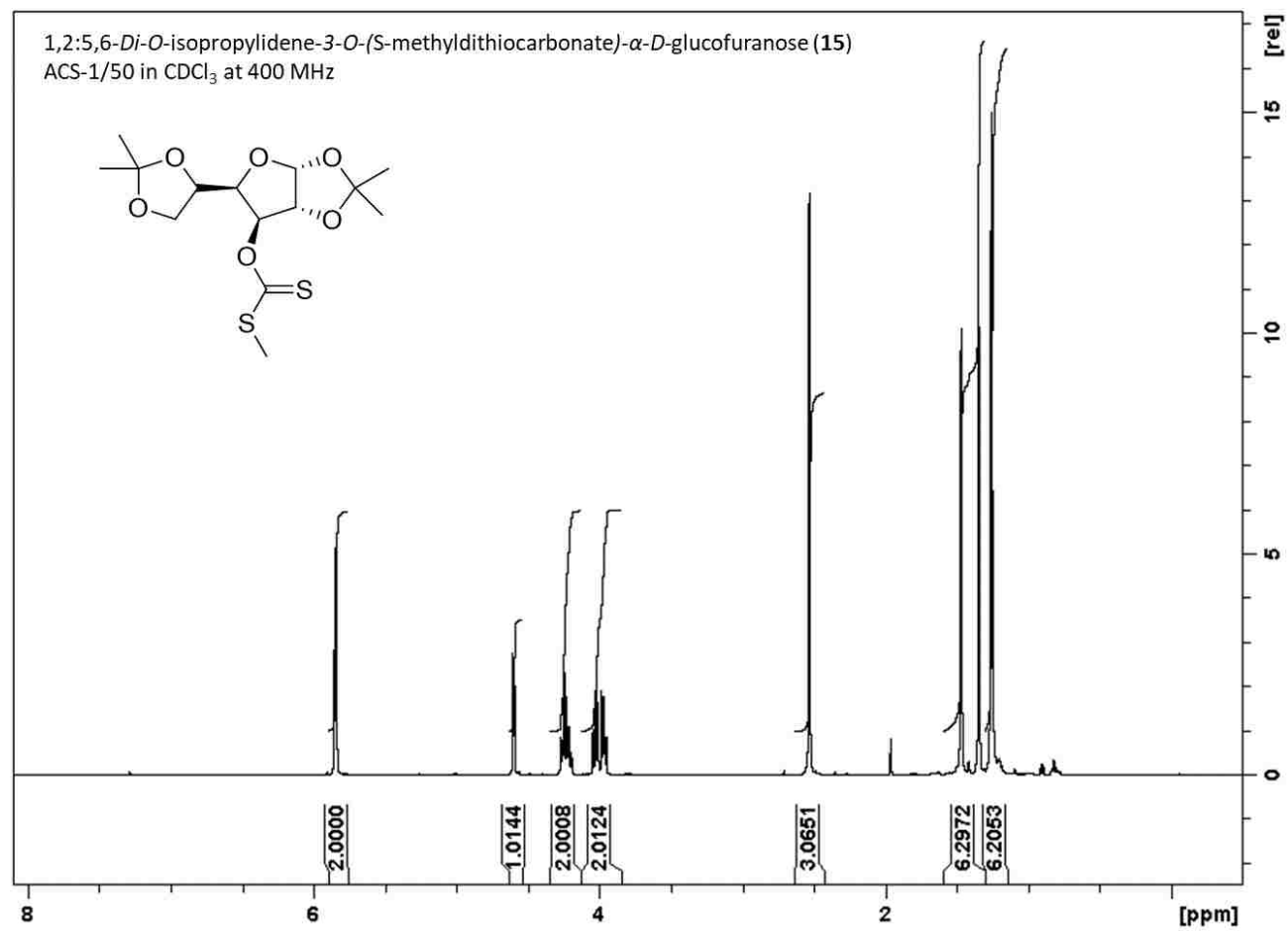


Figure 51.  $^1\text{H}$  NMR spectrum of compound 15

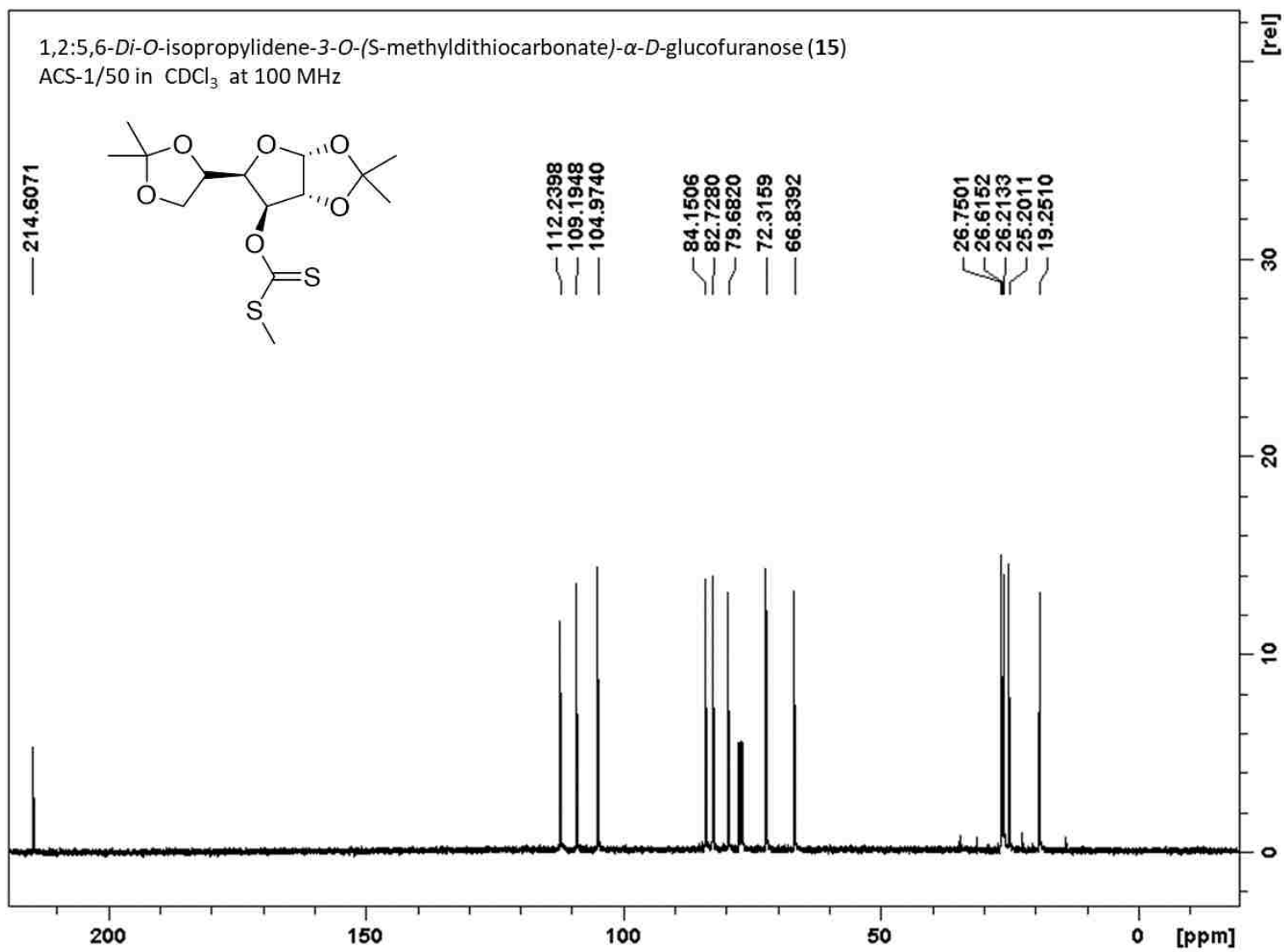


Figure 52. <sup>13</sup>C NMR spectrum of compound 15

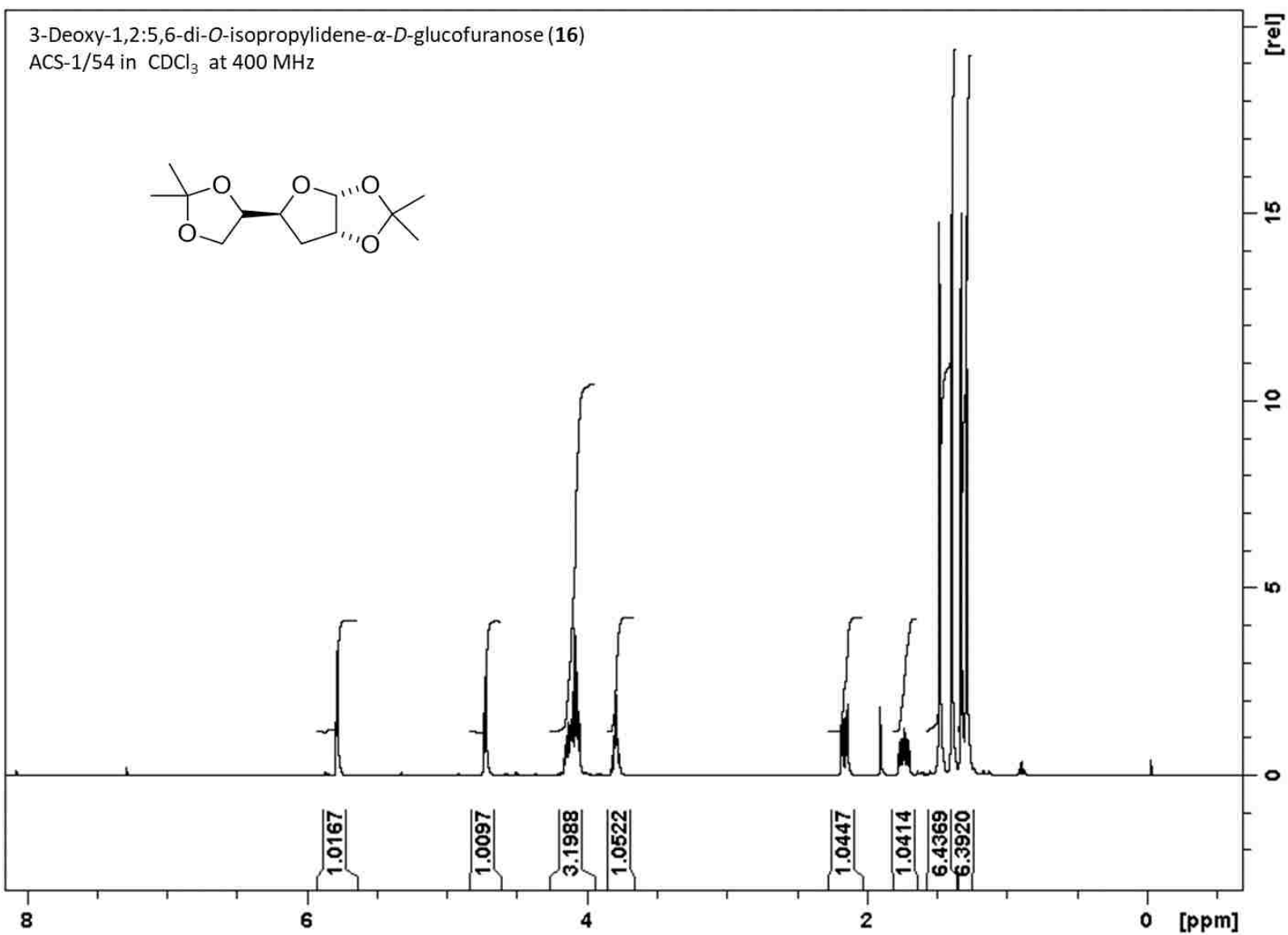


Figure 53. <sup>1</sup>H NMR spectrum of compound **16**

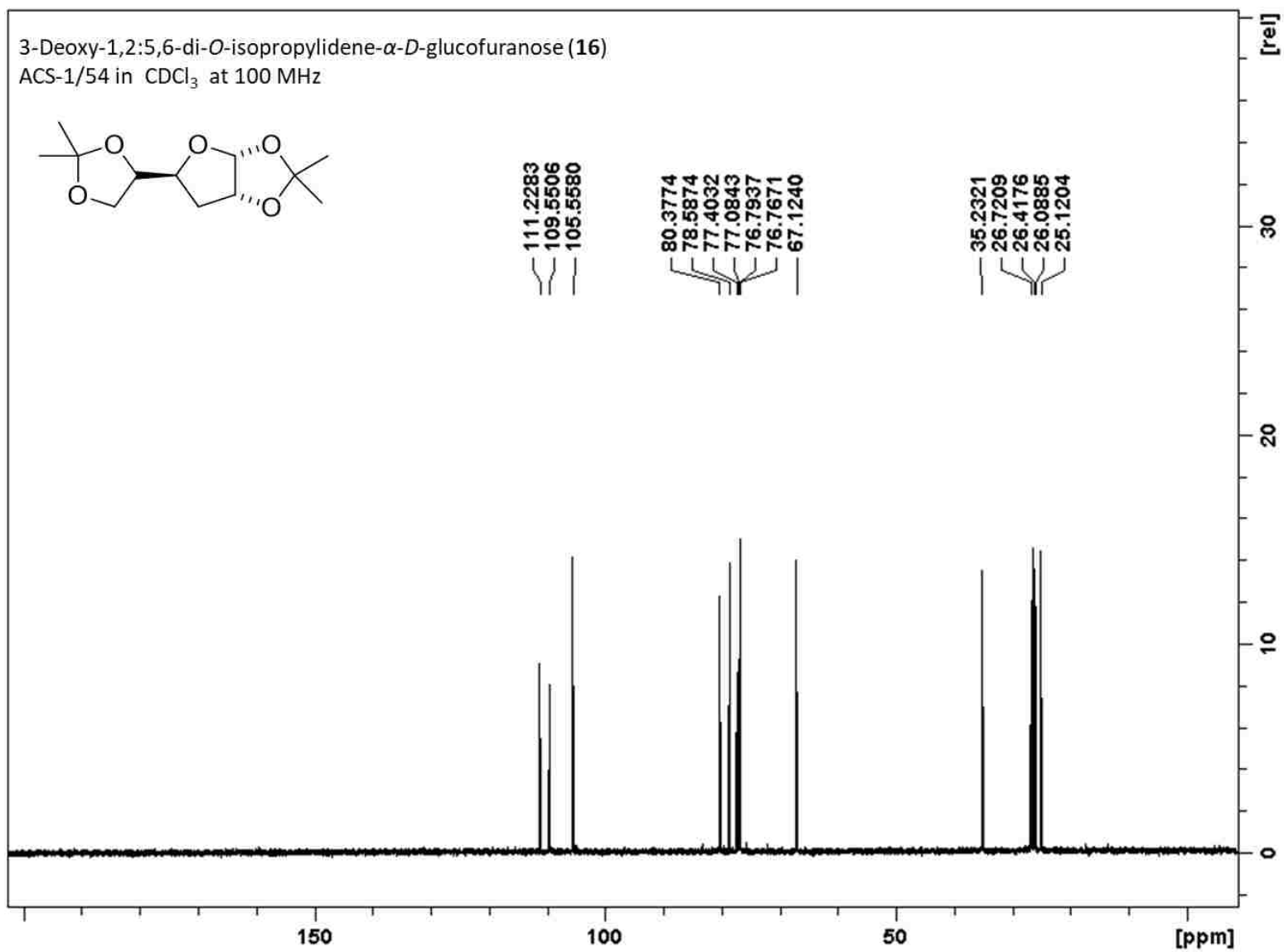


Figure 54. <sup>13</sup>C NMR spectrum of compound **16**



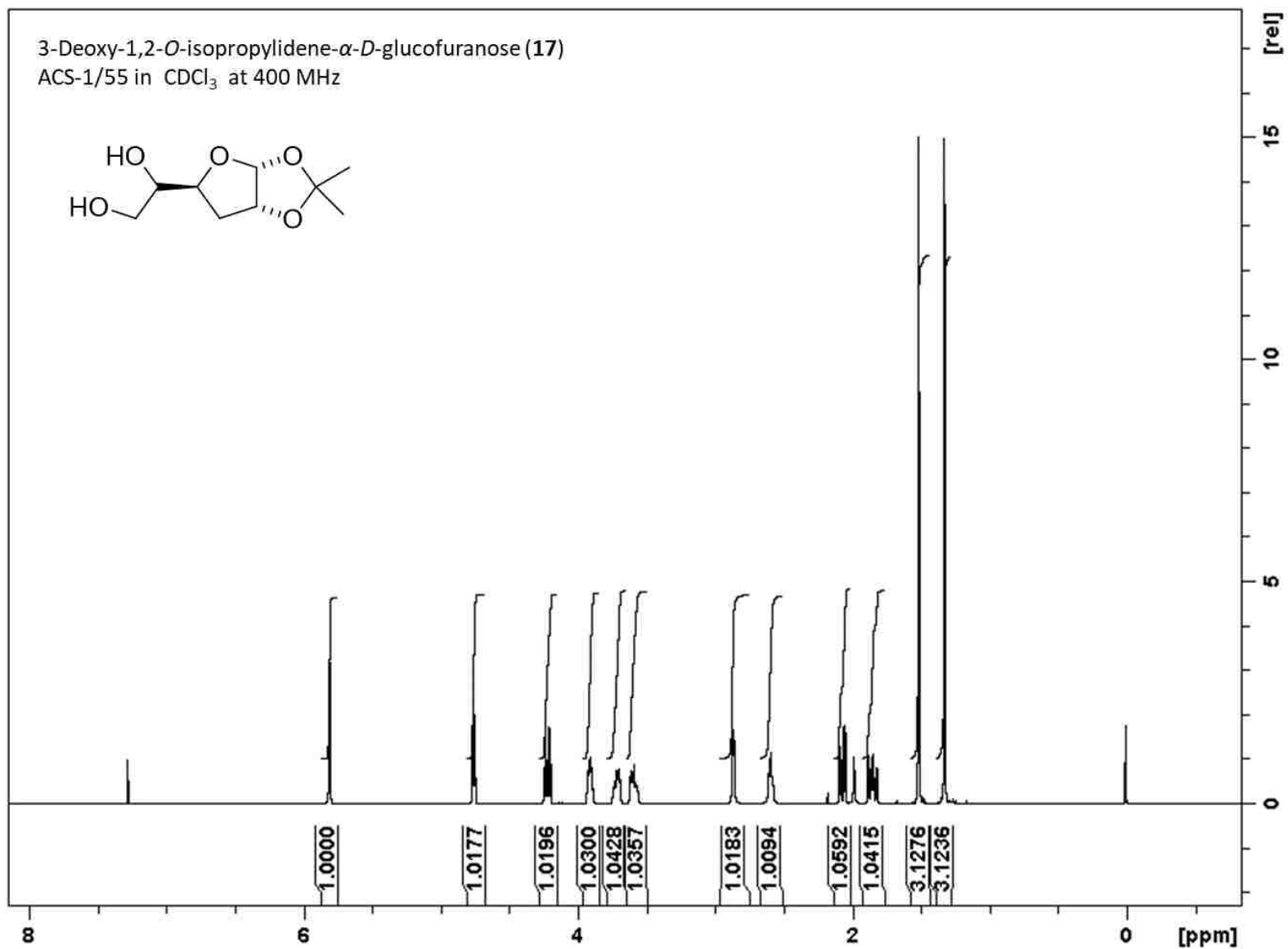


