universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Synthesis of stable isotope-labeled histidine and isoleucine precursors for protein and peptide NMR investigation

verfasst von | submitted by

Alisa Wimmer BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 862
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Chemie
Betreut von | Supervisor:	Mag. Dr. Roman Lichtenecker Privatdoz.

This master thesis was completed under the supervision of Mag. Dr. Roman Lichtenecker at Mag-Lab GmbH and at the Institute of Organic Chemistry at the University of Vienna from September 2023 to April 2024.

Parts of the thesis have been published in J. Mol. Biol. **2025**, *437* (23), 169347.

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Vienna, 16.01.2026

Date

Signature

Acknowledgements

I want to express my deep gratitude to Dr. Roman Lichtenecker for giving me the opportunity to complete my master's thesis under his supervision. I am very thankful for his constant support, encouragement and guidance, as well for the freedom he gave me to explore my interests. His openness, approachability, and willingness to always make time for questions made me feel truly in good hands throughout this journey.

I am also deeply grateful to Matus for teaching and supporting me in the lab every day and for sharing his extensive knowledge of organic chemistry. His guidance helped me take a huge step forward and greatly enhanced my understanding of synthetic chemistry.

A special thanks goes to Sarah, who was always ready to help and encouraged me whenever I doubted myself. Furthermore, I am very grateful for her friendship and good company in the lab.

I would like to thank my family for their love and support throughout my academic journey. I am especially grateful to my mum – my greatest cheerleader – for her constant support, guidance, and encouragement, especially in moments of despair.

I would like to express my sincere gratitude to Dr. Rita Fürst, who gave me the opportunity to complete a fascinating bachelor's thesis under her supervision. Her guidance, encouragement, and praise had a profound positive impact on my self-perception and academic development.

Furthermore, I would like to thank everyone who walked alongside me for part of my academic path, and made this journey both enriching and memorable. I would like to thank the mentors I encountered along the way for teaching, supporting, and guiding me, and for helping me grow academically and personally. Over the years, I have had the honor of working with and learning from incredibly talented people – an experience for which I am deeply grateful.

Table of Contents

Zusammenfassung

Die NMR-Spektroskopie ist heutzutage eine der wichtigsten Methoden um Struktur, Dynamik und Interaktionen von Proteinen auf atomarer Ebene zu untersuchen.[1] Die Analyse großer Biomoleküle mittels NMR wird jedoch durch Einschränkungen in Auflösung und Sensitivität erschwert. Diese Probleme lassen sich durch die gezielte Einführung von ^{13}C-, ^{15}N- und ^{2}H-Isotopen an ausgewählten Positionen überwinden. Dies kann entweder durch uniforme oder selektive Isotopenmarkierung erreicht werden.[2, 3] Ziel dieses Projekts ist die Entwicklung und Optimierung effizienter mehrstufiger Syntheserouten für isotopenmarkierte metabolische Vorläufer der Aminosäuren Histidin und Isoleucin, die den gezielten Einbau von ^{13}C, ^{15}N und ^{2}H in die entsprechenden Aminosäuren eines Proteins ermöglichen.

α-Ketobutyrat bzw. 2-Aceto-2-hydroxybutanoat sind metabolische Vorstufen von Isoleucin und wurden bereits erfolgreich für die selektive Isotopenmarkierung von Isoleucin in *E. coli*-Überexpressionssystemen eingesetzt.[4, 5] Das gezielte Isoleucin-Labeling von Proteinen in Säugtierzellen erfordert jedoch eine andere Strategie, da diesen Zellen ein biosynthetischer Weg zur Herstellung von Isoleucin fehlt. 3-Methyl-2-oxopentanoat ist das erste Intermediat des Isoleucin-Abbaus und Produkt einer reversiblen Transaminierung. Diese Verbindung kann *in vivo* zu Isoleucin umgewandelt werden und stellt somit einen potentiellen Vorläufer für die selektive Markierung in Säugetierzell-basierten Überexpressionssystemen dar.[2]

Ein Ziel dieses Projektes ist die Entwicklung einer geeigneten Syntheseroute für die Herstellung eines Isotopologs von 3-Methyl-2-oxopentanoat, welches eine protonierte δ-$^{13}CH_3$-Gruppe in einer ansonsten deuterierten Seitenkette enthält. Zudem soll untersucht werden, ob dieser Precursor *in vivo* durch eine Transaminase-katalysierte Reaktion in Isoleucin umgewandelt und in weiterer Folge effizient in Proteine eingebaut werden kann.

Für die selektive Histidin-Markierung wird ein ähnlicher Ansatz verfolgt. 2-Hydroxy-3-(1*H*-imidazol-4-yl)acrylsäure wurde bereits als Precursor für die selektive und effiziente Isotopenmarkierung von Histidin in *E. coli*-Überexpressionssystemen eingesetzt.[6] Diese Verbindung ist ein Enol-Tautomer des Histidinabbauprodukts Imidazol-4-pyruvat und wird selektiv über die reversible, transaminasekatalysierte Reaktion in den gewünschten Aminosäurerest umgewandelt.

In diesem Fall besteht das Ziel darin, verschiedene Isotopologe des beschriebenen Vorläufers mit unterschiedlichen Spinsystemen, nämlich ^{1}H-^{13}C, ^{1}H-^{15}N oder ^{13}C-^{15}N, zu synthetisieren. Je nach Markierungsmuster eignen sich diese Verbindungen für unterschiedliche NMR-Anwendungen, um Informationen über Proteinstruktur, -dynamik und -interaktionen zu gewinnen.

Abstract

NMR spectroscopy is one of the most valuable methods to investigate conformational properties and dynamic processes of proteins at an atomic resolution.[1] However, using NMR to study large biomolecules is hampered by limitations in resolution and sensitivity. These issues can be overcome by strategically introducing ^{13}C, ^{15}N and ^{2}H at specific atomic positions resulting in enhanced signal resolution and simplified spectra.[2, 3] This can be achieved by either uniform protein labelling, segmental isotope labeling or by selective side-chain labelling.[7]

The aim of this project is the development of economic and efficient multistep synthetic routes for new isotopologues of selective histidine and isoleucine precursors for protein and peptide NMR investigation. The target molecules are produced from commercially available sources of heavy isotopes using organic synthesis and then applied in *E. coli* or/and mammalian cell-based overexpression systems to access the corresponding isotope labeled protein samples.

α-Ketobutyrate or 2-aceto-2-hydroxybutanoate are early metabolic precursors of isoleucine and have been described as suitable candidates for selective isoleucine labeling in bacterial overexpression systems.[4, 5] The labeling of mammalian proteins, however, demands a different strategy since mammalian cells lack a biosynthetic pathway for isoleucine. 3-Methyl-2-oxopentanoate is the corresponding α-ketoacid of isoleucine and represents a possible precursor for selective labeling in mammalian overexpression systems. As the first intermediate of isoleucine degradation and the product of a reversible transamination, this compound can be converted back into the target amino acid *in vivo*.[2]

This project aims to develop a synthetic route for an isotopologue of 3-methyl-2-oxopentanoate containing a protonated δ-$^{13}CH_3$ group in an otherwise deuterated side chain. Furthermore, it needs to be proven whether the selected precursor can be effectively incorporated into proteins by utilizing a transaminase-catalyzed *in vivo* conversion to the corresponding target amino acid.

For selective histidine labeling, a similar approach was envisioned. Lichtenecker *et al.* introduced 2-hydroxy-3-(1*H*-imidazol-4-yl) acrylic acid as a suitable precursor for efficient histidine labeling in *E. coli* overexpression systems. This compound is an enol tautomer of the histidine degradation product imidazole-4-pyruvate, and selectively converted into the target residue by taking advantage of the reversible nature of the corresponding transaminase activity.[6] In this case, the objective is to synthesize different isotopologues of the previously described precursor containing distinct spin systems, namely ^{1}H-^{13}C, ^{1}H-^{15}N, or ^{13}C-^{15}N. Depending on the labeling pattern, different NMR applications can be envisaged to obtain information about protein structure, dynamics and interaction.

1. Introduction

Proteins are vital molecules that play an essential role in enabling and sustaining life. These biomolecules perform a wide variety of functions, including the enzymatic catalysis of metabolic reactions, DNA replication, regulation of gene expression, transport processes, and chemical messaging. Therefore, a profound understanding of their structure, dynamics and interactions is fundamental for unravelling the basic principles and processes of life.[8] Since proteins are involved in basically every metabolic process, they also play a fundamental role in the onset and progression of numerous diseases. Hence, studying the role of proteins in signaling cascades and interaction mechanisms is crucial for elucidating pathogenic developments, and moreover, for identifying drug targets and efficiently guiding drug design.[9]

1.1. Protein NMR Spectroscopy

The three most valuable experimental tools for structure elucidation of proteins are X-ray crystallography, cryogenic electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) spectroscopy. Each method has its own strengths and limitations, but especially in combination, they enable elucidation of protein structures at atomic resolution, including insights into protein dynamics and protein-ligand interactions.[10]

For the past decades, X-ray crystallography has been the most popular and productive method for uncovering 3D structures of proteins. This technique is based on the diffraction of X-rays at a protein crystal resulting in a unique diffraction pattern, and enables structure elucidation at an atomic resolution.[11] However, X-ray crystallography is limited by the need for homogeneous protein crystals, which can be difficult and time-consuming to produce. Especially insoluble proteins or intrinsically disordered proteins are quite challenging if not impossible to crystallize.[12] Moreover, hydrogen atoms are invisible in most X-ray structures due to their low electron density. As a result, essential interactions like hydrogen bonds cannot be observed. Furthermore, about 20% of protein-bound water molecules, which are crucial for protein-ligand interactions, are not captured by X-ray diffraction.[13] Last but not least, X-ray crystallography captures highly populated low-energy conformations, leaving the molecular dynamics of proteins unresolved.[12, 13]

Thanks to constant improvements in electron microscopy, detection, and image processing, cryo-EM has become an essential tool for protein structure determination. With steadily increasing resolution, currently ranging between 2 and 4 Å, cryo-EM is beginning to rival X-ray crystallography.[14, 15] In cryo-EM, the sample molecule in solution is rapidly frozen at cryogenic temperatures, preserving the proteins native structure. The frozen sample is then exposed to

an electron beam and the detected images are translated into an electron density map.[16] This method enables the visualization of very large complexes or insoluble proteins, which are difficult to crystallize and therefore inaccessible to X-ray crystallography.[15] However, this technique is limited to proteins and peptides larger than 50 kDa. Moreover, cryo-EM is also unable to provide information about molecular dynamics of proteins since the image is generated based on a snapshot of the protein.[17]

Over the past three decades, NMR has evolved into a powerful tool for determining 3D structures of biomolecules in solution. This progress has been driven by advances in isotope labeling strategies, improvements in magnet technology providing higher magnetic fields (and therefore, greater sensitivity and resolution), the introduction of cryogenic probes (to improve signal-to-noise ratio), innovations in spectrometer hardware, and the development of new pulse sequences, as well as a wide range of sophisticated NMR experiments.[17] Furthermore, NMR spectroscopy has one main advantage over X-ray crystallography and cryo-EM, namely the ability to capture protein dynamics.[13, 18]

Proteins are flexible molecules and their dynamic properties are fundamental for their unique abilities and functions. These conformational changes take place in signal transduction, during ligand-binding or the process of enzymatic catalysis, and in response to a variety of other cellular signals. These sometimes short-lived and highly energetic conformational intermediates can be investigated by NMR spectroscopy.[18]

In addition, NMR spectroscopy is able to capture hydrogen atoms and deliver detailed information on hydrogen atom positions, hydrogen bonding networks, and the protonation states of residues. Moreover, protein-bound water molecules can be investigated, which are fundamental for understanding protein-ligand interactions since they stabilize the binding event. Furthermore, NMR enables characterization of water molecules in protein hydration shells, providing information on their lifetimes and hydrogen bonding strengths.[13]

However, like any technique, protein NMR spectroscopy comes with its challenges and limitations. The study of large biomolecules is complicated by severe signal overlap, which can make the interpretation of spectra difficult or even impossible. Moreover, transverse relaxation in large molecules happens faster, which leads to increased line widths, and therefore, to a low signal-to-noise ratio, or even no signals at all.[3, 19] Furthermore, natural abundance NMR measurements can be quite time-consuming due to the low natural concentration of NMR-active isotopes, which demands a high concentration of protein sample and a high number of scans for spectra acquisition.[20] Moreover, this approach is limited to the study of small proteins of 7-10 kDa.[21] Luckily, these issues can be overcome by strategically introducing isotope labels resulting in enhanced signal resolution and simplified spectra.[9] Furthermore, multidimensional

NMR experiments provide reduced spectral complexity and deliver information about connectivity and spatial proximity of atoms.[22, 23]

Several isotope labeling strategies have been developed to address different challenges. Uniform ^{13}C and ^{15}N isotope enrichment of the protein leads to an improved sensitivity ratio and shorter measurement times.[17, 24] The insufficient resolution caused by fast transverse relaxation times (T_2) can be overcome by protein deuteration, which leads to a reduction of transverse ^{1}H-relaxation and an improved signal-to-noise ratio.[19] The overlap/number of signals can be decreased by introducing selective isotope labels, which lead to a reduction in the number of spectral resonance lines and, therefore, simplified spectra. Furthermore, highly selective labeling pushed the molecular weight limit of proteins that can be studied using NMR spectroscopy.[9]

1.1.1. The NMR Experiment

NMR spectroscopy relies on the ability of certain atomic nuclei to absorb and re-emit radio-frequency energy when placed in an external magnetic field. When nuclei with a non-zero spin are placed in an external magnetic field (B_0), the nuclear magnetic dipoles are aligned in two different orientations relative to the direction of the magnetic field. These orientations correspond to different discrete nuclear spin energy levels. The magnetic dipole moment of a spin-half nucleus (e.g. ^{1}H, ^{13}C, ^{15}N) can assume one of two orientations: In the ground state, it is aligned with the direction of the magnetic field, and in the excited state, it is oriented in the opposite direction.[21, 25]

Transitions from the ground to the excited state can be induced by irradiating the sample with radiofrequency waves. This excitation field (B_1) can be applied either by scanning through multiple wavelengths (continuous wave NMR), or as a short burst of radio frequency that excites a broad range of transitions (pulsed NMR).[21]

Before the pulse, the magnetic dipoles precess about B_0 at a frequency that is equal to their resonance frequency. During the excitation, the bulk magnetic moment of the sample is tipped from the z-axis to the x-y plane. After the pulse, when the system relaxes back toward equilibrium, the bulk magnetization precesses about the direction of B_0 inducing a time dependent current in the receiver coil. This signal is called the free induction decay (FID) and represents bulk magnetization that exists in the x-y plane.[21]

Excited nuclear spin dipoles experience two different relaxation processes. The spin-lattice (or longitudinal) relaxation describes the re-alignment of the longitudinal component of the magnetization vector M_z with the direction of the magnetic field. In this process, the excited

spins transfer their energy to the surroundings, also called lattice. The rate of this process is characterized by the relaxation time T_1.[21]

The spin-spin relaxation (or transverse) relaxation arises from dipolar couplings between magnetic dipoles and describes the loss of x- and y-magnetization – the mechanisms by which the transverse component of the magnetization vector M_{xy} relaxes towards its equilibrium value.[25] Fast transverse relaxation times, characterized by the constant T_2, result in increased linewidths causing a decrease in resolution and signal-to-noise.[21]

The frequency of the absorbed radio waves depends on the strength of the magnetic field, the isotope, and the chemical environment of the nucleus. As the electron density around the nucleus increases, the effective field decreases, leading to lower resonance frequencies. The observed frequency is referred to as a chemical shift.[21, 22]

As already mentioned, NMR characterization of proteins is hindered by severe signal overlap and a low signal-to-noise ratio. Some of these issues can be alleviated by introducing an additional frequency dimension. In two-dimensional NMR, the position of a peak is defined by the resonance frequencies of two coupled spins. The introduction of an additional frequency dimension allows for improved signal resolution and separation of overlapping peaks.[21] Furthermore, multi-dimensional NMR is very useful for identifying the connectivity and spatial proximity of nuclei.[22]

Two-dimensional NMR experiments consist of a sequence of radio frequency pulses with delay periods in between. These pulse-sequences can be divided into four steps (Figure 1).[25]

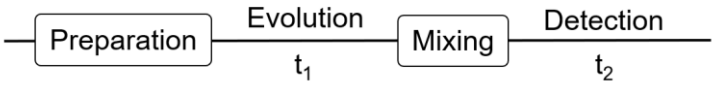


Figure 1: Scheme of a 2D-NMR experiment.

During preparation, the sample is excited by one or more pulses. In the following evolution period, the resulting magnetization is allowed to precess freely for a variable time period t_1. In the mixing period, the sample is excited once more, which leads to magnetization transfer between two spins. Lastly, the signal is recorded as a function of the second time variable, t_2. This sequence is repeated multiple times, whereby the time delay t_1 is incremented by a fixed Δt_1 each time. For each value of t_1, a FID is recorded as a function of t_2.[25]

1.1.2. Structure Determination of Proteins

Protein NMR spectroscopy uses a variety of experiments with diverse pulse sequences, each designed to investigate specific structural elements, in combination with sophisticated labeling strategies to elucidate protein structure and dynamics. In this way, the labeling pattern of a protein can be tailored to the specific requirements of the experiment.[17, 24]

Multi-dimensional HSQC (Heteronuclear single-quantum coherence) experiments couple the spins of nuclei connected by a single bond via through-bond scalar couplings. ^{1}H,^{15}N-HSQC provides backbone amide correlations and produces a signal for the backbone amide group of each amino acid residue. In addition, side-chain NH-groups are also visible in the HSQC spectrum. ^{15}N-HSQC experiments are often performed using uniformly ^{15}N-labeled proteins.[23]

NOESY (Nuclear Overhauser effect spectroscopy) exploits the Nuclear Overhauser Effect to visualize through-space dipolar couplings between nuclear spins. NOESY couples spins that are spatially close to each other, even if they are far apart in the protein sequence.[22] As a result, this technique reveals spatial proximity between nuclei less than 5 Å apart, and delivers information about the two- and three-dimensional fold of the peptide chain.[21, 22]

The intensity of a NOESY cross-peak depends on the distance between two ^{1}H nuclei. Therefore, these signals can be translated to inter-proton distances, usually referred to as NOEs.[21, 22] Medium-range NOEs deliver information about the local conformation of the protein backbone and are used to determine secondary structure elements, while long-range NOEs elucidate the arrangement of secondary structure elements and are helpful for the determination of the tertiary structure.[22] Moreover, NOEs are sensitive reporters of short-range through-space intra- and intermolecular interactions and deliver information about the internal geometry of receptor bound ligands.[26]

However, conventional two-dimensional NOE spectra of large biomolecules are quite complex and difficult to analyze. Isotope-edited NOESY, on the other hand, is based on the selective observation of protons attached to labeled nuclei and shows selected correlations.[27] ^{13}C-edited NOESY detects NOEs between ^{1}H nuclei that are attached to ^{13}C atoms and can, for example, visualize methyl-methyl distances, which provide information about global folds of proteins.[28] Similarly, ^{15}N-edited NOESY correlates ^{1}H nuclei that are attached to ^{15}N-labeled nitrogen and is useful for determining the three-dimensional fold of the polypeptide chain.[22, 29]

The TOCSY (Total correlation spectroscopy) experiment correlates all protons within the same spin system (e.g. an amino acid residue) with each other through scalar couplings. In this way, this technique provides cross-peaks for protons that are part of the same spin system and produces a characteristic resonance pattern for each amino acid. The comparison of a 2D-

TOCSY and a 2D-NOESY spectrum, for example, enables the discrimination between intra- and inter-residual proton cross-peaks.[22]

With increasing protein size, 2D spectra become even more complex due to signal overlap.[22] Furthermore, excited spins relax faster in large molecules leading to a broadening of the signals and poor sensitivity.[1] TROSY (Transverse relaxation-optimized spectroscopy) reduces transverse relaxation by exploiting destructive interference effects between relaxation mechanisms.[28] In this way, TROSY generates spectra with enhanced resolution and signal-to-noise, and thereby, greatly extends the molecular size range accessible to NMR.[19] The introduction of $^{13}CH_3$ groups in an otherwise deuterated protein allows the study of biomolecules with molecular masses of more than hundreds of kilodaltons.[30]

Another approach to overcoming the size problem is the introduction of an additional frequency dimension. 3D experiments can be constructed by combining the pulse sequences of two 2D experiments, for example NOESY and HSQC, or TOCSY and HSQC. In a 3D spectrum, the peaks are distributed over three frequency axes, which alleviates the problem of signal-overlapping.[22] To exemplify, ^{15}N-NOESY-HSQC correlates protons close in space while maintaining a link to the attached nitrogen, while ^{15}N-TOCSY-HSQC correlates protons within an amino acid and connects them to their intra-residual amide group.[31, 32] The corresponding ^{13}C experiments (^{13}C-NOESY-HSQC and ^{13}C-TOCSY-HSQC, respectively) are highly useful to identify NOE signals of side-chain protons.[22]

Another form of three-dimensional NMR spectroscopy is the triple resonance experiment, which correlates the chemical shifts of three different nuclei and involves magnetization transfer between ^{1}H, ^{15}N, and ^{13}C atoms. The various triple resonance experiments employ different magnetization transfer pathways to correlate nuclei. For example, the HNCA experiment transfers magnetization from the backbone amide proton to the adjacent ^{15}N nucleus, and then to the $^{13}C_\alpha$ of both, the same residue and the preceding residue before moving back to the amide proton. In contrast, HCA(CO)N transfers magnetization from the C_α-hydrogen to the directly bonded $^{13}C_\alpha$ nucleus and then via the carbonyl carbon to the amide nitrogen. Although the carbonyl carbon participates in magnetization transfer, its chemical shift is not detected. These experiments are especially useful for studying larger proteins (> 15 kDa) since each amino acid residue only produces a few signals.[22]

By comparing and combining the data provided by the various NMR experiments, each signal can be assigned to its corresponding nucleus. This process is referred to as sequence specific assignment and is a crucial part of protein structure determination.

In addition to determining protein structure, NMR spectroscopy is also able to capture protein dynamics. Proteins are dynamic molecules and populate a range of interconvertible

conformational states including local backbone fluctuations, side-chain flips, rearrangements of structural domains and even changes in oligomeric subunits.[18] In order to understand the functions and dysfunctions of proteins in health and disease, a characterization of the accessible conformational states is required. Especially the investigation of sparsely populated and highly-energetic intermediates, which are essential for protein function, is of immense importance for understanding the structure-function relationship of proteins.

NMR spectroscopy allows for the investigation of protein dynamics on different time scales ranging from femtoseconds to hours and even days. Fast time scale dynamics, such as bond vibrations, side-chain rotations, or loop motions are determined by measuring spin relaxation rates, including longitudinal and transversal relaxation rates, and heteronuclear NOEs.[30, 33] Slow time scale motions such as protein folding and unfolding, domain motions, protein-ligand binding, and enzyme catalysis are investigated using techniques like relaxation dispersion or exchange experiments (EXSY, ZZ-exchange). The relaxation rates of backbone amide groups deliver information about backbone flexibility, whereas side-chain methyl groups (of Leu, Ile, and Val) are sensitive reporters of side-chain dynamics.[23]

The measurement of amide proton exchange rates can also provide useful information about structural dynamics.[34] For that purpose, the protein is dissolved in D_2O and the H/D exchange rates of amide protons are determined by following the loss of signal intensity. Amide protons involved in H-bonds stabilizing secondary structure elements as well as protons located in the center of the protein have slower exchange rates, whereas amide protons on the protein surface exchange faster.[22]

Another useful technique in protein NMR spectroscopy is chemical shift perturbation (CSP), which detects changes in the chemical shifts of a protein upon interaction with another molecule. Theses changes can be used to determine the amino acids involved in ligand binding, the dissociation constant of the ligand and the structure of the protein-ligand complex.[35]

Furthermore, NMR is useful for the study of intrinsically disordered proteins (IDPs). Nearly one third of human proteins are IDPs or contain intrinsically disordered regions that lack a fixed tertiary structure. Instead, they populate a range of rapidly interconverting structural states.[18] However, in some cases the presence of a binding partner can lead to the folding of IDPs. In recent studies, scientists prepared a sample of an isotopically labeled IDP and added sub-stoichiometric amounts of a binding partner, which led to complex formation and the generation of a sparsely populated bound state. The chemical shifts from the invisible bound state could be obtained through chemical exchange NMR methods.[36-38]

In summary, protein NMR spectroscopy is a powerful technique for elucidating protein structure, dynamics, and interactions at atomic resolution. Advanced pulse sequences, isotope labeling, and multi-dimensional experiments allow studies of protein motions, ligand binding, and even sparsely populated conformational states. Consequently, NMR spectroscopy is an indispensable tool in modern structural biology.

1.1.3. NMR Spectroscopy in Drug Discovery and Design

NMR spectroscopy has evolved into an important technique to guide drug-development processes as it provides site specific information about molecular interactions at atomic resolution. Furthermore, weak and reversible interaction such as van der Waals, electrostatic, and hydrogen bonding interactions can be studied, which contribute significantly to the affinity of protein-ligand interactions.[39] For that purpose, several techniques have been developed.

Chemical shift perturbation (CSP) detects changes in chemical shifts caused by ligand binding and enables the identification of the binding site and the interacting amino acids. In the standard experiment, an unlabeled ligand is titrated to a ^{13}C- and/or ^{15}N-labeled protein. At each titration point, the chemical shifts are measured by acquiring a 2D HSQC spectrum. Chemical shift changes are very sensitive to structural changes, which means that even subtle ligand interactions cause observable CSPs. By analyzing how the peaks move throughout titration, the ligand binding site can be identified. Moreover, the dissociation constant of the ligand can be determined by analyzing the parameters of the titration curve.[35] The selective labeling of amino acid side-chains with isolated ^{1}H-^{13}C spin systems allows the identification of the residues involved in ligand binding.[13]

Other methods detect ligand binding by monitoring changes in the transverse relaxation rate T_2. Small molecules usually exhibit longer transverse relaxation rates compared to macromolecules such as proteins. However, upon binding to a protein, the T_2 of a ligand increases. By measuring the T_2 values of a potential ligand before and after the addition of a protein target, binding events can be detected, whereby a broadening of the resonance lines of the ligand indicates ligand binding. Similarly, changes in the longitudinal relaxation rate T_1 can serve as reporters for ligand binding. If the ligand is not bound to the protein, small positive NOEs can be observed. Upon binding, small negative NOEs can be detected.[26]

Another indispensable tool in NMR-supported drug discovery is saturation transfer difference (STD) NMR. This technique is quite useful for the screening of compound libraries to identify their binding activities to target proteins and to characterize the binding epitopes.[40]

STD NMR is based on the magnetization transfer from a saturated protein receptor to any bound ligand.[41] For that purpose, two spectra are recorded. First, the protein is exposed to a selective pulse, which saturates only a few protein resonances (*on-resonance* spectrum). This saturation is spread over the entire protein by intramolecular saturation transfer. Ligand molecules, which bind to the receptor, are saturated by intermolecular saturation transfer, while non-interacting molecules are not affected. By subtracting the *on-resonance* spectrum from a spectrum without pre-irradiation (*off-resonance* spectrum), interacting ligand molecules can be identified. The difference spectrum only shows the resonance of hydrogen atoms of the binding ligand in close contact with the protein.[40, 42]

STD NMR enables the identification of binding ligands from mixtures containing different molecules varying in concentration and size. It is insensitive to ligand excess and minimizes the risk of detecting false positives since only saturation transferred to the ligand is detected, Furthermore, this technique is not limited by the size of the protein. In fact, the sensitivity increase with protein size due to a more efficient intra- and intermolecular saturation transfer.[40] WaterLOGSY is a related technique, that is based on magnetization transfer from bulk water to the protein binding site and on to an interacting ligand.[26]

An alternative method for identifying high-affinity ligands is SAR (structure-activity relationship) by NMR – a linked-fragment approach, where ligands are constructed from individual building blocks that have been optimized to bind to proximal protein binding sites.[43]

The process starts with the screening of a ligand library to identify an initial lead compound, which is then optimized by testing structurally related analogues. Subsequently, a second ligand is identified that binds to a nearby binding pocket. The analysis of chemical shift changes allows the identification of the approximate location of the second pocket. This step is followed by the optimization of the second ligand by screening of analogues. Once the two lead fragments have been identified, their binding locations and orientations are determined. Finally, new compounds are synthesized by linking the two fragments with the goal of creating a ligand with increased binding affinity.[43]

SAR by NMR is usually based on the interpretation of ^{15}N–HSQC experiments to detect fragment binding to the ^{15}N-labeled protein and to identify the different binding site locations. The ^{15}N-HSQC spectrum only shows signals from the ^{15}N-labeled protein, while unlabeled ligands are not observed. This eliminates background interferences and allows detection of binding events even at high ligand concentrations. On the other hand, this approach is limited to relatively small proteins (< 30 kDa) that can be produced in large quantities (typically around 200 mg). Despite this limitation, many proteins within this size range represent important drug targets. Additionally, the ligands used in these studies must be sufficiently soluble in water.[43]

Modern drug discovery focuses on the optimization of weak, reversible interactions, such as van der Waals, electrostatic, and hydrogen bonding interactions, to design highly selective and potent compounds that target proteins implicated in the onset and progression of various diseases. CH-π interactions are a special type of non-covalent forces that occur between the π-electrons of aromatic ring systems and the CH-groups of aromatic or aliphatic hydrocarbons. These hydrogen-bonding interactions fundamentally determine affinity and selectivity in protein-ligand interactions. NMR spectroscopy combined with the introduction of isolated ^{1}H-^{13}C spin systems into amino acid side-chains enables the identification of CH-π interactions relevant for ligand binding. By analyzing the chemical shift changes induced by such interactions, the strength of each individual CH-π interaction can be determined. This approach allows to specifically fine-tune CH-π interactions in protein-ligand complexes and opens new doors in drug design and lead optimization.[39]

In summary, protein NMR is a powerful tool for investigating protein-ligand interactions. Chemical shift mapping allows the localization of ligand binding sites by analyzing the chemical shifts changes caused by the binding event.[35] STD NMR and waterLOGSY are useful for screening compound libraries and identifying binding ligands, while SAR by NMR uses a fragment-based approach to create new high affinity ligands.[40, 43] Finally, NMR spectroscopy can be used to optimize CH-π interactions involved in ligand binding, thereby enhancing ligand affinity.[39]

1.2. Isotope Labeling Strategies

Protein structure elucidation by NMR spectroscopy would not be possible without the constant development of stable isotope labeling techniques. The incorporation of isotope labels at strategic positions enables investigation of structure, dynamics and interactions of large proteins by enhancing signal resolution and spectral simplicity.[44, 45]

Several methods for isotope labeling of proteins have been developed. In uniform protein labeling, every atom of a given element is exchanged for another, less abundant isotope, that is useful for NMR applications. Carbon, nitrogen, and hydrogen are the atoms forming the framework of proteins and are the nuclei of choice for performing uniform protein labeling. Their naturally predominant isotopes ^{12}C, ^{14}N, and ^{1}H are substituted by ^{13}C, ^{15}N, and ^{2}H, respectively.[23]

Incorporation of ^{13}C and ^{15}N nuclei is achieved by supplementing the growth media of a host organism with labeled nutrients, usually ^{13}C-glucose and ^{15}N-ammonium salts. These substrates are then used by the host organism for protein biosynthesis. The incorporation of the spin ½ nuclei ^{13}C and ^{15}N allows for an improved intensity and signal-to-noise ratio. Nitrogen-15 nuclei are very useful for studying protein structure and backbone dynamics by measuring the relaxation rates of the backbone amide groups in 2D [^{15}N, ^{1}H] HSQC experiments.[23] For triple resonance experiments, uniform ^{13}C and ^{15}N labeling is required.[22]

Hydrogen atoms are omnipresent in proteins and represent the main source of relaxation. Deuteration reduces the proton density within the protein and leads to a reduction of transverse relaxation, resulting in enhanced resolution and sensitivity. However, hydrogens are also sensitive reporters of structural information, therefore complete deuteration is not useful. As a compromise, proteins can be deuterated to a certain extent (such as 70%).[19] Alternatively, the introduction of an isolated ^{1}H-^{13}C spin system in an otherwise deuterated protein allows site-specific observation of that labeled position.[28] Deuteration of proteins is achieved by simply using deuterated glucose and D_2O for protein expression. If necessary, the backbone amide protons can be reintroduced by using an H_2O-based buffer.[23]

Another approach for reducing spectral complexity is segmental isotope labeling, in which only a specific protein domain is enriched with NMR active nuclei. As a result, only a single domain is visible in the NMR spectrum, which facilitates the analysis of large multidomain proteins. For this purpose, the different protein domains are expressed separately and are finally linked together. For domain ligation different methods are available including expressed protein ligation (EPL), protein trans-splicing (PTS), and native chemical ligation (NCL). Moreover, enzyme-mediated ligation techniques have been developed.[7] Segmental isotope labeling is a useful tool for investigating interdomain interactions within multidomain proteins,

conformational changes and ligand binding. Furthermore, this technique facilitates the resonance assignment of large proteins and thereby enables protein structure determination.[46]

Selective protein labeling describes the isotope exchange of a specific atom or a specific subset of atoms within a defined amino acid. For that purpose, labeled metabolic precursors of the desired amino acid are synthesized from simple stable isotope sources and added to the growth medium of the protein-expressing organism (Figure 2). These compounds are then incorporated into the protein of interest within biosynthetic pathways *in vivo*.[9] Alternatively, selective labeling can also be performed using cell-free *in vitro* expression systems, by supplementing cell lysates with a labeled amino acid precursor.[47]

Many amino acids can be selectively labeled by applying the corresponding α-keto acid, which represents a direct metabolic precursor of the target residue. The use of α-keto acids circumvents the synthesis of labeled amino acids and the introduction of a stereocenter at the Cα position.[9, 48]

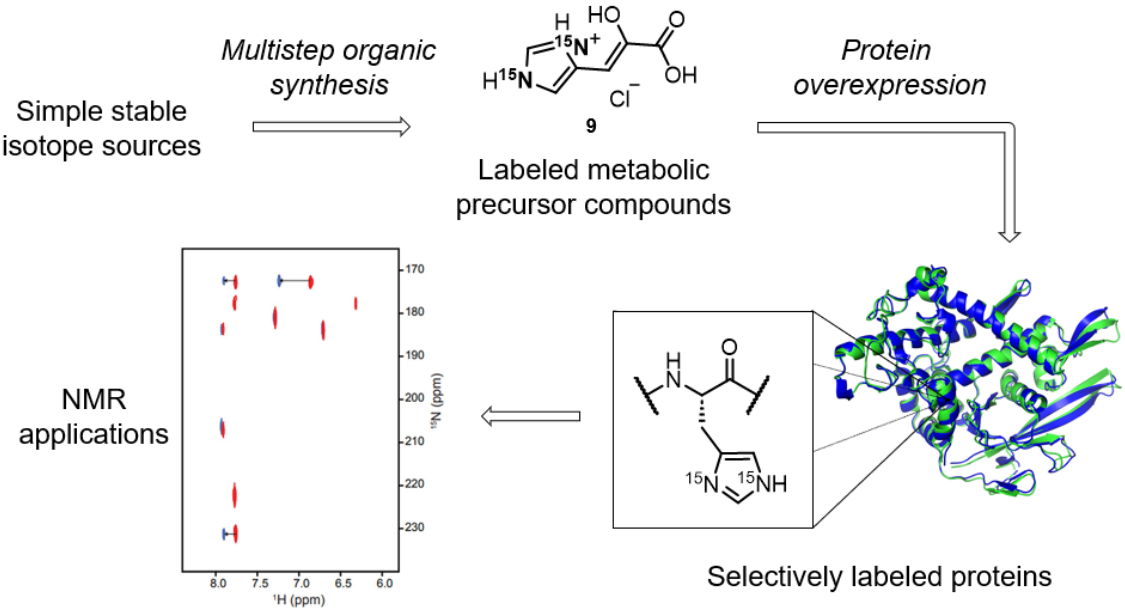


Figure 2: Scheme of the selective labeling process. The depicted NMR spectrum shows a ^{1}H-^{15}N SOFAST-HMQC NMR spectrum of histidine labeled KRAS4B G12D in the absence (red) and presence (blue) of a ligand.[49]

In the overexpressed protein, the native amino acid is substituted with the labeled amino acid. The NMR spectrum in figure 2 shows a ^{1}H-^{15}N SOFAST-HMQC NMR spectrum of KRAS4B G12D selectively histidine labeled (by applying precursor **9**) in the absence (red) and presence (blue) of a ligand.[49] Depending on the research question, which may concern protein structure, protein-ligand interactions, protein folding dynamics, or mechanisms of enzymatic catalysis, different labeling patterns can be employed.[3]

1.2.1. Aliphatic Amino Acid Labeling

The methyl groups of aliphatic side-chain residues including alanine, valine, leucine, and isoleucine are attractive labeling targets due to various reasons. Methyl groups frequently occur in hydrophobic cores of proteins, which makes them sensitive reporters of protein structure and dynamics, as well as protein folding mechanisms.[44, 50, 51] Furthermore, they are highly abundant at the active sites of enzymes and at protein interaction surfaces, and therefore, represent excellent probes for studying ligand binding or protein-protein interactions.[23, 44] Moreover, aliphatic CH groups act as hydrogen-bond donors in CH-π interactions, which are crucial determinants of affinity in ligand binding.[39] In addition, NOEs between methyl protons provide useful structural information by connecting residues that are far apart in the primary sequence.[45] Finally, methyl groups exhibit unique relaxation properties leading to intense and well separated resonances.[28] Participation in biomolecular interactions leads to a change in the chemical environment of the involved methyl groups resulting in chemical shift perturbations, which provide detailed qualitative and quantitative information about the interaction.[50] The introduction of ^{1}H,^{13}C-labeled methyl groups in an otherwise deuterated protein combined with relaxation-optimized techniques (such as methyl TROSY) enables the study of high molecular weight systems.[28, 44]

In *E. coli* overexpression systems, isoleucine can be selectively labeled by adding either α-ketobutyrate or (S)-2-aceto-2-hydroxybutanoate to the growth medium. These metabolic precursors are selectively transformed to isoleucine *in vivo*.[51]

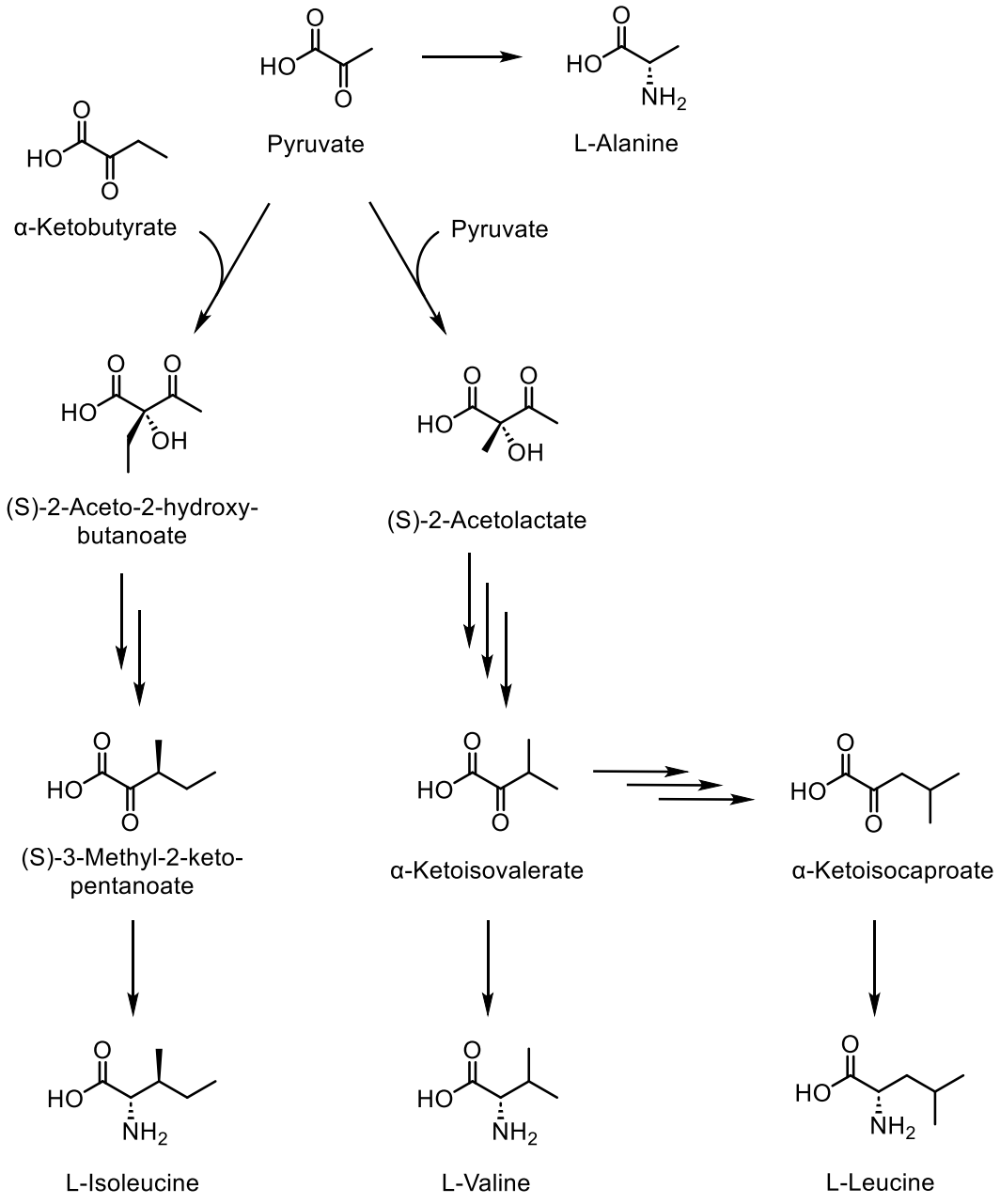
The biosynthetic pathways of leucine and valine, however, are interconnected (Scheme 1). Therefore these residues are simultaneously labeled by using either α-ketoisovalerate or (S)-2-acetolactate. Lichtenecker *et al.* introduced α-ketoisocaproate as selective precursor for L-leucine.[51] For selective valine labeling, α-ketoisovalerate can be applied with simultaneous addition of unlabeled α-ketoisocaproate to saturate the metabolic branch of leucine biosynthesis.[52]

α-Ketoisovalerate and α-ketoisocaproate are direct metabolic precursors of L-valine and L-leucine, respectively, and are converted into the target amino acid in a single amino-transferase catalyzed step. It was proven, that these precursors transform into the corresponding amino acid residues without observable metabolic scrambling.[51, 52]

Alanine labeling presents a challenge because its direct metabolic precursor, pyruvate, also feeds the biosynthesis of the branched-chain amino acids valine, leucine, and isoleucine. Since alanine is generated by the reversible transamination of pyruvate, simply supplying labeled alanine leads to isotope scrambling through these interconnected pathways. Nevertheless, alanine can be selectively labeled with minimal scrambling by adding the labeled

amino acid together with specific metabolic supplements that saturate the connected metabolic branches, thereby preventing isotope scrambling.[45]

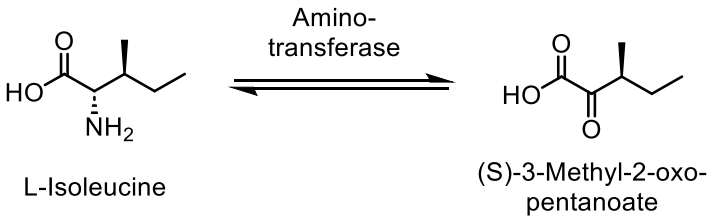
For aliphatic amino acid labeling using mammalian overexpression systems, another strategy is required. Since valine, leucine and isoleucine are essential amino acids in mammals, the amino acids themselves must be applied. Recent experiments show that transaminase activity in mammalian expression systems can often be exploited to enable the use of the direct α-ketoacid precursors.[2]



Scheme 1: Biosynthetic pathway of aliphatic amino acids.[53]

1.2.1.1. Isoleucine Labeling

In bacterial overexpression systems, α-ketobutyrate or acetohydroxybutyrate can be applied as additives for selective Cδ-methyl labeling of isoleucine residues. These compounds are direct metabolic precursors of isoleucine and are converted to the desired amino acid without detectable cross-labeling.[51] Mammalian cells, on the other hand, do not have a biosynthetic pathway for the essential amino acid isoleucine. Hence, a different strategy is required for selective isoleucine labeling of mammalian proteins. (S)-3-Methyl-2-oxopentanoate is the first intermediate of isoleucine degradation in mammalian cells and the substrate of the transaminase-catalyzed conversion to the corresponding target amino acid (Scheme 2). Since isoleucine transamination is a reversible reaction, an efficient transformation of this precursor into the desired residue could be an option for effective isoleucine isotope labeling.[2] Furthermore, this precursor allows the selective introduction of methyl labels at position Cγ2 or Cδ.

