TECHNISCHE UNIVERSITÄT MÜNCHEN
Department Chemie
Lehrstuhl für Organische Chemie II

A chemical proteomic strategy for studying pyridoxal phosphate-dependent enzymes

Annabelle Dorothée Hoegl

TECHNISCHE UNIVERSITÄT MÜNCHEN
Department Chemie
Lehrstuhl für Organische Chemie II

Eine proteomische Strategie zur Untersuchung Pyridoxalphosphat-abhängiger Enzyme

Annabelle Dorothée Hoegl

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Aymelt Itzen
Prüfer der Dissertation	1. Prof. Dr. Stephan A. Sieber
	2. Prof. Dr. Kathrin Lang

Diese Dissertation wurde am 11.01.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 07.03.2018 angenommen.

Abstract

Pyridoxal phosphate (PLP) is an enzyme cofactor required for the chemical transformation of biological amines in various facets of cellular metabolism. PLP-dependent enzymes (PLP-DEs) are ubiquitous throughout evolution and catalytically diverse, which complicates their classification using current methods. We present a novel chemical proteomic platform for the global identification and characterization of PLP-DEs in cells. This platform uses functionalized pyridoxal cofactor mimics designed to be integrated into cellular PL uptake mechanisms and metabolic processing to access and report on the full complement of PLP-DEs in *Staphylococcus aureus*. Reductive amination of the probes to PLP-DE binding sites mediates an irreversible link, thus permitting conjugation to reporter tags for enrichment and quantitative proteomic analysis.

A small library of chemical probes for binding to cellular PLP-DEs was synthesized, which comprised minimally-modified pyridoxal derivatives bearing either an alkyne or azide tag. In so doing, an effective route for alkylating the PL-scaffold at the 2'-position was developed. Subsequent biological evaluation of the probes using recombinant proteins and cell-based experiments established their suitability as chemical reporters of PLP binding. Here, kinetic and structural studies revealed effective probe phosphorylation by *S. aureus* pyridoxal kinase as well as the ability of the activated synthetic cofactors to function catalytically within the model PLP-DE alanine racemase. Furthermore, targeted metabolomics and cellular labeling experiments in *S. aureus* confirmed probe uptake, phosphorylation at physiological rates and probe-dependent protein labeling by fluorescence SDS-PAGE. Proteomic analysis of labeled proteins upon enrichment and MS-measurement revealed known as well as novel PLP-DE family members. Here, a two-tiered filtering strategy based on LFQ-intensity profiles was developed to add confidence to our data and to discern from non-specific binding. Our PL-probe toolbox was able to collectively access 73% of the *S. aureus* PLPome (33 proteins) and consequently identified two novel PLP-DE family members. Cofactor binding was confirmed for select known and uncharacterized PLP-DEs, and preliminary insights into the latter proteins could be gained. In addition, various applications for the method were explored, such as the assessment of *in situ* inhibitor specificity of the antibiotic D-cycloserine across the set of PLP-DEs and comparison of enzyme active site architectures using elaborated probe structures. Finally, labeling in other organisms such as human cells and the Gram-negative bacterium *Pseudomonas aeruginosa* was explored in order to expand the application of the method.

This thesis describes the development of a method for profiling an enzyme family of primary biological significance using an interdisciplinary approach, which can be used for genome annotation, comparative profiling and drug development.

Zusammenfassung

Pyridoxalphosphat (PLP) wird als Cofaktor für diverse enzymatische Reaktionen von biologischen Aminen benötigt und ist essenziell für den zellulären Metabolismus. PLP-abhängige Enzyme (PLP-DEs) sind daher weit verbreitet und weisen eine enorme katalytische Vielfalt auf, was die Klassifizierung dieser Proteinfamilie verkompliziert. In dieser Arbeit wurde eine chemisch proteomische Methode zur globalen Identifizierung und Charakterisierung zellulärer PLP-DEs entwickelt. Dafür wurden Pyridoxal-Derivate als chemische Sonden genutzt. Diese Cofaktor-Analoga sollen von Zellen aufgenommen werden und in PLP-abhängige, biologische Netzwerke eingebaut werden. Durch reduktive Aminierung können die Protein-gebundenen Sonden irreversibel an PLP-DEs geknüpft werden, was die darauffolgende Kopplung an Affinitätsmarker ermöglicht. Die so isolierten Proteine können mithilfe quantitativer Massenspektrometrie identifiziert werden.

PL-Cofaktor-Derivate mit minimalen Alkin- oder Azid-Funktionen wurden hier als Sonden zur Untersuchung der PLP-DEs aus *Staphylococcus aureus* synthetisiert. Hierzu wurde eine effiziente Syntheseroute entwickelt um an der 2'-Position verschiedene Alkyl-Einheiten einzuführen. Die Cofaktor-Analoga wurden durch *in-vitro*-Assays mit rekombinanten Proteinen und zellbasierten Experimenten validiert. Beispielsweise konnte mit der Aufnahme von Enzymkinetiken gezeigt werden, dass die Sonden durch die *S. aureus* Pyridoxalkinase phosphoryliert und damit aktiviert werden. Zusätzlich konnte mit der Aktivität des PLP-DE Alaninracemase erfolgreiches PLP-Mimikry nachgewiesen werden. Metabolomische Studien mit *S. aureus* bestätigten die Aufnahme und Phosphorylierung der Sonden auch unter physiologischen Bedingungen, und Fluoreszenz-SDS-PAGE wies auf die Bindung zellulärer Proteine hin. Diese Proteine konnten in nachfolgenden Proteomik-Experimenten angereichert und mittels Massenspektrometrie vermessen werden. Die Ergebnisse offenbarten bekannte wie auch unbekannte PLP-DEs. Eine zweistufige Datenfilteranalyse, die auf den quantitativen Intensitätsprofilen der Proteine fußte, diskriminierte mögliche PLP-DEs von Hintergrundproteinen. Mithilfe aller Sonden konnten 73% des beschriebenen *S. aureus* PLPoms (33 Proteine) gefunden werden. Darüberhinaus wurden zwei neue PLP-DEs identifiziert. Angeschlossene Validierungsstudien bewiesen tatsächlich eine in-vitro-PLP-Bindung. Neben dieser Entdeckung weiterer Mitglieder der PLP-DE-Proteinfamilie, wurden zusätzliche Anwendungen der Methode erforscht. Es konnte somit die *in-situ*-Spezifität des Antibiotikums D-Cycloserin unter PLP-DEs bewertet werden und die aktiven Zentren verschiedener PLP-DEs mithilfe weiterer PL-Derivate verglichen werden. Des Weiteren wurden erste Experimente in anderen Organismen durchgeführt, wie z.B. mit Humanzellen oder mit dem gramnegativen Bakterium *Pseudomonas aeruginosa*.

Die vorliegende Arbeit umfasst die Entwicklung einer Methode zur Untersuchung einer Enzymfamilie mit hoher biologischer Relevanz. Sie trägt zur Genom-Annotierung bei und kann bei der Enzymcharakterisierung oder auch in der Medikamentenentwicklung Anwendung finden.

Acknowledgements

Contributions

Chapter 2   The project was designed by A. Hoegl, Dr. M. Nodwell, and Prof. Dr. S. A. Sieber. Dr. M. Nodwell established the synthesis of PL1 and PL2 probes. A. Hoegl reproduced the synthesis of PL1 and PL2, and developed the syntheses of PL3 and PL4 probes.

Chapter 3   A. Hoegl and Dr. M. Nodwell purified recombinant proteins and Alr holoenzymes, and performed Alr kinetics assays as well as *in vitro* PLK probe phosphorylation experiments. M. Pfanzelt performed PLK kinetics assays as well as gel-based labeling experiments and UV-Vis measurements of Alr holoenzymes as a research intern under the supervision of A. Hoegl. Dr. S. Schneider performed crystallization and determined X-ray structures of Alr. A. Hoegl carried out intact-protein MS measurements, cellular labeling experiments, *S. aureus* growth curves and metabolomic experiments. V.C. Kirsch prepared and analyzed targeted metabolomic samples and recorded HRMS measurements of phosphorylated probes.

Chapter 4   The proteomic strategy was designed and developed by A. Hoegl, Dr. N.C. Bach and Prof. Dr. S.A. Sieber. A. Hoegl performed and analyzed proteomic experiments.

Chapter 5   A. Hoegl cloned and purified recombinant proteins, performed intact-protein MS measurements, UV-Vis measurements, analytical size-exclusion chromatography, enzyme assays, binding-site identification, as well as proteomic and metabolomic experiments. M. Pfanzelt performed select UV-Vis measurements and gel-based labeling experiments as a research intern under the supervision of A. Hoegl. Dr. N.C. Bach and M. Stahl analyzed PLP binding-site data and proteomic data. V.C. Kirsch performed and analyzed metabolomics experiments. C. Scheidler and Dr. S. Schneider performed crystallization and determined the X-ray structure of the uncharacterized protein Q2FF14.

Chapter 6   Project applications were designed by A. Hoegl and Prof. Dr. S. A. Sieber. A. Hoegl performed all proteomic experiments and data analysis. M. Pfanzelt synthesized PL5 and PL6 probes under the supervision of A. Hoegl.

Publications

Journal Publication

A. Hoegl, M.B. Nodwell, V.C. Kirsch, N.C. Bach, M. Pfanzelt, M. Stahl, S. Schneider, S.A. Sieber. Mining the cellular inventory of pyridoxal phosphate-dependent enzymes with functionalized cofactor mimics. *(Manuscript in revision process)*

Conferences

A. Hoegl, M.B. Nodwell, N.C. Bach, V.C. Kirsch, M. Stahl, M. Pfanzelt, S. Schneider, S.A. Sieber. A chemical proteomic strategy for the study of pyridoxal phosphate-dependent enzymes. *Quantitative proteomics: Strategies and tools to probe biology. European Molecular Biology Laboratory (EMBL) course*. Heidelberg, Germany, June 19-23, 2017. (Poster presentation)

A. Hoegl, M.B. Nodwell, N.C. Bach, V.C. Kirsch, M. Stahl, M. Pfanzelt, S. Schneider, S.A. Sieber. Using chemical proteomics to study pyridoxal phosphate-dependent enzymes. *100th Canadian Chemistry Conference and Exhibition*. Toronto, Canada, May 28 - June 1, 2017. (Conference talk)

A. Hoegl, M.B. Nodwell, S. Schneider, S.A. Sieber. A chemical proteomic strategy for the study of pyridoxal phosphate-dependent enzymes. *Bioorganic Chemistry Gordon Research Conference: Chemical Approaches for Unraveling Biology*. Andover, NH, USA, June 5-10, 2016. (Poster presentation)

x

Table of Contents

List of Figures

List of Schemes

List of Tables

1. Introduction

Pyridoxal phosphate (PLP, Figure 1.1), a bioactive component of vitamin B_6, is a versatile and ubiquitous enzyme cofactor required by all organisms for the metabolism of amino acids and amine-containing compounds. PLP-dependent enzymes (PLP-DEs) use the PLP cofactor to catalyze a range of chemical reactions important for essential cellular processes such as glucose, lipid and amino acid metabolism, as well as heme, nucleotide and neurotransmitter production.[rev.1-3] Accounting for 238 distinct catalytic activities, PLP-dependent chemistry, including racemization, transamination, decarboxylation, and carbon-carbon bond cleavage and formation, is estimated to represent 4% of all known catalytic activities classified to date.[4] The study of PLP-DEs and their catalytic mechanisms has provided important insights into the evolutionary diversity, therapeutic relevance and encompassing roles of this enzyme family.
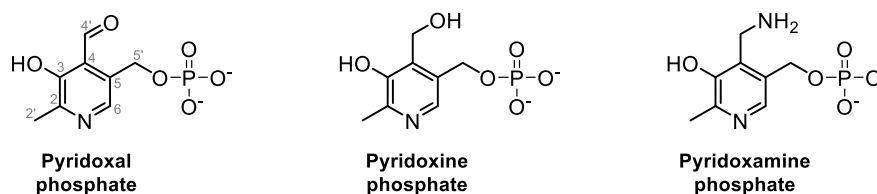
Figure 1.1 Pyridoxal phosphate and related vitamin B_6 derivatives.

1.1 PLP chemistry

PLP catalysis hinges on basic chemical principles which enable a modular, effective and diversified biocatalytic system. All PLP-dependent reactions share common mechanistic features which are elaborated towards different chemical outcomes through the enzyme active site environment. In the resting state, PLP-DEs bind the cofactor at an active site lysine residue via a reversible imine bond, termed the *internal aldimine* (Figure 1.2). The electron-withdrawing effect of the heteroaromatic pyridine ring, which is often protonated, polarizes the molecule and activates it towards imine formation. The internal aldimine is stabilized by resonance effects and hydrogen bonding with the nearby C3 phenoxide. Transaldimation with an incoming substrate amine proceeds through the formation of a geminal diamine intermediate to generate the substrate-

bound *external aldimine*. The latter is the catalytically competent complex which facilitates and directs chemical reactions. PLP functions as an electrophilic catalyst by stabilizing negative charge formation at C_α of the substrate, thus making adjacent sigma bonds labile towards heterolytic cleavage. The resulting carbanion is delocalized across an extended π-system in the form of a coplanar *quinonoid* intermediate, which is resonance stabilized by pyridine and Schiff base protonation.
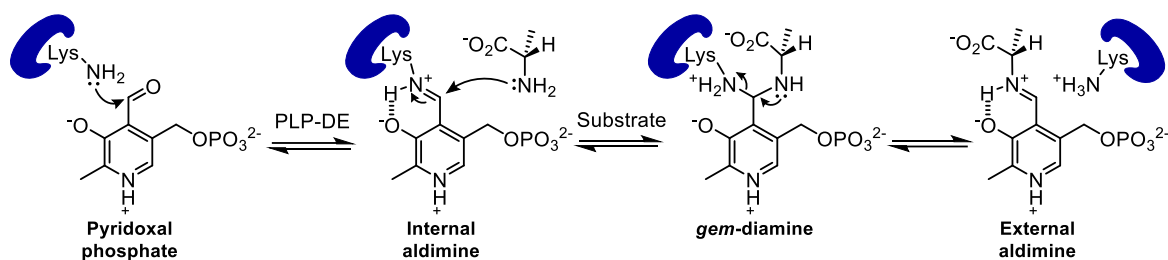
Figure 1.2 Internal and external aldimine formation. Active site lysine residues bind PLP via a reversible imine bond, which is displaced by the incoming substrate to yield the external aldimine via *gem*-diamine formation.

The external aldimine represents a key point of diversification for PLP chemistry, in which the formation of one of three different quinonoid intermediates dictates the downstream reaction path (Figure 1.3).[5,6] In the majority of cases, deprotonation at C_α (C_α-H bond cleavage) yields a carbanion that can subsequently undergo a variety of reactions. Reprotonation of C_α at the opposite face results in substrate racemization or epimerization, while reprotonation at C4' of the cofactor generates a ketimine intermediate which, upon hydrolysis, generates pyridoxamine phosphate (PMP) for transamination reactions (Figure 1.4). Alternatively, the carbanion can act as a nucleophile to undergo aldol and Claisen condensations, or can displace a suitable leaving group at C_β to generate an aminoacrylate intermediate, as demonstrated in cysteine and tryptophan biosynthesis. Decarboxylation (C_α-CO_2 bond cleavage), on the other hand, is generally followed by reprotonation of the quinonoid intermediate to produce biogenic amines required for hormone, alkaloid, nucleic acid and protein biosynthesis. Finally, side-chain cleavage (C_α-C_β bond cleavage) leads to retro-aldol-type condensations, which are used by serine hydroxymethyltransferase or threonine aldolase enzymes. Certain enzymes have also been shown to catalyze combinations of these reactions, as exemplified by dialkylglycine decarboxylase (decarboxylative transamination).[7]

Enzymes control PLP reaction specificity through their active site environment. According to *Dunathan*,[8] stereoelectronic effects within the external aldimine direct which of the three σ-bonds of the substrate is initially cleaved (Figure 1.3). The most labile bond is that oriented parallel to neighbouring *p*-orbitals of the π-system due to optimal charge delocalization and stabilization. Thus, enzymes affect the reaction outcome by directing the substrate binding orientation through their active sites. In addition, residues within the enzyme pocket provide the necessary functionality for catalysis and can tune the electrophilic strength of the system by stabilizing

specific protonation states of the external aldimine.[6] Both the number of protons as well as their placement can affect the stability and reactivity of the carbanion.[9] The evolution of distinct active site features has influenced the specificity of PLP-DEs for catalyzing certain reaction types at the expense of others. Therefore, active site amino acid substitutions are known to affect both the reaction rate as well as the reaction path.[10,11] While PLP is able to catalyze reactions independently of PLP-DEs,[12,13] particularly when assisted by metal ion chelation, enzymes represent an optimal and precise way to achieve a diversified catalytic output.
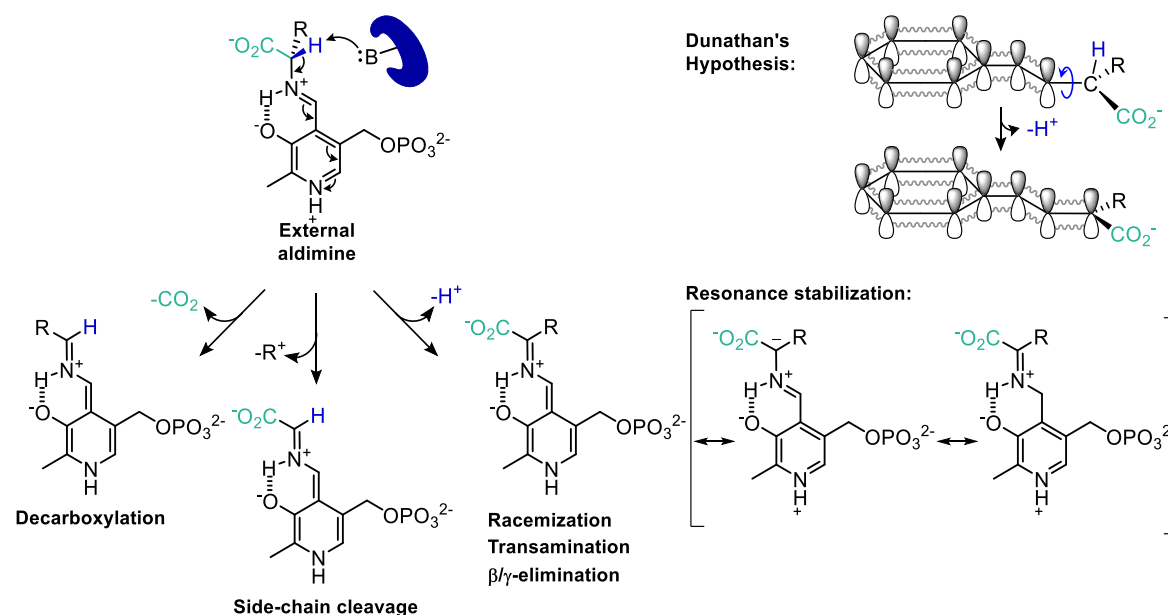
Figure 1.3 Controlling PLP reaction specificity. In the external aldimine, heterolytic cleavage of a σ-bond adjacent to Cα of the substrate results in the formation of a resonance-stabilized quinonoid intermediate. Enzymes control specificity through stereoelectronic effects (Dunathan's hypothesis) and their active site environments.

1.2   PLP-dependent enzymes

Enzymes that use PLP for catalysis originated early in evolution and were already specialized towards basic metabolic pathways in the universal ancestor (~ 1500 million years ago).[14,15] PLP-DEs can be traced to five independent ancestral protein lineages, whose common mechanistic features likely arose by chemical necessity and are reflective of convergent evolution. These subfamilies can be classified based on structural fold type and are named after their most representative protein member.[16-18] Fold type I (aspartate aminotransferase) represents the largest, most diverse PLP-DE subgroup, which is subdivided into six classes including three aminotransferase types.[rev.2,3] Fold type II (tryptophan synthase) enzymes, which often catalyze β-elimination reactions, share overarching similarities with the fold type I group, but differ in their active site placement and the presence of an additional regulatory domain.
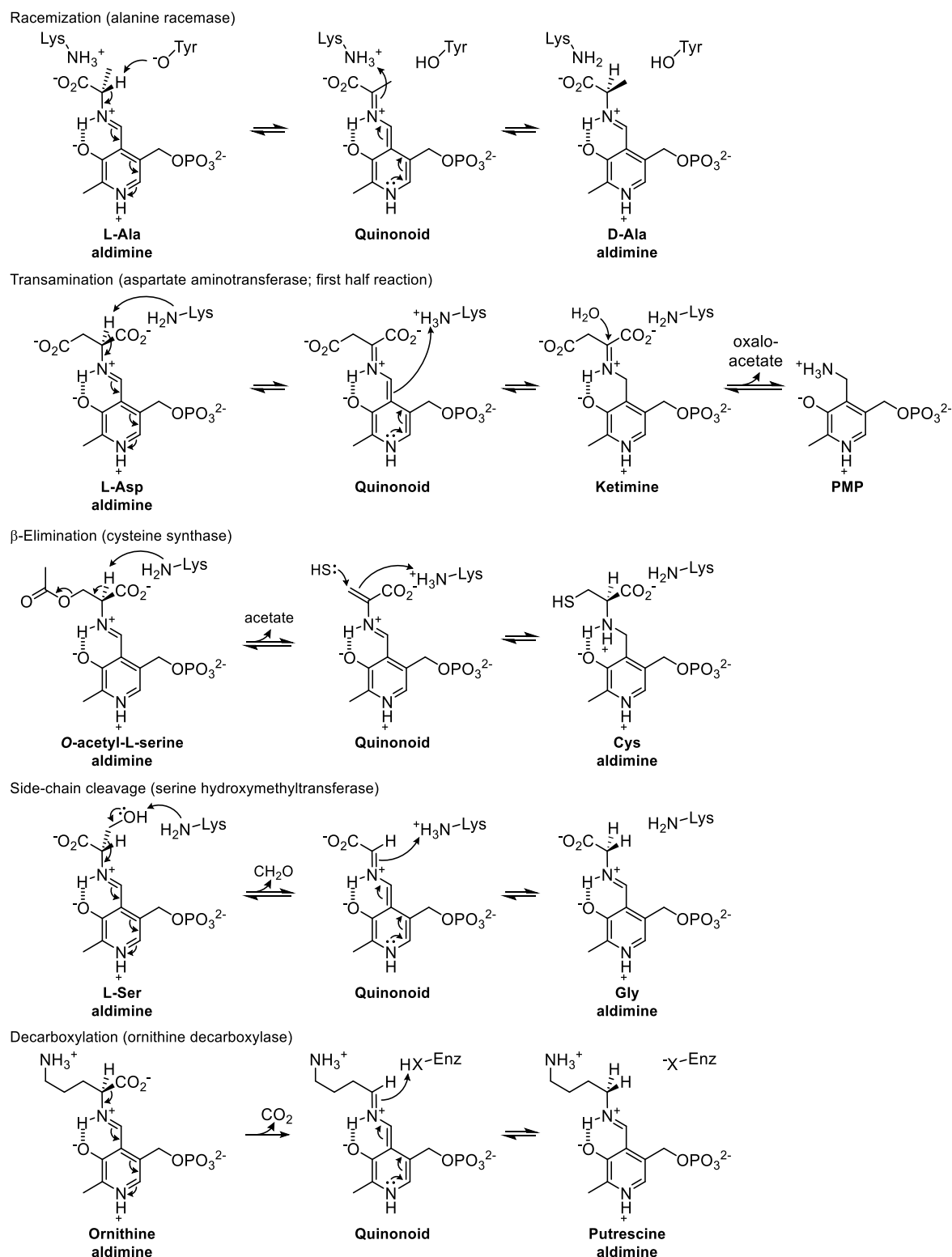
Figure 1.4 Mechanisms of PLP-catalyzed reactions. Racemization: alanine racemase interconverts D- and L-alanine isomers using active site Tyr and Lys residues as local acids/bases. Transamination (first half reaction): proton abstraction from aspartate yields a ketimine, which is hydrolyzed to generate PMP and oxaloacetate. The PMP amine is transferred onto another substrate to regenerate PLP in the second half reaction. β-Elimination: elimination of acetate from *O*-acetylserine by cysteine synthase generates an aminoacrylate intermediate, to which hydrogen sulfide adds to form the product cysteine. Side-chain cleavage: serine hydroxymethyltransferase initiates the retro-aldol cleavage of serine by deprotonating the hydroxyl group, releasing formaldehyde. Protonation by an active site lysine releases glycine as the product. Decarboxylation: loss of CO_2 from substrate amino acids generates biogenic amines, such as putrescine from ornithine.

Fold type IV (D-amino acid aminotransferase) enzymes resemble both fold types I and II, but often display a near mirror image of their active sites. By contrast, fold type III (alanine racemase) PLP-DEs are characterized by distinct α/β-barrel cores and are found among bacterial racemases and eukaryotic decarboxylases. Fold type V enzymes (glycogen phosphorylase) exhibit a unique catalytic mechanism using the PLP phosphate group for acid/base chemistry. The structural classification of PLP-DEs facilitates their identification, but is not prognostic of protein function since each fold type has multiple catalytic activities. Early PLP-DEs catalyzed several reactions to cover a range of functions and have since evolved towards a narrower substrate scope and reaction specificity.[14] Several enzymes are still promiscuous in their ability to catalyze more than one reaction type at a physiologically-relevant rate. This phenomenon manifests itself especially throughout *in vitro* experiments with purified proteins and has found significant use in biotechnological applications for the synthesis of unnatural products and metabolites.[1]

The field of PLP-DE enzymology is dynamic and sustained by continuous discoveries of novel enzymes, structures, chemical mechanisms and biological roles. For example, a recently discovered indolmycin biosynthetic enzyme was found to be capable of coupling O_2 and PLP reactivity in order to efficiently oxidize an unactivated carbon-carbon bond.[19] Other studies have identified new PLP-DE fold types among aminomutases, which do not closely resemble any of the currently known PLP-DE structures.[20,21] Moreover, regulatory roles of PLP have been elucidated, including enzyme inhibition (PPAT decarboxylase, mevalonate kinase glyceraldehyde-3-phosphate dehydrogenase, pyridoxal kinase),[22-26] transcriptional regulation (MocR/GabR family)[27-29] and chaperone-like capacities.[30] These discoveries not only expand our knowledge of the versatility of PLP as a cofactor but also enrich our toolbox for biocatalytic, biotechnological, biomedical and biochemical applications.

1.3 PLP-based drug development

Due to their widespread involvement in basic metabolic processes, PLP-DEs have been implicated in human disease and are recognized as important drug targets.[rev.2,31,32] PLP-DEs are well-suited to mechanism-based inhibition, whereby reactive substrate derivatives bind and rearrange to modify the enzyme covalently or to generate tight-binding PLP-substrate complexes in the active sites. Examples include β-chloroalanine, which eliminates to an electrophilic aminoacrylate intermediate, vinylglycine, which forms a conjugated ketimine (Michael acceptor), and propargyl-glycine, which rearranges to a reactive allene intermediate (Figure 1.5). Lack of specificity limits the therapeutic utility of these compounds and is also a general challenge in the development of PLP-DE inhibitors, which are often associated with numerous side effects.

a   PLP-DE Inhibitors

β-Chloroalanine   Propargylglycine   Vinylglycine

Gabapentin   Eflornithine   Carbidopa   D-Cycloserine   Vigabatrin

b   Example of Reaction Mechanisms

Vinylglycine
aldimine   Quinonoid   Ketimine   Inactivated
enzyme

Propargylglycine
ketimine   Enamine
intermediate   Allene
intermediate   Inactivated
enzyme
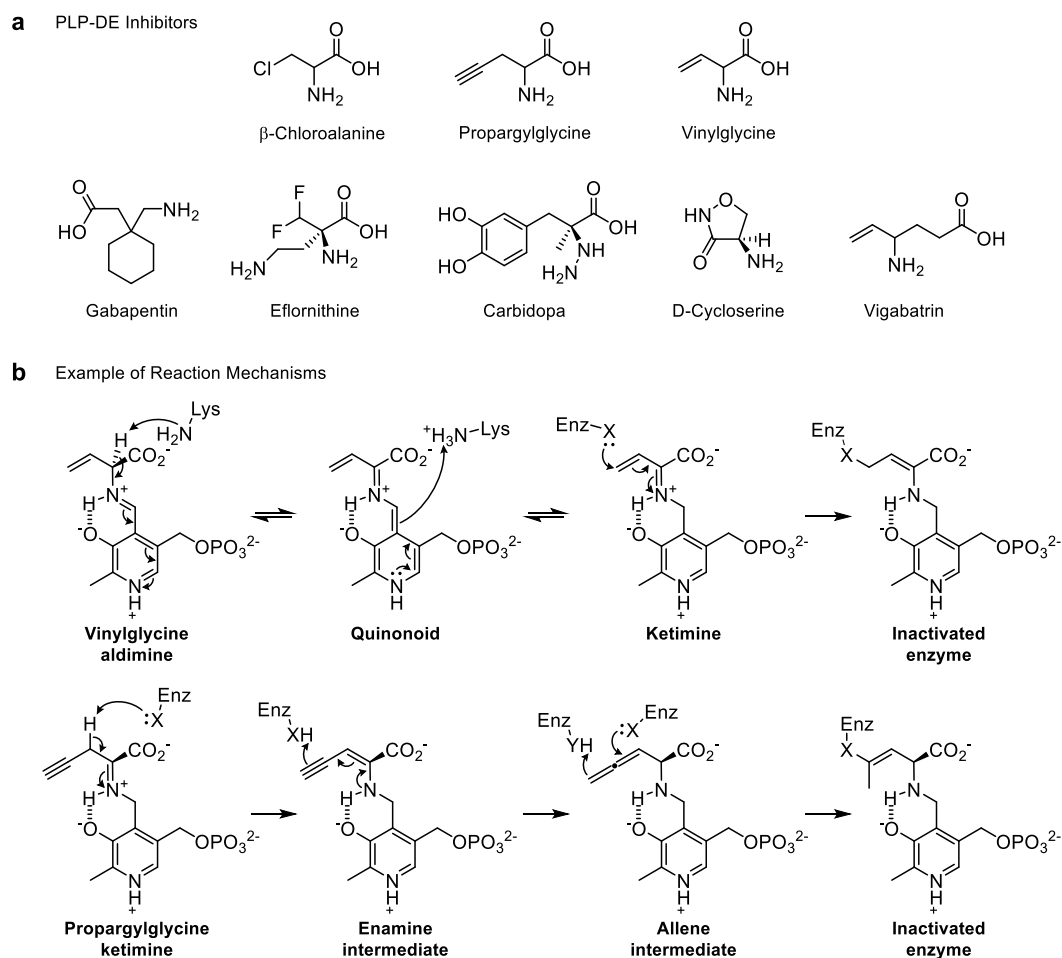
Figure 1.5 Inhibitors of PLP-DEs. (**a**) Mechanism-based inhibitors of PLP-DEs (top) and select clinical drugs targeting PLP-DEs for the treatment of neurological disorders and infectious diseases (bottom). (**b**) Mechanism of inhibition of vinylglycine and propargylglycine.

Inhibitors of human PLP-DEs frequently target their roles in neurological disorders, cancer, inflammation and metabolic dysfunctions. Enzymes for which clinical drugs have been developed include serine hydroxymethyltransferase (Mimosine; malaria, tumours), GABA aminotransferase (Vigabatrin; epilepsy), DOPA decarboxylase (Carbidopa; Parkinson's), and branched-chain amino acid aminotransferase (Gabapentin; GABA/glutamate equilibrium pathologies).[31] Many of these drugs function by resolving imbalances in important CNS neurotransmitter levels. Differences in the mechanisms and types of PLP-DEs in humans versus bacteria or protozoa can also be exploited to treat infectious diseases. Clinically druggable targets include alanine racemase (D-cycloserine; antibacterial) and ornithine decarboxylase (Eflornithine; African sleeping sickness). Other pathways emerging as potential antibacterial targets include peptidoglycan synthesis (*N*-acetylornithine aminotransferase, diaminopimelate decarboxylase), trans-sulfuration (cystathionine γ-synthase, cystathionine β-lyase, methionine γ-lyase, cystalysin, cysteine synthase), as well as amino acid (threonine synthase, tryptophan synthase) and cofactor (heme, biotin) biosynthesis. Mining the cell for further essential metabolic pathways can aid the discovery of new drug targets.

1.4 The PLP-DE family

Gaining a global overview of PLP-DEs has been the subject of several bioinformatics-driven studies aimed at both their identification and functional annotation. Despite notoriously low sequence similarity among PLP-DEs as a result of evolutionary disparity, limited structural diversity has facilitated their classification into five distinct fold types. Here, genomics-based methods capable of extracting structural information from gene sequences have been useful in predicting PLP-DE fold type, identifying new PLP-DEs and drawing preliminary functional conclusions.[4,14] An advantage of these algorithms is their ability to detect distant homology between proteins despite low degrees of sequence identity. This has helped to characterize the molecular evolution of the PLP-DE family and has enabled comparison of their number and variety in different organisms. It was thereby determined that bacteria and archaea contain approximately 20-40 genes for PLP-DEs, representing 1.5% of most prokaryotic genomes.[4] Although higher organisms have a few more PLP-DE genes, their genomic proportion decreases with greater genomic size and complexity, presumably because of their basic metabolic involvement. Results from these analyses have been published in the form of an online B_6-database which integrates bioinformatics with available experimental information (http://bioinformatics.unipr.it.B6db).[33] However, genomic analyses of this type are restricted to known sequence-structure relationships and are imprecise at functionally annotating PLP-DEs. Other bioinformatics approaches for classifying PLP-DE function have therefore compared the structures of active sites, which show high conservation of catalytic residues essential to enzymatic function.[34] These structure-based analyses were able to detect defined active site architectures among the different PLP-DE fold types and were further able to discriminate between unique variations in active site geometry that clustered to specific PLP-DE functions. These studies, and extensions thereof, have helped to define functional evolutionary connections and to decipher the roles of new PLP-DEs.

Nonetheless, our current understanding of the PLPome is incomplete and many PLP-DEs remain unidentified whilst even more lack functional characterization.[4] Although genomic and structure-based bioinformatic analyses have established an overview of the PLP-DE family, these strategies rely on the availability of prior knowledge. In a more direct approach, Western-blotting using an anti-pyridoxine antibody was used to visualize the PLPome, but was unable to identify or quantify distinct PLP-DEs.[35] MS-based approaches attempting to establish a method for the direct detection of PLP-modified residues in proteomes are limited by low intensity and poor stability of the PLP modification during measurement,[36] highlighting the need for a strategy to enrich PLP-bound proteins or peptides. Profiling the family of PLP-DEs would not only help to address important questions in the field concerning evolutionary relationships and novel PLP-dependent functions, but would also enable a range of comparative PLPome analyses.

1.5   Chemical proteomic profiling strategies

Although genomic data contributes a wealth of knowledge to biological investigations, understanding protein function in a cellular context requires further information at the molecular level. To this end, global protein profiling platforms capable of reporting on protein expression, molecular interactions, posttranslational modifications, and localization are useful for studying proteins in their native biological environments. Complementary to these approaches, activity-based protein profiling (ABPP) is an emerging technique which uses active site-directed chemical probes to report on enzyme activities in complex samples.[rev.37]

Activity-based probes (ABPs) contain functionality for covalently binding designated protein targets as well as a reporter group. Incorporating electrophilic moieties such as Michael acceptors or strained rings facilitates irreversible reaction with nucleophilic residues in the enzyme active sites. Alternatively, probes can be equipped with a photo-labile crosslinker which reacts with local residues in the enzyme pocket upon UV-irradiation.[38] Fluorescence (ex. rhodamine) or enrichment (ex. biotin) tags are integrated into the ABP in order to report on binding. A small alkyne or azide handle can be installed in place of the reporter group to improve cell permeability and protein binding, which permits the downstream attachment of tags via copper-catalyzed azide-alkyne 1,3-dipolar cycloaddition (CuAAC).[39] In addition, azide-based probes can be conjugated via strain-promoted azide-alkyne cycloaddition (SPAAC)[40] for *in vivo* labeling applications or via Staudinger-Bertozzi ligation[41] using phosphine-based ligands. Upon incubation of live cells or cell lysates with the probe and click chemistry, labeled proteins can be visualized by fluorescence SDS-PAGE or analyzed by MS (Figure 1.6). For MS-based detection, probe-bound proteins are conjugated to a biotin tag and are enriched using avidin beads. Following on-bead tryptic digest, LC-MS/MS analysis of resulting peptides can provide a quantitative readout of enriched proteins. Binding site peptides can be recovered from the beads using tandem-orthogonal proteolysis (TOP-ABPP), in which a TEV- or trypsin-cleavage site is engineered into the biotin-azide linker and samples are digested accordingly.[42]

A host of ABPP-based applications have been reported, which extend from drug development to characterization of enzyme families. For these purposes, probes selective for diverse enzyme classes have been developed, such as serine hydrolases, cysteine proteases, metallohydrolases, kinases, glycosidases, cytochrome p450s and phosphatases.[37] ABPP facilitates the comparison of enzyme activities in different cell types or states (ex. healthy vs. diseased) in order to understand their implications in certain phenotypes at a molecular level and to identify druggable protein targets. In addition, competitive ABPP has surfaced as an effective screening platform for identifying reversible and irreversible enzyme inhibitors by their ability to block probe labeling. Conducting experiments under native biological conditions ensures a more accurate readout of drug efficacy and specificity, and allows parallel analysis of multiple enzymes. Further applications of competitive ABPP include the assessment of cysteine reactivity in cells using broad-reactivity

electrophilic fragments.[43-45] Other drug-related ABPP approaches involve the use of electrophilic natural product-based probes to identify their cellular protein targets, thereby uncovering new druggable pathways and providing a basis for downstream medicinal chemistry studies.[46] Finally, ABPP has facilitated the identification and characterization of new enzymes. Studying the reactivity profile of uncharacterized proteins with libraries of ABPP probes helps to assign them to a specific mechanistic class and provides insights into their active site binding capacities. Taken together, ABPP has proven to be a powerful method for reporting on relevant enzyme activities under native conditions in order to dissect experiments of high complexity and to extract information that cannot be detected with traditional biochemical methods.
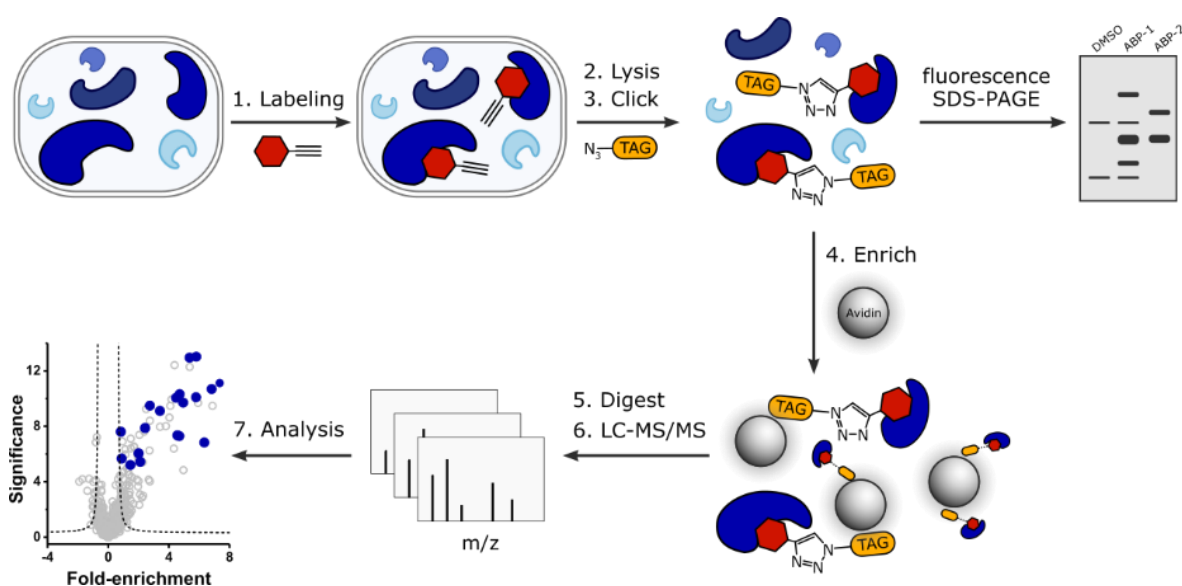
Figure 1.6 Activity-based protein profiling workflow. Labeling of live cells or lysates with an activity-based probe. Cell lysis is followed by click chemistry (CuAAC) with an azide-bearing reporter group (biotin or rhodamine). Rhodamine-labeled proteins are visualized by fluorescence SDS-PAGE. Biotin-labeled proteins are enriched using avidin beads, trypsin digested, measured by LC-MS/MS and analyzed by volcano plot (fold-enrichment vs. *p*-value)

Similar chemical probe strategies have also been valuable for profiling post-translational modifications (PTMs) of proteins to understand their dynamic regulatory networks in the cell.[(rev.47)] Probes for studying PTMs often include alkyne- or azide-modified metabolic precursors that are incorporated into native biological pathways. For example, metabolic incorporation of modified glycans has been useful for profiling glycosylation to understand intracellular regulatory pathways and extracellular glycan composition.[48,49] Furthermore, fatty acid-based probes have served to characterize lipidation and its effects on membrane localization, trafficking and cell signalling.[50] Alternatively, cofactor-based probes can be used to transfer reporter groups onto substrates or bind dedicated proteins directly. Cofactor derivatives capable of transferring metabolic labels onto substrates have been developed for acetyl-CoA (acetylation), *S*-adenosyl-L-methionine (methylation), NAD (mono- and poly-ADP-ribosylation), and ATP (AMPylation and

phosphorylation). Their use has helped to identify the substrates of these modifications and has provided insights into their biological roles. Cofactor probes designed to directly report on protein binding have required the incorporation of a photo-crosslinker to mediate irreversible binding. Thereby, probes for vitamin B_{12} have unraveled new regulatory roles of the cofactor in folate, ubiquinone and methionine metabolism, and probes for other B-vitamins (thiamine, riboflavin and biotin) identified several vitamin transporters.[51,52] Taken together, chemical probes are valuable tools for exploring biological systems which can serve a vast array of discovery-based applications.