Figure 55. <sup>1</sup>H NMR spectrum of compound 17

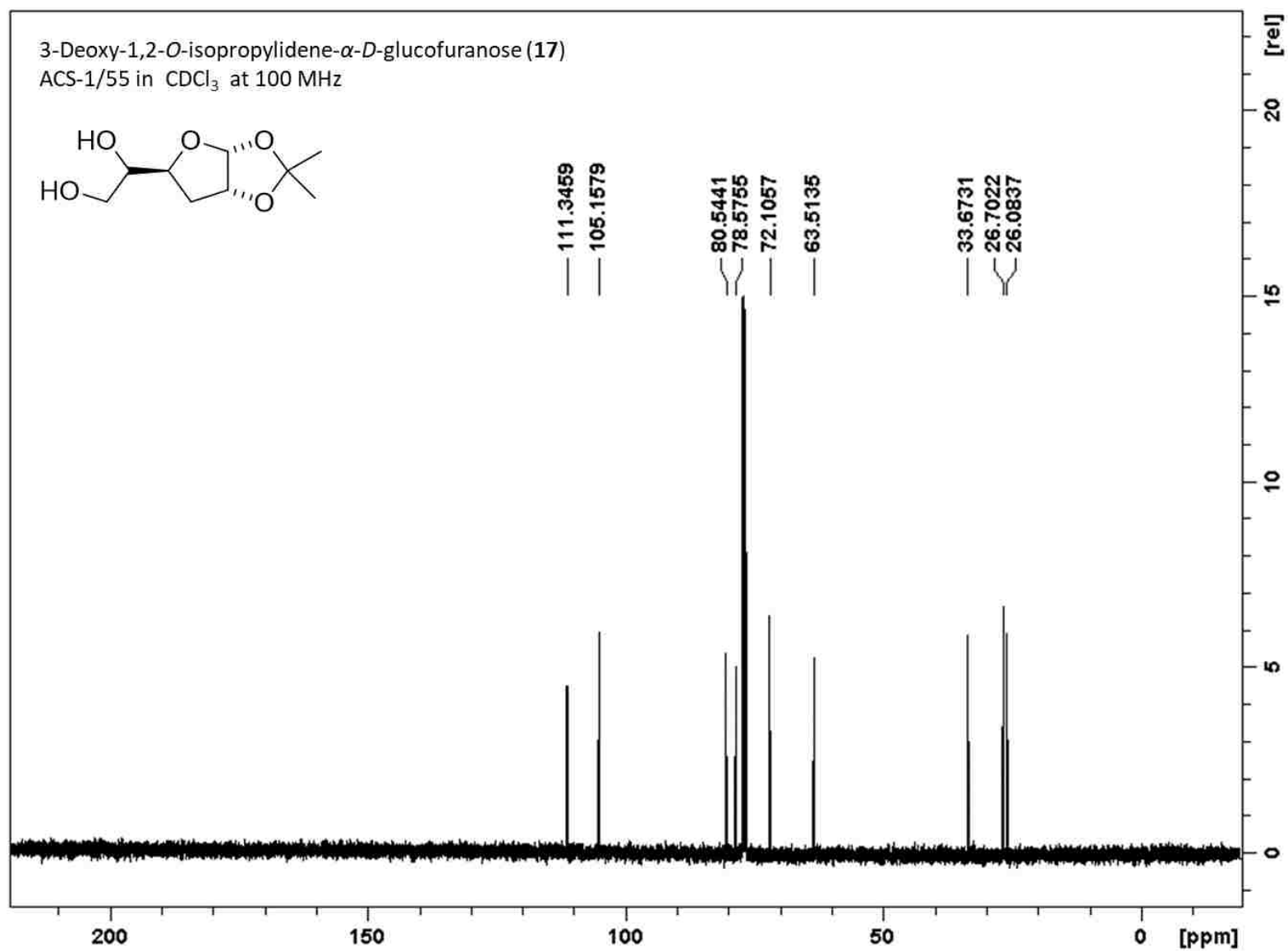


Figure 56. <sup>13</sup>C NMR spectrum of compound 17

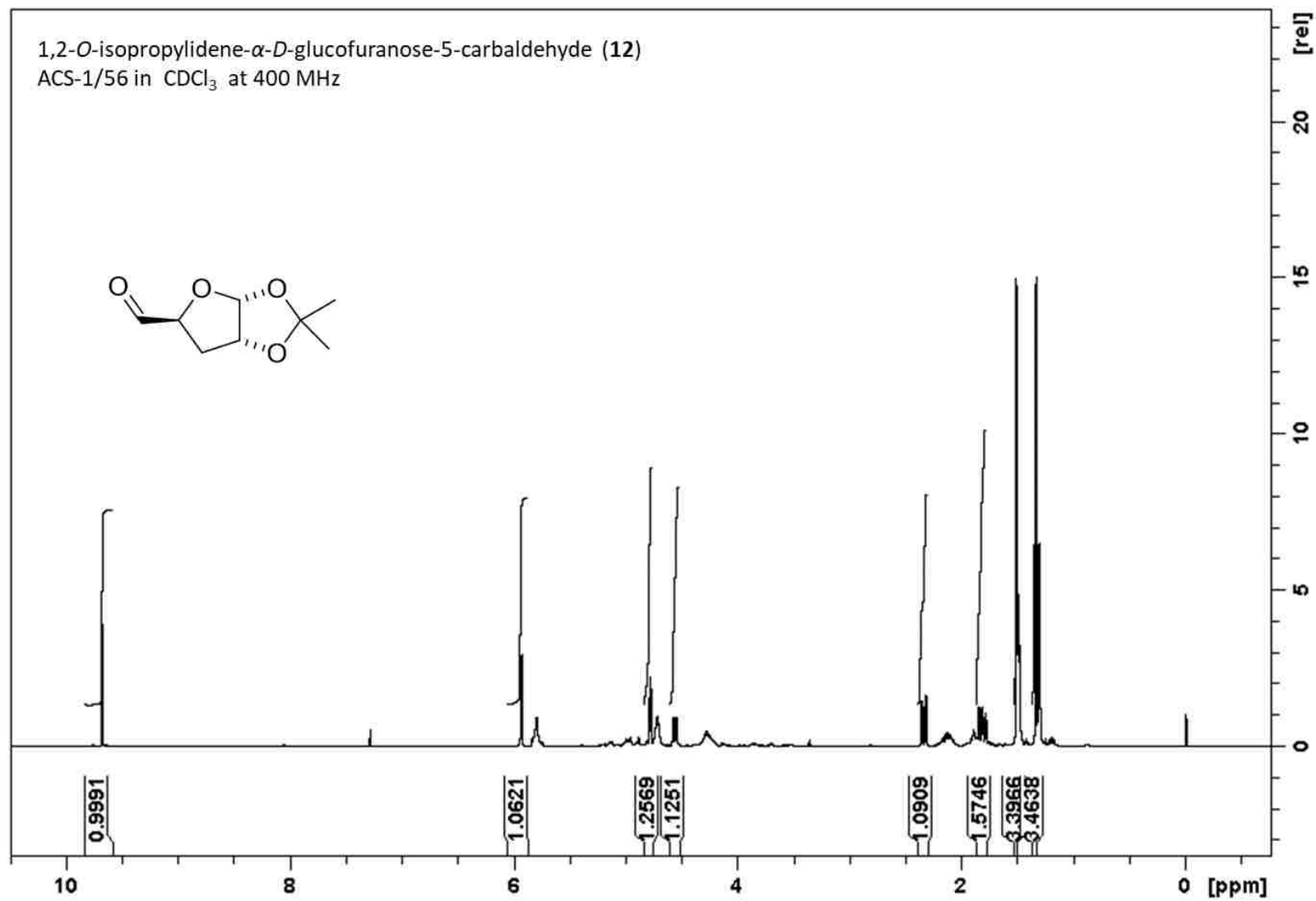


Figure 57. <sup>1</sup>H NMR spectrum of compound **12**

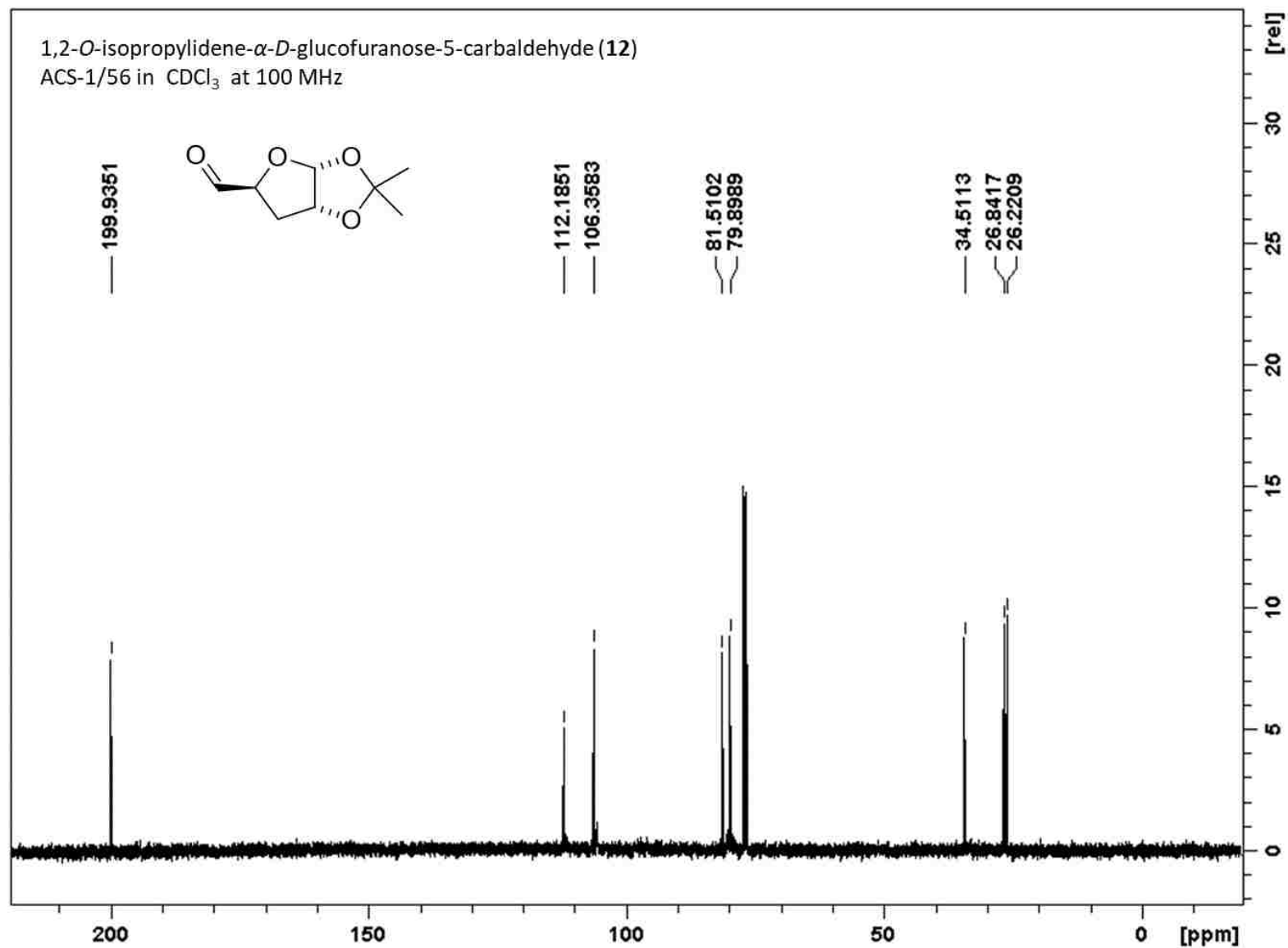


Figure 58. <sup>13</sup>C NMR spectrum of compound 12

B-2  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of compounds in Scheme 4