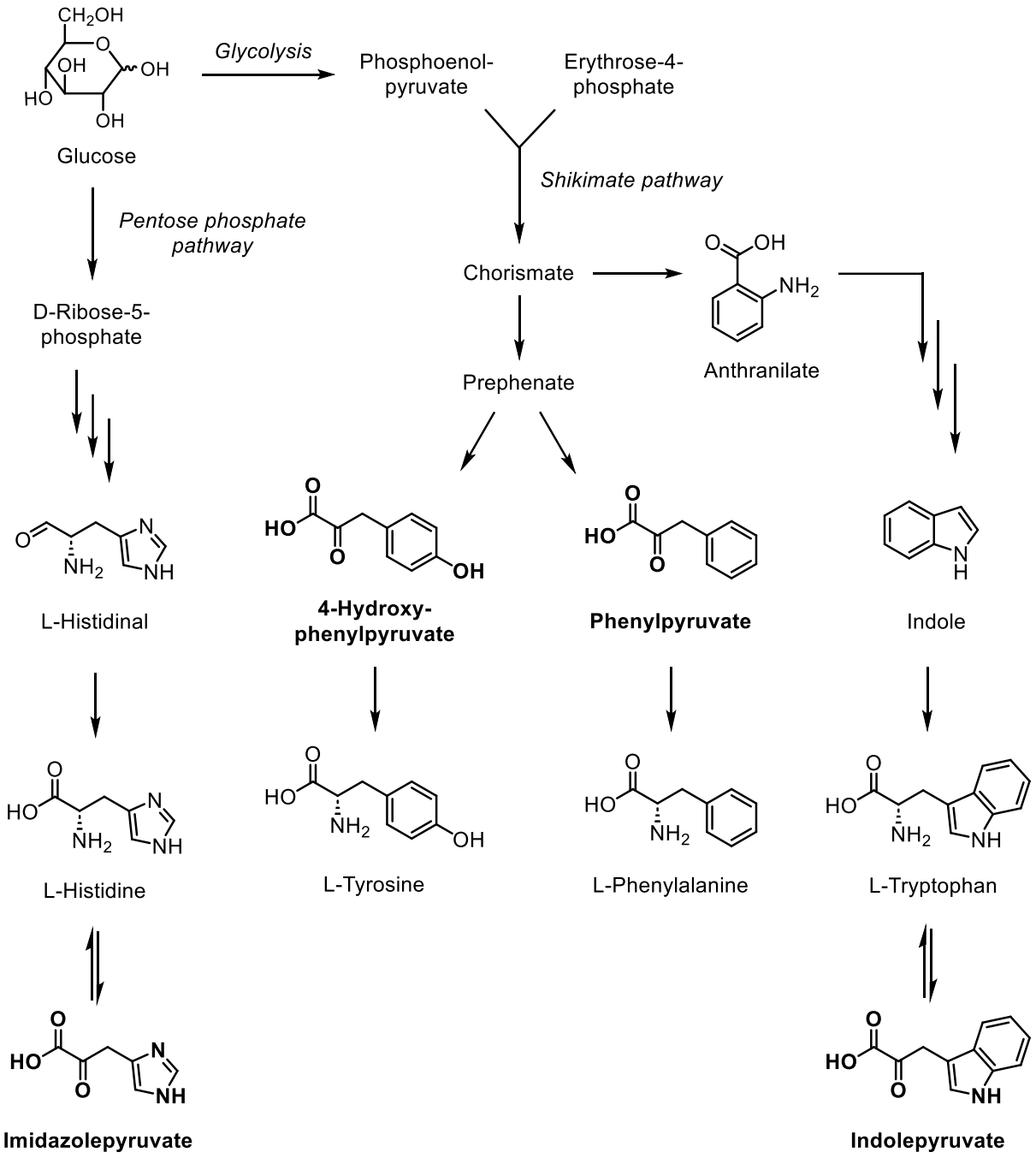


Scheme 2: The first step of isoleucine degradation is the reversible transamination to (S)-3-methyl-2-oxopentanoate.[53]

1.2.2. Aromatic Amino Acid Labeling

Aromatic amino acids, which include phenylalanine, tryptophan, tyrosine, and histidine, are interesting labeling targets because they are highly abundant at protein binding sites and frequently participate in catalytic action. All aromatic residues, and especially histidine with its different charging states in the imidazole side-chain, are valuable players in enzyme catalysis. Furthermore, these residues form a substantial part of the hydrophobic core of proteins and are directly involved in protein stabilization. In addition, they are responsible for specific dynamic processes, such as ring flips and His tautomerization. As a result, they are sensitive reporters of conformational dynamics as well as protein-ligand interactions.[9, 54]

In bacterial overexpression systems, aromatic amino acid residues can be selectively labeled with a defined isotope pattern by adding the corresponding labeled α-ketoacid to the growth medium. These compounds represent direct metabolic precursors of the target residues and are selectively and effectively converted into the corresponding amino acid without detectable isotope scrambling (Scheme 3).[6, 9]



Scheme 3: Metabolism of aromatic amino acids and possible precursors for selective labeling.[9, 55]

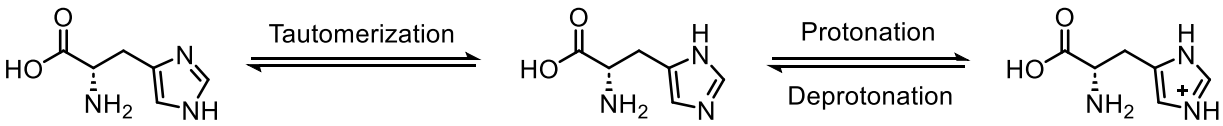
Phenylpyruvate and 4-hydroxyphenylpyruvate are the last intermediates of the biosynthetic pathways of L-phenylalanine and L-tyrosine, respectively, and are converted into the target amino acid in a single transaminase-catalyzed step. Importantly, these compounds are the only non-chiral molecules within their pathways, which makes them particularly convenient for isotopologue synthesis.[9]

In case of L-tryptophan labeling, the use of indolepyruvate, which is the first intermediate of tryptophan degradation, ensures selective incorporation of isotope labels. The reversible

nature of transaminase *EC* 2.6.1.27 allows efficient conversion of this compound into the target residue. In addition, anthranilate and indole can be used for selective tryptophan labeling.[9]

1.2.2.1. Histidine Labeling

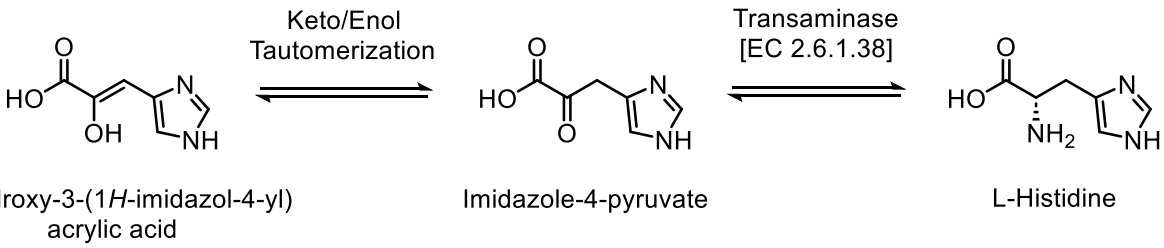
The amino acid histidine has several unique and functionally important properties that make this amino acid a useful target for isotope labeling. First of all, histidine is highly abundant at the active sites of enzymes, and frequently participates in catalytic action. In the catalytic triad, histidine can act as an acid or base catalyst.[49] Furthermore, the imidazole group can exist in a neutral or protonated form under physiological conditions (Scheme 4). The neutral imidazole ring can participate in cation-π interactions with a cationic side-chain such as Arg^+ or Lys^+. The positively charged imidazole ring, on the other hand, can form interactions with other aromatic residues, which are crucial for protein stability and facilitate protein-ligand interactions. In addition, the imidazole side-chain can engage in π-π stacking and hydrogen-π interactions.[49]



Scheme 4: The imidazole side-chain of histidine can exist in two neutral tautomeric forms and a protonated form. Aromaticity is maintained in every form.

Moreover, histidine can act as a nucleophile, as a proton shuttle, and as a hydrogen bond donor and acceptor. In addition, His can form chelate complexes with metal ions using its ring nitrogen atoms for coordination. Finally, histidine phosphorylation plays an important role in signal transduction pathways.[6]

Lichtenecker *et al.* introduced 2-hydroxy-3-(1*H*-imidazol-4-yl) acrylic acid as a precursor for selective histidine labeling in *E. coli* overexpression systems (Scheme 5). This compound is the enol tautomer to imidazole-4-pyruvate, which is the first intermediate of the minor histidine degradation pathway and substrate of histidine transaminase. The reversible nature of this enzyme enables effective incorporation without detectable isotope scrambling.[6]



*Scheme 5: 2-Hydroxy-3-(1*H*-imidazol-4-yl) acrylic acid represents a selective precursor for histidine labeling and is converted into the target amino acid in a single step.[49]*

Different synthesis routes for various ^{13}C- and ^{15}N-labeled isotopologues of this precursor have been developed, including reaction steps that have been optimized within the scope of this thesis. Commercially available compounds such as ^{15}N-ammonium chloride or ^{13}C-formaldehyde were used as isotope sources. To highlight possible applications, the synthesized precursors were used to selectively label pharmaceutically relevant targets, including WDR5, PLCγ1-SH2, and KRAS. It could be shown that the addition of 2-hydroxy-3-(1H-imidazol-4-yl) acrylic acid to *E. coli*-based expression media leads to selective and quantitative incorporation of labeled histidine residues. Furthermore, the residues involved in certain protein-ligand interactions could be identified by analyzing ligand-induced chemical shift perturbations.[49]

1.3. Production of Isotope Labeled Proteins

The production of isotope labeled proteins can be achieved either by *in vivo* or by cell-free protein synthesis.[56] In the former, also called cell-based methods, a suitable host organism is transformed with a plasmid carrying the gene for the protein of interest. The host organism is cultured in media containing a labeled metabolic precursor of the desired amino acid. This supplemented compound is then converted into the corresponding amino acid in biosynthetic pathways *in vivo* and finally incorporated into the protein. The protein can be fused with suitable solubility or affinity tags to facilitate protein purification or increase yield/solubility.[57]

E. coli is the most commonly used host organism in cell-based protein expression due to the ease of genetic manipulation, rapid growth, high expression yields, and cost effectiveness. However, *E. coli* based protein production faces limitations tied to dysfunctional folding of proteins, as well as absence or underdevelopment of cell organelles required for post-translational modification.[57, 58] The production of more complex proteins originating from higher organisms is performed with higher eukaryotic expression hosts like yeast, insect, or mammalian cells, which contain the more advanced cellular machinery for protein folding, post-translational modification, or membrane insertion.[58]

However, cell-based expression systems have various limitations, such as limited cellular uptake, elaborate purification of the produced protein, or toxic effects of synthesized proteins or protein complexes on the host cell.[59]

Some of these issues can be circumvented by switching to cell-free approaches. Cell-free protein synthesis (CFPS) is performed in cell lysates that mimic the natural cellular cytoplasmic environment while allowing facile addition of compounds.[60] These cell lysates can be produced from prokaryotic cells like *E. coli*, archaea, or eukaryotic cells including yeast, wheat germ or

insect cells. Each systems has its strengths and limitations, and the appropriate lysate is selected according to the protein's production requirements.[47]

Cell lysates are generated by lysis of suitable cells (mostly *E. coli*) followed by elimination of endogenous DNA and mRNA, cell walls, and most of the intracellular membranes.[47, 56] The complexity of the cellular machinery is reduced by this process and the final lysate only contains the essential protein expression apparatus including ribosomes and aminoacyl-tRNA synthetases. Essential substrates like amino acids, energy substrates, nucleoside triphosphates (NTPs), protease inhibitors, tRNAs as well as the DNA or mRNA templates for the protein of interest have to be supplied.[56]

The open and accessible nature of cell-free approaches enables active monitoring, rapid sampling, and direct manipulation of the production environment.[61] Nutrients, bio-active compounds like ligands, substrates, or inhibitors as well as stabilizers, solubilization agents and scrambling inhibitors can be added to the reaction at any time to optimize reaction conditions and improve folding and stability of the protein.[56, 60] Protein labeling is achieved by adding the isotope labeled amino acids to the cell lysates. Importantly, the labeled amino acids will be exclusively incorporated into the protein of interest, and isotope scrambling is significantly reduced.[56] Furthermore, this cell-free methods allow the incorporation of non-canonical amino acids, and enables the production of toxic proteins, complex integral membrane proteins and protein assemblies.[56, 59] Moreover, cell-free protein synthesis allows the production of a single protein product by optimizing the cell extract.[61]

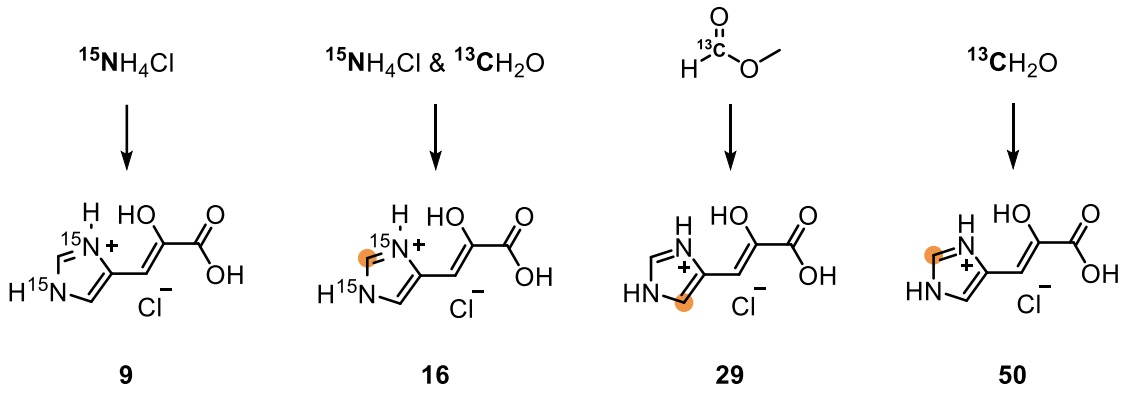
Identifying the optimal production conditions for a certain protein resulting in its correct and stable folding is quite challenging.[56] Selecting a suitable cell lysate is crucial for the successful production of functional proteins. Protein production using *E. coli* extracts promises high protein yield, simple cultivation, fast cell growth and lysate preparation, as well as simple genetic engineering. However, this method is hindered by limited post-translational modifications and the absence of membrane structures for the synthesis of integral membrane proteins. While yeast extracts enable certain post-translational modifications, they generally suffer from low protein yields and cannot reproduce mammalian-type modifications. Tobacco-, insect-, and mammalian-based cell-free systems, on the other hand, contain endogenous microsomal structures facilitating protein folding and post-translational modifications. Therefore, these eukaryotic systems show an improved capacity of producing functionally active proteins. [47]

2. Results and Discussion

The aim of this project is the development and optimization of economic and efficient multistep synthetic routes for ^{13}C- and/or ^{15}N- labeled isotopologues of histidine and isoleucine precursors for protein and peptide NMR investigation.

2.1. Synthesis of Selective Histidine Precursors

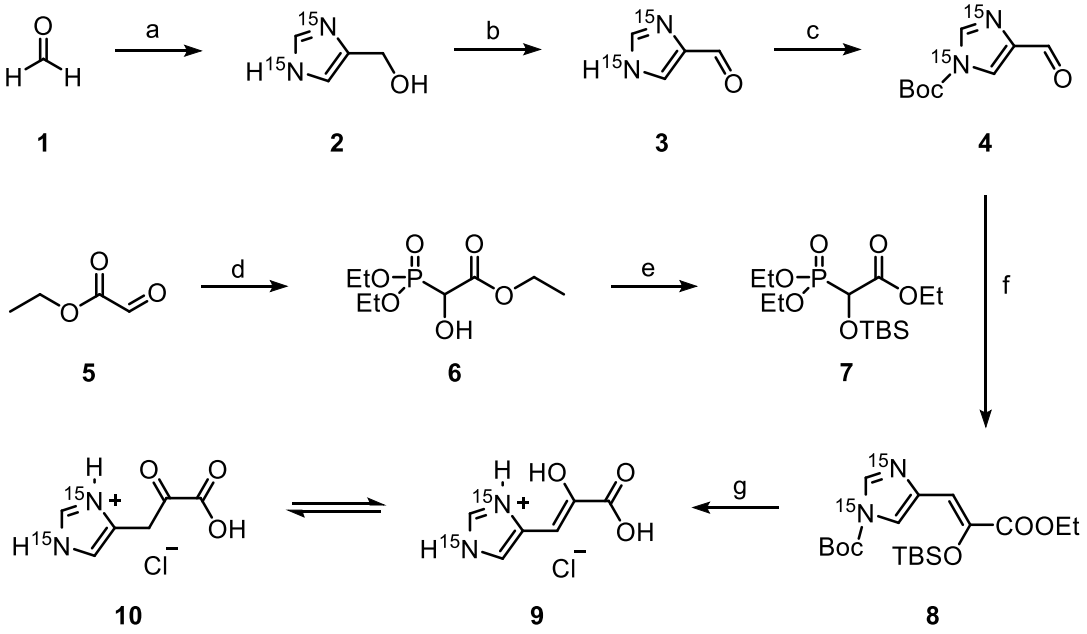
The histidine precursor 2-hydroxy-3-(1*H*-imidazol-4-yl)acrylic acid hydrochloride salt, which is converted to the target amino acid histidine in a transaminase catalyzed reaction *in vivo*, has been proven to be a suitable candidate for selective histidine labeling of proteins.[6, 49] Isotope labels can be introduced in various atomic positions: ^{15}N isotopes can be incorporated in place of either one or both nitrogen atoms, and ^{13}C-labels can be introduced at either Cβ, Cγ, Cδ2 or Cε1. These isotope labels can be combined with each other, resulting in a variety of possible isotopologues of the shown histidine precursor. Depending on the desired isotope pattern, different synthesis routes must be employed. Scheme 6 shows possible isotopologues of the 2-hydroxy-3-(1*H*-imidazol-4-yl)acrylic acid hydrochloride salt as well as the heavy isotope sources used in their synthesis. Precursor **9** and **16** were synthesized in the course of this master thesis. Furthermore, the synthesis of precursor **27**, which was developed by Julia Schörghuber in her PhD thesis, was further optimized.[62]



Scheme 6: Different isotopologues of the selective histidine precursor 2-hydroxy-3-(1H-imidazol-4-yl)acrylic acid hydrochloride salt and the respective heavy isotope sources used in their synthesis. The colored dots indicate ^{13}C labeling.

The following reaction scheme (Scheme 7) shows the synthesis of the ^{15}N-labeled histidine precursor **9**, which is based on the procedure for the ^{13}C-ε1-labeled isotopologue **50** published by Lichtenecker *et al.* in 2017.[6] In the case of the ^{13}C-ε1-labeled precursor, NH_3 was used as nitrogen source in the first step of the synthesis – the formation of the imidazole ring. However, for the synthesis of ^{15}N-labeled precursor **9**, $^{15}NH_4Cl$ was used as heavy isotope source, since ^{15}N-labelled ammonia is quite expensive and ^{15}N-ammonium chloride presents a more economic and convenient source or nitrogen-15. Nonetheless, since $^{15}NH_3$ is formed in situ

from $^{15}NH_4Cl$ and NaOH, several equivalents of the ^{15}N-source are required to achieve sufficient yields.



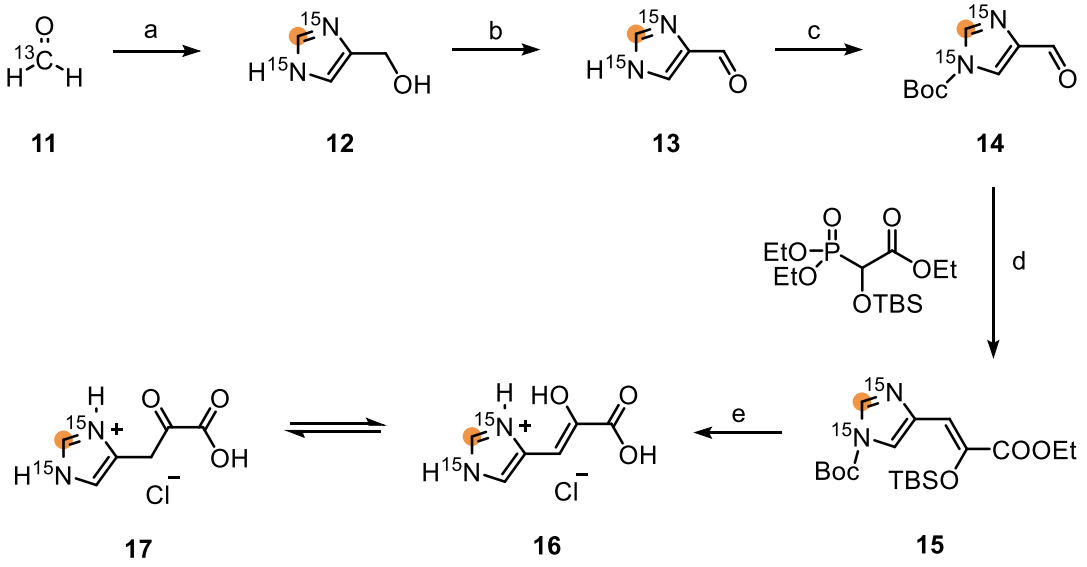
*Scheme 7: Synthesis of precursor **9**. a) (i) $^{15}NH_4Cl$, NaOH, $CuCO_3$•$Cu(OH)_2$, dihydroxyacetone dimer, 100 °C, 4 h; (ii) thioacetamide, 50 °C, 2 h, 53%; b) MnO_2, 80 °C, 3 h, 87%; c) Boc_2O, DMAP cat., rt, overnight, 66%; d) triethyl phosphonoacetate, p-toluenesulfonic acid, 135 °C, overnight, 56%; e) TBSCl, imidazole, rt, overnight, 89%; f) LiHMDS, -78 °C to rt, overnight, 49%; g) 6M HCl, rt, overnight, 93%.*

The first step of the synthesis was the assembly of ^{15}N-labeled (hydroxymethyl)imidazole (**2**) using formaldehyde (**1**), $^{15}NH_4Cl$, NaOH, and dihydroxyacetone dimer as reagents. Thereby, ^{15}N-labeled ammonia was formed *in situ* from $^{15}NH_4Cl$ and NaOH. The presence of Cu^{II} ions enabled the formation of the imidazole ring and subsequent treatment with thioacetamide gave compound **2** in 53% yield. In earlier attempts, synthesis of compound **2** was carried out as a one-pot reaction in the microwave at 100 °C giving compound **2** in 33% yield. By performing the reaction in the fume hood and allowing the generation of NH_3 to complete before introducing the other reagents, the yield could be increased to 53%.

The oxidation of the hydroxy group to acquire aldehyde **3** was achieved by treatment with MnO_2 in 87% yield. The following DMAP-catalyzed Boc-protection of the ring nitrogen gave compound **4** in 66% yield. This step suffered from poor reproducibility, with yields varying from 20 to 80%, likely due to the moisture sensitivity of the reagent Boc_2O and the moderate solubility of the substrate **3** in the used solvents DMF and acetonitrile.

Compound **8** was prepared in a Horner-Wadsworth-Emmons reaction under argon atmosphere using LiHMDS as base and was acquired in 49% yield. This step showed inconsistent results as well, with yields ranging from 20% to 80%, which might result from variability in the quality of the LiHMDS batches used. Finally, acidic deprotection and simultaneous ester hydrolysis

gave the desired precursor **9** in an overall yield of 16%. Keto-enol tautomerization leads to the corresponding α-ketoacid **10** .



*Scheme 8: Synthesis of precursor **16**. The synthesis was performed as described for precursor **9**, with the exception that ^{13}C-formaldehyde **11** was used as substrate in the first step. a) 40% in fume hood (**12a**), 37% in microwave (**12b**); b) 78%; c) 80%; d) 38%; e) 93%. Overall yield 13%.*

Scheme 8 shows the synthesis route for precursor **16**, which contains an additional ^{13}C label in between the two ^{15}N atoms. The synthesis was carried out as described for precursor **9**, with the exception that ^{13}C labeled formaldehyde was used as the substrate in the first step.

The incorporation rate of ^{13}C in step *a* was not quantitively but varied with reaction conditions. It was found that ^{13}C incorporation can be increased from 54% to 68-74% when performing the initial step, the imidazole formation, in closed vials under microwave irradiation. The ^{13}C content of compound **12** was analyzed by NMR signal integration (Figure 3).

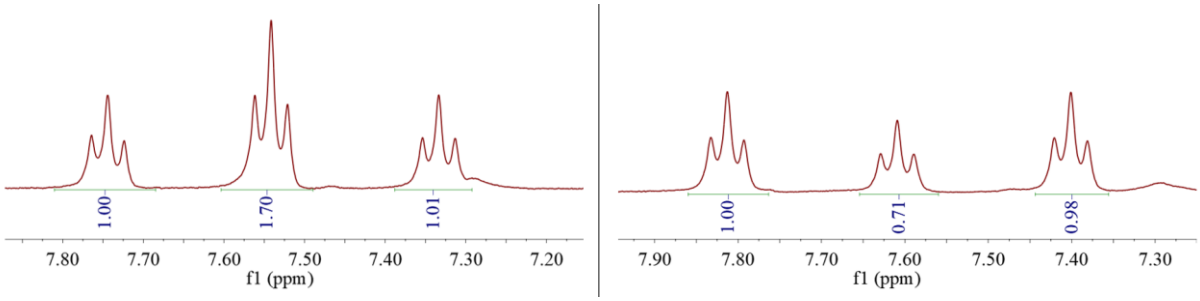


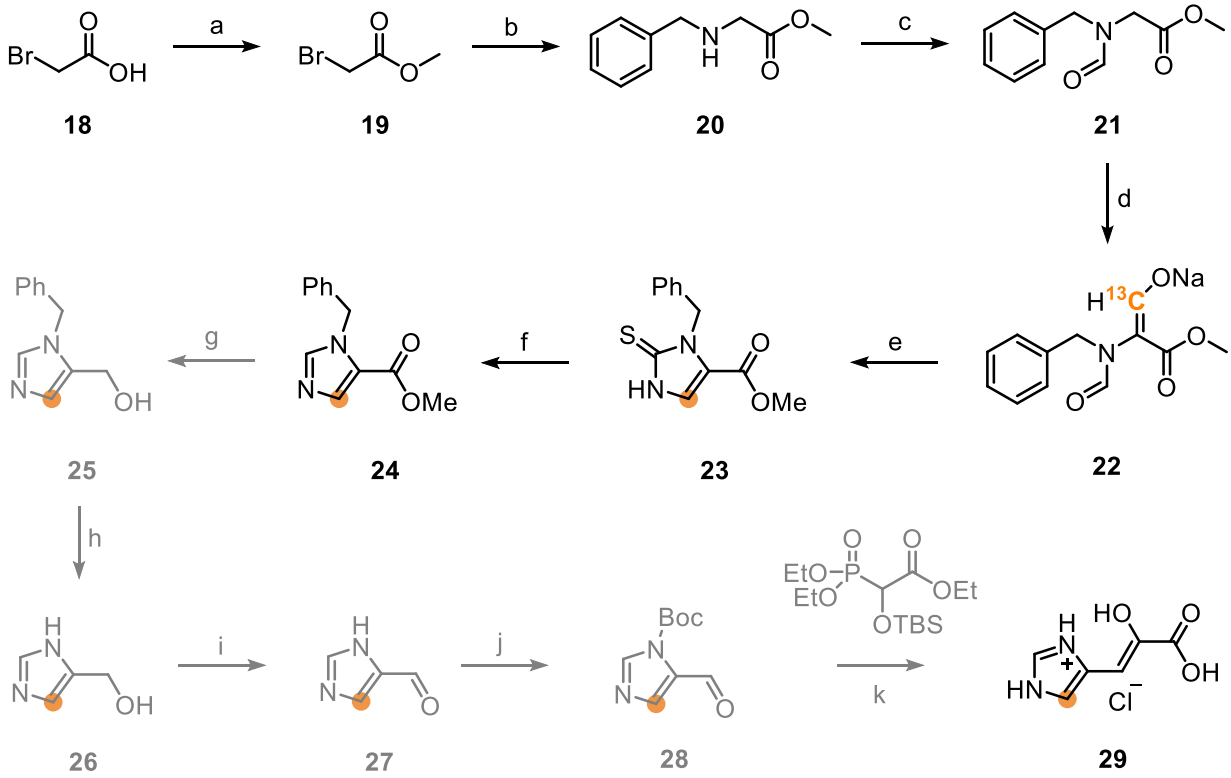
*Figure 3: ^{1}H-NMR peaks of the Cε1- and ^{13}Cε1-bound protons of ^{13}C-^{15}N-labeled compound **12a** (left) and **12b** (right). Compound **12b**, which was synthesized using a microwave-assisted method, exhibits 74% ^{13}C incorporation, whereas compound **12a** shows 54% ^{13}C incorporation.*

Figure 3 shows the ^{1}H-NMR signals corresponding to the ^{12}Cε1- and ^{13}Cε1-bound protons of ^{13}C,^{15}N-labeled compound **12**. The signal in the middle was assigned to the proton bound to

^{12}Cε1 of the unlabeled product fraction. The two outer peaks correspond to the ^{13}Cε1-bound proton. The splitting into these two signals is caused by one-bond ^{1}H-^{13}C coupling.

Compound **12a**, which was synthesized under standard conditions in the fume hood, exhibits a ^{13}C incorporation of 54%. By performing the reaction in the microwave, the heavy isotope percentage could be increased to 74%. The mechanism for the formation of the ^{12}C containing product fraction still requires further investigation.

Scheme 9 shows the synthesis route for the ^{13}Cδ2 labeled precursor **29**, whereby ^{13}C is introduced in step *d* by using methyl formate-1-^{13}C as heavy isotope source. Furthermore, this synthesis pathway presents the possibility to introduce a ^{13}C label at position Cε1 by using KS^{13}CN in step *e*. In addition, the use of 2-bromoacetic-1-^{13}C acid or 2-bromoacetic-2-^{13}C acid yields the ^{13}Cβ- or the ^{13}Cγ-labeled precursor, respectively.[62]



*Scheme 9: Synthesis of precursor **29**. a) (i) oxalyl chloride, 60 °C, 3 h; (ii) MeOH, rt, 30 min, 88%; b) triethylamine, benzylamine, rt, overnight, 59%; c) formic acid, acetic anhydride, 100 °C, 1 h, 89%; d) methyl formate-1-^{13}C, sodium methoxide, rt, 2 h; e) HCl, KSCN, 80 °C, 2 h, 28% over 2 steps; f) $NaNO_2$, HNO_3, rt, 2 h, 98%. Steps g-k have not been performed in the course of this thesis, whereby the steps i-k are carried out as described for compound **9**.*

Moreover, this route can be employed to selectively introduce ^{15}N on either Nδ1 or Nε2 using either benzylamine-^{15}N (step *b*) or KSC^{15}N (step *e*) as isotope sources, respectively. In the course of this thesis steps *a* to *f* were conducted towards the synthesis of ^{13}Cδ2 labeled compound **29**.

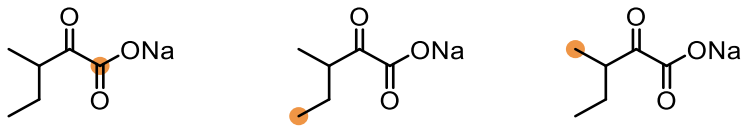
The synthesis started with the esterification of bromoacetic acid **18** giving compound **19** in 88% yield followed a nucleophilic substitution reaction with benzylamine to obtain compound **20** in 59% yield. Subsequent N-formylation (step *c*) was conducted using acetic formic anhydride as formylating agent, which was prepared *in situ* from formic acid and acetic anhydride, and yielded compound **21** in 89%.

In the following C-formylation, NaOMe was used for deprotonation of the substrate, which immediately reacted with methyl formate in a Claisen condensation to form compound **22**. This step presents an opportunity to introduce a ^{13}C label by using methyl formate-1-^{13}C as heavy isotope source. The subsequent formation of thioimidazole was achieved by treatment with KSCN under acidic conditions giving compound **23** in 28% yield (over 2 steps). This reaction step would also give access to $^{15}N\varepsilon 2$ or $^{13}C\varepsilon 1$ when employing ^{15}N- or ^{13}C-labeled KSCN.

Oxidative desulfurization was achieved by using aqueous $NaNO_2$ and HNO_3 and gave compound **24** in 98% yield. Initially, crude compound **23** was used in this step without previous purification as described by Schörghuber.[62] However, using this procedure in step *f* proved problematic, as the reaction showed either incomplete conversion of the substrate or no conversion at all. By introducing an additional purification step in the synthesis of **23** and by using the purified substrate in the desulfurization, the yields improved significantly. This suggests that the initial problem may have been caused by impurities, that interfered with the reaction.

2.2. Synthesis of Selective Isoleucine Precursors

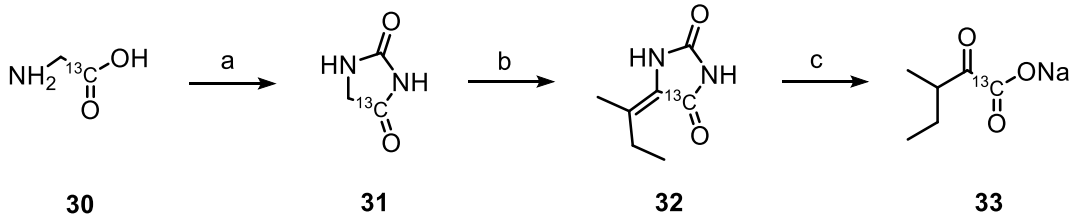
As already described, α-ketobutyrate represents a selective precursor for isoleucine labeling using *E. coli* overexpression systems. Isoleucine labeling of proteins using mammalian cells, however, demands a different strategy since mammalian cells do not have a biosynthetic route for the essential amino acid isoleucine.[63] (S)-3-methyl-2-oxopentanoate is the corresponding α-ketoacid of isoleucine and the product of the first step of the isoleucine degradation pathway. Since isoleucine transamination is a reversible reaction, efficient conversion of this precursor into the desired residue could provide an effective route for isoleucine isotope labeling, raising the key question of whether mammalian cells possess sufficient transaminase activity to catalyze this conversion efficiently *in vivo*. Moreover, this precursor enables labeling of both the Cγ2 and Cδ methyl group, while α-ketobutyrate only allows the introduction of an isotope label at the Cδ position. However, the synthesis yields a racemic mixture of 3-methyl-2-oxopentanoate, and the transaminase converts only the (S)-enantiomer into isoleucine.



Scheme 10: Possible isotopologues of the isoleucine precursor sodium 3-methyl-2-oxopentanoate. The colored dots indicate ^{13}C labeling.

Sodium 3-methyl-2-oxopentanoate-1-^{13}C **33**, which carries the ^{13}C label at position C1, was successfully synthesized in the course of this project. Furthermore, it was shown that this precursor is effectively incorporated into proteins synthesized using mammalian overexpression systems. Furthermore, efforts were made to develop a synthetic route towards precursor **51** carrying the isotope label on position Cδ.

The following reaction scheme (Scheme 11) shows the synthesis of the isoleucine precursor **33** carrying the ^{13}C-label at the carbonyl carbon. The first step of the synthesis was the formation of hydantoin using ^{13}C-labeled glycine, KOCN and HCl as reagents, and gave compound **31** in 53% yield. The subsequent condensation of hydantoin and 2-butanone using ethanolamine as base gave compound **32** in 61% yield. This step could be improved from 20% to 84% by performing the reaction in the microwave and by using 2-butanone as solvent instead of water/THF. However, when performing the reaction with labeled substrates a yield of only 61% was achieved.



*Scheme 11: Synthesis of precursor **33**. a) (i) KOCN, 100 °C, 3 h; (ii) HCl conc., 100 °C, 3 h, 57%; b) 2-butanone, ethanolamine, microwave irradiation, 100° C, 5 h, 61%; c) NaOH 20%, 100 °C, 5 h, argon stream, 87%.*

Different solvent mixtures consisting of THF/water 1:1 to 10:1 were tested to determine the reactions tolerance for water. As shown in table 1 it was found, that the presence of water hinders the condensation of hydantoin and 2-butanone. The less water is present in the reaction mixture, the higher the yields. By using 100% THF as solvent, a yield of 50% could be achieved, whereas using water as solvent resulted in only 20%. However, the highest yields of 84% were achieved by using the reagent 2-butanone as solvent.

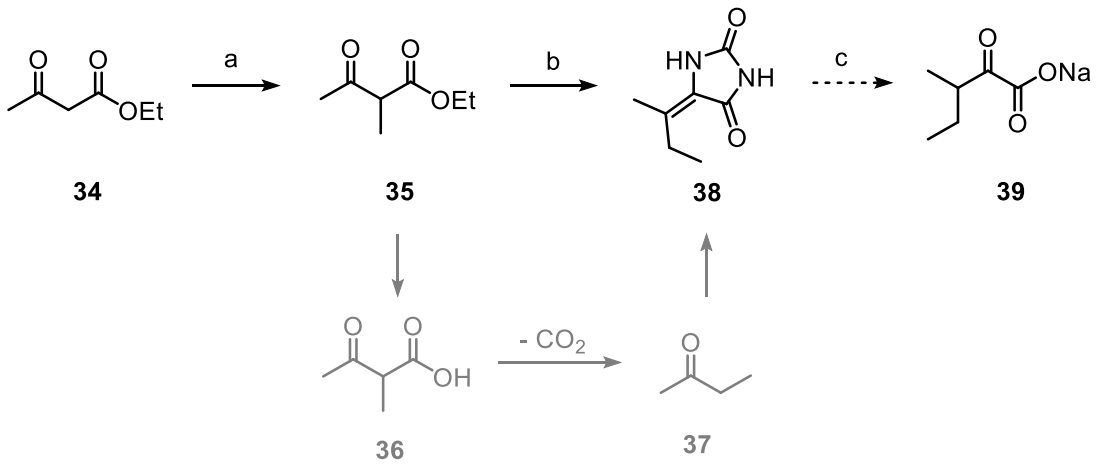
*Table 1: Effects of the solvent composition on the efficiency of the hydantoin condensation yielding compound **32**.*

Solvent mixture	Yield [%]
Water	20
THF/water 1:1	27
THF/water 1:5	31
THF/water 1:10	32
THF	48
2-Butanone	84

Each reaction was conducted using the following conditions: A microwave vial was loaded with 2-butanone (5.56 mmol, 400 mg, 1.0 eq), hydantoin (7.21 mmol, 722 mg, 1.3 eq), ethanolamine (8.32 mmol, 508 mg, 1.5 eq) and 2 mL of solvent. A microwave reaction was performed at 100 °C for 6 h. Afterwards, the solvent was evaporated under vacuum, the residue was taken up in a small amount of water and acidified with 1M HCl (pH 3). The mixture was stored at 4 °C overnight, the precipitated product was filtered off and dried under vacuum.