1.6   Objectives

Enzymes utilizing PLP for catalysis are ubiquitous throughout evolution and functionally diverse. The importance of this cofactor in sustaining essential cellular pathways can hardly be understated, though our current understanding of this enzyme family is incomplete. To date, there is no comprehensive method for profiling the cellular PLPome and it is thus estimated that many PLP-DEs remain unidentified while even more lack complete functional classification. Profiling the family of PLP-DEs could help to uncover novel biological roles of PLP as well as potential druggable pathways, and would enable diverse comparative applications.

This thesis describes the development and evaluation of a chemical proteomic platform for the global identification and characterization of PLP-DEs in cells. We envisaged the use of functionalized cofactor mimics that are able to integrate into cellular PL uptake mechanisms and metabolic processing to access and report on the full complement of PLP-DEs (Figure 1.7). A small library of PL-derivatives containing an alkyne or azide tag as chemical probes was therefore designed. Here, the 4'-aldehyde acts as an effective handle for mediating an irreversible link to PLP-DEs upon sodium borohydride reduction of the internal aldimine. A stable covalent bond is a prerequisite for downstream proteomic manipulation in order to attach reporter or affinity tags by CuAAC. This permits the enrichment of probe-bound proteins for subsequent identification and quantification by MS-analysis. We aimed to establish a robust proteomic method supported by in-depth biochemical evaluation to ensure specific and authentic probe binding to PLP-DEs, which has the potential for application in diverse experiments. The clinically-relevant Gram-positive bacterial pathogen *Staphylococcus aureus* was selected in order to develop our experimental workflow. Our goals were to use our designed method to identify novel PLP-DE functions or cellular cofactor roles and to investigate its applicability to therapeutically-relevant scenarios, such as off-target screening of PLP-DE drugs and global profiling of enzyme active sites. The method could ideally be further expanded to other organisms and provides an effective tool for probing PLP-related biological questions.
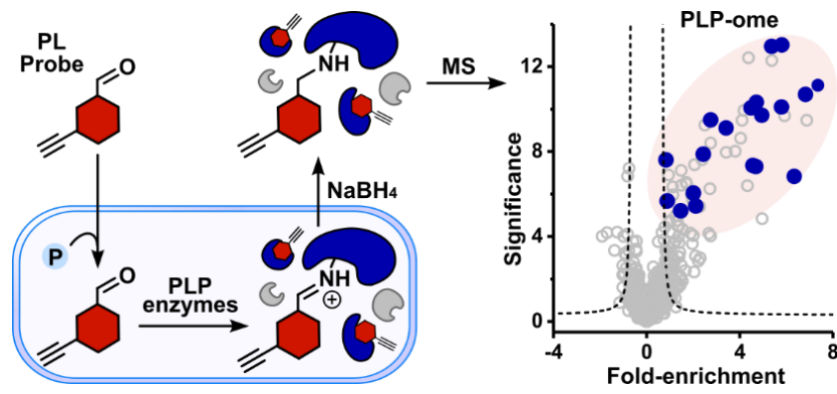
Figure 1.7 Overview of chemical proteomic strategy for profiling the family of PLP-DEs. PL-based probes were designed to be taken up by bacterial cells and integrated into cellular metabolism. Irreversible modification and subsequent MS-based analysis of probe-bound proteins serves to report on PLP-DEs.

2. Design & Synthesis

2.1 Introduction

Reporting on natural PLP binding events requires a functional, minimally-modified synthetic cofactor probe capable of accessing the complement of cellular PLP-DEs and labeling them covalently. Previous investigations into the effects of modifying the PLP scaffold on cofactor viability have helped to define the core functionality required for PLP catalysis and guided our probe design. It was found that PLP-DEs tolerate modification of the cofactor to varying extents. In some cases, the PLP derivatives bound to the enzymes and supported catalysis, though often at reduced rates. In other cases, altering the PLP structure was detrimental to catalysis either because essential cofactor functionality was removed or an ensuing change in binding geometry was not favourable for catalysis. Interestingly, several cofactor derivatives that were non-activating still maintained the capacity to bind PLP-DE active sites and were even competitive for PLP binding. As a result, a number of PLP-derivatives have been synthesized and evaluated for antivitamin and inhibitory properties.[53-55]

The functional groups essential for PLP catalysis include the 3-phenol, 4'-aldehyde and 5'-phosphate. The 4'-aldehyde mediates cofactor binding to PLP-DEs and substrates as previously described and the Schiff base aldimines are stabilized through hydrogen bonding and resonance with the 3-phenolate, which is deprotonated under physiological conditions. Analogues modifying the phenol, such as 3-methoxy-PLP or 3-deoxy-PLP (Figure 2.1), are generally unable to activate apo-PLP-DEs for catalysis but maintain some binding ability.[56,57] The 5'-phosphate of PLP serves as a stable anchoring point to PLP-DEs, mediating up to 9 hydrogen-bonding interactions with a defined phosphate-binding motif conserved among PLP-DE active sites.[58] Modifications at the 5'-position that disrupt phosphorylation (5'-deoxy-PL) or the electrostatic capacity of the phosphate group (methyl-phosphonate, carboxymethyl etc.) significantly decrease binding, while close mimics (5'-deoxy-methylenephosphonate, ethylene phosphate or 5'-methyl-phosphate) bind relatively well and support catalysis to varying degrees.[56,57,59] Pyridine protonation has widely been assumed to be necessary for PLP catalysis by acting as an electron sink. However, recent studies have shown that N-modified PLP derivatives (N-oxide-PLP, de-aza-PLP) support binding

and certain PLP-DE functions.[56,60] For example, a de-aza-PLP analogue yielded functionally competent alanine racemase and cysteine synthase, but not aspartate aminotransferase.[60] In combination with computational studies of enzyme reactions and activation barriers, these studies indicate that not all PLP-DEs require the electrophilicity of the protonated pyridine for carbanion stabilization. Instead, the electrophilic propensity of PLP is tuned towards specific reactions by residues lining the enzyme active site. Finally, it has been found that removal of the 2-methyl group (nor-PLP) or limited expansion thereof (ethyl, propyl) was tolerated by several enzymes for both catalysis and binding, though sometimes with reduced efficacy.[57,61-63] Similarly, methylation at the 6 position (6-methyl PLP) is tolerated.[57,61] The 2-methyl group does not have a distinct catalytic function and is believed to fulfill a spatial role in directing the conformation of cofactor binding. Its modification with bulky substituents can therefore affect the cofactor binding geometry within the enzyme pocket to the point of disrupting catalysis.

In summary, PLP-DEs show varying tolerances towards modification of the PLP cofactor. The 2-position is amenable to alteration without significant or broad repercussions on both binding and catalysis. We therefore reasoned that this site is suitable for the introduction of a small alkyne or azide handle in order to generate a PLP-probe that mimics the natural cofactor and permits integration into PLP-dependent biological pathways.

Modifications at 2-position

nor-PLP   ω-methyl-PLP   2-*n*-propyl-PLP   2-*n*-butyl-PLP   2',2'-dimethyl-PLP
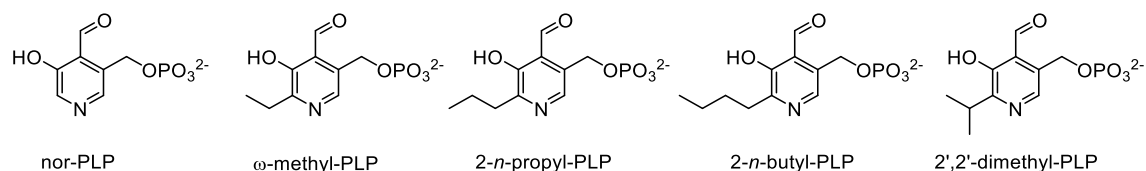
Modifications at 3-position   **Modifications at 5'-position**
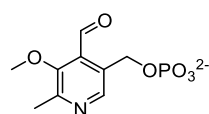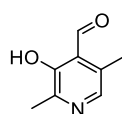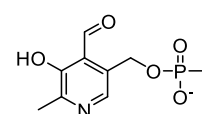
3-*O*-methoxy-PLP   3-deoxy-PLP   5'-deoxy-PL   5'-methylphosphonate   5'-deoxymethylene phosphonate
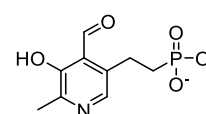
Modifications at 6-position   **Modifications of pyridine**
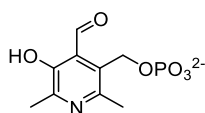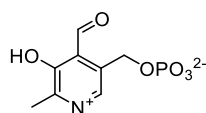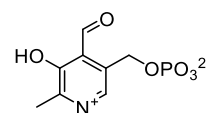
6-methyl-PLP   2-nor-6-methyl-PLP   *N*-methyl-PLP   *N*-oxide-PLP   de-aza-PLP
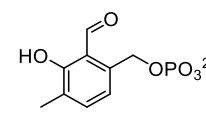
Figure 2.1 Chemical structures of PLP derivatives. Various PLP derivatives have been synthesized and evaluated for their ability to bind to PLP-DEs and facilitate catalysis.

2.2   Experimental Design

Following the example of activity-based protein profiling,[37] we designed PL-probes containing a small alkyne tag either directly on the pyridine ring (**PL1**, **PL2**) or with an ethylene spacer (**PL4**), as well as a 2'-azide analogue (**PL3**) to account for chemical preferences within protein binding sites (Figure 2.2a). These probes are intended to make use of cellular PL-uptake mechanisms and metabolism to yield phosphorylated PLP derivatives capable of binding PLP-DEs (Figure 2.2b). Our strategy harnesses the intrinsic reactivity of the internal aldimine to anchor the probe to the enzyme irreversibly upon sodium borohydride (NaBH$_4$)-mediated reduction,[36] circumventing the need for additional reactive groups. A stable covalent bond is a prerequisite for downstream proteome manipulation, in which denaturing conditions are necessary for protein identification. Subsequent bioorthogonal ligation of the alkyne tag to biotin-azide or *vice versa* allows gel- and MS-based detection of the PLPome. This methodology was developed with the goal of generating effective chemical tools for the discovery, characterization and exploration of PLP-binding events in the complex biological context of the living cell.

PL-Probes
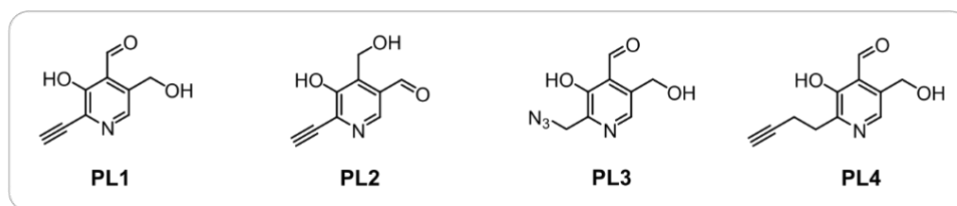
Labeling Strategy
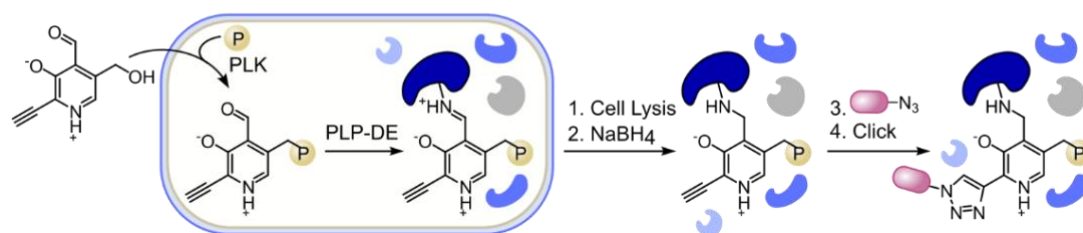
Figure 2.2 Proteomic strategy for profiling PLP-DEs. (**a**) Chemical structures of PL-based probes as tools for reporting on PLP-binding. (**b**) PLPome detection strategy: PL-probes are taken up by live *S. aureus* cells, phosphorylated and incorporated into PLP-DEs. Upon cell lysis, NaBH$_4$ reduction of the imine bond and click chemistry with fluorescent or enrichment tags permit proteomic identification of labeled PLP-DEs.

2.3   Synthesis

2.3.1   Synthesis of PL1 and PL2

A semisynthetic strategy starting from pyridoxine was developed for **PL1** and **PL2**, which was inspired by previous chemistry by Korytnyk *et al.*[54] (Scheme 2.1). While a 2'-alkynylated derivative of pyridoxine similar to **PL1** has previously been reported, its cofactor activity was not assessed.[54] For the synthesis, pyridoxine hydrochloride was first reacted with three equivalents of TBSCl to yield protected intermediate **1**. *N*-oxidation using *m*CPBA was followed by a Boekelheide rearrangement[64] using trifluoroacetic anhydride and quenching with MeOH. This served to install the 2'-alcohol of intermediate **2** and resulted in loss of the phenolic TBS group. The primary alcohol was then oxidized to the corresponding aldehyde (**3**) using MnO_2 and the phenol was reprotected using MOMCl. Reaction of **4** with the Ohira-Bestmann reagent[65] (dimethyl-1-diazo-2-oxopropylphosphonate) in the presence of K_2CO_3 gave terminal alkyne **5**. Upon removal of silane protecting groups, oxidation of **6** with MnO_2 in methanol gave a 2.5:1 mixture of isomers, **7a** and **7b**, which were separated by preparative HPLC and assigned by 2D NMR spectroscopy. A structural isomer of **PL1** with inverse functionality at the 4' and 5' positions that is electronically incapable of PLP catalysis (**PL2**) was obtained as a synthetic by-product and was later used as a control to test for non-specific reactivity of the PL-scaffold. Hydrolysis under acidic conditions yielded the final **PL1** and **PL2** probes. Based on the NMR spectra, the probes were primarily present in the favoured hemiacetal form.[53]
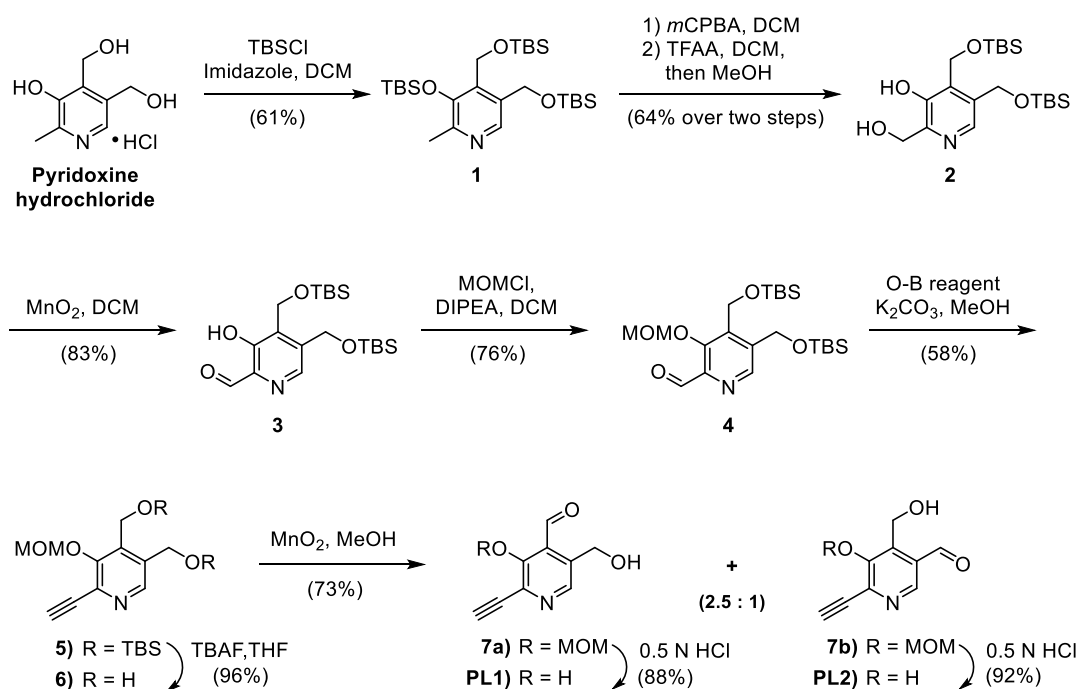
Scheme 2.1 Synthesis of PL1 and PL2. Pyridoxine hydrochloride was TBS-protected and oxidized at the 2-methyl position. An Ohira-Bestmann reaction was used to install the alkyne handle of probes **PL1** and **PL2**. Upon deprotection and MnO_2 oxidation to the aldehyde, isomeric probe precursors were separated by HPLC.

2.3.2 Synthesis of PL3

To access **PL3**, pyridoxine was first protected as the 3,4'-cyclic acetal **8** in acetone under acidic conditions[66] (Scheme 2.2). Protection of the 5'alcohol using PMBCl yielded intermediate **9**. Following *N*-oxidation, the Boekelheide rearrangement[64] was used to introduce a hydroxyl group at the 2' position of the PL-scaffold (**10**). Installation of a chloride at this position using $SOCl_2$ enabled subsequent displacement with sodium azide to generate **11**. Upon hydrolysis of the acetal, the primary alcohol was oxidized to the corresponding aldehyde **12** using MnO_2. Final deprotection under acidic conditions removed the PMB group to afford the **PL3** probe.
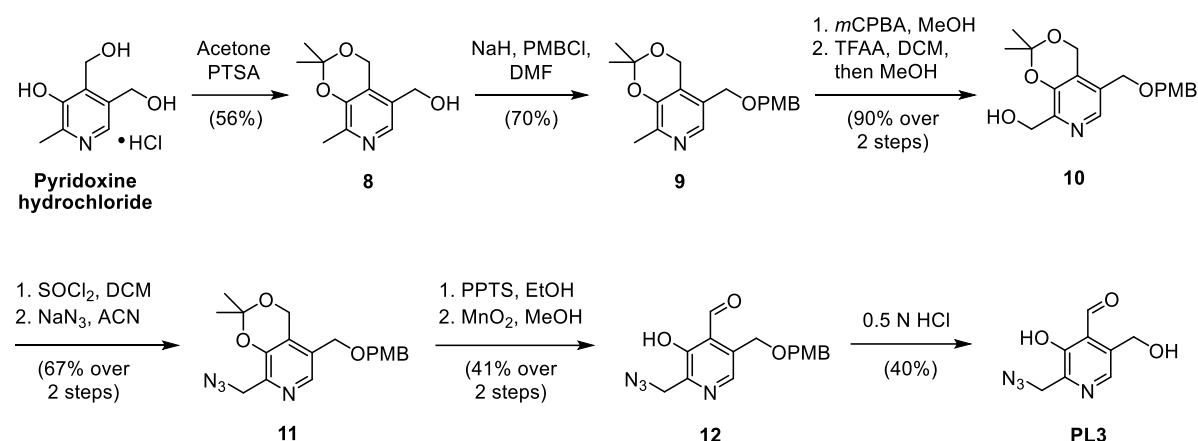
Scheme 2.2 Synthesis of PL3. Pyridoxine was protected as an acetal and PMB-ether, and oxidized at the 2-methyl position. The azide moiety was introduced by displacement of a transiently installed chloride using sodium azide.

2.3.3 Synthesis of PL4

In the case of **PL4**, we devised an efficient new synthetic route using an alkylation strategy inspired by previous work on methyl-pyridine derivatives.[67] **PL** was first protected as the cyclic monomethyl acetal[68] followed by MOM-protection of the phenol to generate intermediate **13** (Scheme 2.3). Deprotonation of the acidic 2-methyl position using LDA followed by reaction with propargyl bromide afforded the alkylated product **14**, which was then deprotected under acidic conditions to yield **PL4**. The protecting group strategy used in this synthesis was necessary in order to deactivate the 4' position of pyridoxal and prevent alkylation at this site. Prior attempts using the 3,4'-cyclic acetal system (intermediate **9**) resulted in 4'-alkylation. Our method provides a modular and facile route for accessing alkyl derivatives of PL modified at the 2-methyl position.
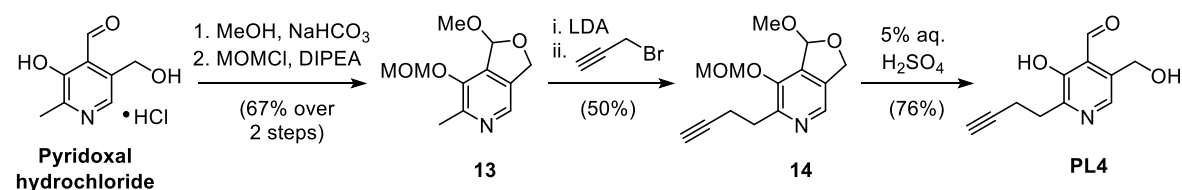
Scheme 2.3 Synthesis of PL4. Acetal protection of pyridoxal deactivates the 4'-position and makes the 2-methyl group available for deprotonation by LDA, followed by alkylation using alkyl halide reagents.

3. Biological Evaluation

3.1 Introduction

Vitamin B_6 is a collective term used to describe pyridoxal (PL), pyridoxine (PN) and pyridoxamine (PM), as well as their 5'-phosphorylated forms (PLP, PNP and PMP). A series of biosynthetic, catabolic and salvage enzymes generate and interconvert the different forms of vitamin B_6 to suit the needs of the cell (Figure 3.1).[69] While microorganisms and plants are capable of *de novo* PLP biosynthesis, animals rely on vitamin B_6 uptake from the diet. PL, PN and PM are believed to be absorbed through facilitated diffusion or active transport, though the exact mechanism of cellular uptake remains unclear.[70,71] A salvage pathway, consisting of pyridoxal kinases (PLKs, *PdxK*), pyridoxine-5'-phosphate oxidase (PNPOx, *PdxH*) and pyridoxal phosphatases, is used by all cells to recycle PLP liberated from enzymatic turnover or from nutritional uptake.[72] PLKs use ATP to phosphorylate the 5' alcohol of each of the three vitamers (PL, PN and PM), while PNPOx oxidizes PNP and PMP to form PLP. The salvage enzymes are believed to maintain homeostasis by regulating cellular PLP levels through feedback inhibition as well as the direct distribution of PLP to corresponding apoenzymes, though this is not fully understood.[22,23,72] The PLP aldehyde is much more reactive than its unphosphorylated counterpart, PL, which exists primarily in the hydrated acetal form. The concentration of PLP is therefore maintained at low levels to prevent undesired binding to cellular nucleophiles.[23,73] Studies have shown that imbalances in PLP metabolism and homeostasis can cause severe neurological pathologies.[74]

Two distinct PLP biosynthetic pathways have recently been characterized.[69,72] The DXP-dependent pathway is found among members of the γ-subdivision of proteobacteria, which include many Gram-negative bacteria (ex. *Escherichia coli*, *Pseudomonas aeruginosa* and *Salmonella enterica*). In this pathway, pyridoxine synthase (PdxJ) condenses de-oxyxylulose 5-phosphate (DXP) with 3-hydroxy-1-aminoacetone phosphate to generate PNP, which is then converted to PLP by PNPOx (PdxH). Although it was discovered first, the DXP-pathway is not widely distributed and rather complex. The R5P pathway is more common and employs a PLP synthase complex, consisting of Pdx1 (PdxS) and Pdx2 (PdxT) subunits, to condense D-ribose 5-phosphate (R5P), glutamine and glyceraldehyde 3-phosphate, generating PLP directly. PLP catabolism proceeds through oxidation

of PL, mainly to 4-pyridoxic acid, and excretion in the urine by humans and animals. Microorganisms can further degrade the cofactor to usable metabolites such as succinic semialdehyde or 2-(hydroxymethyl)-4-oxobutanoate, acetate, ammonia and carbon dioxide.
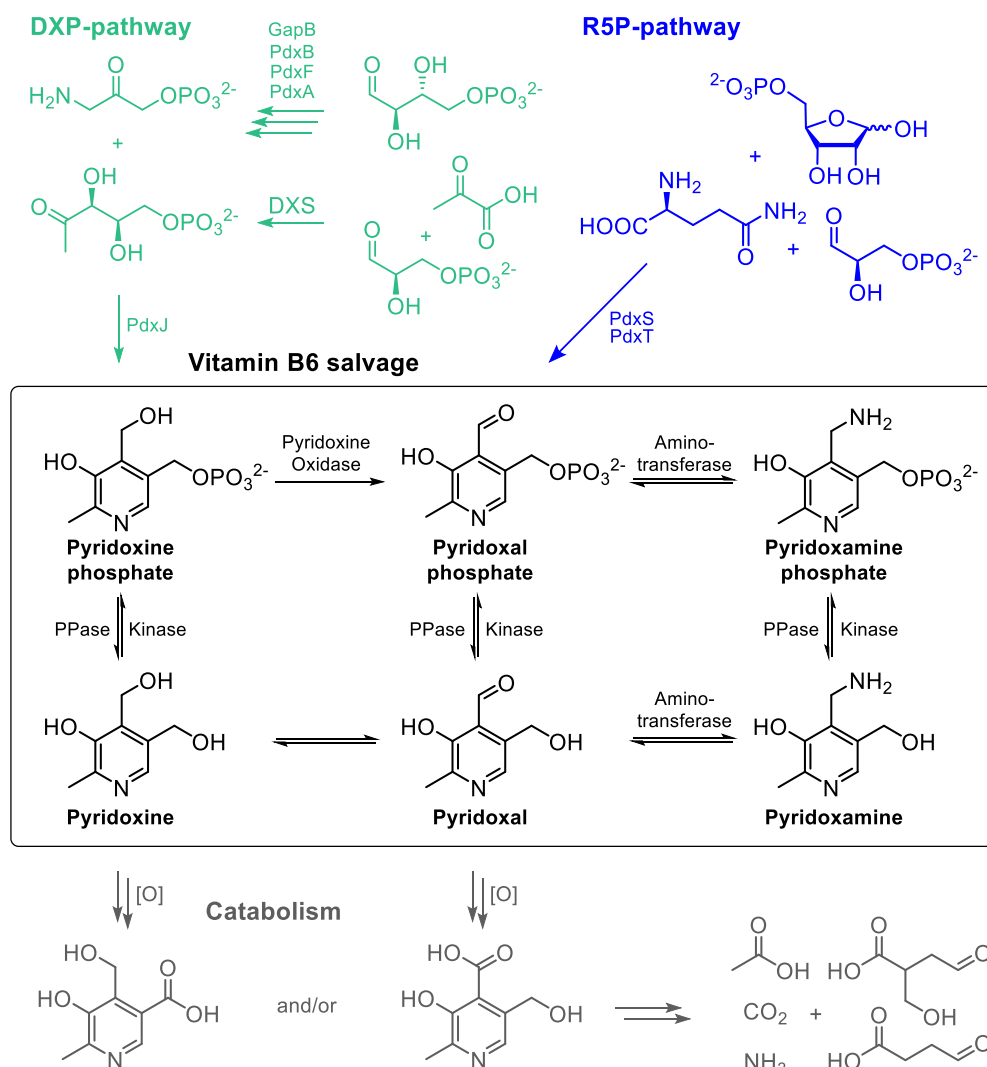
Figure 3.1 PLP metabolism. Microorganisms synthesize PLP via DXP- or R5P-dependent pathway. A vitamin B_6 salvage pathway found in all organisms interconverts the different forms of the cofactor as needed by the cell. Catabolism occurs through oxidation of PL to pyridoxic acid, followed by excretion (humans, animals) or further degradation (microorganisms).

3.2   Probe phosphorylation by *S. aureus* pyridoxal kinase

PLP is the main bioactive component of vitamin B_6 and it is generated through the phosphorylation of PL by PLKs. The 5'-phosphate group is important for binding PLP-DE active sites, where it helps to anchor and correctly position the cofactor through hydrogen-bonding interactions.[58] *S. aureus*

PLK (*PdxK*) is capable of phosphorylating all forms of vitamin B_6 (PL, PN, PM) and was recently discovered to have roles in both PLP salvage as well as thiamine biosynthesis.[75] Structural studies revealed *S. aureus* PLK to adopt a classic ribokinase fold and lent insights into its unique mechanism of action. A flexible loop of PLK engages the pyridoxal aldehyde as a hemithioacetal at a highly conserved cysteine residue and folds down in order to sequester the phosphorylation reaction within the catalytic pocket of the enzyme.[75]

We first studied the ability of *S. aureus* PLK to accept our synthetic probes as substrates in order to yield activated cofactor mimics that can bind to PLP-DEs. The kinetics of probe phosphorylation by PLK were therefore determined using a previously described assay that couples the ATP-dependent activity of PLK to the consumption of NADH (340 nm) via a pyruvate kinase/lactate dehydrogenase system.[75] Both **PL3** and **PL4** were efficiently converted to phosphorylated products at levels comparable to PL (Figure 3.2). Formation of **PL3P** and **PL4P** could be confirmed by MS upon overnight incubation of the probes with PLK in the presence of ATP to generate the phosphorylated products. Although significant overlap in the absorption spectra of NADH (340 nm) and **PL1** (325 nm, extended conjugation) precluded kinetic analysis, MS-based detection of **PL1P** served to confirm phosphorylation by PLK. In contrast, phosphorylation could not be detected for **PL2**, the negative control probe.

MS Measurement

Substrate	MW$_{calculated}$ (g/mol)	MW$_{found}$ (g/mol)
PL1P	258.01620	258.01666
PL2P	258.01620	N/A
PL3P	289.03325	289.03403
PL4P	286.04750	286.04819

PLK Kinetics

Substrate	k_{cat} (min^{-1})	K_m (µM)	V_{max} (µM/min)
PL	6.72 ± 0.60	81 ± 29	2.34 ± 0.30
PL1	N/A	N/A	N/A
PL3	3.94 ± 0.58	86 ± 35	1.97 ± 0.29
PL4	4.66 ± 0.50	72 ± 23	2.33 ± 0.25
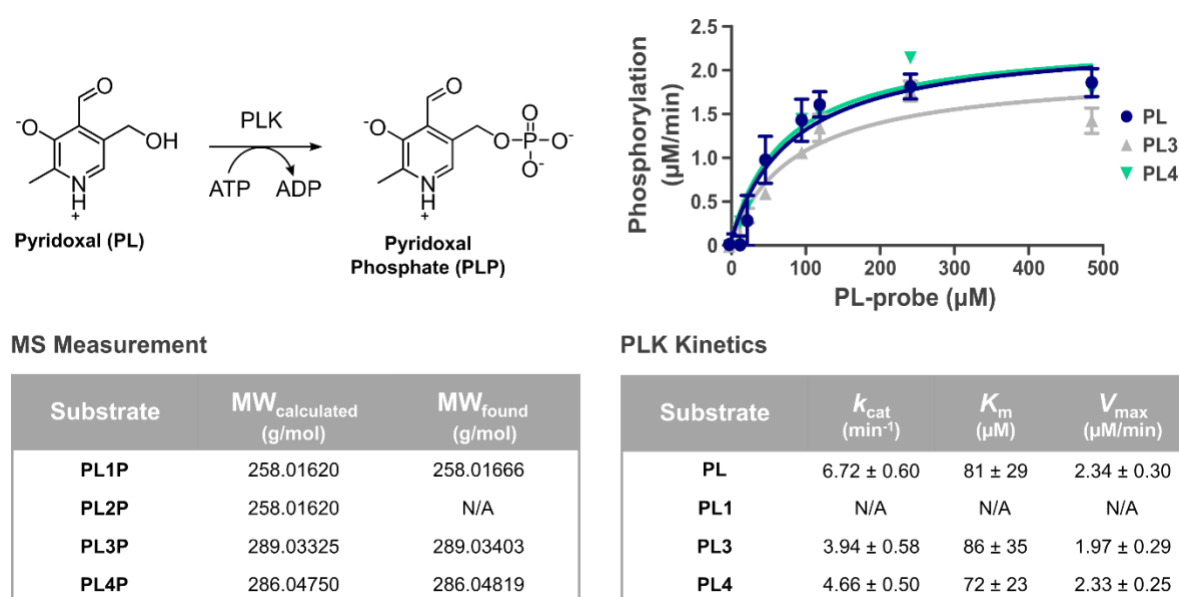
Figure 3.2 Probe phosphorylation by *S. aureus* PLK. PLK uses ATP to phosphorylate the 5'-OH of PL. Kinetics of probe phosphorylation compared to PL formation are shown (*n* = 4, error bars: mean ± standard deviation) as well as HRMS confirmation of phosphorylated products. Kinetic analysis of **PL1** (325 nm) was not possible due to significant spectral overlap with NADH (340 nm).

3.3   Cofactor viability of probes with alanine racemase

Next, we examined the compatibility of the phosphorylated probes with alanine racemase (Alr) as a model PLP-DE since it is well-studied and universally present in bacteria. Alr catalyzes the interconversion of L- and D-alanine, an essential component for bacterial peptidoglycan synthesis, and is therefore considered an important antibiotic target.[76] Recombinant, strep-tagged *S. aureus* Alr (Uniprot ID: P63479) was first purified as the PLP-holoenzyme using affinity chromatography and was subsequently transformed to its apo-form (Figure 3.3a). Methods for resolving the PLP cofactor from PLP-DEs are enzyme-dependent, but have included nucleophilic displacement by small molecule amines or extensive dialysis.[77] In some cases, harsher treatments such as partial unfolding or precipitation of the enzyme are required in order to remove the cofactor completely.[59,78] On-column nucleophilic displacement of bound PLP using hydroxylamine was effective at generating the alanine racemase apoenzyme (Apo-Alr), and its formation could be monitored spectrally by the loss of absorbance of the internal aldimine (410-420 nm).[60] Upon reconstitution of Apo-Alr with the phosphorylated probes, UV-Vis peaks indicative of internal aldimine formation were observed (Figure 3.3c). Probe binding and the degree of cofactor loading within the different holoenzymes were further confirmed by MS (Figure 3.3d). Therefore, the holoenzymes were reduced using $NaBH_4$ and measured by intact-protein MS, revealing protein adduct formation (80-100%) corresponding to the respective masses of the phosphorylated probes. Having confirmed reductive amination of the internal aldimine to stably fix the probes to Alr, we subsequently investigated whether Alr could be chemically labeled with a fluorescent reporter tag using copper-catalyzed azide-alkyne 1,3-dipolar cycloaddition (CuAAC, click chemistry) with rhodamine-azide.[79-81] Visualization by fluorescence SDS-PAGE revealed specific labeling of Alr by **PL1P** and **PL4P** (Figure 3.3e), while unphosphorylated probes and the control probe **PL2** helped gauge nonspecific binding of the reactive aldehyde. For **PL3P**, CuAAC with rhodamine alkyne proved more specific over strain-promoted click chemistry (SPAAC) with DBCO-rhodamine for labeling Alr, presumably due to nucleophilic attack of other enzyme residues on the strained alkyne system.[82]

Subsequently, we investigated the capacity of the phosphorylated probes to support enzyme function. The ability of the individual probe-bound Alr holoenzymes to catalyze alanine racemization was therefore characterized kinetically. The synthetic cofactors yielded catalytically competent Alr-holoenzymes with nearly identical efficiency compared to the PLP-holoenzyme (Figure 3.3b). **PL1P** even showed higher turnover compared to PLP, which may result from increased stabilization due to its extended conjugation or perhaps favoured binding within the Alr active site. Such hyper-activity has previously been described for other cofactor derivative-enzyme systems.[63] The remarkable structural tolerance within the Alr active site was consequently reaffirmed by X-ray crystallography.
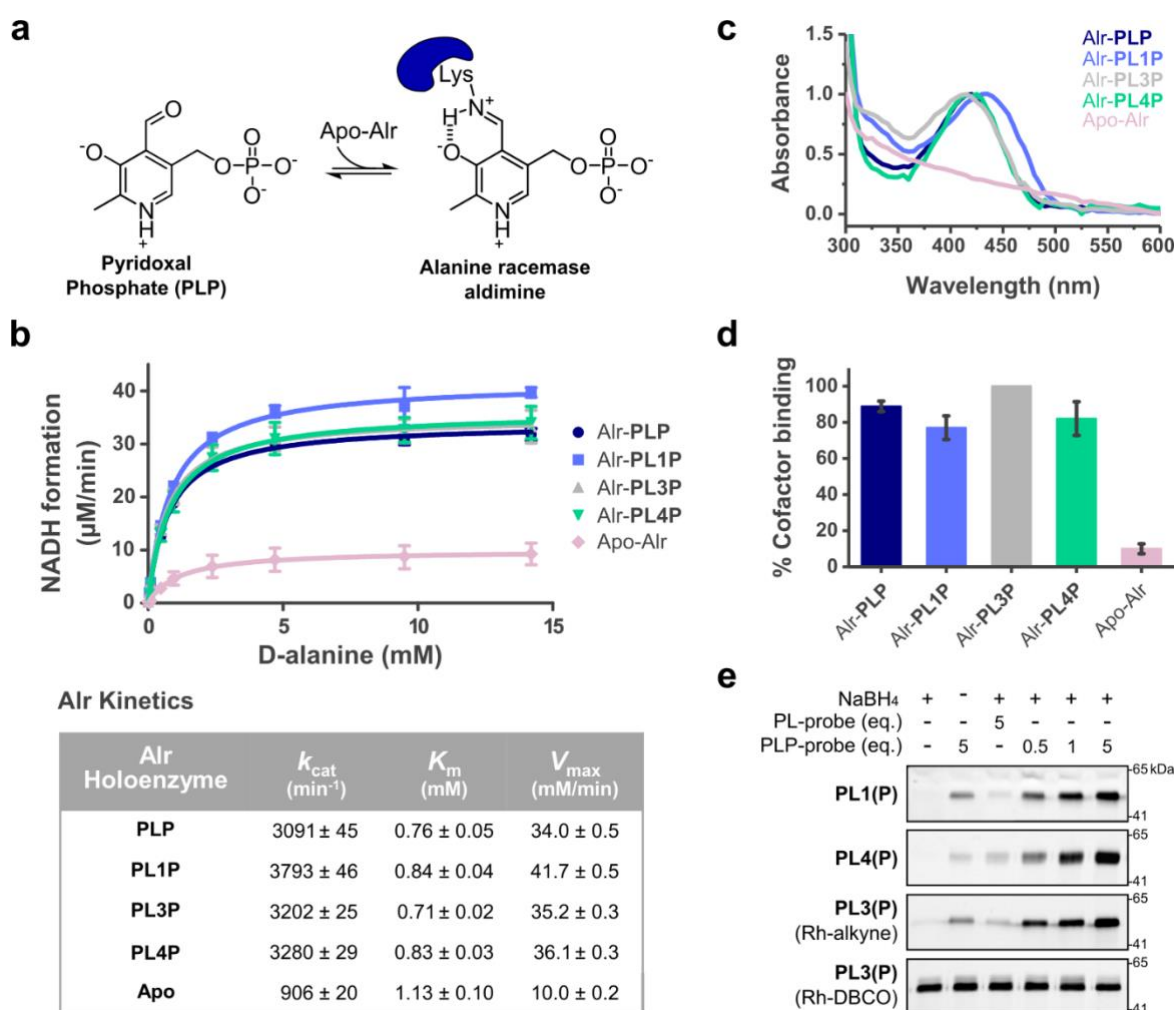
Figure 3.3 Reconstitution and activity of Alr with cofactor probes. (**a**) Reconstitution of PLP or phosphorylated probes with Alr. (**b**) Kinetics of alanine racemization (D- to L-Ala direction) by *S. aureus* Alr holoenzymes (0.01 µM) containing PLP or cofactor analogues (*n* = 3, error bars: mean ± standard deviation). (**c**) UV-Vis absorption spectra of different *S. aureus* Alr holoenzymes (100 µM). Data was normalized to absorbance of the internal aldimine (420-430 nm). (**d**) Percent of *S. aureus* holoenzymes bound by PLP or cofactor analogues upon NaBH$_4$ reduction and measurement by intact protein MS (*n* = 3, error bars: mean ± standard deviation). (**e**) Fluorescence SDS-PAGE of *S. aureus* Alr (10 µM) labeling using PLP-probes (0.5, 1, 5 equivalents) or unphosphorylated PL probes (5 equivalents) upon click chemistry to rhodamine-azide or alkyne. DBCO = dibenzyl cyclooctyne. See appendix for full gels and Coomassie staining.

In order to understand the binding mode of the PLP-derivatives within the Alr active site, we determined the X-ray crystal structures of Alr bound to **PL1P** and **PL4P**, representing the smallest and largest probes respectively. Several crystal structures of Alr have been resolved, which revealed a conserved overall fold (type III) consisting of an N-terminal α/β barrel and a predominantly β-stranded C-terminus. Alr functions as a homodimer, forming two active sites that are lined with positive charge at the dimer interface, and PLP binds to a highly conserved lysine residue (Lys39) near the mouth of the barrel. Both **PL1P** and **PL4P** bound the structurally equivalent position of the natural PLP cofactor (comparison with PDB: 4A3Q,[83] Figure 3.4), and

clear electron density was indicative of internal aldimine formation with Lys39. The phosphate groups are coordinated by a conserved hydrogen-bonding network present in all available structures of Alr and its homologues.[83] Though the active site geometry generally matched the PLP-holoenzyme, a slight realignment of the Lys131 and Arg138 sidechains was necessary to fit the alkyne handle of the PLP-analogues. Our studies confirm the adequate space within the Alr aldimine binding pocket for accommodating structural changes imparted to the synthetic cofactor without influencing catalytic activity, and led us to conclude that the probes represent good cofactor mimics for binding and reporting on PLP-DEs.