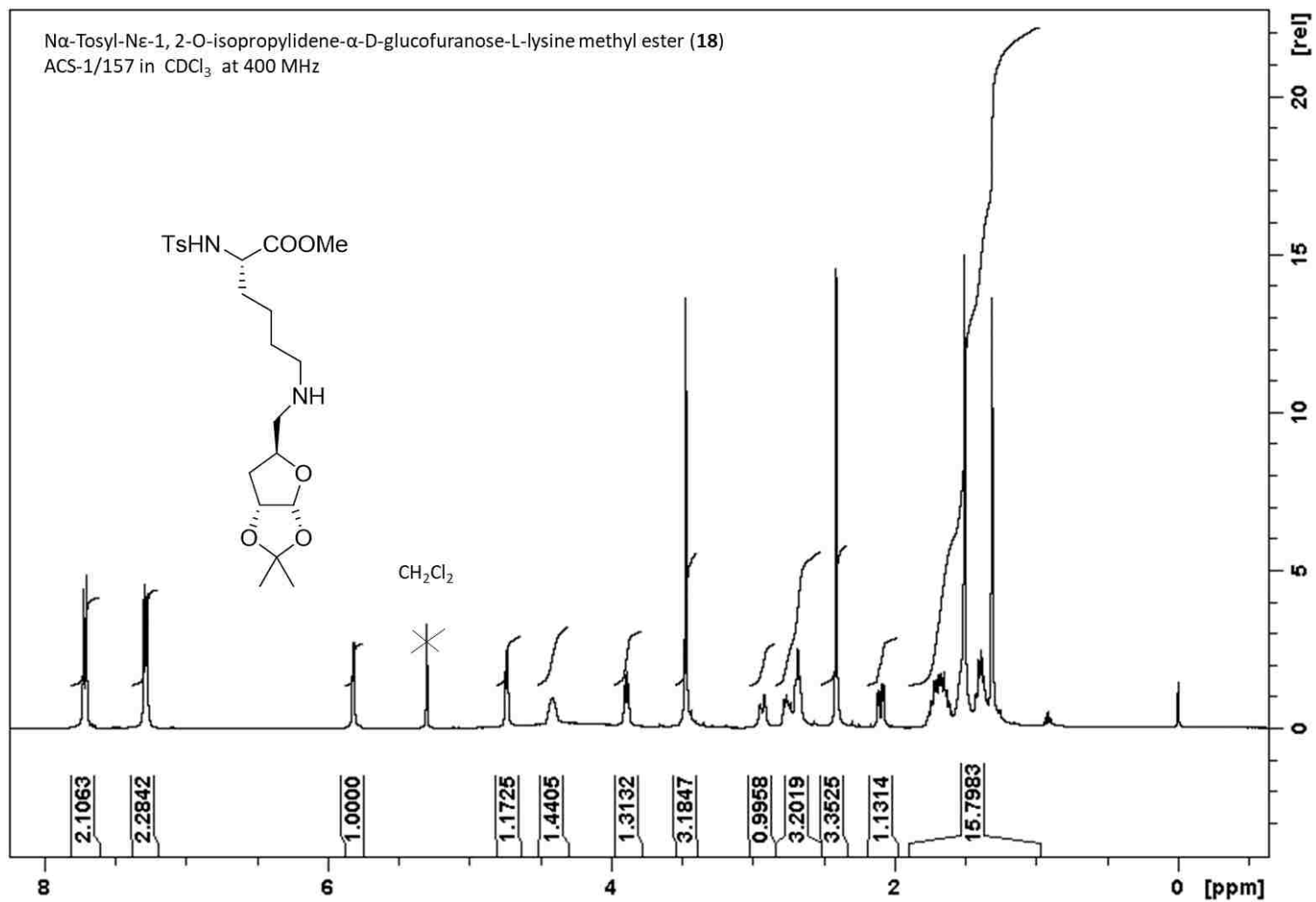


Figure 59.  $^1\text{H}$  NMR spectrum of compound **18**

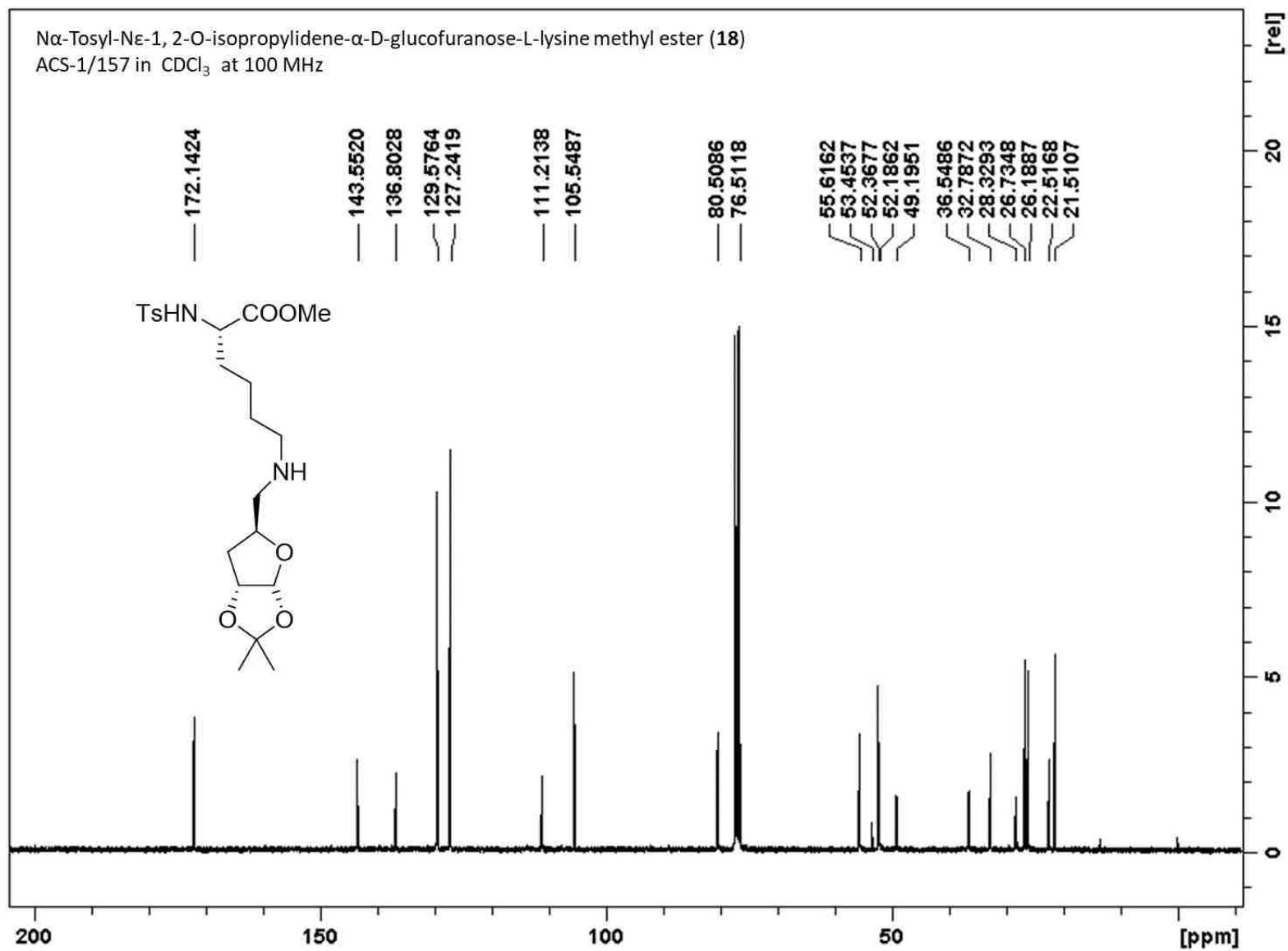


Figure 60. <sup>13</sup>C NMR spectrum of compound **18**

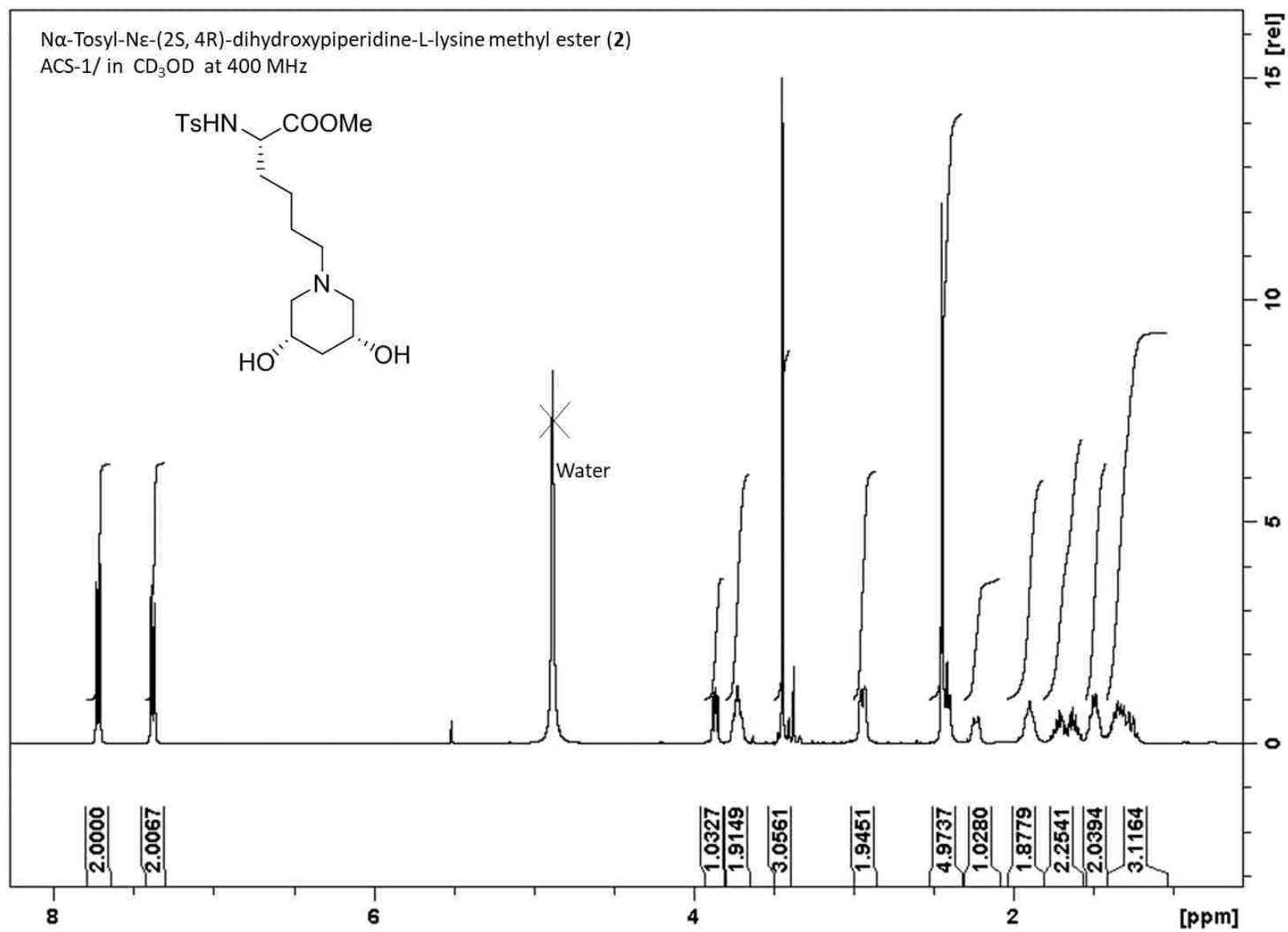


Figure 61. <sup>1</sup>H NMR spectrum of compound 2

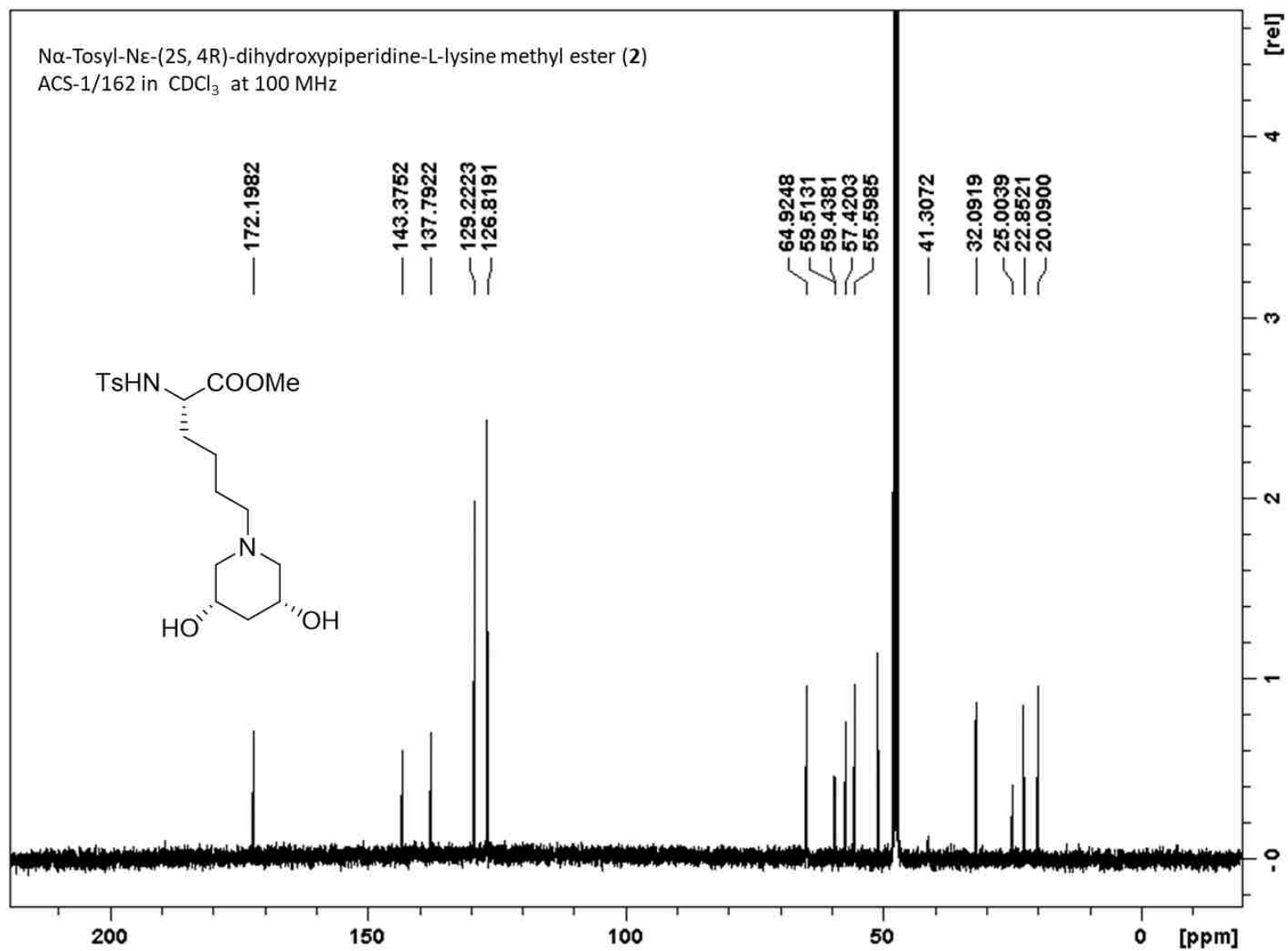


Figure 62. <sup>13</sup>C NMR spectrum of compound 2



B-3  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of compounds in Scheme 5