Finally, the cleavage of compound **32** using 20% aqueous NaOH under argon stream gave sodium 3-methyl-2-oxopentanoate-1-^{13}C **33** in an overall yield of 30%. This precursor can for example be applied for isoleucine backbone labeling in both, *E. coli* and mammalian cell-based overexpression.

If the ^{13}C label is to be introduced in the aliphatic side-chain, different synthesis routes must be employed. The following approach (Scheme 12) presents one of several possibilities for the introduction of ^{13}C at position Cδ and was tested for feasibility and efficiency.



*Scheme 12: Attempts to synthesize precursor **39**. This synthesis was performed with unlabeled substrates only. a) K_2CO_3, CH_3I, 40 °C, overnight, 77%; b) (i) NaOH 10%, microwave irradiation, 100 °C for 2 h; (ii) hydantoin, ethanolamine, microwave irradiation, 100 °C for 5 h, 16%; c) NaOH 20%, 100 °C, 5 h, argon stream. The final step c was not performed.*

Methylation of ethyl acetoacetate **34** was achieved by treatment with CH_3I in the presence of K_2CO_3 giving ester **35** in 77% yield. The following three-step process started with the hydrolysis of **35** under basic conditions to form the corresponding acid **36**, which is directly

decarboxylated to give 2-butanone **37**. Subsequently, 2-butanone undergoes a condensation reaction with hydantoin to give compound **38** in 16% yield. Finally, **38** is cleaved under basic conditions as described for **33** to give the α-ketoacid **39**. If ^{13}C-enriched iodomethane is used in step *a*, the δ-^{13}C-labeled isoleucine precursor sodium 3-methyl-2-oxopentanoate-5-^{13}C is obtained.

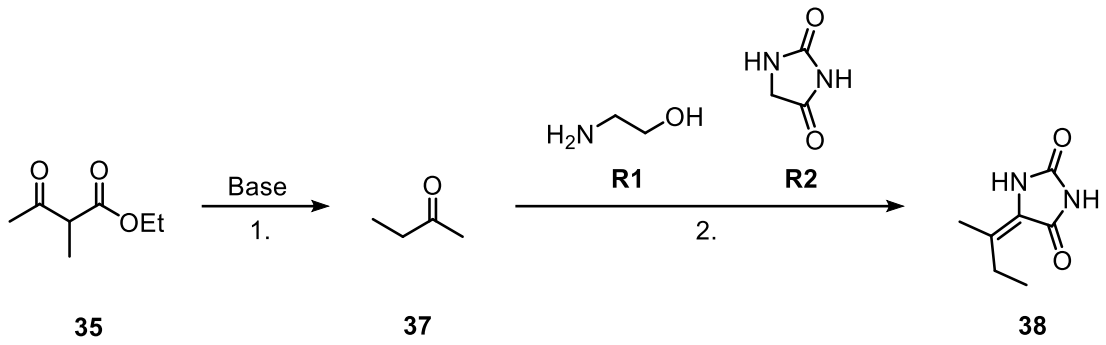
A drawback of this route are the poor yields in step *b*, which could be optimized to only 16%. However, this limitation can be compensated by the time efficiency of this synthesis and the fact that it requires only three reaction steps, and that both the first and the last step achieve yields over 75%.

Efforts were made to optimize step *b* by varying the base used for hydrolysis and decarboxylation, whereby either 10% NaOH or ethanolamine **R1** was employed, the amounts of the reagents ethanolamine **R1** and hydantoin **R2**, the solvent composition as well as the reaction time and temperature.

There are several factors, that could be responsible for the inefficiency of step *b*. While the first reaction of this process, the hydrolysis of the β-keto ester **35**, proceeds completely under basic conditions, the efficiency of the decarboxylation is hard to estimate, since it is difficult to monitor. Since the reaction is performed in a closed microwave vial, it is possible, that the formed CO_2 cannot escape, leading to a build-up of pressure and the formation of an equilibrium, which keeps the decarboxylation reaction from proceeding completely.

Performing the reaction as an open system does not solve the problem either because temperatures have to be kept below 68 °C to prevent evaporation of 2-butanone. However, while hydrolysis can be performed at this temperature, higher temperatures are needed for a successful decarboxylation of the β-keto ester.

The last step of this three-step process, the condensation of the formed 2-butanone with hydantoin, might be the success-limiting reaction, since the presence of water has proven to lead to low yields below 30% (see Table 1). However, using water as solvent cannot be avoided either since water is essential for hydrolysis and D_2O is required for deuteration. Another yield-limiting factor may be the aldol self-condensation of the generated 2-butanone, as the basic conditions and high temperatures used in the reaction are known to promote aldol condensation.



Scheme 13: Reaction scheme of step b. The optimization of this step was attempted by varying reagents, solvent compositions, reaction temperatures and reaction times.

Table 2 summarizes the conditions tested towards the optimization of step *b*. Experiment 1 was conducted to investigate whether ethanolamine can be used as base for hydrolysis. Although TLC indicated complete hydrolysis, the reaction afforded only a 9.5% yield.

In experiment 2, ethanolamine was added as base and the temperature of step 1 was increased to 140 °C. THF served as solvent and 4 equiv. of water were added for hydrolysis. Despite 4 h at 140 °C, conversion of ester **35** remained incomplete, indicating that the amount of water was insufficient.

*Table 2: Conditions for the optimization of the conversion of ester **35** to hydantoin **38**.*

	Reagents	Conditions	Solvent	Yield [%]
1	1. R1 (2.5 eq) 2. R1 (1.2 eq), R2 (1.2 eq)	1. 100 °C, 2 h 2. 100 °C, 5 h	Water	9.5
2	1. R1 (3.0 eq), H_2O (4.0 eq) 2. R1 (1.5), R2 (eq)	1. 140 °C, 4 h 2. 100 °C, 6 h	THF	9.0
3	1. NaOH (1.2 eq) 2. R1 (1.4 eq), R2 (1.2 eq)	1. 100 °C, 2 h 2. 100 °C, 5 h	Water	**16**
4	1. NaOH (1.0 eq) 2. R1 (1.4 eq), R2 (1.2 eq)	1. 120 °C, 4h 2. 100 °C, 4h	Water	6.0
5	1. NaOH (1.0 eq) 2. R1 (1.5 eq), R2 (1.3 eq)	1. 120 °C, 4 h 2. 100 °C, 4 h	1. Water 2. Water/THF 1:2	7.0
6	1. NaOH (1.0 eq) 2. R1 (1.5), R2 (1.2 eq)	1. 75 °C, 3 h, hood 2. 80 °C, overnight	Water	< 5.0

*Each experiment was conducted using the following procedure: A microwave vial was loaded with ester **35**, solvent (1 mL) and base. Hydrolysis and decarboxylation were performed under microwave irradiation (unless noted otherwise). Subsequently, reagents R1 and R2 were added and hydantoin condensation was carried out under microwave irradiation. Temperature and reaction times were varied.*

Experiment 3 employed 10% NaOH as base in step 1 with water as solvent. The temperature was kept at 100 °C throughout the process. Under these conditions, the yield improved to 16%, representing the best result obtained.

Experiment 4 investigated whether a higher temperature (120 °C) and longer reaction time (4 h) in step 1 would improve the outcome. In a similar experiment (5), THF was added in step 2. Although TLC indicated complete conversion of ester **35**, the overall yields decreased to 6 and 7%, respectively, suggesting that the tested conditions may not be suitable. The reaction temperature of 120 °C might promote aldol condensation of the formed 2-butanone.

Experiment 6 was performed in an open system in the fume hood to allow the formed CO_2 to escape. Although TLC indicated complete hydrolysis of ester **35** after 3 h, the overall yield was <5%. This low yield may be because the temperature was high enough for 2-butanone to evaporate, or too low to promote decarboxylation and the following condensation.

In summary, none of the tested conditions led to satisfying yields and the reaction requires further investigation and optimization. Alternatively, different, more promising synthesis routes could be explored.

Furthermore, the feasibility of the deuteration of compound **38** was investigated by performing the reaction in D_2O. Figure 17*a* shows the NMR spectrum of unlabeled compound **38**. This compound exists as E/Z isomers, which is reflected by two populations of signals in the NMR spectrum. The protons of methyl group 1 correspond to the doublet of triplets on the right side of the spectrum. The protons of the second methyl group (3) give rise to two singlets, and the CH_2 protons (2) correspond to the two quartets.

Spectrum *b* shows the NMR signal of deuterated compound **40**. The signals of the CH_3 and CH_2 protons at positions 3 and 2, respectively, have decreased significantly suggesting that they have been replaced by deuterons by approximately 60-75%. Methyl group 1 is not deuterated under the used conditions and is still visible in the spectrum. Instead of a triplet, however, a singlet can be observed as the neighboring protons have been exchanged for deuterons. Since deuteration was incomplete, a second peak is visible due to residual coupling.

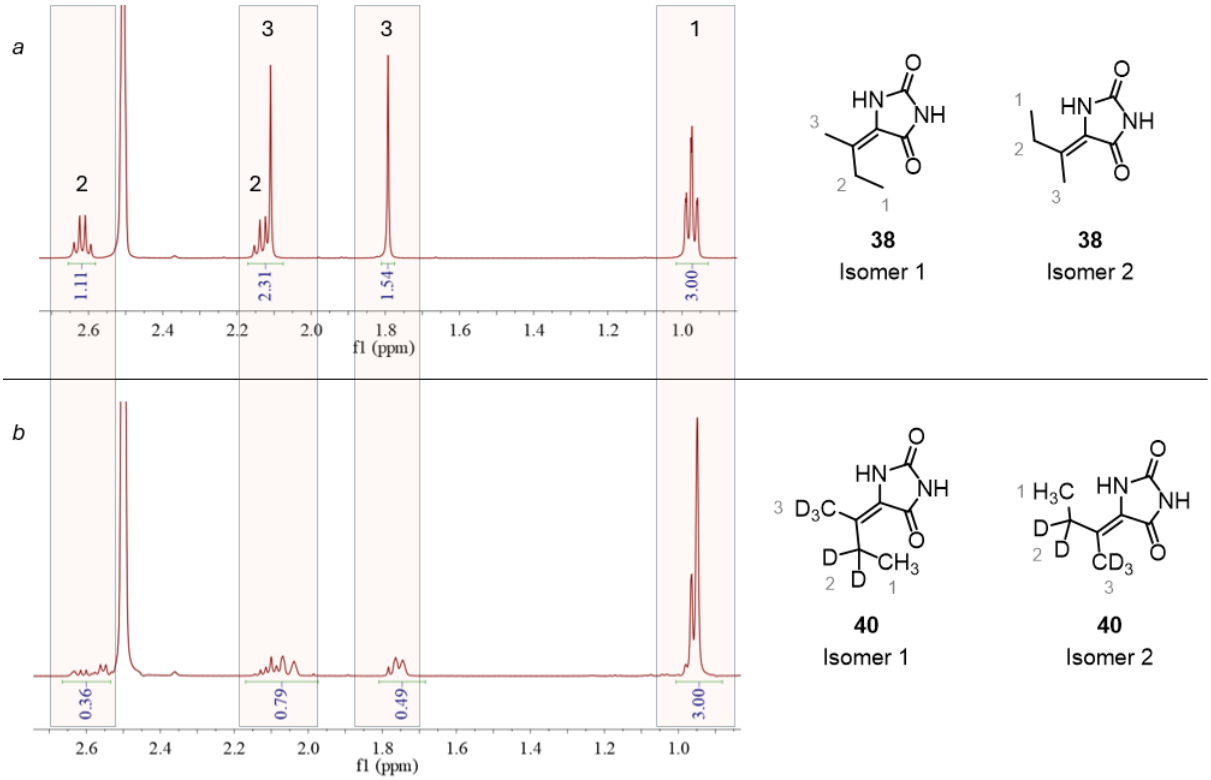


*Figure 4: 1H-NMR spectra of compound **38** (a) and deuterated compound **40** (b).*

In summary, precursor **33** was synthesized with an overall yield of 41%, whereby the synthesis of compound **32** could be improved from 20% to 84% by performing the reaction in the microwave and by using 2-butanone as solvent instead of water/THF. Synthesis of precursor **39** was attempted only with unlabeled substrates in a 3-step synthesis. Ester hydrolysis of **35**, subsequent decarboxylation and the condensation reaction leading to compound **38** were combined in a single step, which could be optimized to 16%.

3. Conclusion and Outlook

In conclusion, the histidine precursors **9** and **16** were successfully synthesized with an overall yield of 16% and 13%, respectively. In the case of the ^{13}Cε1,$^{15}N_2$-labeled precursor **16**, microwave irradiation in the first synthetic step increased ^{13}C incorporation from 54% to 68-74%. Furthermore, the synthesis of the ^{13}Cδ2-labeled histidine precursor **29** could be optimized. The oxidative desulfurization leading to compound **24** initially showed incomplete or absent conversion. However, the yield was improved to 98% by purifying the substrate (**23**) instead of using the crude compound. The histidine precursors **9** and **16** will be incorporated into proteins using *E. coli* expression systems and thereby, used for the investigation of protein-protein interactions or protein-ligand binding by NMR.

Isoleucine precursor **33** was synthesized with an overall yield of 41%, whereby the synthesis of compound **32** could be improved from 20% to 84% by optimizing the reaction conditions. Precursor **33** was tested in mammalian cell-based expression systems and was effectively incorporated into proteins upon transaminase-catalyzed conversion into the target amino acid.

Moreover, a new synthetic route toward precursor **39** was explored. This synthesis would allow the introduction of a ^{13}C label at the δ-position and has not been previously reported. Ester hydrolysis of **35**, subsequent decarboxylation and the condensation reaction leading to compound **38** were combined in a single step, which could be optimized to 16%. However, further investigation is required before this route can be successfully applied.

4. Experimental Part

4.1. General Remarks

Synthetic Methods: Oxygen- and moisture sensitive reactions were carried out under argon atmosphere and in dry solvents. Sensitive liquids and dry solvents were transferred by needle and syringe through septa. All reactions were stirred magnetically, unless otherwise noted. All synthesis steps were developed and optimized by using non-isotope labeled reagents.

Solvent and Chemical Purification: All reagents were purchased from commercial suppliers and were utilized as obtained without further purification. Dry solvents were stored under argon atmosphere over molecular sieves (4Å).

Thin-Layer Chromatography: Reaction progress was monitored via thin layer chromatography (TLC) using pre-coated TLC sheets ALUGRAM® Xtra SIL G/UV254. Compounds were visualized by UV-light irradiation (wavelength = 254 nm) and/or by applying basic $KMnO_4$ solution, or ninhydrin solution in ethanol as dying reagents and subsequent heating using a heat gun.

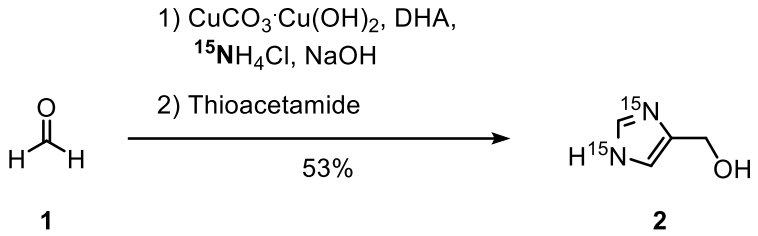
Flash Column Chromatography: Preparative Flash Column Chromatography was performed manually using silica gel 60 from Merck (particle size 40-63 µm) as the stationary phase. Details about the used eluents can be found in the experimental procedures.

Lyophilization: Lyophilization was conducted using a Christ Alpha 1-2 LDplus freeze dryer.

Nuclear Magnetic Resonance Spectroscopy: NMR spectra were recorded on a Bruker Avance NEO 500 MHz, Bruker Avance III 400 MHz, Avance III 600 MHz or Avance III HD 700 MHz spectrometer. NMR spectra were analyzed using MestreNova. Chemical shifts δ are referenced to the residual proton and carbon signal of the deuterated solvent ($CDCl_3$: δ = 7.26 ppm (1H), 77.16 ppm (^{13}C); 6d-DMSO: δ = 2.50 ppm (1H), 39.52 ppm (^{13}C); CD_3OD: δ = 3.31 ppm (^{1}H), 49.00 ppm (^{13}C); D_2O: δ = 4.79 ppm (1H)). Chemical shifts δ are given in ppm (parts per million) and coupling constants J in Hz (Hertz). Signal multiplicities are abbreviated as s (singlet), br s (broad singlet), d (doublet), dd (doublet of doublet), td (triplet of doublet), t (triplet), dt (doublet of triplet), q (quadruplet) and m (multiplet). Deuterated solvents for nuclear resonance spectroscopy were purchased from euriso-top®.

4.2. Synthesis of Histidine Precursors

*4.2.1. Synthesis of 4-(2-carboxy-2-hydroxyvinyl)-1H-imidazol-3-ium-1,3-$^{15}N_2$ chloride **9**:*



1 **2**

A two-neck round-bottomed flask was loaded with sodium hydroxide (32.0 mmol, 1.28 g, 8.0 eq) and water (10 mL). Ammonium-^{15}*N*-chloride (32.2 mmol, 1.75 g, 8.0 eq) was added, and the mixture was stirred at 50 °C for 45 min. After cooling to rt, $CuCO_3 \cdot Cu(OH)_2$ (4.00 mmol, 885 mg, 1.0 eq) and formaldehyde **1** (16% solution in water, 4.20 mmol, 0.8 mL, 1.05 eq) were added, and the reaction mixture was stirred at rt for 15 min. Dihydroxyacetone dimer (5.0 mmol, 900 mg, 1.3 eq) in water (3 mL) was added to the deep blue solution over 15 min and the reaction was stirred at 100 °C for 4 h. The mixture was cooled to rt and stored at 4 °C overnight.

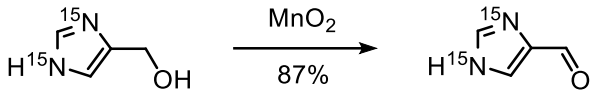
The brown imidazole-copper-complex was filtered off and washed with cold water (15 mL). The complex was suspended in water (20 mL), thioacetamide (4.0 mmol, 300 mg, 1.0 eq) was added, and the mixture was stirred at 50 °C for 2 h. The solid was filtered off and washed with water. The filtrate was concentrated under vacuum and the residue was purified by flash column chromatography (ethyl acetate/MeOH/NH_4OH (28%) 6:2.5:0.5, $KMnO_4$). The product (1*H*-imidazol-4-yl-$^{15}N_2$)methanol **2** was obtained as an orange oil (212 mg, 53%).

The formaldehyde solution was prepared by dissolving paraformaldehyde in water and 2 drops of 1M NaOH. The mixture was stirred at 60 °C for 2 h.

R_f = 0.54 (ethyl acetate/MeOH/NH_4OH (28%) 6:2.5:0.5, $KMnO_4$)

^{1}H-NMR (500 MHz, D_2O) δ 8.06 (t, *J* = 7.70 Hz, 1H), 7.20 (s, 1H), 4.59 (s, 2H) ppm.

^{13}C-NMR (176 MHz, 6d-DMSO) δ 136.18, 118.57, 56.81 (d, *J* = 5.00 Hz) ppm.

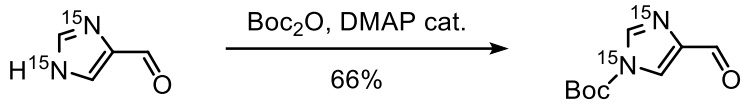


2 **3**

A round-bottomed flask equipped with a reflux condenser was loaded with compound **2** (1.7 mmol, 170 mg, 1.0 eq). MnO_2 (13.8 mmol, 1.2 g, 8.1 eq) and dioxane (20 mL) were added and the mixture was stirred at 80 °C for 3 h. The hot mixture was filtered over celite and washed with hot methanol (50 mL). The filtrate was concentrated in vacuo yielding 1*H*-imidazol-4-carbaldehyde-1,3-$^{15}N_2$ **3** as yellow solid (145 mg, 87%).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 9.74 (s, 1H), 8.02 (s, 1H), 7.95 (s, 1H) ppm.

^{13}C-NMR data are reported elsewhere.[49]



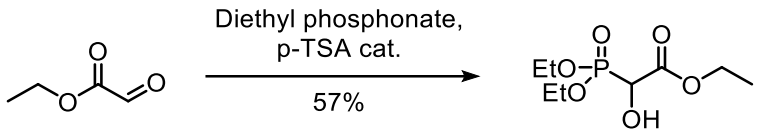
3 **4**

This reaction was performed under inert conditions. In a dry round-bottomed flask set under argon atmosphere, compound **3** (1.5 mmol, 145 mg, 1.0 eq) and DMAP (0.3 mmol, 40 mg, 0.2 eq) were dissolved in dry acetonitrile (30 mL) and DMF (4 mL). Boc_2O (2.6 mmol, 570 mg, 1.8 eq) was dissolved in DMF (2 mL) and added to the solution in small portions over 20 min. The mixture was stirred at rt overnight. The solvents were evaporated under reduced pressure giving the crude product as an orange oil, which was purified by flash column chromatography (ethyl acetate/heptane 1:1). *Tert*-butyl 4-formyl-1*H*-imidazol-1-carboxylate-1,3-$^{15}N_2$ **4** was obtained as white crystals in 66% yield (192 mg).

R_f = 0.45 (ethyl acetate/heptane 1:1)

^{1}H-NMR (500 MHz, $CDCl_3$) δ 9.93 (s, 1H), 8.14 (dd, *J* = 11.7, 7.95 Hz, 1H), 8.03 (d, *J* = 3.30 Hz, 1H), 1.65 (s, 9H) ppm.

^{13}C-NMR (150.9 MHz, $CDCl_3$) δ 207.08, 186.14 (dd, *J* = 8.81, 1.01 Hz), 146.38 (d, *J* = 22.2 Hz), 138.08 (dd, *J* = 8.44, 1.10 Hz), 122.21 (dd, *J* = 14.4, 0.98 Hz), 87.61, 27.98 ppm.



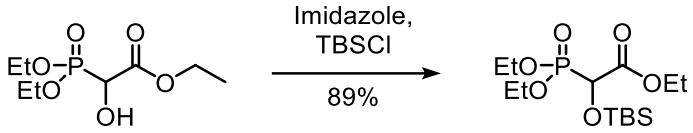
5 **6**

Diethyl phosphonate (11.8 mmol, 1.62 g, 1.2 eq) and pTSA (0.09 mmol, 16 mg, 0.01 eq) were dissolved in toluene (15 mL). A 50% *w/w* solution of ethyl glyoxylate **5** (9.80 mmol, 1.00 g, 1.0 eq) in toluene was added slowly and the mixture was refluxed at 135 °C overnight. The solvent was evaporated *in vacuo* and the crude product was purified by flash column chromatography (CH_2Cl_2/MeOH 15:1) giving ethyl-2-(diethoxyphosphoryl)-1-hydroxyacetate **6** in 57% yield (1.33 g).

R_f = 0.40 (CH_2Cl_2/methanol 79:1)

^{1}H-NMR (500 MHz, $CDCl_3$) δ 4.53 (dd, *J* = 15.7, 6.85 Hz, 1H), 4.33 (t, *J* = 7.10 Hz, 2H), 4.22 (m, 3H), 3.46 (t, *J* = 6.30 Hz, 1H), 1.36-1.31 (m, 10H) ppm.

^{13}C-NMR data are reported elsewhere.[49]



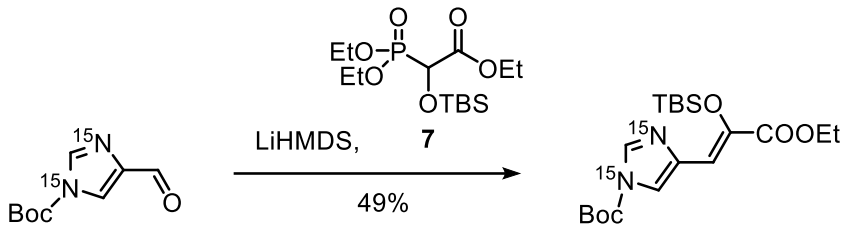
6 **7**

Compound **6** (5.54 mmol, 1.33 g, 1.0 eq) was dissolved in CH_2Cl_2 (25 mL), cooled to 0 °C and treated with imidazole (9.97 mmol, 679 mg, 1.8 eq). Afterwards, TBSCl (7.76 mmol, 1.16 g, 1.4 eq) in CH_2Cl_2 (2 mL) was added. After 1 h the mixture was brought to rt and stirred overnight. The reaction was quenched with saturated aqueous NH_4Cl (20 mL) and the aqueous phase was extracted with CH_2Cl_2 (5x 20 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated under vacuum. The crude product was purified by flash column chromatography (heptane/ethyl acetate 1:1) and the product ethyl 2-((*tert*-butyldimethylsilyl)oxy)-2-(diethoxyphosphoryl)acetate **7** was obtained as a colorless oil (1.75 g, 89%).

R_f = 0.40 (ethyl acetate/heptane 1:1)

^{1}H-NMR (500 MHz, $CDCl_3$) δ 4.57 (d, *J* = 18.0 Hz, 1H), 4.26-4.15 (m, 6H), 1.34-1.28 (m, 10H), 0.91 (s, 9H), 0.10 (d, *J* = 4.15 Hz, 6H) ppm.

^{13}C-NMR (150.9 MHz, $CDCl_3$) δ 168.70 (*C*(O)), 70.92 (d, *J* = 161 Hz, P*C*H), 63.68 (d, *J* = 6.65 Hz, PO*C*H_2), 63.58 (d, *J* = 6.85 Hz, PO*C*H_2), 61.81 (C(O)O*C*$H_2$$CH_3$), 25.67 (SiC(*C*$H_3$)$_3$), 16.57 (d, *J* = 2.57 Hz, POCH$_2$*C*H_3), 16.53 (d, *J* = 2.67 Hz, POCH$_2$*C*H_3), 14.23 (C(O)OCH$_2$*C*H_3), -5.21 (Si*C*H_3), -5.35 (Si*C*H_3) ppm.



4 **8**

This reaction was carried out under inert conditions. A dry two-neck flask flushed with argon was loaded with compound **7** (1.7 mmol, 620 mg, 1.2 eq). Dry THF (10 mL) was added, and the solution was cooled to -78 °C (using dry ice and acetone). LiHMDS (1M solution in THF; 1.9 mmol, 1.9 mL, 1.3 eq) was added slowly over 10 min and the mixture was stirred at -78 °C for 40 min. Compound **4** (1.5 mmol, 290 mg, 1.0 eq) was dissolved in dry THF and added to the mixture slowly. The reaction mixture was allowed to warm to rt and stirred overnight.

The reaction was quenched with saturated aqueous NH_4Cl (40 mL) and diluted with CH_2Cl_2 (40 mL). The phases were separated, and the aqueous phase was extracted with CH_2Cl_2 (5x 40 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated under reduced pressure. The crude product (brown oil) was purified by flash column chromatography (EA/heptane 1:4) yielding compound **8** (*tert*-butyl 4-(2-((tert-butyldimethylsilyl)oxy)-3-ethoxy-3-oxoprop-1-en-1-yl)-1*H*-imidazol-1-carboxylate-1,3-$^{15}N_2$) as colorless oil (286 mg, 49%) with an E/Z ratio of 1:1.

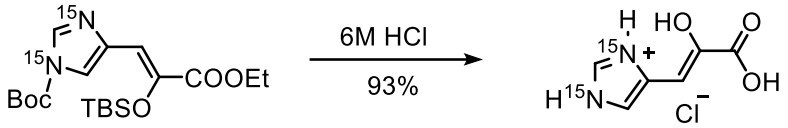
R_f = 0.56 / 0.48 (ethyl acetate/heptane 1:2)

E-Isomer: ^{1}H-NMR (600 MHz, 6d-DMSO) δ 8.01 (ddd, *J* = 11.6, 7.44, 0.96 Hz, 1H), 7.87 (dd, J = 4.02, 0.66 Hz, 1H), 6.39 (s, 1H), 4.25 (q, *J* = 7.14 Hz, 2H), 1.62 (s, 9H), 1.32-1.36 (m, 3H), 0.97 (s, 9H), 0.20 (s, 6H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO) δ 164.66, 147.04 (d, *J* = 1.92 Hz), 140.64 (d, *J* = 3.64 Hz), 136.63 (dd, *J* = 6.61, 1.28 Hz), 136.29 (dd, *J* = 9.10, 1.37 Hz), 117.84 (dd, *J* = 13.4, 1.19 Hz), 112.54 (d, *J* = 8.06 Hz), 85.74, 61.10, 28.05, 25.74, 18.42, 14.31, -4.63 ppm.

Z-Isomer: ^{1}H-NMR (600 MHz, 6d-DMSO) δ 8.08 (ddd, *J* = 11.6, 7.44, 0.96 Hz, 1H), 8.07 (dd, *J* = 4.02, 0.66 Hz, 1H), 6.92 (s, 1H), 4.28 (q, *J* = 7.14 Hz, 2H), 1.63 (s, 9H), 1.32-1.36 (m, 3H), 0.99 (s, 9H), 0.25 (s, 6H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO) δ 164.94, 147.19 (d, *J* = 1.86 Hz), 140.81 (d, *J* = 3.12 Hz), 137.44 (dd, *J* = 6.55, 1.25 Hz), 136.78 (dd, *J* = 9.43, 1.24 Hz), 118.12 (dd, *J* = 13.5, 1.43 Hz), 116.13 (d, *J* = 8.62 Hz), 85.91, 61.43, 28.11, 26.24, 18.89, 14.39, -3.12 ppm.



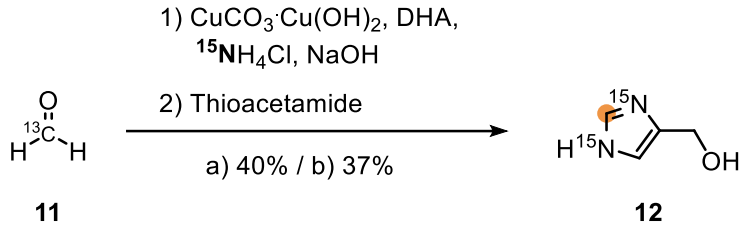
8 **9**

Compound **8** was treated with 6M aqueous HCl solution (12 mL) and the reaction was stirred at rt overnight. Subsequently, the solvent was evaporated *in vacuo* and the crude product was purified by trituration (Et_2O + a few drops ethyl acetate) giving 4-(2-carboxy-2-hydroxyvinyl)-1*H*-imidazol-3-ium-1,3-$^{15}N_2$ chloride **9** as a white solid in 93% yield (129 mg).

^{1}H-NMR (600 MHz, 6d-DMSO) δ 8.86 (dt, *J* = 5.82, 1.02 Hz, 1H), 7.67 (t, *J* = 3.84 Hz, 1H), 6.37 (d, *J* = 2.22 Hz, 1H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO) δ 165.09, 144.93 (d, *J* = 2.04 Hz), 134.05 (t, *J* = 15.0 Hz), 128.42 (dd, *J* = 10.9, 1.06 Hz), 118.41 (dd, *J* = 10.1, 1.34 Hz), 96.16 ppm.

4.2.2. *Synthesis of 4-(2-carboxy-2-hydroxyvinyl)-1H-imidazol-3-ium-2-^{13}C-1,3-$^{15}N_2$ chloride **16**:*



11 **12**

Compound **12a** was synthesized as described for compound **2** but using formaldehyde-^{13}C (16% solution in water, 4.20 mmol, 126 mg, 1.05 eq) as a substrate. The reaction afforded (1*H*-imidazol-4-yl-2-^{13}C-$^{15}N_2$)methanol **12a** as a yellow oil in 40% yield (162mg) with a ^{13}C content in the ε position of 54%.

Alternatively, a microwave vial was loaded with $^{15}NH_4Cl$ (16.0 mmol, 870 mg, 8.0 eq) and an aqueous NaOH solution (16.0 mmol, 640 mg in 5 mL of water, 8.0 eq) was added. The solution was stirred at 50 °C for 30 min. After cooling to rt, $CuCO_3{\cdot}Cu(OH)_2$ (2.0 mmol, 442 mg, 1 eq) and formaldehyde-^{13}C (16% solution in water, 2.1 mmol, 65 mg, 1.05 eq) were added. Subsequently, dihydroxyacetone dimer (2.5 mmol, 450 mg, 1.25 eq) in 2 mL water was added slowly over 10 min. The reaction was performed in a microwave device at 100 °C for 3.5 h. Afterwards, the vial was stored at 4 °C for 2 days. The synthesis was continued as described for compound **2**, and afforded (1*H*-imidazol-4-yl-2-^{13}C-$^{15}N_2$)methanol **12b** as a yellow oil in 37% yield (148 mg) with a ^{13}C content in the ε position of 74%.

The formaldehyde-^{13}C solution was prepared by dissolving paraformaldehyde-^{13}C in water and 2 drops of 1M NaOH. The mixture was stirred at 60 °C for 2 h.

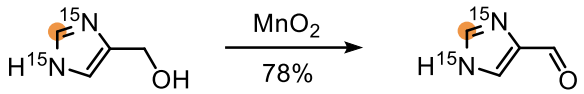
R_f = 0.54 (ethyl acetate/MeOH/NH_4OH (28%) 6:2.5:0.5, $KMnO_4$)

^{1}H-NMR (500 MHz, 6d-DMSO, **12a**) δ 7.55 (dt, *J* = 205, 10.0 Hz, 0.5 H, ^{13}CH), 7.54 (t, *J* = 10.0 Hz, 0.5 H, ^{12}CH), 6.86 (s, 1H), 4.37 (s, 2H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO, **12a**) δ 171.81, 134.90 (t, *J* = 5.70 Hz), 117.94, 55.98 ppm.

^{1}H-NMR (500 MHz, 6d-DMSO, **12b**) δ 7.55 (dt, *J* = 205, 10.0 Hz, 0.7H, ^{13}CH), 7.55 (t, *J* = 10.0 Hz, 0.3H, ^{12}CH), 6.86 (s, 1H), 4.37 (s, 2H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO, **12b**) δ 173.44, 135.46 (t, *J* = 5.36 Hz), 118.25, 56.29 ppm.

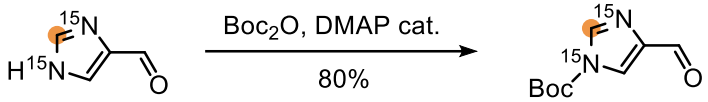


12 **13**

Compound **12** (2.4 mmol, 242 mg, 1.0 eq) was taken up in dioxane (30 mL), and MnO_2 (19.1 mmol, 1.66 g, 8.0 eq) was added. The mixture was stirred at 80 °C for 3 h. The hot mixture was filtered over celite and washed with hot MeOH. The filtrate was evaporated *in vacuo* giving 1*H*-imidazol-2-^{13}*C*-1,3-^{15}*N*$_2$-4-carbaldehyde **13** as yellow solid (78%, 184 mg).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 9.74 (s, 1H), 7.99 (s, 1H), 7.92 (t, *J* = 10.5 Hz, 0.3H, ^{12}C*H*), 7.91 (dt, *J* = 210, 10.5 Hz, 0.7H, ^{13}C*H*) ppm.

^{13}C-NMR data are reported elsewhere.[49]



13 **14**

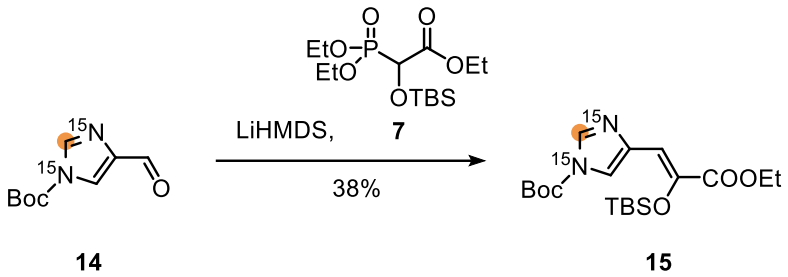
This reaction was performed under inert conditions. In a dry round-bottomed flask set under argon atmosphere, compound **13** (1.86 mmol, 145 mg, 1.0 eq) and DMAP (0.53 mmol, 65 mg, 0.3 eq) were dissolved in dry acetonitrile (30 mL) and DMF (4 mL). Boc_2O (3.5 mmol, 760 mg, 1.9 eq) was dissolved in dry acetonitrile (4 mL) and added to the solution in small portions over 5 min. The mixture was stirred at rt overnight.

The solvents were evaporated under reduced pressure giving the crude product as an orange oil, which was purified by flash column chromatography (ethyl acetate/heptane 1:1). *Tert*-butyl 4-formyl-1*H*-imidazol-1-carboxylate-2-^{13}C-1,3-$^{15}N_2$ **14** was obtained as white crystals in 80% yield (297 mg).

R_f = 0.45 (ethyl acetate/heptane 1:1)

^{1}H-NMR (500 MHz, $CDCl_3$) δ 9.93 (s, 1H), 8.14 (ddd, *J* = 219, 8.30, 3.30 Hz, 0.5H, ^{13}CH), 8.14 (dd, *J* = 8.30, 3.30 Hz, 0.5H, ^{12}CH), 8.03 (d, *J* = 2.15 Hz, 1H), 1.65 (s, 9H) ppm.

^{13}C-NMR (150.9 MHz, $CDCl_3$) δ 186.15 (d, *J* = 9.02 Hz), 146.38 (d, *J* = 22.0 Hz) , 138.08 (d, *J* = 8.72 Hz), 122.21 (d, *J* = 14.3 Hz), 87.62, 27.98 ppm.



14 **15**

This compound was synthesized as described for compound **8**. Compound **7** (1.7 mmol, 600 mg, 1.1 eq) was dissolved in dry THF (10 mL) and the solution was cooled to -78 °C (using dry ice and acetone). LiHMDS (1M solution in THF; 2.0 mmol, 2.0 mL, 1.3 eq) was added slowly over 10 min and the mixture was stirred at -78 °C for 40 min. Compound **13** (1.5 mmol, 297 mg, 1.0 eq) was dissolved in dry THF and added to the mixture slowly. The reaction was stirred and allowed to warm to rt overnight.

The reaction mixture was worked up as described for compound **8** yielding *tert*-butyl 4-(2-((tert-butyldimethylsilyl)oxy)-3-ethoxy-3-oxoprop-1-en-1-yl)-1*H*-imida-zole-1-carboxylate-2-^{13}C-1,3-$^{15}N_2$ **15** as a colorless oil (227 mg, 38%).

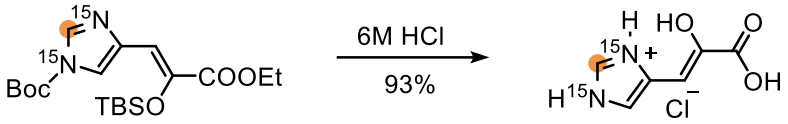
R_f = 0.56 / 0.48 (ethyl acetate/heptane 1:2)

E-isomer: ^{1}H-NMR (500 MHz, $CDCl_3$) δ 8.08 (ddd, *J* = 215, 7.85, 3.55 Hz, 0.4H, ^{13}C*H*), 8.08 (t, *J* = 7.55 Hz, 0.6H, ^{12}C*H*), 7.86 (s, 1H), 6.91 (s, 1H), 4.30-4.22 (m, 2H), 1.62 (s, 9H), 1.36-1.32 (m, 3H), 0.99 (s, 9H), 0.25 (s, 6H) ppm.

^{13}C-NMR (150.9 MHz, $CDCl_3$) δ 164.66, 147.03, 140.64 (d, *J* = 3.20 Hz), 137.42 (d, *J* = 1.57 Hz), 136.28 (d, *J* = 8.63 Hz), 117.88-117.76 (m), 112.58-112.49 (m), 85.73, 61.10, 28.04, 25.73, 18.41, 14.30, -4.63 ppm.

Z-isomer: ^{1}H-NMR (500 MHz, $CDCl_3$) δ 8.02 (dd, *J* = 7.85, 3.30 Hz, 0.14H, ^{12}C*H*), 8.01 (ddd, *J* = 215, 7.50, 3.00 Hz, 0.2H, ^{13}C*H*), 7.86 (s, 1H), 6.39 (s, 1H), 4.30-4.22 (m, 2H), 1.62 (s, 9H), 1.36-1.32 (m, 3H), 0.97 (s, 9H), 0.19 (s, 6H) ppm.

^{13}C-NMR (150.9 MHz, $CDCl_3$) δ 164.93, 147.19, 140.81 (d, *J* = 3.33 Hz), 137.47 (d, J = 1.64 Hz), 136.77 (d, J = 8.69 Hz), 118.16-118.03 (m), 116.16-116.10 (m), 85.91, 61.42, 28.11, 26.24, 18.88, 14.39, -3.12 ppm.



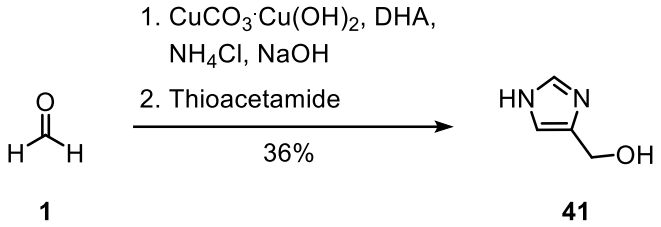
15 **16**

Compound **15** was treated with 6M aqueous HCl solution (8 mL) and stirred at rt overnight. Subsequently, the solvent was evaporated *in vacuo* and the crude product was purified by trituration (Et_2O + a few drops EA) giving the final product 4-(2-carboxy-2-hydroxyvinyl)-1*H*-imidazol-3-ium-2-^{13}C-1,3-$^{15}N_2$ chloride **16** as a white solid in 93% yield (102 mg).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 9.02 (dt, *J* = 220 Hz, 0.5 H), 9.02 (t, 0.5 H), 7.69 (s, 1H), 6.42 (s, 1H) ppm.