Figure 3.4 Alr crystallography. (**a**) Crystal structure of *S. aureus* Alr bound to **PL1P** (left) and **PL4P** (center). Simulated annealing omit electron density map contoured at 2.5 σ. Superimposed active sites (right) of PLP-bound Alr (purple, PDB 4A3Q),[83] Alr-**PL1P** (grey) and Alr-**PL4P** (blue). Residues lining the active site as well as the **PL1P** (grey) and **PL4P** (gold) cofactors are shown in stick models; hydrogen bonds are drawn as dashed lines. (**b**) Surface representation of the *S. aureus* Alr dimer (blue, grey) in complex with **PL4P** (gold, stick representation); the two active sites are indicated by red arrows. (**c**) Two-dimensional LIGPLOT1-representation of the interactions between *S. aureus* Alr and **PL4P**. Figures courtesy of Dr. S. Schneider.

3.4 Cellular growth, metabolism and labeling by PL-probes

The ability of our PL probes to function globally in native biological systems was assessed using bacterial growth experiments. To promote the cellular uptake and use of our probes as PL surrogates, we used a transposon mutant strain of *S. aureus* that is incapable of *de novo* PLP biosynthesis but is able to salvage exogenous PL and convert it to PLP. Since *S. aureus* employs the R5P-biosynthetic pathway, which requires the functioning of a PLP-synthase complex consisting of PdxS and PdxT subunits to generate PLP, we reasoned that disruption of the *pdxS* gene would provide the desired phenotype. The *S. aureus* USA300 TnPdxS transposon mutant (Nebraska transposon mutant library)[84] was therefore selected. Bacteria were incubated in defined, PL-free media supplemented with various concentrations of the probe and growth was measured after 24 h. We found that < 1 μM PL was required to support full bacterial growth, while absence of PL or incubation with **PL2** prohibited growth (OD_{600} < 0.1), thereby confirming the transposon phenotype. Both **PL1** and **PL3** were able to support bacterial growth to the same maximal level as PL, though required much higher concentrations (approximately 25 μM) (Figure 3.5). However, *S. aureus* cells were unable to grow using **PL4** as a cofactor substitute under the conditions tested. These results suggest that small modifications of the cofactor are tolerated for a range of cellular functions, while the increased steric bulk of the **PL4** alkyne linker may preclude its use for certain essential biological pathways.
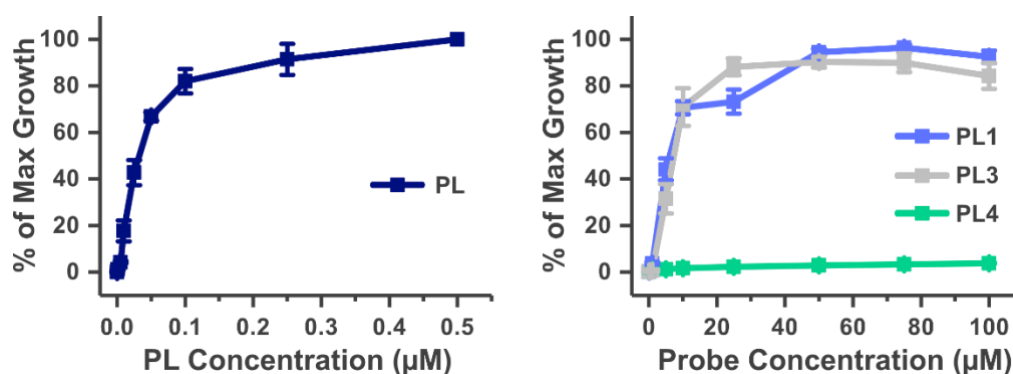
Figure 3.5 Bacterial growth using synthetic cofactor mimics. Growth curves of *S. aureus* USA300 TnPdxS (Nebraska transposon library)[84] in defined, PL-free media supplemented with PL (left) or PL probes (right). Bacterial growth was measured (OD_{600}) after 24 h and data was obtained from three biological replicates (error bars: mean ± standard deviation).

Our studies on the growth of *S. aureus* TnPdxS with our probes as cofactor surrogates indicated that forcing conditions would be required to compete with endogenous PLP levels for PLP-DE binding. However, as the probes were unable to support sufficient bacterial growth on synthetically practical amounts for labeling experiments, we adopted an alternative strategy in which bacteria were first grown to stationary phase in PL-rich media and subsequently transferred to chemically-defined media containing 100 μM of the respective probes in place of PL. Upon completion of a two-hour labeling period with the probes, the cells were lysed and treated with NaBH$_4$ in order to fix the probes to the enzymes by reductive amination. Conjugation to rhodamine azide by CuAAC and fluorescence SDS-PAGE analysis revealed various extents of protein labeling by the different probes (Figure 3.6a). While **PL4** labeled a large number of proteins very strongly, the **PL1** probe showed only weak labeling above **PL2** or DMSO treatment (controls). We further investigated the reason for the poor labeling efficiency of **PL1** and found labeling of purified Alr-**PL1P** to decrease significantly when spiked into cell lysate, media or reaction mixtures containing nucleophilic amino acids such as cysteine (Figure 3.6b). Labeling could be rescued to a certain degree by quenching reactive nucleophiles using iodoacetamide alkylation[82] prior to click chemistry. Furthermore, adduct formation could be observed by MS when incubating **PL1** with *N*-acetyl-cysteine in the presence of copper (data not shown). We therefore reasoned that copper may coordinate to the **PL1** pyridine and activate the 2'-alkyne, which is in conjugation with the aromatic system, towards conjugate addition by cellular nucleophiles. Future proteomic experiments were therefore conducted using an additional iodoacetamide treatment for **PL1**-labeled samples. Similarly, **PL3**-based labeling proved weak over background. The reasons for this are unclear, but may stem from lower specificity and efficiency of inverse click reactions.[80]
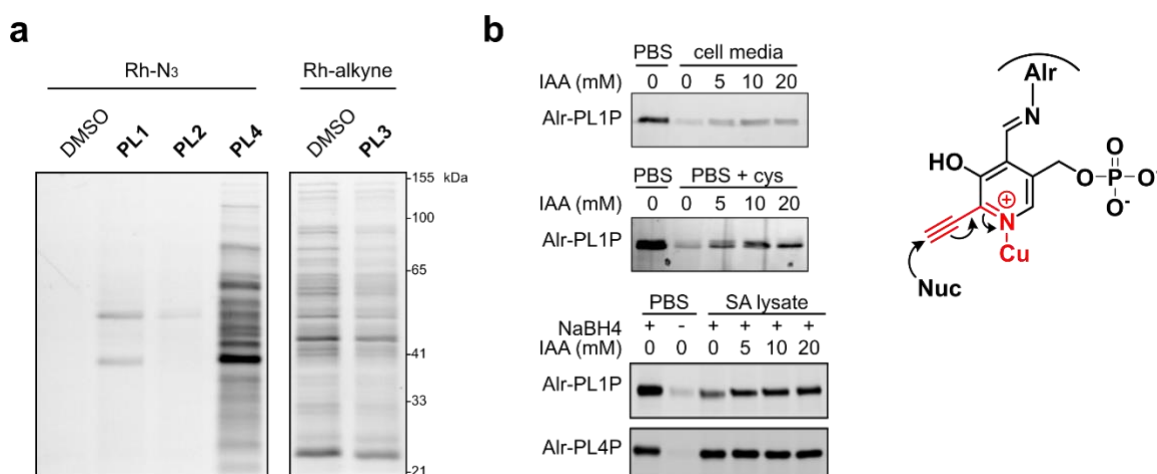
Figure 3.6 Cellular probe labeling. (**a**) Fluorescence SDS-PAGE showing probe labeling (100 μM for 2 h at 37°C) in *S. aureus* USA300 TnPdxS. See appendix for Coomassie stained gels. (**b**) Optimizations of **PL1**-labeling. Fluorescence SDS-PAGE showing labeling of Alr-**PL1P** in the presence of cell media (CDM, top), PBS containing 1 mM cysteine (middle), or *S. aureus* lysate (bottom). **PL1P**-bound Alr holoenzyme (5 μM) was reduced with NaBH$_4$ (10 mM) and treated with iodoacetamide (IAA) prior to CuAAC with rhodamine azide. The proposed mechanism of inactivation of **PL1P** by conjugate addition of cellular nucleophiles in the presence of copper is illustrated.
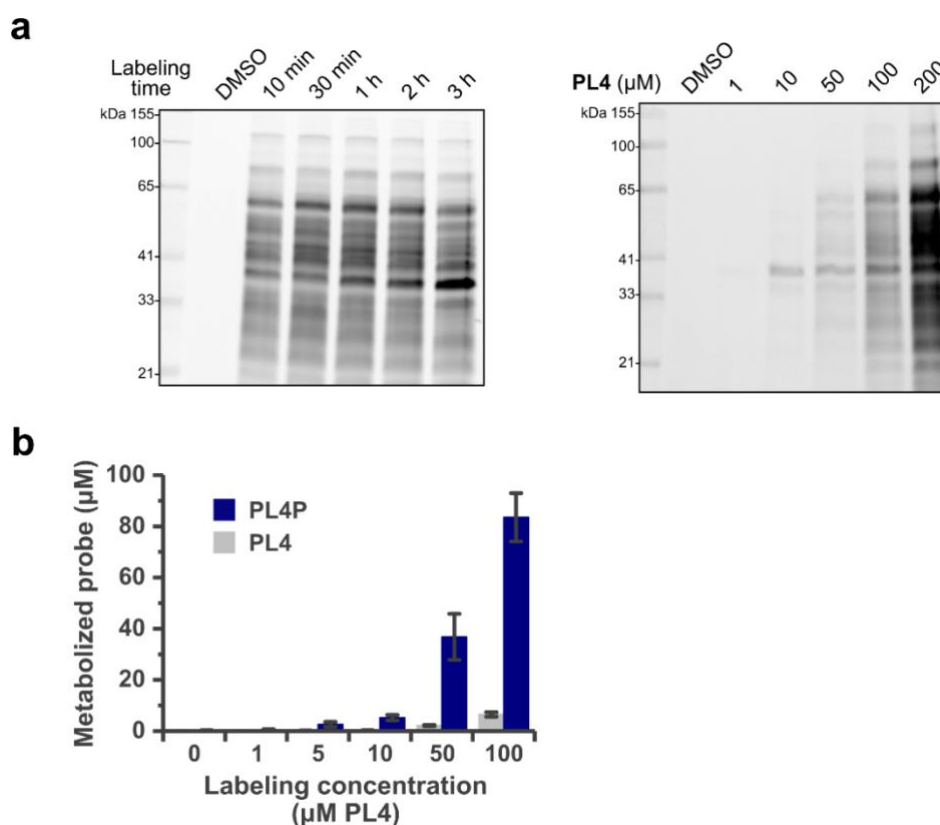
Figure 3.7 Optimization of PL4-labeling. (**a**) Time-dependent (left) and concentration-dependent labeling (right) of **PL4** in *S. aureus* USA300 TnPdxS by fluorescence SDS-PAGE. See appendix for Coomassie stained gels. (**b**) *In situ* metabolism of **PL4** to **PL4P** (*n* = 3, error bars: mean ± standard deviation) in *S. aureus* USA300 TnPdxS (100 µM for 2 h at 37°C).

Further optimizations with regards to labeling time and probe concentration were directed at **PL4** due to its robust labeling profile. We found that incubation time does not significantly affect labeling efficiency and that **PL4**-based labeling is concentration-dependent (Figure 3.7a). Using a targeted metabolomics approach, we also verified *in situ* probe phosphorylation under these experimental conditions. It was found that approximately 80% of the probe was converted to its phosphorylated form within 2 h of incubation with the bacteria, and formation of **PL4P** increased with higher probe concentrations (Figure 3.7b). A concentration of 100 µM of **PL4** was eventually selected as a good balance between labeling strength and non-specific background binding, which was later reaffirmed through proteomics experiments.

3.5 Conclusion

In summary, the biological activity and suitability of three PL-derivatives as chemical probes for reporting on PLP-binding were evaluated. Each probe is efficiently phosphorylated to its biologically active form by the *S. aureus* PLK at levels comparable to PL, and binding was assessed

by reconstituting the model PLP-DE Alr with the activated probes. The catalytic activity of artificial Alr holoenzymes was not affected by the chemical modification of the cofactor, which can be attributed to structural accommodation within the Alr binding site. Labeling experiments with Alr-**PL3P** revealed significant background labeling linked to the reactivity of alkyne- and DBCO-rhodamine reagents, while cellular **PL1**-labeling suffered from the conjugate addition of biological nucleophiles. Although these probes showed the best biological uptake and incorporation, as evidenced by their ability to support bacterial growth as cofactor surrogates, chemical factors limit their use as effective and specific probes under the present experimental conditions. By comparison, the extended bulk of **PL4** prevented it from substituting for all essential cellular PLP-functions, but strong gel-based labeling led to its selection for further optimizations and cell-based metabolic experiments. We were encouraged by previous observations that PLP-derivatives are sometimes unable to support enzymatic catalysis though are nonetheless able to bind PLP-DEs. With a robust PLP-labeling strategy in hand, we proceeded with MS-based proteomic investigation in order to identify and quantify cellular proteins accessed by the PL-probes in *S. aureus*.

4. Proteomic Analysis

4.1 Introduction

Mass spectrometry represents a powerful analytical platform for proteomic investigations. The development of advanced MS-based methods that are capable of identifying and quantifying proteins have made possible the detailed and global comparison of biological systems. Furthermore, in combination with chemical tools to selectively enrich protein subtypes, MS-analysis can be used to study distinct cellular events of interest. Recent technological advances in peptide separation techniques, mass spectrometer instruments and data analysis have been critical in establishing MS-methods for routine use in diverse applications and proteomic analyses.

Quantitative proteomic analysis has traditionally relied on the use of dyes, fluorophores or radioactive tags in combination with high-resolution protein separation techniques, such as 2D gel electrophoresis.[85] However, these methods are limited by their inability to simultaneously identify proteins, a shortcoming which has been addressed by modern LC-MS/MS techniques. A challenge for the absolute quantification of proteins by MS is the range in physicochemical properties of proteolytic peptides (size, charge, hydrophobicity). As such, strategies for the relative quantification of peptides between different samples using stable isotope tags have been developed.[86-88] These labels are introduced either at the level of the protein or peptide by metabolic, enzymatic or chemical means, and resulting peptides behave identically to their non-labeled counterparts throughout LC but can be distinguished by MS.[89] For metabolic incorporation, labels are introduced to the living cell by substituting heavy- or light-isotope amino acids into the growth medium (ex. SILAC; stable isotope labeling by amino acids in cell culture).[90] The early-stage introduction of the labels allows samples to be pooled and processed together, minimizing handling or instrumental errors to produce high-accuracy results. Alternatively, labels can be introduced chemically at the peptide level. For example, in a method termed dimethyl labeling, deuterated and ^{13}C-labeled formaldehyde and sodium cyanoborohydride ($NaCNBH_3$) are used to methylate lysine residues by reductive amination.[91] While dimethyl labeling is affordable and effective, only three experimental conditions can be compared due to the limited number of reagent combinations that produce a detectable mass difference (≥ 4 Da) and minor shifts during

chromatography can complicate data analysis. These challenges have been addressed by isobaric mass tagging techniques, such as iTRAQ (isotope tags for relative and absolute quantification)[92] and tandem mass tags (TMT).[93] Here, labeled peptides co-migrate exactly during LC and are only distinguished and quantified upon MS2 fragmentation, while multiplexing of up to 10 samples is possible. Finally, label-free quantification (LFQ) strategies have recently gained popularity due to their ability to compare unlimited samples at a low cost. Since comparative samples are processed independently, errors in sample handling and MS measurement are reflected in the obtained data and advanced data analysis and statistical evaluation methods are required to extract reliable quantitative data from LFQ-based MS measurements.[94] In conclusion, the field of MS-based proteomic quantification is rapidly developing and, with it, technological advances in mass spectrometry instruments as well as methods for bioinformatic data analysis are keeping pace.

Previous investigations into the cellular PLPome have labeled and visualized PLP-binding proteins by Western blot using an α-pyridoxine antibody.[35] However, these reports were unable to identify and quantify distinct PLP-DEs, which could be addressed through MS-based analysis. Direct analysis of PLP-modified peptides from cell mixtures is complicated by their suppressed ionization efficiency and poor detection, reminiscent of phosphopeptides.[36,95] Neutral-loss scanning for phosphate and PLP moieties during mass fingerprinting has been effective at localizing the PLP-binding site of purified protein.[36] An additional enrichment strategy using immobilized metal ion affinity chromatography (IMAC) coupled to PLP phosphatase treatment identified only few sites in *E. coli* lysate.[95] We present an MS-based proteomic approach using chemical probes for the enrichment and identification of PLP-DEs. We apply supplementary data analysis strategies in order to add confidence to LFQ data and to improve the accuracy of the method for our purposes.

4.2 *En route* to an optimized proteomic workflow

Having established a viable labeling strategy supported by in-depth biochemical characterization (*Chapter 3*), we proceeded with the development of an MS-based proteomic workflow to evaluate PL-probe binding in *S. aureus* USA300 TnPdxS cells. Experiments were initially optimized using the **PL4** probe due to its promising gel-based labeling pattern, and subsequently applied to **PL1**, **PL2** and **PL3**. A stationary-phase bacterial culture was therefore treated with **PL4** in defined media, devoid of PL, over the course of a two-hour labeling period. Upon cell lysis and reductive fixation of the probe to PLP-DEs, a protein precipitation step followed by washing of the protein fraction was necessary to remove excess NaBH₄. The labeled proteins were then solubilized (0.4% SDS in PBS) and conjugated by CuAAC to a trifunctional rhodamine-biotin-azide linker[96] for enrichment using avidin-beads. The introduction of a fluorescent tag onto labeled proteins was critical during experimental optimizations in order to qualitatively assess protein enrichment by gel. Samples were subsequently trypsin digested and the resulting peptides were analyzed by LC-MS/MS.

Experiments were initially performed using dimethyl labeling as a basis for proteomic quantification of samples treated with 100 µM **PL4** *versus* DMSO. Originally, bacteria were grown to stationary phase directly in CDM containing minimal amounts of PL (1 µM), in order to promote maximal probe uptake and metabolism. Our first experiment revealed enrichment of a few PLP-DEs upon analysis by volcano plot, in which the quantitative fold-difference of a protein between **PL4**-enriched and control (DMSO) samples is plotted *versus* significance (*p*-value) (Figure 4.1a). While this preliminary finding hinted to us that our method was working, poor reproducibility and low PLP-DE enrichment led us to pursue alternative quantitation strategies. We therefore performed subsequent experiments using label-free quantification[94] in order to avoid problems with the labeling efficiency of the dimethyl-labeling reagents.
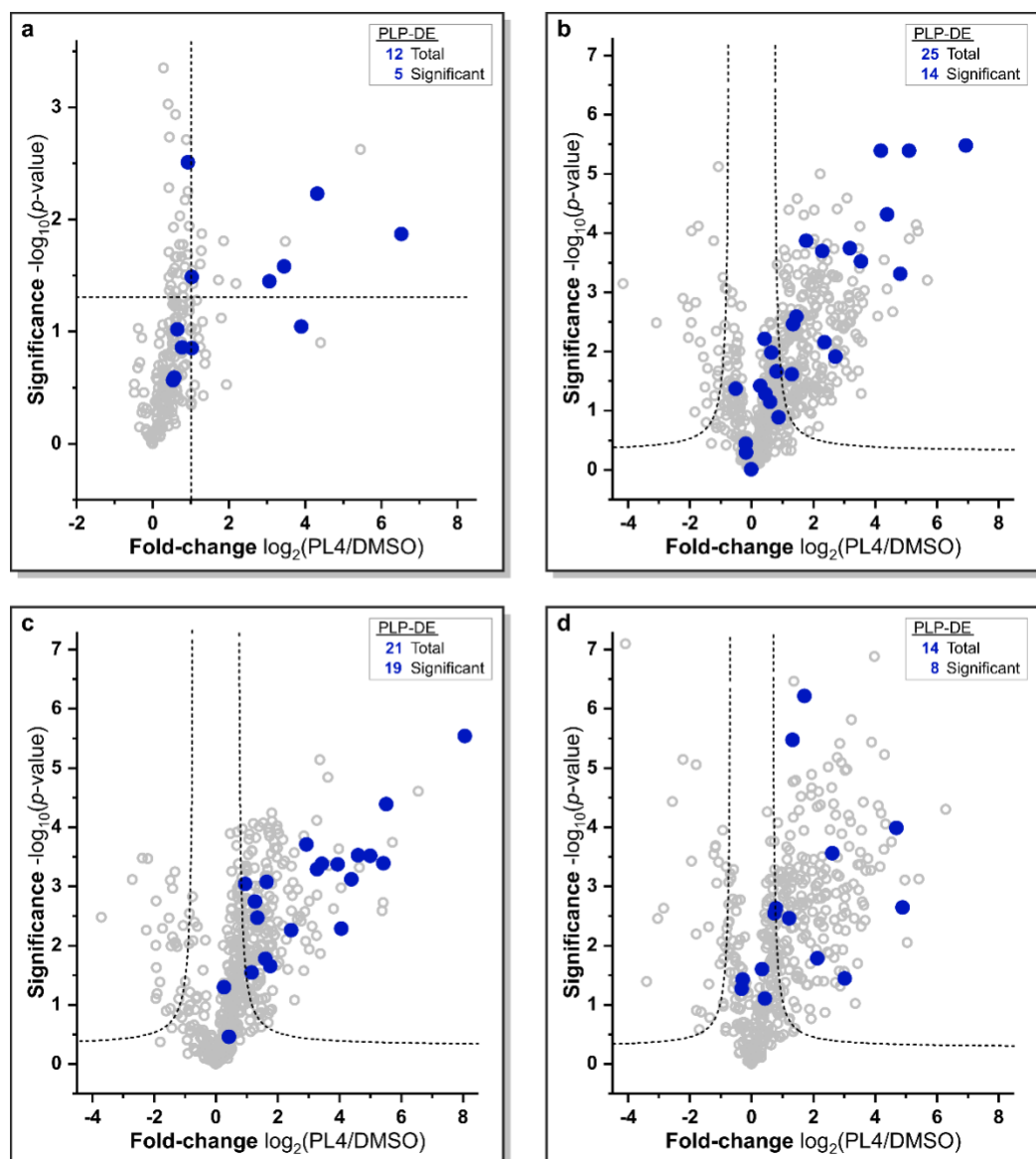
Figure 4.1 Proteomics-based optimization of PL4-labeling in *S. aureus*. Volcano plots showing enrichment of proteins using 100 µM **PL4** (*n* = 3 biological replicates) and significance upon performing a student's *t*-test (FDR 0.05, s0 1). (**a**) *S. aureus* USA300 TnPdxS grown in CDM + PL (1 µM) and analysis using dimethyl labeling. (**b**) *S. aureus* USA300 TnPdxS grown in CDM + PL (1 µM) and LFQ. (**c**) *S. aureus* USA300 TnPdxS grown in rich medium, followed by **PL4**-labeling in CDM and analysis using LFQ. (**d**) *S. aureus* NCTC8325 grown in rich medium and LFQ.

Repetition of the above experiment with LFQ analysis led to the identification of a higher number of proteins in total, including 25 PLP-DEs of which 14 were significantly enriched (Figure 4.1b). In order to increase the number of identified PLP-DEs further, we focused on optimizing the bacterial growth and labeling conditions. Since *S. aureus* USA300 TnPdxS grew very slowly and to a lesser overall extent under the stress of minimal medium, we opted to first grow healthy cultures in rich medium and later transfer them to CDM for the labeling reaction. Using these conditions, we found an increase in the number of significantly enriched PLP-DEs and improved their separation over background (Figure 4.1c). Comparison of labeling in wild-type *S. aureus* (NCTC8325) showed an overall lower enrichment of PLP-DEs (Figure 4.1d), presumably because the bacteria are able to synthesize PLP endogenously in order to out-compete our synthetic cofactor probes. This highlighted the need to use the TnPdxS transposon mutant of *S. aureus* for proteomic labeling experiments. Using this method, we found high reproducibility between different experiments and were able to significantly improve the enrichment statistics of PLP-DEs over background binding with increasing replicate numbers (9 biological replicates total, Figure 4.2a,b).
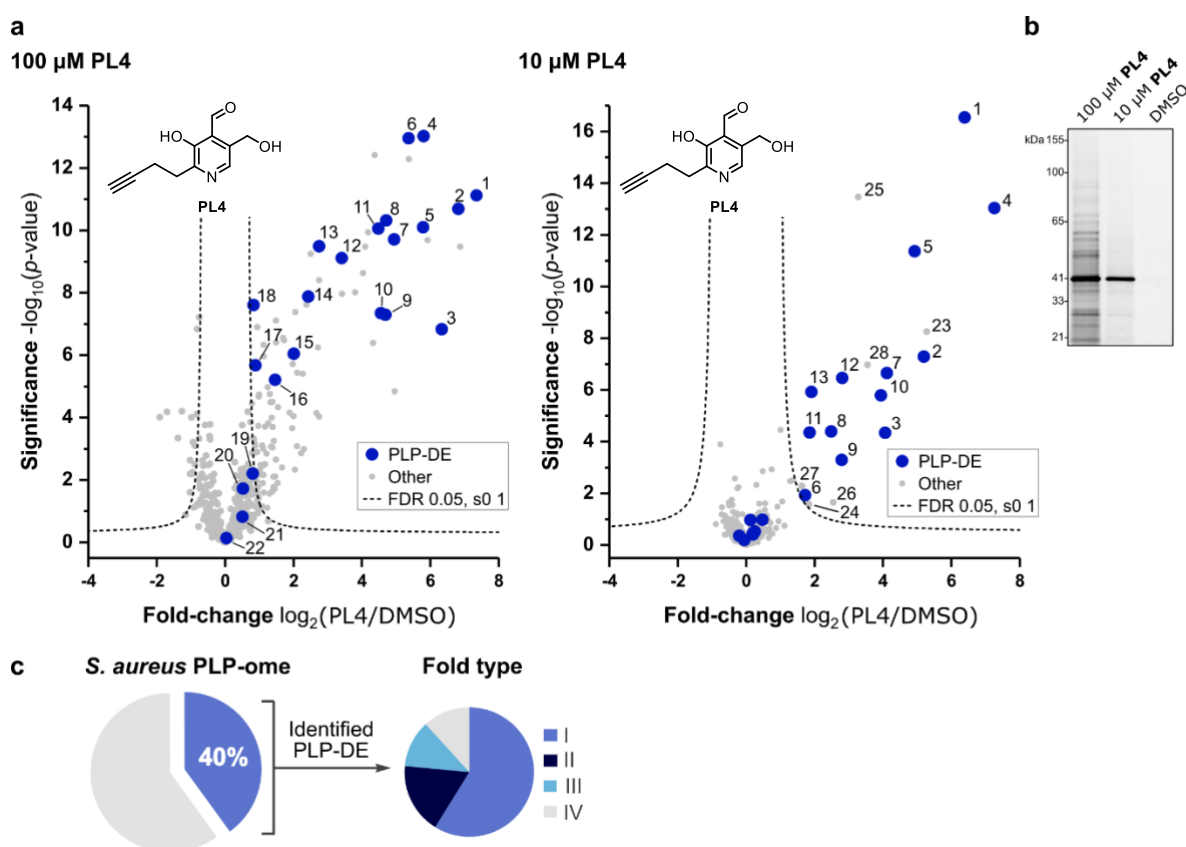
Figure 4.2 PL4 labeling 100 µM, 10 µM. (**a**) Volcano plot visualization of fold-enrichment using 100 µM (left) or 10 µM (right) **PL4** compared to DMSO *versus* significance upon performing a two-sample *t*-test (FDR 0.05, s0 1; *n* = 9 biological replicates). Known PLP-DEs are highlighted with blue dots and are numerically annotated according to Table 4.1. (**b**) Fluorescence SDS-PAGE of **PL4**-labeling in proteomic experiments upon CuAAC to trifunctional rhodamine-biotin-azide tag and avidin bead enrichment. (**c**) Proportion of the *S. aureus* PLPome (40%) significantly identified in proteomic experiments and distribution of PLP-DE fold-types.

Under the optimized conditions, we observed a striking enrichment of proteins annotated to bind or utilize PLP (18 out of the 104 significant proteins) by gene ontology (GO-term: 0030170; PLP-binding) or EC classification (GO overrepresentation analysis for pyridoxal phosphate binding at 100 µM: enrichment factor 4.2, *p*-value 1.3e-10). PLP-DEs of diverse functionality and fold type were identified, accounting for 40% of all PLP-DEs annotated in *S. aureus* USA300 so far (45 total) (Figure 4.2c). To improve the reliability of our LFQ method, we established a filtering strategy in order to explore our proteomic data and to accurately assign PLP-binding to unknown proteins with high confidence.

4.3 Data processing and filtering strategy

With the goal of identifying novel PLP-DEs, we attempted to distinguish PLP binding from background within our data set by establishing a two-tiered filtering strategy based on the LFQ-intensity profiles of proteins throughout proteomic experiments (Figure 4.3a). First, we grouped proteins based on their similarity in behaviour to known PLP-DEs and assigned them a confidence ranking. Therefore, experiments were conducted with two different probe concentrations (100 and 10 µM **PL4**) and DMSO as a control (Figure 4.2). As expected from analytical experiments (*Chapter 3*), a large number of proteins were significantly enriched using 100 µM **PL4**, while fewer were detected at 10 µM. We observed that known PLP-DEs were reproducibly and concentration-dependently enriched over the course of these experiments. Having reasoned that putative PLP-binding proteins should interact specifically with our probe and should display similar LFQ-intensity profiles to PLP-DEs, we extracted the 10 most similar profiles (Perseus, Pearson correlation) of each known PLP-DE and grouped them together. Among these, we further defined high-confidence PLP-binding candidates as those proteins showing both similar profile behaviour and significant enrichment at the lower, and thus more selective, 10 µM probe concentration (18% of 104 significant proteins). These proteins were the same as the most significantly enriched proteins in the 100 µM experiment (approximate cut-off: fold-change ≥ 8; *p*-value ≤ 0.015). The remainder of the proteins from our profile search were assigned to a medium-confidence interval (28% of 104 significant proteins). Proteins whose intensity profiles resembled that of PLP-DEs less stringently (11 ≤ n ≤ 20 most similar profiles) were classified as low-confidence (19% of 104 significant proteins) and dissimilar profile behaviour was deemed background (35% of 104 significant proteins).

These four confidence levels were mapped by colour onto the corresponding 100 µM volcano plot (Figure 4.3b). In total, nearly half of the significantly enriched proteins fell into the medium and high confidence categories (Figure 4.3c), which include known PLP-DEs, uncharacterized proteins and proteins currently annotated to have non-PLP-dependent functions (here termed "alternative function" proteins). Based on our data, we hypothesize that proteins of the latter two categories

are putatively PLP-binding and represent leads for uncovering novel PLP-dependent roles within the cell. In fact, literature reports indicate that two of the alternative function proteins, mevalonate kinase (EC 2.7.1.36)[25] and PPAT decarboxylase (EC 4.1.1.36),[26] are inhibited by PLP. This illustrates the potential value of our probe in identifying transient or regulatory roles of PLP in the cell, on top of identifying catalytic PLP functions. The complete list of significantly enriched proteins identified in this experiment can be found in Table 4.1.
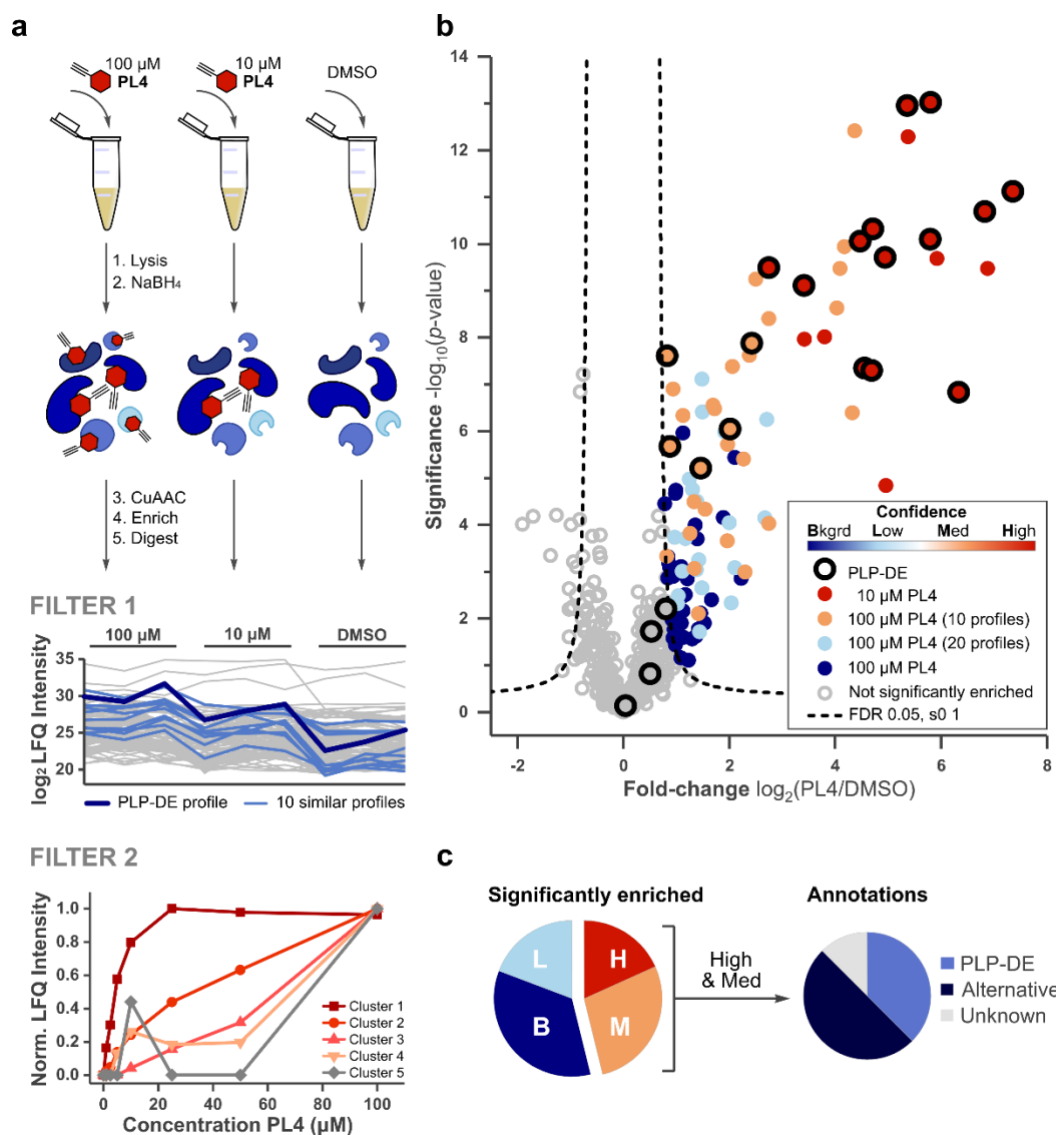
Figure 4.3 Filter 1 and confidence plot. (**a**) Proteomic workflow and filtering strategy. *S. aureus* USA300 TnPdxS cells were labeled with **PL4** (100 and 10 μM) or DMSO and cell lysates were reduced with NaBH₄. Labeled proteins were conjugated to a biotin-azide tag by CuAAC, enriched, trypsin digested and analyzed by quantitative LC-MS/MS. **Filter 1**: LFQ-intensity profiles of enriched proteins were plotted and the 10 most similar profiles (Pearson correlation) of each PLP-DE were identified. **Filter 2**: Proteomic experiments over eight concentrations revealed distinct binding behaviour of proteins, which were clustered (Pearson correlation) into 5 profile shape groups. (**b**) Volcano plot visualization of fold-enrichment using 100 μM **PL4** compared to DMSO *versus* significance upon performing a two-sample t-test (FDR 0.05, s0 1; *n* = 9 biological replicates). Confidence levels: High (10 μM **PL4**, red), Medium (100 μM **PL4** + 10 profiles, orange), Low (100 μM **PL4** + 20 profiles, light blue), Background (100 μM **PL4**, dark blue), not significant (grey). (**c**) Proportion of proteins in each confidence class and functional annotation of high- and medium-confidence proteins.

Table 4.1 PLP-DEs and select proteins of interest identified in volcano plot 4.3b. See Table 4.1 continued (p. 36) for remaining significant proteins. (Red = High confidence, Orange = Medium confidence, Grey = Non-significant).

		Fold-change	Significance	Name	Uniprot ID	Cluster
PLP-DEs	**1.**	7.36	11.12	Cystathionine gamma-synthase (EC 4.4.1.8)	A0A0H2XFF8	1
	2.	6.82	10.69	Diaminopimelate decarboxylase (EC 4.1.1.20)	A0A0H2XFD9	2
	3.	6.34	6.83	Cysteine desulfurase (EC 2.8.1.7)	A0A0H2XHJ5	2
	4.	5.81	13.02	Cysteine synthase (EC 2.5.1.47)	A0A0H2XFQ3	1
	5.	5.79	10.10	Aminotransferase, class I (EC 2.6.1.-)	A0A0H2XFA8	2
	6.	5.37	12.96	Threonine synthase (EC 4.2.3.1)	A0A0H2XH24	3
	7.	4.94	9.71	Orn/Lys/Arg decarboxylase (EC 4.1.1.18)	A0A0H2XII6	1
	8.	4.71	10.32	Aminotransferase, class V	A0A0H2XKI8	2
	9.	4.69	7.29	L-threonine dehydratase catabolic TdcB (EC 4.3.1.19)	Q2FH01	2
	10.	4.56	7.34	Uncharacterized protein	A0A0H2XHH8	1
	11.	4.48	10.06	Branched-chain-amino-acid aminotransferase (EC 2.6.1.42)	A0A0H2XIS2	2
	12.	3.41	9.11	D-alanine aminotransferase (EC 2.6.1.21)	A0A0H2XHU6	1
	13.	2.75	9.49	Alanine racemase (EC 5.1.1.1)	Q2FF55	1
	14.	2.43	7.88	Aminotransferase, class I (EC 2.6.1.-)	A0A0H2XFY9	3
	15.	2.01	6.04	Probable glycine dehydrogenase subunit 2 (EC 1.4.4.2)	Q2FGI7	3
	16.	1.46	5.21	Serine hydroxymethyltransferase (EC 2.1.2.1)	Q2FF15	3
	17.	0.89	5.68	Ornithine aminotransferase (EC 2.6.1.13)	A0A0H2XIK4	2
	18.	0.83	7.61	Putative pyridoxal phosphate-dependent acyltransferase	A0A0H2XJW8	2
	19.	0.81	2.20	Probable glycine dehydrogenase subunit 1 (EC 1.4.4.2)	Q2FGI6	
	20.	0.53	1.72	Aminotransferase	A0A0H2XJK0	
	21.	0.51	0.82	Cysteine synthase/cystathionine beta-synthase (EC 2.5.1.47)	A0A0H2XG73	
	22.	0.04	0.13	Glutamate-1-semialdehyde 2,1-aminomutase 1 (EC 5.4.3.8)	Q2FG69	
Other	**23.**	6.88	9.47	Pseudouridine-5'-phosphate glycosidase (EC 4.2.1.70)	A0A0H2XJC6	2
	24.	5.92	9.69	Mevalonate kinase (EC 2.7.1.36)	A0A0H2XHE9	3
	25.	5.38	12.29	Putative heme-dependent peroxidase (EC 1.11.1.-)	Q2FJ56	2
	26.	4.96	4.83	Ferrochelatase (EC 4.99.1.1)	A0A0H2XGS5	2
	27.	3.81	8.02	Glutamyl-aminopeptidase (EC 3.4.11.7)	A0A0H2XJ82	2
	28.	3.42	7.96	UPF0340 protein SAUSA300_2068	Q2FF14	1
	29.	4.37	12.41	Uncharacterized protein	A0A0H2XGP0	2
	30.	4.33	6.39	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36)	A0A0H2XFW7	3
	31.	1.55	4.33	D-alanine-D-alanine ligase (EC 6.3.2.4)	Q2FF43	3

Having classified enriched proteins based on their similarity to PLP-DE profile behaviour, we next focused on the properties of probe binding for individual proteins as a secondary filter. Proteomic experiments were therefore conducted over a larger range of probe concentrations, and the profile plots of all proteins significantly enriched in the above confidence plot (Filter 1, Tabel 4.1) were examined. Upon plotting the normalized LFQ-intensities over the different probe concentrations, defined binding curves could be revealed for proteins interacting specifically with the probe. The curve shapes were clustered into five groups of similar behaviour (Perseus, Pearson correlation, Figure 4.4a,b) as illustrated by representative examples in Figure 4.3a (bottom). For some protein populations, binding was saturated at low probe concentrations (cluster 1), while others never reached saturation under the conditions tested but still bound the probe in a concentration-dependent manner (clusters 2-3). Proteins that interacted with the probe non-specifically showed sporadic, unpredictable LFQ intensities over the range of probe concentrations (clusters 4-5). Most PLP-DEs could be assigned to clusters 1 or 2, which depict the most pronounced binding curve behaviour and also include the majority of high-confidence binding proteins determined by Filter 1 (Figure 4.5a-d). While these data cannot be directly correlated to protein affinity for **PL4P** because the cellular protein concentrations are unknown, we suspect that these trends reflect binding affinity (ie. clusters 1 and 2 = higher affinity).

Table 4.1 continued Additional proteins significantly enriched using **PL4** (100 µM). (Orange = Medium confidence, Light Blue = Low confidence, Dark Blue = Background, Grey = Non-significant).