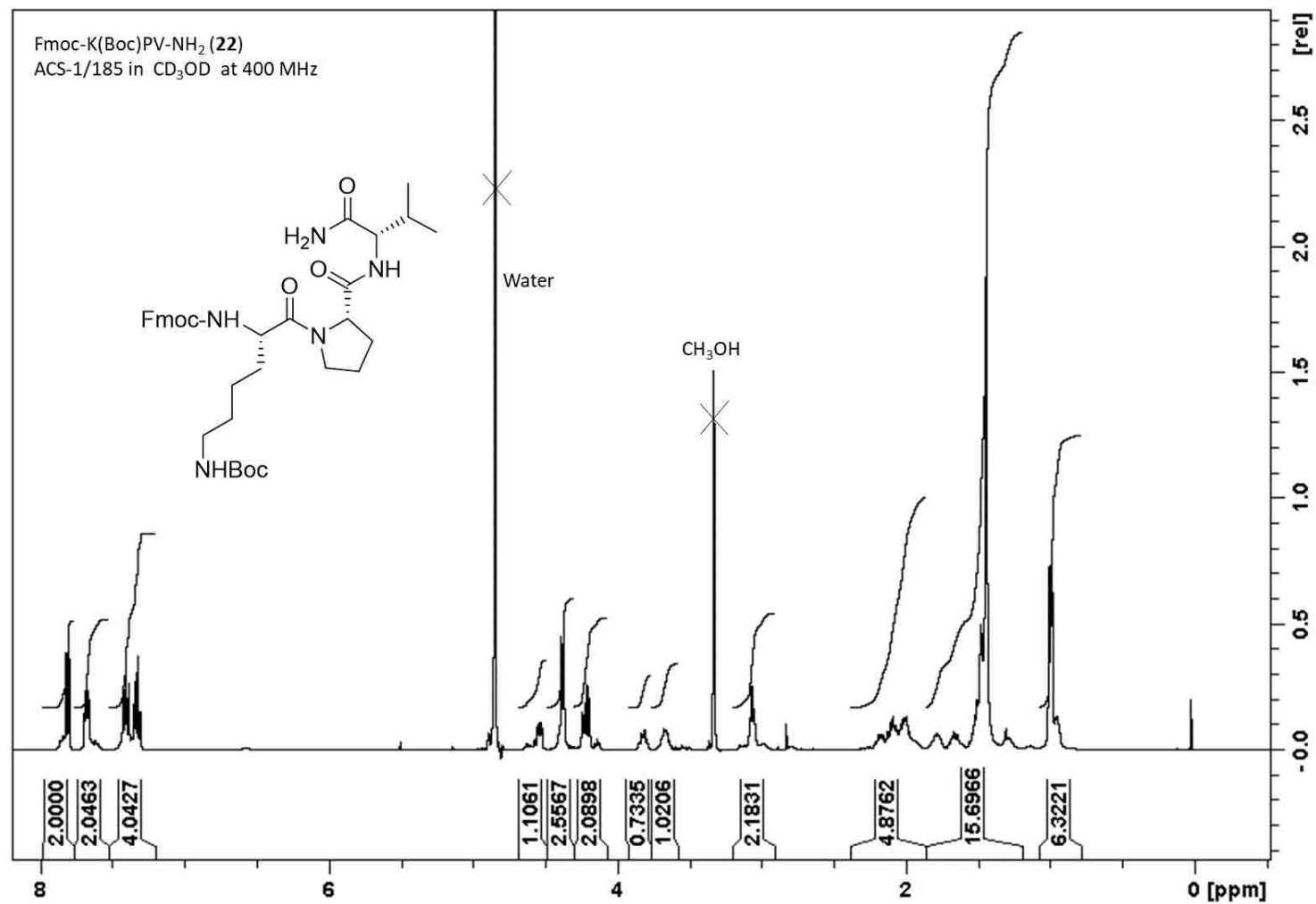


Figure 63.  $^1\text{H}$  NMR spectrum of compound 22

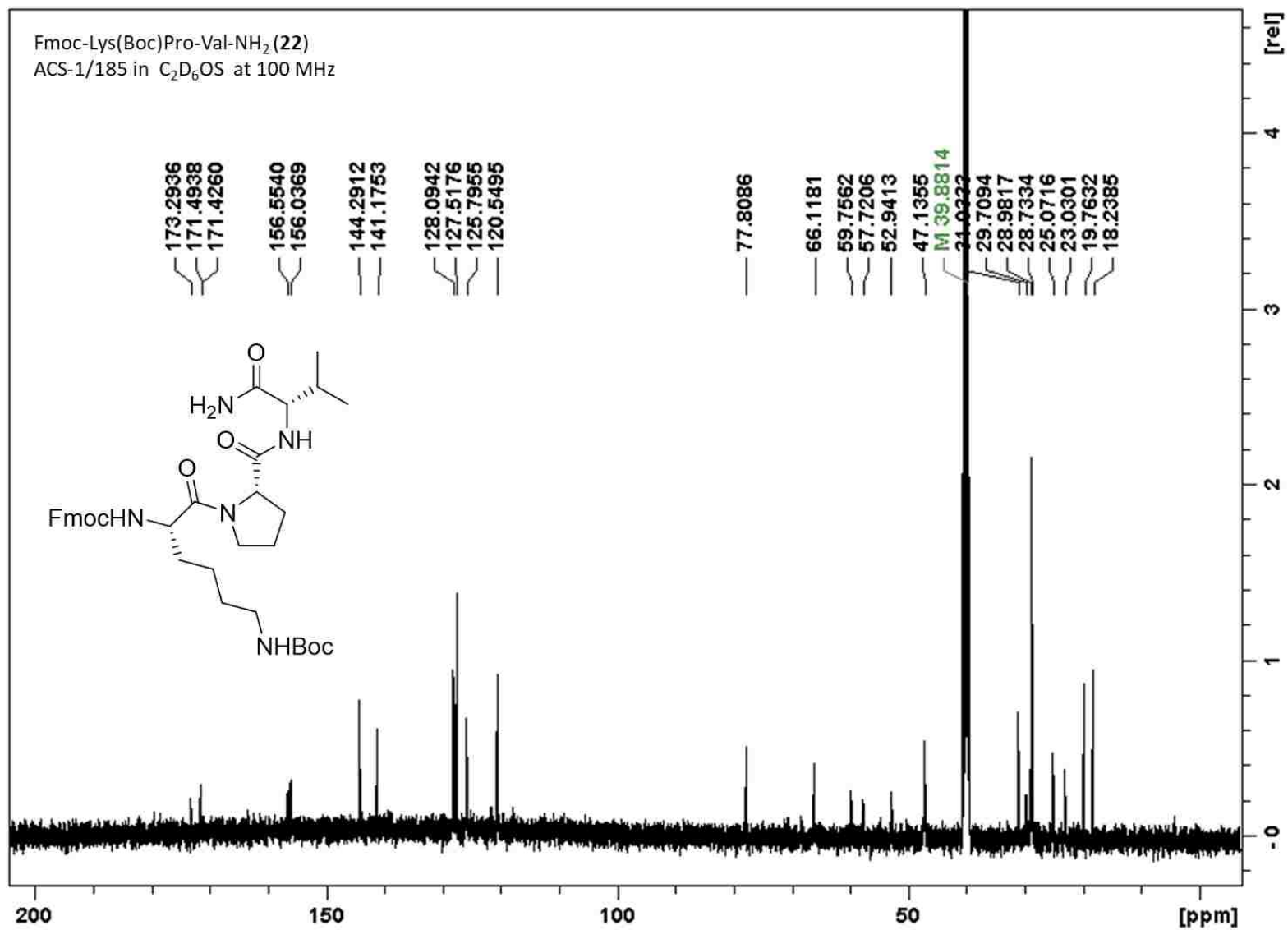


Figure 64. <sup>13</sup>C NMR spectrum of compound 22

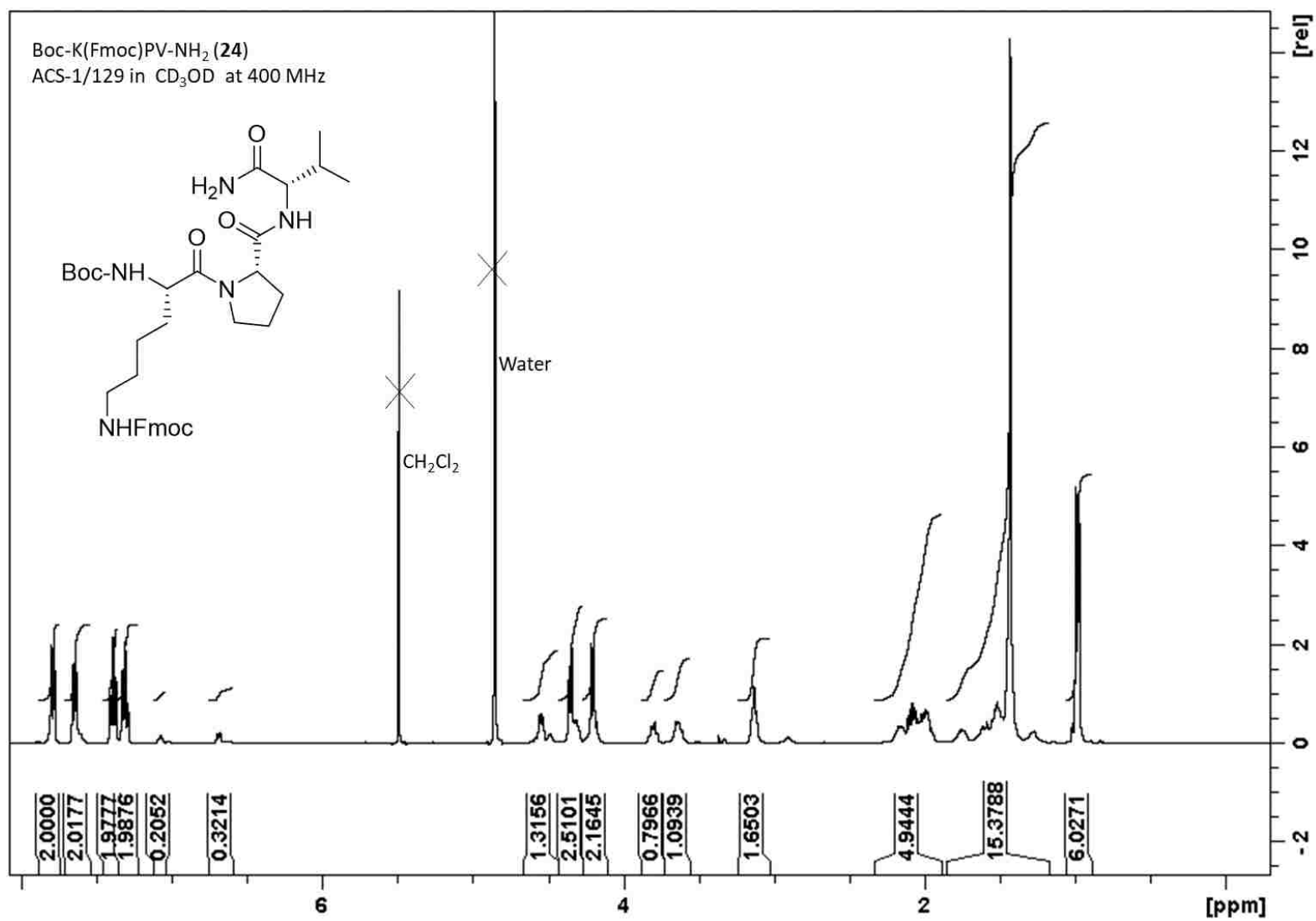


Figure 65. <sup>1</sup>H NMR spectrum of compound **24**

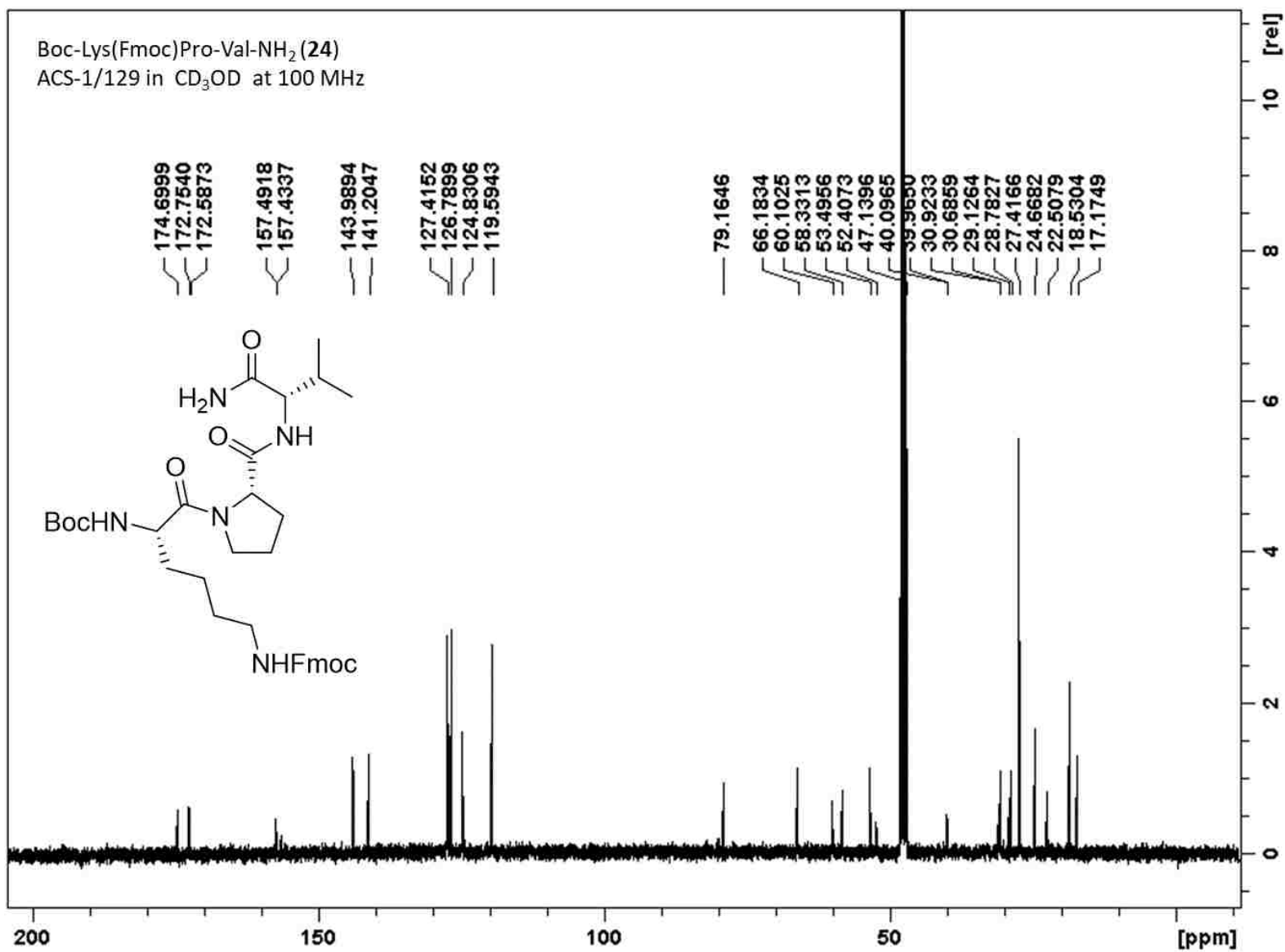


Figure 66. <sup>13</sup>C NMR spectrum of compound 24

B-4  $^1\text{H}$  and  $^{13}\text{C}$  NMR Spectra of compounds in Scheme 6