^{13}C-NMR data are reported elsewhere.[49]

*4.2.3. Synthesis of 4-(2-carboxy-2-hydroxyvinyl)-1H-imidazol-3-ium chloride **45**:*



1 **41**

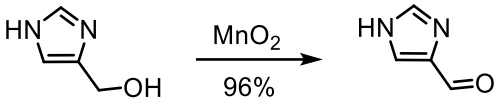
A microwave vial was loaded with formaldehyde **1** (16% solution in water, 4.0 mmol, 750 mg, 1.0 eq), 1,3-dihydroxy-acetone dimer (4.8 mmol, 865 mg, 1.2 eq), $CuCO_3$•$Cu(OH)_2$ (4.8 mmol, 1062 mg, 1.2 eq), NH_4Cl (40 mmol, 2140 mg, 10 eq), NaOH (40 mmol, 1600 mg, 10 eq) and water (15 mL). A microwave reaction was performed (100 °C, 1.5 h) and the flask was stored at 4 °C overnight to allow precipitation of the intermediate complex. The brown copper-imidazole complex was filtered off, washed with ice-cold water and taken up in water (20 mL). Thioacetamide (3.2 mmol, 240 mg, 0.8 eq) was added and the mixture was stirred at 50 °C for 2 h. The brown solid was filtered off, washed with water and the filtrate was evaporated under vacuum. The residue was purified by flash column chromatography (ethyl acetate/MeOH/ NH_4OH (28%) 6:2.5:0.5) giving compound **41** as a yellow oil (134 mg, 36%).

The formaldehyde solution was prepared by dissolving paraformaldehyde in water and 2 drops of 1M NaOH. The mixture was stirred at 60 °C for 2 h.

R_f = 0.54 (ethyl acetate/MeOH/NH_4OH (28%) 6:2.5:0.5, $KMnO_4$)

^{1}H-NMR (500 MHz, 6d-DMSO) δ 7.60 (s, 1H), 6.89 (s, 1H), 4.37 (s, 2H) ppm.

^{13}C-NMR data are reported elsewhere.[62]

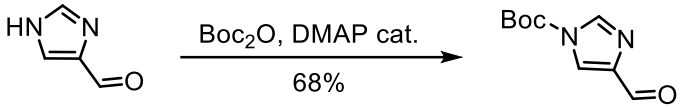


41 **42**

MnO_2 (5.75 mmol, 500 mg, 6.2 eq) was added to compound **41** (0.93 mmol, 93 mg, 1.0 eq) in 1,4-dioxane and the reaction stirred at 80 °C overnight. The hot mixture was filtered over celite and washed with hot methanol. The filtrate was evaporated *in vacuo* giving compound **42** as a bright yellow solid (96%, 88 mg).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 12.86 (s, 1H), 9.75 (s, 1H), 8.04 (s, 1H), 7.89 (s, 1H) ppm.

^{13}C-NMR data are reported elsewhere.[62]



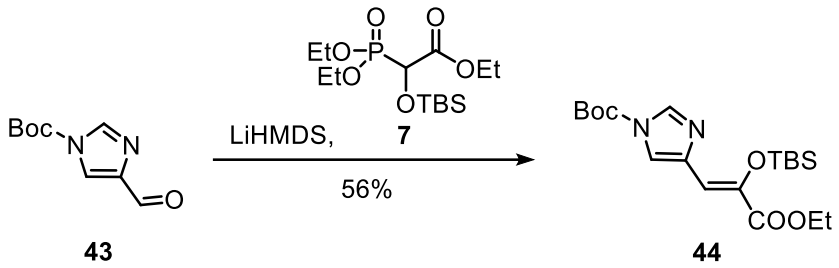
42 **43**

This reaction was carried out under argon atmosphere. In a dry 2-neck round-bottomed flask flushed with argon, compound **42** (0.62 mmol, 60 mg, 1.0 eq) and DMAP (0.098 mmol, 12 mg, 0.15 eq) were dissolved in dry THF (4 mL). Boc_2O (1.15 mmol, 250 mg, 1.85 eq) was dissolved in dry acetonitrile (3 mL) and added to the mixture slowly. The reaction was stirred at rt overnight. The solvents were evaporated under reduced pressure and the crude product was purified by flash column chromatography (heptane/ethyl acetate 1:1) giving compound **43** as a white solid (83 mg, 68%).

R_f = 0.45 (ethyl acetate/heptane 1:1)

^{1}H-NMR (500 MHz, $CDCl_3$) δ 9.93 (s, 1H), 8.14 (s, 1H), 8.03 (s, 1H), 1.65 (s, 9H) ppm.

^{13}C-NMR data are reported elsewhere.[62]



43 **44**

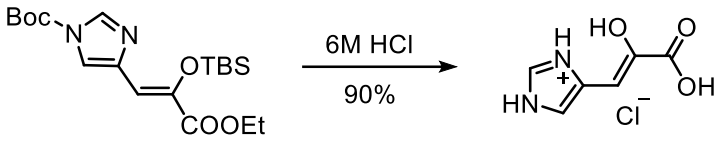
This reaction was carried out under inert conditions. In a dry round-bottomed flask set under argon atmosphere, compound **7** (0.5 mmol, 180 mg, 1.2 eq) was dissolved in dry THF (8 mL). The solution was cooled to -78 °C using dry ice/acetone and LiHMDS (1M solution in THF; 0.6 mmol, 1.4 eq) was added slowly over 15 min. The mixture was stirred for 30 min, before compound **43** (0.42 mmol, 83 mg, 1.0 eq) dissolved in dry THF (4 mL) was added. The mixture was allowed to warm to rt and stirred at rt overnight. The reaction was quenched by the addition of saturated aqueous NH_4Cl (20 mL) and extracted with CH_2Cl_2 (5x 20 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated under reduced pressure. The crude product (brown oil) was purified by flash column chromatography (ethyl acetate/heptane 1:4) yielding compound **44** as colorless oil (94 mg, 56%) with an E/Z ratio of 1.3:1.

R_f = 0.56 / 0.48 (ethyl acetate/heptane 1:2)

E-Isomer: ^{1}H-NMR (500 MHz, $CDCl_3$) δ 8.08 (s, 1H), 7.86 (s, 1H), 6.91 (s, 1H), 4.29-4.22 (m, 2H), 1.62 (s, 9H), 1.36-1.31 (m, 3H), 0.99 (s, 9H), 0.25 (s, 6H) ppm.

Z-Isomer: ^{1}H-NMR (500 MHz, $CDCl_3$) δ 8.08 (s, 1H), 8.01 (s, 1H), 6.93 (s, 1H), 4.29-4.22 (m, 2H), 1.62 (s, 9H), 1.36-1.31 (m, 3H), 0.97 (s, 9H), 0.19 (s, 6H) ppm.

^{13}C-NMR data are reported elsewhere.[62]



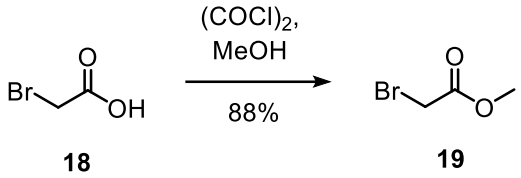
44 **45**

Compound **44** (0.24 mmol, 94 mg, 1.0 eq) was treated with 6M aqueous HCl solution (12 mL) and the reaction was stirred at rt overnight. Subsequently, the solvent was evaporated *in vacuo* and the crude product was purified by trituration (Et_2O + a few drops ethyl acetate) giving 4-(2-carboxy-2-hydroxyvinyl)-1*H*-imidazol-3-ium chloride **45** as a white solid in 90% yield (41 mg).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 14.45 (s, 2H), 9.05 (s, 1H), 7.70 (s, 1H), 6.41 (s, 1H) ppm.

^{13}C-NMR data are reported elsewhere.[62]

*4.2.4. Synthesis of methyl 1-benzyl-1H-imidazol-5-carboxylate **24**:*



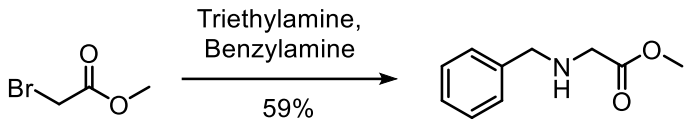
18 **19**

This reaction was carried out under argon atmosphere. A two-neck flask loaded with 2-bromoacetic acid **18** (38.0 mmol, 5.30 g, 1.0 eq) was equipped with a reflux condenser and the system was flushed with argon. Oxalyl chloride (49.6 mmol, 4.35 mL, 1.3 eq) was added in a dropwise manner and the solution was stirred at 60 °C for 3 h. Subsequently, dry MeOH was added slowly at 0 °C and the mixture was stirred at rt for 30 min. The reaction was quenched with water (40 mL), and the aqueous phase was extracted with Et_2O (5x 40 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated under reduced pressure (>200 mbar). Methyl 2-bromoacetate **19** was obtained as a yellow oil in 88% yield (5.12 g).

R_f = 0.47 (ethyl acetate/heptane 1:4)

^{1}H-NMR (500 MHz, $CDCl_3$) δ 3.84 (s, 2H), 3.78 (s, 3H) ppm.

^{13}C-NMR data are reported elsewhere.[62]



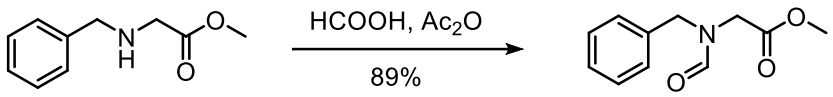
19 **20**

Methyl-2-bromoacetate **19** (33.5 mmol, 5.12 g, 1.0 eq) was dissolved in THF (40 mL). The solution was slowly treated with triethylamine (60.2 mmol, 8.4 mL, 1.8 eq) followed by the addition of benzylamine (36.8 mmol, 4.0 mL, 1.1 eq) in THF (5 mL). The mixture was stirred at rt overnight. The white precipitate was filtered off, washed with THF and the filtrate was concentrated under reduced pressure. The crude product was purified by flash column chromatography (ethyl acetate/heptane 3:2) giving methyl benzylglycinate **20** as a partly crystallized colorless oil (3.52 g, 59%).

R_f = 0.30 (ethyl acetate/heptane 3:2)

^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.26 (s, 2H), 7.25 (s, 2H), 7.21-7.17 (m, 1H), 3.73 (s, 2H), 3.66 (s, 3H), 3.36 (s, 2H) ppm.

^{13}C-NMR (150.9 MHz, $CDCl_3$) 173.00, 139.54, 128.60, 128.39, 127.32, 53.40, 51.90, 50.02 ppm.



20 **21**

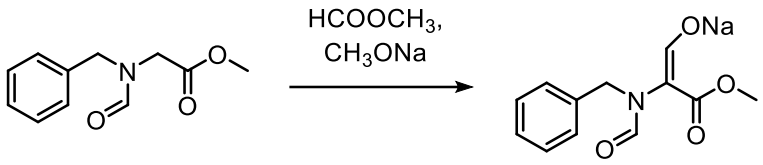
Methyl benzylglycinate **20** (19.6 mmol, 3.50 g, 1.0 eq) was slowly treated with formic acid (16 mL, 95%) followed by the dropwise addition of acetic anhydride (16.6 mL). The mixture was stirred at 100 °C for one hour and subsequently cooled to rt. The solvents were evaporated completely *in vacuo* and the crude product was taken up in CH_2Cl_2 (25 ml). The resulting solution was washed with aqueous saturated $NaHCO_3$ (2x 25 mL), water (25 mL), and brine (25 mL), and dried over $MgSO_4$. The solvent was evaporated under reduced pressure giving methyl *N*-benzyl-*N*-formylglycinate **21** as a red oil (3.64 g, 89%) with an E/Z ratio of 4:1.

Rotamer 1: ^{1}H-NMR (400 MHz, $CDCl_3$) δ 8.36 (s, 1H), 7.39-7.30 (m, 3H), 7.24-7.20 (m, 2H), 4.53 (s, 2H), 3.97 (s, 2H), 3.71 (s, 3H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO) δ 168.53, 163.12 (d, *J* = 16.9 Hz), 135.91, 128.48, 127.95, 127.67, 51.67, 50.90, 45.74 ppm.

Rotamer 2: ^{1}H-NMR (400 MHz, $CDCl_3$) δ 8.18 (s, 1H), 7.39-7.30 (m, 3H), 7.24-7.20 (m, 2H), 4.61 (s, 2H), 3.85 (s, 2H), 3.70 (s, 3H) ppm.

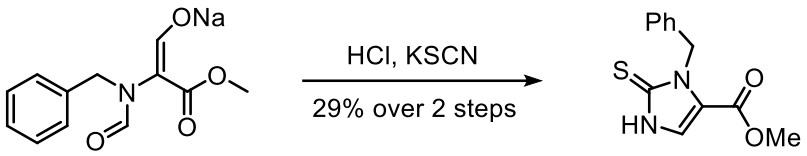
^{13}C-NMR (150.9 MHz, 6d-DMSO) δ 169.65, 163.62 (d, *J* = 15.9 Hz), 136.05, 128.32, 127.84, 127.27, 51.84, 47.50, 42.97 ppm.



21 **22**

This reaction was carried out under argon atmosphere. A dry 3-neck flask flushed with argon was loaded with compound **21** (9.65 mmol, 2.0 g, 1.0 eq) and methyl formate (29.0 mmol, 2.0 mL, 3.0 eq). After cooling to 0 °C, sodium methoxide (13.5 mmol, 730 mg, 1.4 eq) was added in small portions over one hour, followed by the addition of toluene (4 mL). The mixture was stirred at rt for 2 h and stored at 4 °C overnight. The solvents were evaporated in vacuo giving crude sodium 2-(*N*-benzylformamido)-3-methoxy-3-oxoprop-1-en-1-olate **22**, which was used in the next step without further purification.

NMR data are reported elsewhere.[62]



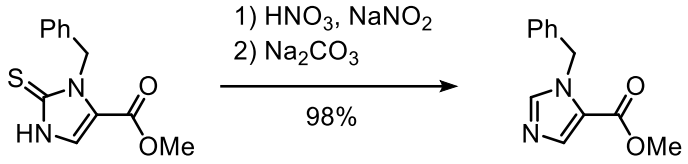
22 **23**

Crude compound **22** (9.65 mmol, 1.0 eq) was taken up in MeOH (6 mL) and water (6 mL). HCl conc. (2 mL) was slowly added at 0 °C and the mixture was stored in the fridge for 1.5 h. After warming to rt, KSCN (11.6 mmol, 1.13 g, 1.2 eq) was added and the mixture was stirred at 80 °C for 2 h. The mixture was diluted with water and the pH was set to 8 using 1M NaOH. The mixture was extracted with Et_2O (5x 20 mL) and the combined organic extracts were dried over $MgSO_4$. The solvent was evaporated under reduced pressure giving the crude product as an orange oil, which was purified by recrystallization. For that purpose, the residue was dissolved in a small amount of hot ethyl acetate and the product was allowed to crystallize at 4 °C overnight. The product was carefully washed with a small amount of heptane and dried under vacuum. The mother liquor was concentrated under reduced pressure and purified again by flash column chromatography (ethyl acetate/heptane 1:3) followed by recrystallization giving methyl 3-benzyl-2-thioxo-2,3-dihydro-1*H*-imidazol-4-carboxylate **23** as white crystals (700 mg, 29% yield over 2 steps).

R_f = 0.45 (ethyl acetate/heptane 1:1)

^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.45 (s, 1H), 7.37 (d, *J* = 7.16 Hz, 2H), 7.32-7.29 (m, 3H), 5.71 (s, 2H), 3.79 (s, 3H) ppm.

^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 165.63, 158.77, 136.61, 128.69, 127.97, 127.87, 122.93, 121.37, 52.23, 48.93 ppm.



23 **24**

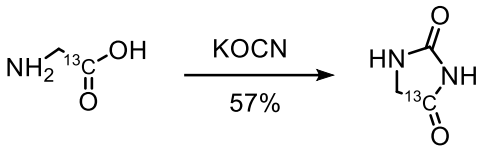
Compound **23** (1.61 mmol, 400 mg, 1.0 eq) was dissolved in THF (8 mL), aqueous $NaNO_2$ solution (2.0 mg/mL; 0.46 mmol, 16 mL, 2.8 eq) was added and the mixture was homogenized under ultrasonication (5 min). HNO_3 35% (1.6 mL) was added carefully and the reaction was stirred at rt until TLC (ethyl acetate 100%) showed complete conversion of the substrate (approximately 2 h). Saturated aqueous Na_2CO_3 solution (10 mL) was added and the reaction was heated to reflux at 130 °C oil bath temperature for 1.5 h. After cooling to rt, the aqueous phase was extracted with Et_2O (5x 30 mL) and the combined organic extracts were dried over $MgSO_4$. The solvents were evaporated under reduced pressure giving methyl 1-benzyl-1*H*-imidazol-5-carboxylate **24** as a bright yellow solid in 98% yield (350 mg).

^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.78 (s, 1H), 7.63 (s, 1H), 7.36-7.30 (m, 3H), 7.17 (d, *J* = 6.96 Hz, 2H), 5.52 (s, 2H), 3.82 (s, 3H) ppm.

^{13}C-NMR (150.9 MHz, $CDCl_3$) δ 159.95, 143.60, 137.50, 137.44, 128.64, 127.63, 126.90, 51.43, 48.92 ppm.

4.3. Synthesis of Isoleucine Precursors

*4.3.1. Synthesis of sodium 3-methyl-2-oxopentanoate-1-^{13}C **33**:*

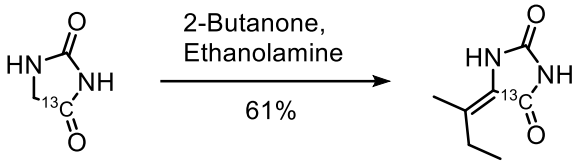


30 **31**

A round-bottomed flask equipped with a reflux condenser was loaded with glycine-*1*-^{13}C **30** (6.60 mmol, 500 mg, 1.0 eq), potassium cyanate (9.86 mmol, 800 mg, 1.5 eq) and water. The mixture was stirred at 100 °C for 4 h. HCl conc. (1.6 mL) was added at 0 °C and the mixture was stirred at 100 °C for 2h. The solvent was evaporated completely and the white residue was suspended in water (2 mL). The flask was stored at 4 °C overnight for further precipitation of the product. The white solid was filtered off and washed with cold water giving hydantoin-*1*-^{13}C **31** as white solid (372 mg, 57%).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.62 (s, 1H), 7.72 (d, *J* = 7.00 Hz, 1H), 3.85 (d, *J* = 5.50 Hz, 2H) ppm.

^{13}C-NMR (176 MHz, 6d-DMSO) δ 173.91, 158.37 (d, *J* = 8.96 Hz), 47.25 (d, *J* = 46.2 Hz) ppm.

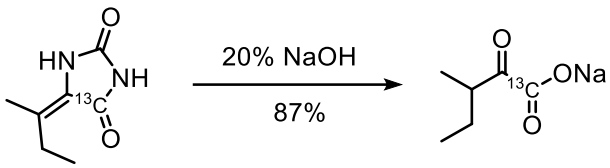


31 **32**

A microwave vial (2-5 mL) was loaded with hydantoin-*1*-^{13}C **31** (3.25 mmol, 325 mg, 1.0 eq), 2-butanone (13.0 mmol, 1.16 mL, 4.0 eq), and ethanolamine (3.25 mmol, 200 µL, 1.0 eq), and the mixture heated to 100 °C using a microwave device for 5 hours. The reaction mixture was concentrated *in vacuo* and the residue was taken up in water (1 mL), which led to the formation of a white precipitate. The pH was set to 3-4 using 1M HCl and the flask was stored at 4 °C overnight for further precipitation of the product. The white solid was filtered off and washed with cold water (3x 10 mL) yielding 5-(butan-2-ylidene)imidazolidine-2,4-dione-*4*-^{13}C **32** as an off-white solid (309 mg, 61%).

Isomer 1: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.81 (s, 1H), 9.75 (s, 1H), 2.12 (q, *J* = 7.50 Hz, 2H), 1.79 (s, 3H), 0.97 (dt, *J* = 7.50, 1.35 Hz, 3H) ppm.

Isomer 2: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.81 (s, 1H), 9.81 (s, 1H), 2.61 (q, *J* = 7.50 Hz, 2H), 2.10 (s, 3H), 0.97 (dt, *J* = 7.50, 1.35 Hz, 3H) ppm.



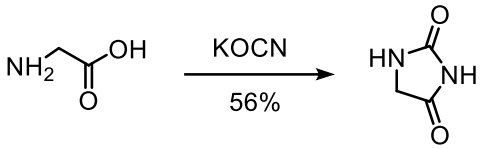
32 **33**

Compound **32** was treated with an aqueous solution of sodium hydroxide (20%, 6 mL) and the reaction was stirred under an argon stream at 100 °C for 5 h. After cooling to rt, the reaction mixture was diluted with water. The aqueous phase was washed with diethyl ether and subsequently acidified to pH 1 by adding HCl conc. at 0 °C. The acidic solution was extracted with diethyl ether, the combined organic extracts were dried over $MgSO_4$. The solvent was evaporated carefully under reduced pressure (> 200 mbar) yielding sodium 3-methyl-2-oxopentanoic-*1*-^{13}C acid as a red oil. The residue was taken up in water and pH was set to 7-8 using 0.5M NaOH. Lyophilization was performed yielding sodium 3-methyl-2-oxopentanoate-*1*-^{13}C **33** as a white powder (262 mg, 87%).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 2.89 (sextet, *J* = 6.75 Hz, 1H), 1.71-1.62 (m, 1H), 1.47-1.38 (m, 1H), 1.06 (d, *J* = 7.00 Hz, 3H), 0.86 (t, *J* = 7.50 Hz, 3H) ppm.

^{13}C-NMR (150.9 MHz, D_2O) δ 211.39 (d, *J* = 61.4 Hz), 172.25, 43.70 (d, *J* = 12.1 Hz), 24.46, 13.72, 10.58 ppm.

*4.3.2. <u>Synthesis of sodium 3-methyl-2-oxopentanoate **49**:</u>*

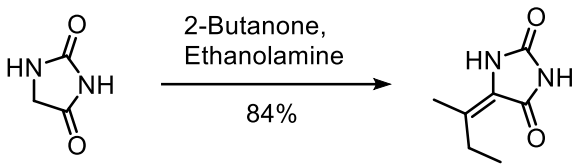


46 **47**

Synthesis of compound **47** was performed as described for compound **31**, but by using unlabeled substrates. The product was obtained as a white solid in 56% yield (371 mg).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.62 (s, 1H), 7.72 (s, 1H), 3.85 (s, 2H) ppm.

^{13}C-NMR (176 MHz, 6d-DMSO) δ 175.53, 159.65, 48.26 ppm.



47 **48**

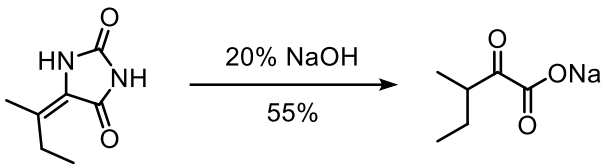
Compound **48** was prepared as described for compound **32** by using 500 mg of hydantoin **47** (5.0 mmol, 1.0 eq), 305 µL of ethanolamine (5.0 mmol, 1.0 eq) and 2 mL of 2-butanone. The product **48** was obtained as an off-white solid in 84% yield (647 mg).

Isomer 1: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.83 (s, 1H), 9.74 (s, 1H), 2.12 (q, *J* = 7.50 Hz, 2H), 1.78 (s, 3H), 0.97 (dt, *J* = 7.50, 1.50 Hz, 3H) ppm.

^{13}C-NMR (176 MHz, 6d-DMSO) δ 164.26, 154.02, 129.24, 124.34, 24.08, 15.31, 11.49 ppm.

Isomer 2: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.83 (s, 1H), 9.80 (s, 1H), 2.61 (q, *J* = 7.50 Hz, 2H), 2.10 (s, 3H), 0.97 (dt, *J* = 7.50, 1.50 Hz, 3H) ppm.

^{13}C-NMR (176 MHz, 6d-DMSO) δ 164.87, 154.08, 129.81, 124.77, 26.86, 17.99, 12.60 ppm.



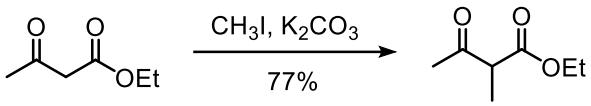
48 **49**

The opening of the hydantoin ring to obtain compound **49** was achieved as described for compound **33** and by using 660 mg of **48** (4.29 mmol, 1.0 eq) and 13 mL of 20% aqueous NaOH. The non-lyophilized product **49** was obtained in 55% yield (358 mg).

^{1}H-NMR (500 MHz, 6d-DMSO) δ 3.03 (sextet, *J* = 6.75 Hz, 1H), 1.67-1.61 (m, 1H), 1.41-1.32 (m, 1H), 1.03 (d, *J* = 6.95 Hz, 3H), 0.84 (t, *J* = 7.45 Hz, 3H) ppm.

^{13}C-NMR (150.9 MHz, D_2O) δ 221.42, 172.25, 43.72, 24.47, 13.73, 10.59 ppm.

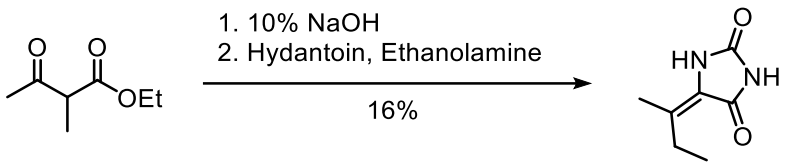
*4.3.3. Synthesis of 5-(butan-2-ylidene)imidazolidine-2,4-dione **38**:*



34 **35**

A round-bottomed flask was loaded with ethanol abs. (60 mL), ethyl acetoacetate **34** (46 mmol, 5.8 mL, 1.02 eq) and potassium carbonate (52 mmol, 7.2 g, 1.15 eq). Iodomethane (45 mmol, 2.8 mL, 1.0 eq) was added and the mixture was stirred at 40 °C for 90 h. The reaction mixture was filtered over celite and washed with diethyl ether. The solvent was evaporated under reduced pressure, the residue was taken up in diethyl ether (20 mL) and washed with water (20 mL). The aqueous phase was extracted with Et_2O (3x 20 mL), the combined organic extracts were dried over $MgSO_4$ and evaporated under reduced pressure, giving ethyl 2-methyl-3-oxobutanoate **35** as colorless liquid (77%, 5.0 g).

^{1}H-NMR (400 MHz, 6d-DMSO) δ 4.19 (q, *J* = 7.12 Hz, 2H), 3.49 (q, *J* = 7.16 Hz, 1H), 2.23 (s, 3H), 1.33 (d, *J* = 7.16 Hz, 3H), 1.27 (t, *J* = 7.24 Hz, 3H) ppm.



35 **38**

A microwave vial was loaded with 10% aqueous NaOH (4.16 mmol, 1.2 eq) and compound **35** (2.77 mmol, 400 mg, 1.0 eq), and the mixture was stirred in the microwave for 2 h at 100 °C. Afterwards, hydantoin (3.33 mmol, 333 mg, 1.2 eq) and ethanolamine (3.88 mmol, 235 µL, 1.4 eq) were added, and the mixture was stirred at 100 °C for another 6 h. The solvent was evaporated and the crude product was purified with flash column chromatography (Acetone/Heptane/EA 1:2:4) giving compound **38** as a white solid (68 mg, 16%).

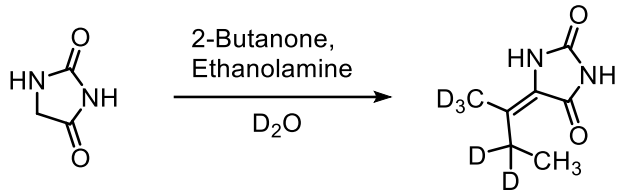
Isomer 1: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.82 (s, 1H), 9.74 (s, 1H), 2.12 (q, *J* = 7.50 Hz, 2H), 1.78 (s, 3H), 0.96 (dt, *J* = 7.50, 1.25 Hz, 3H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO) δ 164.67, 154.46, 130.52, 124.52, 24.48, 15.71, 11.81 ppm.

Isomer 2: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.82 (s, 1H), 9.80 (s, 1H), 2.61 (q, *J* = 7.50 Hz, 2H), 2.10 (s, 3H), 0.96 (dt, *J* = 7.50, 1.25 Hz, 3H) ppm.

^{13}C-NMR (150.9 MHz, 6d-DMSO) δ 165.29, 154.51, 131.05, 124.95, 27.27, 18.38, 13.01 ppm.

*4.3.4. Synthesis of 5-(butan-2-ylidene-1,1,1,3,3-d_5)imidazolidine-2,4-dione **40**:*



40

This reaction was carried out as described for compound **32** but using 722 mg of commercially available hydantoin (7.21 mmol, 1.3 eq), 508 mg of ethanolamine (8.32 mmol, 1.5 eq), 400 mg of 2-butanone (5.55 mmol, 1.0 eq) and 2 mL of D_2O as solvent. The reaction afforded 5-(butan-2-ylidene-*1,1,1,3,3-d_5*)imidazolidine-2,4-dione-*4*-^{13}C **40** as an off-white solid. The yield was not determined.

Isomer 1: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.79 (s, 1H), 9.74 (s, 1H), 0.95 (s, 3H) ppm.

Isomer 2: ^{1}H-NMR (500 MHz, 6d-DMSO) δ 10.79 (s, 1H), 9.80 (s, 1H), 0.95 (s, 3H) ppm.

5. Resources

(1) Kay, L. E. NMR studies of protein structure and dynamics. *J Magn Reson* **2011**, *213* (2), 477-491.

(2) Rosati, M.; Barbieri, L.; Hlavac, M.; Kratzwald, S.; Lichtenecker, R. J.; Konrat, R.; Luchinat, E.; Banci, L. Towards cost-effective side-chain isotope labelling of proteins expressed in human cells. *Journal of Biomolecular NMR* **2024**, *78* (4), 237-247.

(3) Rowlinson, B.; Crublet, E.; Kerfah, R.; Plevin, M. J. Specific isotopic labelling and reverse labelling for protein NMR spectroscopy: using metabolic precursors in sample preparation. *Biochemical Society Transactions* **2022**, *50* (6), 1555-1567.

(4) Fasshuber, H. K.; Demers, J. P.; Chevelkov, V.; Giller, K.; Becker, S.; Lange, A. Specific 13C labeling of leucine, valine and isoleucine methyl groups for unambiguous detection of long-range restraints in protein solid-state NMR studies. *J Magn Reson* **2015**, *252*, 10-19.

(5) Ruschak, A. M.; Velyvis, A.; Kay, L. E. A simple strategy for 13C,1H labeling at the Ile-γ2 methyl position in highly deuterated proteins. *J Biomol NMR* **2010**, *48* (3), 129-135.

(6) Schörghuber, J.; Geist, L.; Platzer, G.; Konrat, R.; Lichtenecker, R. J. Highly Selective Stable Isotope Labeling of Histidine Residues by Using a Novel Precursor in E. coli-Based Overexpression Systems. *Chembiochem* **2017**, *18* (15), 1487-1491.

(7) Freiburger, L.; Sonntag, M.; Hennig, J.; Li, J.; Zou, P.; Sattler, M. Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *Journal of Biomolecular NMR* **2015**, *63* (1), 1-8.

(8) Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level*; Wiley, 2016.

(9) Schörghuber, J.; Geist, L.; Platzer, G.; Feichtinger, M.; Bisaccia, M.; Scheibelberger, L.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Late metabolic precursors for selective aromatic residue labeling. *J Biomol NMR* **2018**, *71* (3), 129-140.

(10) Cerofolini, L.; Fragai, M.; Ravera, E.; Diebolder, C. A.; Renault, L.; Calderone, V. Integrative Approaches in Structural Biology: A More Complete Picture from the Combination of Individual Techniques. *Biomolecules* **2019**, *9* (8), 370.

(11) Dubach, V. R. A.; Guskov, A. The Resolution in X-ray Crystallography and Single-Particle Cryogenic Electron Microscopy. *Crystals* **2020**, *10* (7), 580.

(12) Zheng, H.; Handing, K. B.; Zimmerman, M. D.; Shabalin, I. G.; Almo, S. C.; Minor, W. X-ray crystallography over the past decade for novel drug discovery. *Expert Opin. Drug Discov.* **2015**, *10* (9), 975-989.

(13) Platzer, G.; Mayer, M.; McConnell, D. B.; Konrat, R. NMR-driven structure-based drug discovery by unveiling molecular interactions. *Commun Chem* **2025**, *8* (1), 167.

(14) Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **2020**, *587* (7832), 157-161.

(15) Bai, X. C.; McMullan, G.; Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci* **2015**, *40* (1), 49-57.

(16) Egelman, E. H. The Current Revolution in Cryo-EM. *Biophys J* **2016**, *110* (5), 1008-1012.

(17) Bax, A.; Clore, G. M. Protein NMR: Boundless opportunities. *J Magn Reson* **2019**, *306*, 187-191.

(18) Alderson, T. R.; Kay, L. E. Unveiling invisible protein states with NMR spectroscopy. *Curr Opin Struct Biol* **2020**, *60*, 39-49.

(19) Fernandez, C.; Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* **2003**, *13* (5), 570-580.

(20) Dumez, J.-N. NMR methods for the analysis of mixtures. *Chem. Commun.* **2022**, *58*, 13855-13872.

(21) Rule, G. S.; Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy*; Springer, 2006.

(22) Kurreck, J.; Engels, J. W.; Lottspeich, F. *Bioanalytik*; Springer, 2021.

(23) Kawale, A. A.; Burmann, B. M. Advanced NMR spectroscopy methods to study protein structure and dynamics. In *Advanced Spectroscopic Methods to Study Biomolecular Structure and Dynamics*, 2023; pp 125-152.

(24) Boisbouvier, J.; Kay, L. E. Advanced isotopic labeling for the NMR investigation of challenging proteins and nucleic acids. *J Biomol NMR* **2018**, *71* (3), 115-117.

(25) Keeler, J. *Understanding NMR spectroscopy*; Wiley, 2002.

(26) Pellecchia, M.; Sem, D. S.; Wuthrich, K. NMR in drug discovery. *Nat Rev Drug Discov* **2002**, *1* (3), 211-219.

(27) Fesik, S. W.; Luly, J. R.; Erickson, J. W.; Abad-Zapatero, C. Isotope-edited proton NMR study on the structure of a pepsin/inhibitor complex. *Biochemistry* **1988**, *27* (22), 8297-8301.

(28) Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E. Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* **2003**, *41* (10), 843-852.

(29) Zuiderweg, E. R. P.; Fesik, S. W. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **1989**, *28* (6), 2387-2391.

(30) Hu, Y.; Cheng, K.; He, L.; Zhang, X.; Jiang, B.; Jiang, L.; Li, C.; Wang, G.; Yang, Y.; Liu, M. NMR-Based Methods for Protein Analysis. *Anal Chem* **2021**, *93* (4), 1866-1879.

(31) Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.beta. *Biochemistry* **1989**, *28* (15), 6150-6156.

(32) Marion, D.; Kay, L. E.; Sparks, S. W.; Torchia, D. A.; Bax, A. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *Journal of the American Chemical Society* **1989**, *111* (4), 1515-1517.

(33) Kleckner, I. R.; Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **2011**, *1814* (8), 942-968.

(34) Gobl, C.; Madl, T.; Simon, B.; Sattler, M. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog Nucl Magn Reson Spectrosc* **2014**, *80*, 26-63.

(35) Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **2013**, *73*, 1-16.

(36) Charlier, C.; Bouvignies, G.; Pelupessy, P.; Walrant, A.; Marquant, R.; Kozlov, M.; Ioannes, P. D.; Bolik-Coulon, N.; Sagan, S.; Cortes, P.; *et al.* Structure and Dynamics of an Intrinsically Disordered Protein Region That Partially Folds upon Binding by Chemical-Exchange NMR. *J Am Chem Soc* **2017**, *139* (35), 12219-12227.

(37) Delaforge, E.; Kragelj, J.; Tengo, L.; Palencia, A.; Milles, S.; Bouvignies, G.; Salvi, N.; Blackledge, M.; Jensen, M. R. Deciphering the Dynamic Interaction Profile of an Intrinsically Disordered Protein by NMR Exchange Spectroscopy. *J Am Chem Soc* **2018**, *140* (3), 1148-1158.

(38) Ocasio, V. J.; Corrêa, F.; Gardner, K. H. Ligand-induced folding of a two-component signaling receiver domain. *Biochemistry* **2015**, *54* (6), 1353-1363.

(39) Platzer, G.; Mayer, M.; Beier, A.; Bruschweiler, S.; Fuchs, J. E.; Engelhardt, H.; Geist, L.; Bader, G.; Schorghuber, J.; Lichtenecker, R.; *et al.* PI by NMR: Probing CH-pi Interactions in Protein-Ligand Complexes by NMR Spectroscopy. *Angew Chem Int Ed Engl* **2020**, *59* (35), 14861-14868.

(40) Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed.* **1999**, *38* (12), 1784-1788.

(41) Monaco, S.; Angulo, J.; Wallace, M. Imaging Saturation Transfer Difference (STD) NMR: Affinity and Specificity of Protein-Ligand Interactions from a Single NMR Sample. *J Am Chem Soc* **2023**, *145* (30), 16391-16397.

(42) Bhunia, A.; Bhattacharjya, S.; Chatterjee, S. Applications of saturation transfer difference NMR in biological systems. *Drug Discov Today* **2012**, *17* (9-10), 505-513.

(43) Shuker, S. B.; Hajdilk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR *Science* **1996**, *274* (5292), 1531-1534.

(44) Tugarinov, V.; Kay, L. E. Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *Chembiochem* **2005**, *6* (9), 1567-1577.

(45) Ruschak, A. M.; Kay, L. E. Methyl groups as probes of supra-molecular structure, dynamics and function. *J Biomol NMR* **2010**, *46* (1), 75-87.

(46) Skrisovska, L.; Schubert, M.; Allain, F. H. T. Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. *Journal of Biomolecular NMR* **2010**, *46* (1), 51-65.

(47) Zemella, A.; Thoring, L.; Hoffmeister, C.; Kubick, S. Cell-Free Protein Synthesis:Pros and Cons of Prokaryotic and Eukaryotic Systems. *ChemBioChem* **2015**, *16* (17), 2420-2431.

(48) Lichtenecker, R. J.; Weinhäupl, K.; Schmid, W.; Konrat, R. a-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *J Biomol NMR* **2013**, *57* (4), 327-331.

(49) Bruschweiler, S.; Hlavac, M.; Kratzwald, S.; Schorghuber, J.; Siess, K. M.; Wimmer, A.; Platzer, G.; Konrat, R.; Lichtenecker, R. J. Isotopologues of a Metabolic Precursor for Selective N-15 and C-13 Histidine Labeling. *J Mol Biol* **2025**, *437* (23), 169347.

(50) Wiesner, S.; Sprangers, R. Methyl groups as NMR probes for biomolecular interactions. *Curr Opin Struct Biol* **2015**, *35*, 60-67.

(51) Lichtenecker, R. J.; Coudevylle, N.; Konrat, R.; Schmid, W. Selective isotope labelling of leucine residues by using alpha-ketoacid precursor compounds. *Chembiochem* **2013**, *14* (7), 818-821.

(52) Lichtenecker, R. J.; Weinhaupl, K.; Reuther, L.; Schorghuber, J.; Schmid, W.; Konrat, R. Independent valine and leucine isotope labeling in Escherichia coli protein overexpression systems. *J Biomol NMR* **2013**, *57* (3), 205-209.

(53) Valine, leucine and isoleucine biosynthesis. In *KEGG: Kyoto Encyclopedia of Genes and Genomes*, Kanehisa Laboratories, 2025; Updated 28.12.2025.

(54) Weininger, U. Optimal Isotope Labeling of Aromatic Amino Acid Side Chains for NMR Studies of Protein Dynamics. *Methods in Enzymology* **2019**, *614*, 67-86.

(55) Phenylalanine, tyrosine and tryptophan biosynthesis. In *KEGG: Kyoto Encyclopedia of Genes and Genomes*, Kanehisa Laboratories, 2025; Updated 28.12.2025.

(56) Hoffmann, B.; Lohr, F.; Laguerre, A.; Bernhard, F.; Dotsch, V. Protein labeling strategies for liquid-state NMR spectroscopy using cell-free synthesis. *Prog Nucl Magn Reson Spectrosc* **2018**, *105*, 1-22.

(57) Mondal, S.; Shet, D.; Prasanna, C.; Atreya, H. S. High yield expression of proteins in E.coli for NMR studies. *Advances in Bioscience and Biotechnology* **2013**, *04* (06), 751-767.

(58) Takahashi, H.; Shimada, I. Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells. *J Biomol NMR* **2010**, *46* (1), 3-10.

(59) Dondapati, S. K.; Stech, M.; Zemella, A.; Kubick, S. Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. *BioDrugs* **2020**, *34* (3), 327-348.

(60) Schneider, B.; Junge, F.; Shirokov, V. A.; Durst, F.; Schwarz, D.; Dotsch, V.; Bernhard, F. Membrane protein expression in cell-free systems. *Methods Mol Biol* **2010**, *601*, 165-186.

(61) Carlson, E. D.; Gan, R.; Hodgman, C. E.; Jewett, M. C. Cell-free protein synthesis: applications come of age. *Biotechnol Adv* **2012**, *30* (5), 1185-1194.

(62) Schörghuber, J. New Concepts in Isotope Labeling for Protein NMR – 13C, 15N and 2H Incorporation with High Selectivity and Efficiency. PhD Thesis, University of Vienna, 2018.

(63) Hofurthner, T.; Toscano, G.; Kontaxis, G.; Beier, A.; Mayer, M.; Geist, L.; McConnell, D. B.; Weinstabl, H.; Lichtenecker, R.; Konrat, R. Synthesis of a (13)C-methylene-labeled isoleucine precursor as a useful tool for studying protein side-chain interactions and dynamics. *J Biomol NMR* **2024**, *78* (1), 1-8.

ChatGPT (OpenAI) was used to assist with language editing and improving clarity of the text.