	Fold-change	Significance	Name	Uniprot ID	Cluster
33.	4.18	9.93	Dihydroorotate dehydrogenase (quinone) (EC 1.3.5.2)	A0A0H2XFW6	3
34.	4.10	9.47	Signal peptidase I (EC 3.4.21.89)	A0A0H2XEA7	2
35.	4.03	8.63	Glutamyl-aminopeptidase	A0A0H2XFP6	3
36.	2.75	4.03	Methicillin resistance protein FemB	A0A0H2XHG6	2
37.	2.75	8.40	GTPase Obg (EC 3.6.5.-)	Q2FG83	2
38.	2.51	9.25	Phosphoglucosamine mutase (EC 5.4.2.10)	Q2FEX1	2
39.	2.38	7.61	Oxidoreductase, aldo/keto reductase family	A0A0H2XG74	2
40.	2.30	2.99	Uncharacterized protein	A0A0H2XJQ7	5
41.	2.27	5.40	3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (EPSPS)	Q2FGX6	2
42.	2.06	7.38	Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)	A0A0H2XHS2	3
43.	1.98	5.72	Elongation factor P (EF-P)	Q2FGJ3	3
44.	1.97	3.65	Putative NAD(P)H nitroreductase	Q2FDY2	2
45.	1.73	6.48	Uncharacterized protein	A0A0H2XGS0	3
46.	1.70	6.55	Uncharacterized protein	A0A0H2XJ48	4
47.	1.42	2.10	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4)	A0A0H2XFP3	1
48.	1.35	3.06	Lipoate--protein ligase (EC 2.7.7.63)	A0A0H2XFJ1	3
49.	1.34	4.49	Bifunctional protein FolD (EC 1.5.1.5); (EC 3.5.4.9)	Q2FI15	2
50.	1.26	3.82	Uncharacterized protein	A0A0H2XEY2	2
51.	1.13	6.34	Uracil phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase)	Q2FF16	4
52.	0.94	6.90	Transcription termination/antitermination protein NusA	A0A0H2XEX6	1
53.	0.82	3.33	Single-stranded DNA-binding protein (SSB)	A0A0H2XJY5	3
54.	2.72	6.25	GTP pyrophosphokinase (EC 2.7.6.5)	A0A0H2XGF7	4
55.	2.68	4.14	2-dehydropantoate 2-reductase (EC 1.1.1.169)	A0A0H2XJU5	3
56.	2.11	3.08	Uncharacterized protein	A0A0H2XI47	4
57.	2.04	2.33	Uroporphyrinogen decarboxylase (UPD) (EC 4.1.1.37)	Q2FFR2	4
58.	2.01	4.05	Protoporphyrinogen oxidase (EC 1.3.3.4)	A0A0H2XFK8	3
59.	1.92	5.98	Glutamate dehydrogenase	A0A0H2XHC2	2
60.	1.50	6.41	Transcription termination/antitermination protein NusG	A0A0H2XEI4	3
61.	1.49	7.10	Protein GrpE (HSP-70 cofactor)	Q2FGE2	2
62.	1.49	2.65	Chorismate synthase (CS) (EC 4.2.3.5)	Q2FGX4	2
63.	1.45	1.71	Glutathione peroxidase	A0A0H2XEC3	3
64.	1.43	3.25	50S ribosomal protein L5	Q2FEQ1	5
65.	1.39	4.50	Transcriptional regulator SarA (Staphylococcal accessory regulator A)	Q2FJ20	4
66.	1.39	3.04	Uncharacterized protein	A0A0H2XI18	4
67.	1.31	4.75	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMPDCase)	Q2FHN4	2
68.	1.25	4.97	50S ribosomal protein L27	Q2FG82	5
69.	1.17	3.71	Putative universal stress protein	Q2FG28	2
70.	1.12	3.01	Putative dipeptidase (EC 3.4.13.-)	Q2FFY7	3
71.	1.04	2.48	ATP-dependent protease ATPase subunit HslU (Unfoldase HslU)	Q2FHI4	5
72.	1.02	2.32	6,7-dimethyl-8-ribityllumazine synthase (Lumazine synthase) (EC 2.5.1.78)	Q2FFX3	2
73.	0.97	3.73	Chorismate mutase/phospho-2-dehydro-3-deoxyheptonate aldolase (EC 5.4.99.5)	A0A0H2XI85	2
74.	2.23	2.85	Inositol monophosphatase family protein (EC 3.1.3.25)	A0A0H2XHI3	3
75.	2.11	5.44	Serine acetyltransferase (EC 2.3.1.30)	A0A0H2XK24	3
76.	1.89	4.15	Phosphate acetyltransferase (EC 2.3.1.8)	A0A0H2XHL4	2
77.	2.23	2.85	Inositol monophosphatase family protein (EC 3.1.3.25)	A0A0H2XHI3	3
78.	2.11	5.44	Serine acetyltransferase (EC 2.3.1.30)	A0A0H2XK24	3
79.	1.89	4.15	Phosphate acetyltransferase (EC 2.3.1.8)	A0A0H2XHL4	2
80.	1.66	2.40	UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.3.1.98) (UDP-N-acetylmuramate dehydrogenase)	Q2FIQ3	4
81.	1.53	1.90	Ribulokinase (EC 2.7.1.16)	Q2FJ88	5
82.	1.47	2.11	30S ribosomal protein S20	Q2FGD8	5
83.	1.42	1.78	Pyridine nucleotide-disulfide oxidoreductase (EC 1.8.1.9)	A0A0H2XJR2	5
84.	1.39	3.69	50S ribosomal protein L25 (General stress protein CTC)	Q2FJE0	4
85.	1.37	1.62	Orotate phosphoribosyltransferase (OPRTase) (EC 2.4.2.10)	A0A0H2XH75	5
86.	1.36	3.99	SUF system FeS assembly protein, NifU family	A0A0H2XJC0	2
87.	1.30	1.56	Uncharacterized protein	A0A0H2XJG9	5
88.	1.24	1.11	UPF0457 protein SAUSA300_2132	Q2FEV9	3
89.	1.21	2.84	Tryptophan--tRNA ligase (EC 6.1.1.2) (Tryptophanyl-tRNA synthetase)	A0A0H2XGX6	3
90.	1.17	2.51	Uncharacterized N-acetyltransferase SAUSA300_1070 (EC 2.3.1.-)	A0A0H2XHJ3	3
91.	1.14	1.60	S1 RNA binding domain protein	A0A0H2XFU0	5
92.	1.13	5.96	Protein RecA (Recombinase A)	A0A0H2XFW9	4
93.	1.13	2.16	Uncharacterized protein	A0A0H2XH52	3
94.	1.12	2.47	50S ribosomal protein L6	Q2FEQ4	3
95.	1.10	1.90	Enoyl-[acyl-carrier-protein] reductase [NADPH] (ENR) (EC 1.3.1.39)	A0A0H2XIJ1	5
96.	1.09	3.11	Uncharacterized protein	A0A0H2XFW2	3
97.	1.09	1.15	50S ribosomal protein L9	Q2FKP1	5
98.	1.05	2.44	Staphylococcal accessory regulator	A0A0H2XHA2	5
99.	1.03	2.16	Bifunctional protein PyrR [Includes: Pyrimidine operon regulatory protein;Uracil phosphoribosyltransferase (UPRTase) (EC 2.4.2.9)]	Q2FHP0	1
100.	1.00	4.74	UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	Q2FE05	3
101.	0.99	4.67	50S ribosomal protein L17	Q2FER6	3
102.	0.98	1.43	Uncharacterized protein	A0A0H2XH92	5
103.	0.97	1.76	50S ribosomal protein L24	Q2FEQ0	5
104.	0.94	1.58	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8)	A0A0H2XEG5	5
105.	0.94	2.89	Ribosome-binding factor A	Q2FHG8	4
106.	0.92	1.97	Uncharacterized protein	A0A0H2XH40	2
107.	0.92	2.27	Alkaline shock protein 23	Q2FEV0	4
108.	0.88	3.10	Phosphoglycerate kinase (EC 2.7.2.3)	Q2FIM0	4
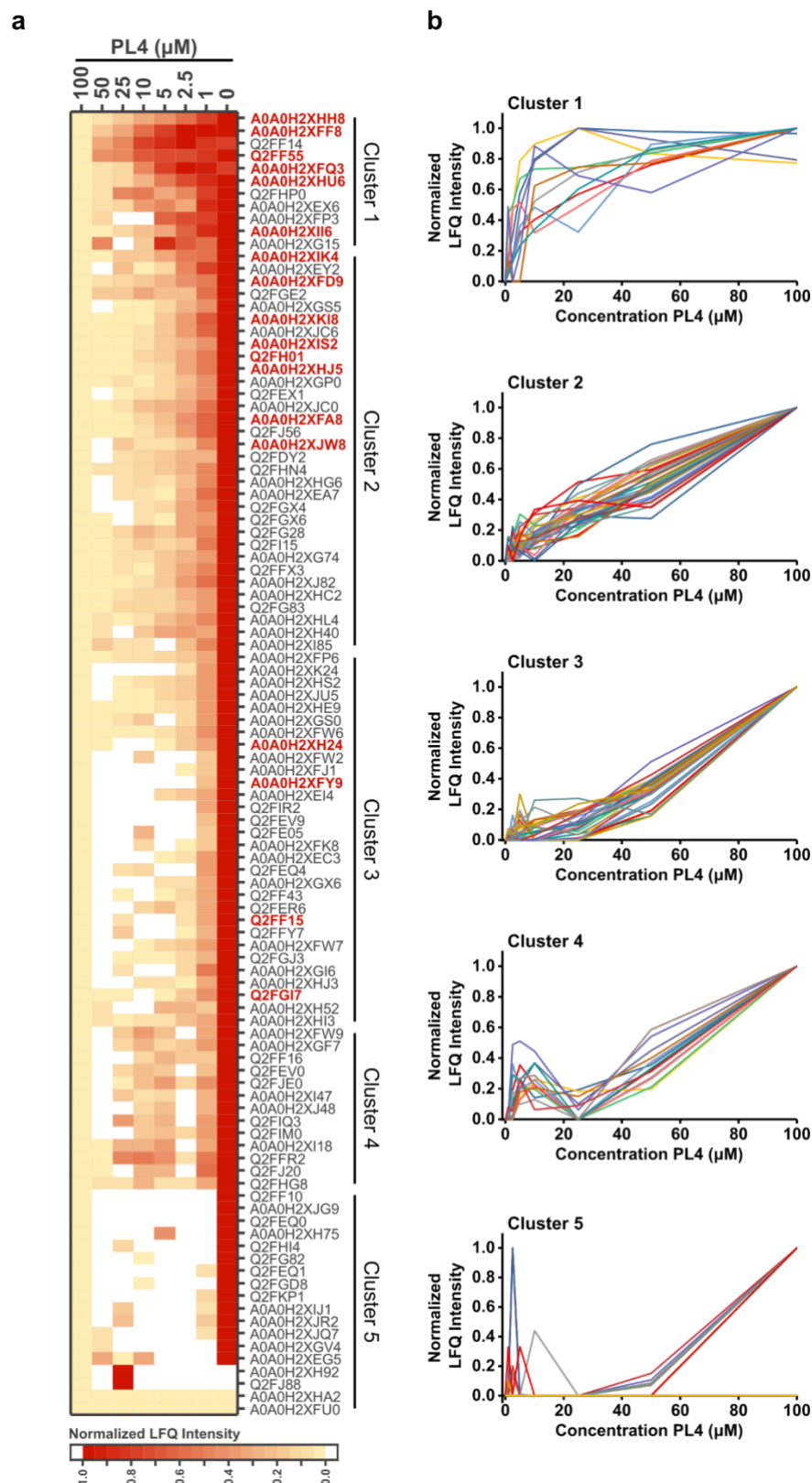
Figure 4.4 Filter 2: Concentration-dependent chemical proteomics of PL4 in *S. aureus* USA300 TnPdxS. (**a**) Heatmap representation of the mean LFQ intensities of significant proteins (from Figure 4.3b) over several **PL4** concentrations (100, 50, 25, 10, 5, 2.5, 1, and 0 µM) performed as a separate experiment (*n* = 3 biological replicates). Data were normalized to DMSO (= 0, light red) and absolute max binding (= 1, red) within individual protein datasets; proteins negatively enriched compared to DMSO (< 0) are shown in white. Proteins are labeled by their Uniprot ID and PLP-DEs are bolded in red. (**b**) Clustering of significantly enriched proteins based on LFQ-intensity profiles (hierarchical clustering, Pearson correlation, 5 groups).
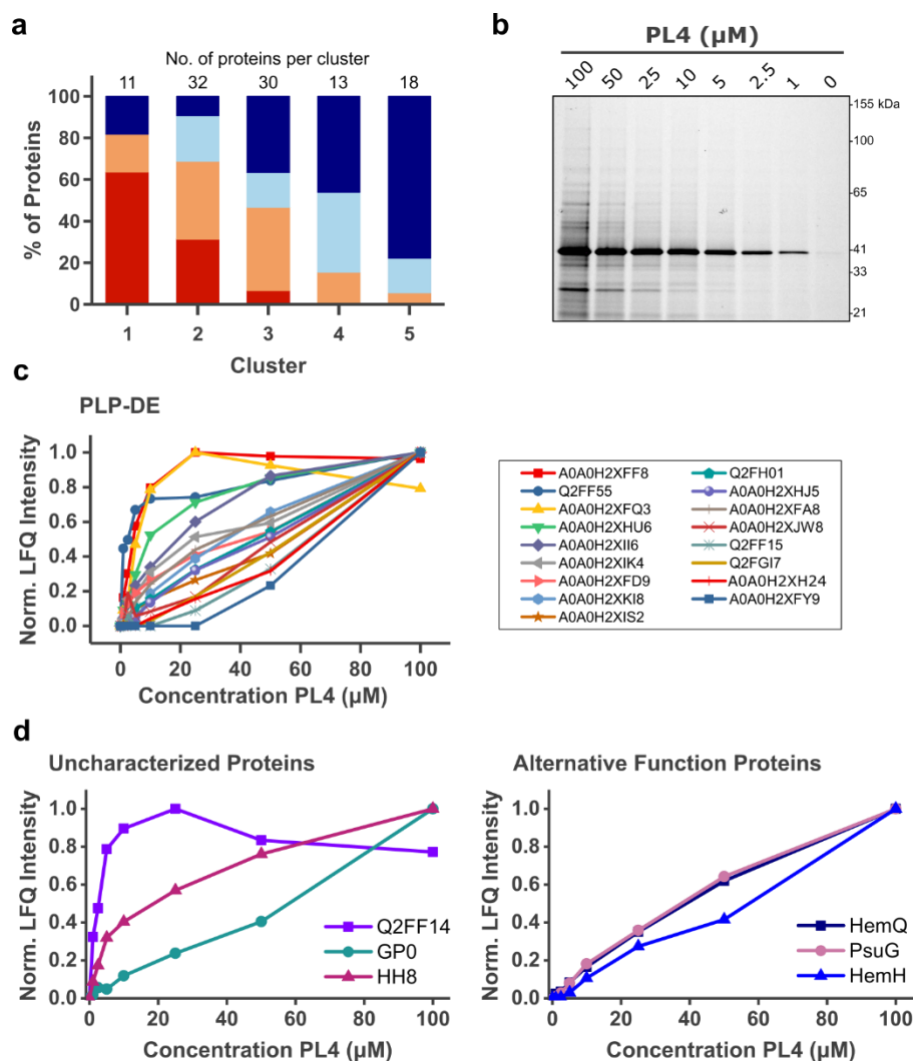
Figure 4.5 Filter 2 Continued: Concentration-dependent chemical proteomics of PL4 in *S. aureus* USA300 TnPdxS. (**a**) Percent of proteins from each confidence class (Filter 1) per cluster (Filter 2). The number of proteins per cluster are written above each respective bar. (**b**) Fluorescence SDS-PAGE showing concentration-dependent **PL4** labeling during proteomic experiments upon enrichment. (**c**) LFQ intensity profiles of all PLP-DEs identified. (**d**) LFQ intensity profiles of uncharacterized proteins (left) and alternative enzymes (right) selected for validation.

The application of our two-fold filtering strategy helped to narrow down the selection of uncharacterized (Q2FF14, A0A0H2XGP0, A0A0H2XHH8) as well as alternative function proteins (HemQ: Q2FJ56, PsuG: A0A0H2XJC6, HemH: A0A0H2XGS5) as candidates for further characterization (*see Chapter 5*). First, LFQ-intensity profile searches were used to correlate the data of known PLP-DEs to unknown proteins and thereby define PLP-DE confidence levels. Secondly, assessing the defined binding curve shapes of proteins over several probe concentrations highlighted PLP-binding over random probe binding. Proteins selected for further investigation were all identified with high or medium confidence (Table 4.1) and showed strong concentration-dependent enrichment (Figure 4.5d). Our study highlights the effectiveness with which concentration-dependent profile analysis can add certainty to proteomic investigations and can help to distinguish weak binding from background in complex experiments.

4.4   MS-based analysis of PL1 vs. PL2 proteomic labeling

We next investigated chemical proteomics using **PL1** and **PL2** in *S. aureus* TnPdxS. Only a few protein hits could be identified using **PL1** (Figure 4.6a,b), which included 5 PLP-DEs as well as the uncharacterized protein Q2FF14 and the putative heme-dependent peroxidase (Q2FJ56), previously identified with **PL4** (high-confidence). Generally, there was little overlap between the proteins identified using the inactive control probe **PL2** (Figure 4.6a) in comparison to **PL1** and **PL4**, supporting the labeling specificity of the probes. Nonetheless, cysteine synthase (A0A0H2XFQ3) and the putative PLP-dependent acyltransferase (A0A0H2XJW8) were significantly enriched with **PL2**, suggesting that these proteins bind non-specifically or are able to tolerate the structural and chemical alterations of **PL2**. While **PL1** represents a specific chemical probe for identifying PLP-DEs in cells, its chemical lability appears to prevent it from labeling as efficiently as **PL4** under comparable experimental conditions.
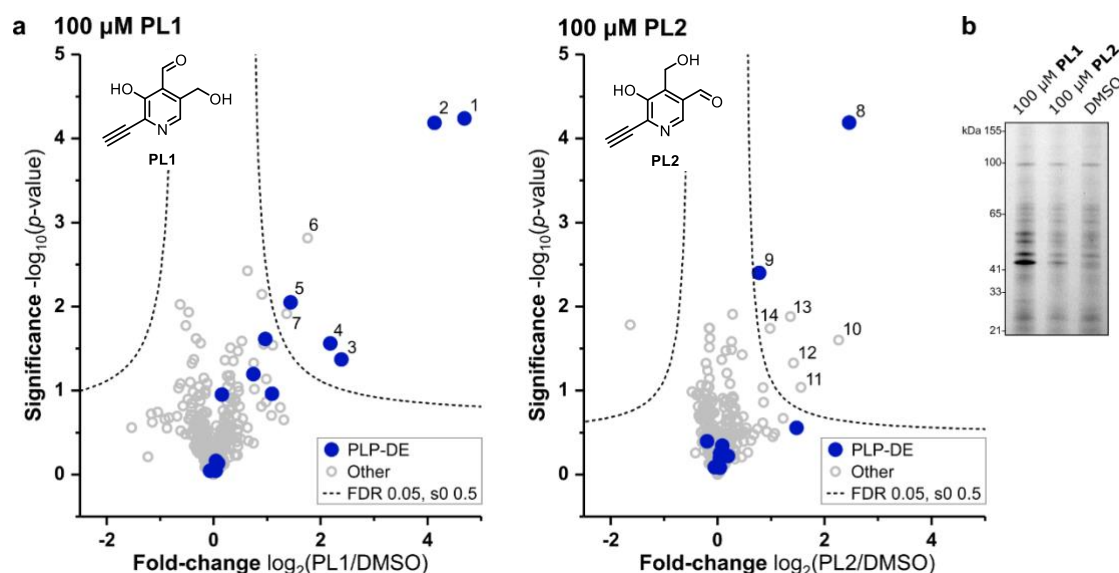
Figure 4.6 Chemical proteomics with PL1 and PL2 in *S. aureus* USA300 TnPdxS. (**a**) Volcano plots showing enrichment of proteins using 100 µM **PL1** (left) or **PL2** (right) (*n* = 3 biological replicates). Known PLP-DEs are highlighted with blue dots and are numerically annotated according to Table 4.2. (**b**) Fluorescence SDS-PAGE showing **PL1** and **PL2** labeling during proteomic experiments upon avidin bead enrichment.

Table 4.2 Proteins significantly enriched using PL1 and PL2 in *S. aureus* USA300 TnPdxS. Corresponding volcano plots in Figure 4.6a.

		Fold-change	Significance	Name	Uniprot ID
PL1	**1.**	4.69	4.24	Putative pyridoxal phosphate-dependent acyltransferase	A0A0H2XJW8
	2.	4.13	4.19	Cysteine synthase (EC 2.5.1.47)	A0A0H2XFQ3
	3.	2.39	1.37	Cysteine desulfurase (EC 2.8.1.7)	A0A0H2XHJ5
	4.	2.18	1.56	Uncharacterized protein	A0A0H2XHH8
	5.	1.44	2.05	Aminotransferase	A0A0H2XJK0
	6.	1.76	2.81	UPF0340 protein SAUSA300_2068	Q2FF14
	7.	1.37	1.91	Putative heme-dependent peroxidase (EC 1.11.1.-)	Q2FJ56
PL2	**8.**	2.46	4.19	Putative pyridoxal phosphate-dependent acyltransferase	A0A0H2XJW8
	9.	0.78	2.40	Cysteine synthase (EC 2.5.1.47)	A0A0H2XFQ3
	10.	2.27	1.60	Protein-tyrosine-phosphatase PtpA (EC 3.1.3.48)	Q2FFL4
	11.	1.56	1.04	Putative thioredoxin	A0A0H2XHS5
	12.	1.42	1.33	Lipoate--protein ligase (EC 6.3.1.20)	A0A0H2XFJ1
	13.	1.36	1.88	ATP-dependent protease ATPase subunit HslU (Unfoldase HslU)	Q2FHI4
	14.	0.98	1.74	Transferrin receptor	A0A0H2XKM7

In order to improve **PL1** labeling, we attempted to increase its metabolic incorporation into biological processes by growing bacteria directly in CDM supplemented with **PL1**. Previous experiments showed that **PL1** acts as a global PL-surrogate and is capable of supporting the growth of *S. aureus* USA300 TnPdxS in CDM in the absence of PL (*see* Figure 3.5). Therefore, the probe appears to function in a vast amount of cellular PLP-dependent processes. For proteomic experiments, bacterial cultures were inoculated into CDM containing 25 µM **PL1**, which was the concentration required to support maximum bacterial growth with respect to PL-substituted media. Cultures grown in CDM containing 0.5 µM PL served as a control in proteomic experiments. Due to the minimal growth conditions, long incubation times (30 h) and aeration (baffled flasks) were required for the culture to reach a maximum OD_{600} (~ 1.3-1.5). At this point, bacteria were harvested, lysed and processed as described previously, including an additional IAA alkylation step prior to click chemistry to improve **PL1**-conjugation. Upon enrichment and LC-MS/MS analysis, the fold change in proteins from bacteria grown in the presence of **PL1** *versus* PL were plotted (Figure 4.7a,b). We found significant enrichment of PLP-DEs – 36% of the annotated *S. aureus* PLPome – with an enrichment factor of 13.8, $p = 1.9e^{-17}$. Considerably less background was observed using this method by comparison to our 2 h labeling protocol with **PL4**. Furthermore, we were able to access 7 additional PLP-DEs not previously identified using the **PL4** probe, while the remaining 9 PLP-DEs matched prior experiments (*see comparison in* Figure 4.11).
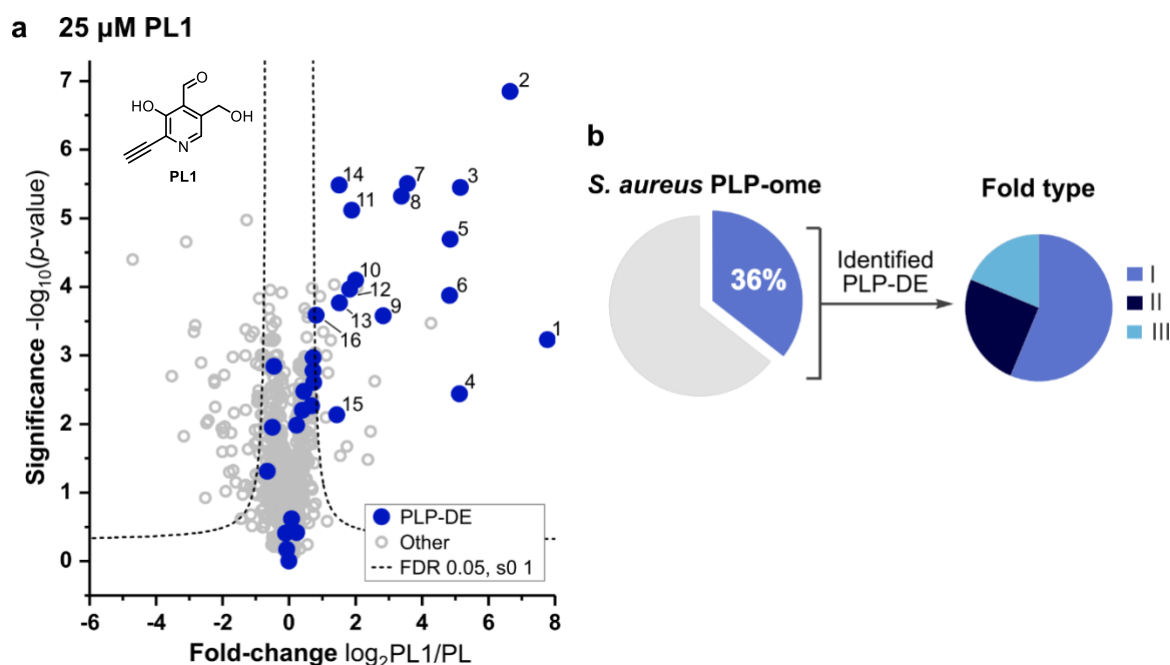
Figure 4.7 Chemical proteomics with *S. aureus* USA300 TnPdxS grown in CDM containing 25 µM PL1. (**a**) Volcano plot showing enrichment of proteins from cultures grown in CDM containing **PL1** (25 µM) versus PL (0.5 µM) (*n* = 3 biological replicates). Known PLP-DEs are highlighted with blue dots and are numerically annotated according to Table 4.3. (**b**) Proportion of the *S. aureus* PLPome (36%) significantly identified in proteomic experiments and distribution of PLP-DE fold-types.

Table 4.3 Proteins significantly enriched from *S. aureus* USA300 TnPdxS grown in CDM containing 25 µM PL1. PLP-DEs (blue) and other significant proteins (grey); corresponding volcano plot in Figure 4.7a.

	Fold-change	Significance	Name	Uniprot ID
1.	7.78	3.23	Aminotransferase, class V	A0A0H2XKI8
2.	6.65	6.85	Putative pyridoxal phosphate-dependent acyltransferase	A0A0H2XJW8
3.	5.15	5.45	Alanine racemase (EC 5.1.1.1)	Q2FF55
4.	5.13	2.44	Tryptophan synthase beta chain (EC 4.2.1.20)	Q2FH64
5.	4.85	4.69	L-threonine dehydratase biosynthetic IlvA (EC 4.3.1.19) (Threonine deaminase)	Q2FF63
6.	4.83	3.87	Uncharacterized protein	A0A0H2XHH8
7.	3.57	5.50	Cysteine desulfurase (EC 2.8.1.7)	A0A0H2XHJ5
8.	3.38	5.32	Cysteine synthase (EC 2.5.1.47)	A0A0H2XFQ3
9.	2.83	3.58	Aminotransferase, class V	A0A0H2XGZ7
10.	2.01	4.09	Trans-sulfuration enzyme family protein	A0A0H2XG37
11.	1.88	5.12	Serine hydroxymethyltransferase (SHMT) (Serine methylase) (EC 2.1.2.1)	Q2FF15
12.	1.82	3.96	Alanine racemase (EC 5.1.1.1)	A0A0H2XI95
13.	1.52	3.77	Cys/Met metabolism PLP-dependent enzyme (EC 4.4.1.8)	A0A0H2XFH9
14.	1.51	5.48	Threonine synthase (EC 4.2.3.1)	A0A0H2XH24
15.	1.44	2.13	Transcriptional regulator, gntR family protein	A0A0H2XF80
16.	0.82	3.59	Aminotransferase, class I (EC 2.6.1.-)	A0A0H2XFY9
17.	4.27	3.47	UPF0340 protein SAUSA300_2068	Q2FF14
18.	2.58	2.62	Bacterial luciferase family protein	A0A0H2XF54
19.	2.46	1.89	NADPH-dependent oxidoreductase (EC 1.6.-.-)	Q2FJN3
20.	2.37	1.48	UPF0348 protein SAUSA300_1025	Q2FHV5
21.	2.05	3.99	Transcriptional repressor, ArsR family	A0A0H2XG84
22.	1.75	1.67	Glyoxalase family protein (EC 1.13.11.39)	A0A0H2XI94
23.	1.54	1.54	Peptide methionine sulfoxide reductase MsrA (EC 1.8.4.11)	A0A0H2XEQ6
24.	1.36	4.03	Bifunctional purine biosynthesis protein PurH [Includes: Phosphoribosyl-aminoimidazolecarboxamide formyltransferase (EC 2.1.2.3); IMP cyclohydrolase (EC 3.5.4.10)	Q2FI05
25.	1.26	3.22	Organic hydroperoxide resistance protein-like	Q2FIJ2
26.	1.18	2.74	Lantibiotic epidermin immunity protein F	A0A0H2XGC6
27.	1.12	3.00	Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)	Q2FI12
28.	1.11	2.09	Uncharacterized oxidoreductase SAUSA300_2422 (EC 1.-.-.-)	Q2FE21
29.	1.03	3.34	Putative heme-dependent peroxidase SAUSA300_0569 (EC 1.11.1.-) (UPF0447 protein SAUSA300_0569)	Q2FJ56
30.	0.93	3.86	UPF0173 metal-dependent hydrolase SAUSA300_1653	Q2FG31

In summary, using conditions in which bacteria are allowed to grow directly in the presence of **PL1** significantly improved labeling, probably by increasing the metabolic incorporation of **PL1** as a cofactor surrogate. This enabled the identification of additional PLP-DEs that could not be identified in **PL4** experiments. Interestingly, we also significantly identify the uncharacterized protein Q2FF14 and the alternative function protein Q2FJ56, which were both highly enriched and presumed to have PLP-dependent roles based on proteomic experiments with the **PL4** probe. Therefore, our findings with **PL1** support the hypothesis that these proteins may interact with PLP.

4.5 PL3 labeling using the Staudinger ligation

Bioorthogonal chemical ligations are useful for studying biological processes and engineering novel biochemical interactions. Prerequisites for bioorthogonal reactions include the ability to selectively form covalent adducts within a complex biological matrix, as well as fast reaction rates and high yields under biological conditions. Azides represent good biorthogonal functionalities which can easily be introduced into diverse biological molecules by metabolic incorporation of azide-bearing substrates. Bioorthogonal reactions applicable for conjugating azide-bearing activity-based probes include both copper-catalyzed[39] and copper-free[40] Huisgen [2+3] cycloadditions,[79,81,97] as well as the Staudinger ligation.[41] Comparison of CuAAC directionality has

shown that the ABP-≡/Rh-N$_3$ reaction proceeds more selectively, but slowly, than ABP-N$_3$/Rh-≡ systems.[80] In the latter system, significant background labeling is attributed to non-specific reactivity of Rh-≡ in cells, which is in agreement with our previous gel-based analysis of **PL3** labeling in cells (*see* Figure 3.6a). Copper-free click reactions were later developed in order to circumvent the limited *in vivo* applicability of CuAAC due to copper (I) cytotoxicity.[98] Here, the Bertozzi lab capitalized on ring strain as a thermodynamic driver of click conjugation by using cyclooctyne moieties (SPAAC – strain-promoted azide alkyne cycloaddition).[40] The reactivity of these systems was later improved upon through the attachment of additional aromatic rings to increase ring strain (DBCO, Figure 4.8)[99] and electron withdrawing moieties (DIFO, difluorocyclooctyne; BARAC, biarylazacyclooctyne).[100,101]

The chemical Staudinger reaction classically reduces azides to amines under mild conditions using phosphine reagents.[102] The reaction proceeds via formation of an unstable aza-ylide, which hydrolyzes to the phosphine oxide and amine in the presence of water. Based on this reaction, the Staudinger *ligation* was later developed in order to couple an azide with a triarylphosphine derivative via concomitant formation of an amide bond.[41,103] In this reaction, a phosphine reagent modified with an appropriately placed electrophilic trap (ie. methyl ester) captures the unstable aza-ylide intermediate by intramolecular cyclization and produces a stable amide bond upon hydrolysis (Figure 4.8).[41] Both reaction partners are abiotic and chemically orthogonal to native cellular components. Its application as an effective biorthogonal ligation was subsequently demonstrated using a water-soluble, tetraethyleneglycol-modified phosphine reagent containing a biotin tag (Figure 4.8). This reagent enabled fluorescent labeling of human cells, upon metabolic incorporation of an azido-sugar on the cell surface.[41]

DBCO-PEG$_4$-Rhodamine   **Phosphine-PEG$_3$-Biotin**
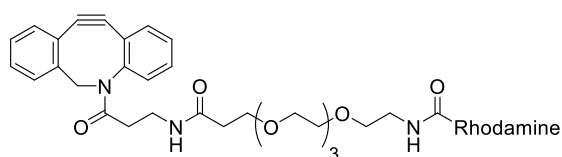
Staudinger Ligation
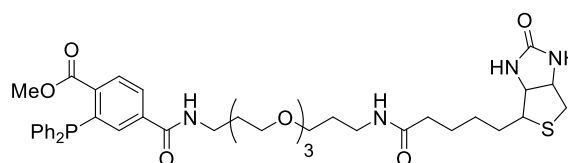
Phosphine with   Aza-ylide   Intramolecular   Amide bond
electrophilic trap   cyclization
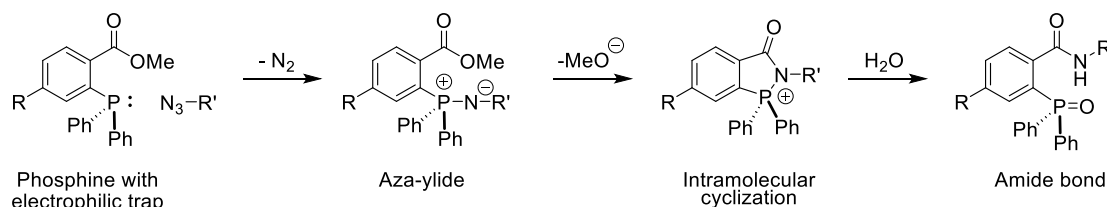
Figure 4.8 Chemical ligation tools for ABPP with azide-based probes. DBCO (Dibenzylcyclooctyne) conjugated to a rhodamine tag was used in initial gel-based **PL3** labeling experiments though showed high background labeling (*Chapter 3*). The phosphine-biotin reagent was envisaged as an alternative for conjugation of the **PL3** probe to an enrichment tag by Staudinger ligation. The mechanism of the Staudinger ligation[41] is shown, which involves amide bond formation between an azide and phosphine reagent via unstable aza-ylide.

Our earlier biochemical investigations proved **PL3** to have suitable probe qualities, though significant background due to nonspecific reactivity of the alkyne- and DBCO-rhodamine reagents observed by SDS-PAGE (Figure 3.6) precluded its application for our method. Similar problems in the literature have been addressed through the use of the Staudinger ligation, which showed improved specificity.[48] We therefore pursued **PL3**-based chemical proteomics using the commercially available Staudinger biotin-phosphine reagent (Figure 4.8). **PL3**-labeled samples (100 and 10 μM, 2 h, 37°C) were reacted with this reagent (24 h incubation) and enriched using avidin beads. MS-analysis revealed enrichment of 10 PLP-DEs with nearly no background labeling at 100 μM **PL3**, and accordingly fewer proteins at the lower probe concentration of 10 μM (Figure 4.9). Significant proteins corresponded well to previous **PL4**-based experiments, and also included several of the non-PLP-DEs selected for further characterization.
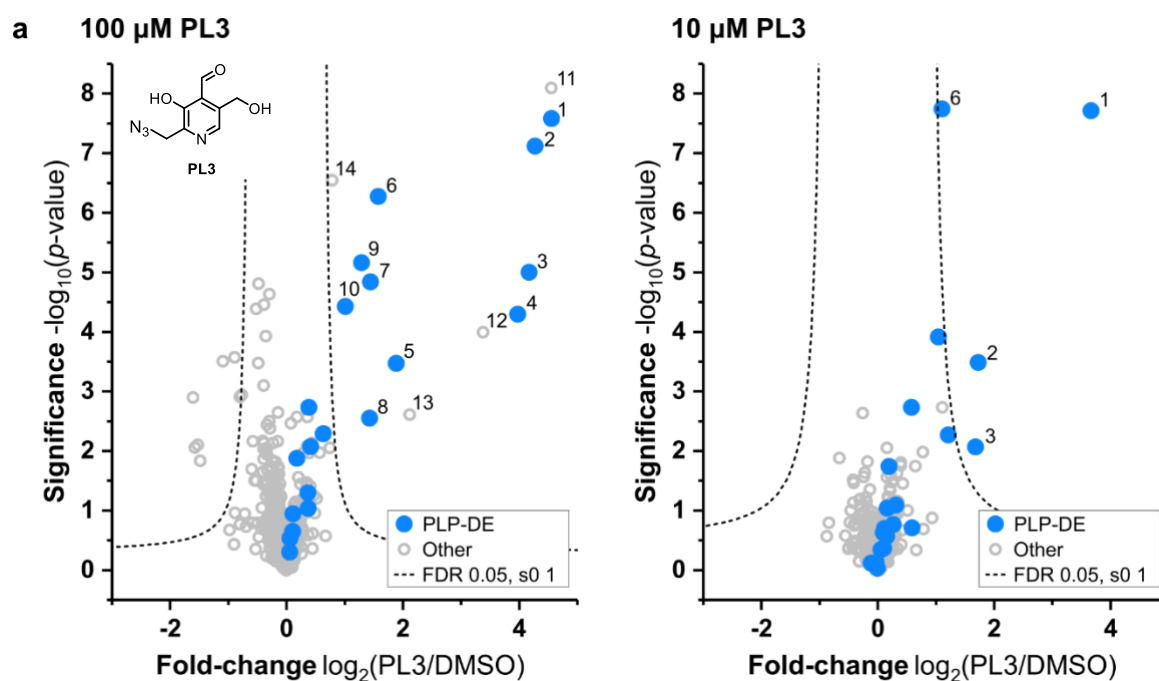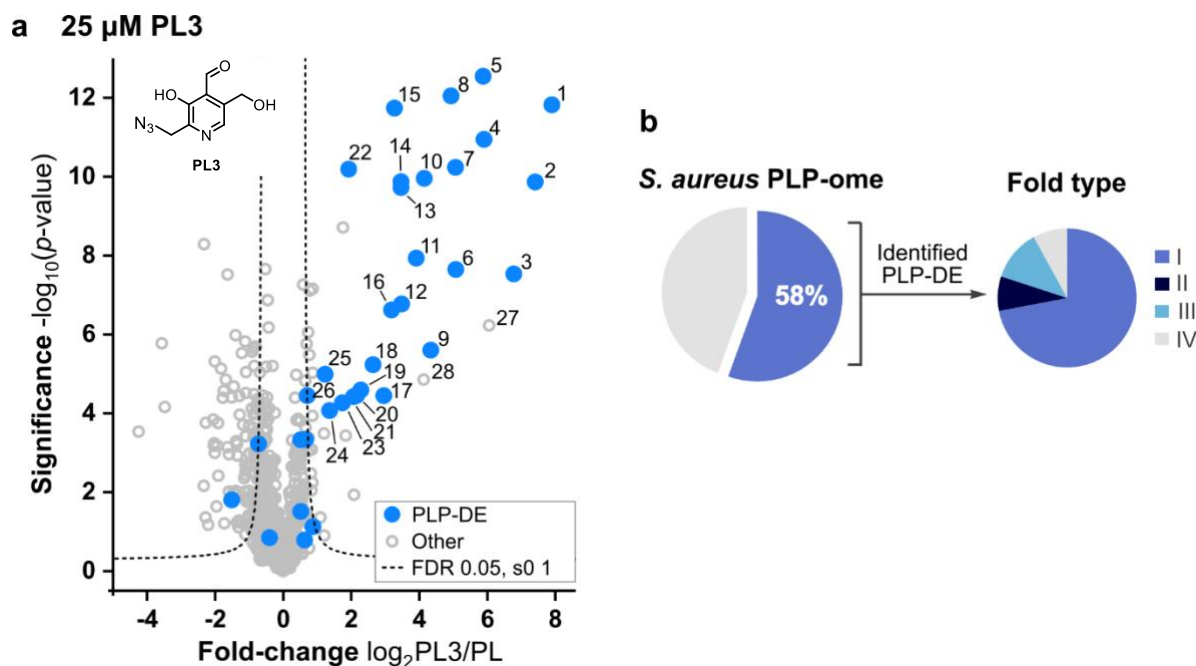

Figure 4.9 Chemical proteomics with PL3 in *S. aureus* USA300 TnPdxS using the Staudinger-Bertozzi ligation. (**a**) Volcano plots showing enrichment of proteins with 100 μM (left) or 10 μM (right) **PL3**. (*n* = 6 biological replicates). Known PLP-DEs are highlighted with blue dots and are numerically annotated according to Table 4.4.

Table 4.4 Proteins significantly enriched by chemical proteomics using PL3 (100, 10 μM).