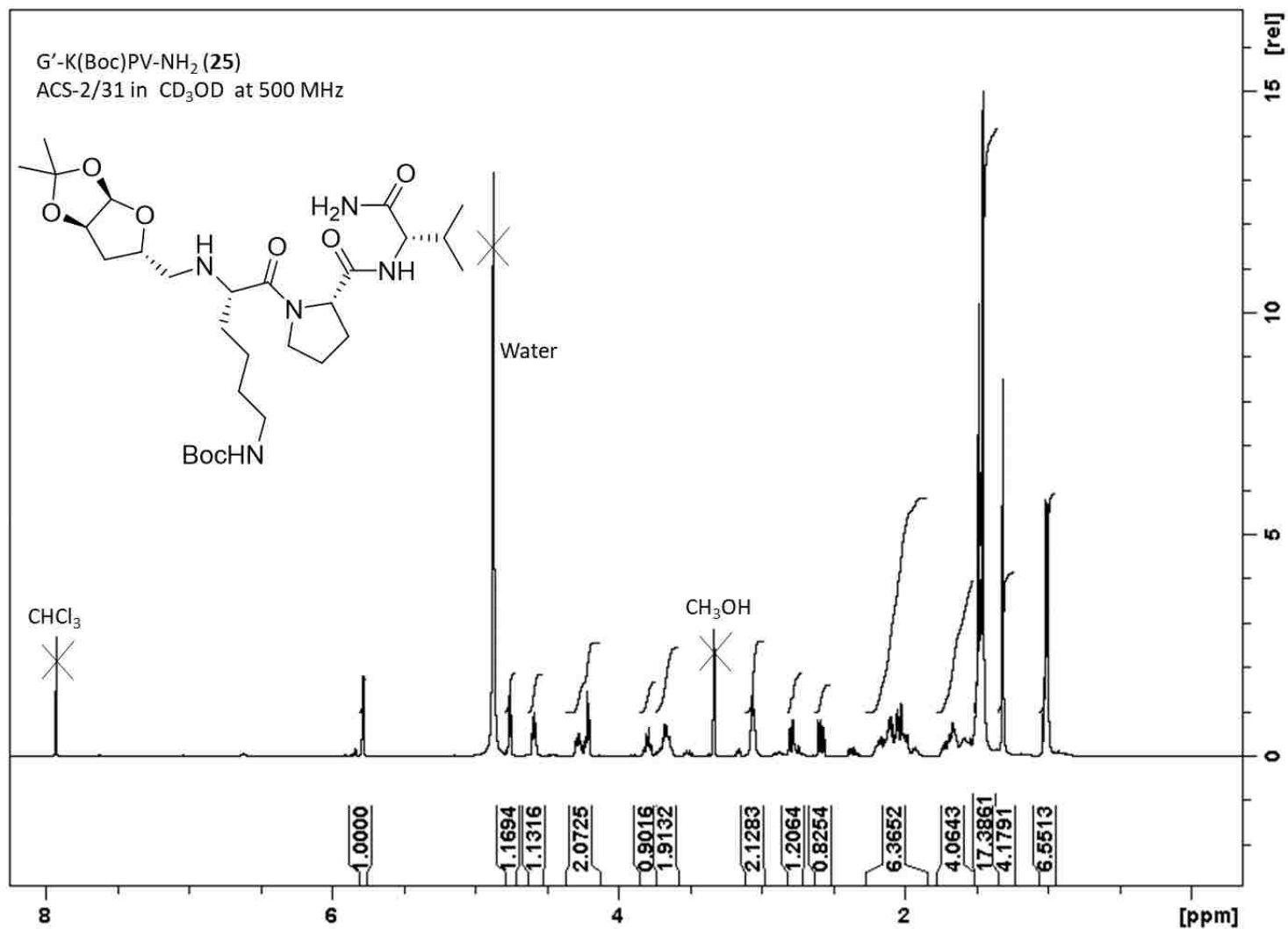


Figure 67.  $^1\text{H}$  NMR spectrum of compound 25

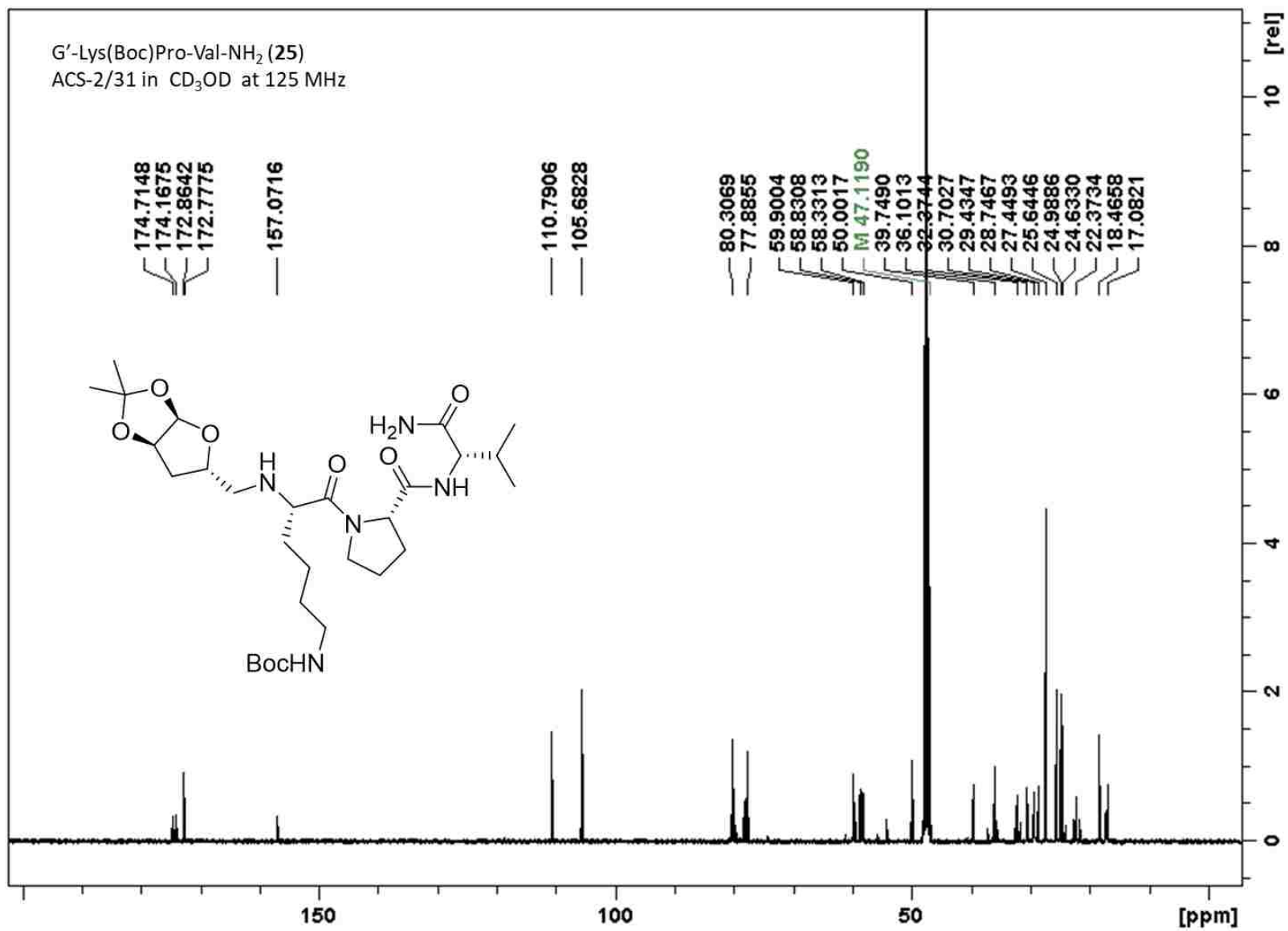


Figure 68. <sup>13</sup>C NMR spectrum of compound 25

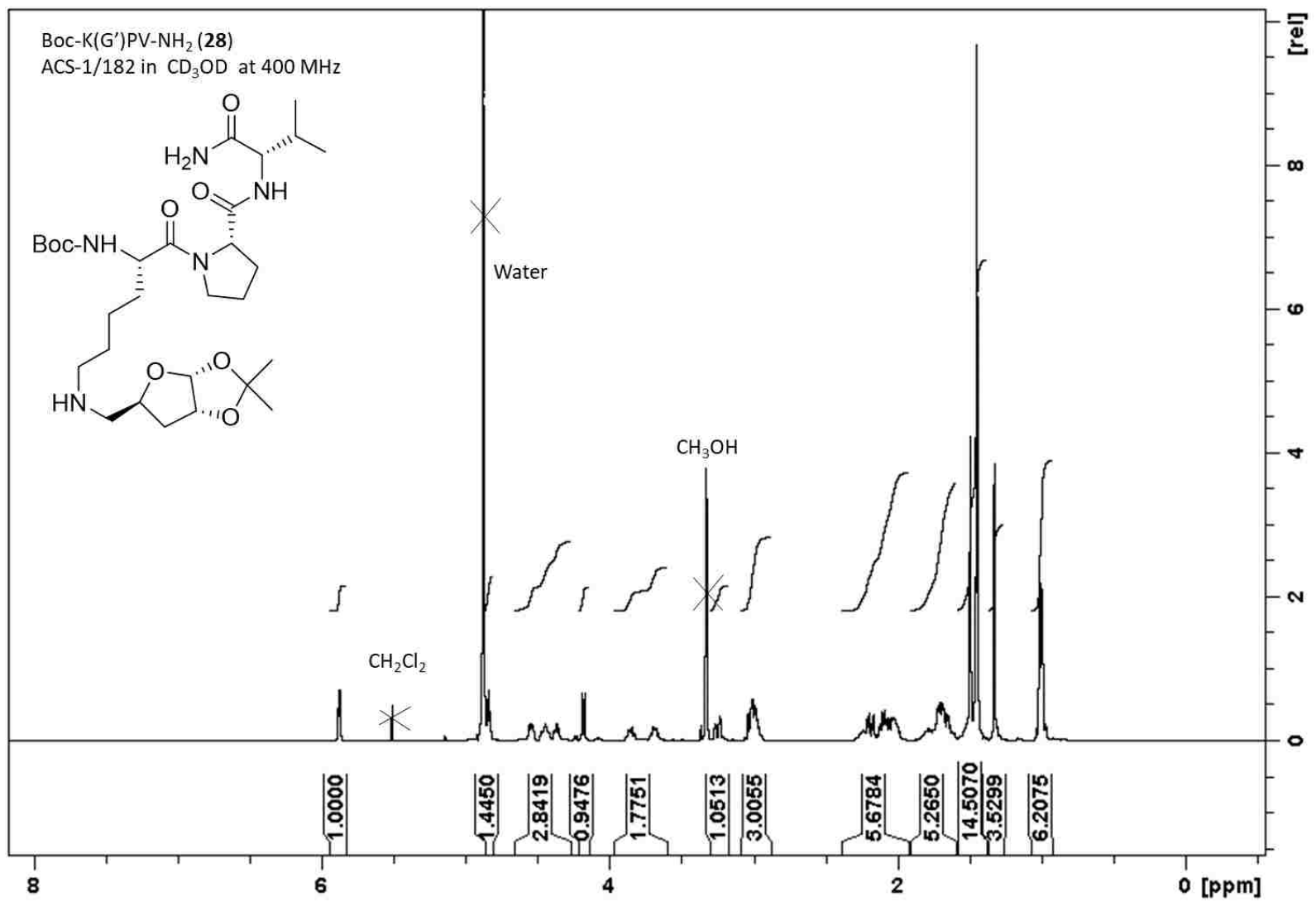


Figure 69. <sup>1</sup>H NMR spectrum of compound **28**

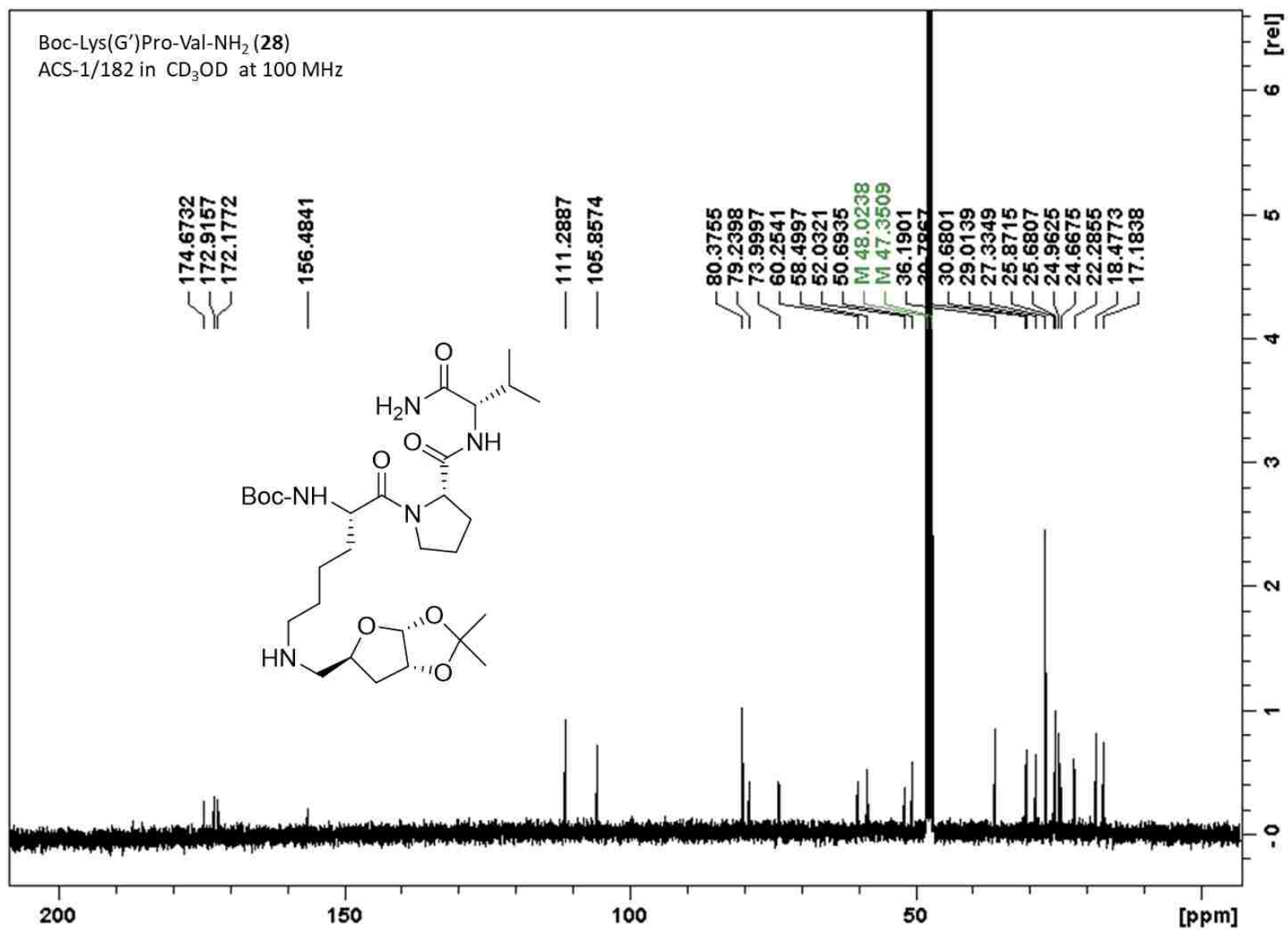


Figure 70. <sup>13</sup>C NMR spectrum of compound 28