6. Appendix

6.1. List of Abbreviations

2D-NMR	Two-dimensional NMR
3D-NMR	Three-dimensional NMR
CFPS	Cell-free protein synthesis
COSY	2D correlated spectroscopy
CSP	Chemical shift perturbations
DHA	Dihydroxyacetone
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E. coli	Escherichia coli
EM	Electron microscopy
EPL	Expressed protein ligation
EXSY	Exchange Spectroscopy
FID	Free induction decay
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single-Quantum Coherence
HWE	Horner-Wadsworth-Emmons
IDP	Intrinsically disordered protein
LiHMDS	Lithium bis(trimethylsilyl)amide
mRNA	Messenger ribonucleic acid
NCL	Native chemical ligation
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NTP	Nucleoside triphosphates
PTM	Posttranslational modification
PTS	Protein trans-splicing

pTSA	*Para*-Toluenesulfonic acid
RT	Room temperature
SAR	Structure-activity relationship
SOFAST	band-Selective Optimized-Flip-Angle Short-Transient
STD	Saturation transfer difference
T_1	Longitudinal relaxation time
T_2	Transverse relaxation time
TBSCl	*tert*-Butyldimethylsilyl chloride
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TOCSY	Total correlation spectroscopy
TROSY	Transverse relaxation-optimized spectroscopy
tRNA	Transfer ribonucleic acid
WaterLOGSY	Water-Ligand Observed via Gradient Spectroscopy