No.	Fold-change	Significance	Name	Uniprot ID
1.	4.56	7.58	Cysteine synthase (EC 2.5.1.47)	A0A0H2XFQ3
2.	4.28	7.12	Cysteine desulfurase (EC 2.8.1.7)	A0A0H2XHJ5
3.	4.17	5.00	Aminotransferase, class V	A0A0H2XKI8
4.	3.98	4.29	Orn/Lys/Arg decarboxylase (EC 4.1.1.18)	A0A0H2XII6
5.	1.89	3.47	Threonine synthase (EC 4.2.3.1)	A0A0H2XH24
6.	1.58	6.27	Putative pyridoxal phosphate-dependent acyltransferase	A0A0H2XJW8
7.	1.44	4.84	Branched-chain-amino-acid aminotransferase (EC 2.6.1.42)	A0A0H2XIS2
8.	1.43	2.55	Cysteine synthase/cystathionine beta-synthase (EC 2.5.1.47)	A0A0H2XG73
9.	1.29	5.16	Alanine racemase (EC 5.1.1.1)	Q2FF55
10.	1.01	4.43	L-threonine dehydratase catabolic TdcB (EC 4.3.1.19) (Threonine deaminase)	Q2FH01
11.	4.55	8.09	Pseudouridine-5'-phosphate glycosidase (PsiMP glycosidase) (EC 4.2.1.70)	A0A0H2XJC6
12.	3.38	3.99	Signal peptidase I (EC 3.4.21.89)	A0A0H2XEA7
13.	2.12	2.61	UPF0340 protein SAUSA300_2068	Q2FF14
14.	0.79	6.54	Putative heme-dependent peroxidase SAUSA300_0569 (EC 1.11.1.-)	Q2FJ56

Similar to **PL1**, the azide-bearing **PL3** probe was previously able to support the growth of *S. aureus* USA300 TnPdxS in CDM, in the absence of PL (*see* Figure 3.5). We therefore applied the same strategy to increase **PL3**-based metabolic incorporation and labeling by growing bacteria directly in the presence of **PL3**. Using this method, we were able to identify 26 PLP-DEs significantly with an enrichment factor of 12.3, p = 1.3e^{-28} (Figure 4.10). This accounts for 58% of the *S. aureus* PLPome and represents the highest level of labeling achieved so far using our PL-probes. Our results showed minimal background labeling, arguing for the specificity of the probe and method. We identify much of the same PLP-DE population previously found using **PL1** and **PL4** probes, in addition to 8 further PLP-DEs (Table 4.5). Interestingly, neither the uncharacterized proteins Q2FF14 and A0A0H2XGP0 nor alternative proteins (Q2FJ56, A0A0H2XJC6), commonly identified using **PL1** and **PL4**, were identified using **PL3**. However, two new uncharacterized proteins – A0A0H2XIC2 and A0A0H2XIU9 – were identified among the significantly enriched proteins, which will be investigated for PLP-binding in future studies.

Figure 4.10 Chemical proteomics of *S. aureus* USA300 TnPdxS grown with PL3 supplementation. (**a**) Volcano plot showing enrichment of proteins from *S. aureus* USA300 TnPdxS cultures grown in CDM supplemented with with 25 µM **PL3** versus 0.5 µM PL (*n* = 6 biological replicates). Known PLP-DEs are highlighted with blue dots and are numerically annotated according to Table 4.5. (**b**) Proportion of the *S. aureus* PLPome (58%) significantly identified in proteomic experiments and distribution of PLP-DE fold-types.

Table 4.5 Proteins significantly enriched from bacterial cultures grown using PL3 supplementation (25 µM).

No.	Fold-change	Significance	Name	Uniprot ID
1.	7.90	11.82	Glutamate-1-semialdehyde 2,1-aminomutase 2 (GSA 2) (EC 5.4.3.8)	Q2FFN1
2.	7.41	9.86	Aminotransferase	A0A0H2XKB4
3.	6.78	7.53	Aminotransferase, class V	A0A0H2XKI8
4.	5.91	10.94	D-alanine aminotransferase (EC 2.6.1.21)	A0A0H2XHU6
5.	5.88	12.54	Putative pyridoxal phosphate-dependent acyltransferase	A0A0H2XJW8
6.	5.08	7.64	Trans-sulfuration enzyme family protein	A0A0H2XG37
7.	5.07	10.23	Cysteine desulfurase (EC 2.8.1.7)	A0A0H2XHJ5
8.	4.94	12.04	Ornithine aminotransferase (OAT) (EC 2.6.1.13)	A0A0H2XIK4
9.	4.34	5.60	Cystathionine gamma-synthase (EC 4.4.1.8)	A0A0H2XFF8
10.	4.15	9.95	Probable glycine dehydrogenase (decarboxylating) subunit 2 (EC 1.4.4.2) (Glycine dehydrogenase (aminomethyl-transferring) subunit 2)	Q2FGI7
11.	3.91	7.93	Serine hydroxymethyltransferase (SHMT) (Serine methylase) (EC 2.1.2.1)	Q2FF15
12.	3.49	6.77	Cysteine synthase (EC 2.5.1.47)	A0A0H2XFQ3
13.	3.47	9.72	Cys/Met metabolism PLP-dependent enzyme (EC 4.4.1.8)	A0A0H2XFH9
14.	3.46	9.87	Branched-chain-amino-acid aminotransferase (EC 2.6.1.42)	A0A0H2XIS2
15.	3.27	11.73	Glutamate-1-semialdehyde 2,1-aminomutase 1 (GSA 1) (EC 5.4.3.8)	Q2FG69
16.	3.19	6.61	Cysteine synthase/cystathionine beta-synthase (EC 2.5.1.47)	A0A0H2XG73
17.	2.96	4.45	Ornithine aminotransferase (OAT) (EC 2.6.1.13)	A0A0H2XHG0
18.	2.65	5.23	Aminotransferase, class V	A0A0H2XGZ7
19.	2.28	4.59	Orn/Lys/Arg decarboxylase (EC 4.1.1.18)	A0A0H2XII6
20.	2.16	4.45	Aminotransferase, class V (EC 1.12.-.-)	A0A0H2XJX6
21.	2.08	4.42	Diaminopimelate decarboxylase (DAP decarboxylase) (EC 4.1.1.20)	A0A0H2XFD9
22.	1.93	10.19	Alanine racemase (EC 5.1.1.1)	Q2FF55
23.	1.75	4.26	Uncharacterized protein	A0A0H2XHH8
24.	1.37	4.06	Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol-phosphate transaminase)	Q2FIR7
25.	1.24	4.99	Aminotransferase	A0A0H2XJK0
26.	0.71	4.44	Aminotransferase, class I (EC 2.6.1.-)	A0A0H2XFA8
27.	6.05	6.22	Uncharacterized protein	A0A0H2XIC2
28.	4.13	4.85	Uncharacterized protein	A0A0H2XIU9
29.	2.09	1.93	PTS system, glucose-specific IIA component (EC 2.7.1.69)	A0A0H2XJ98
30.	1.84	3.43	Iron-sulfur cluster repair protein ScdA	Q2FK11
31.	1.76	8.70	Bifunctional purine biosynthesis protein PurH [Includes: Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3);IMP cyclohydrolase (EC 3.5.4.10) (IMP synthase) (Inosinicase)]	Q2FI05

4.6 Conclusion

In conclusion, our chemical proteomic strategy was effective at labeling and enriching PLP-DEs from *S. aureus* USA300 TnPdxS. The use of three PL-derivatives with different chemical and biological properties as probes for PLP-binding enabled labeling of a combined total of 73% of the *S. aureus* PLPome (Figure 4.11). From these results, protein candidates were selected for further validation as putative PLP-DEs. Several of these candidates were identified throughout multiple proteomic experiments and across different probe-labeled samples, which adds additional confidence to their interaction with PLP. Nonetheless, we find marked differences in our labeling results, the types of PLP-DEs identified and the methodological requirements. **PL4** shows effective enrichment of PLP-DEs during a 2-hour *in situ* labeling period, which can be coupled to a two-tiered confidence filtering strategy to extract reliable results. While **PL1** was less effective using this protocol due to chemical lability, its ability to support bacterial growth enabled an alternative labeling strategy which increased its metabolic incorporation and led to the identification of additional PLP-DEs not labeled by **PL4**. The azide-containing **PL3** probe requires the employment of the Staudinger ligation to achieve labeling specificity suitable for proteomic experiments.

Labeling is highly specific during the 2-hour labeling protocol, but fewer proteins are identified than with **PL4**. However, when bacteria are grown directly in the presence of **PL3** as a cofactor substitute, the most complete and specific PLPome labeling profile thus far could be accessed. This enabled the identification of two additional uncharacterized proteins, which were not previously identified, as candidates for further characterization. The validation of PLP-binding to known PLP-DEs, uncharacterized and alternative proteins is described in *Chapter 5*.
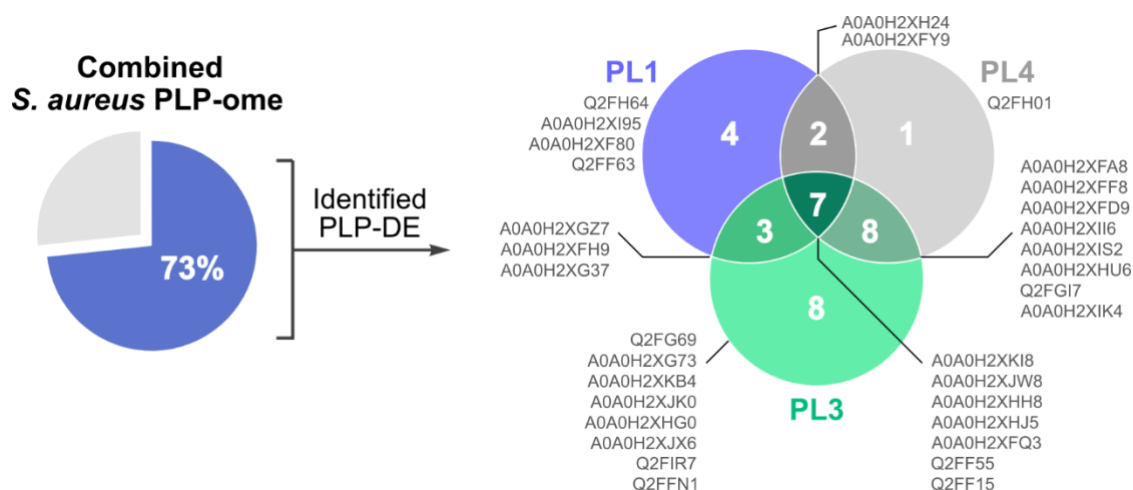
Figure 4.11 Combined total of *S. aureus* USA300 TnPdxS PLP-DEs detected with PL1, PL3 and PL4. A total of 33 proteins, ie. 73 % of the *S. aureus* PLPome could be identified with our probe-labeling strategy. A Venn diagram illustrates the differences in proteins significantly enriched (Uniprot identifiers) using the different probes. The comparison comprises data from the following volcano plots: **PL1** (Figure 4.7a, CDM supplementation with **PL1**), **PL3** (Figure 4.10a, CDM supplementation with **PL3**), and **PL4** (Figure 4.3b).

5. Validation & Enzyme Characterization

5.1 Introduction

Recent estimates indicate that about 30-40% of genes in each genome remain unannotated or only putatively assigned a function.[104] The evolutionary conservation of many uncharacterized proteins (70%) implies that they are functionally relevant or metabolically important.[105,106] Chemical profiling strategies are able to put these proteins into a functional context by their reactivity or affinity, and thereby provide starting points for further investigation of protein function. Current strategies for the full characterization of proteins require the integration of multiple approaches, including bioinformatics, structural, biochemical and cell-based methods.

Bioinformatic strategies for the functional analysis of proteins rely on sequence- or structure-homology searches within corresponding databases. While sequence-based searches have evolved advanced algorithms for detecting distant structural similarities and phylogenetic relationships among proteins, they are not accurate at discriminating protein function. For PLP-DEs, sequence-based analyses has led to their classification into different fold types, but could not assign specific catalytic function due to overlap and multiplicity among the different subgroups.[18] In an alternative approach, the use of reference search models tailored to specific catalytic activities rather than structural fold provided a more detailed genomic overview of PLP-DE activities across different organisms.[4] Search methods that rely on structural data are more accurate at assigning protein function. Strategies that have been applied to PLP-DEs include multiple structure alignments (3DM database analysis and overlays) as well as comparisons of active site architecture.[34,107,108] A combination of these methods was applied by Fleischmann *et al.* in order to establish a robust workflow for characterizing PLP-DEs using three uncharacterized proteins from the human microbiome.[107] Initial secondary structure matching (SSM) identified similarity to known aminotransferases and a Pfam search[109] (protein family database search) was able to assign the corresponding aminotransferase subtypes. Upon structural superimposition with a subset of representative family members, the best matches were selected for structure-based multiple sequence alignment in order to reveal conservation of distinct residues. It was found that active site motifs characteristic of aminotransferases were strictly conserved. PLP-DE

active site structures are often similar due to common mechanistic features associated with the cofactor. However, unique arrangements of residues essential for catalysis remain highly conserved throughout evolution and can be used to infer enzyme function. Active site fingerprinting of slow-evolving features using the program CPASS (comparison of active site structures) has been used to cluster PLP-DE function based on conserved residues proximal to the PLP ligand (6 Å).[34] Structural data and *in silico* analysis thus represent a powerful combination for obtaining an initial hypothesis to pursue in subsequent biological investigations.

Substantial literature precedence (since the 1940's) for the family of PLP-DEs facilitates their study by providing important information on their diverse catalytic mechanisms, unique characteristics and kinetic activities. A common method for ascertaining cofactor binding and deriving qualitative insights into protein function is UV-Vis spectroscopy, which permits the observation of characteristic catalytic intermediates such as the internal aldimine (330-360 nm; 420-430 nm), quinonoid (500 nm), aminoacrylate (470 nm) or PMP (330 nm).[60,110] This strategy was applied by Fleischmann *et al*. to assess substrate preferences of their aminotransferases, revealing considerable substrate promiscuity (≥ 8 amino acids each).[107] The natural enzyme substrates were narrowed down using kinetics assays to screen possible amine acceptors (ex. ketoglutarate, oxaloacetate, pyruvate), as well as slow-cooling molecular dynamics simulation of the modeled external aldimines. A variety of coupled assays have been reported for studying PLP-DE activity, and stopped-flow kinetics are frequently used to directly monitor quinonoid formation or other characteristic intermediates. Additionally, PLP-DE substrates can be determined by screening libraries of compounds and analyzing reaction products by LC-MS or HPLC. Nonetheless, standard *in vitro* experiments do not account for complex biological settings, such as protein interactions, cryptic natural substrates or protein regulation. In particular, studying the roles of PLP-DEs *in vitro* is complicated by substrate promiscuity and functional redundancy.

In vivo or *ex vivo* studies are useful for understanding protein function directly in a native biological environment. Here, the application of proteomic and transcriptomic methods can help to dissect the roles of proteins within biological pathways or networks, as well as their regulation, interaction partners and cellular localization. Moreover, coupling these methods to metabolomic analysis can provide powerful datasets for gaining direct insights into the state of a cell.[111] For example, these strategies can be applied to study the function of a protein by analyzing its cellular implications upon over- or in-activation (ie. overexpression or genetic disruption). This often triggers the activation of compensatory mechanisms, and the resulting proteomic or metabolomic changes can thereby provide clues as to the cellular function of a protein. Alternatively, metabolomics can be used to measure the activity of a protein under near-native conditions using a strategy termed activity-based metabolomic profiling (ABMP).[112,113] The purified protein is therefore incubated with extracted metabolome or cell lysate, and differences in metabolite levels are measured by MS as a readout of enzyme activity. Using native metabolomes ensures the

screening of physiologically-relevant metabolites from the parent organism and provides access to non-commercial substrates. There are several challenges associated with metabolomics, in which compound identification and annotation rely on matching to corresponding database entries (MS, MS/MS). Technical difficulties include the inability to discern isobaric compounds or to account for in-source fragmentation, and selecting the correct measurement technique for metabolites of interest. Furthermore, high variability between samples due to cellular dynamics, compound stability, and metabolite extraction complicate analysis.

Using a suite of bioinformatic, structural, biochemical and cell-based methods, we attempted to confirm the PLP-dependent function of select proteins identified from chemical proteomic experiments (*Chapter 4*). Techniques for validating PLP-binding were first established using known PLP-DEs and then applied to study uncharacterized and alternative-function proteins. We further studied the uncharacterized proteins using several of the strategies outlined above in order to gain functional insights. The identification of novel PLP-dependent roles in the cell represents a major application of our PLPome profiling method.

5.2 Known PLP-DEs

To validate the binding of PLP to proteins identified in our proteomic experiments, we first established a PLP-detection strategy with known PLP-DEs using UV-Vis spectroscopy, gel-based labeling, and intact-protein MS. Therefore, five significantly enriched (high and medium confidence) PLP-DEs of binding clusters 1 and 2 were selected from **PL4**-based proteomic experiments. Different types of PLP-DEs were chosen, including two class I aminotransferases (AT1 and AT2; A0A0H2XFY9 and A0A0H2XFA8), D-alanine aminotransferase (DAT; A0A0H2XHU6), ornithine/lysine/arginine decarboxylase (OKR-DC; A0A0H2XII6) and cysteine synthase (CS; A0A0H2XFQ3), to study a range of PLP-dependent processes. These proteins were recombinantly expressed and purified on an analytical scale by streptavidin affinity chromatography. UV-Vis spectra indicated the presence of the PLP cofactor bound as an internal aldimine (360 or 420 nm, Figure 5.1a) for each protein (Figure 5.1b, 5.2b). We further used UV-Vis in order to observe characteristic spectral properties of PLP-dependent catalytic intermediates.[60,110] For example, upon incubation of CS with the substrate *O*-acetylserine (OAS), formation of the aminoacrylate intermediate (470 nm)[60] was observed (Figure 5.1a,b). It also enabled the qualitative assessment of AT1, AT2 and DAT substrate preferences *in vitro* upon incubating each aminotransferase with various amino acids in the presence of PLP and monitoring PMP formation (330 nm, Figure 5.2a,b). We observed that DAT converts many D-amino acids (7 out of 11 tested), while AT1 and AT2 converted only L-amino acids (3 and 6 out of 17 tested, respectively; Figure 5.2c).
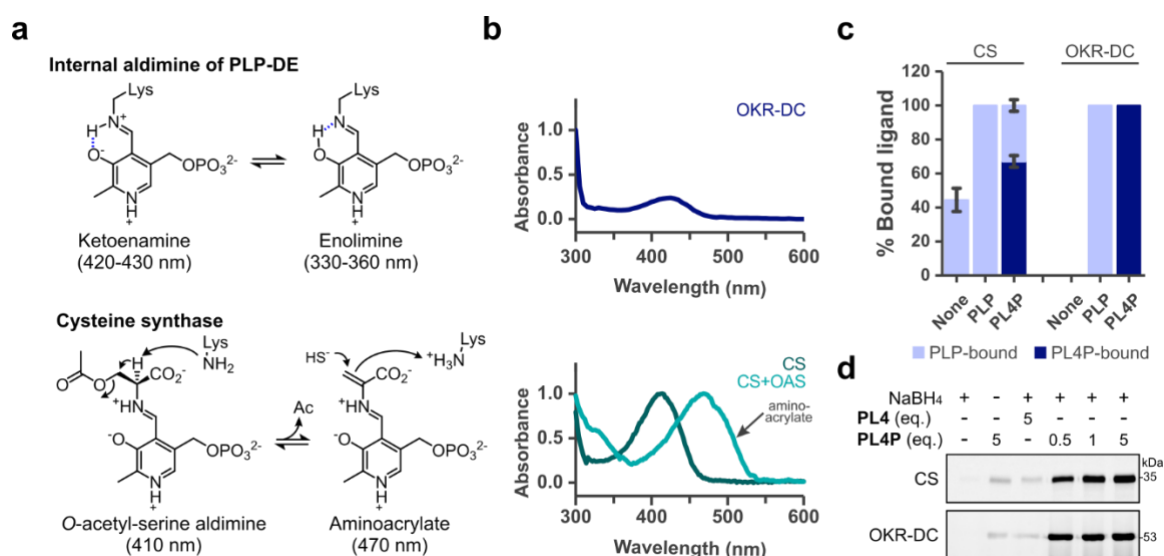
Figure 5.1 Validation of known PLP-DEs OKR-DC and CS by UV-Vis, MS and gel-based analysis. (a) Characteristic spectral properties of internal aldimine and aminoacrylate intermediates.[60,110] (**b**) UV-Vis spectra of OKR-DC (top) and CS ± OAS (10 equivalents) to observe aminoacrylate formation (470 nm, bottom). Curves were normalized to absorbance at 300 nm. (**c**) Proportion of PLP-DE bound upon incubation with PLP or cofactor analogues (4 equivalents), NaBH$_4$ reduction and measurement by intact protein MS (*n* = 3, error bars: mean ± standard deviation). Up to 6% nonspecific addition was observed for CS (PL, **PL4**). (**d**) Fluorescence SDS-PAGE showing *in vitro* labeling of OKR-DC and CS with **PL4P** (0.5, 1, 5 equivalents) or **PL4** (5 equivalents) as a negative control. See appendix for full gels and Coomassie staining.

Our findings are in line with previous reports of the *in vitro* promiscuity of PLP-DEs, and are a starting point for more detailed kinetic investigations of substrate scope. Spectroscopic measurements therefore represent a reliable and sensitive method to confirm the activity of an enzyme and to gain preliminary functional insights, which could be coupled to stopped-flow kinetic experiments for more detailed analysis.

We next corroborated PLP and **PL4P** binding to the selected PLP-DEs using intact-protein mass spectrometry. Incubation of the proteins with excess cofactor followed by NaBH$_4$ reduction resulted in quantifiable, single-fold adduct formation (Figure 5.1c, 5.2d). Control experiments with non-phosphorylated cofactors (PL and **PL4**) did not show adduct formation, thereby proving the specificity of the interaction. For most PLP-DEs, cofactor binding could not be detected without prior incubation with excess PLP. MS-based analysis of binding thus appears to be less suitable for monitoring probe binding than UV-Vis, which detected bound cofactor in all PLP-DEs, perhaps due to poor efficiency of the NaBH$_4$ reduction step. CS and AT1 were approximately 40% bound by PLP and **PL4P** appeared to bind the remaining fraction of apoenzyme without significantly displacing the cofactor from the PLP-holoenzymes. Probe binding and specificity was further reflected in gel-based labeling experiments, in which only the active probe labeled proteins significantly (Figure 5.1d, 5.2e). We therefore conclude that the methods presented here are simple and effective at detecting authentic PLP binding, and can be used across a panel of PLP-DEs.
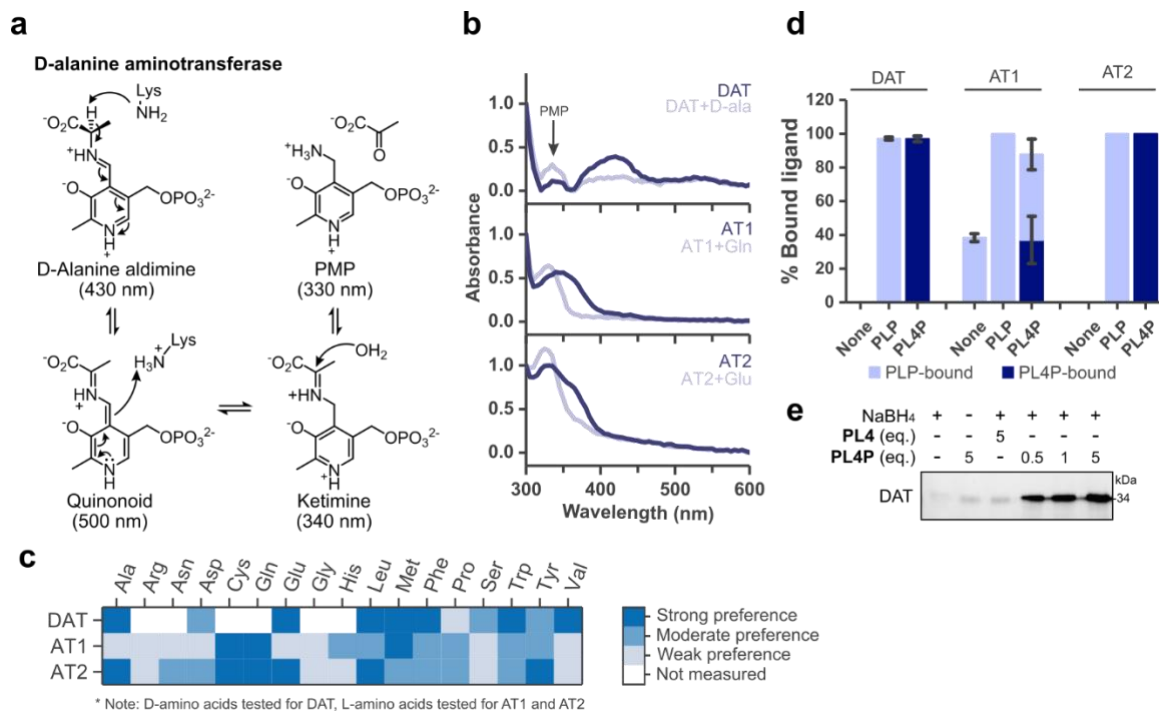
Figure 5.2 Validation of PLP-dependent aminotransferases by UV-Vis, MS and gel-based analysis. (**a**) Characteristic spectral properties of PLP/PMP intermediates during transamination. (**b**) UV-Vis spectra of DAT, AT1 (A0A0H2XFY9) and AT2 (A0A0H2XFA8) ± substrate amino acids (10 equivalents) to observe PMP formation (330 nm). Curves were normalized to absorbance at 300 nm. (**c**) Heatmap representation of *in vitro* substrate preferences of aminotransferases observed by spectral shifts. DAT was tested with D-amino acids; AT1 and AT2 were tested with L-amino acids. (**d**) Proportion of PLP-DE bound upon incubation with PLP or cofactor analogues (4 equivalents), NaBH₄ reduction and measurement by intact protein MS ($n = 3$, error bars: mean ± standard deviation). (**e**) Sample fluorescence SDS-PAGE showing *in vitro* labeling of DAT with **PL4P** (0.5, 1, 5 equivalents) or **PL4** (5 equivalents) as a negative control. See appendix for full gel and Coomassie staining.

5.3 Uncharacterized PLP-DEs

Having validated cofactor binding for several known PLP-DEs, we used the above techniques to assess the PLP-binding of uncharacterized proteins identified from **PL4**-based proteomics. These included Q2FF14 (Uniprot ID; high-confidence; cluster 1), GP0 (Uniprot: A0A0H2XGP0; medium-confidence; cluster 2), and HH8 (Uniprot: A0A0H2XHH8; high-confidence; cluster 1). While the latter protein is annotated to be functionally uncharacterized, it is predicted to bind PLP by homology.[108,114,115] Each of the three proteins appears to bind PLP and **PL4P** based on initial UV-Vis, MS and gel-based labeling experiments (Figure 5.3). UV-Vis spectra were recorded from large-scale protein batches, in which *E. coli* expression lysate was treated with PLP (100 µM, 20 min, 4°C) prior to purification to saturate PLP binding and improve the spectroscopic signal strength. We next proceeded with biochemical characterization according to the strategies outlined in the introduction, in order to gain insights into potential roles of these proteins.
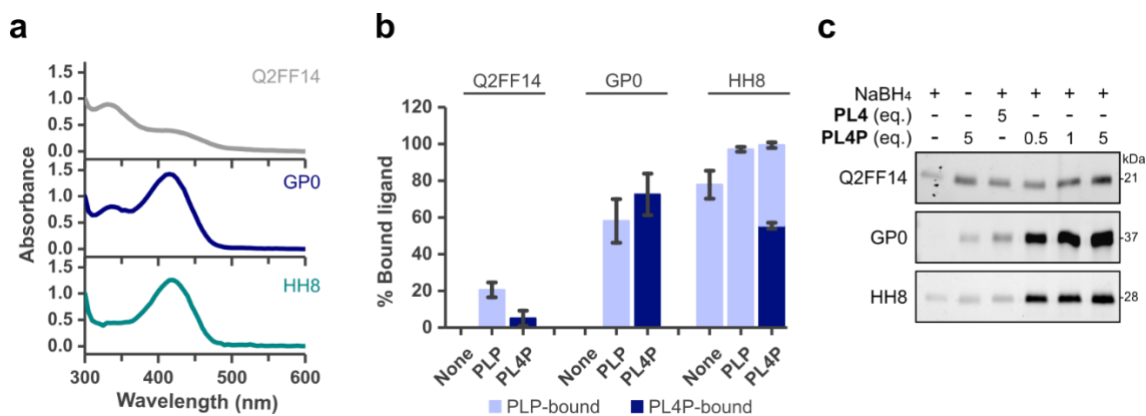
Figure 5.3 Validation of PLP binding for uncharacterized proteins identified using PL4. (**a**) UV-Vis spectra of uncharacterized proteins saturated with PLP (100 µM) in *E. coli* expression lysate; curves were normalized to absorbance at 300 nm. (**b**) Proportion of protein bound upon incubation with PLP or cofactor analogues (Q2FF14, GP0: 4 equivalents, HH8: 1 equivalent), $NaBH_4$ reduction and measurement by intact protein MS (*n* = 3, error bars: mean ± standard deviation). (**c**) Fluorescence SDS-PAGE showing labeling of uncharacterized proteins with **PL4P** (0.5, 1, 5 equivalents) or **PL4** (5 equivalents) as a control. See appendix for full gels and Coomassie staining.

5.3.2   Q2FF14

Q2FF14 is currently annotated as a member of the YwlG superfamily (uncharacterized protein family UPF0340; domain of unknown function DUF436), comprising 68 highly conserved proteins found in Gram-positive bacteria. The protein is 174 residues in length and has a molecular weight of 18.9 kDa. A sequence alignment of Q2FF14 with homologues from other species shows distinct regions of conservation (Figure 5.4). The corresponding gene (SAUSA300_2068) is located upstream from the PLP-DE serine hydroxymethyltransferase, in an operon together with UDP-GlcNAc 2-epimerase and uracil phosphoribosyltransferase (Figure 5.5a), which are important for one-carbon, aminosugar and pyrimidine metabolism, respectively.

Q2FF14 shows a strong absorbance at 330-340 nm, indicative of either the enolimine form of the internal aldimine or a PL-derivative saturated at the C4A carbon, in addition to a modest keto-enamine peak at 420 nm (Figure 5.3a). The peaks are very weak when isolating the protein directly from *E. coli* expression lysate, but prior incubation of the lysate with PLP generates a protein that is visibly yellow and has a correspondingly strong signal. However, the apparent holoenzyme is unstable and readily precipitates at 4°C within 1 h. Our MS-based analysis suggests that Q2FF14 binds PLP and **PL4P** weakly, but specifically over non-phosphorylated counterparts (Figure 5.3b). This weak binding is also reflected in gel-based labeling, in which a 5-fold excess of probe yields only minimal fluorescent signal (Figure 5.3c). These results match our observation that Q2FF14 may be unstable in its PLP-bound form, or has low affinity towards PLP, though requires further confirmation.
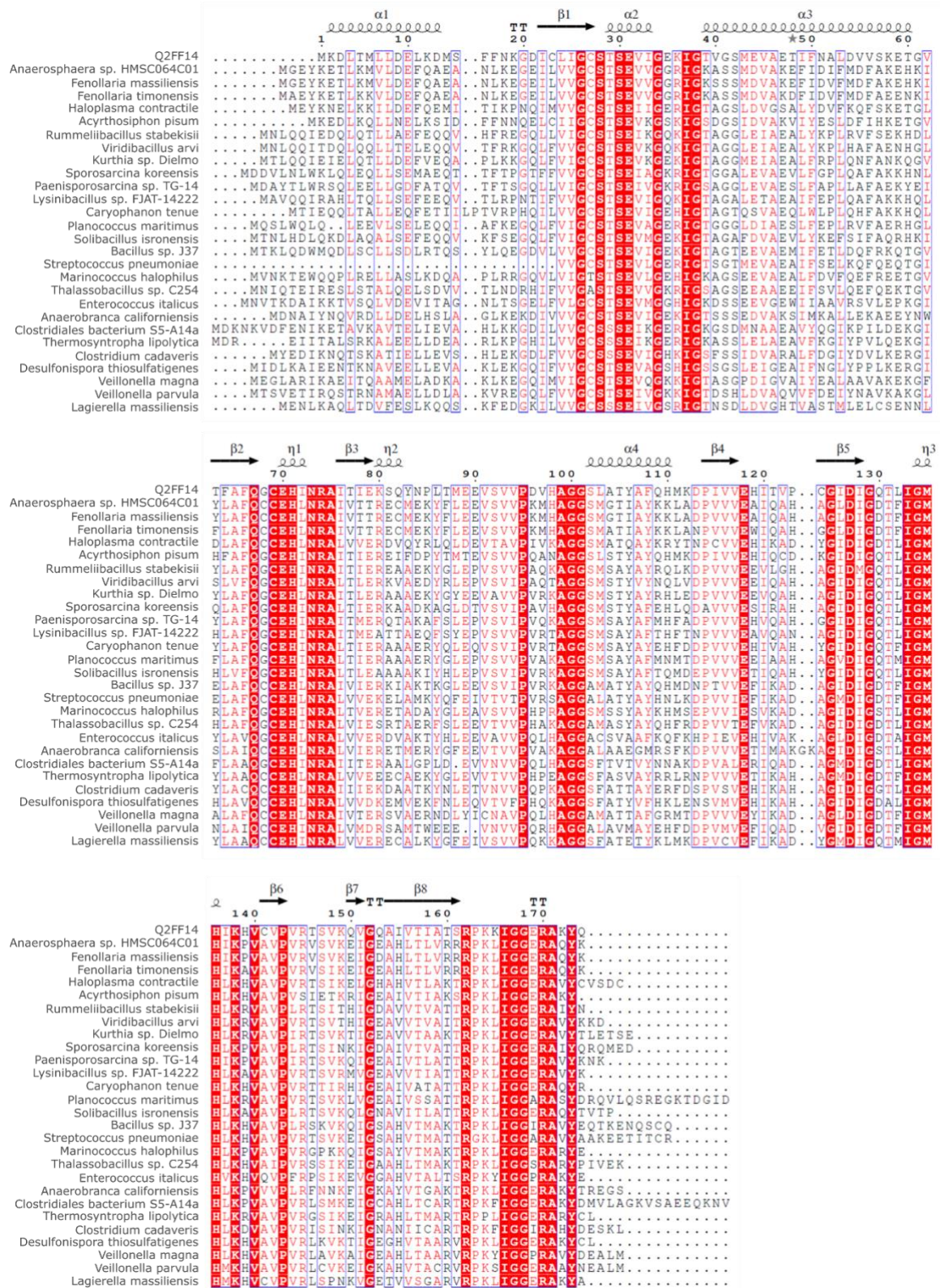
Figure 5.4 Sequence alignment of Q2FF14 with homologous sequences from Gram-positive bacteria. Sequences were selected by Blast search of the Q2FF14 protein sequence, excluding *S. aureus*. Sequences were aligned using ClustalX[116] and processed using ESPript.[117] Figure courtesy of Dr. S. Schneider.

We determined the X-ray crystal structure of Q2FF14 to 1.75 Å resolution and did not detect bound PLP or other ligands (Figure 5.5b,c,d). The protein fold consists of a central 8-stranded β-sheet decorated with α-helices and a long C-terminal extension (Figure 5.5b). Q2FF14 forms a hexameric structure that is made up of two stacked trimeric rings (Figure 5.5d), which matched solution-based SEC analysis (Figure 5.5e). Structural alignment search (DALI)[118] reveals high similarity to the conserved hypothetical protein TT1679 from *Thermus thermophilus* HB8 (PDB: 1v8d),[119] with which it shares 40% sequence identity,[119] as well as distant similarity (11% sequence identity) to glutamate dehydrogenases (e.g. PDB: 1v9l).[120]

We next attempted to elucidate the function of Q2FF14 using substrate screens and cell-based experiments. Incubation of apo-Q2FF14 with a mixture of D/L-amino acids in the presence of PLP and other cofactors, however, did not result in any characteristic spectral peaks nor changes by MS (data not shown). We conclude that Q2FF14 does not transform amino acid substrates or that the protein is inactive under the conditions tested. An activity-based metabolomic profiling approach using *S. aureus* metabolome may therefore be more suitable for identifying a possible substrate under more biologically-relevant conditions.

Cell-based experiments were performed to understand the functional implications of Q2FF14 in a native environment. A strain of *S. aureus* USA300 in which the gene encoding Q2FF14 was disrupted by transposon insertion was obtained from the Nebraska transposon mutant library.[84] The availability of this strain and its growth viability indicate that Q2FF14 is not essential. A whole-proteome analysis served to compare stationary-phase cultures of TnQ2FF14 and wild-type *S. aureus* USA300 (Figure 5.5f). We find that all of the downregulated proteins are within the same operon, initiated by Q2FF14. It is therefore possible that Q2FF14 regulates the expression of other genes within its operon; however, this observation may also result from non-specific mRNA processing effects due to transposon insertion (3.2 kb). Only one uncharacterized protein (probable β-lactamase, A0A0H2XJY4) was significantly upregulated. Similarly, we did not observe any significant dysregulations of metabolites in corresponding metabolomic profiling experiments (data not shown). This suggests that Q2FF14 influences cellular pathways that cannot be significantly detected under the experimental conditions.

To summarize, the conserved uncharacterized protein Q2FF14 appears to bind PLP or a derivative thereof, but subsequently shows *in vitro* instability. We are unable to draw functional conclusions based on the obtained structural data, substrate analysis or cell-based studies. Some of these experiments may require optimizations of conditions, such as cellular growth, in order to clearly distinguish an effect of Q2FF14 activity.
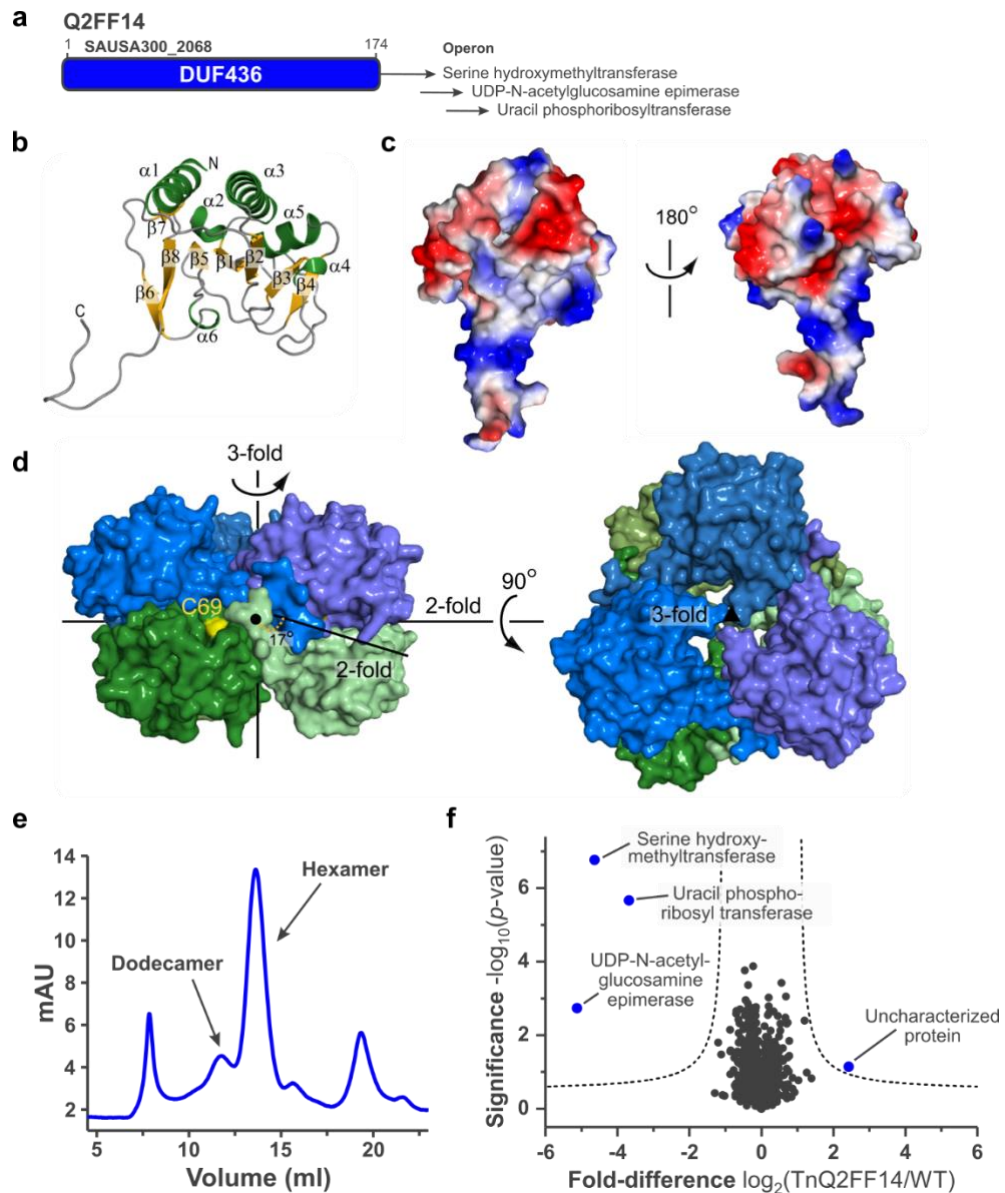
Figure 5.5 X-ray crystallography and structural analysis of Q2FF14. (**a**) Q2FF14 (174 residues) contains a domain of unknown function (DUF436) and initiates a 4-gene operon. (**b**) Secondary structure and protein fold of a Q2FF14 monomer (ribbon representation). (**c**) Electrostatic surface representation of a Q2FF14 monomer. (**d**) Assembly of Q2FF14 into a hexamer composed of two stacked trimers. Crystallography figures courtesy of Dr. S. Schneider. (**e**) Size exclusion chromatography of Q2FF14 shows hexamer formation. (**f**) Volcano plot showing fold-difference between *S. aureus* USA300 Q2FF14 transposon mutant versus wild-type upon performing a two-sample t-test (FDR 0.5, s0 0.5; *n* = 3 replicates).