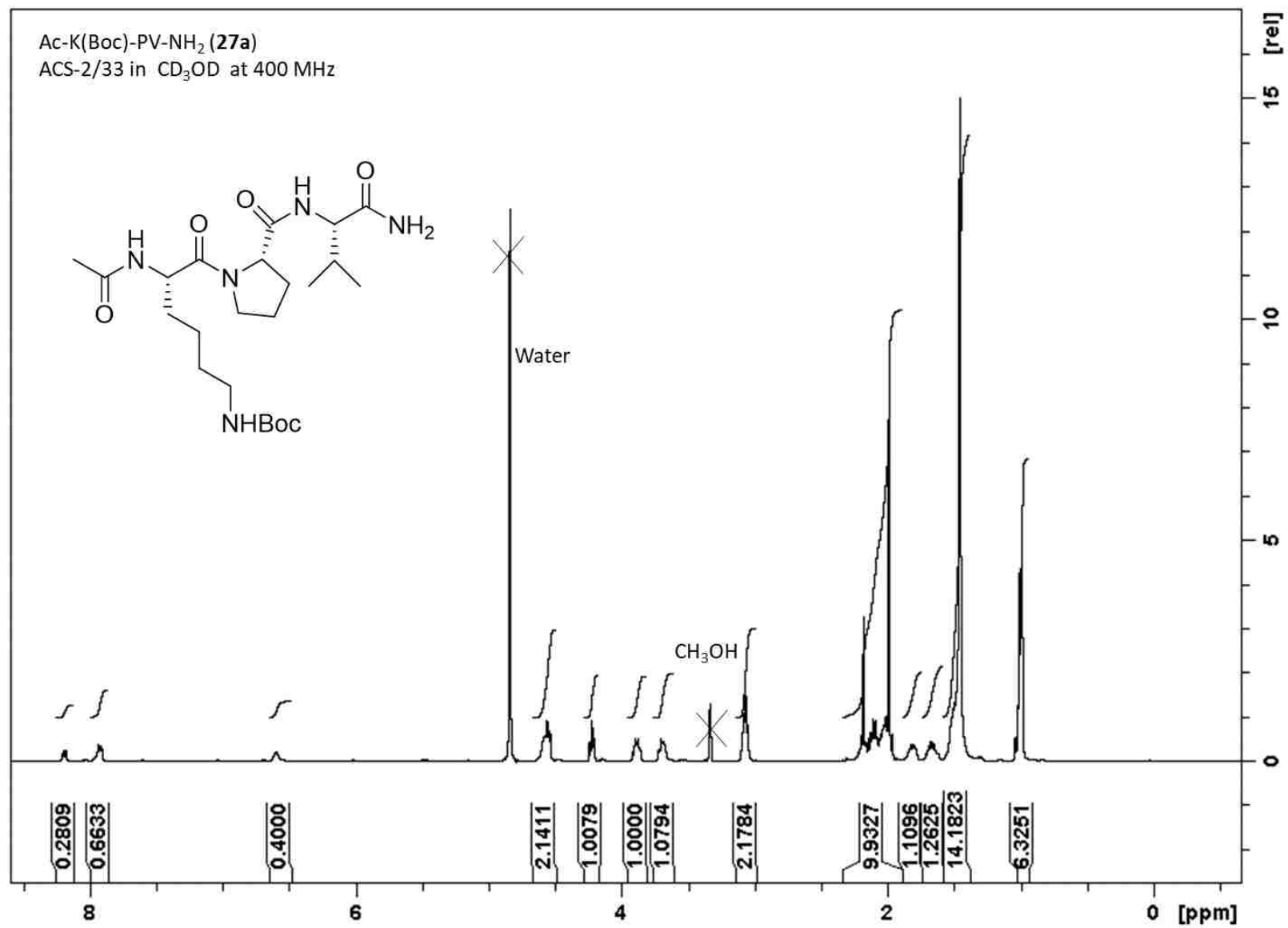


Figure 71. <sup>1</sup>H NMR spectrum of compound **27a**

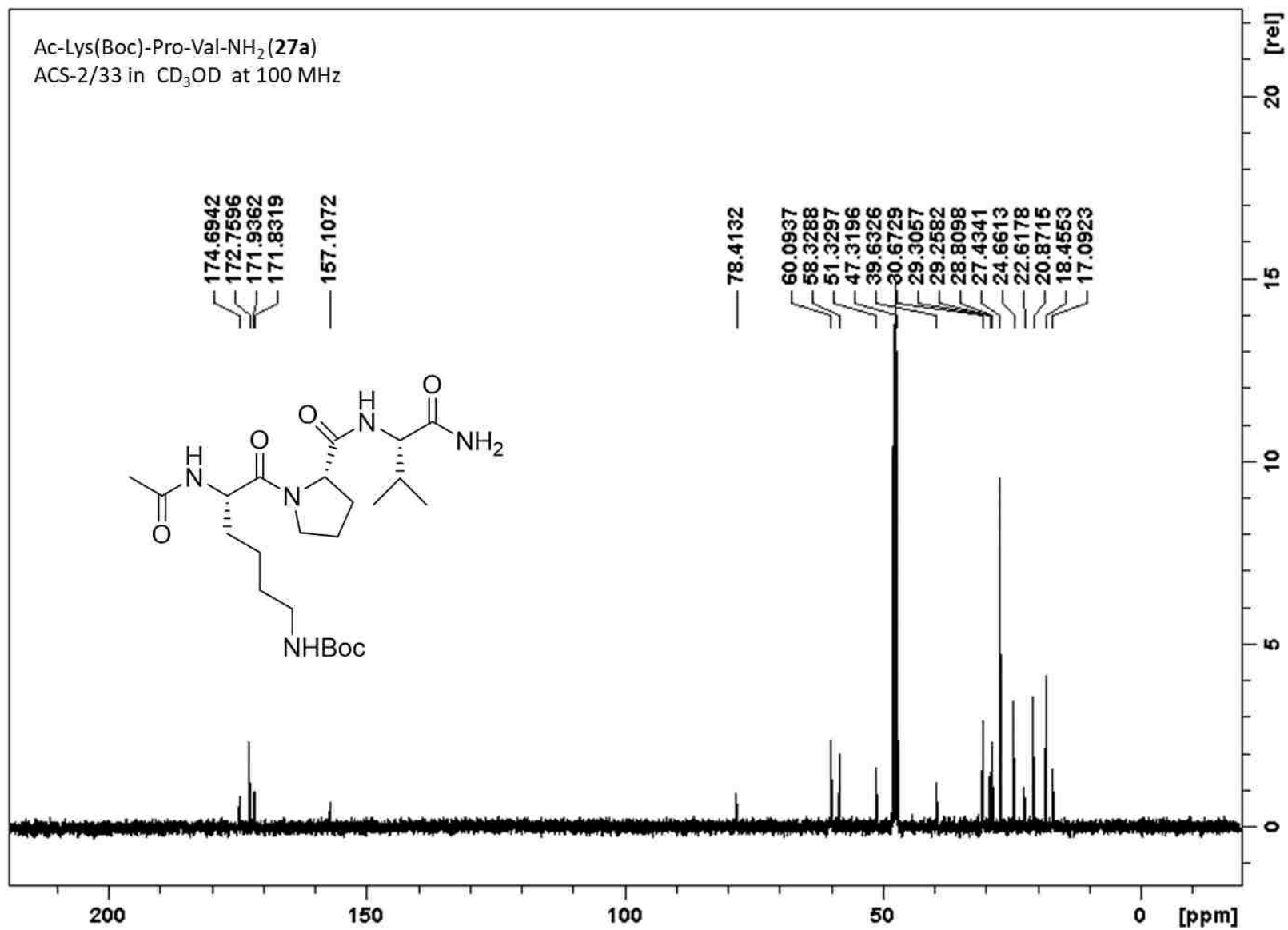


Figure 72. <sup>13</sup>C NMR spectrum of compound 27a

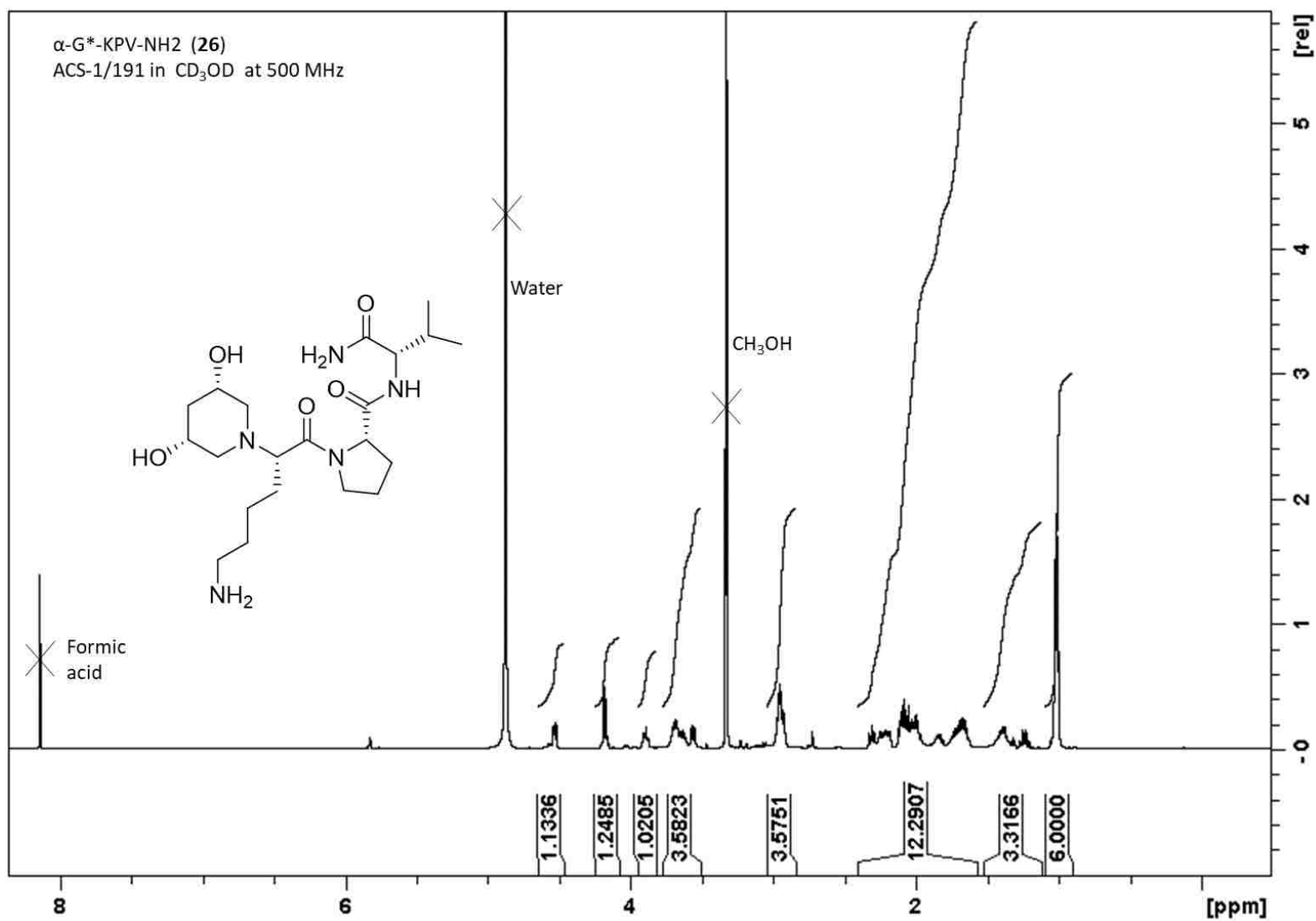


Figure 72. <sup>1</sup>H NMR spectrum of compound **26**

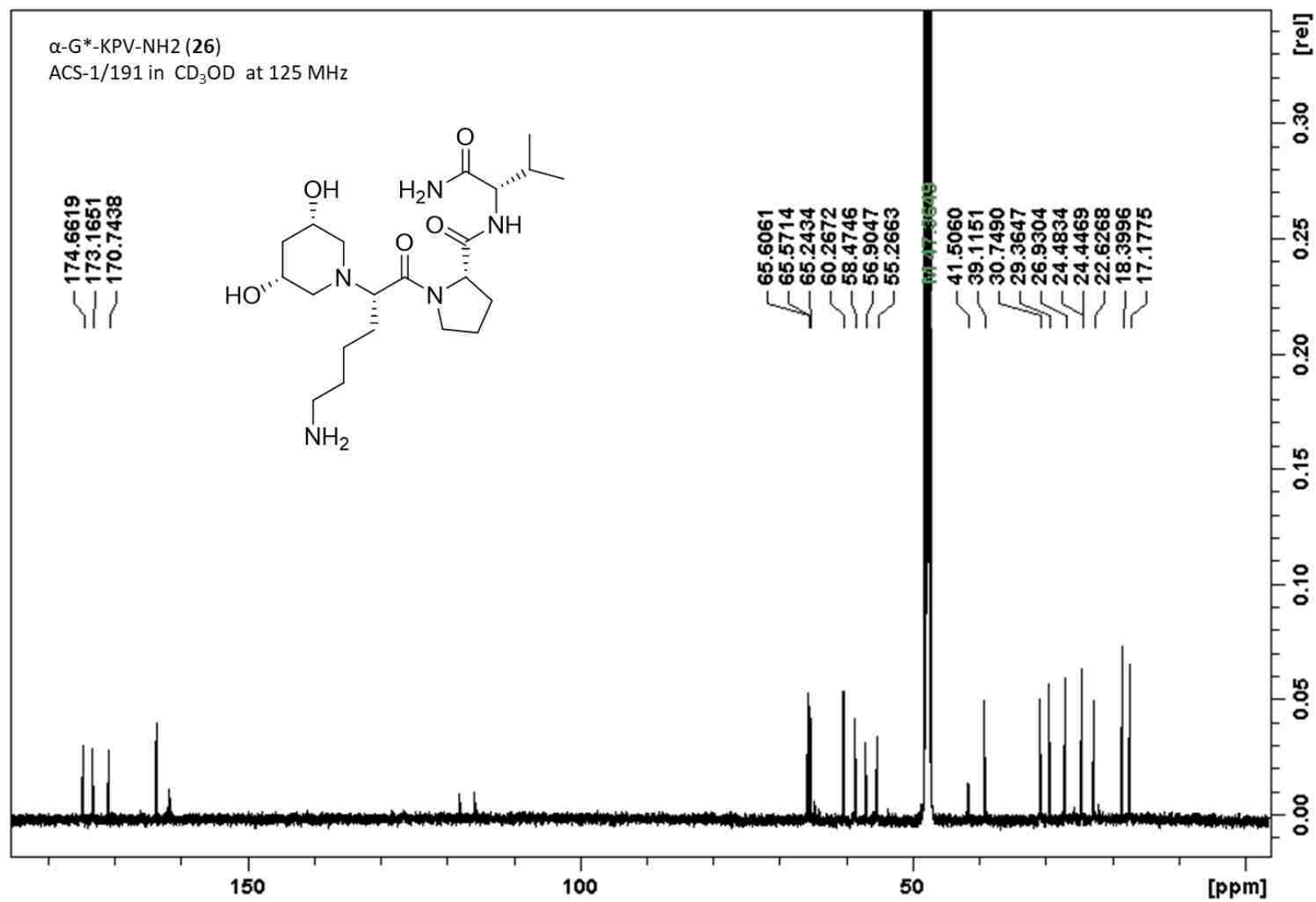


Figure 74. <sup>13</sup>C NMR spectrum of compound 26

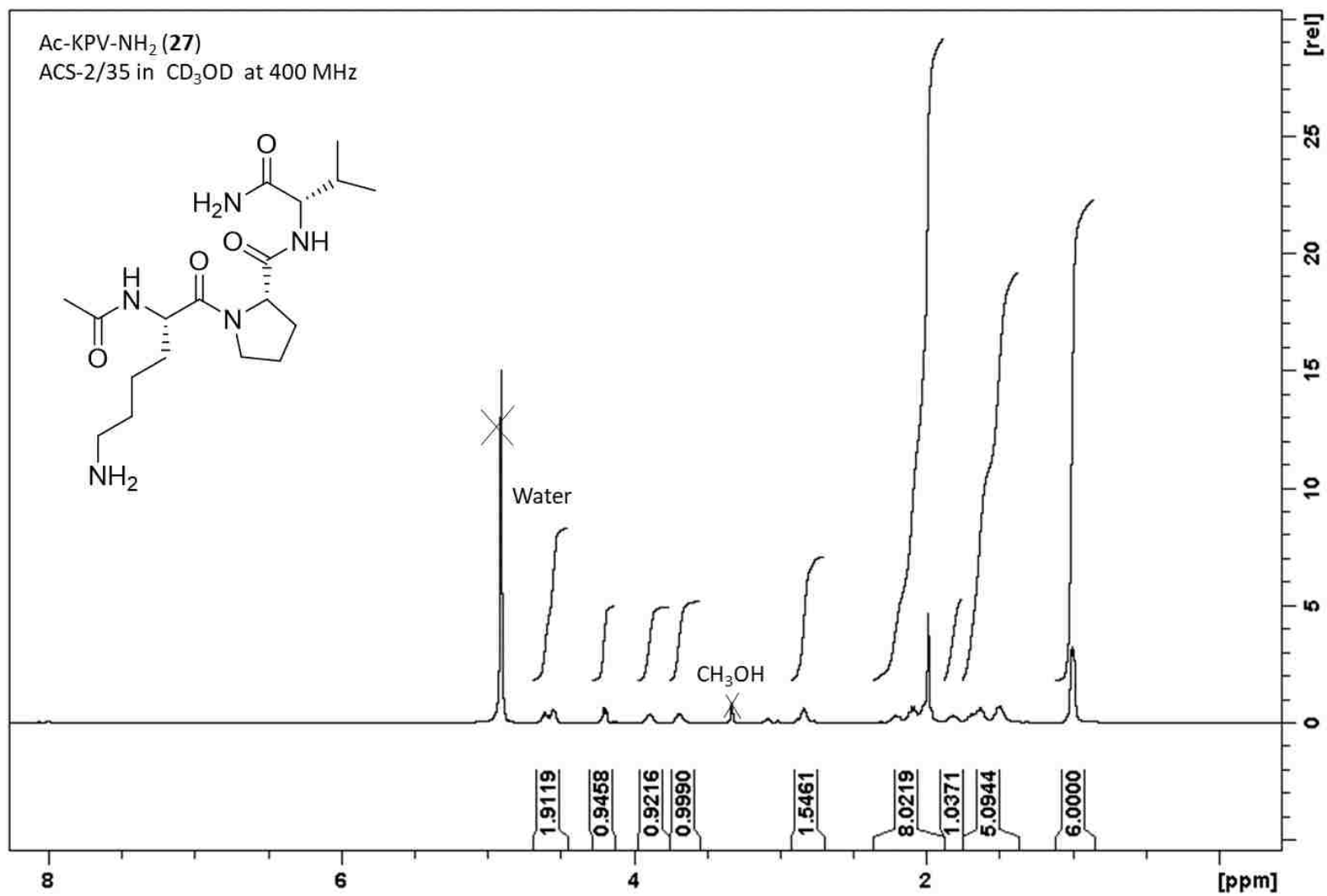