6.2. NMR Spectra

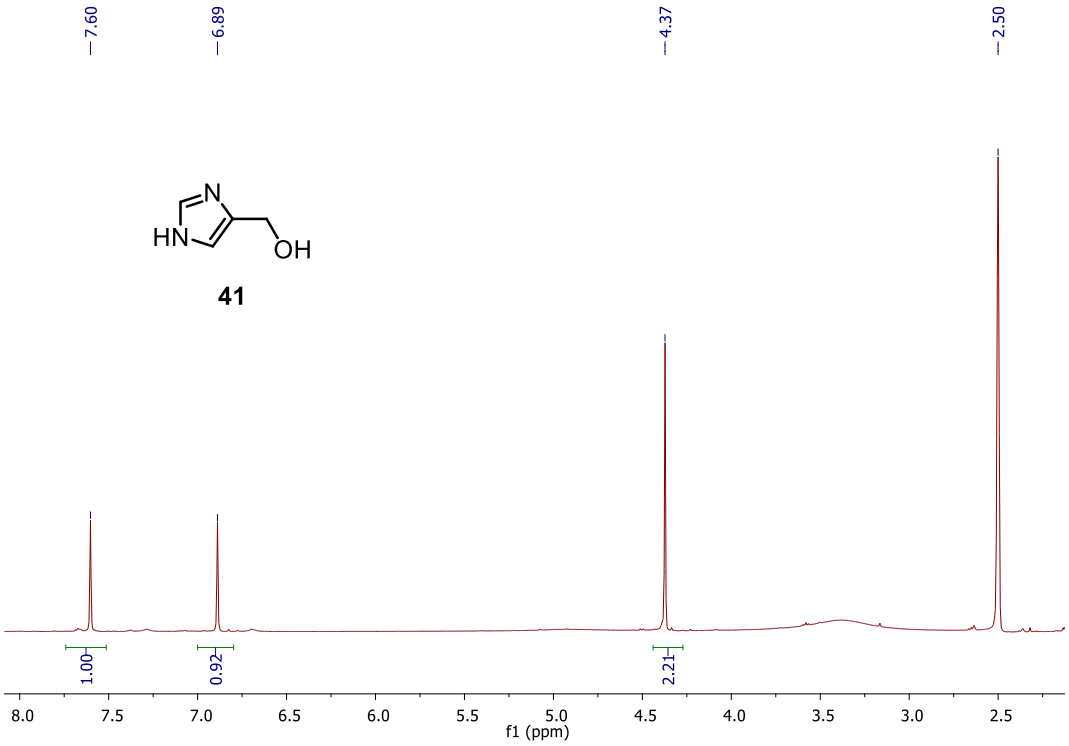


*Figure 5: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **41**.*

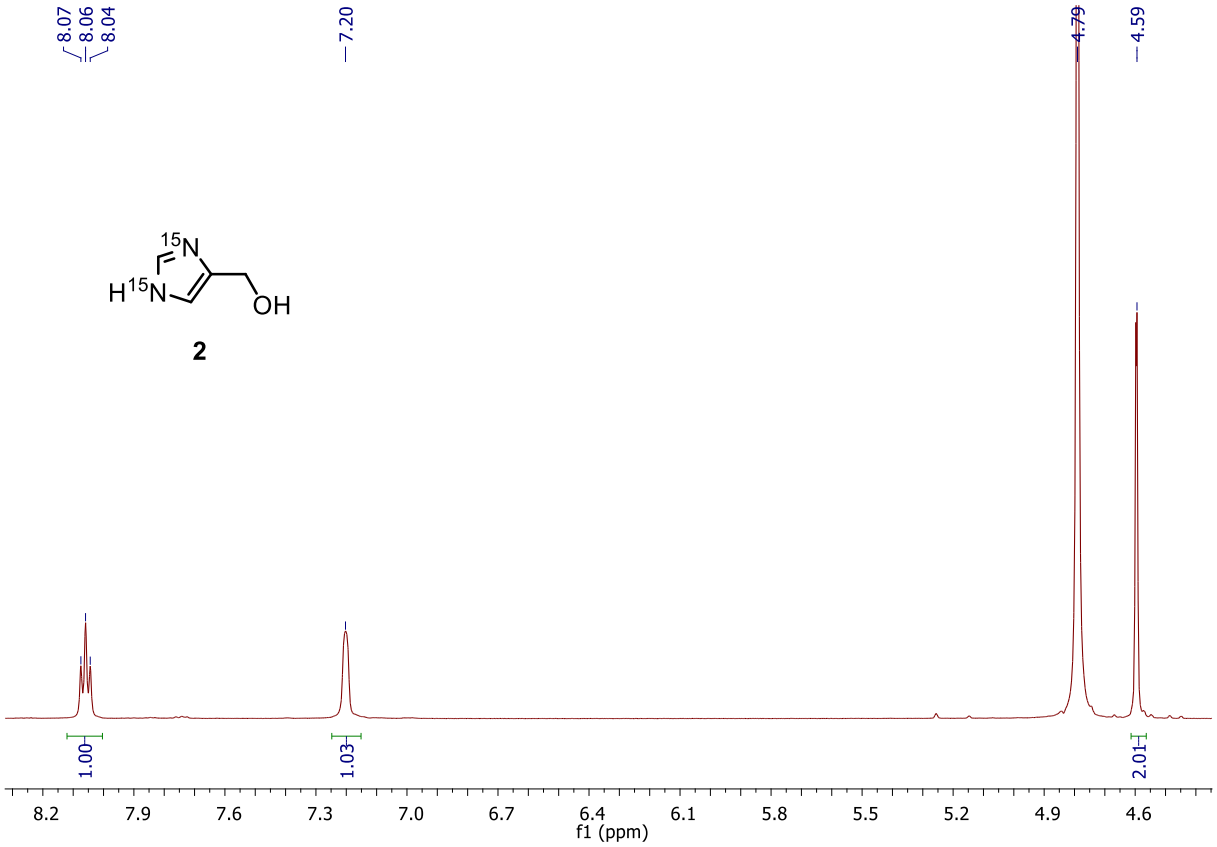


*Figure 6: ^{1}H-NMR (500 MHz, D_2O) of compound **2**.*

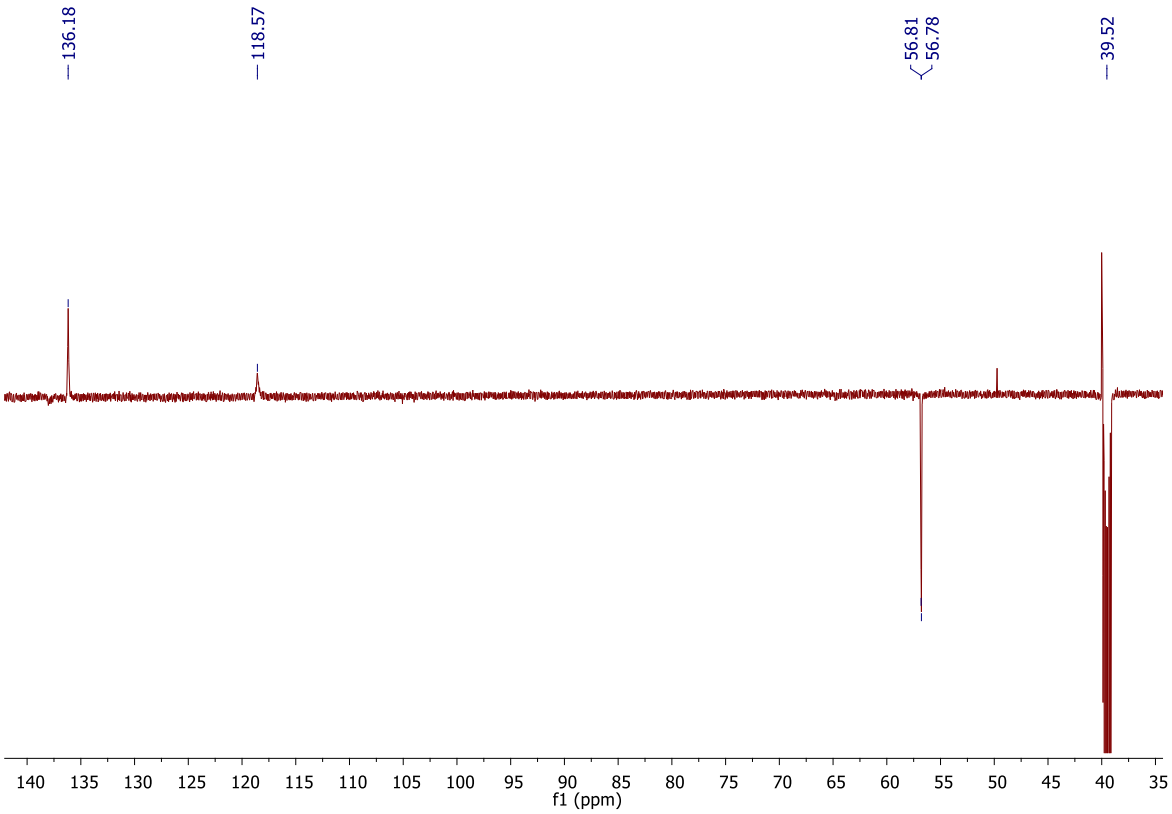


*Figure 7: ^{13}C-NMR (176 MHz, 6d-DMSO) of compound **2**.*

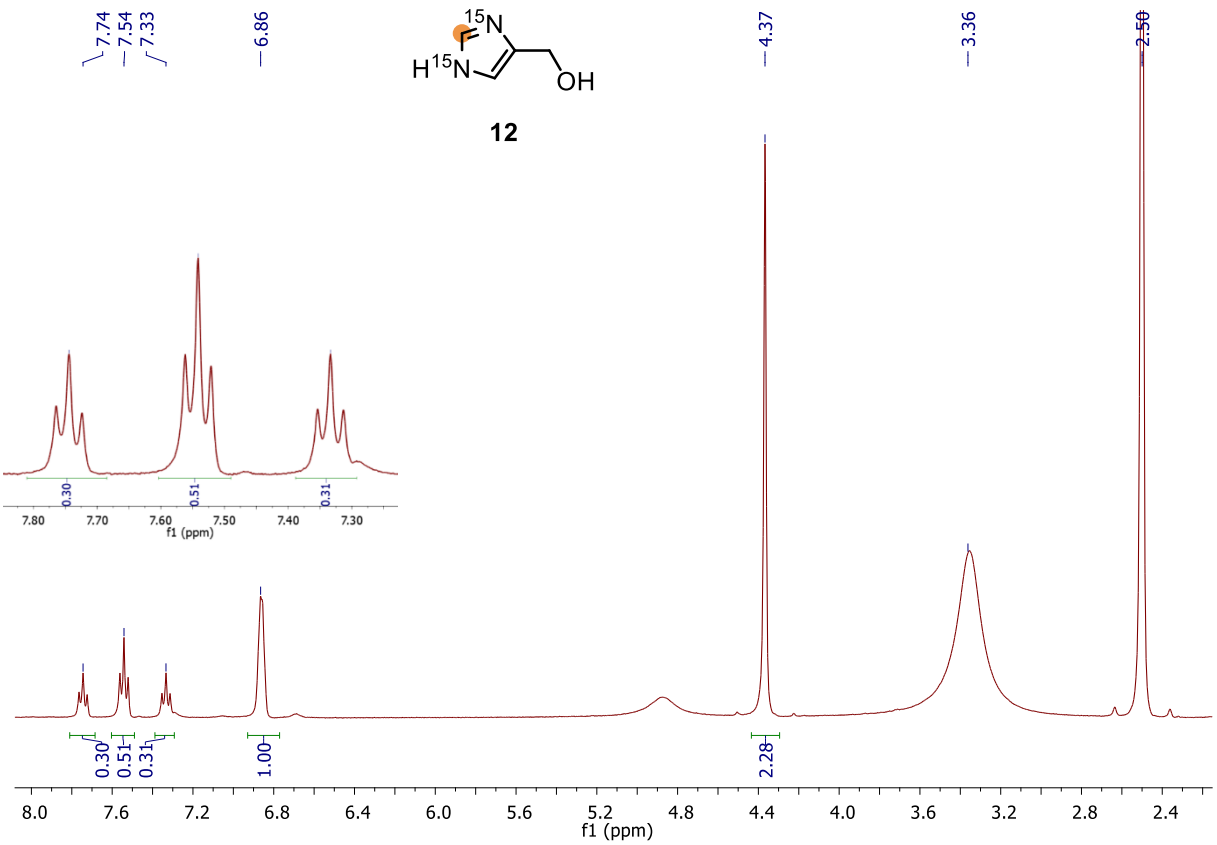


*Figure 8: [1]H-NMR (500 MHz, 6d-DMSO) of compound **12a**.*

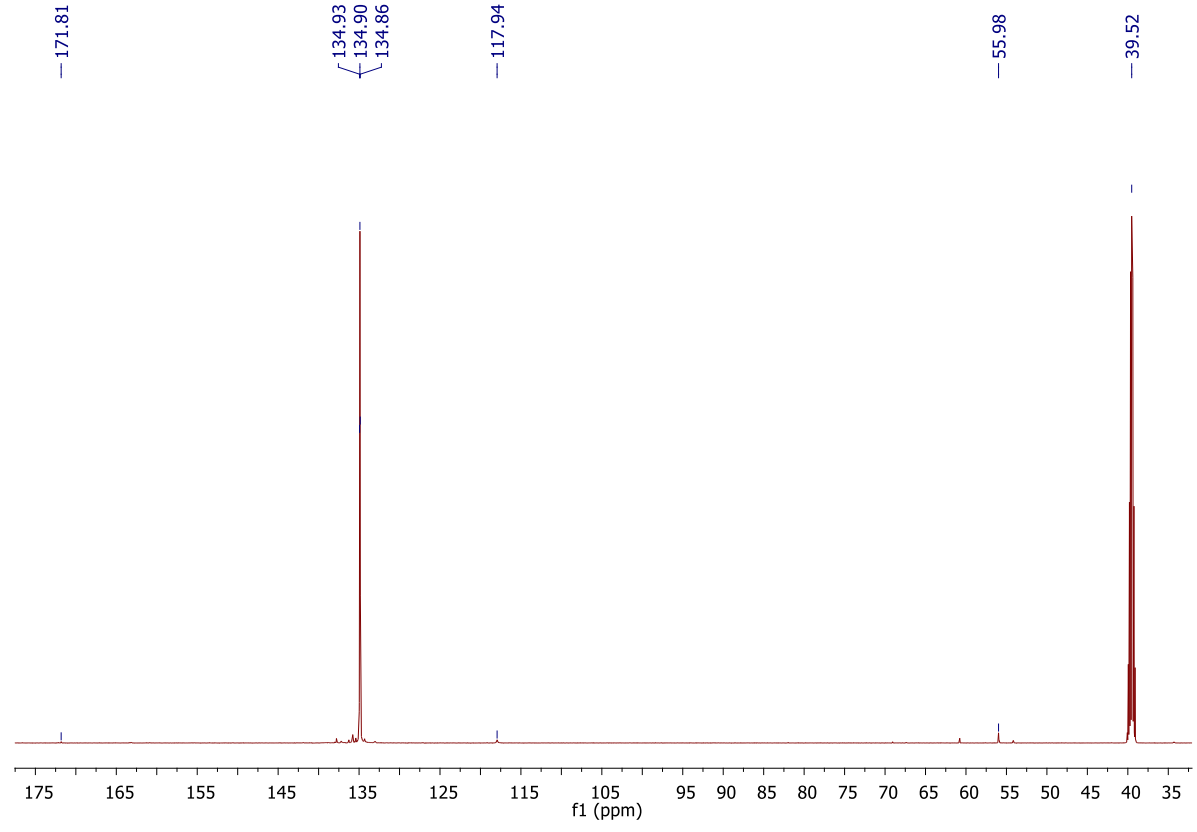


*Figure 9: [13]C-NMR (150.9 MHz, 6d-DMSO) of compound **12a**.*

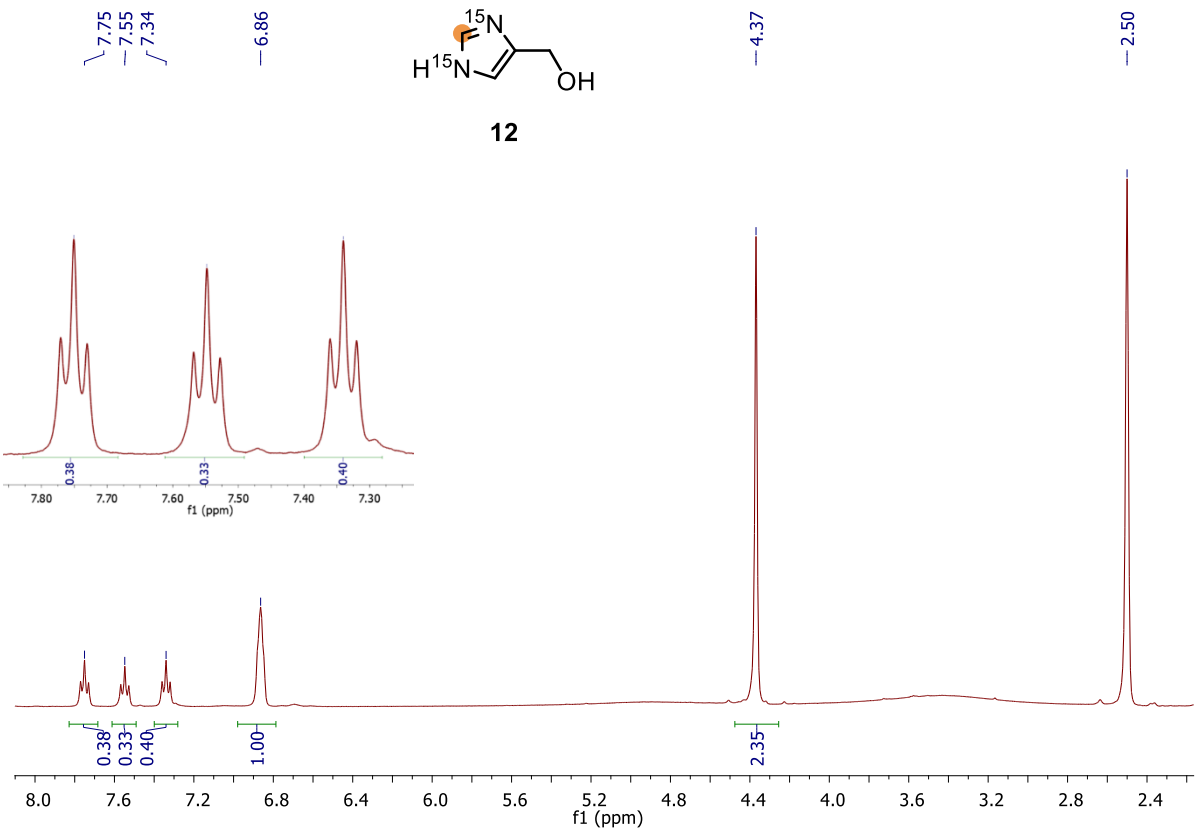


*Figure 10: [1]H-NMR (500 MHz, 6d-DMSO) of compound **12b**.*

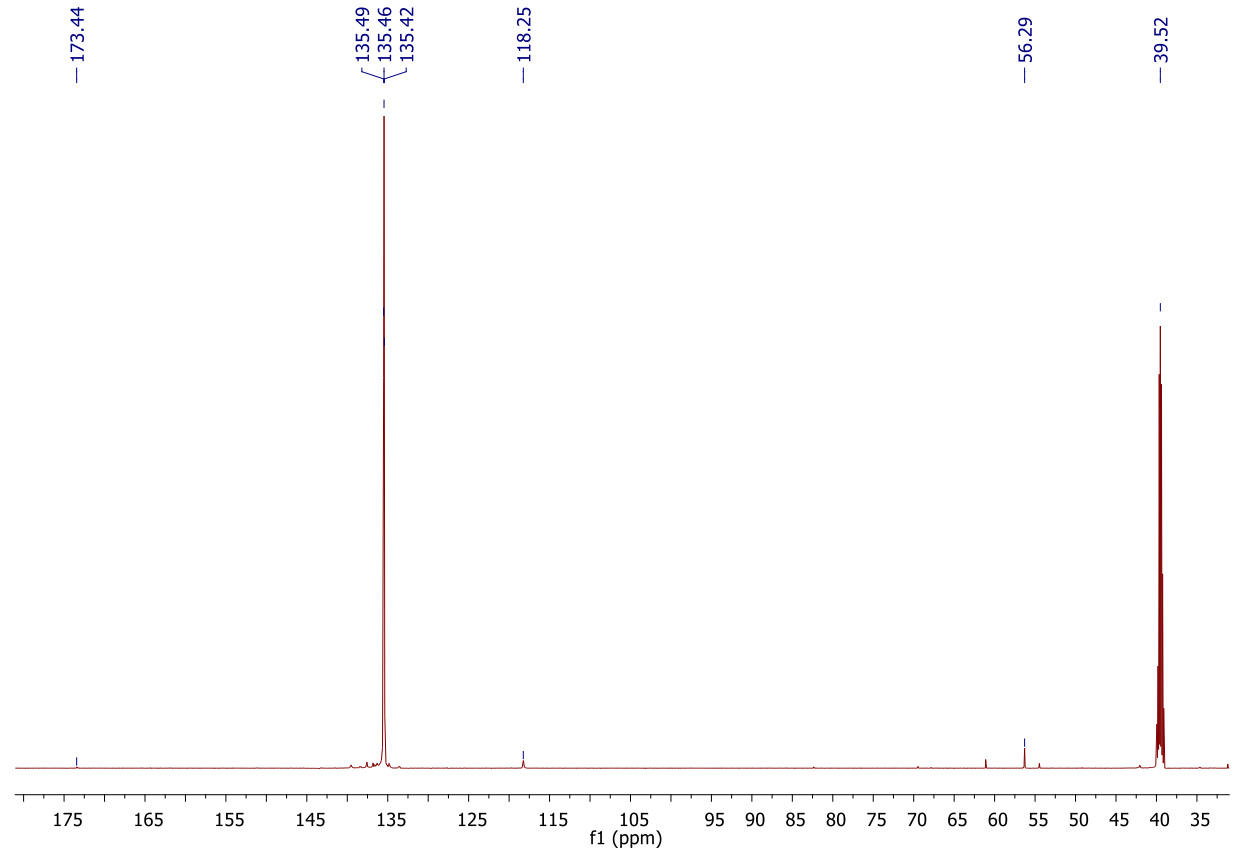


*Figure 11: [13]C-NMR (150.9 MHz, 6d-DMSO) of compound **12b**.*

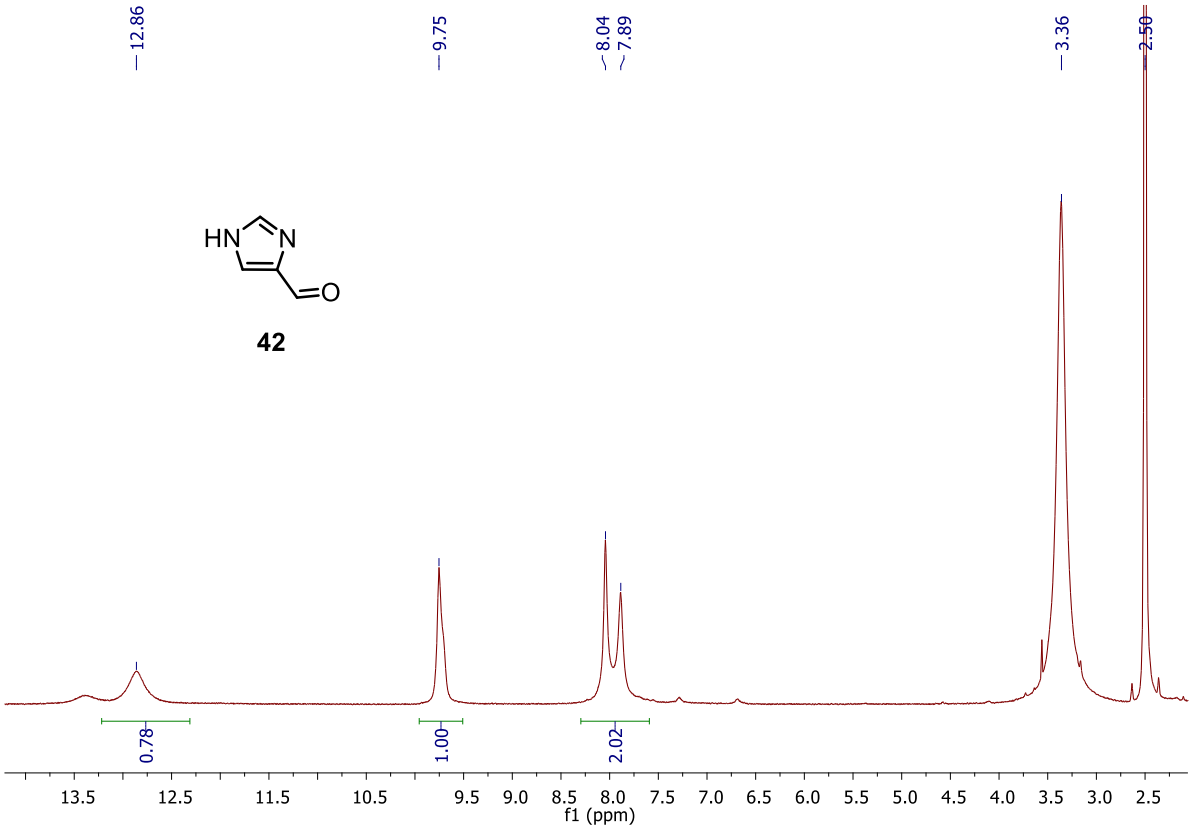


*Figure 12: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **42**.*

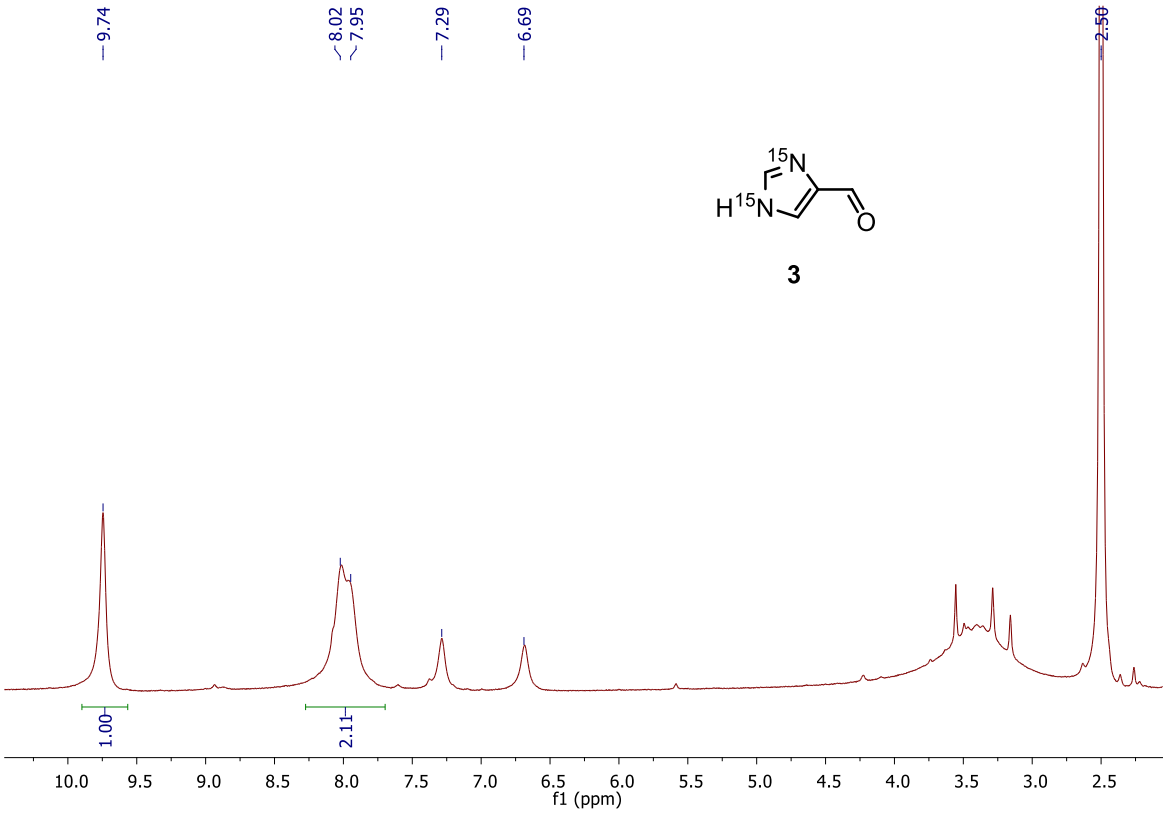


*Figure 13: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **3**.*

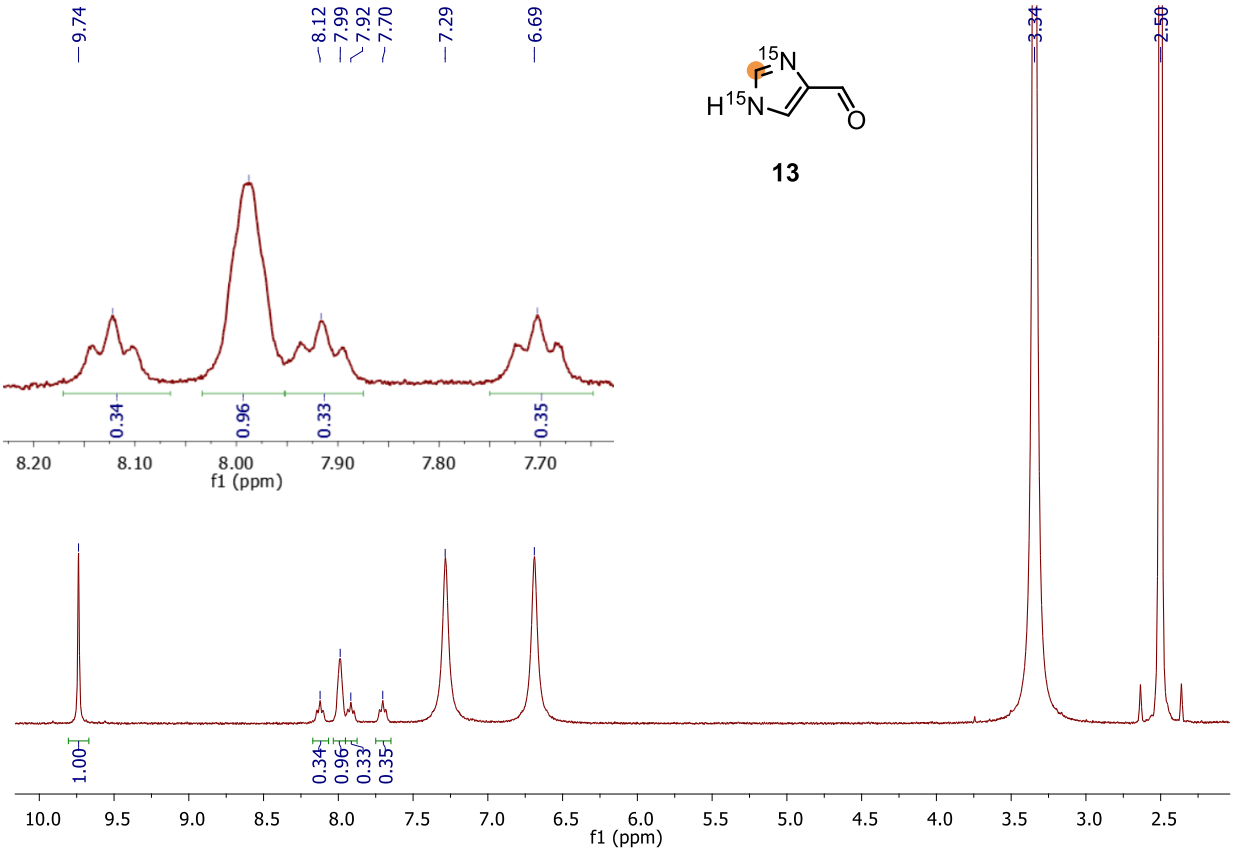


*Figure 14: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **13**.*

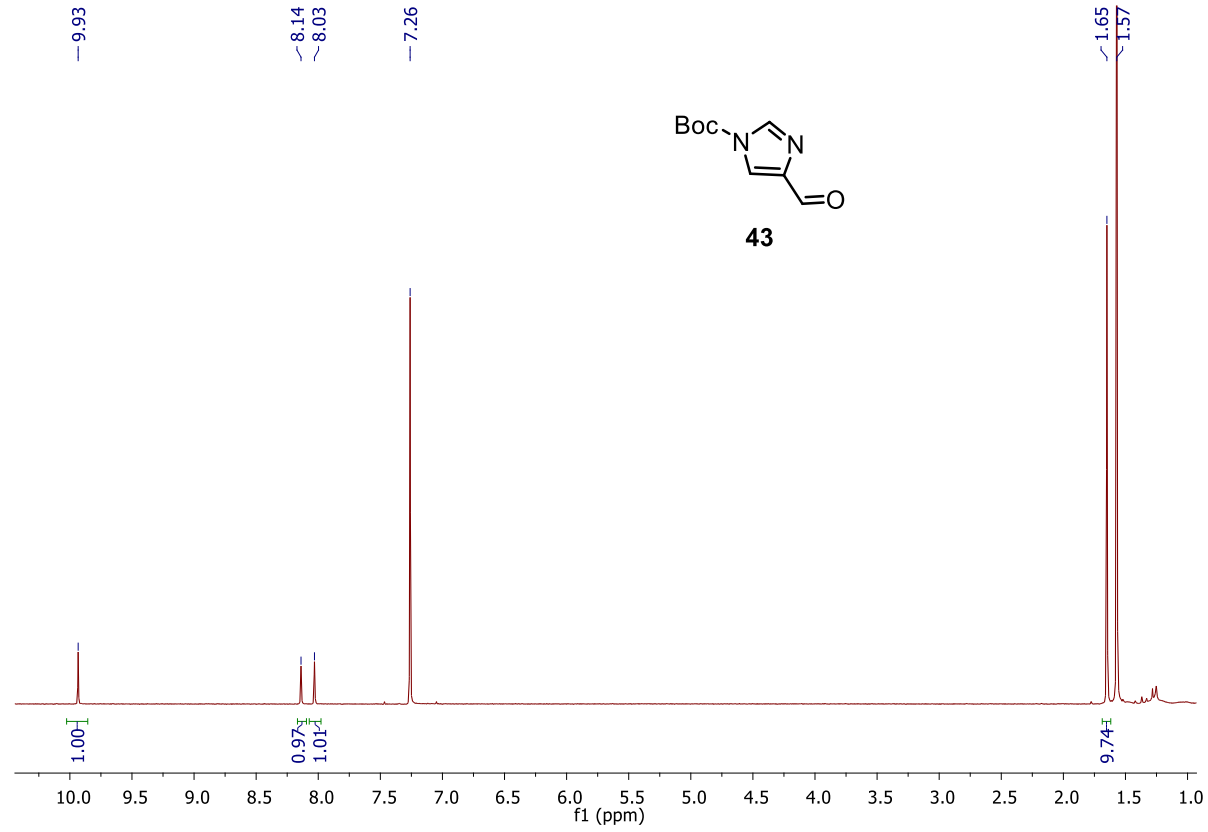


*Figure 15: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **43**.*

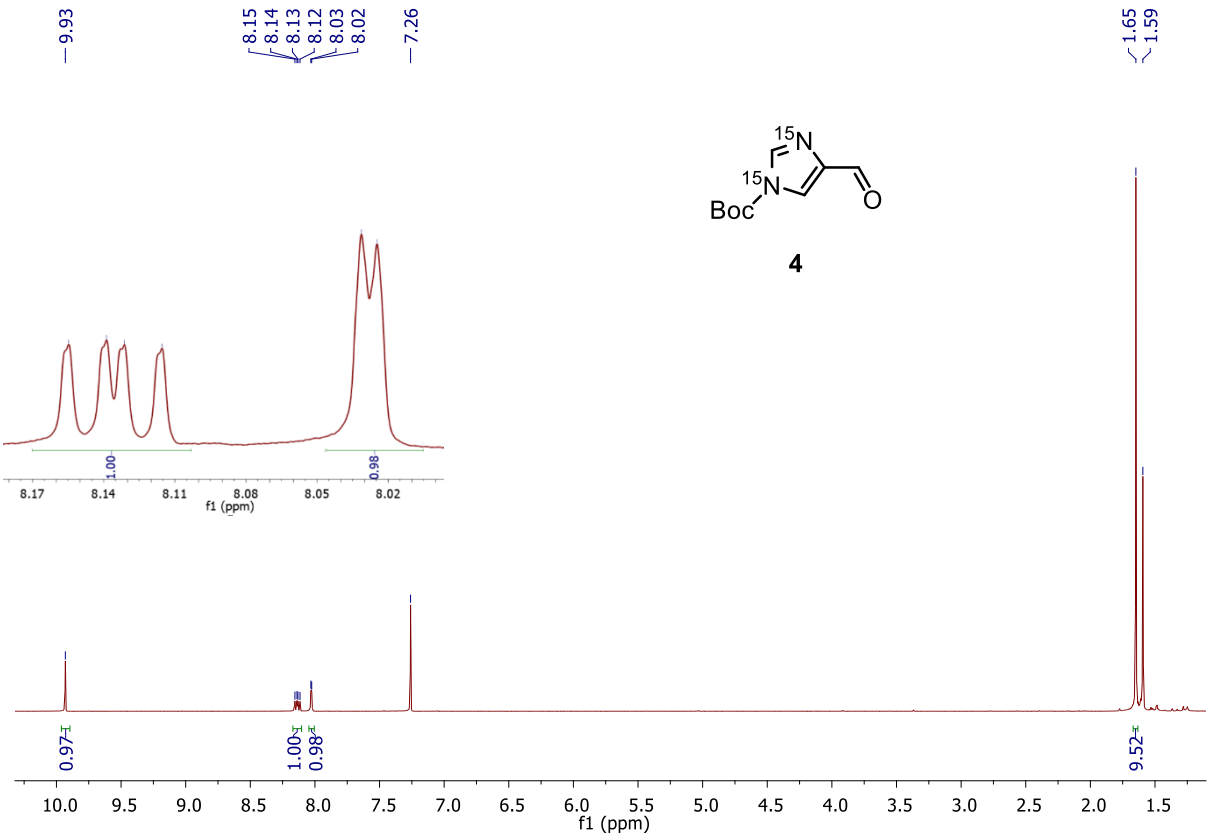


*Figure 16: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **4**.*

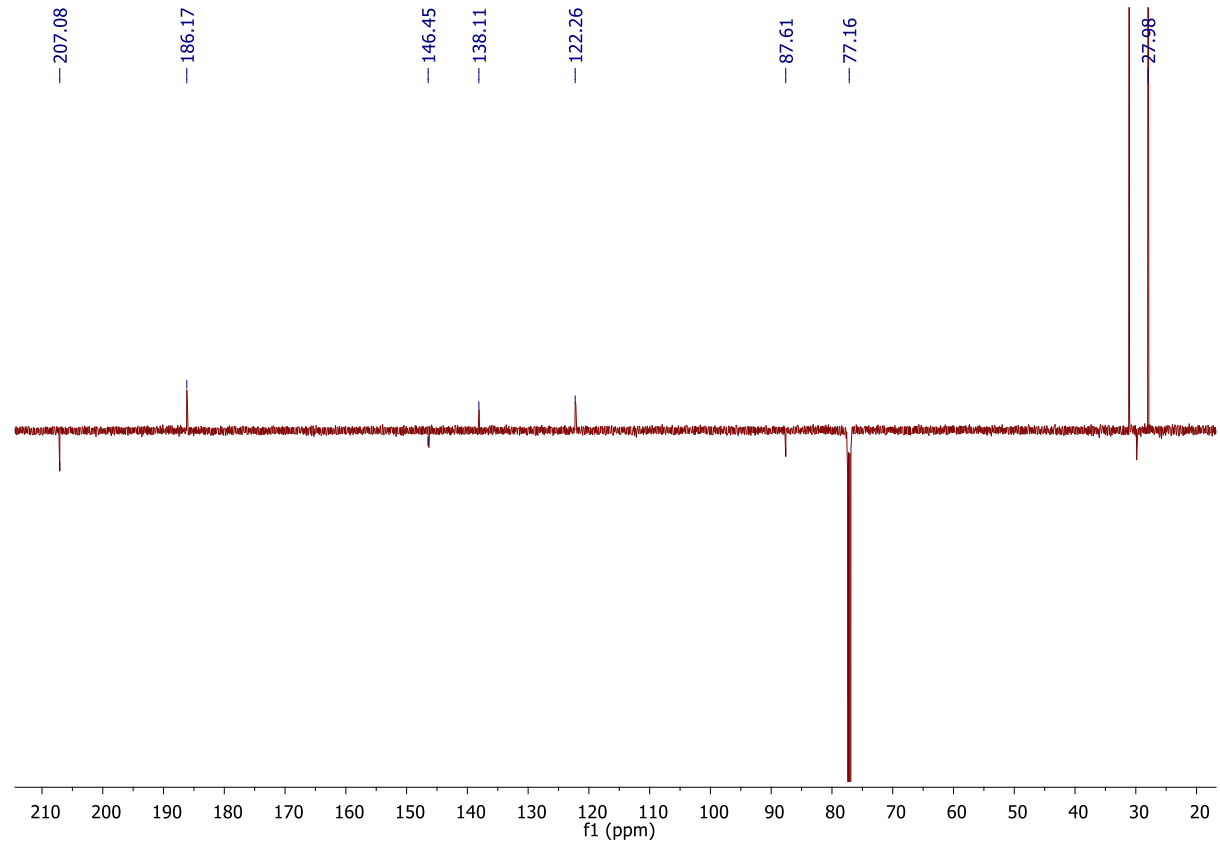


*Figure 17: ^{13}C-NMR (150.9 MHz, $CDCl_3$) of compound **4**.*

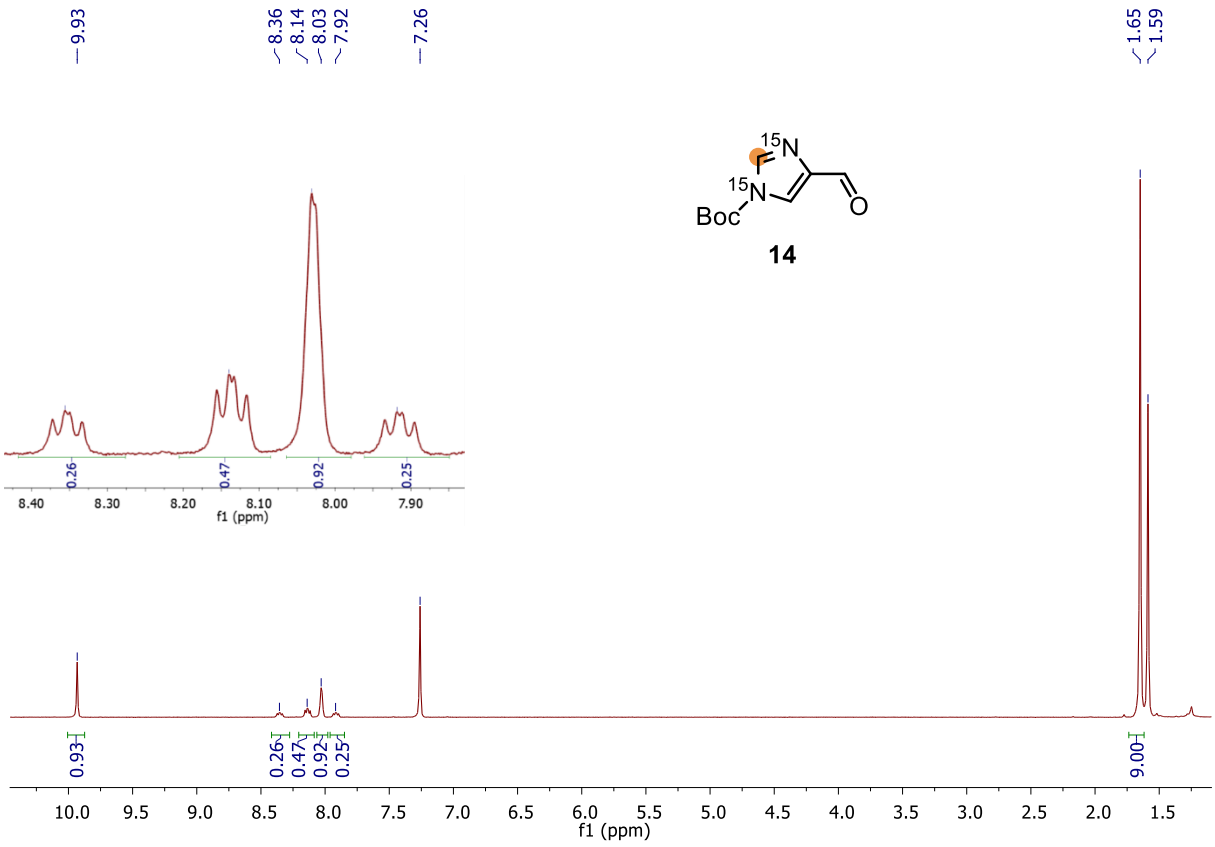


*Figure 18: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **14**.*

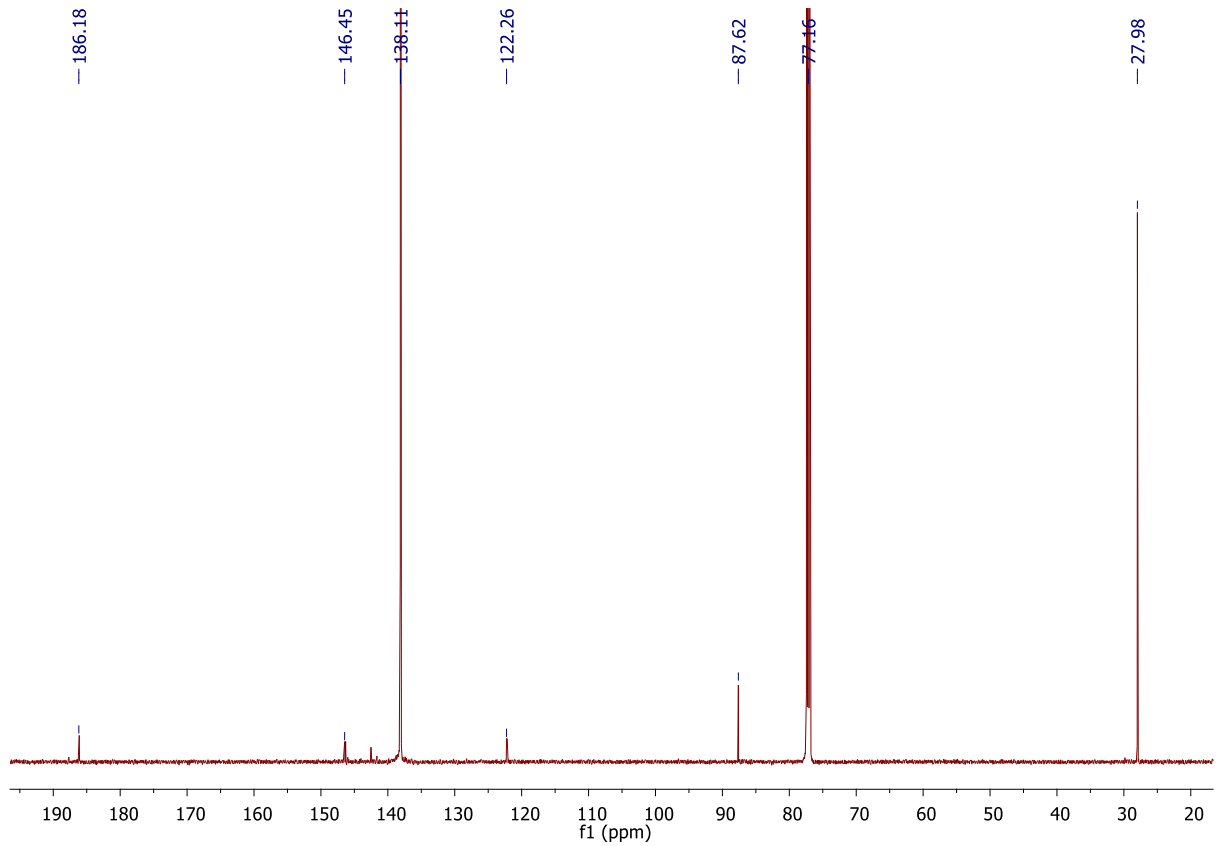


*Figure 19: ^{13}C-NMR (150.9 MHz, $CDCl_3$) of compound **14**.*

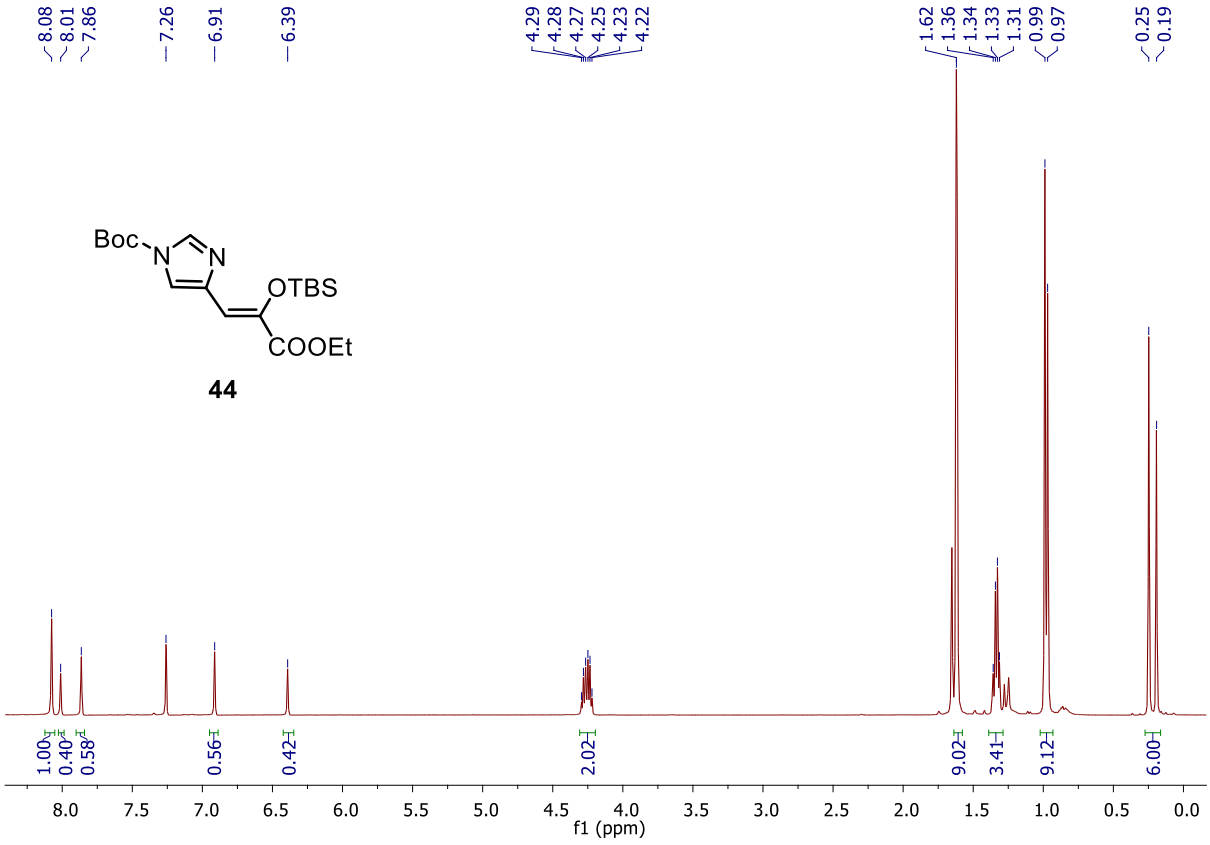


*Figure 20: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **44**.*

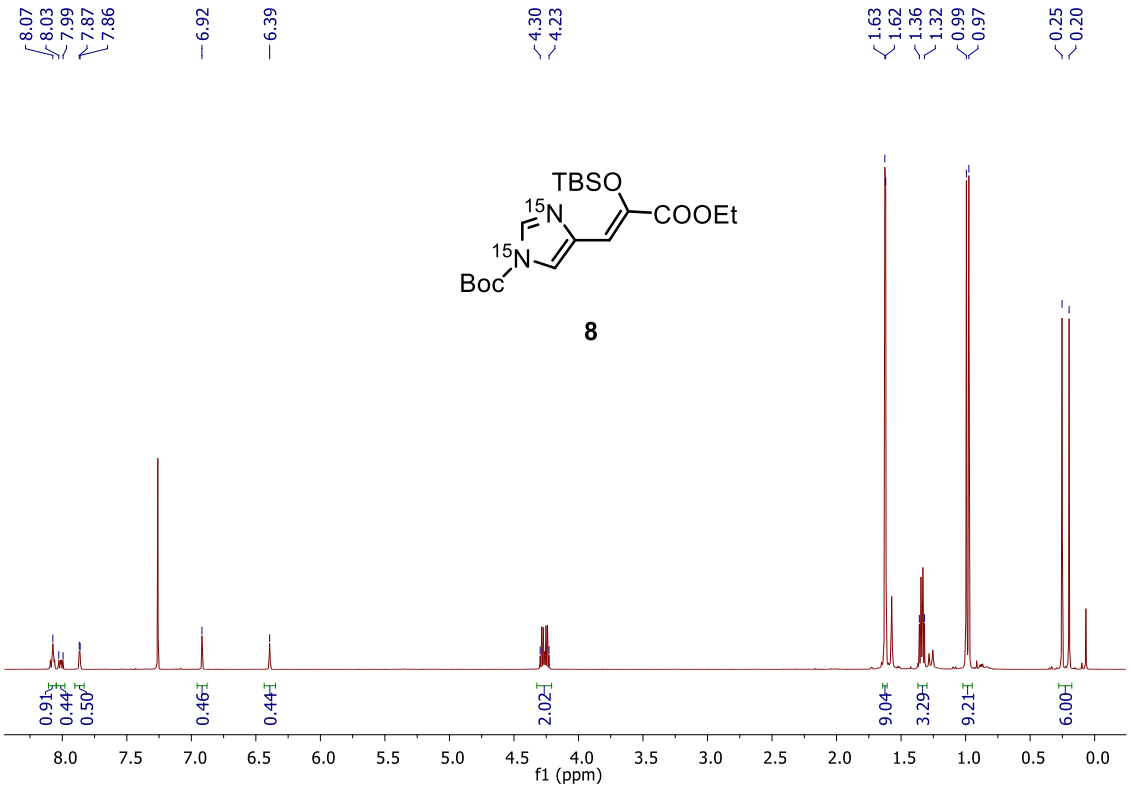


*Figure 21: ^{1}H-NMR ($CDCl_3$) of compound **8**.*

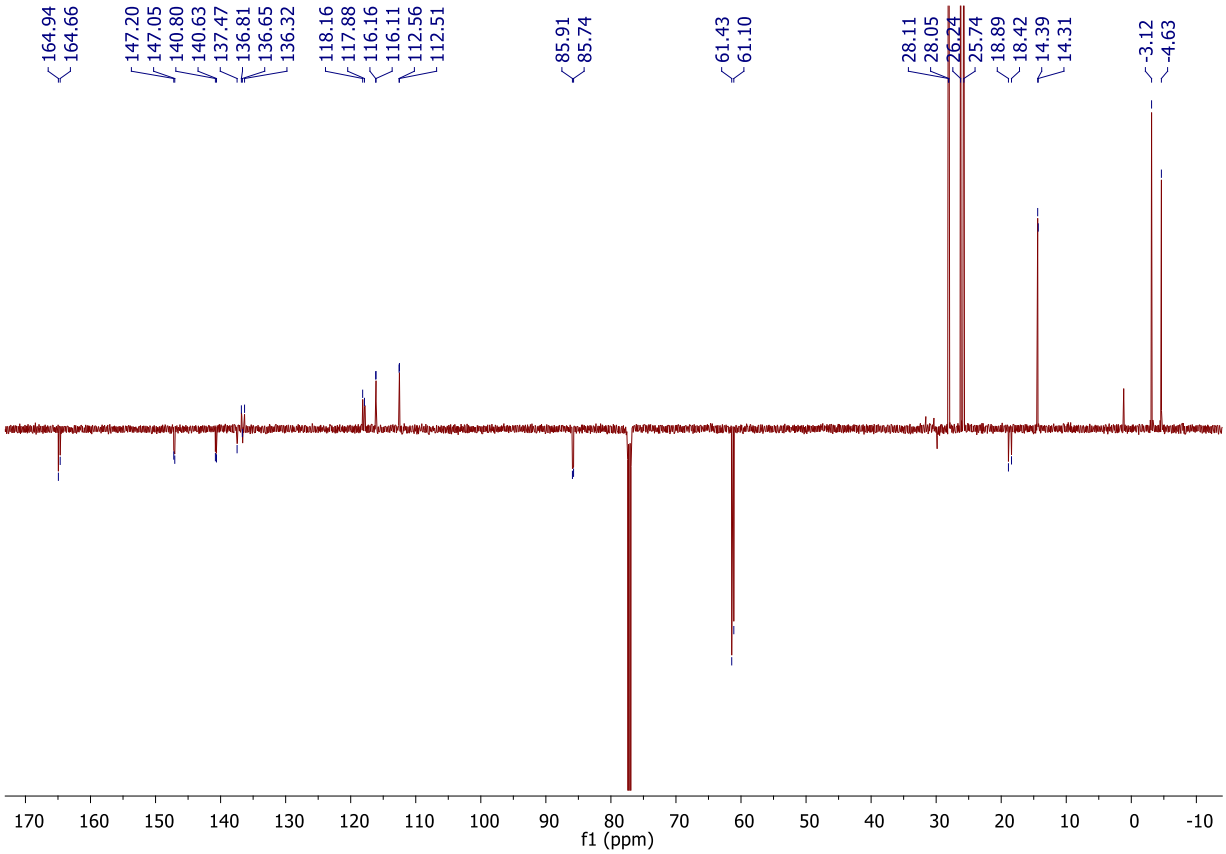


*Figure 22: ^{13}C-NMR (150.9 MHz, $CDCl_3$) of compound **8**.*

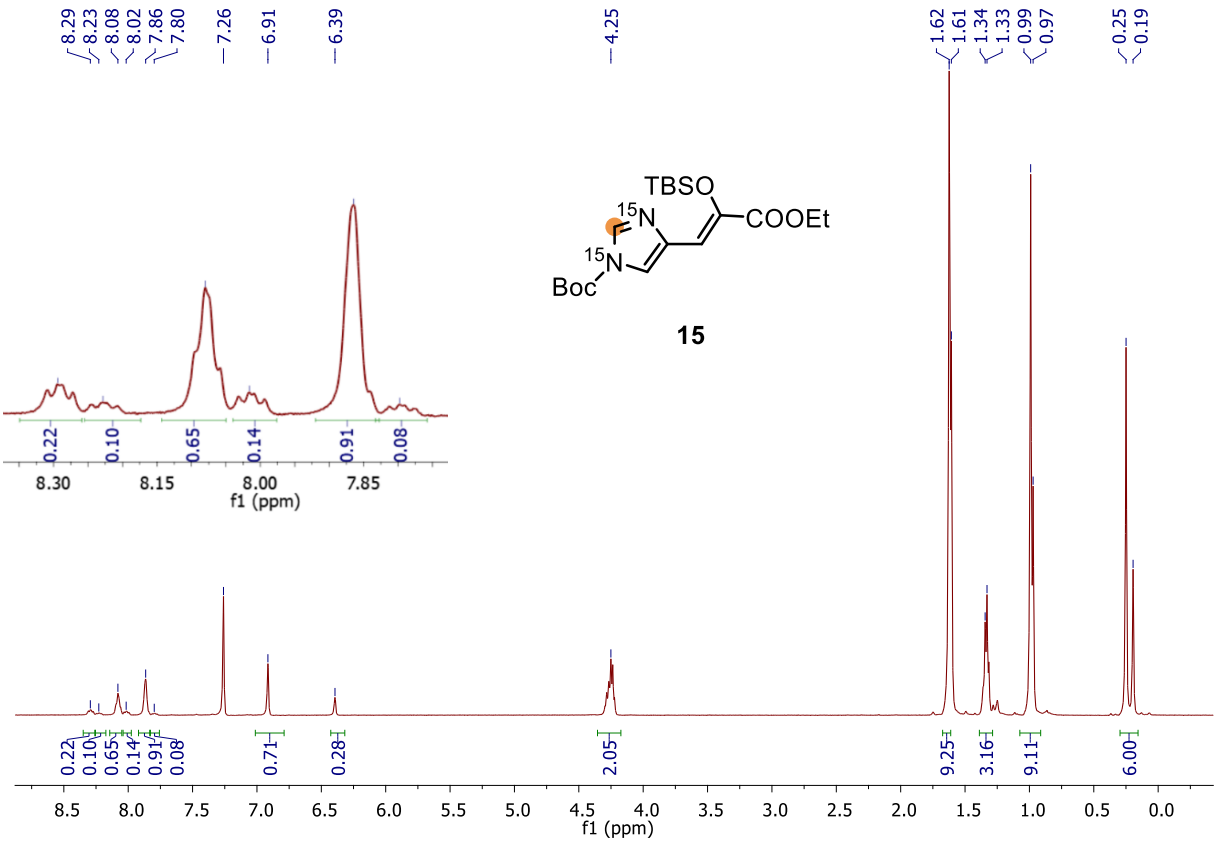


*Figure 23: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **15**.*

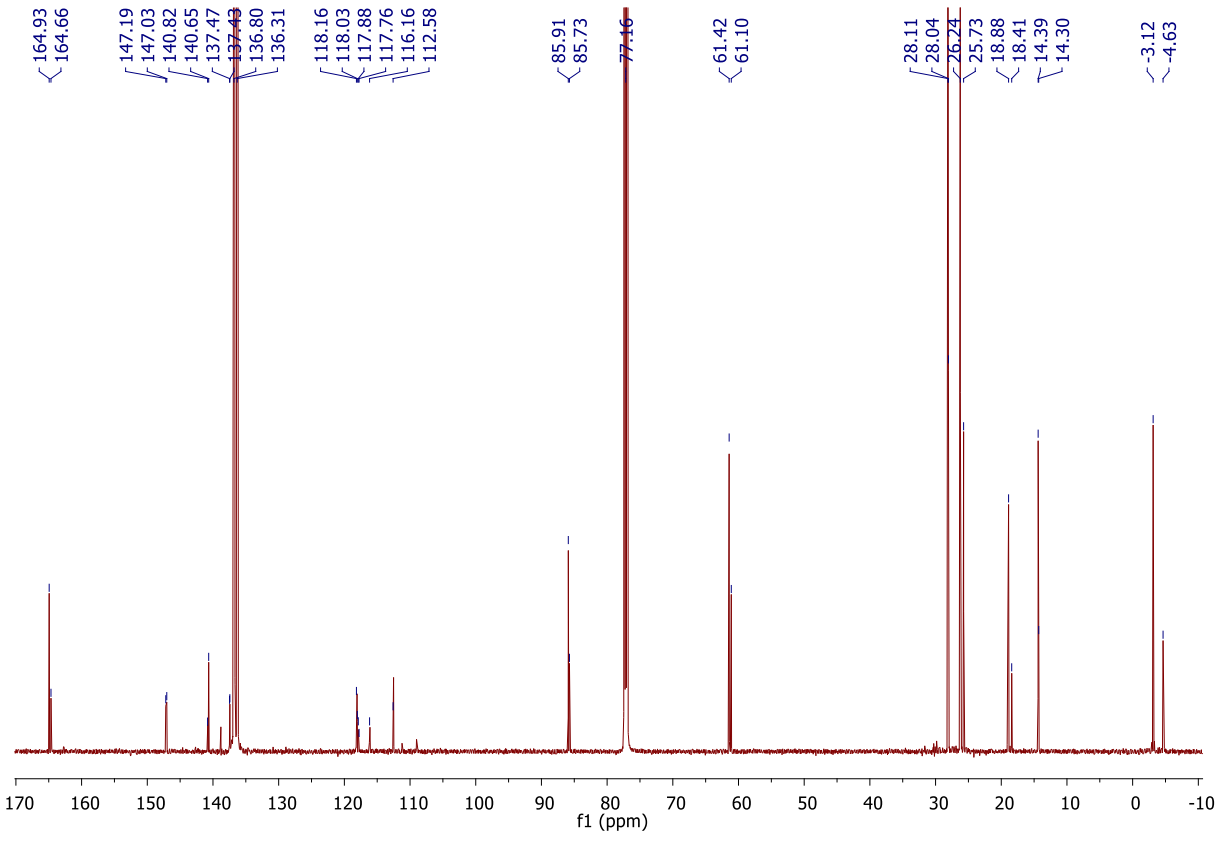


*Figure 24: ^{13}C-NMR (150.9 MHz, $CDCl_3$) of compound **15**.*

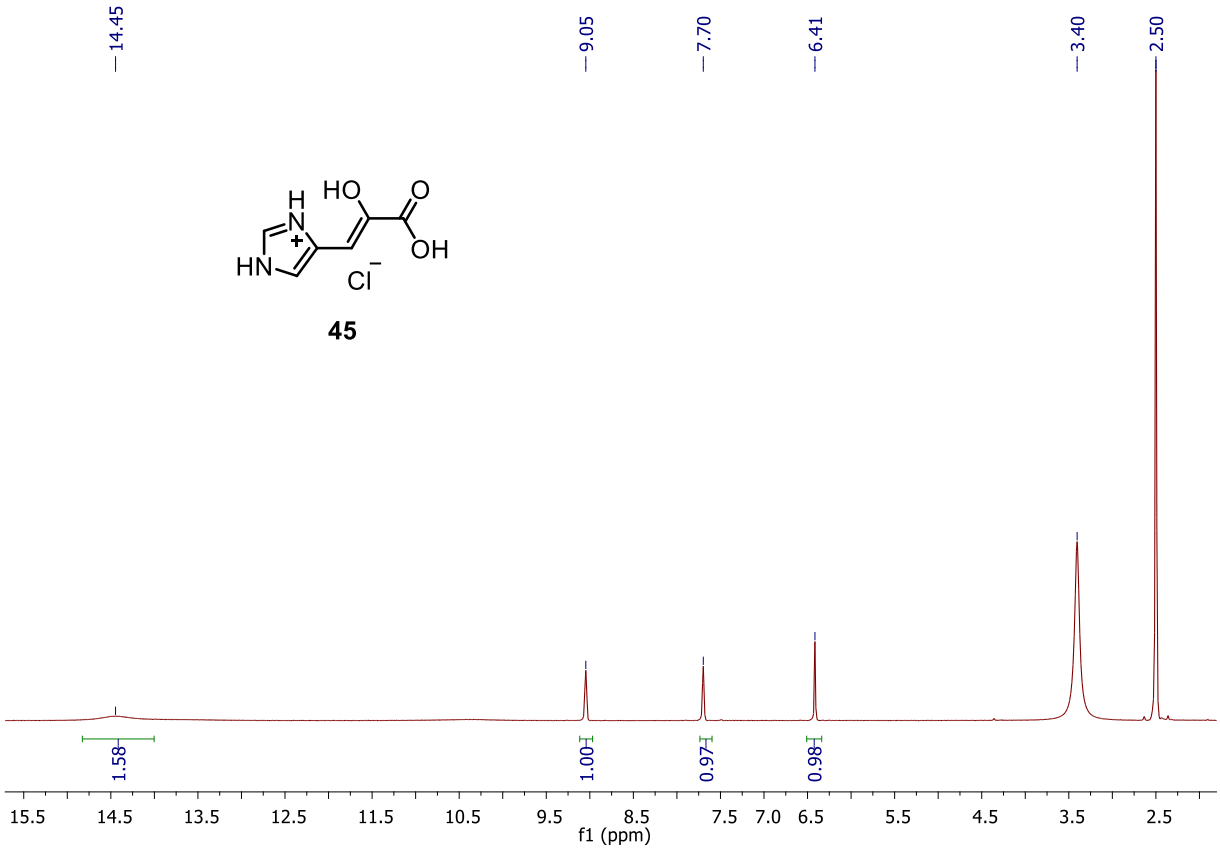
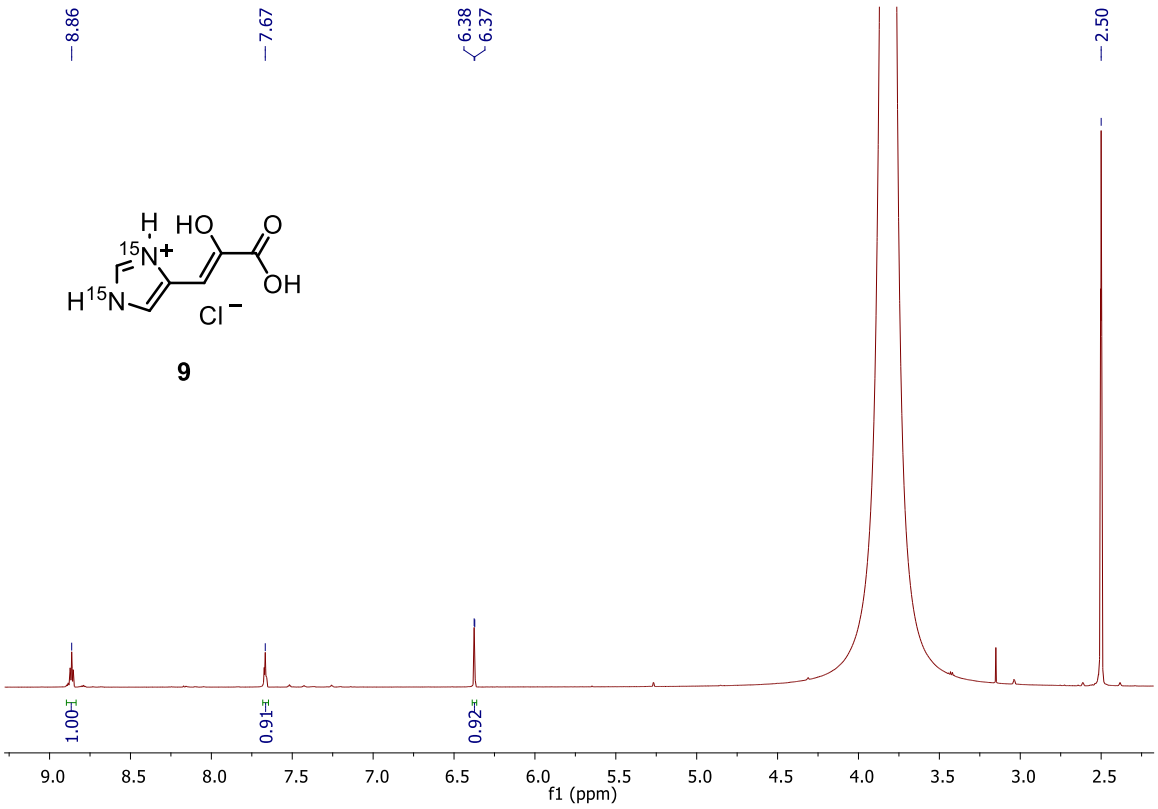