5.3.2 A0A0H2XGP0

GP0 is an uncharacterized protein composed of 295 residues (34.8 kDa), which contains a C-terminal domain (150-295) predicted to have *N*-acyltransferase function (InterPro, Figure 5.6a). Specifically, this domain belongs to the family of Gcn5-related *N*-acetyltransferases (GNAT, Interpro IPR000182) which use acetyl-CoA to transfer an acetyl group onto primary amines of small molecules and proteins. GNAT proteins have diverse cellular functions in both prokaryotes

and eukaryotes, and acetylate a variety of substrates such as aminoglycoside antibiotics (resistance mechanism), glucosamine-6-phosphate (bacterial cell wall biosynthesis), histones (transcriptional activation), serotonin and spermidine. They are characterized by a structurally conserved fold comprising an arrangement of 6-7 β-strands and 4 α-helices (β_1-α_1-α_2-β_2-β_3-β_4-α_3-β_5-α_4-β_6-β_7). The gene encoding GP0 (SAUSA300_2209) is conserved among staphylococci and is within its own operon, frequently located in proximity to the sugar transporter glcU (SAUSA300_2210). *S. aureus* contains 35 putative GNAT enzymes, many of which are functionally and structurally uncharacterized.[121]

UV-Vis, MS and gel-based analyses revealed a definitive interaction of GP0 with the PLP cofactor (Figure 5.3). In solution, the protein exists primarily as a monomer with small amounts of homodimerization, a common feature of PLP-DEs (Figure 5.6c).[2] Preliminary substrate studies using a pool of D/L-amino acids did not show any conversion by UV-Vis or MS (data not shown). The previous annotation of GP0 as an acetyltransferase led us to conduct a broad substrate screen for acetyltransferase activity with amino acids, antibiotics, and diverse other metabolites (*for exact substrates see* Table 8.3). The assay was based on a literature protocol, using Ellman's reagent (5,5'-dithiobis(2-nitrobenzoic acid), DTNB) as a quantitative, spectrophotometric reporter of the production of free CoA thiol.[122] Using this plate-assay with 48 possible substrates available in our lab (*see* Table 8.3), we did not detect any GP0-based acetylation activity above controls. The application of an activity-based metabolomic profiling approach using extracted *S. aureus* metabolome would enable screening of a broader substrate scope and may therefore lend further insights into GP0 function.

We next assessed the cellular effect of GP0 disruption by comparing a TnGP0 transposon mutant of *S. aureus* USA300 to wild-type on both a proteomic and metabolomic level. We obtained the TnGP0 mutant strain from the Nebraska transposon mutant library,[84] which grew similarly to the wild-type strain. In proteomic experiments, a significant downregulation of proteins involved in amino acid import and biosynthesis was observed (Figure 5.6b, Table 5.1). Particularly, pathways of Ser/Thr, Met/Cys and Val/Ile/Leu were affected, as observed by a decrease in the methionine importer (Q2FJI0) and aspartokinase (A0A0H2XKG7). Furthermore, several lipoproteins and ABC transporters were also downregulated. Only the uncharacterized protein A0A0H2XHF8 was significantly upregulated in the absence of GP0. Considerable alterations in amino acid biosynthesis could also be detected by metabolomic profiling, in which levels of amino acids, acetylated amino acids and short peptides appear to be altered (Figure 5.6d, Table 5.2). Putative pathways that were significantly affected include lysine, proline and tryptophan biosynthesis, as well as UDP-*N*-acetyl-D-glucosamine biosynthesis (biochemical precursor for cell wall components) and cell wall recycling (anhydromuropeptides) (Figure 5.6e)

a **A0A0H2XGP0** (SAUSA300_2209)

1   150   295

GNAT domain

```
          10         20         30         40         50         60         70         80         90        100
MSIVQLYDIT QIKSFIEHSN YESASYLYKL PQQYNEIDVL ITDAIESPGV FSIKENDSIK AIILSFAYDK NKFKVIGPFV ADNYVLSVDT FETLFKAMTS
         110        120        130        140        150        160        170        180        190        200
NQPDDAVFNF SFEEGIQQYK PLMKVIQASY NFTDYYIEAR TRLEEDMHQP NIIPYHKGFY RAFSKLHTTT FKYQAQSPQD IIDSLDDHHH LFLFVSEGLL
         210        220        230        240        250        260        270         80        290
KGYLYLEIDS QQSIAEIKYF SSHVDYRLKG IAFELLAYAL QYAFDNFDIR KVYFKIRNKN NKLIERFNGL GFHINYEYIK FKFESRNVKD QTIPE
```

Figure 5.6 Overview of GP0 and metabolomics analysis. (**a**) GP0 (295 residues) contains a C-terminal Gcn5-related acetyltransferase (GNAT) domain. (**b**) Volcano plot showing fold-difference of proteins between *S. aureus* USA300 GP0 transposon mutant versus wild-type upon performing a two-sample t-test (FDR 0.5, s0 0.5; *n* = 3 replicates). (**c**) Size exclusion chromatography shows GP0 to exist mainly as a dimer. (**d**) Volcano plot showing fold-difference of metabolites between *S. aureus* USA300 GP0 transposon mutant versus wild-type (significance cut-offs shown at *p*-value ≤ 0.5, fold-difference ≥ 2; *n* = 7 replicates). (**e**) Cloud plot representation of putative metabolic pathways influenced by the absence of GP0 in cell-based metabolomic profiling experiments generated with XCMS online.

Table 5.1 Whole proteome comparison of *S. aureus* TnGP0 vs. wild-type. Corresponding volcano plot in Figure 5.6b.

Uniprot ID	Name	Gene	Gene Ontology
	Up-regulated proteins		
A0A0H2XHF8	Uncharacterized protein	SAUSA300_1489	
	Down-regulated proteins		
Q2FF67	2-isopropylmalate synthase (EC 2.3.3.13) (Alpha-isopropylmalate synthase)	SAUSA300_2010 (leuA)	leucine biosynthesis [GO:0009098]
A0A0H2XKG7	Aspartokinase (EC 2.7.2.4)	SAUSA300_1225	lysine biosynthesis via DAP [GO:0009089]; threonine biosynthesis process [GO:0009088]
A0A0H2XHB0	Uncharacterized protein	SAUSA300_0177	
A0A0H2XI71	Peptide ABC transporter, ATP-binding protein	SAUSA300_0200	peptide transport [GO:0015833]
A0A0H2XJL5	Uncharacterized protein	SAUSA300_0173	
A0A0H2XFT9	Lipoprotein	SAUSA300_0437	
Q2FJI0	Methionine import ATP-binding protein MetN 1 (EC 3.6.3.-)	SAUSA300_0435 (metN1)	methionine transport [GO:0015821]
A0A0H2XJJ1	Putative lipoprotein	SAUSA300_0175	
A0A0H2XIF6	Putative lipoprotein	SAUSA300_0203	

Table 5.2 Whole metabolome comparison of *S. aureus* TnGP0 vs wild-type. Select up/downregulated features from volcano plot Figure 5.6d identified using the Metlin database.[123]

Feature	Fold-difference	Significance	Predicted Metabolite #1	Predicted Metabolite #2	Predicted Metabolite #3
		Up-regulated metabolites			
374.2245	2.17	1.08	-	-	-
491.3030	1.91	1.42	CysLeuLeuLys	GlyAsnLysArg	
118.0853	1.84	2.94	L-Valine		
231.1330	1.50	5.94	Hydroxyprolylvaline	Valylhydroxyproline	
189.1224	1.40	1.89	*N*-ε-acetyllysine	ValAla	
175.1069	1.36	11.85	*N*-acetylornithine	ValGly	*N*-acetylproline
90.0544	1.26	7.31	L-Alanine		
		Down-regulated metabolites			
442.2354	-2.79	1.38	AspProLeuPro		
441.2327	-2.31	1.55	IleLeuSerSer		
441.2319	-2.27	1.46	AspLeuProPro		
104.1062	-1.19	1.26	-	-	
203.1018	-1.00	4.09	SerPro	Alanylhydroxyproline	*N*-acetyl-β-glucosamine

In summary, we can confirm GP0 to bind PLP and could detect alterations in PLP-dependent processes, such as amino acid biosynthesis, on a cellular level in its absence. However, it is difficult to draw specific function from cell-based proteomic and metabolomic analyses, in which multiple mechanisms may compensate for the loss of a protein and the interrelatedness of biological networks makes interpretation difficult. Though GP0 is predicted to have an acetyltransferase domain similar to GNAT proteins, we do not detect acetyltransferase activity across a range of possible substrates. Limitations in structural data acquisition currently prevent us from confirming this structure-based prediction and drawing conclusions as to the N-terminal half of the protein, which often serves substrate recognition in GNAT proteins. To date, the only reported PLP-dependent acetyltransferase activity is that of glycine C-acetyltransferase, also known as 2-amino-3-ketobutyrate CoA ligase (A0A0H2XJW8).

5.3.3   A0A0H2XHH8

HH8 (A0A0H2XHH8) is annotated as an uncharacterized protein (UPF001) which is predicted to be part of a functionally unknown subgroup of the alanine racemase family (Pfam: PF01168). The universal distribution and evolutionary conservation of proteins belonging to this subgroup argue for its importance in cellular functions. Crystal structures of homologues from *Saccharomyces cerevisiae* (YBL036C, PDB: 1CT5/1B54)[114] and *Escherichia coli* (YggS, PDB: 1W8G)[124] have revealed similarity to bacterial alanine racemase and eukaryotic ornithine decarboxylase, which were previously not detected by sequence-based searches. However, unlike Alr, HH8 homologues are single-domain monomeric structures which lack the Alr active site base Tyr265.[114] Furthermore, they are unable to racemize amino acids[114,115] or showed only weak activity upon overnight incubation,[108] leading to the hypothesis that they are non-specific racemases with a yet unknown substrate. However, other functional investigations have implicated this protein class in PLP homeostasis.[115,125] Here, it was shown that the ΔYggS *E. coli* strain accumulated PNP and that high levels of PN (but not PM, PL or PLP) were toxic to this strain.[125] Comparative genomics in *E. coli* further linked this protein subfamily to PLP salvage, cell division and cell wall biosynthesis.[125] In a previous genomic clustering experiment, the human homologue PROSC (proline synthetase co-transcribed protein) was linked to proline synthesis, though this likely relates to another function. In humans, deficiency of PROSC causes vitamin B6-dependent epilepsy.[126] This strengthens its functional link to PLP homeostasis, and its dysregulation may thus influence secondary processes such as amino acid metabolism.

In our studies, we confirm the *S. aureus* USA300 enzyme HH8 to bind PLP based on UV-Vis and MS experiments (Figure 5.3). Interestingly, we found HH8 to be nearly completely saturated with PLP upon purification from the *E. coli* expression strain, even without any prior addition of PLP to the lysate. This was not the case for any other PLP-DE studied here, where maximally half of the protein population showed PLP-binding (*see* Figures 5.1c, 5.2d). Furthermore, PLP adduct formation could be detected even in the absence of NaBH$_4$ reduction (data not shown). These observations argue for a strong interaction of HH8 with PLP and a high cellular binding affinity. In agreement with previous studies, SEC analysis indicates that HH8 is present exclusively in its monomeric form (Figure 5.7b), and we do no detect Alr activity at a relevant timescale using our coupled assay (data not shown). Studies have suggested that HH8 homologues are essential for some organisms (ex. *P. aeruginosa* and *S. aureus*) but are dispensable for others.[125] The lack of a corresponding HH8 transposon mutant in the *S. aureus* USA300 Nebraska transposon mutant library supports this conclusion. In *S. aureus*, we find the gene encoding HH8 (SAUSA300_1082) to be located in an operon with three other hypothetical proteins (Figure 5.7a) and therefore cannot draw any conclusions from its genomic environment. In summary, our results match previous literature reports of this PLP-dependent uncharacterized protein family, whose function remains elusive though likely includes a role in PLP homeostasis.
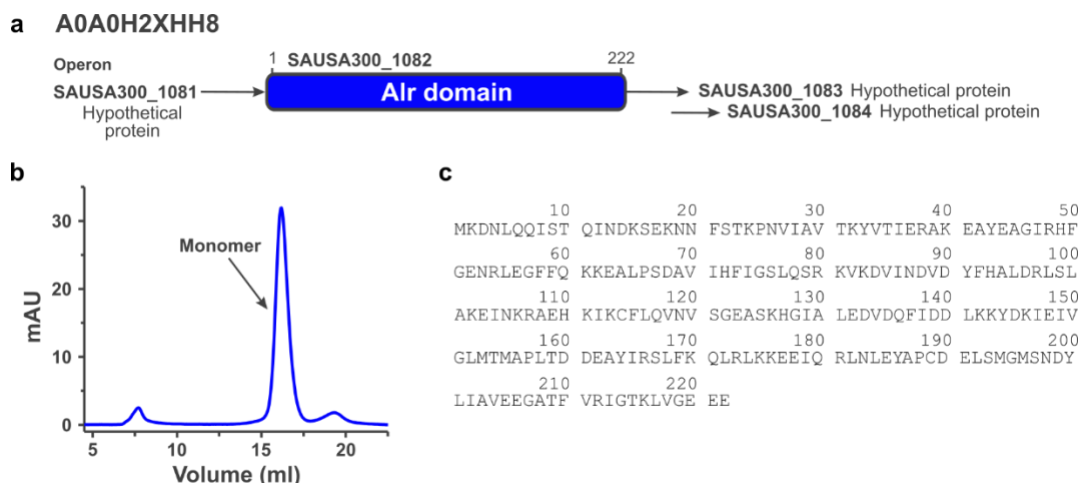
Figure 5.7 Overview of HH8. (**a**) HH8 contains a domain similar to Alr and is part of a 4-gene operon with hypothetical proteins of unknown function. (**b**) Size exclusion chromatography of HH8 shows it is present exclusively as a monomer. (**c**) Protein sequence of HH8 (222 residues).

5.4 Alternative function proteins

Finally, we proceeded with the analysis of three alternative function proteins, which represent proteins identified using our probe strategy that are annotated to have non-PLP-related functions. These included coprohaem decarboxylase (HemQ; Q2FJ56), pseudouridine-5'-phosphate glycosidase (PsuG; A0A0H2XJC6) and ferrochelatase (HemH; A0A0H2XGS5). HemQ was initially annotated as a putative haem peroxidase[127] and was recently classified as coprohaem decarboxylase.[128] This enzyme forms a non-classical haem biosynthetic pathway found in Gram positive bacteria together with HemY and HemH,[128,129] the latter of which has been suggested to interact with PLP for ferrous iron chelation.[130] PsuG, on the other hand, is part of a large family of pseudouridine-5'-phosphate glycosidases, which are uniquely present in bacteria and catalyze the metabolism of the nonclassical RNA nucleotide pseudouridine (5-ribosyluracil).[131,132] While the role of pseudouridine is unclear, it appears to be important in the rigidification of higher-order RNA structures. Not all proteins annotated as PsuG have the same function, as exemplified by a subclass involved in the biosynthesis of blue pigments.[132] The selected proteins were ranked as high-confidence binders based on their significant and reproducible appearance throughout multiple proteomic replicates and experimental setups. We therefore hypothesized that they may have a yet unknown interaction with PLP or have additional unannotated roles.

The alternative proteins were purified on a small scale and their binding to PLP was assessed using UV-Vis, MS and fluorescence SDS-PAGE methods. Based on MS- and gel-based labeling experiments, HemQ appears to bind the PL scaffold non-specifically since both phosphorylated

(**PL4P**) and non-phosphorylated (**PL4**) probes bound the protein to equal extents (Figure 5.8c). Contrastingly, HemH and PsuG bound the phosphorylated cofactors PLP and **PL4P** weakly, but specifically over inactive forms by MS-analysis, and showed enhanced binding to the probe over the natural cofactor. These trends were further reflected in gel-based experiments; however, PLP binding was not observed spectrally for any of the purified proteins. Lack of UV-Vis absorption suggests that the proteins do not carry PLP in their pockets as an internal aldimine, and it is therefore possible that an interaction with PLP may instead be transient. Due to the unclear binding trends observed in our experiments, we conclude that the alternative enzymes studied here either bind our probe non-specifically, are inactive using the current purification protocol, or require certain conditions not present in our *in vitro* system to incite PLP-binding.
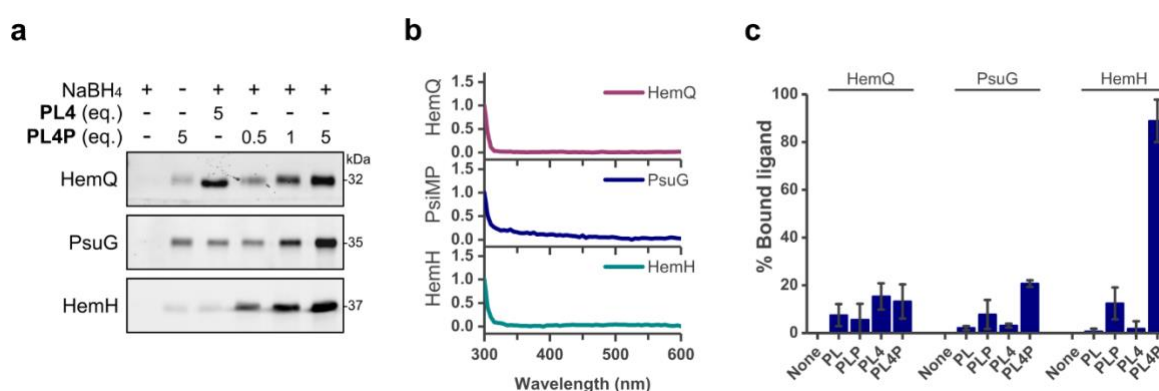
Figure 5.8 Analysis of PLP binding for alternative function proteins. (**a**) Fluorescence SDS-PAGE showing labeling of select alternative function proteins with **PL4P** (0.5, 1, 5 equivalents) or **PL4** (5 equivalents) as a negative control. (**b**) UV-Vis spectra of alternative function proteins. Curves were normalized to absorbance at 300 nm. (**c**) Percent of alternative function proteins bound upon incubation with PL, PLP, **PL4**, **PL4P** (4 equivalents), NaBH$_4$ reduction and measurement by intact protein MS (n = 3, error bars: mean ± standard deviation). See appendix for full gels and Coomassie staining.

5.5 Binding site identification

MS-based methods have successfully been used to map a variety of posttranslational modifications to distinct protein residues. PLP-binding sites have classically been determined using laborious methods such as point mutational studies coupled with UV-Vis spectroscopy and enzyme activity assays,[133,134] or crystallography. Recent reports have attempted to establish MS-methodology for directly assigning PLP-binding sites to recombinant proteins[36] as well as protein mixtures from cell lysates.[95] These studies have underscored the challenges as well as the unique characteristics associated with PLP-bound peptides that are fixed by reductive amination throughout MS-measurement. For example, PLP-peptides exhibit distinct spectral features in CID that correspond to the neutral loss of phosphate [H$_3$PO$_4$] (-98 Da) and the PLP moiety [C$_8$H$_{10}$NO$_5$P] (-231 Da) from the precursor peptide ion. These diagnostic peaks can be used during peptide mass

fingerprinting in order to identify PLP-sites by neutral loss scanning. However, a corresponding lack in backbone fragmentation often hinders subsequent sequence interpretation, an effect also observed for phosphopeptides. Thus, the application of alternative fragmentation techniques such as ETD (electron transfer dissociation), ECD (electron capture dissociation) or neutral loss-triggered MS3 acquisition may be required, which have proven effective for the analysis of other labile modifications. Furthermore, PLP-peptides exhibit suppressed precursor ionization efficiency over non-modified peptides, making their identification in complex mixtures difficult. Therefore, enrichment techniques such as IMAC (immobilized metal-affinity chromatography) or fractionation by liquid chromatography can be used to simplify sample peptide mixtures for MS-measurement. In a recent report, enrichment of PLP-bound peptides using Ti^{4+}-IMAC was followed by treatment with an alkaline phosphatase (ALP) to selectively remove the PL-phosphate moiety for improved ionization and to distinguish from co-enriched phosphopeptides. This led to the identification of one PLP binding site from *E. coli* lysate (K270, tryptophanase). In summary, measuring PLP-binding sites on proteins directly using MS is accompanied by several technical complications which are exacerbated in complex mixtures such as cell lysate.

In light of these challenges, we attempted to use MS-based methods in order to localize PLP binding to specific lysine residues of purified proteins. Both purified PLP-DEs and uncharacterized proteins were thus reduced with $NaBH_4$ in the presence of excess PLP. The proteins were then digested using chymotrypsin or GluC as alternatives to trypsin, which cannot cleave at PLP-modified sites.[135,136] During MS measurement of the resulting peptide mixture, we found that conducting both the full scan as well as MS/MS fragmentation by HCD in the orbitrap was optimal. A representative example of an MS/MS spectrum depicting the known PLP binding site of Alr (K39) is shown in Figure 5.9a; all remaining spectra are shown in the appendix. Using our method, we did not detect the neutral loss of phosphate or the PLP moiety. Despite also observing low detection of PLP-modified peptides, we were able to identify binding sites with high scores and posterior error probability (PEP) (Figure 5.9b). Our data showed excellent sequence coverage and the experimentally-determined sites were largely in agreement with literature reports for the known PLP-DEs (Interpro).[137] The observed inconsistency of the predicted versus the identified site for D-amino acid aminotransferase (A0A0H2XHU6) was analyzed several times under various digestion and measurement conditions, and the predicted site K146 was repeatedly found to be unmodified. Based on our data, we propose K172 and K58 to be possible binding sites of the uncharacterized PLP-binding proteins A0A0H2XGP0 and Q2FF14, respectively.

Figure 5.9 Binding site identification. (**a**) Sample MS/MS spectrum (MaxQuant) for the PLP binding site peptide of Alr (Lys39). All other spectra are shown in the appendix. (**b**) Table of highest confidence PLP binding sites for known and uncharacterized proteins. PEP = posterior error probability (MaxQuant). Seq. Cov. = protein sequence coverage. Predicted sites were obtained using InterPro;[137] [a] known site;[83] [b] predicted site K146 was measured as unmodified.

5.6 Conclusion

In conclusion, the application of our chemical probe strategy facilitated the identification of authentic PLP-binding proteins in cells. Our studies with known PLP-DEs helped to outline a simple and robust sequence of steps in order to validate PLP-binding to purified proteins *in vitro*, and can be used across a panel of PLP-DEs. We can confirm the uncharacterized proteins Q2FF14 and GP0 to interact with PLP, as well as the predicted PLP-DE HH8 to have an unusually strong interaction with the cofactor. Our proteomics experiments also identified proteins with alternative, non-PLP-related functions, some of which may be inhibited by PLP. The proteins selected for study appear to bind the PL-probe scaffold non-specifically or require native biological conditions to observe binding.

We propose enzymes of the UPF0340 family to have a unique, PLP-related function due to its unique structure, while GP0 may represent a putative PLP-dependent acetyltransferase based on

its sequence similarity to Gcn5-related acetyltransferases. Our functional investigations are only preliminary and require further experimental optimization to draw useful conclusions. Currently, we do no detect any cellular repercussions of the Q2FF14 transposon mutation, suggesting its function cannot be detected under the current experimental setup. The crystal structure of Q2FF14 does not show significant similarity to other known structures and does not contain bound PLP, which appears to bind Q2FF14 but renders it unstable. GP0 contains a domain that is highly similar to Gcn5-related acetyltransferases, but it cannot be ascribed a corresponding function based on a broad-substrate screen for GNAT activity. It is possible that the enzyme acetylates other substrates not tested in our experiments or transfers alternative acyl groups. GP0 gene disruption moderately affects amino acid biosynthetic pathways, though a precise function cannot yet be pinpointed.

The detailed functional annotation of proteins requires significant experimental work and we have applied several state-of-the-art techniques to understand the function of two uncharacterized proteins. Additional experiments worthy of pursuit include activity-based metabolomics as a native, broad substrate screen of enzyme activity. Furthermore, future experiments will also investigate the uncharacterized proteins A0A0H2XIC2 and A0A0H2XIU9, identified in **PL3**-based proteomic experiments as putative PLP-DEs. In conclusion, we have applied a method for extrapolating knowledge from the proteome in order to annotate proteins as part of the PLP-DE family without prior knowledge and provide a starting point for further investigations of UPF0340 proteins and PLP-dependent acetyltransferases.

6. Applications

With a functional proteomic strategy for studying PLP-DEs in hand, we next investigated diverse applications of the method beyond the identification of novel PLP-DEs. These include drug development-based applications, such as enzyme active site mapping and inhibitor off-target screening, as well as expansion of the method for use in other organisms.

6.1 Comparison of PLP-DE active site constraints

The family of PLP-DEs has evolved different active site features which help to direct the specificity of chemical reactions. Here, the architecture of the binding pocket influences the orientation of substrate binding and inititally directs reactions through stereoelectronic effects. Furthermore, active site residues control PLP-chemistry by providing the necessary functionality for reaction and tuning the electrophilicity of the system. As a result of the active site variability, PLP-DEs are able to accommodate modification of the PLP scaffold to varying extents. Previous studies into the effects of structurally modifying PLP on cofactor viability have helped to define the core functionality required for catalysis (*see Chapter 2 Introduction).* These include the 3-phenol, 4'-aldehyde and 5'-phosphate, which are essential for binding to active sites and catalysis, through hydrogen bonding and resonance.

Activity-based probe libraries have been useful for studying enzyme active sites and their structural binding preferences in native biological environments. In order to profile structural constraints within PLP-binding sites on a global level, we generated two additional PL probes bearing increased steric bulk at the 2-position, including an extended alkyl chain (**PL5**) and an aromatic group (**PL6**) (Scheme 6.1). The probes were synthesized according to the alkylation strategy developed for **PL4**. Therefore, acetal-protected PL-intermediate **13** was reacted with either TMS-protected 6-bromohexyne or 4-ethynyl-benzyl iodide, to generate intermediates **15** and **17**, respectively. TMS-deprotection of the alkyne using potassium carbonate was followed by acidic deprotection of the acetal and MOM groups of both intermediates in order to produce the final probes **PL5** and **PL6**.
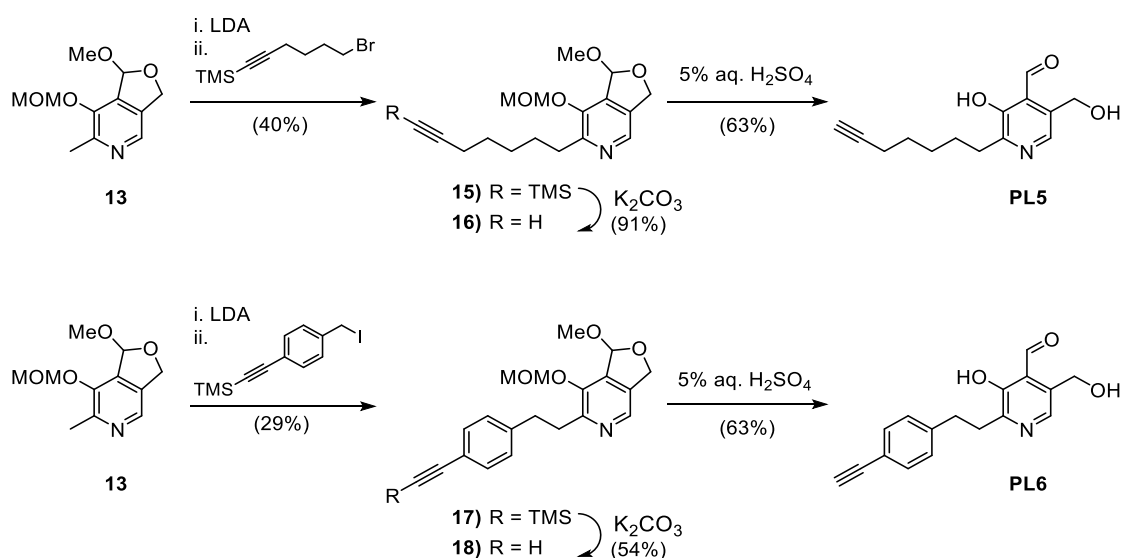
Scheme 6.1 Synthesis of PL5 and PL6. Extended linkers of **PL5** and **PL6** were installed by alkylation of **13** using LDA and corresponding alkyl halides. Two-step deprotection using K_2CO_3 followed by dilute aqueous H_2SO_4 yielded the respective probes.

To determine the influence of these modifications on phosphorylation by the *S. aureus* PLK, we performed kinetic and MS-based assays (Figure 6.1). We found that PLK converts the bulkier PL-derivatives at reduced rates, likely due to steric clash within the enzyme active site. Nonetheless, probe phosphorylation could be detected by MS following incubation of the probes with the kinase.

PLK Kinetics

Substrate	k_{cat} (min^{-1})	K_m (µM)	V_{max} (µM/min)
PL5	0.84 ± 0.11	68 ± 27	0.42 ± 0.06
PL6	0.28 ± 0.03	47 ± 16	0.14 ± 0.01

MS Measurement

Substrate	$MW_{calculated}$ (g/mol)	MW_{found} (g/mol)
PL5P	328.09445	328.09518
PL6P	362.07880	362.07956
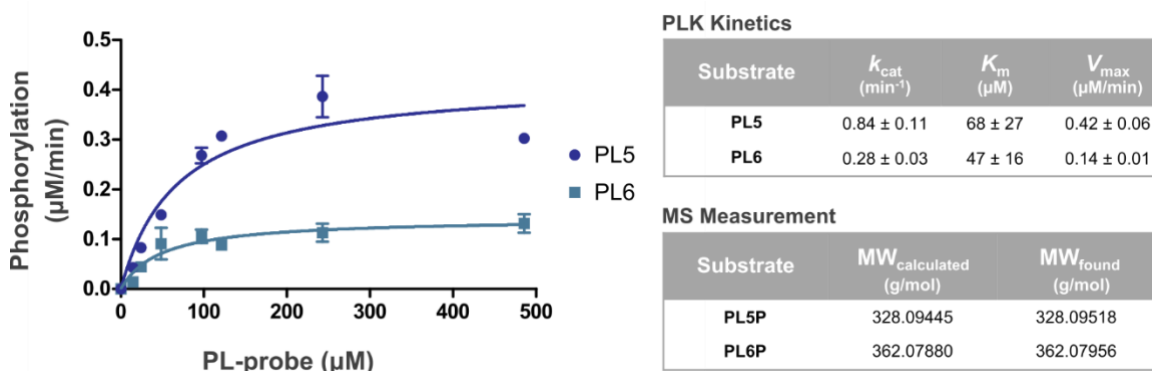
Figure 6.1 Phosphorylation of PL5 and PL6 by *S. aureus* PLK. Kinetics of probe phosphorylation by PLK. (*n* = 4, error bars: mean ± standard deviation) and HRMS confirmation of phosphorylated products.

Significant gel-based labeling of *S. aureus* USA300 TnPdxS using **PL5** and **PL6** prompted us to explore these probes using MS-based proteomics (Figure 6.2, Table 6.1). Comparison of the relative LFQ-intensities of PLP-DEs enriched by **PL4**, **PL5** and **PL6** within the same experiment revealed trends in structural tolerance among the enzymes. Generally, PLP-DEs were identified at successively lower intensities as structural bulk at the 2-methyl position increased and significantly more background labeling was observed. Interestingly, proteins such as cysteine synthase

(A0A0H2XFQ3), the class-I aminotransferase (A0A0H2XFY9),cystathionine gamma-synthase (A0A0H2XFF8), and uncharacterized putative PLP-homeostasis protein (A0A0H2XHH8) all tolerated the extended chain of **PL5** well and could be identified with intensities in the range of **PL4**-based labeling. Similarly, alanine racemase (Q2FF55), which was proven to accommodate structural alterations of the cofactor, was also significantly enriched using **PL5**. Other proteins, including threonine synthase (A0A0H2XH24), threonine deaminase (Q2FH01) and the uncharacterized protein A0A0H2XGP0 preferred the aromatic bulk of **PL6**. A few proteins showed higher intensities with **PL5** (aminotransferases A0A0H2XFA8 and A0A0H2XIS2) or **PL6** (putative PLP-dependent acyltransferase A0A0H2XJW8 and putative aluminum resistance protein A0A0H2XHV8) than with **PL4**. However, nearly half of the proteins were only strongly enriched using **PL4**, including the uncharacterized protein Q2FF14. This data therefore reaffirms that limited modification of the PLP-scaffold permits access to the largest number and variety of PLP-DEs and underscores the evolution of this protein family towards binding a defined PLP structure.
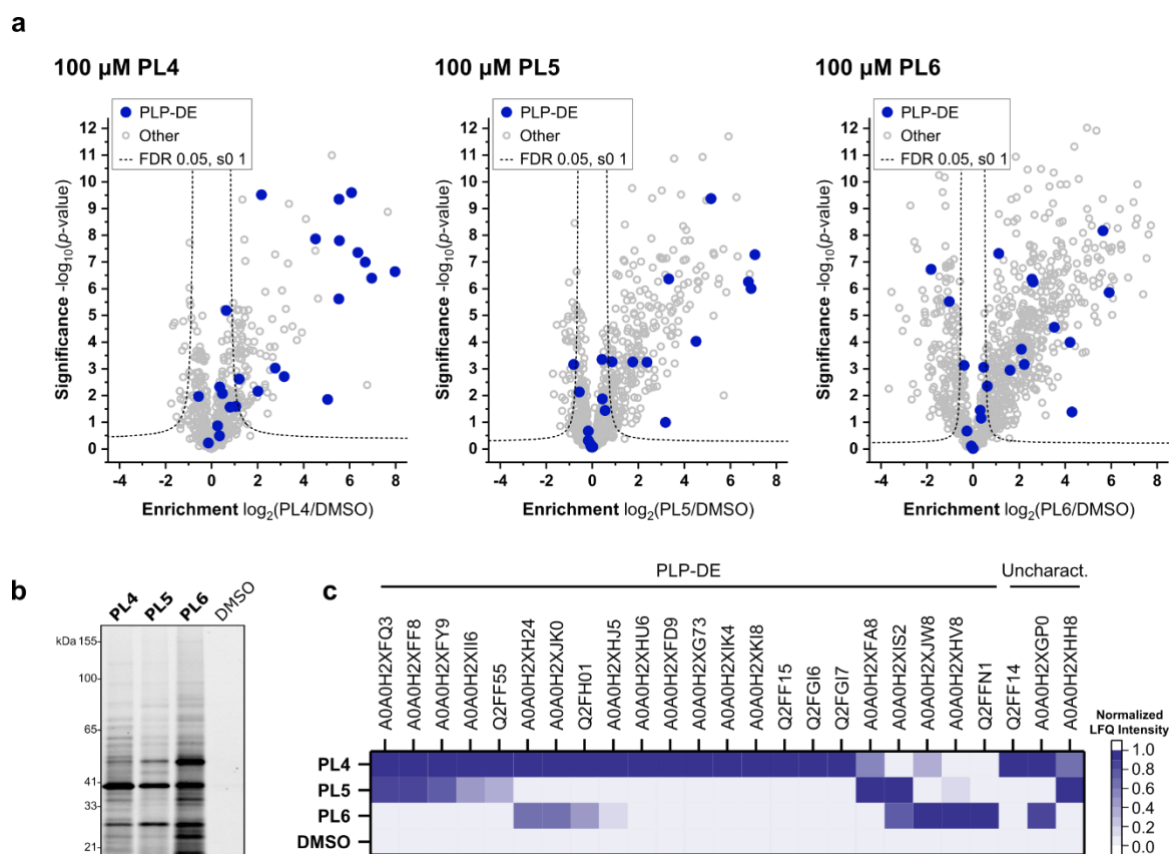
Figure 6.2 Chemical proteomics of PL4, PL5 and PL6 in *S. aureus* USA300 TnPdxS. (**a**) Volcano plots showing enrichment of proteins with 100 μM **PL4**, **PL5** and **PL6** (*n* = 6 biological replicates). Known PLP-DEs are highlighted with blue dots and are listed in Table 6.1. (**b**) Fluorescence SDS-PAGE showing labeling of **PL4**, **PL5** and **PL6** (100 μM each) in *S. aureus* USA300 TnPdxS upon CuAAC to a trifunctional rhodamine-biotin-azide tag and avidin bead enrichment. (**c**) Heatmap representation of proteomic data showing normalized LFQ intensities of PLP-DEs and uncharacterized proteins identified.

Table 6.1 Comparison of PLP-DEs significantly enriched using PL4, PL5 and PL6 in *S. aureus* USA300 TnPdxS. Corresponding volcano plots Figure 6.2a. PLP-DEs significantly enriched with respective probes are marked (+).

	PL4 (sig.)	PL5 (sig.)	PL6 (sig.)	Name	Uniprot ID
1.	+	+		Alanine racemase (EC 5.1.1.1)	Q2FF55
2.	+	+		Aminotransferase, class I (EC 2.6.1.-)	A0A0H2XFY9
3.	+	+	+	Aminotransferase, class I (EC 2.6.1.-)	A0A0H2XFA8
4.	+			Aminotransferase, class V	A0A0H2XKI8
5.	+	+	+	Branched-chain-amino-acid aminotransferase (EC 2.6.1.42)	A0A0H2XIS2
6.	+	+		Cystathionine gamma-synthase (EC 4.4.1.8)	A0A0H2XFF8
7.	+		+	Cysteine desulfurase (EC 2.8.1.7)	A0A0H2XHJ5
8.	+	+	+	Cysteine synthase (EC 2.5.1.47)	A0A0H2XFQ3
9.	+			Cysteine synthase/cystathionine beta-synthase (EC 2.5.1.47)	A0A0H2XG73
10.	+			D-alanine aminotransferase (EC 2.6.1.21)	A0A0H2XHU6
11.	+	+	+	Diaminopimelate decarboxylase (DAP decarboxylase) (DAPDC) (EC 4.1.1.20)	A0A0H2XFD9
12.	+	+	+	L-threonine dehydratase catabolic TdcB (EC 4.3.1.19) (Threonine deaminase)	Q2FH01
13.	+	+	+	Orn/Lys/Arg decarboxylase (EC 4.1.1.18)	A0A0H2XII6
14.	+		+	Threonine synthase (EC 4.2.3.1)	A0A0H2XH24
15.	+	+	+	Uncharacterized protein	A0A0H2XGP0
16.	+	+	+	Uncharacterized protein	A0A0H2XHH8
17.	+			UPF0340 protein SAUSA300_2068	Q2FF14
18.				Probable glycine dehydrogenase (decarboxylating) subunit 2 (EC 1.4.4.2)	Q2FGI7
19.		+		Putative aluminium resistance protein	A0A0H2XHV8
20.		+		Putative pyridoxal phosphate-dependent acyltransferase	A0A0H2XJW8
21.				Aminotransferase	A0A0H2XJK0
22.				Glutamate-1-semialdehyde 2,1-aminomutase 2 (GSA 2) (EC 5.4.3.8)	Q2FFN1
23.				Ornithine aminotransferase (OAT) (EC 2.6.1.13)	A0A0H2XIK4
24.				Probable glycine dehydrogenase (decarboxylating) subunit 1 (EC 1.4.4.2)	Q2FGI6
25.				Serine hydroxymethyltransferase (SHMT) (Serine methylase) (EC 2.1.2.1)	Q2FF15

This study acts as a proof-of-concept, illustrating how PLP-derivatives can be applied to actively probe space within PLP binding sites on a proteome-wide level. Analyses of this type can also help to define structure-activity and structure-selectivity relationships across a panel of PLP-DEs and can provide useful information for their specific targeting by PLP-DE inhibitors.