Figure 75. <sup>1</sup>H NMR spectrum of compound 27

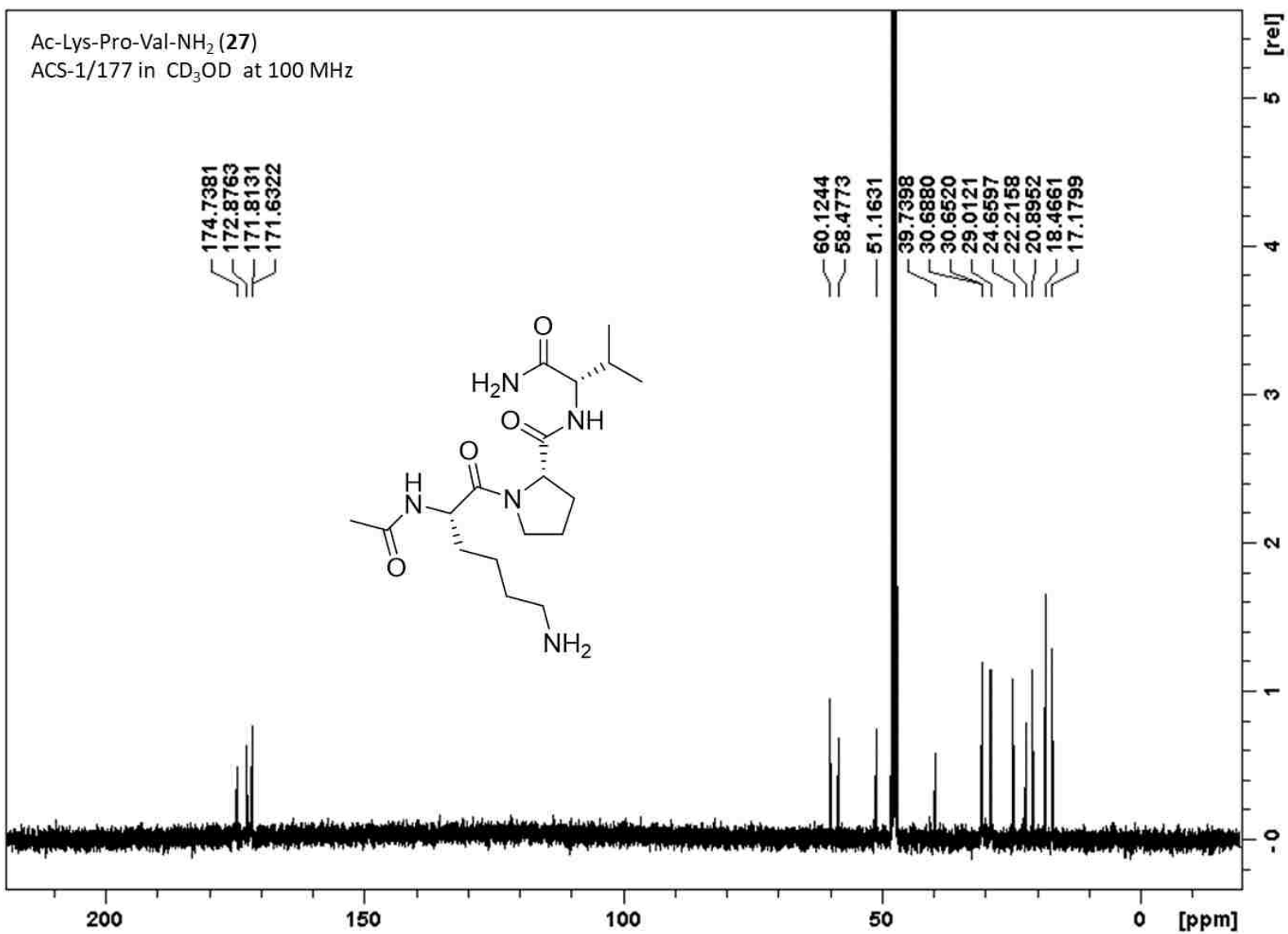


Figure 76. <sup>13</sup>C NMR spectrum of compound 27

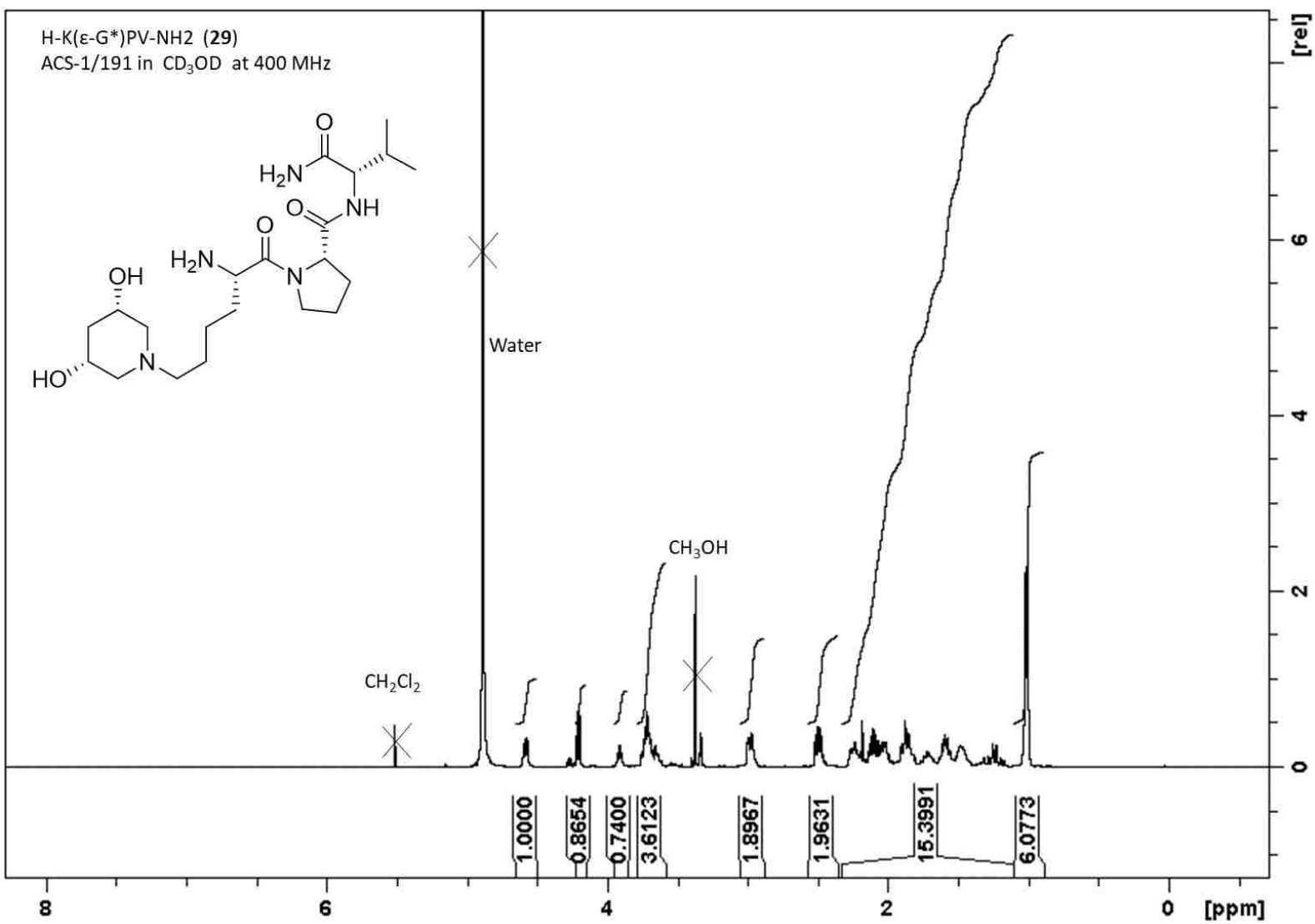


Figure 77. <sup>1</sup>H NMR spectrum of compound **29**

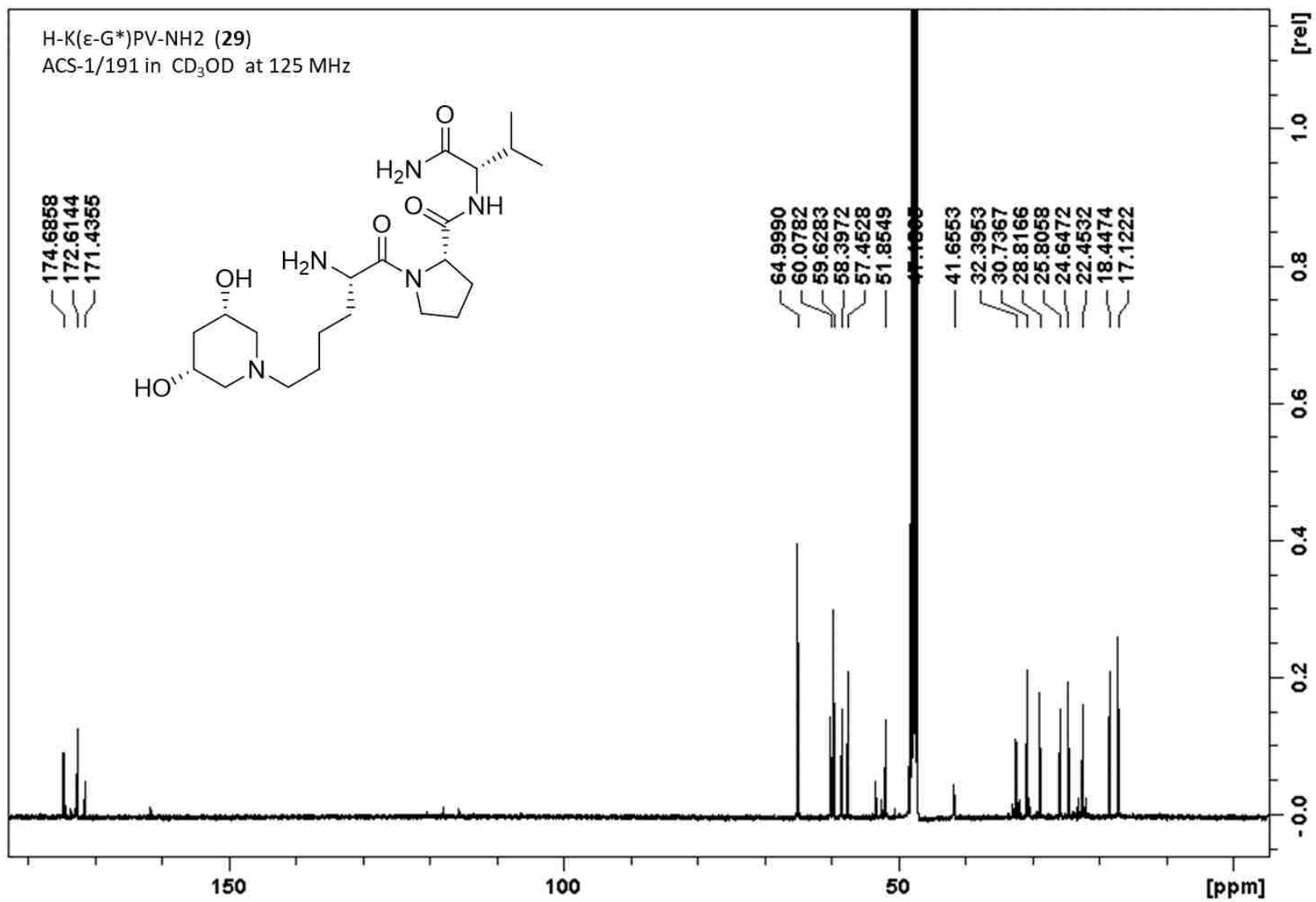


Figure 78. <sup>13</sup>C NMR spectrum of compound **29**



B-5 <sup>1</sup>H NMR spectra of compounds 26, 27, and 29 on pronase cleavage reaction