*Figure 25: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **45**.*



9

*Figure 26: [1]H-NMR (500 MHz, 6d-DMSO) of compound **9**.*

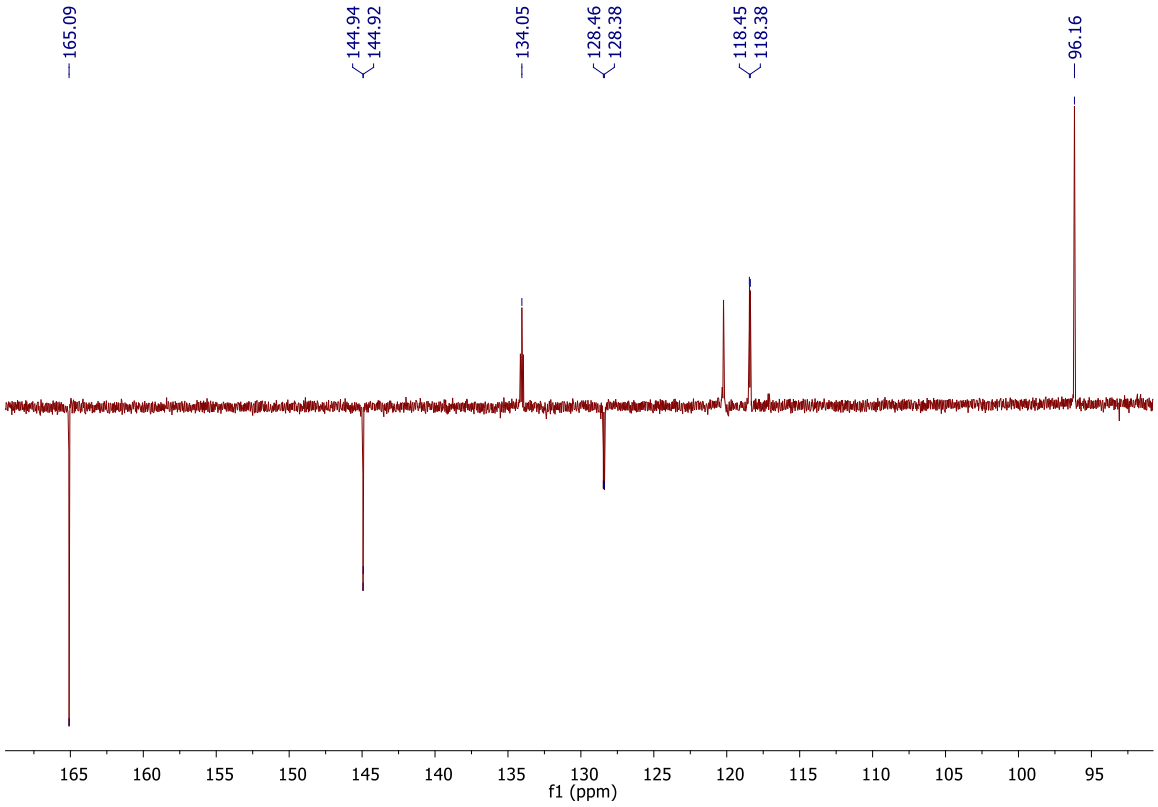


*Figure 27: [13]C-NMR (150.9 MHz, 6d-DMSO) of compound **9**.*

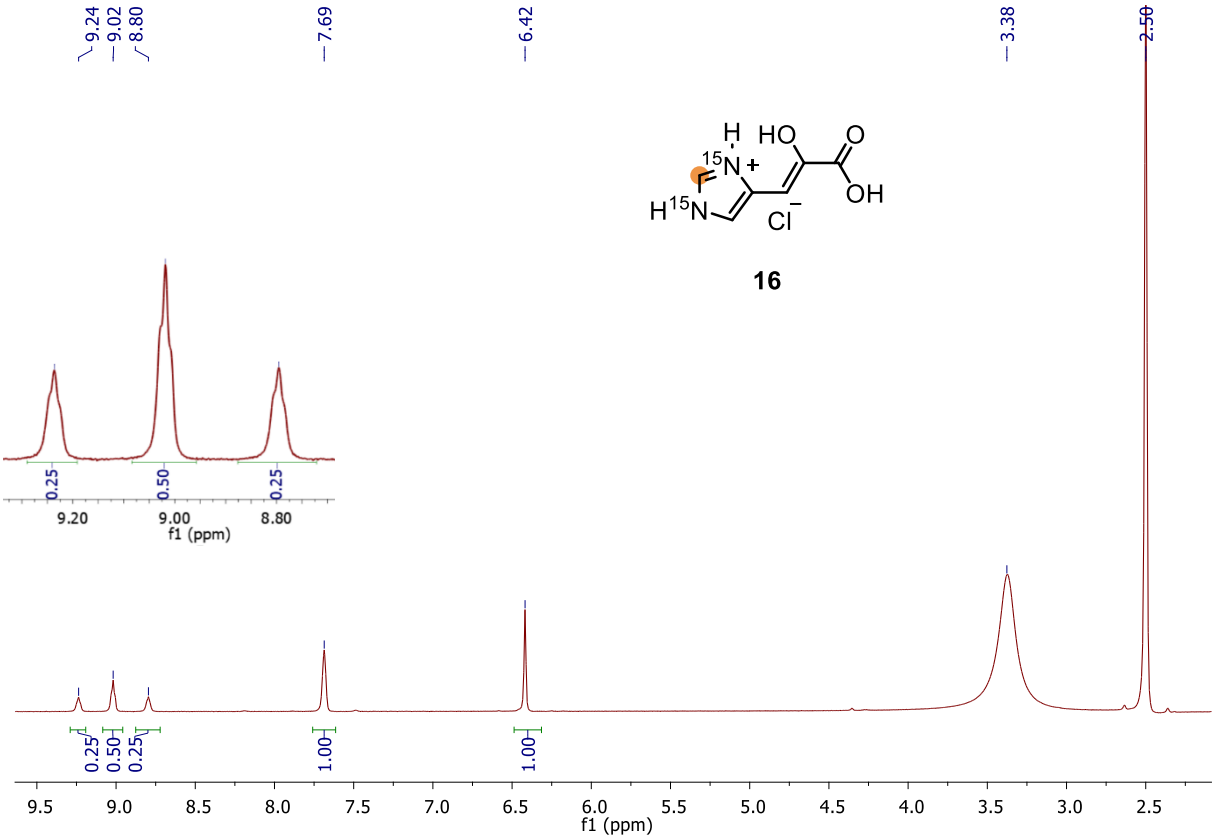


*Figure 28: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **16**.*

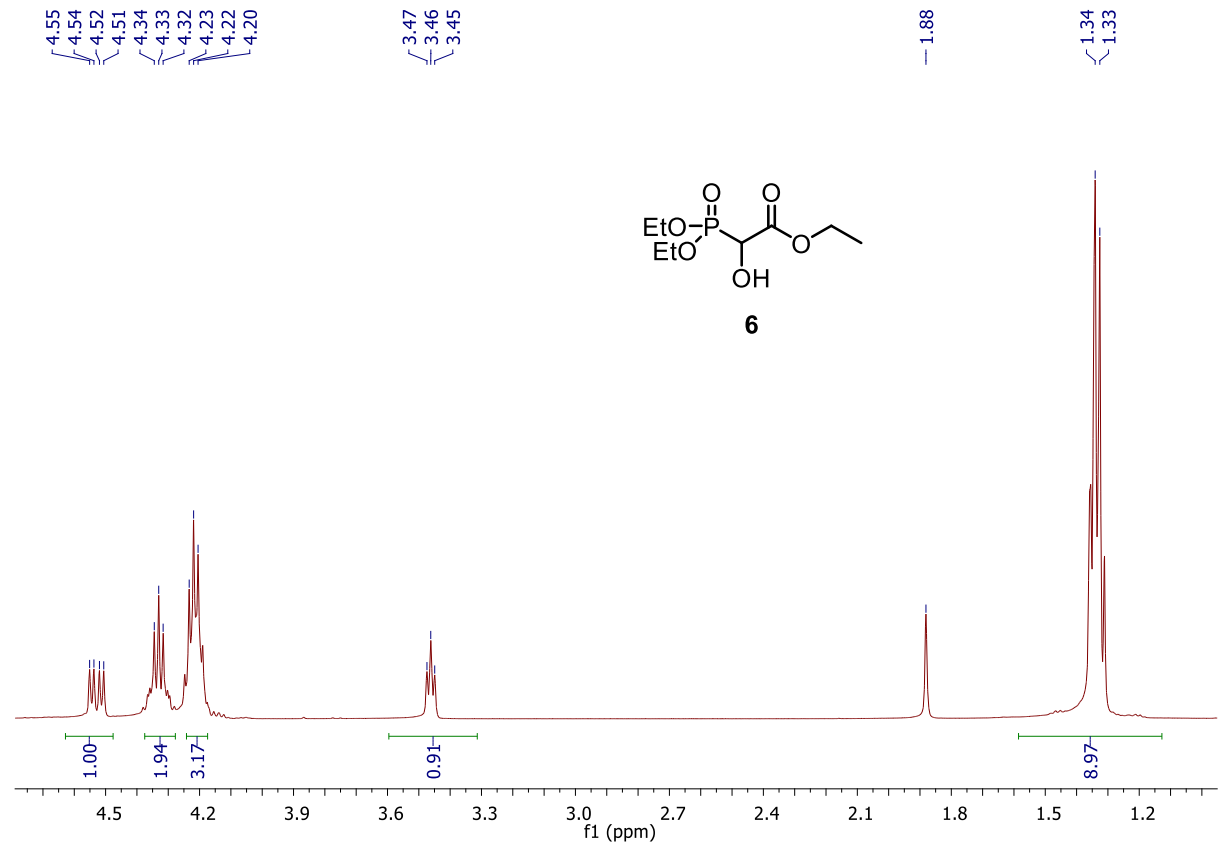


*Figure 29: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **6**.*

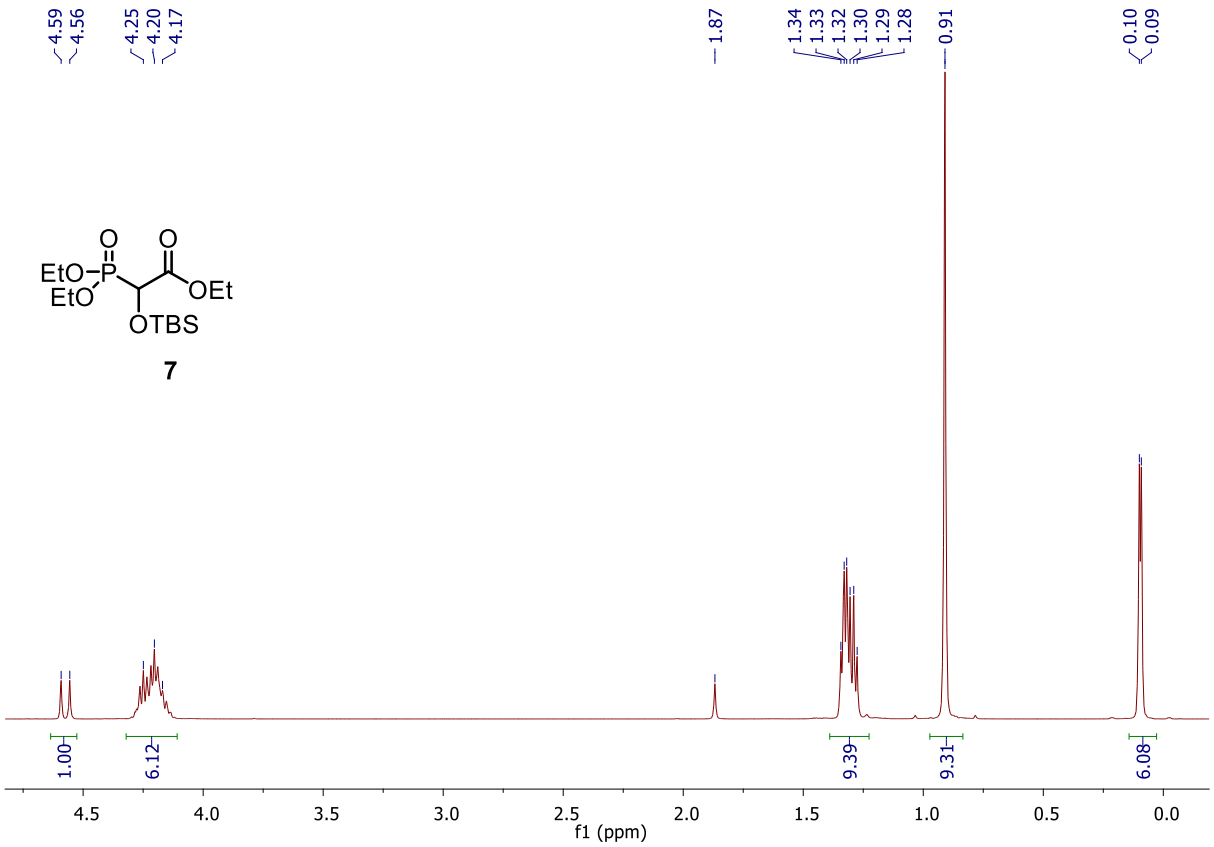


*Figure 30: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **7**.*

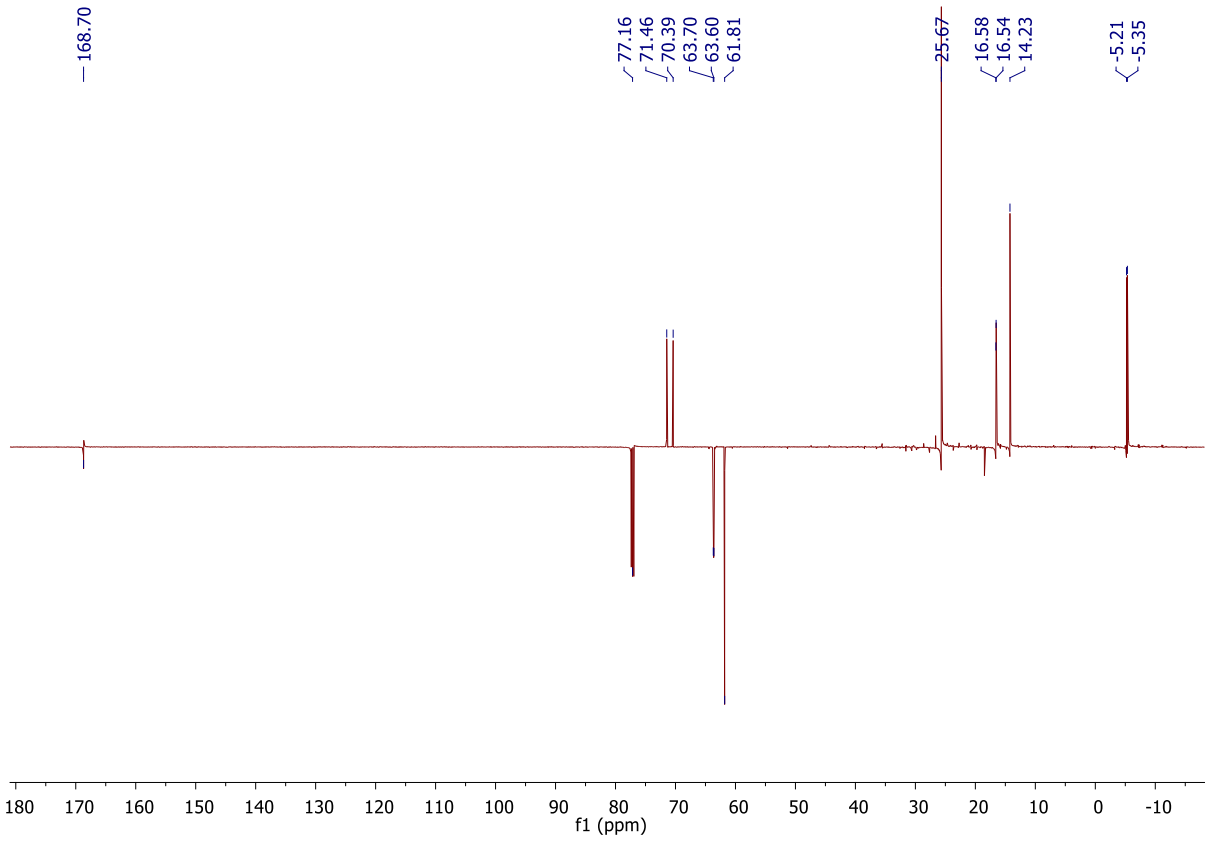


*Figure 31: ^{13}C-NMR (150.9 MHz, $CDCl_3$) of compound **7**.*

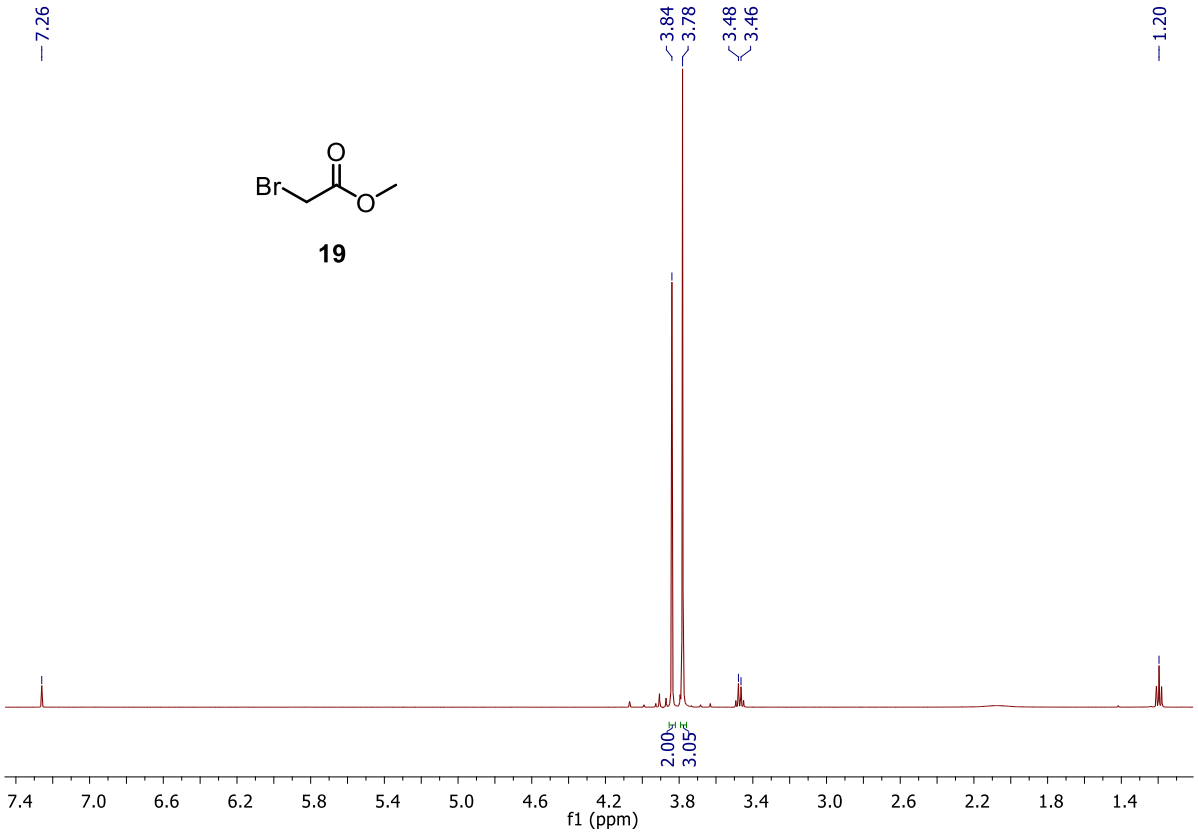


*Figure 32: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **19**.*

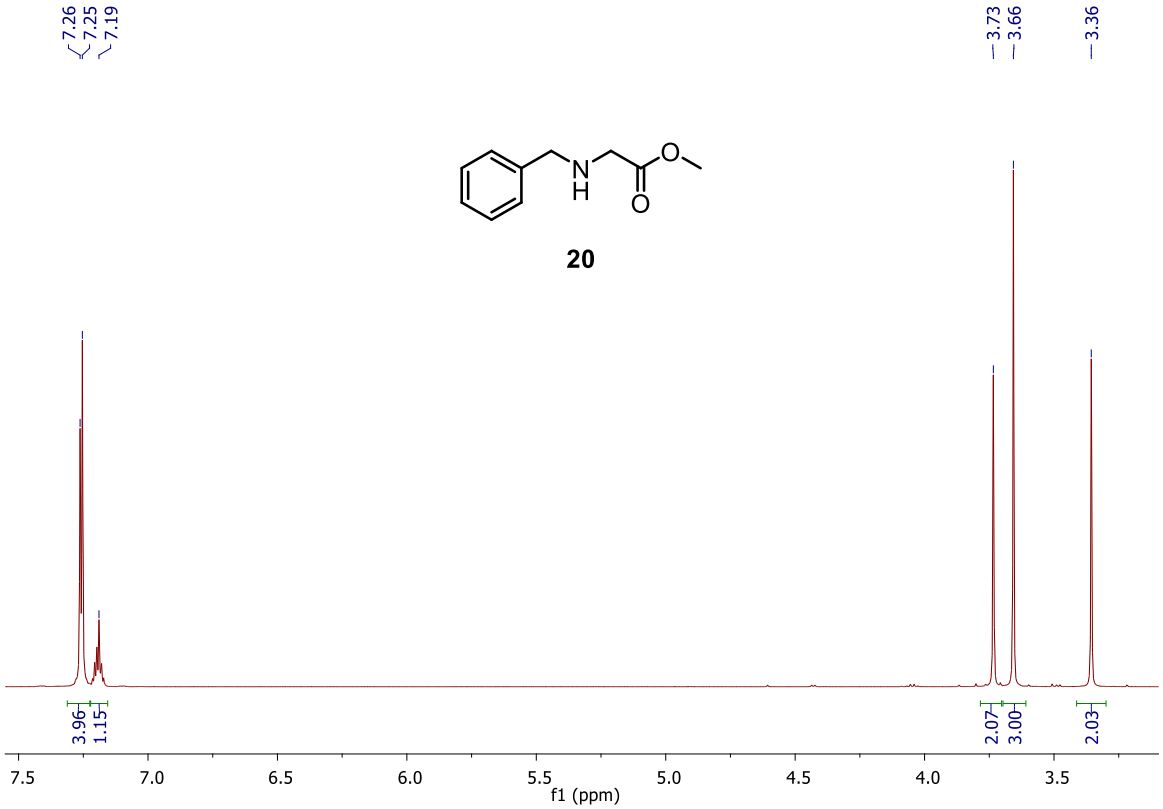


*Figure 33: ^{1}H-NMR (500 MHz, $CDCl_3$) of compound **20**.*

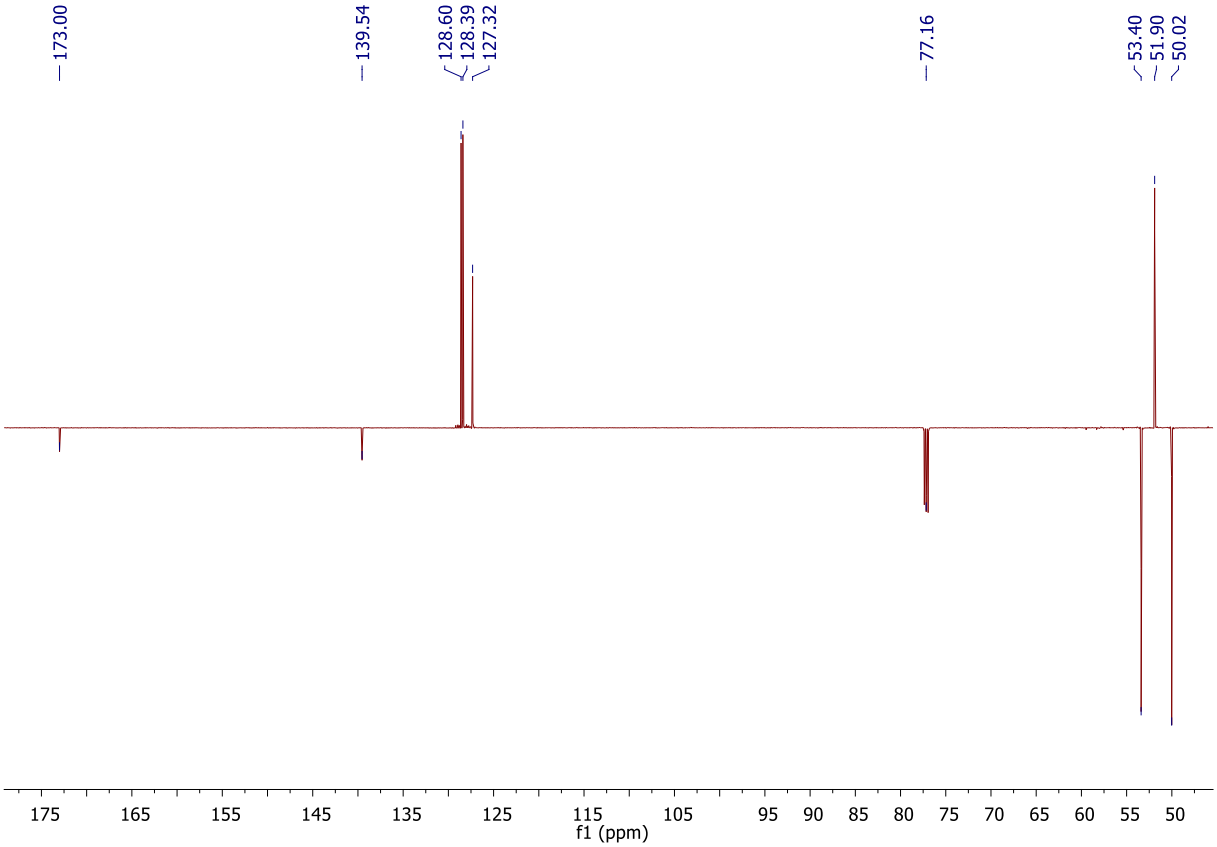


*Figure 34: ^{13}C-NMR (150.9 MHz, $CDCl_3$) of compound **20**.*

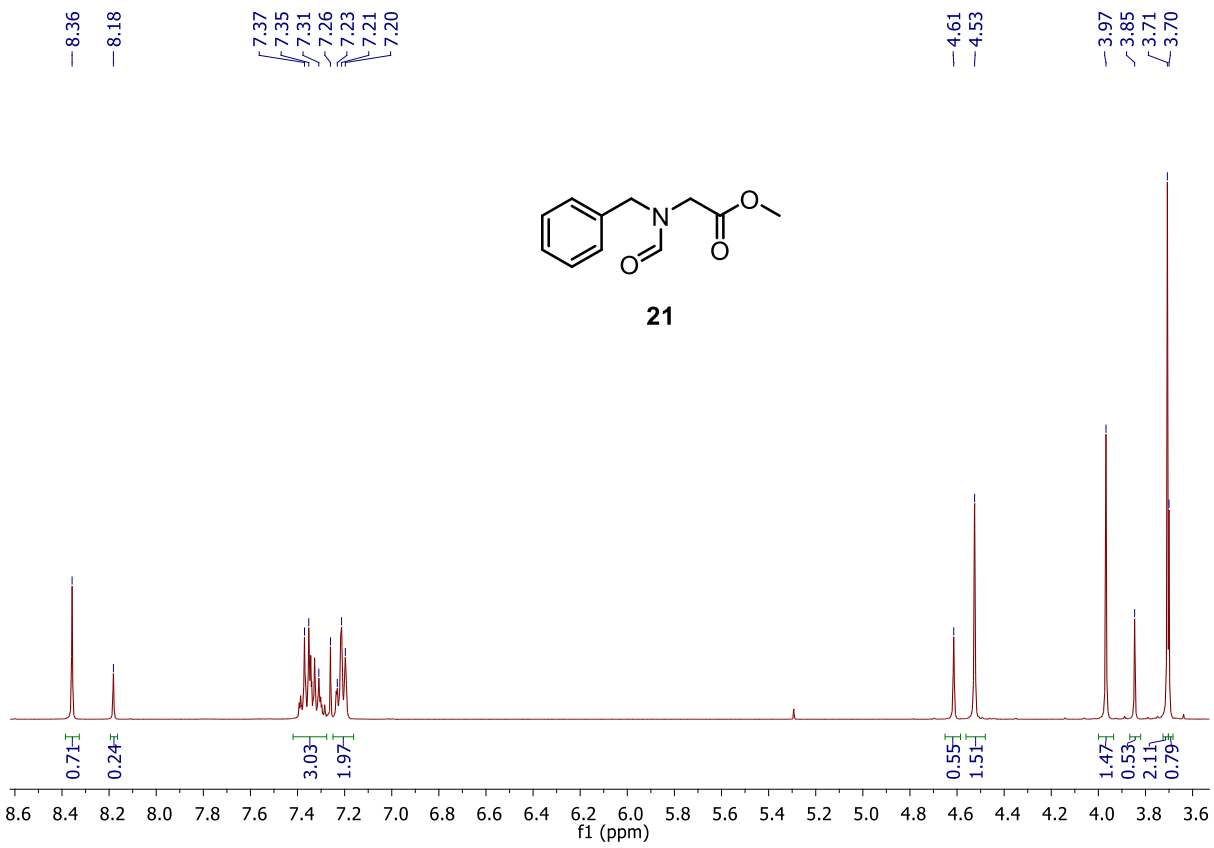


*Figure 35: [1]H-NMR (400 MHz, CDCl3) of compound **21**.*

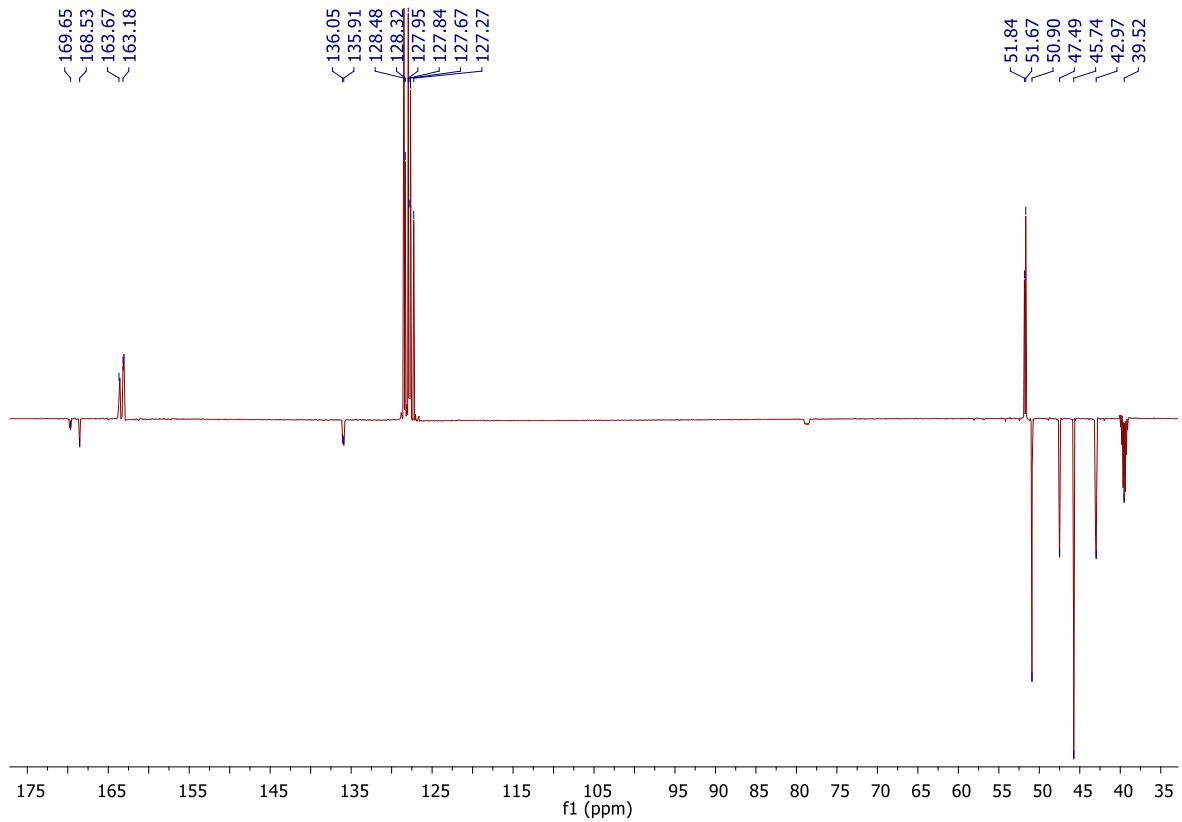


*Figure 36: [13]C-NMR (150.9 MHz, 6d-DMSO) of compound **21**.*

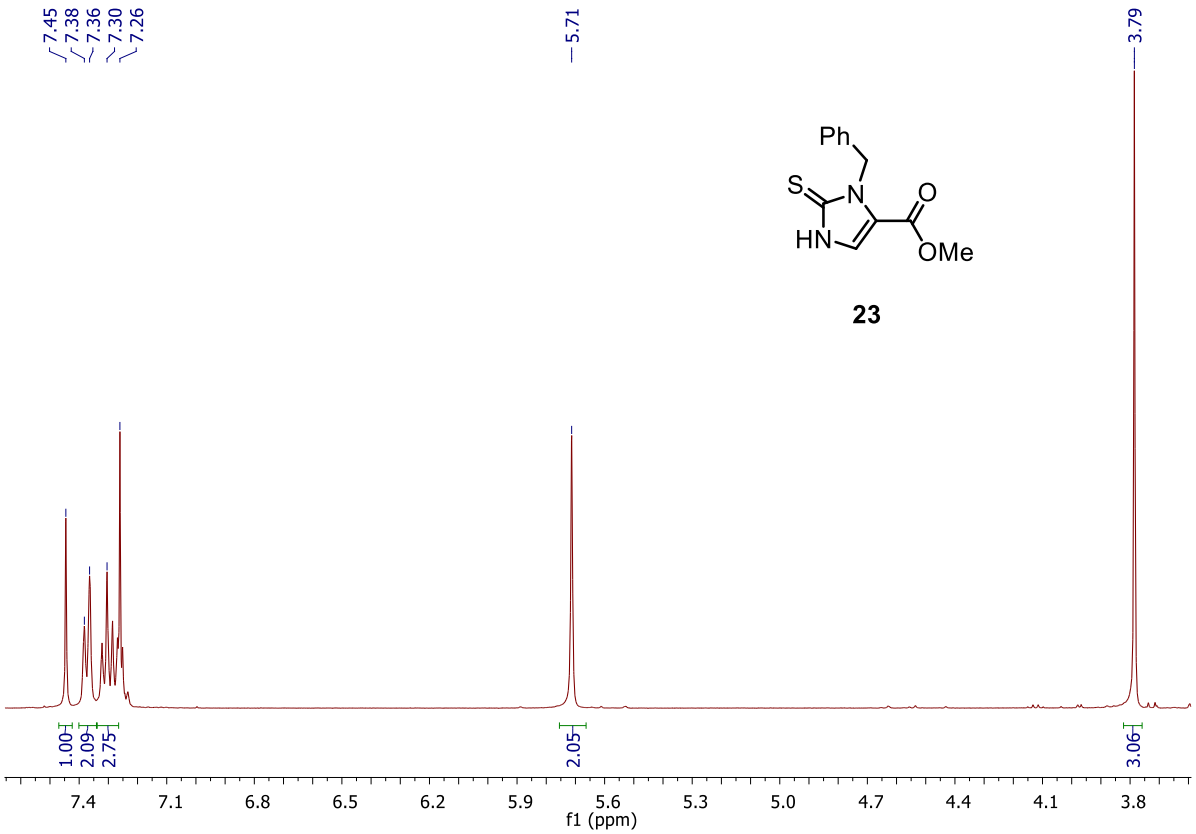


*Figure 37: ^{1}H-NMR (400 MHz, $CDCl_3$) of compound **23**.*

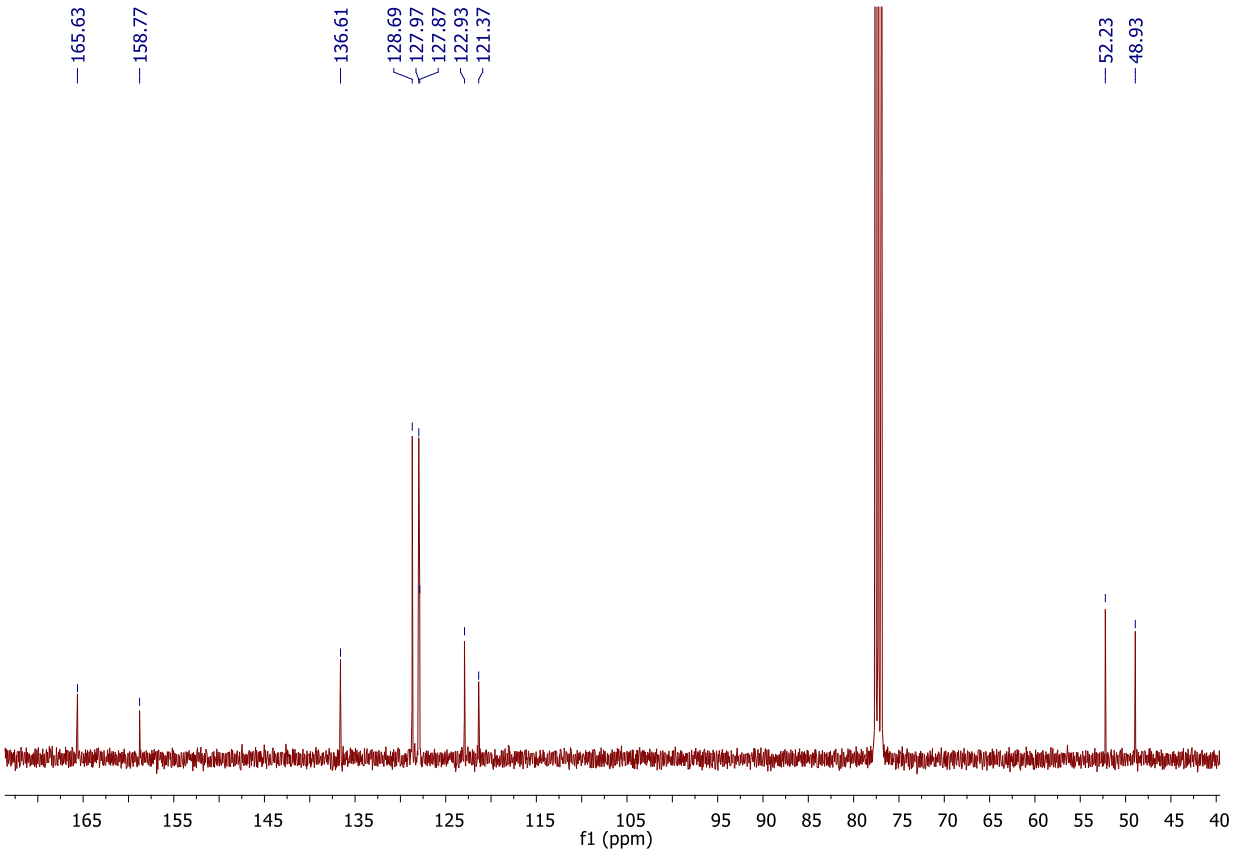


*Figure 38: ^{13}C-NMR (100.6 MHz, $CDCl_3$) of compound **23**.*

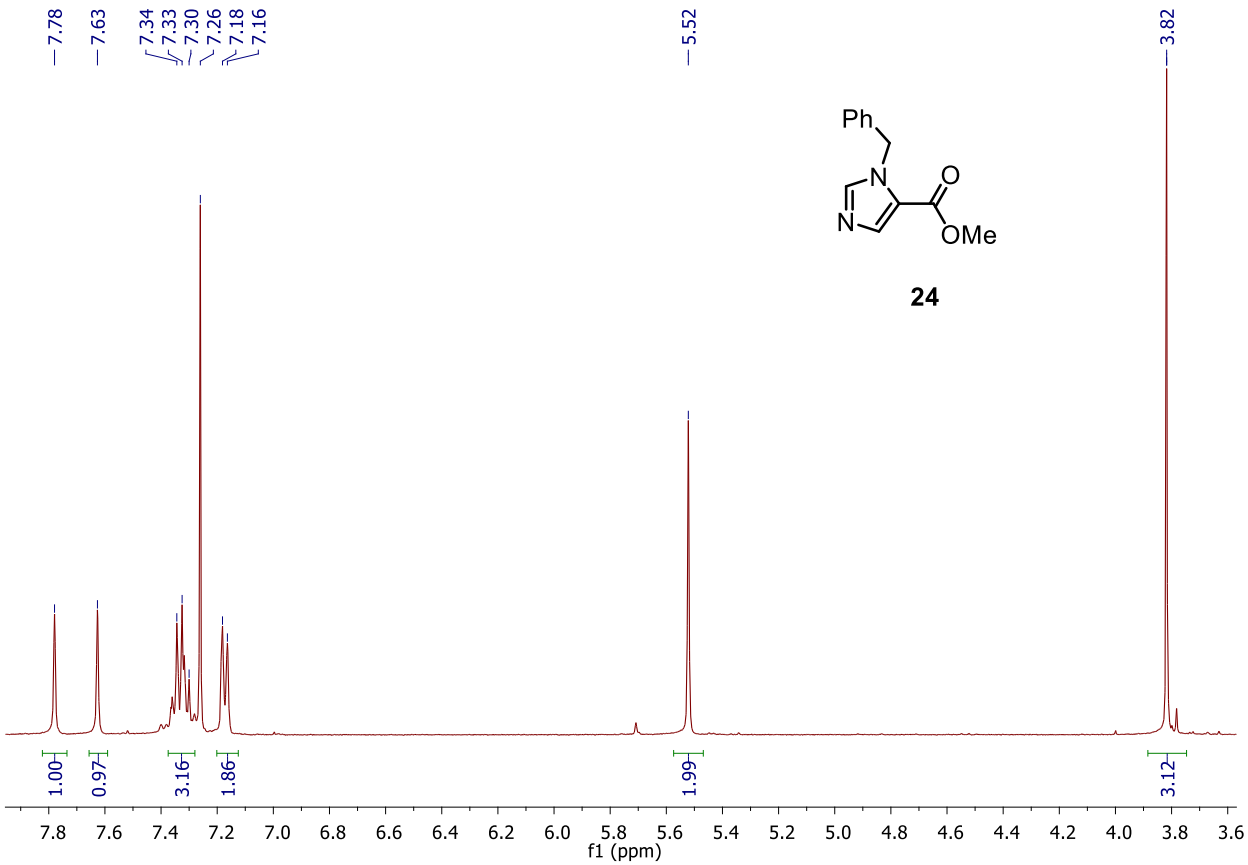


*Figure 39: ^{1}H-NMR (400 MHz, $CDCl_3$) of compound **24**.*

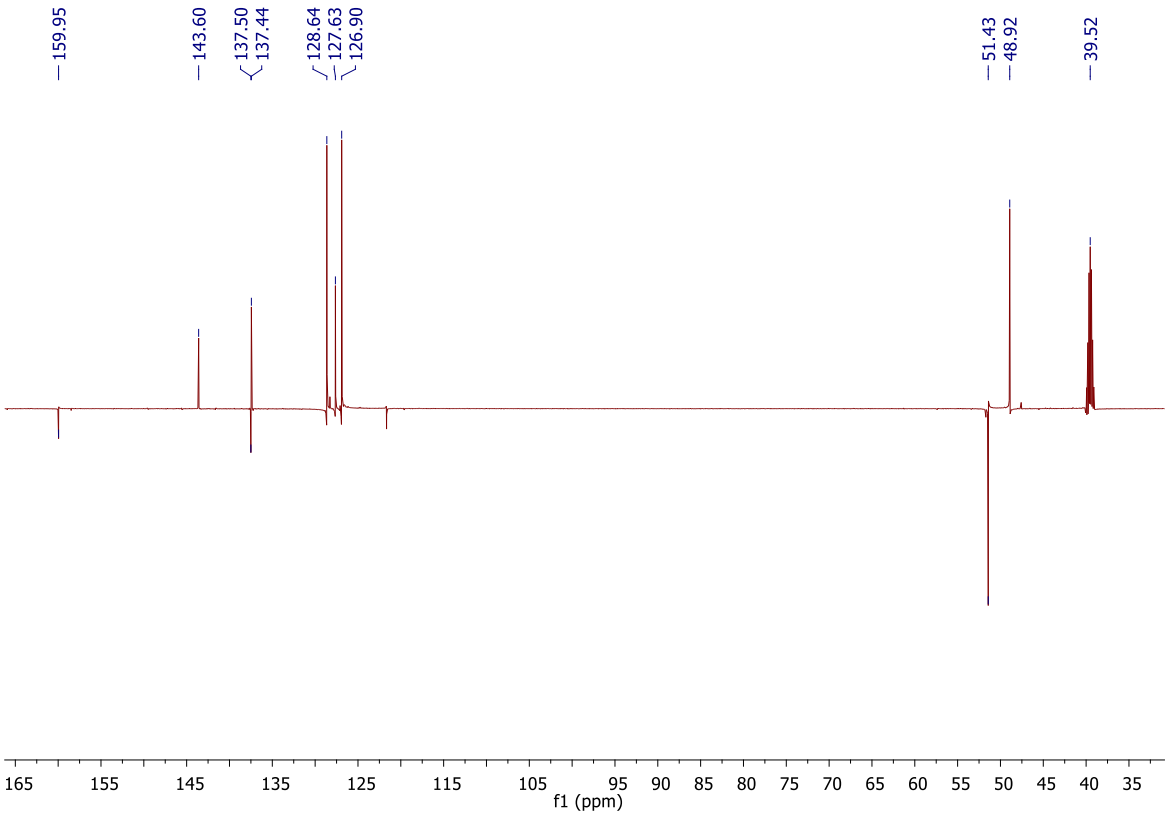