6.2   In situ inhibitor specificity of the antibiotic D-cycloserine

The natural product D-cycloserine, produced by the bacterial strains *Streptomyces garyphalus* and *lavendulae*, is known to inhibit the PLP-DE alanine racemase (Alr).[76,138] Alr generates the essential peptidoglycan precursor D-alanine by interconverting the two amino acid isoforms. The peptidoglycan is a bacterial cell wall component, composed of a meshwork of long carbohydrate chains (alternating *N*-acetylglucosamine and *N*-acetylmuramic acid) crosslinked by short peptide bridges (D-ala, D-glu, L-lys), which provides cellular rigidity and support.[139] Since Alr is universal to bacteria but absent from most eukaryotes, it is considered a promising therapeutic target. Although several bacteria (ex. *E. coli*, *P. aeruginosa*, *S. aureus*) contain two alanine racemase isozymes – one that is constitutively expressed (Alr) and one that is inducible by the presence of L-alanine (DadX)[140,141] – Alr knockout appears to be lethal to bacteria in the absence of an external source of D-alanine.[142] Several mechanism-based inhibitors of Alr have been reported, including the reactive alanine analogues β-fluoroalanine, β-chloroalanine, β-trifluoroalanine and alanine phosphonate. However, lack of target specificity prevents their therapeutic use as antibacterials. DCS is currently the only clinically marketed drug targeting Alr,[31] though it is reserved as a second-line defense against resistant tuberculosis due to severe neurological side-effects.[143-145]

Crystal structures of Alr from various species have revealed a highly conserved Alr active site containing a wide entryway.[146] The structure of the *Geobacillus stearothermophilus* Alr in complex with DCS helped to unravel its mechanism of inhibition, whereby DCS first binds to the PLP cofactor through transamination to generate the external aldimine (Figure 6.3a).[147] Following a 1,3-prototropic shift, the resulting ketimine intermediate tautomerizes to form a stable isoxazole ring which is covalently attached to the PLP cofactor (Figure 6.3b). The aromatic PLP-derivative remains tightly bound within enzyme active sites, though extensive dialysis against PLP at neutral pH can eventually re-activate the enzyme.[148,149]

DCS acts synergistically with other peptidoglycan inhibitors known to target Alr (ex. β-chloroalanine)[150] and is known to also inhibit other key enzymes in peptidoglycan biosynthesis. These include D-ala-D-ala ligase (Ddl),[151] which catalyzes the formation of D-alanine dipeptides, and D-amino acid aminotransferase (DAT), a PLP-DE which converts D-alanine into D-glutamate.[148,149] DCS inhibits DAT in the same manner as for Alr, through the formation of a stable covalent PLP-adduct which binds the enzyme tightly.[148] Other PLP-DEs reported to be inhibited by DCS include dialkylglycine decarboxylase,[152] a unique enzyme which performs decarboxylative transamination of its substrate, as well as serine palmitoyltransferase,[153] in which DCS undergoes decarboxylative ring-opening to modify and thereby inhibit the enzyme irreversibly.
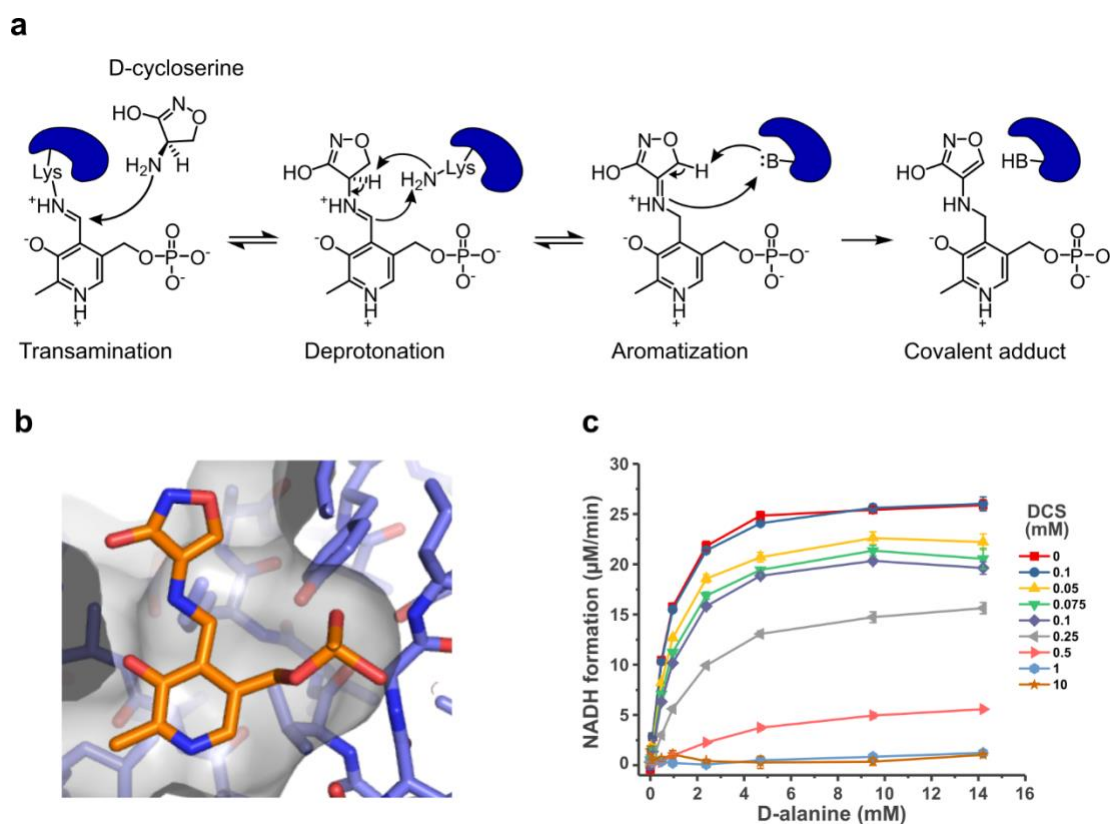
Figure 6.3 Mechanism of D-cycloserine inhibition of Alr. (**a**) Mechanism of DCS inhibition by formation of a stable aromatic PLP adduct within enzyme pockets. (**b**) Crystal structure of *Geobacillus stearothermophilus* Alr bound to the DCS-PLP inhibitor complex (PDB: 1EPV).[147] (**c**) Kinetics of alanine racemization (D- to L-Ala direction) by *S. aureus* Alr (0.01 µM) in the presence of DCS (*n* = 3, error: mean ± standard deviation).

Since DCS is promiscuous in its inhibition of PLP-DEs, we attempted to evaluate the complete set of DCS off-targets directly in *S. aureus* cells using our chemical proteomic method. We first confirmed DCS inhibition in our hands to match literature reports,[154,155] observing concentration-dependent inhibition of DCS on Alr activity *in vitro* (Figure 6.3c) and a minimum inhibitory concentration (MIC) of 0.5 mM in *S. aureus*. To screen for other PLP-DEs inhibited by DCS, we designed a competitive proteomic experiment in which *S. aureus* USA300 TnPdxS was incubated with DCS (10 or 1 mM) for 30 min prior to conducting our 2h labeling protocol with **PL4**. Analysis of non-treated *versus* DCS-treated samples by volcano plot revealed a handful of enzymes significantly influenced by DCS in a concentration-dependent manner (Figure 6.4, Table 6.2). Among these was the known target DAT, as well as two previously undescribed PLP-DE targets: ornithine/lysine/arginine decarboxylase (OKR-DC) and diaminopimelate decarboxylase (DAP-DC). OKR-DC catalyzes the decarboxylation of amino acid substrates to generate putrescine, agmatine and cadaverine. These amine building blocks are used for polyamine (spermine and spermidine) biosynthesis, which are important for cell growth and bacterial virulence.[156] DAP-DC decarboxylates meso-diaminopimelate to the amino acid lysine, which is important for protein and peptidoglycan biosynthesis. Lysine biosynthesis is recognized as an antibacterial target since it is essential but absent in mammals.[157]

Interestingly, Alr labeling was not significantly outcompeted by DCS under these conditions. While this is in disagreement with several previous *in vitro* studies,[138,155] the known *in vitro* substrate promiscuity of PLP-DEs may account for the observed reaction with DCS.[158,159] In support of this conclusion, recent metabolomics studies in *mycobacteria* have reported that Alr is only weakly inhibited by DCS in a cellular context, while Ddl is the primary lethal target in cells.[111,160] We were surprised to significantly identify Ddl in our experiment, which, despite being a known target of DCS, is not known to interact with PLP. Ddl catalyzes the formation of D-ala:D-ala dipeptides in two steps, in which D-ala first binds to a high-affinity site (N-terminal) and is phosphorylated at its carboxylate group using ATP.[161] Then, a second D-ala residue binds at the lower-affinity site (C-terminal) and undergoes nucleophilic attack on the phosphorylated D-ala to complete the coupling reaction. DCS inhibits Ddl competitively and reversibly, and is believed to bind to the high-affinity D-ala site.[162,163] Based on earlier proteomic data (*see* Figure 4.3b, Table 4.1), we find Ddl to be significantly enriched (medium-confidence), implying that it interacts with our **PL4P** probe and may thus bind PLP by extension. Although a direct role of cofactor binding to Ddl remains speculative, our data suggests that PLP and DCS compete for binding to Ddl. Interestingly, Ddl has been reported to be under PLP-based transcriptional regulation by the MocR/GabR-type protein DdlR in *Bacillus brevis*.[29] Although the exact mechanism of this regulator are not fully understood, the protein binds PLP as an internal aldimine and is able to form an external aldimine with the D-ala-D-ala dipeptide.
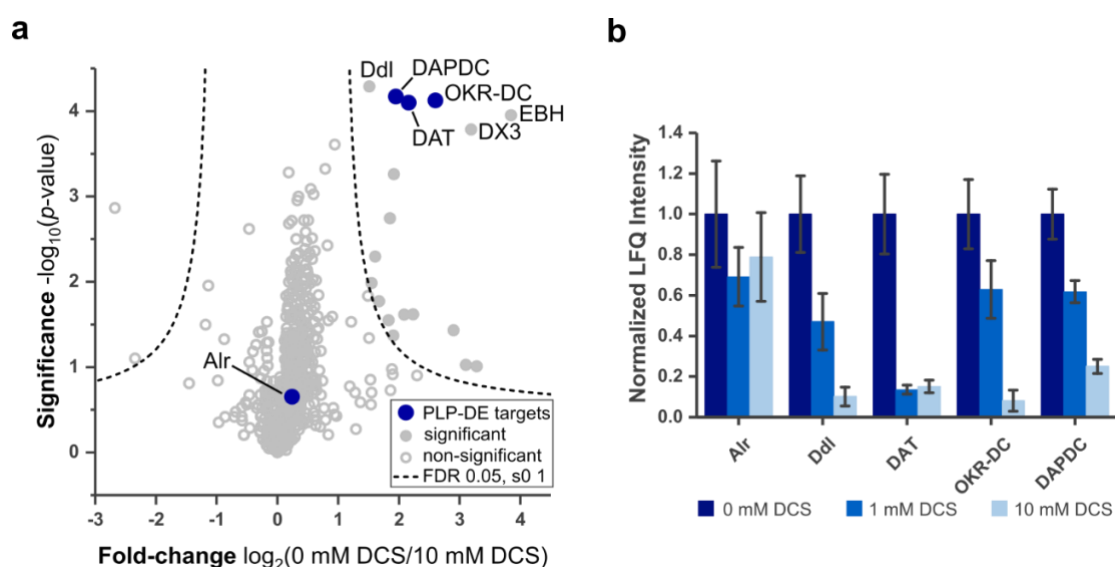
Figure 6.4 Competitive proteomic profiling reveals PLP-DE off-targets of D-cycloserine. (**a**) Volcano plot showing competition of **PL4P** binding by DCS (10 mM) (*n* = 3 biological replicates) in *S. aureus* USA300 TnPdxS. Proteins with highly significant fold-change between the two samples are labeled and remaining significant proteins are listed in Table 6.2. (**b**) Normalized LFQ intensities of select DCS targets from proteomic data (10 and 1 mM DCS). Mean raw LFQ intensities were normalized to DMSO (= 0) and max intensity (= 1) within individual protein sets (error bars: mean ± s.e.m.).

Table 6.2 Proteins significantly depleted by DCS treatment in *S. aureus* USA300 TnPdxS. Corresponding volcano plot in Figure 6.4a.

	Fold-change	Significance	Name	Uniprot ID
1.	1.95	4.17	**DAPDC**: Diaminopimelate decarboxylase (EC 4.1.1.20)	A0A0H2XFD9
2.	2.60	4.12	**OKR-DC**: Orn/Lys/Arg decarboxylase (EC 4.1.1.18)	A0A0H2XII6
3.	2.16	4.10	**DAT**: D-alanine aminotransferase (EC 2.6.1.21)	A0A0H2XHU6
4.	0.24	0.65	**Alr**: Alanine racemase (EC 5.1.1.1)	Q2FF55
5.	1.51	4.29	**Ddl**: D-alanine-D-alanine ligase (EC 6.3.2.4)	Q2FF43
6.	3.85	3.95	**EBH**: Extracellular matrix-binding protein (ECM-binding protein homologue)	Q2FH04
7.	3.28	1.01	Uncharacterized protein	A0A0H2XHI0
8.	3.19	3.78	**DX3**: Uncharacterized protein	A0A0H2XDX3
9.	3.10	1.02	PTS system, glucose-specific IIA component (EC 2.7.1.69)	A0A0H2XJ98
10.	2.90	1.43	Octanoyltransferase LipM (EC 2.3.1.181)	A0A0H2XFB3
11.	2.24	1.62	Uncharacterized protein	A0A0H2XE84
12.	2.09	1.61	Uncharacterized protein	A0A0H2XFE3
13.	1.92	3.26	Uncharacterized protein	A0A0H2XDZ0
14.	1.91	1.37	Deoxynucleoside kinase family protein (EC 2.7.1.113)	A0A0H2XGW9
15.	1.86	2.74	Glycosyl transferase, group 1 family protein (EC 2.4.1.-)	A0A0H2XIE0
16.	1.83	1.55	Hydrolase, alpha/beta hydrolase fold family (EC 3.4.11.5)	A0A0H2XGC7
17.	1.67	1.77	Uncharacterized protein	A0A0H2XIX9
18.	1.61	2.29	Uncharacterized protein	A0A0H2XEN1
19.	1.55	1.98	Carboxylesterase (EC 3.1.1.1)	A0A0H2XJL0

Taken together, our competitive PLP-DE profiling method enables the analysis of drug off-targets among a subset of cellular enzymes. Analyzing the targets of a drug directly under native cellular conditions provides a more accurate readout of its activity and selectivity in cells. Off-target reactivity of PLP-DE inhibitors has been a major challenge in the development of drugs for this enzyme class, particularly because inhibitors are often derived from amino acids and the active sites of PLP-DEs are often similar. PLP-DE inhibitors are therefore frequently associated with severe side-effects, which limits their therapeutic use. Thus, understanding inhibitor off-targets, especially in human cells, can help to fine-tune drug specificity and to improve safety.

6.3   PLPome profiling in Gram-negative bacteria

Discovering new therapeutic strategies against pathogenic bacteria is an urgent and unmet need in the face of antibiotic resistance. Gram-negative bacteria are particularly challenging to target due to their insensitivity to many commonly used antibiotics. Moreover, a recently reported subset of Gram-negative "ESKAPE" pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumanii, Pseudomonas aeruginosa, and Entero-bacter species*) cause the majority of nosocomial infections worldwide and are rapidly developing multi-drug resistance.[164] Drug-resistant bacteria are a major public health concern which is forcing clinicians to fall back on less safe drug options (ex. colistin) due to lack of available alternatives. Most antibiotics discovered during the golden era (1940-1960) stem from a limited set of molecular scaffolds whose useable lifespan has been maximized through synthetic derivatization (Figure 6.5). However, only few novel developments have arisen in recent years (since 2000). Strategies currently being pursued to aid antibiotic drug discovery and development include target-based discovery (high throughput screening, rational drug design), searching for novel molecular scaffolds by exploring untapped natural product resources and repurposing synthetic libraries for antibiotics.[165,166]   However, these platforms are hindered by multiple challenges associated with antibiotic development, such as cell permeability and unpredictable drug properties, high toxicity, and low return on cost.[165,167]
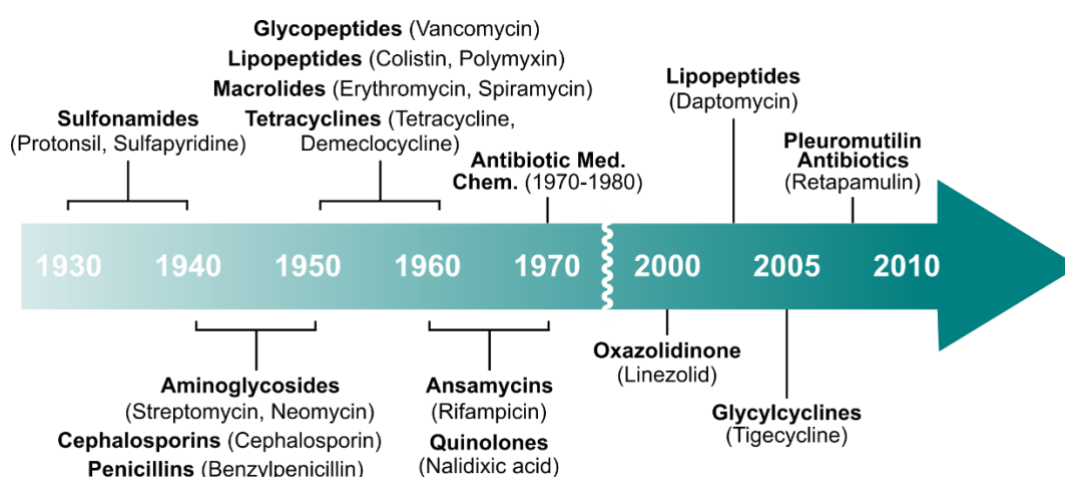
Figure 6.5 Timeline of antibiotic discoveries. Most modern antibiotics originated during the "Golden Era" of antibiotic discovery (1940-1960). This was proceeded by the rise of antibiotic medicinal chemistry (1970-1980), ie. synthetic derivatization to maximize the lifespan of antibiotics against the threat of bacterial resistance. Few drug candidates have been found since, resulting in a discovery void. Figure adapted from *G.D. Wright.*[168]

The ability of Gram-negative bacteria to keep out small molecules is particularly troublesome for antibiotic development. Bacteria of this type have a highly effective cellular barrier comprising two membranes, including a thin peptidoglycan layer in the periplasmic space, as well as a lipo-polysaccharide coat on the outer membrane (Figure 6.6).[169] Small molecules can enter through porin channels or by active transport, though can be actively extruded via efflux pumps once inside.[170] By comparison, Gram-positive bacteria contain only one cellular membrane, but are characterized by a thick peptidoglycan layer and are less restrictive to the entry of small molecules. Currently, only few targets of the thousands of natural antibiotics are known, and the ability of some to traverse the cell wall of Gram-negative bacteria is not understood. The idea of species-selective drug screening is re-emerging as a possible antibacterial strategy in order to broaden the chances of finding drugs active against at least a subset of bacteria.[165] Studying and understanding metabolic variability among bacteria can shed light on important differences when targeting diverse species.
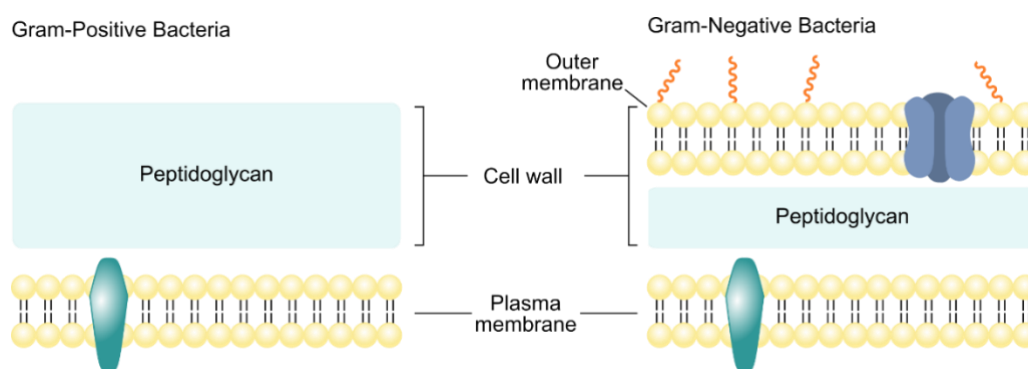
Figure 6.6 Cell walls of Gram-positive and Gram-negative bacteria. The cell wall of Gram-positive bacteria (ex. *S. aureus*) is composed of a thick peptidoglycan layer, which provides rigidity and support. Gram-negative bacteria (ex. *P. aeruginosa*) have a secondary outer membrane and a thinner peptidoglycan layer in the periplasmic space. The outer membrane is coated in lipopolysaccharide and porin channels allow selective entry to small molecules.

In order to uncover yet unknown metabolic enzymes important for cellular survival as potential drug targets, we envisaged applying our PLP-profiling strategy to Gram-negative bacteria. Our initial attempts focused on *Pseudomonas aeruginosa* (PAO1), a ubiquitous and opportunistic pathogen which is able to survive and adapt to a variety of nutrient sources by virtue of its metabolic diversity.[171] Since *P. aeruginosa* utilizes the DXP-dependent pathway of PLP biosynthesis found among members of the γ-subdivision of proteobacteria,[172] a transposon mutant of the pyridoxine phosphate (PNP) synthase enzyme (PdxJ) was instead used for labeling experiments to promote probe incorporation. PNP synthase condenses deoxyxylulose 5-phosphate (DXP) with 3-hydroxy-1-aminoacetone phosphate to make PNP (*see* Figure 3.1), which is further oxidized to PLP by PNP oxidases (PdxH). The *P. aeruginosa* TnPdxJ (strain: PW2400, ORF: PA0773) transposon mutant was obtained from the Manoil laboratory (University of Washington) and the insertion was validated by PCR according to literature methods (*see* Figure 8.2).[173] As per

our labeling protocol, the bacteria were grown in rich medium until stationary phase, upon which they were substituted into defined, PL-free MOPS medium[171] for the labeling reaction (100 or 10 µM **PL4**, 37°C, 2h). Following click conjugation and enrichment with a trifunctional rhodamine-azide-biotin linker, fluorescence SDS-PAGE revealed probe concentration-dependent protein labeling above the DMSO background. Upon LC-MS/MS measurement and volcano plot analysis, we found 17 significantly enriched PLP-DEs out of a total number of 40 PLP-DEs identified in our proteomic experiment using 100 µM of the probe (Figure 6.7a, Table 6.3). This represents 24% of the global PLPome (70 proteins) currently annotated in *P. aeruginosa* by gene ontology (GO-term: 0030170, PLP-binding), Uniprot (cofactor search) or EC classification. As expected, fewer proteins were significantly enriched (18 total) using 10 µM **PL4**, including 5 PLP-DEs (Figure 6.7b,c). Several aminotransferases were identified among the significant PLP-DEs from these experiments, including two aminotransferases involved in sugar biosynthesis. Overlap between the PLP-DE homologues identified from *P. aeruginosa* and *S. aureus* was modest and, in total, more PLP-DEs were identified in *P. aeruginosa*, though several annotations are putative. Furthermore, we were able to identify the *P. aeruginosa* homologue of the putative pyridoxal phosphate homeostasis protein (P24562), whose functional characterization remains incomplete though appears to be important throughout multiple species. Nonetheless, substantial background labeling and lack of significant enrichment of many PLP-DEs suggest that labeling conditions can be optimized. Thus, additional experiments are necessary in order to determine the ideal probe labeling time and bacterial growth conditions.
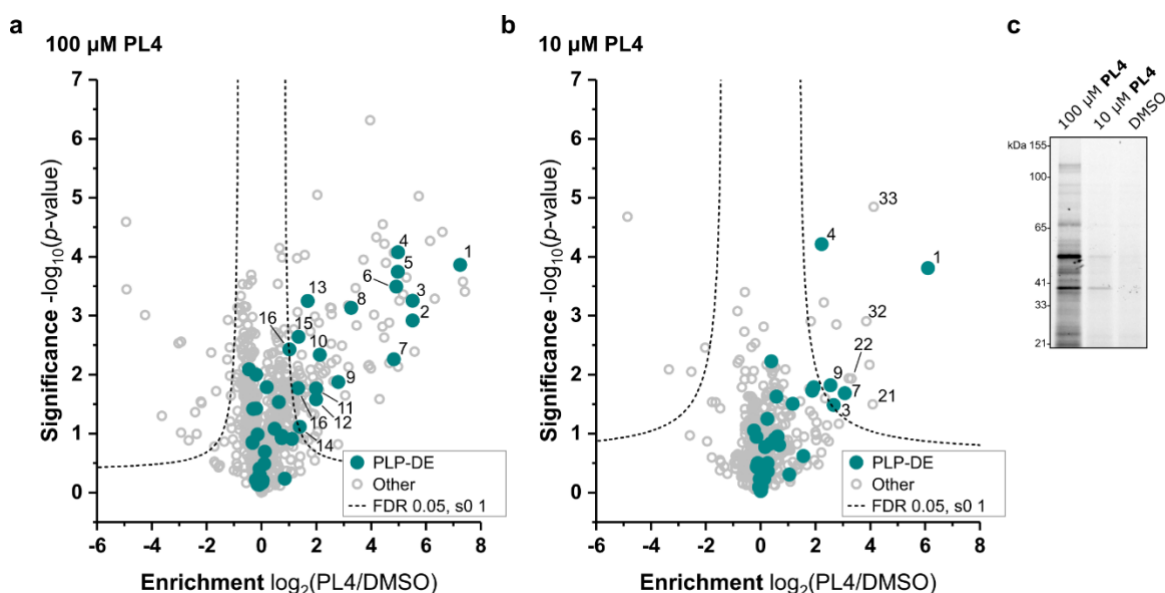
Figure 6.7 Chemical proteomics with PL4 in *P. aeruginosa* TnPdxJ. Volcano plots showing enrichment of proteins with 100 µM (**a**) or 10 µM (**b**) **PL4** (*n* = 3 biological replicates). Known PLP-DEs are highlighted with green dots and are numerically annotated according to Table 6.3. (**c**) Fluorescence SDS-PAGE showing **PL4** labeling during proteomic experiments upon avidin bead enrichment.

Table 6.3 Proteins significantly enriched using 100 µM PL4 in *P. aeruginosa* TnPdxJ cells. Corresponding volcano plot in Figure 6.7a. Green = PLP-DEs; Grey = other significant proteins (cut-offs: enrichment > 3, significance > 3).

No.	Fold-change	Significance	Name	Uniprot ID
1.	7.25	3.86	Probable aminotransferase	Q9HV83
2.	5.51	2.91	UDP-4-amino-4-deoxy-L-arabinose aminotransferase	Q9HY65
3.	5.51	3.25	Tryptophan synthase beta chain (EC 4.2.1.20)	P07345
4.	4.99	4.07	Biosynthetic arginine decarboxylase (ADC) (EC 4.1.1.19)	Q9HUX1
5.	4.98	3.74	L-seryl-tRNA(Sec) selenium transferase (EC 2.9.1.1) (Selenocysteine synthase)	Q9HV01
6.	4.92	3.49	Pyridoxal phosphate homeostasis protein (PLP homeostasis protein)	P24562
7.	4.83	2.26	Probable aminotransferase	Q9HXY0
8.	3.27	3.13	Branched-chain-amino-acid aminotransferase (BCAT) (EC 2.6.1.42)	O86428
9.	2.80	1.87	Aminotransferase (EC 2.6.1.-)	Q9HV76
10.	2.13	2.33	UDP-2-acetamido-2-deoxy-3-oxo-D-glucuronate aminotransferase (UDP-3-oxo-D-GlcNAcA aminotransferase) (EC 2.6.1.98)	Q9HZ76
11.	1.99	1.76	Probable cystathionine gamma-lyase	Q9I6A0
12.	1.99	1.58	O-succinylhomoserine sulfhydrylase (OSH sulfhydrylase) (EC 2.5.1.-)	P55218
13.	1.69	3.25	Aromatic-amino-acid aminotransferase (EC 2.6.1.57)	P43336
14.	1.40	1.11	Alanine racemase, catabolic (EC 5.1.1.1)	Q9HTQ2
15.	1.35	2.64	Uncharacterized protein	Q9HZI4
16.	1.34	1.77	Lysine-specific pyridoxal 5'-phosphate-dependent carboxylase, LdcA	Q9I2S7
17.	1.02	2.43	Ornithine decarboxylase	Q9HVQ3
18.	7.42	3.41	Deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase) (EC 3.6.1.23)	Q9HTN3
19.	7.35	3.57	Uncharacterized protein	Q9HX12
20.	6.60	4.42	Probable oxidoreductase	Q9I2R2
21.	6.32	3.29	Transcriptional regulatory protein AlgP (Alginate regulatory protein AlgR3)	P15276
22.	6.15	4.27	Nitrogen regulatory protein P-II 2	Q9HTR6
23.	5.74	5.03	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase (EC 6.3.2.10)	Q9HVZ7
24.	5.29	3.64	Aerotaxis transducer Aer2	Q9I6V6
25.	5.17	3.36	Transcriptional repressor NrdR	Q9HWX1
26.	5.04	3.25	Uncharacterized protein	Q9I2V2
27.	4.83	4.06	Probable short-chain dehydrogenase	Q9I2R7
28.	4.54	3.90	Probable pyrophosphohydrolase	Q9HW06
29.	4.49	4.22	Acylphosphatase (EC 3.6.1.7) (Acylphosphate phosphohydrolase)	Q9I504
30.	4.43	4.55	Periplasmic serine endoprotease DegP-like (EC 3.4.21.107)	G3XD20
31.	4.19	4.33	Uncharacterized protein	Q9HVV5
32.	4.11	3.01	Uncharacterized protein	Q9HTY8
33.	3.97	6.31	Glutathione peroxidase	Q9I5A2
34.	3.72	3.95	Ribonucleoside-diphosphate reductase (EC 1.17.4.1)	Q9I4I1
35.	3.43	3.47	Uncharacterized protein	Q9I6Z1
36.	3.17	3.17	Probable FAD-dependent monooxygenase	Q9HWJ1

To test whether probe uptake by *P. aeruginosa* was hampered by its effective cellular barrier, labeling experiments were also conducted in bacterial lysate using the phosphorylated probe. *P. aeruginosa* lysate was thus incubated with 10 or 1 µM of **PL4P** (45 min, 4°C), upon which the samples were reduced with NaBH₄, conjugated to the trifunctional linker, enriched and analyzed by LC-MS/MS (Figure 6.8a-c, Table 6.4). In this experiment, 19 PLP-DEs out of 49 total were significantly identified, representing 27% of the annotated *P. aeruginosa* PLPome. As a whole, similar PLP-DEs were identified in comparison to labeling in live cells, with some observable shifts in significance. Notably, the PLP homeostasis protein distinguishes itself from the background in terms of significance, implying considerable binding to the probe over other proteins. This is in agreement with its presumed function of regulating cellular PLP levels by scavenging the reactive aldehyde or passing the cofactor to other proteins. While optimizations are required to improve labeling over background (labeling time, number of replicates) or filtering as detailed in Chapter 4 may be necessary, this work presents a proof-of-concept indicating that **PL4P**-labeling can be performed in bacterial lysates. This strategy thus represents a promising alternative in case a PLP-biosynthesis mutant strain is not available, or if PL-probe uptake limits the use of the labeling method in live cells. Establishing the method for use in Gram-negative bacteria enables comparative analysis of different strains, and may serve to identify essential metabolic enzymes common to certain bacterial subsets as putative drug targets.
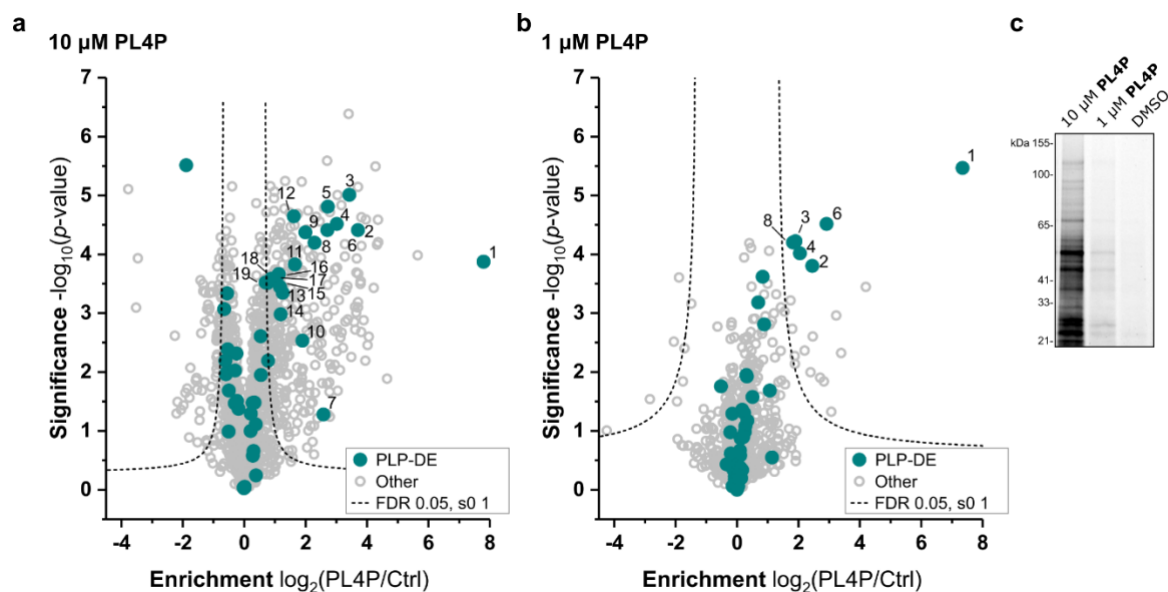
Figure 6.8 Chemical proteomics with PL4P in *P. aeruginosa* lysate. Volcano plots showing enrichment of proteins with (**a**) 10 μM or (**b**) 1 μM **PL4P** (*n* = 3 biological replicates). Known PLP-DEs are highlighted with green dots and are numerically annotated according to Table 6.4. (**c**) Fluorescence SDS-PAGE showing **PL4P** labeling in lysate during proteomic experiments upon avidin bead enrichment.

Table 6.4 Proteins significantly enriched using 10 μM PL4P in *P. aeruginosa* lysate. Corresponding volcano plot in Figure 6.8a. Green = PLP-DEs; Grey = other significant proteins (cut-offs: enrichment > 3.5, significance > 3.5).

No.	Fold-change	Significance	Name	Uniprot ID
1.	7.80	3.87	Pyridoxal phosphate homeostasis protein (PLP homeostasis protein)	P24562
2.	3.71	4.41	Probable aminotransferase	Q9HXY0
3.	3.42	5.01	Tryptophan synthase beta chain (EC 4.2.1.20)	P07345
4.	3.02	4.51	Probable transcriptional regulator	Q9HZV0
5.	2.73	4.81	Probable aminotransferase	Q9HV83
6.	2.71	4.41	L-seryl-tRNA(Sec) selenium transferase (EC 2.9.1.1) (Selenocysteine synthase)	Q9HV01
7.	2.58	1.28	Probable aminotransferase	Q9I371
8.	2.29	4.20	UDP-2-acetamido-2-deoxy-3-oxo-D-glucuronate aminotransferase (UDP-3-oxo-D-GlcNAcA aminotransferase) (EC 2.6.1.98)	Q9HZ76
9.	1.99	4.37	Serine hydroxymethyltransferase 3 (SHMT 3) (EC 2.1.2.1)	Q9HVI7
10.	1.90	2.53	Serine hydroxymethyltransferase 2 (SHMT 2) (EC 2.1.2.1)	Q9I138
11.	1.65	3.83	Biosynthetic arginine decarboxylase (ADC) (EC 4.1.1.19)	Q9HUX1
12.	1.62	4.64	Ornithine decarboxylase	Q9HVQ3
13.	1.26	3.35	Uncharacterized protein	Q9HZI4
14.	1.20	2.98	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase (EC 2.6.1.87)	Q9HY65
15.	1.17	3.45	Probable transcriptional regulator	Q9I283
16.	1.13	3.66	Probable aminotransferase	Q9I015
17.	1.04	3.57	Aspartate aminotransferase (AspAT) (EC 2.6.1.1)	P72173
18.	0.88	3.58	Probable cystathionine gamma-lyase	Q9I6A0
19.	0.76	3.51	Aminotransferase (EC 2.6.1.-)	Q9HV76
20.	5.66	3.98	Potential phenazine-modifying enzyme	Q9I781
21.	4.37	4.20	Uncharacterized protein	Q9I585
22.	4.36	4.59	Signal recognition particle receptor FtsY (SRP receptor)	Q9I6C1
23.	4.33	4.16	Probable short-chain dehydrogenase	Q9I3P2
24.	4.28	5.49	Acetolactate synthase isozyme III small subunit	Q9HVA1
25.	3.97	4.59	Uncharacterized protein	Q9HX12
26.	3.82	4.63	Uncharacterized protein	Q9I2B4
27.	3.82	4.14	Nonspecific ribonucleoside hydrolase	Q9I6Y9
28.	3.79	3.54	Phosphopantetheine adenylyltransferase (EC 2.7.7.3) (Dephospho-CoA pyrophosphorylase) (Pantetheine-phosphate adenylyltransferase) (PPAT)	Q9I6D1
29.	3.71	3.67	NAD-dependent protein deacylase 1 (EC 3.5.1.-) (Regulatory protein SIR2 homologue 1)	Q9I4L0
30.	3.71	5.14	Uncharacterized protein	Q9HVT3
31.	3.67	3.93	3-oxoacyl-[acyl-carrier-protein] reductase FabG (EC 1.1.1.100) (3-ketoacyl-acyl carrier protein reductase)	O54438
32.	3.66	3.69	Nucleoid-associated protein PA1533	Q9I3I0

6.4 Profiling the Human PLPome

Humans rely on the dietary uptake of vitamin B_6 (PN, PM and PL), for which major sources include whole grain and nuts, vegetables, bananas, milk and certain types of meat. While non-phosphorylated B_6 vitamers are directly absorbed in the intestine by passive diffusion, the phosphorylated forms first require removal of the phosphate group by an intestinal alkaline phosphatase in order to pass cell membranes.[174] On entering cells, the human pyridoxal kinase (PDXK) phosphorylates the PL-derivatives to generate the bioactive PLP cofactor. PLP-DEs use the cofactor throughout essential metabolic processes, such as amino acid and aminosugar metabolism, as well as cofactor, nucleotide and lipid biosynthesis. Additionally, PLP-DEs are important for the production of biogenic amines in humans, such as the neurotransmitters of the central nervous system (CNS). These include histamine (histidine decarboxylase), serotonin (aromatic L-amino acid decarboxylase), GABA (glutamate decarboxylase) and dopamine (L-3,4-dihydroxyphenylalanine (DOPA) decarboxylase). Vitamin B_6 deficiency, due to diet, malabsorption (celiac disease), increased excretion (renal dialysis) or enzymatic mutations (ex. antiquitin), can cause neurological symptoms such as irritability and seizures.[175] On the other hand, vitamin B_6 abuse, i.e. excessive uptake of PN (9.6 g/day, 3 years), has resulted in modest to severe neurological and dermatological problems which were reversible upon discontinuation.[176,177]

Alterations in vitamin B_6 metabolism are associated with multiple human diseases and PLP-DE inhibitors are clinically prescribed for treating CNS neutrotransmitter imbalances as well as cancer.[31] For example, the PLP-DE GABA aminotransferase controls the levels of γ-aminobutyric acid (GABA), the major inhibitory neurotransmitter of the CNS, by catalyzing its transamination to succinic semialdehyde. Since low levels of GABA can cause convulsions and seizures,[178] which are symptoms of epilepsy, Parkinson's, Huntington's and Alzheimer's diseases, the GABA aminotransferase inhibitor vigabatrin is commonly prescribed.[179,180] However, vigabatrin is associated with severe irreversible side effects, such as loss of peripheral vision, the molecular mechanisms of which are unclear.[181] Similarly, the PLP-DE DOPA decarboxylase controls production of the two major neurotransmitters dopamine and serotonin through the decarboxylation of DOPA and L-5-HTP substrates, respectively. In the treatment of Parkinson's disease, involving the progressive loss of dopamine-producing cells in the midbrain, exogenous DOPA is co-administered with the DOPA decarboxylase inhibitor carbidopa.[31] This drug helps to increase DOPA levels at the site by preventing premature metabolism in the body prior to reaching the CNS, where it cannot pass the blood-brain barrier. Unfortunately, as with many hydrazine-based PLP-DE inhibitors, carbidopa also suffers from poor selectivity. Finally, though the role of vitamin B_6 in cancer is complex and unclear, PLP-dependent pathways appear to be affected by the metabolic rewiring of rapidly proliferating cells.[182] Cancer patients often have decreased levels of circulating B_6 vitamers,[183] due to an increased demand for PLP-dependent metabolism for

sustaining the anabolic needs of highly proliferating cells and for their adaption to adverse conditions during tumour progression.[184] For example, serine hydroxymethyltransferase has a central role in one-carbon metabolism and is thus required for increased DNA synthesis in rapidly proliferating cells.[185,186] Drugs that inhibit SHMT, such as mimosine, are therefore clinically relevant for the treatment of tumours.[187] Contrastingly, in some distinct cases, increased levels of circulating vitamin B_6 also correlated with reduced incidence of certain neoplasms.[188,189] Here, high intratumoural expression of PDXK was prognostic of improved disease outcome, perhaps due to increased activity of the antitumour immune response.[190]

Genomic comparison of eukaryotic and prokaryotic PLPomes has revealed that the proportion of PLP-DE-encoding genes decreases with increasing size and complexity of the genome.[4] These observations are characteristic of enzymes involved in basic metabolic pathways rather than specialized regulatory functions. Furthermore, this analysis also showed that a substantial fraction of identified PLP-DE genes (nearly 20%) remain functionally unclassified or only tentatively classified. Uncovering and functionally characterizing new roles for PLP in the cell using PLP-profiling tools can help to uncover novel metabolic pathways and identify new therapeutic targets. Furthermore, chemical probes can be used to analyze PLP-DE off-targets of known inhibitors to improve therapeutic safety, and to compare and understand PLP-dependent metabolic rewiring in cancer cells for identification of biomarkers or prognostic factors.

In order to access a range of useful experimental applications using our PLPome profiling method, we attempted to establish labeling in human cells. We reasoned that human cells would lend themselves well to probe uptake since they depend on vitamin B_6 absorption from the environment and do not possess PLP biosynthetic machinery. HeLa cells were therefore cultured in standard medium and subsequently labeled using 100 µM **PL4** in PBS. Initial gel-based analysis revealed significant fluorescence labeling and encouraged us to examine **PL4** labeling on a proteomics level. Samples of **PL4**-labeled HeLa cells were thus lysed and reduced using $NaBH_4$, followed by click conjugation to a trifunctional rhodamine-azide-biotin linker and enrichment using avidin beads. LC-MS/MS measurement of samples, data analysis and volcano plot representation of the fold enrichment *versus* significance revealed no distinct enrichment of PLP-DEs (Figure 6.9, Table 6.5). Only five PLP-DEs were identified throughout the experiment, including three that were significantly enriched, but dispersed within a cloud of background. Among the significantly identified PLP-DEs are ornithine aminotransferase, pyridoxal kinase and aspartate aminotransferase.
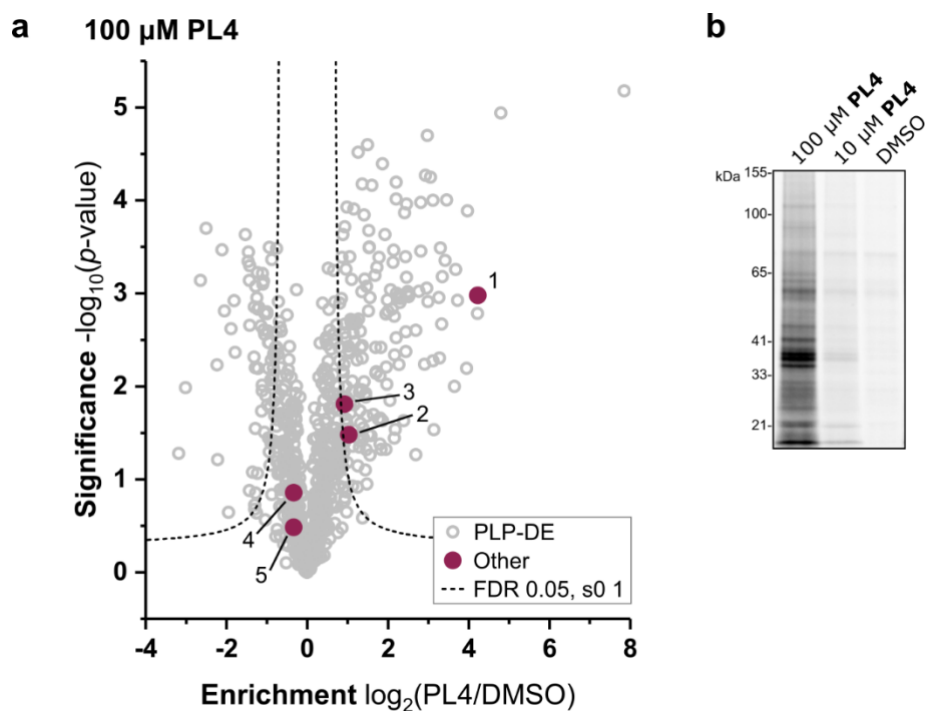
Figure 6.9 Chemical proteomics with PL4 in human cells (HeLa). (**a**) Volcano plot showing enrichment of proteins using 100 μM **PL4** (*n* = 3 biological replicates). Known PLP-DEs are highlighted with purple dots and are numerically annotated according to the table below. (**b**) Fluorescence SDS-PAGE showing **PL4** labeling during proteomic experiments upon CuAAC to a trifunctional rhodamine-biotin-azide tag and avidin bead enrichment.