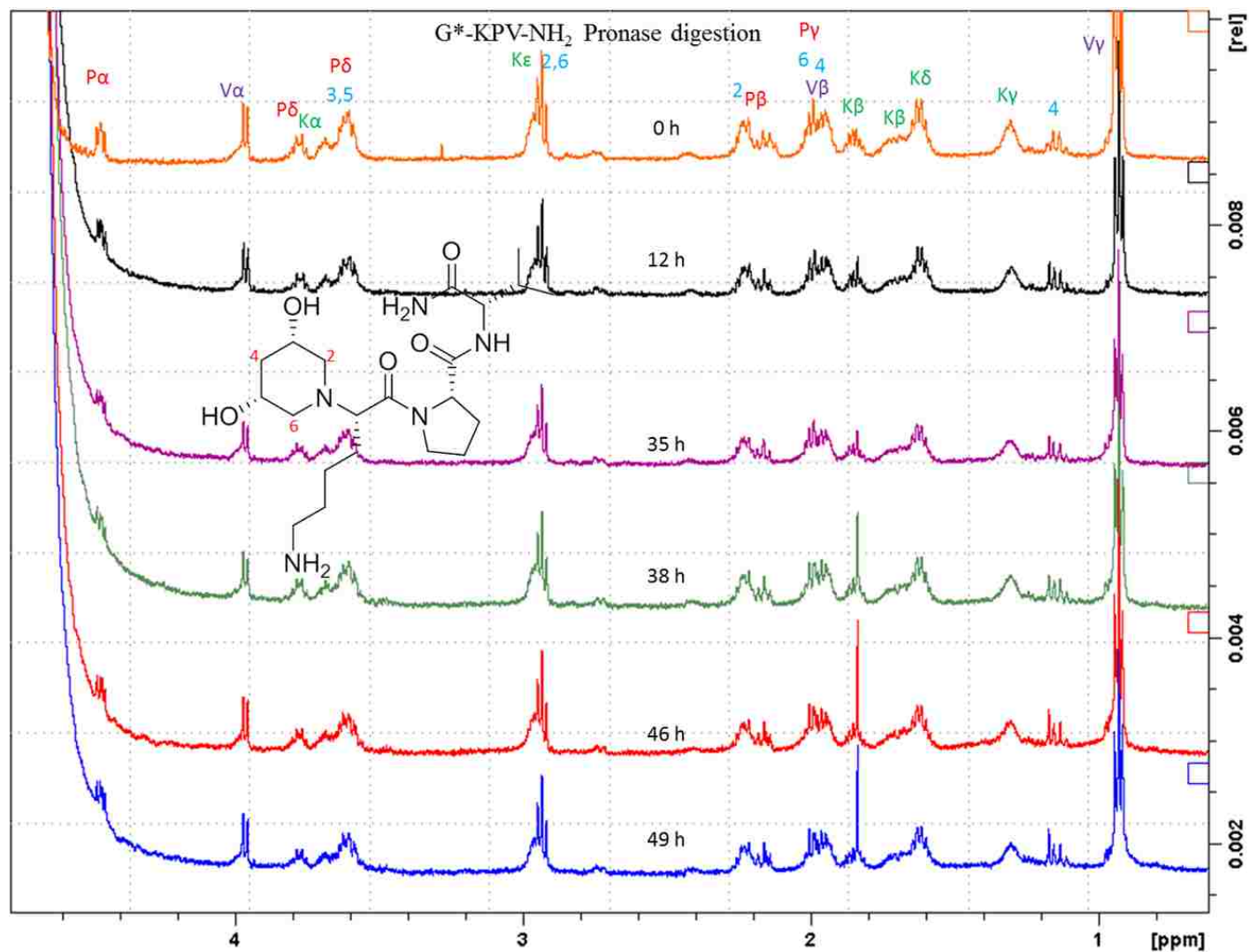


Figure 79. <sup>1</sup>H NMR spectrum of compound 26, pronase reaction

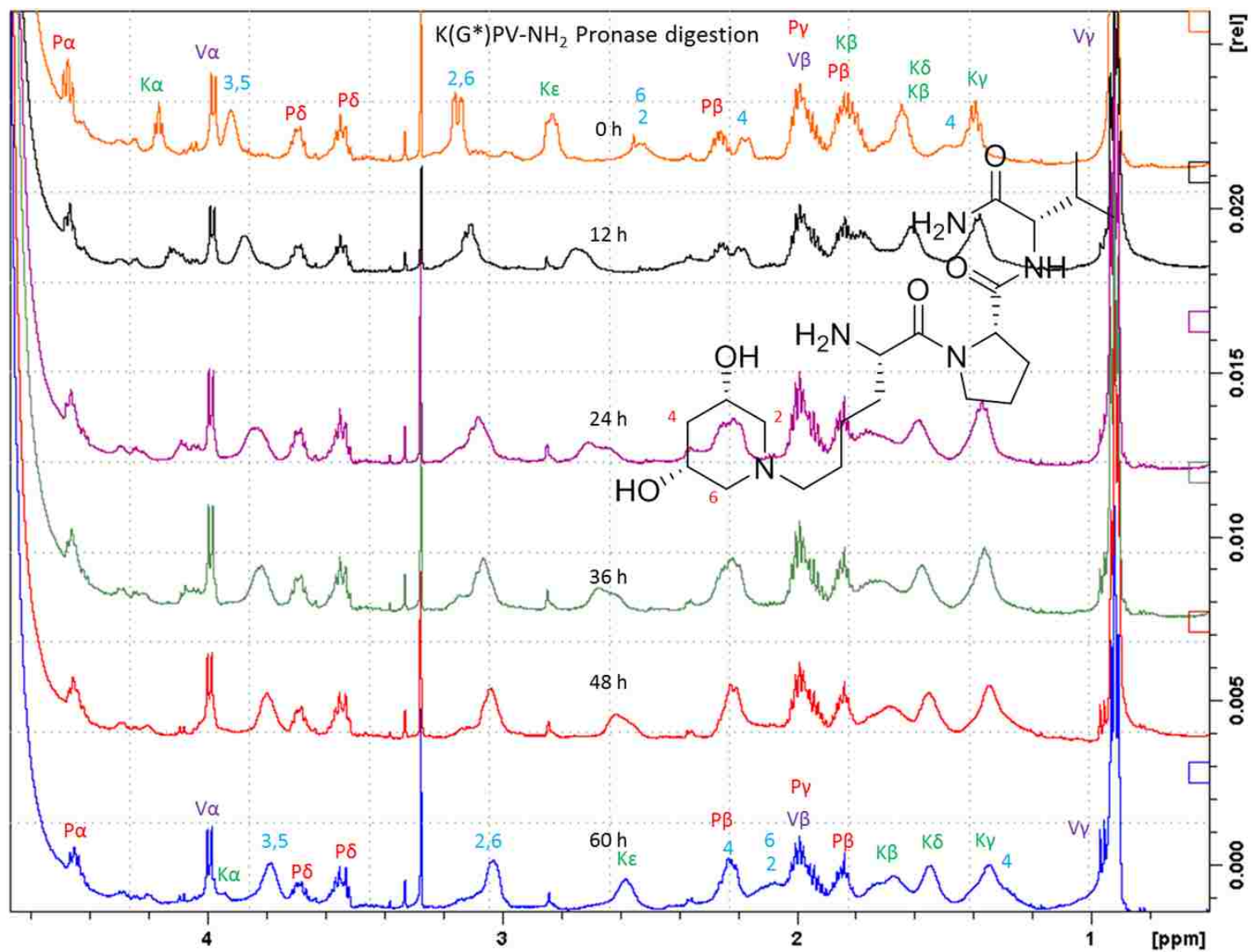


Figure 80. <sup>1</sup>H NMR spectrum of compound 29, pronase reaction

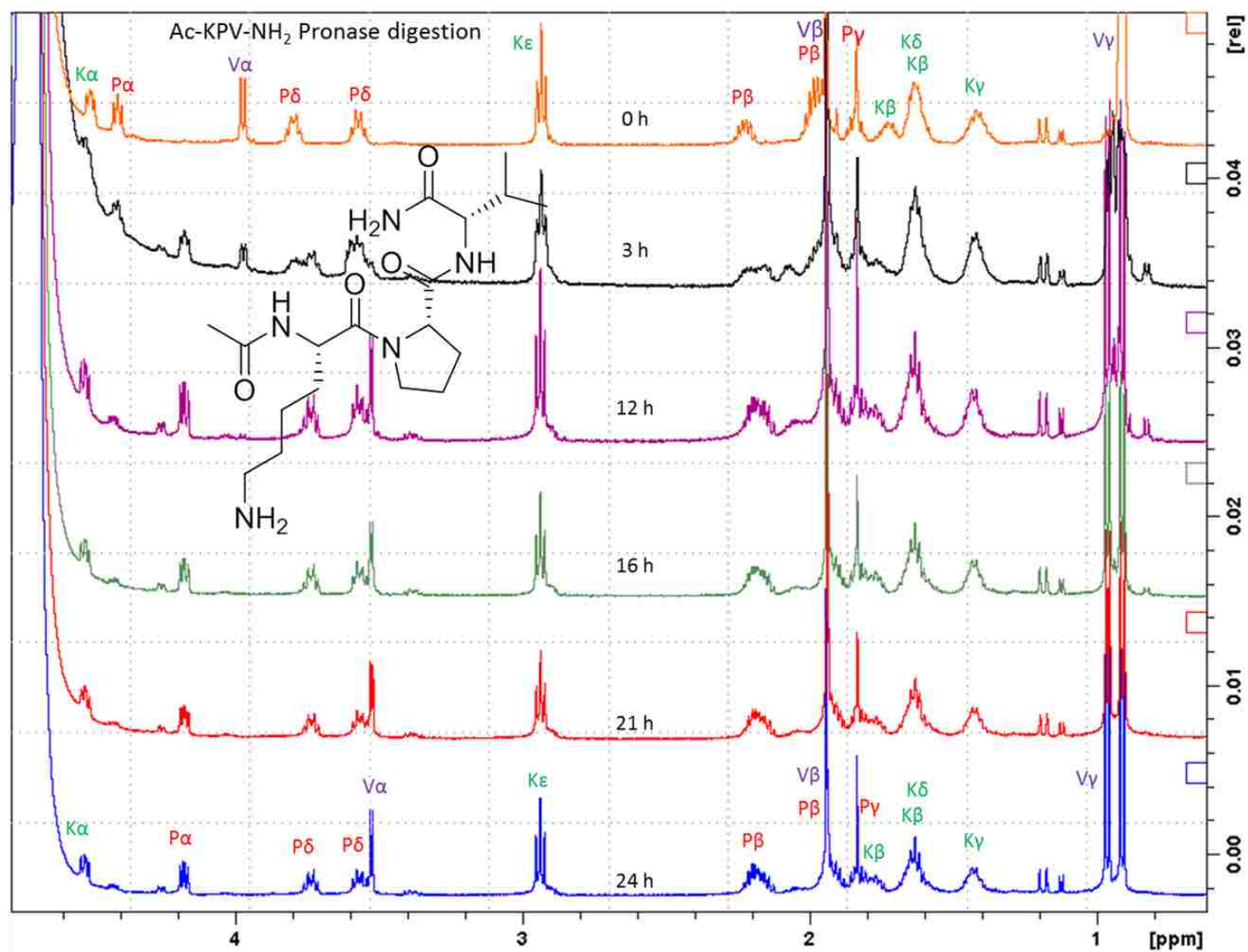


Figure 81. <sup>1</sup>H NMR spectrum of compound 27, pronase reaction

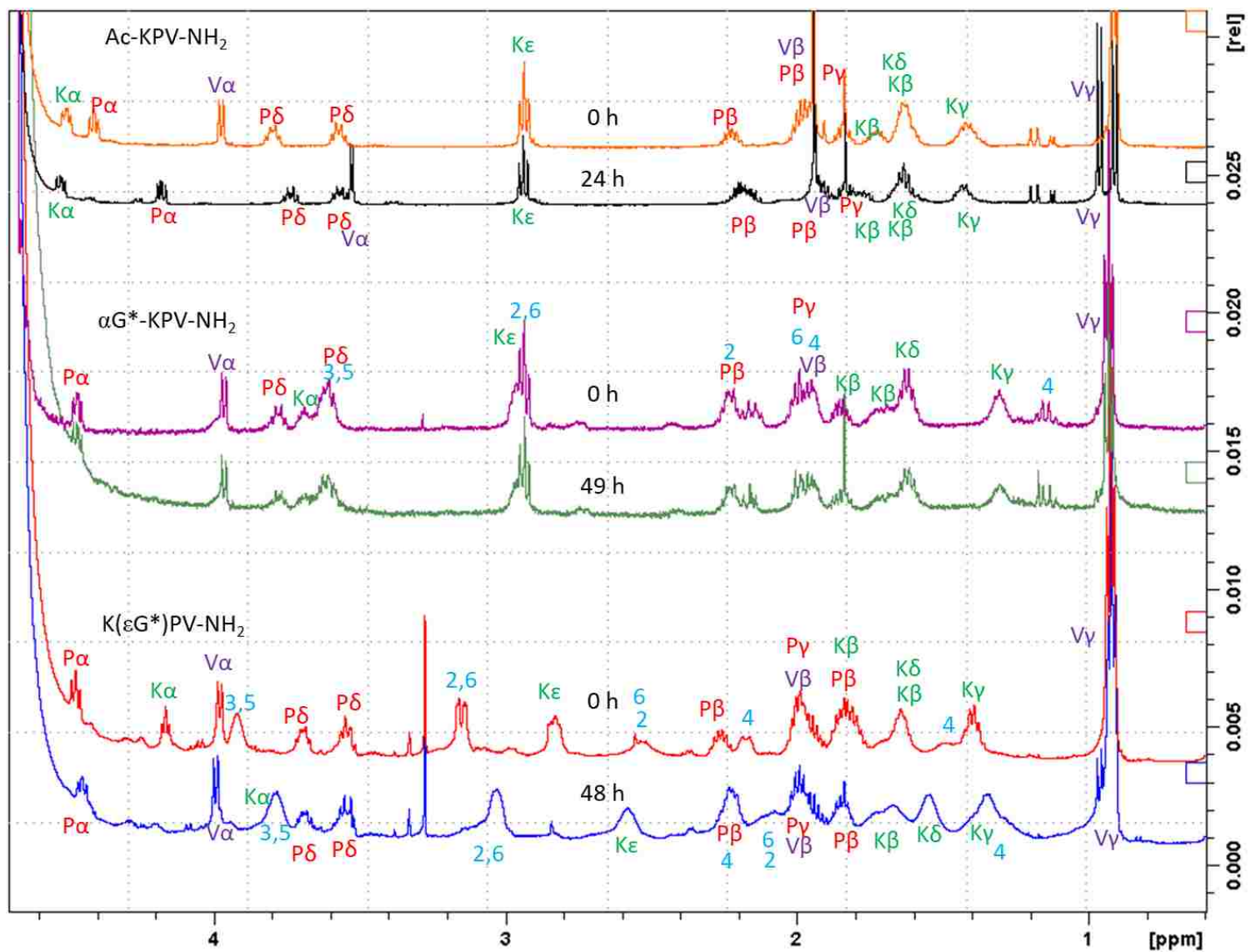


Figure 82. <sup>1</sup>H NMR spectrum of compounds 26, 27, and 29, pronase reaction

## VITA

Abigael Chebichiy was born in Eldoret, Kenya to Leah Jepkemboi. She studied at Kapnyeberai girls high school and joined the University of Nairobi for her undergraduate studies in 2003. She graduated in 2007 with second-class honors, upper division, in Bachelor of Science Chemistry. Abigael taught high school Chemistry, Math, and Biology before joining Louisiana State university in 2012, for her doctoral degree in analytical Chemistry. She joined Dr. Macnaughtan research group and worked under the mentorship of Dr. Megan Macnaughtan and Dr. Carol Taylor. She served as the president for Adventist Christian Fellowship at LSU(ACF@LSU) in 2016-2017 She is a candidate to graduate with her degree of Doctor of Philosophy in Chemistry from Louisiana State University in August 2018.