*Figure 40: ^{13}C-NMR (150.9 MHz, 6d-DMSO) of compound **24**.*

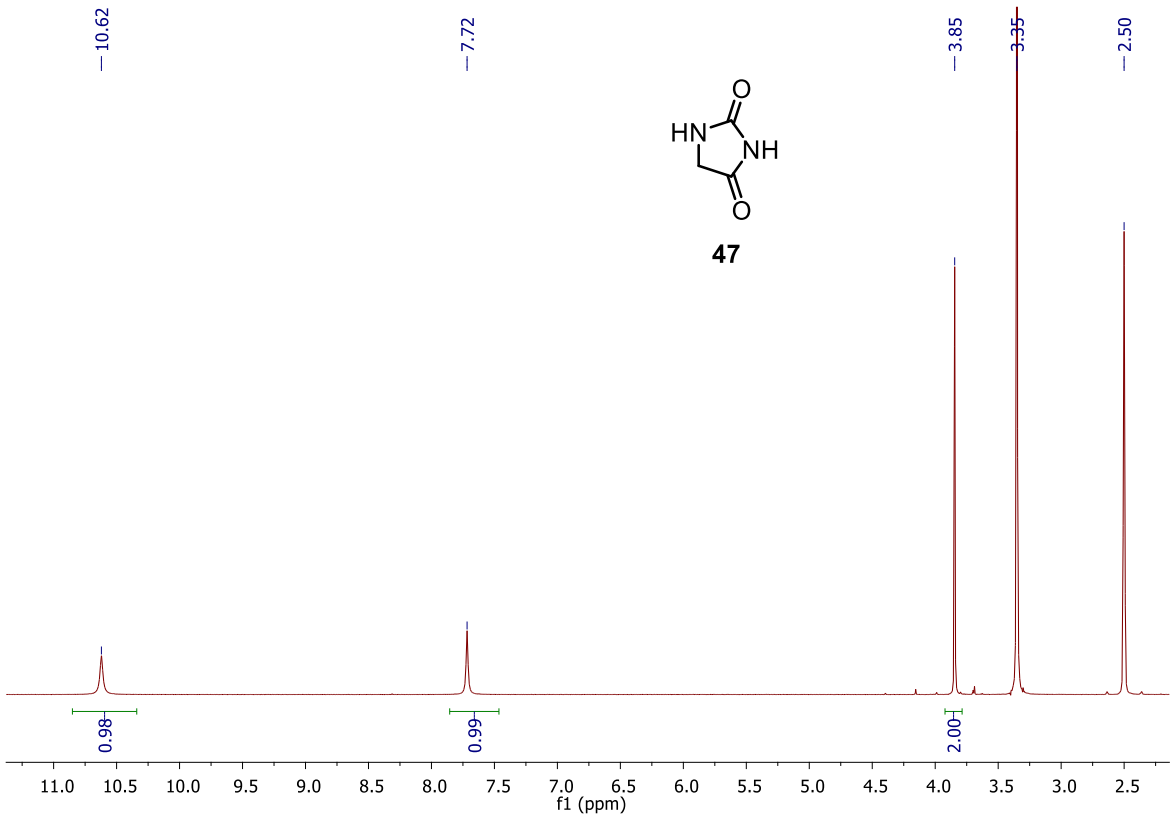


*Figure 41: [1]H-NMR (500 MHz, 6d-DMSO) of compound **47**.*

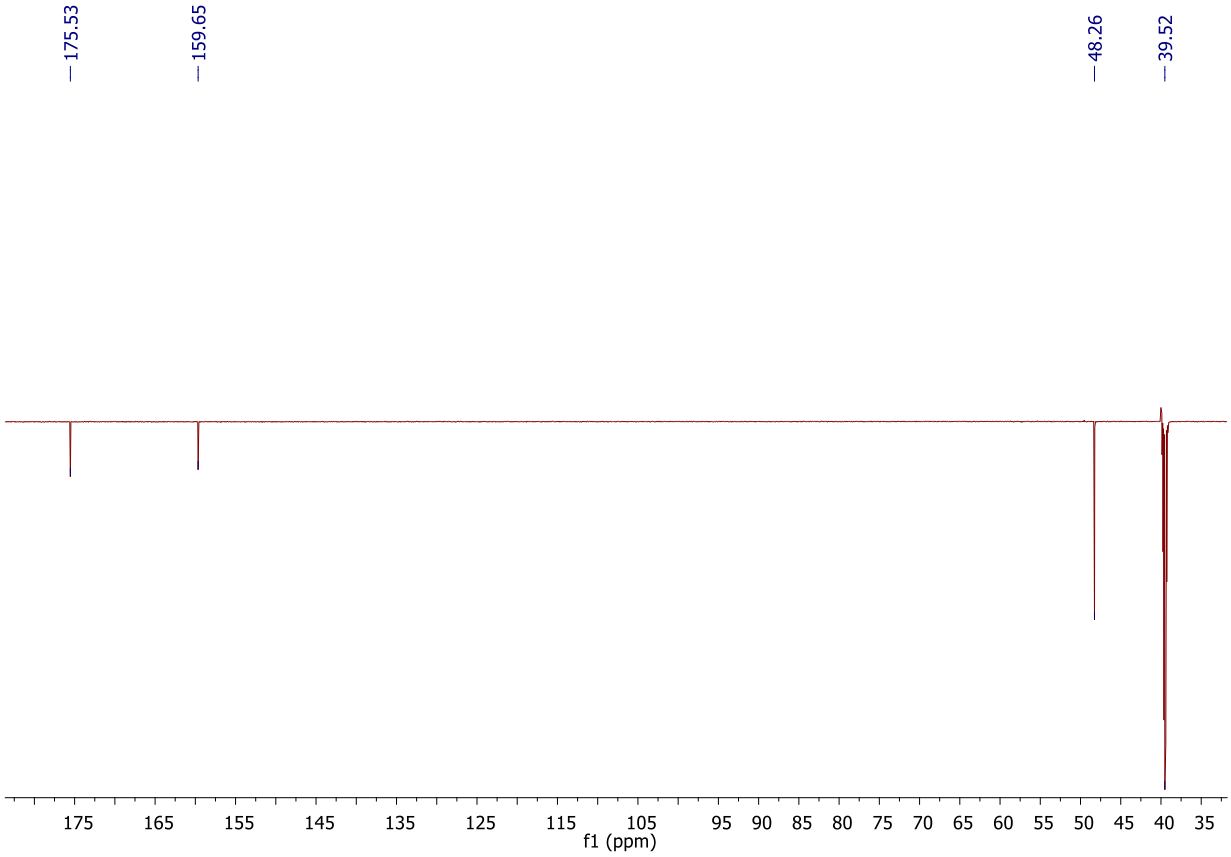


*Figure 42: [13]C-NMR (176 MHz, 6d-DMSO) of compound **47**.*

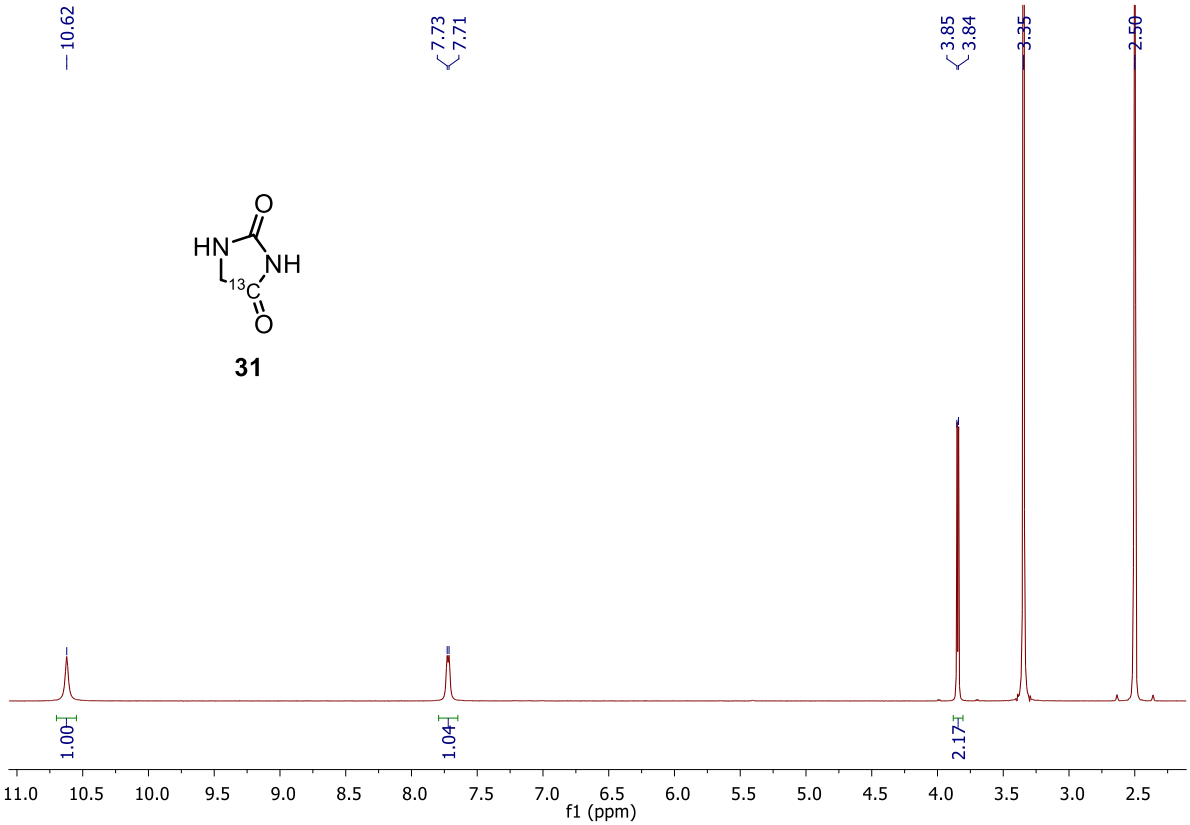


*Figure 43: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **31**.*

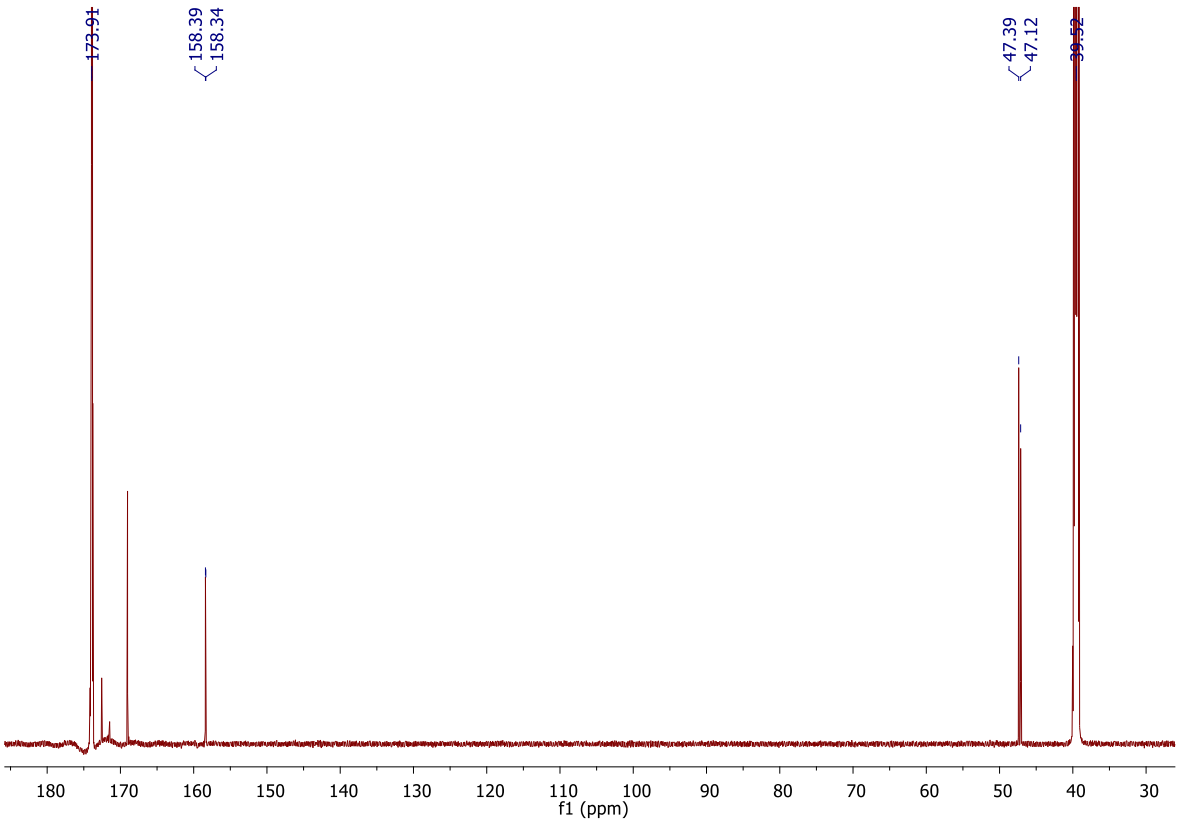


*Figure 44: ^{13}C-NMR (176 MHz, 6d-DMSO) of compound **31**.*

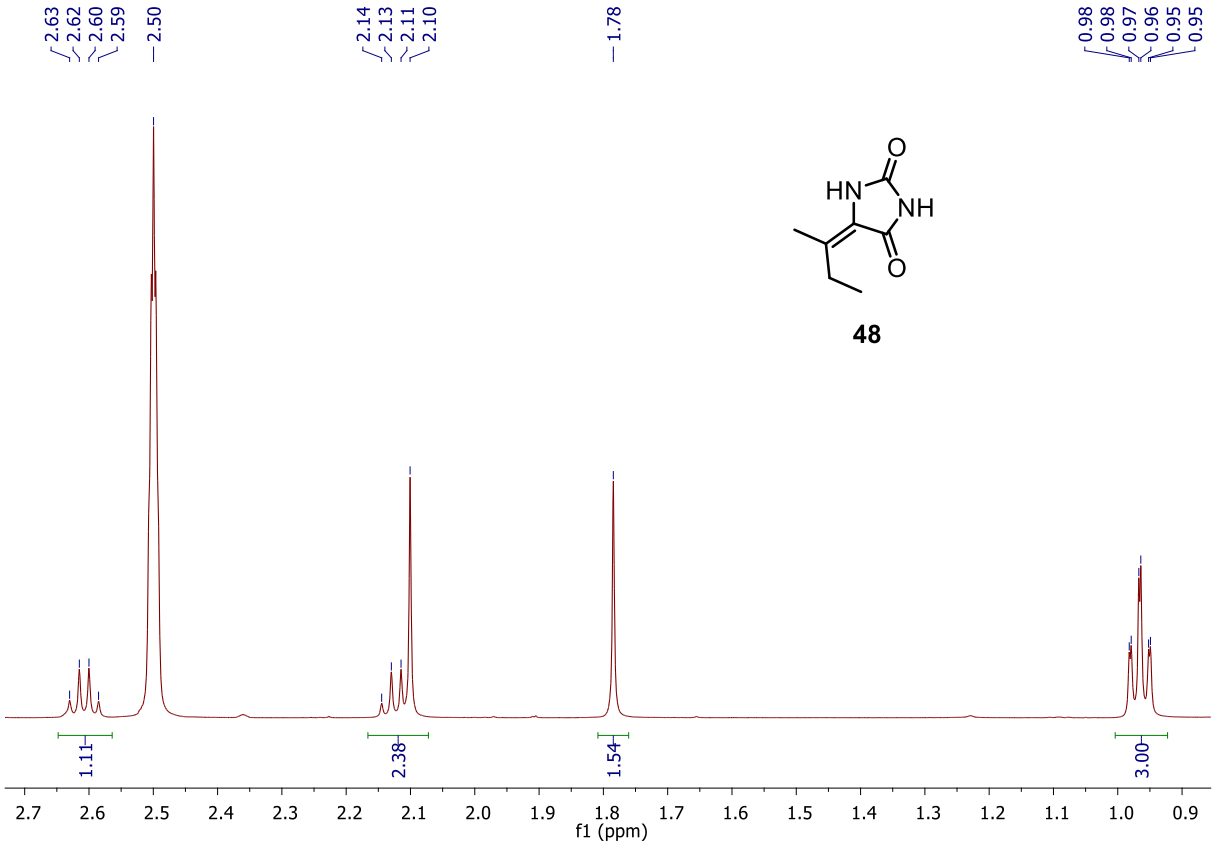


*Figure 45: [1]H-NMR (500 MHz, 6d-DMSO) of compound **48**.*

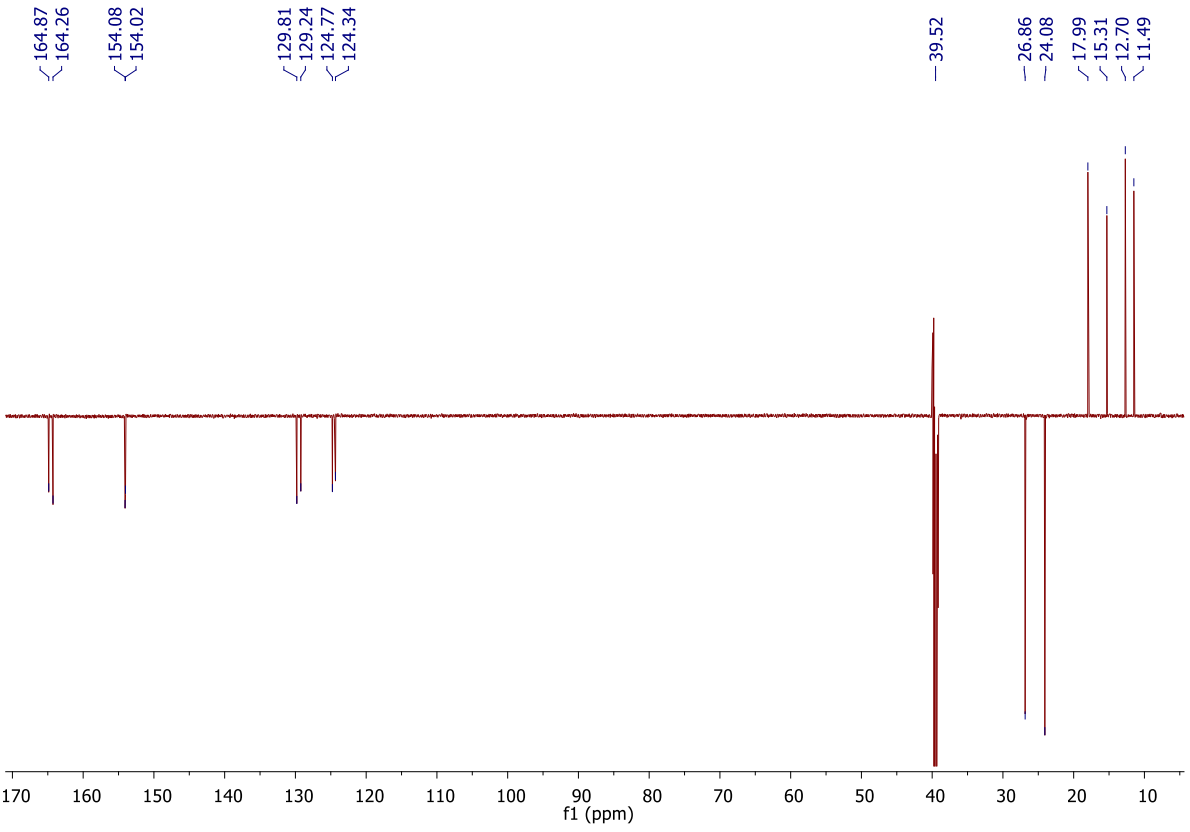


*Figure 46: [13]C-NMR (176 MHz, 6d-DMSO) of compound **48**.*

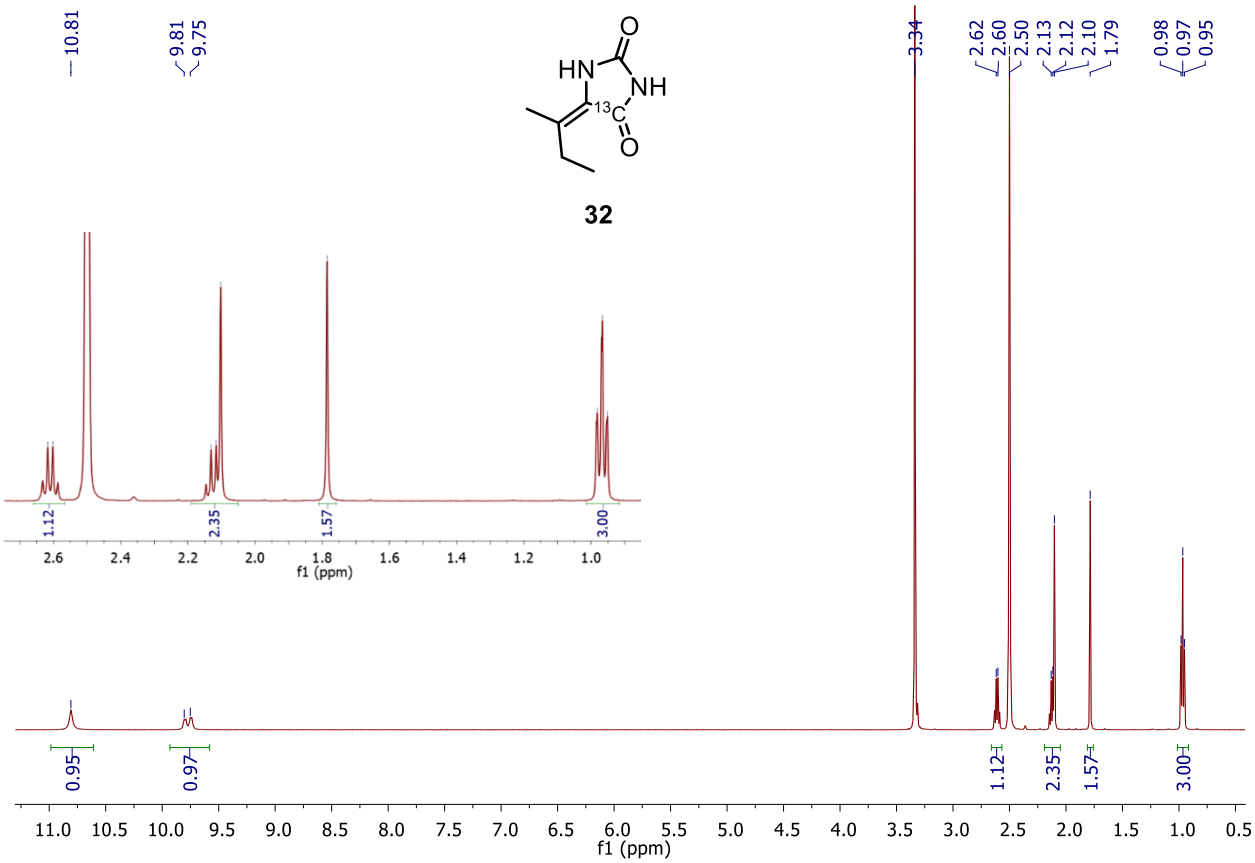


*Figure 47: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **32**.*

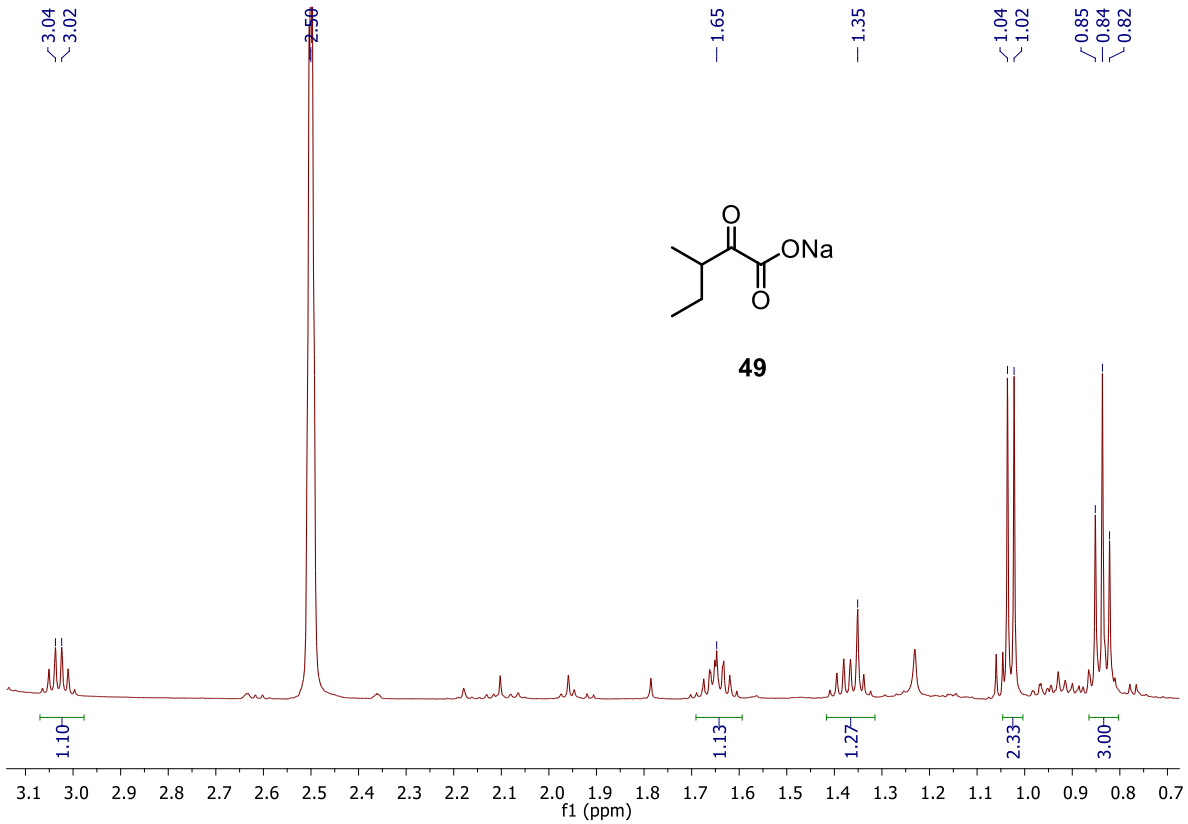


*Figure 48: [1]H-NMR (500 MHz, 6d-DMSO) of compound **49**.*

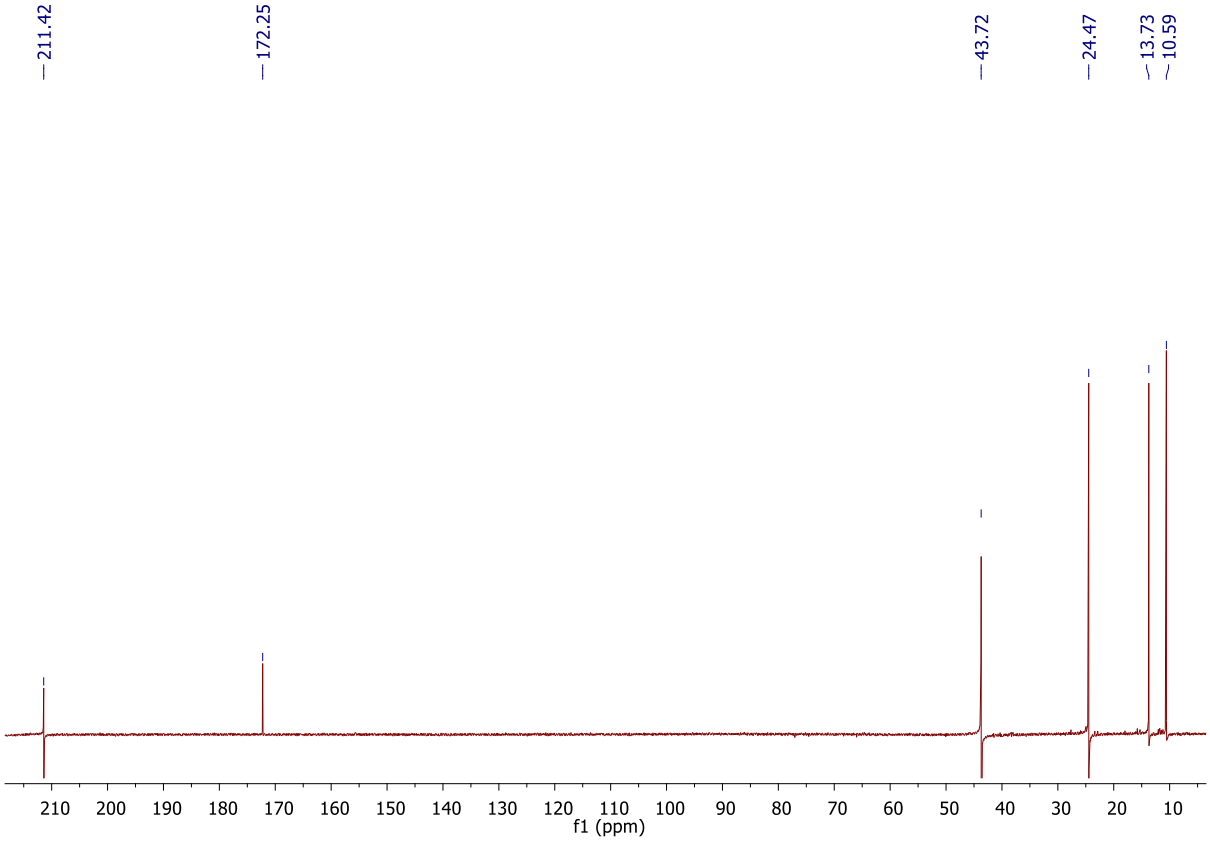


*Figure 49: [13]C-NMR (150.9 MHz, D_2O) of compound **49**.*

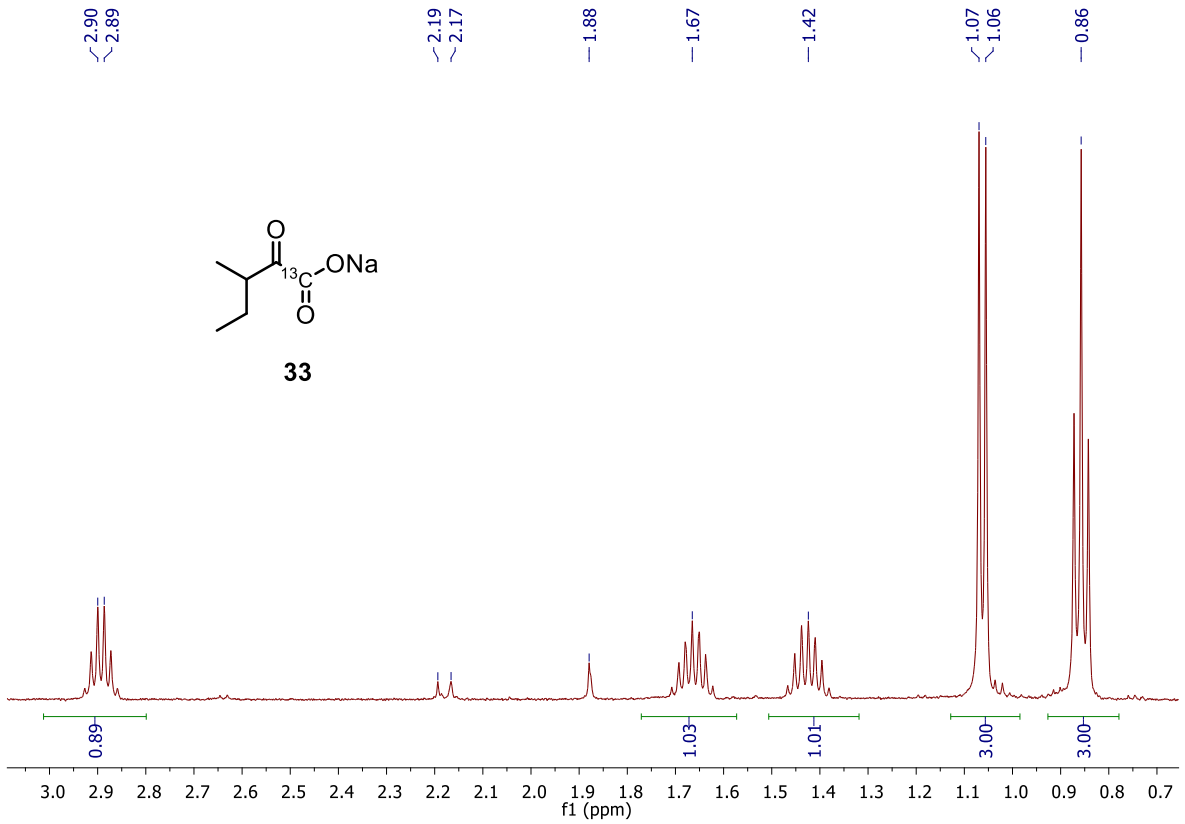


*Figure 50: ^{1}H NMR (500 MHz, D_2O) of compound **33**.*

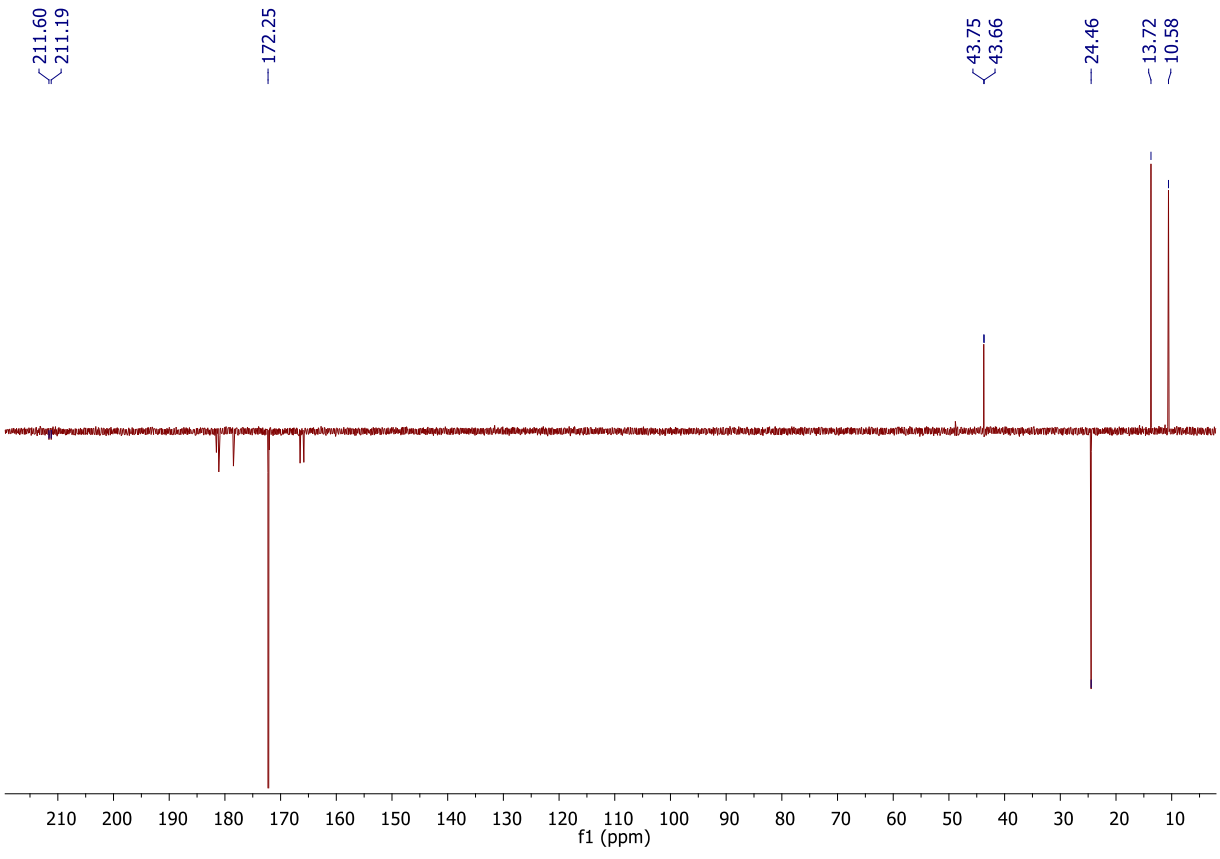


*Figure 51: ^{13}C-NMR (150.9 MHz, D_2O) of compound **33**.*

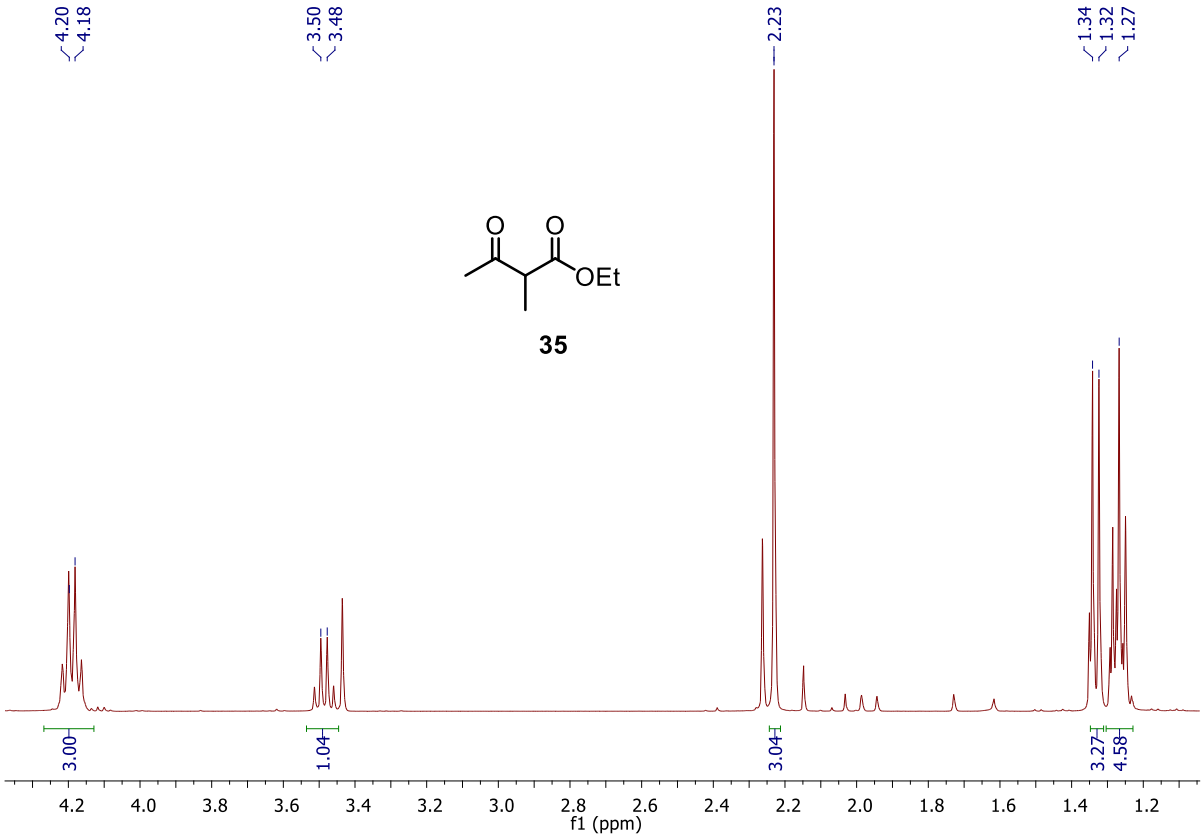


*Figure 52: ^{1}H-NMR (400 MHz, $CDCl_3$) of compound **35**.*

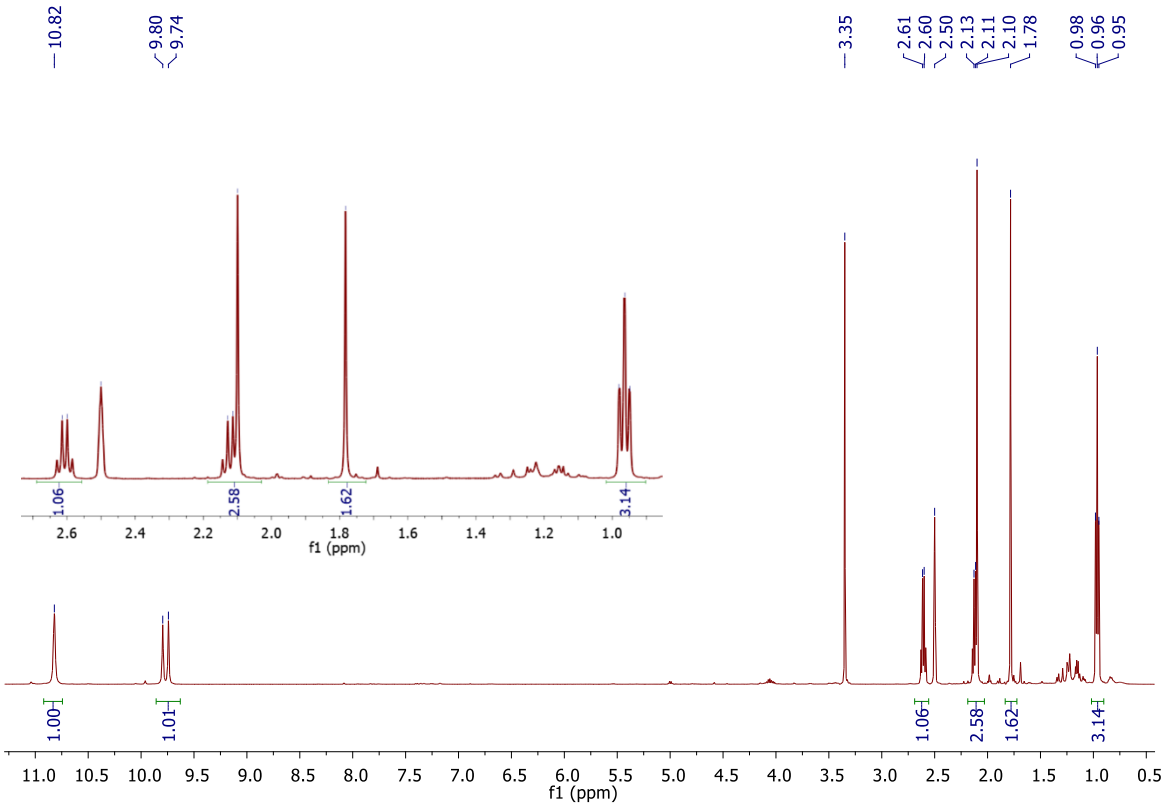


*Figure 53: [1]H-NMR (500 MHz, 6d-DMSO) of compound **38**.*

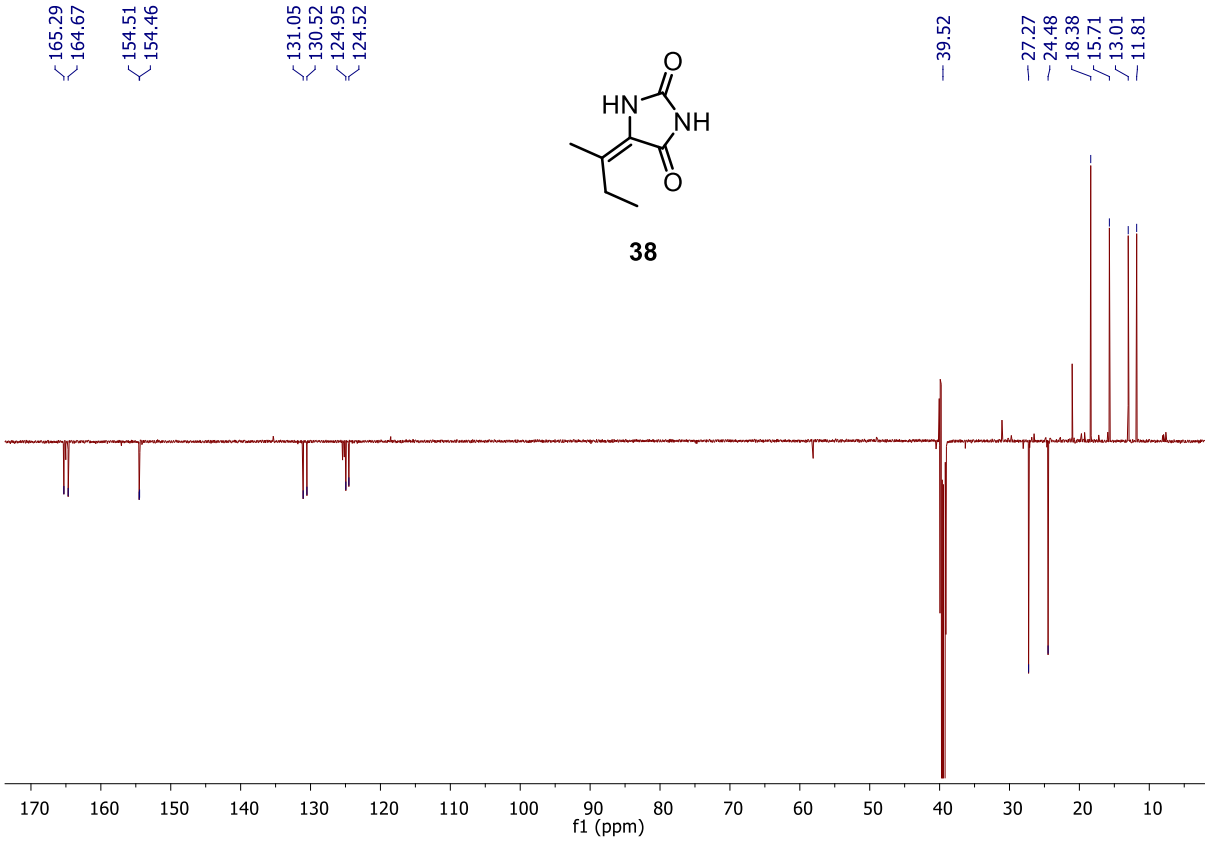


*Figure 54: [13]C-NMR (150.9 MHz, 6d-DMSO) of compound **38**.*

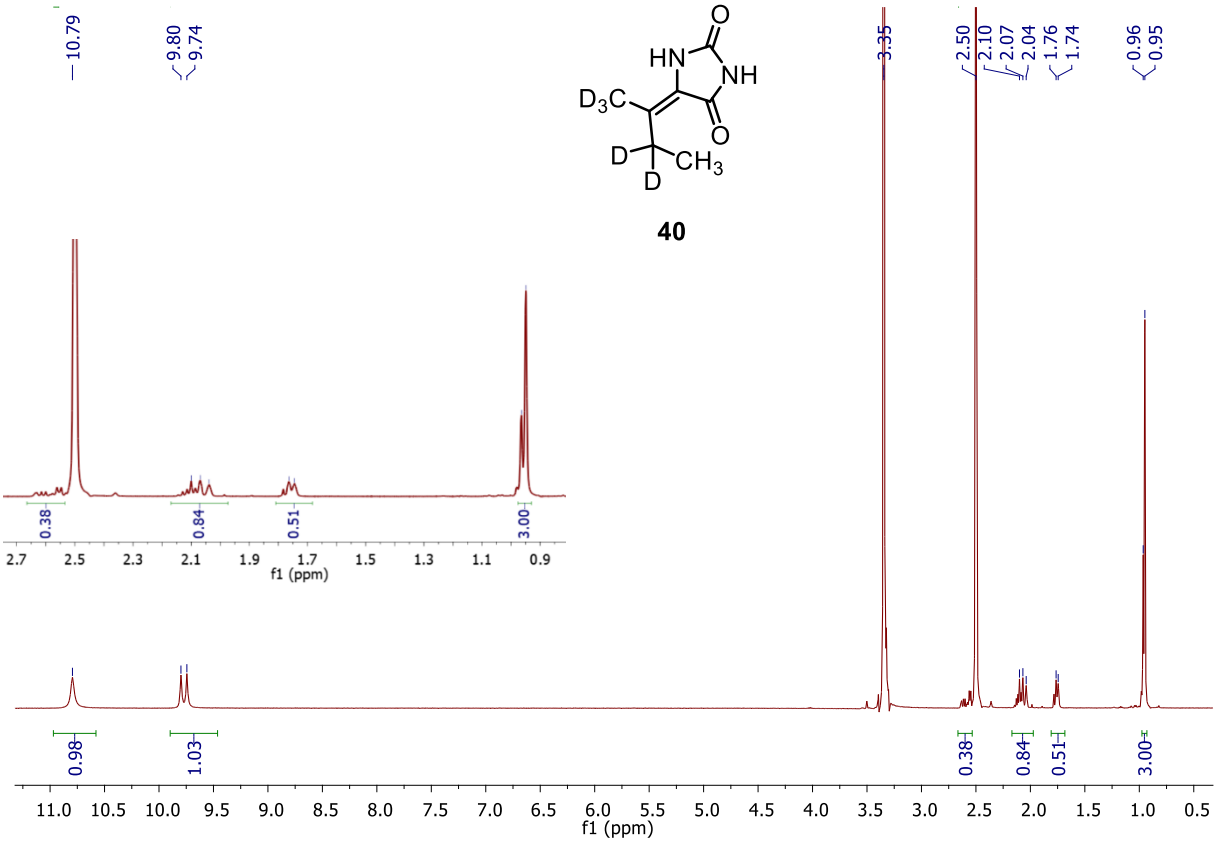


*Figure 55: ^{1}H-NMR (500 MHz, 6d-DMSO) of compound **40**.*