Table 6.5 Proteins significantly enriched using 100 μM PL4 in HeLa cells. Corresponding volcano plot Figure 6.9.

	Fold-change	Significance	Name	Gene
1.	4.22	2.98	Ornithine aminotransferase, mitochondrial	OAT
2.	1.03	1.48	Pyridoxal kinase	PDXK
3.	0.93	1.81	Aspartate aminotransferase, cytoplasmic	GIG18, GOT1
4.	-0.33	0.86	Aspartate aminotransferase, mitochondrial	GOT2
5.	-0.34	0.48	Kynureninase	KYNU

We were surprised to find comparatively low enrichment of PLP-DEs in human cells *versus* bacteria. This result points towards significant differences in the ability of the human cells and bacteria to incorporate PLP analogues into their metabolic systems. Such differences may arise from poor cellular uptake, bioactivation or distribution, for which more complex mechanisms may be in place in human cells. It is also possible that human PLP-DEs have evolved a stronger dependence on the integrity of the PLP scaffold for cofactor binding, and cannot tolerate the structures of the PL-probe derivatives. Additional experiments aimed at optimizing the labeling conditions will be required for profiling the human PLPome. Here, it may be necessary to examine the underlying complexities that hinder the method from functioning effectively in human cells, which are not apparent in bacteria. Further optimizations and method developments can provide access to a variety of exciting applications for drug development and biological discovery.

7. Conclusion

This thesis describes the development of a chemical proteomic strategy for the study of PLP-DEs in live cells for diverse applications. We present a small library of probes as novel tools for profiling the family of PLP-DEs, which are easily accessible, compatible with biological systems and adaptable to suit diverse chemical biology needs.

The synthesis of a small PL-probe library involved core protecting group systems of the PL-scaffold and outlines strategies for the facile derivatization of the 2' position using diverse alkyl halides or nucleophilic groups. Importantly, the probes could be activated to their phosphorylated forms by *S. aureus* pyridoxal kinase and are also accommodated by diverse PLP-DEs. As exemplified by Alr, the additional bulk of the probes did not hinder catalytic activity, which could later be rationalized through crystallographic evidence. Our results indicate that multiple enzymes show a similar phenomenon, since two cofactor probes (**PL1** and **PL3**) were able to globally substitute for essential PLP-processes in order to support bacterial growth. Considering the evolution of PLP-DEs towards binding the specific cofactor structure, the ability of tool compounds to integrate into such diverse catalytic facets is astounding. Although the larger ethylene linker of **PL4** was not tolerated as a full cellular substitute of cofactor activity, it nonetheless appears to bind a multitude of proteins. Using additional probes, some structural preferences could be revealed among PLP-DEs, implying that different probe structures may be able to selectively target parts of the PLPome. A challenge here is maintaining the ability to undergo phosphorylation by PLK, which decreases as bulk of the substituents increases.

Our proteomic strategy capitalizes on the unique mode of cofactor binding to PLP-DEs in the form of an internal aldimine, which can be fixed by reductive amination. This circumvents the need for cumbersome and inherently non-specific photocrosslinkers or electrophilic traps, thus allowing for a more diversified and selective set of probes for reporting on natural PLP-binding events. Several binding controls (**PL2** and unphosphorylated probes) in analytical and proteomic experiments further highlight the specificity and authenticity of probe binding in the complex environment of the living cell. Labeling in bacteria directly cultured on the probes (**PL1** and **PL3**) as a PL supplement provided excellent selectivity and PLPome coverage. Alternatively, **PL4** identified many proteins upon a short, *in situ* labeling treatment (2h, 37°C), and results could be

strengthened using extended MS-based analysis. A two-step filtering strategy based on LFQ-intensity profile searches across multiple probe concentrations was used to cluster confidence and probe binding behaviour. We thus demonstrate the effectiveness with which supplementary analyses can help to distinguish specific and low-affinity binding from background and present general strategies for extracting additional information from proteomic datasets in order to gain deeper experimental insights.

Using our probes, we cumulatively labeled 73% of the current *S. aureus* PLPome and identified several more putative PLP binding events. This enabled the annotation of uncharacterized proteins Q2FF14 and GP0 as novel PLP-DEs, as well as IU9 and IC2 as putative PLP-DEs which require additional validation. A multifaceted exploration of the uncharacterized proteins Q2FF14 and GP0 incorporated *in silico*, *in vitro* and *in vivo* approaches. These studies provided preliminary functional insights into the PLP-related roles of these proteins, though necessitate further investigation. We anticipate that our tool compounds provide a resource for unraveling new PLP roles or yet undiscovered interactions in a cellular context. Additional applications of the method address challenges of drug development for PLP-DEs, several of which are recognized as drug targets, but are difficult to target selectively. A competitive proteomic screening platform was used to assess off-target reactivity of the clinical PLP-DE inhibitor DCS, thereby identifying two decarboxylases as potential targets while suggesting that the widely accepted target Alr is not significantly inhibited *in vivo* in *S. aureus*. Learning about the off-target mechanisms of PLP-DE inhibitors can help to pinpoint ways of improving current therapies and can be applied to other known drugs that are currently not approved due to safety concerns. Furthermore, *in situ* SAR profiling using a panel of structurally diverse probes can serve as a starting point for understanding active site geometries of different PLP-DEs and their selectivity.

A current limitation of our PLP-DE detection strategy is the requirement to promote probe uptake using a bacterial transposon mutant devoid of *de novo* PLP biosynthesis. However, this is allayed by the availability of many bacterial transposon mutant libraries, and the future prospect of labeling in bacterial lysates. In human cells, the lack of PLP biosynthetic machinery and dependence on PL salvage from the environment make it a suitable system for our labeling method, but further investigations are required to understand the underlying mechanisms currently preventing the method from functioning effectively. In addition, structural modifications imparted to the cofactor limit our access to enzymes with restricted binding sites, as evidenced by the inability of **PL4** to promote cellular growth. This challenge has been addressed through the use of a diversified set of probes, which has broadened proteome coverage, but nonetheless does not yet permit complete access to the PLPome.

Profiling the family of PLP-DEs provides an important foundation for the study of diverse topics in the field, including evolutionary relationships of PLP-DEs, cellular PLP homeostasis mechanisms, alternative roles of PLP and comparison of different organisms, cell types or cell states. Thus,

future investigations will aim to expand the method for application in different organisms. The ability to simultaneously monitor multiple PLP-DEs in their natural context and to identify novel family members without prior information has the potential to fill major gaps in knowledge. Nonetheless, full functional characterization of PLP-DEs upon their identification is difficult and requires combined biochemical, structural and bioinformatics approaches. Although significant literature precedence and flowcharts[107] for the functional characterization of PLP-DEs are useful for providing general insights, extensive experimental work is required to account for the diverse catalytic activities. Global cellular profiling techniques, such as proteomics and metabolomics, can further be used to understand cellular implications of PLP-DE transposon mutants in order to unravel their functions and activity-based metabolomics profiling can also help to identify unusual enzyme substrates.

By recent estimates, 30-40% of genes in all genomes remain unannotated or only putatively assigned a function and, of these, a significant fraction is predicted to encode functionally relevant proteins due to their phylogenetic conservation.[104-106] Genomics-based bioinformatic analysis can help to putatively annotate uncharacterized proteins, but can sometimes fail to identify protein similarity or results in misannotation due to low sequence identity. Current methods for directly detecting PLP binding in proteomes are either not quantitative or associated with many technical challenges, highlighting the need for supplementary chemical probe-based enrichment strategies for identifying PLP-binding in complex mixtures. Our methodology provides the first detailed proteomic perspective on the family of PLP-DEs and is concurrent with other recent endeavours in profiling post-translational modifications and cofactor-dependent families.[47,51,52,191-193] Effectively mining proteomes for important metabolic enzymes increases our knowledge of diverse protein families, thereby providing a basis for downstream drug development applications. These studies therefore contribute efficient measures for addressing the wealth of uncharacterized proteins through the exploration of complex biological systems.

8. Experimental procedures

8.1 Chemical Synthesis

General Methods. Reagents were purchased from Sigma-Aldrich, Thermo Fisher Scientific, Merck, TCI Europe, VWR International, and Alfa Aesar, and were used without further purification. All reactions involving air or water sensitive chemicals were carried out in oven-dried flasks under argon atmosphere. Dry solvents were purchased from Sigma-Aldrich and stored over molecular sieves under argon atmosphere. Flash chromatography and TLC (F254) analysis were performed using 60 Å silica gel from Merck (Darmstadt, Germany). TLC plates were visualized by staining with ceric ammonium molybdate (CAM) and phosphomolybdic acid (PMA) stains, or UV-absorption. HRMS spectra were acquired by ESI using a LTQ-FT Ultra mass spectrometer (Thermo Scientific). NMR spectra were recorded on Bruker AV-250, AV-360, AV-500, AVHD-300, AVHD-400, and AVHD-500 instruments, or an AVHD-500 spectrometer coupled to a cryo platform. The chemical shifts (δ) are reported in parts per million (ppm) and spectra are referenced to residual proton and carbon signals of the deuterated solvent. The following abbreviations are used to describe NMR coupling patterns: s – singlet; d – doublet; t – triplet; q – quartet; p – pentet; m – multiplet. The coupling constants, J, are reported in Hertz (Hz).

HPLC Purification. Compounds were purified by preparative, reversed-phase HPLC using a Waters 2545 quaternary gradient module equipped with a Waters 2998 photodiode array detector and fraction collector on a YMC Triart C18 column (250 × 10 mm, 5 μm). Gradients are listed in Table 8.1, using ddH_2O and HPLC-grade acetonitrile (no TFA) as the mobile phase.

Table 8.1 HPLC Gradients.

Time (min)	Method A		Method B		Time (min)	Method C	
	% H_2O	% ACN	% H_2O	% ACN		% H_2O	% ACN
0	98	2	98	2	0	98	2
1	98	2	98	2	12	2	98
30	50	50	80	20	14	2	98
35	2	98	2	98	15	98	2
37	2	98	2	98	17	98	2
40	98	2	98	2			
45	98	2	98	2			

General alkylation protocol. LDA was freshly prepared by adding diisopropylamine (1.5 ml, 10.7 mmol) and *n*-butyllithium (2.5 M in hexanes; 3.9 ml, 9.8 mmol) to a dry flask containing THF (30 ml) under argon at 0°C, and stirring for 15 min. After cooling to -78°C, starting material **13** (1.0 g, 4.4 mmol) dissolved in dry THF (5 ml) was added dropwise. The reaction was stirred for 1 h and then re-cooled to -78°C prior to the addition of alkyl halide (17.8 mmol). Upon stirring the reaction for 1 h at -78°C, the ice bath was removed and the reaction was allowed to warm to room temperature for 1 h. The reaction was quenched by adding satd. NH_4Cl (5 ml) at -78°C, and extra H_2O (30 ml) was added prior to separating the phases. The aqueous layer was extracted with EtOAc (3 x 50 ml), and the combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The product was purified by flash chromatography (5% MeOH/DCM).

Synthesis of PL1 and PL2

3-((tert-butyldimethylsilyl)oxy)-4,5-bis(((tert-butyldimethylsilyl)oxy)methyl)-2-methylpyridine
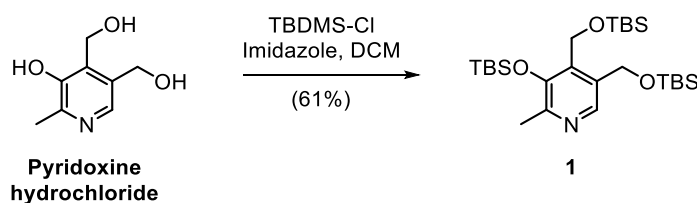
Pyridoxine hydrochloride (2.0 g, 9.7 mmol), imidazole (2.9 g, 43.6 mmol) and TBDMS-Cl (5.11 g, 34.0 mmol) were slurried in DCM (30 ml) and the resulting suspension was stirred overnight. The reaction was filtered to remove solids, washed with DCM and the filtrate was concentrated under

reduced pressure. The resulting oil was dissolved in Et$_2$O (50 ml) and acidified with 4 M HCl in dioxane (2.5 ml, 10 mmol) to yield the HCl salt as a white precipitate. Upon filtering and extensive washing with Et$_2$O, the solid was partitioned between EtOAc and satd. NaHCO$_3$. The organic phase was washed twice with satd. NaHCO$_3$, dried over Na$_2$SO$_4$ and concentrated under reduced pressure to yield **1** (3.0 g, 5.9 mmol, 61% yield) as a colourless oil. R$_f$ = 0.50 (10% EtOAc/Hex). ^{1}H NMR (360 MHz, CDCl$_3$) δ 8.30 (s, 1H), 4.87 (s, 2H), 4.71 (s, 2H), 2.44 (s, 3H), 1.04 (s, 9H), 0.93 (s, 9H), 0.87 (s, 9H), 0.12 (s, 6H), 0.09 (s, 6H), 0.04 (s, 6H); ^{13}C NMR (90 MHz, CDCl$_3$) δ 149.4, 146.3, 140.8, 135.7, 135.0, 61.3, 56.9, 26.1, 25.9, 25.9, 20.8, 18.6, 18.5, 18.3, -3.6, -5.2, -5.3; HRMS for C$_{26}$H$_{53}$NO$_3$Si$_3$ [M+H]$^+$ calcd. 512.3412, obtained 512.3404.

4,5-bis(((tert-butyldimethylsilyl)oxy)methyl)-2-(hydroxymethyl)pyridin-3-ol
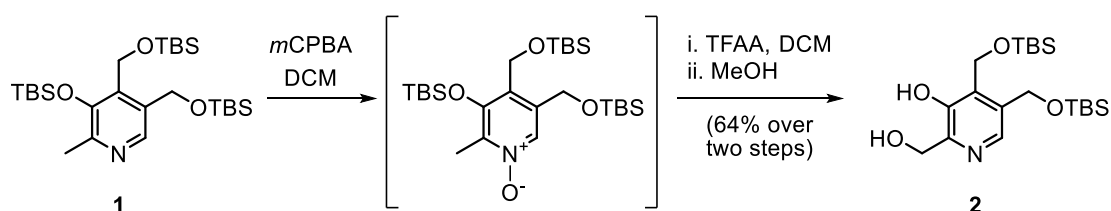
Following a modified procedure by Korytnyk *et al*,[54] *m*CPBA (1.0 g, 77%, 4.5 mmol) was added to a solution of **1** (2.3 g, 4.5 mmol) in DCM (50 ml) at 0°C, and the reaction was stirred until complete consumption of starting material was indicated by TLC (20% EtOAc/Hex). Upon removal of the solvent under reduced pressure, the resulting residue was dissolved in EtOAc and washed with satd. NaHCO$_3$ (3 × 50 ml). The organic phase was dried over Na$_2$SO$_4$ and concentrated to yield the intermediate *N*-oxide, which was used without further purification. Crude *N*-oxide was dissolved in DCM (50 ml) and cooled to 0°C prior to the slow addition of trifluoroacetic anhydride (1.3 ml, 9.0 mmol) via syringe for the Boekelheide rearrangement.[64] The reaction mixture was stirred overnight and allowed to warm to rt. Upon cooling to 0°C, MeOH (10 ml) was slowly added and the reaction was stirred at 0°C for 20 min and then allowed to warm up to rt while stirring for another 20 min. Following removal of solvents under reduced pressure, the residue was dissolved in EtOAc and washed with satd. NaHCO$_3$ (3 × 50 ml). The organic phase was dried over Na$_2$SO$_4$, filtered and concentrated. The crude product was purified by flash chromatography to yield **2** (1.2 g, 2.9 mmol, 64% yield over 2 steps) as a white solid. R$_f$ = 0.41 (50% EtOAc/Hex). ^{1}H NMR (360 MHz, CDCl$_3$) δ 9.16 (bs, 1H), 7.93 (s, 1H), 5.12 (s, 2H), 4.75 (s, 2H), 4.59 (s, 2H), 0.95 (s, 9H), 0.89 (s, 9H), 0.18 (s, 6H), 0.07 (s, 6H); ^{13}C NMR (90 MHz, CDCl$_3$) δ 150.1, 147.8, 138.3, 131.3, 129.2, 63.1, 61.7, 60.5, 26.0, 25.8, 18.3, 18.3, -5.2, -5.5; HRMS for C$_{20}$H$_{39}$NO$_4$Si$_2$ [M+H]$^+$ calcd. 414.2496, obtained 414.2482.

4,5-bis(((tert-butyldimethylsilyl)oxy)methyl)-3-hydroxypicolinaldehyde
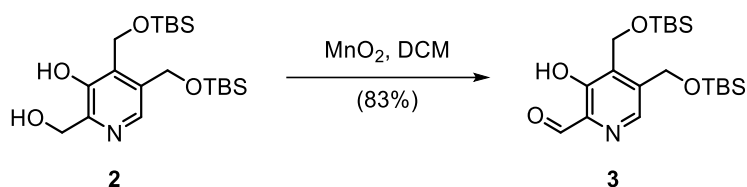
To a solution of alcohol **2** (1.2 g, 2.9 mmol) in DCM (10 ml) was added MnO_2 (0.7 g, 8.7 mmol), and the resulting slurry was stirred vigorously for approximately 3 h until complete consumption of starting material was indicated by TLC (20% EtOAc/Hex). If the reaction stalled, more MnO_2 was added. Once the oxidation was complete, the reaction mixture was filtered through celite and concentrated under reduced pressure to yield aldehyde **3** (1.0 g, 2.4 mmol, 83%) as a colourless oil which did not require further purification. R_f = 0.72 (20% EtOAc/Hex). ^{1}H NMR (360 MHz, CDCl$_3$) δ 11.03 (s, 1H), 10.04 (s, 1H), 8.59 (s, 1H), 5.00 (s, 2H), 4.81 (s, 2H), 0.96 (s, 9H), 0.89 (s, 9H), 0.13 (s, 6H), 0.09 (s, 6H); ^{13}C NMR (90 MHz, CDCl$_3$) δ 198.7, 155.6, 143.9, 141.6, 135.5, 133.6, 61.0, 54.6, 26.0, 26.0, 18.5, 18.4, -5.3, -5.3; HRMS for $C_{20}H_{37}NO_4Si_2$ [M+H]$^+$ calcd. 412.2339, obtained 412.2326.

4,5-bis(((tert-butyldimethylsilyl)oxy)methyl)-3-(methoxymethoxy)picolinaldehyde
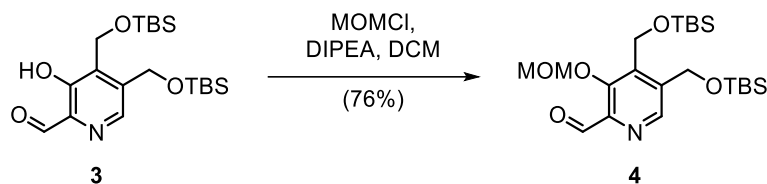
DIPEA (145 μl, 0.83 mmol) and MOMCl (86 μl, 1.13 mmol) were sequentially added to a solution of **3** (310 mg, 0.75 mmol) in DCM (5 ml) and the reaction mixture was stirred at room temperature for 2 h. Upon removal of the solvents under reduced pressure, the crude product was purified by flash chromatography to yield **4** (209 mg, 0.46 mmol, 76% yield) as a colourless oil which crystallizes upon freezing. R_f = 0.30 (20% EtOAc/Hex). ^{1}H NMR (360 MHz, CDCl$_3$) δ 10.17 (s, 1H), 8.82 (s, 1H), 5.10 (s, 2H), 4.99 (s, 2H), 4.84 (s, 2H), 3.57 (s, 3H), 0.95 (s, 9H), 0.89 (s, 9H), 0.12 (s, 6H), 0.10 (s, 6H); ^{13}C NMR (90 MHz, CDCl$_3$) δ 191.7, 152.4, 144.8, 143.9, 142.6, 140.3, 102.4, 61.0, 58.1, 56.1, 26.0, 25.9, 18.5, 18.3, -5.3, -5.4; HRMS for $C_{22}H_{41}NO_5Si_2$ [M+H]$^+$ calcd. 456.2602, obtained 456.2601.

4,5-bis(((tert-butyldimethylsilyl)oxy)methyl)-2-ethynyl-3-(methoxymethoxy)pyridine
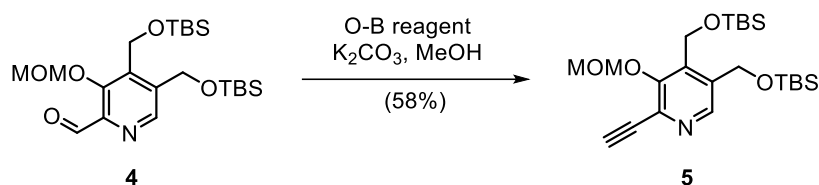
To a solution of **4** (209 mg, 0.46 mmol) in MeOH (5 ml) were added the Ohira-Bestmann reagent[65] (dimethyl-1-diazo-2-oxopropylphosphonate) (131 mg, 0.68 mmol) and solid K_2CO_3 (124 mg, 0.90 mmol), and the reaction was rapidly stirred for 16 h. EtOAc (20 ml) was then added to the reaction mixture and the organic phase was washed with H_2O (2 × 10 ml), dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was purified by flash chromatography to yield **5** (120 mg, 0.27 mmol, 58% yield) as a yellow solid. R_f = 0.47 (20% EtOAc/Hex). ^{1}H NMR (360 MHz, CDCl$_3$) δ 8.55 (s, 1H), 5.20 (s, 2H), 4.92 (s, 2H), 4.79 (s, 2H), 3.58 (s, 3H), 3.31 (s, 1H), 0.93 (s, 9H), 0.88 (s, 9H), 0.11 (s, 6H), 0.09 (s, 6H); ^{13}C NMR (90 MHz, CDCl$_3$) δ 153.1, 144.7, 138.8, 137.9, 135.3, 100.7, 80.8, 60.9, 58.1, 56.3, 26.0, 26.0, 25.9, 18.4, 18.3, -5.3, -5.4; HRMS for $C_{23}H_{41}NO_4Si_2$ [M+H]$^+$ calcd. 452.2652, obtained 452.2635.

(6-ethynyl-5-(methoxymethoxy)pyridine-3,4-diyl)dimethanol
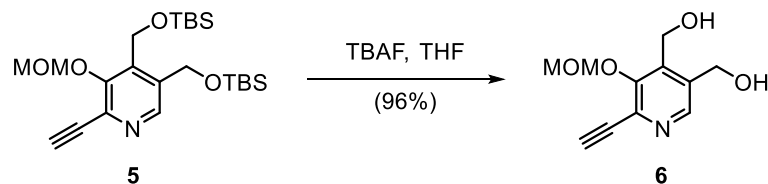
TBAF (514 μl, 1 M in THF, 514 mmol) was added to a solution of **5** (116 mg, 0.257 mmol) in THF (5 ml). The reaction was stirred for 10 min at rt and following removal of solvents under reduced pressure, the crude product was purified by flash chromatography to yield **6** (55 mg, 0.246 mmol, 96% yield) as a tan solid. R_f = 0.48 (10% MeOH/DCM). ^{1}H NMR (360 MHz, MeOD) δ 8.38 (s, 1H), 5.28 (s, 2H), 4.85 (s, 2H), 4.80 (s, 2H), 4.01 (s, 1H), 3.63 (s, 3H); ^{13}C NMR (90 MHz, MeOD) δ 155.5, 145.8, 142.8, 139.4, 137.0, 101.7, 84.1, 80.6, 60.4, 58.3, 55.7; HRMS for $C_{11}H_{13}NO_4$ [M+H]$^+$ calcd. 224.0923, obtained 224.0916.

2-ethynyl-5-(hydroxymethyl)-3-(methoxymethoxy)isonicotinaldehyde   and   6-ethynyl-4-(hydroxylmethyl)-5-(methoxymethoxy)nicotinaldehyde
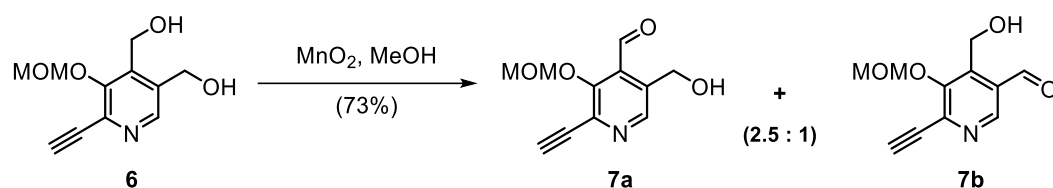
To a solution of **6** (30 mg, 0.13 mmol) in MeOH (3 ml) was added MnO_2 (34 mg, 0.39 mg), and the resulting slurry was stirred vigorously for 1 h. The reaction mixture was subsequently filtered through a 0.4 μM syringe filter and concentrated under reduced pressure. The product was dissolved in H_2O and the two isomers, **7a** and **7b**, were separated by HPLC purification using method A (**7a** t_R = 13.4 min, **7b** t_R = 14.0 min). R_f = 0.53 (10% MeOH/DCM). The product distribution was 2.5:1 and the respective final yields were **7a** (15 mg, 0.067 mmol, 52%) and **7b** (6 mg, 21%). **7a** ^{1}H NMR (500 MHz, MeOD) δ 8.22 (s, 1H), 6.56 (s, 1H), 5.59 (d, *J* = 6.4 Hz, 1H), 5.28 (d, *J* = 6.5 Hz, 1H), 5.23 (dd, *J* = 13.8 Hz, *J* = 1.2 Hz, 1H), 5.05 (d, *J* = 13.8 Hz, 1H), 3.92 (s, 1H), 3.55 (s, 3H); ^{13}C NMR (125 MHz, MeOD) δ 151.8, 140.0, 139.4, 138.3, 135.2, 101.1, 98.1, 83.6, 80.2, 70.6, 57.5; HRMS for $C_{11}H_{11}NO_4$ [M+H]+ calcd. 222.0766, obtained 222.0759. **7b** ^{1}H NMR (500 MHz, MeOD) δ 8.27 (s, 1H), 6.48 (d, *J* = 2.1 Hz, 1H), 5.33 (dd, *J* = 14.2, 2.2 Hz, 1H), 5.23 (s, 2H), 5.16 (d, *J* = 14.2 Hz, 1H), 3.96 (s, 1H), 3.53 (s, 3H); ^{13}C NMR (125 MHz, MeOD) δ 151.4, 140.9, 140.0, 139.8, 135.7, 101.1, 98.1, 84.1, 80.0, 71.0, 57.4; HRMS for $C_{11}H_{11}NO_4$ [M+H]+ calcd. 222.0766, obtained 222.0759.

PL1
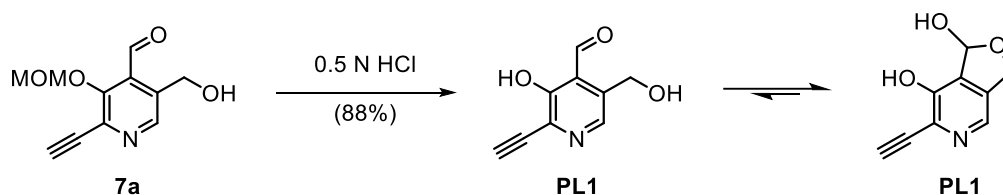
7a (20.7 mg, 0.094 mmol) was dissolved in 0.5 N HCl and stirred at rt for 15 min. The solution was diluted by half with H_2O and purified by HPLC using method B (t_R = 8.6 min). Upon lyophilization, the **PL1** hydrochloride salt (17.5 mg, 0.082 mmol, 88% yield) was obtained as a light yellow, fluffy solid. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.05 (s, 1H), 6.43 (s, 1H), 5.08 (d, *J* = 13.7 Hz, 1H), 4.91 (d, *J* = 13.7 Hz, 1H), 4.29 (s, 1H); ^{13}C NMR (125 MHz, DMSO-d_6) δ 150.6, 137.8, 134.5, 134.5, 129.5, 98.5, 83.7, 80.4, 69.2; HRMS for $C_9H_7NO_3$ [M+H]+ calcd. 178.0504, obtained 178.0498.

PL2
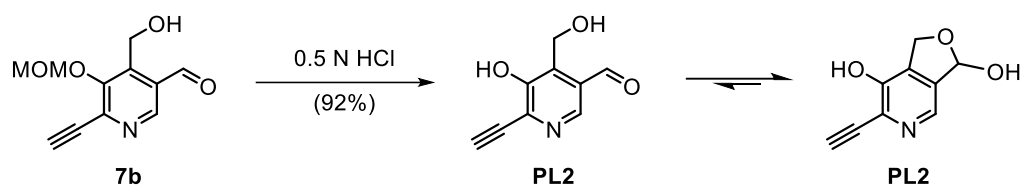
7b (12.3 mg, 0.056 mmol) was dissolved in 0.5 N HCl and stirred at rt for 15 min. The solution was diluted by half with H_2O and purified by HPLC using method B (t_R = 9.1 min). Upon lyophilization, the **PL2** hydrochloride salt (10.9 mg, 0.051 mmol, 92% yield) was obtained as a light yellow, fluffy solid. ^{1}H NMR (500 MHz, MeOD) δ 8.09 (s, 1H), 6.49 (d, *J* = 2.3 Hz, 1H), 5.18 (dd, *J* = 14.3, 2.3 Hz, 1H), 5.01 (d, *J* = 14.3 Hz, 1H), 3.94 (s, 1H); ^{13}C NMR (126 MHz, MeOD) δ 152.1, 139.5, 137.8, 136.4, 130.7, 101.8, 84.7, 79.3, 70.4; HRMS for $C_9H_7NO_3$ [M+H]$^+$ calcd. 178.0504, obtained 178.0498.

Synthesis of PL3

(2,2,8-trimethyl-4H-[1,3]dioxino[4,5-c]pyridin-5-yl)methanol
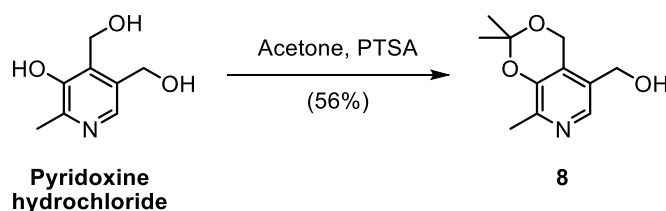
Acetal protection of pyridoxine was performed using a previously described protocol by Yang *et al.*[66] To a suspension of pyridoxine hydrochloride (5.0 g, 24.3 mmol) in dry acetone (100 ml) were added 2,2-dimethoxypropane (75 ml) and *p*-toluenesulfonic acid (18.5 g, 97.2 mmol). The mixture was stirred overnight upon which the dark-brown solution was neutralized with satd. $NaHCO_3$ until light yellow. The aqueous layer was extracted with EtOAc (3 × 100 ml) and the combined organic layers were washed with brine and concentrated under reduced pressure until crystalline. The crude product was recrystallized from EtOH/Et$_2$O (40:60) to yield pure **8** (2.8 g, 13.4 mmol) as a white crystalline solid. R_f = 0.29 (10% MeOH/DCM). ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.00 (s, 1H), 4.95 (s, 2H), 4.62 (s, 2H), 2.45 (s, 3H), 1.57 (s, 6H). ^{13}C NMR (91 MHz, CDCl$_3$) δ 147.6, 146.2, 138.6, 129.8, 126.1, 99.9, 60.0, 58.7, 24.9, 18.2. HRMS for $C_{11}H_{15}NO_3$ [M+H]$^+$ calcd. 210.1130, obtained 210.1123.

5-(((4-methoxybenzyl)oxy)methyl)-2,2,8-trimethyl-4H-[1,3]dioxino[4,5-c]pyridine
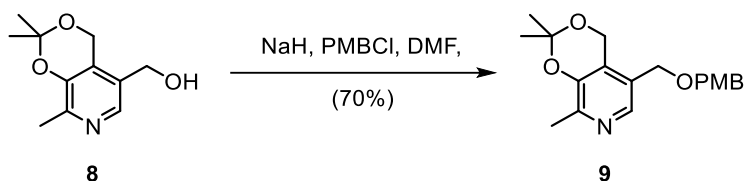
To a dry flask under argon was added NaH (60% dispersion in mineral oil; 1.9 g, 47.8 mmol) followed by dry DMF (20 ml). The suspension was heated to 70°C in an oil bath and **8** (2.0 g, 9.6 mmol) dissolved in DMF was slowly added. Once addition was complete, the reaction was left at 70°C for 5 min and then allowed to cool to room temperature. The reaction was cooled to 0°C for the slow addition of PMBCl (1.6 ml, 11.5 mmol) and was allowed to warm to room temperature while stirring overnight. Upon cooling on ice, the reaction was slowly quenched with H_2O (50 ml), extracted with EtOAc (3 × 50 ml), dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was purified by flash chromatography to yield **9** (2.2 g, 6.7 mmol, 70%) as a clear oil. R_f = 0.43 (10% MeOH/DCM). ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.94 (s, 1H), 7.24 (d, *J* = 8.6 Hz, 2H), 6.88 (d, *J* = 8.7 Hz, 2H), 4.85 (s, 2H), 4.42 (s, 2H), 4.39 (s, 2H), 3.80 (s, 3H), 2.40 (s, 3H), 1.54 (s, 6H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 159.4, 148.2, 146.1, 139.8, 129.6, 126.4, 126.1, 113.9, 99.7, 71.8, 67.1, 58.8, 55.3, 24.8, 18.6. HRMS for $C_{19}H_{23}NO_4$ $[M+H]^+$ calcd. 330.1705, obtained 330.1699.

(5-(((4-methoxybenzyl)oxy)methyl)-2,2-dimethyl-4H-[1,3]dioxino[4,5-c]pyridin-8-yl)methanol
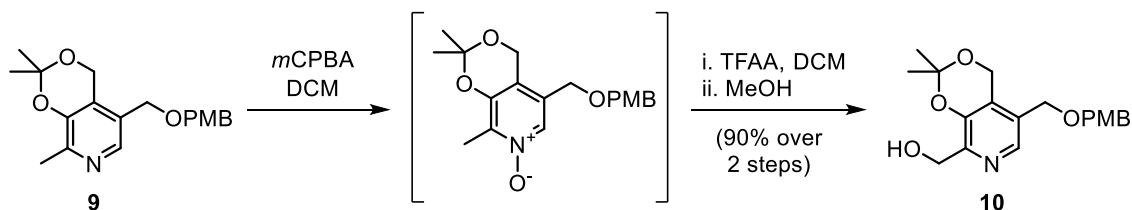
Following a modified procedure by Korytnyk *et al*,[54] *m*CPBA (0.8 g, 77%, 3.7 mmol) was added to a solution of **9** (1.0 g, 3.0 mmol) in DCM (50 ml) and the reaction was stirred overnight at room temperature. Upon removal of the solvent under reduced pressure, the resulting residue was dissolved in EtOAc and washed with satd. $NaHCO_3$ (3 × 30 ml). The organic phase was dried over Na_2SO_4 and concentrated to yield intermediate *N*-oxide, which was used directly in the next step. Crude *N*-oxide was dissolved in DCM (50 ml) and cooled to 0°C prior to the slow addition of trifluoroacetic anhydride (0.86 ml, 6.0 mmol) via syringe for the Boekelheide rearrangement.[64] The ice bath was then removed and the reaction was stirred for 4 h. Upon re-cooling to 0°C, MeOH (10 ml) was slowly added and the reaction was stirred at 0°C for 20 min and then allowed to warm up to rt while stirring for another 20 min. Following removal of solvents under reduced pressure, the residue was dissolved in EtOAc and washed with satd. $NaHCO_3$ (3 × 50 ml). The organic phase

was dried over Na_2SO_4, filtered and concentrated. The crude product was purified by flash chromatography to yield **10** (941 mg, 2.73 mmol, 90% yield over 2 steps) as an off-white crystalline solid. R_f = 0.24 (50% EtOAc/Hex). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.02 (s, 1H), 7.24 (d, *J* = 8.6 Hz, 2H), 6.89 (d, *J* = 8.7 Hz, 2H), 4.87 (s, 2H), 4.69 (s, 2H), 4.44 (s, 2H), 4.42 (s, 2H), 3.80 (s, 3H), 1.53 (s, 6H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 159.5, 147.7, 144.8, 139.1, 129.7, 129.5, 127.8, 126.6, 114.0, 100.2, 72.1, 67.1, 59.8, 58.8, 55.4, 24.8. HRMS for $C_{19}H_{23}NO_5$ [M+H]$^+$ calcd. 346.1655, obtained 346.1649.

8-(azidomethyl)-5-(((4-methoxybenzyl)oxy)methyl)-2,2-dimethyl-4H-[1,3]dioxino[4,5-c]pyridine
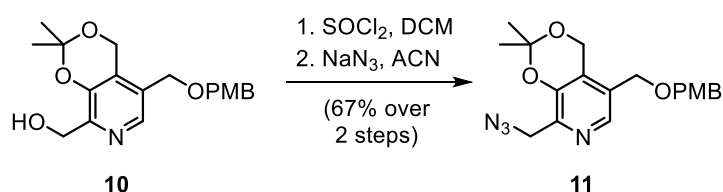
A solution of **10** (0.6 g, 1.7 mmol) in dry DCM (20 ml) was cooled to 0°C prior to the slow addition of $SOCl_2$ (1.3 ml, 17.4 mmol). Upon stirring at 0°C for 20 min and then at room temperature for 20 min, the reaction was quenched by adding satd. NaHCO$_3$ at 0°C until basic. The phases were separated and the organic layer was washed with satd. NaHCO$_3$ (20 ml) and brine prior to concentration under reduced pressure. The resulting chloride was used directly in the next reaction.

To a solution of crude intermediate chloride in ACN (20 ml) was added NaN$_3$ (108 mg, 1.66 mmol) and the reaction was refluxed for 16 h. The reaction was washed with brine, concentrated under reduced pressure and the crude product was purified by flash chromatography to yield **11** (430 mg, 1.16 mmol, 67%) as a clear oil. R_f = 0.20 (5% MeOH/DCM). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.05 (s, 1H), 7.25 (d, *J* = 8.7 Hz, 2H), 6.90 (d, *J* = 8.7 Hz, 2H), 4.88 (s, 2H), 4.46 (s, 2H), 4.43 (s, 2H), 4.42 (s, 2H), 3.80 (s, 3H), 1.57 (s, 6H). ^{13}C NMR (75 MHz, CDCl$_3$) δ 159.4, 146.4, 144.3, 140.2, 129.6, 129.4, 129.03, 127.4, 114.1, 113.9, 100.5, 72.2, 66.9, 58.7, 55.3, 50.8, 24.7. HRMS for $C_{19}H_{22}N_4O_4$ [M+H]$^+$ calcd. 371.1719, obtained 371.1711.

2-(azidomethyl)-3-hydroxy-5-(((4-methoxybenzyl)oxy)methyl)isonicotinaldehyde
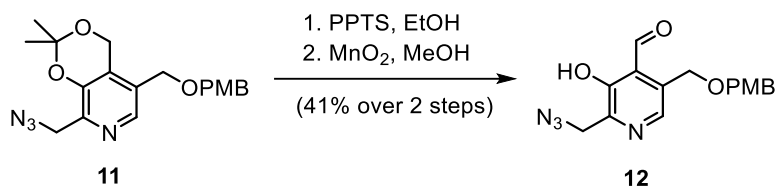
A solution of **11** (350 mg, 0.95 mmol) and PPTS (950 mg, 3.8 mmol) in EtOH (20 ml) was refluxed for 16 h. Upon completion, the reaction was cooled and the solvent was removed under reduced pressure. The residue was dissolved in EtOAc (20 ml) and washed with satd. $NaHCO_3$ (3 × 10 ml). The organic phase was concentrated and the crude diol was used directly in the next step.

To a solution of the intermediate alcohol in DCM (5 ml) was added MnO_2 (739 mg, 8.5 mmol). The reaction was stirred vigorously for 2 h until complete and was subsequently filtered through celite. The filtrate was evaporated and the resulting residue was purified flash chromatography to yield **12** (127 mg, 0.39 mmol, 41% over 2 steps) as a pale yellow oil. R_f = 0.52 (5% MeOH/DCM). ^{1}H NMR (400 MHz, $CDCl_3$) δ 11.49 (s, 1H), 10.40 (s, 1H), 8.17 (s, 1H), 7.25 (td, J = 8.7, 2.1 Hz, 2H), 6.90 (td, J = 8.7, 2.1 Hz, 2H), 4.75 (s, 2H), 4.56 (s, 2H), 4.54 (s, 2H), 3.82 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 197.0, 159.7, 153.7, 148.4, 140.4, 132.5, 129.7, 128.9, 121.7, 114.1, 72.6, 66.1, 55.3, 50.2. HRMS for $C_{16}H_{16}N_4O_4$ $[M+H]^+$ calcd. 329.1250, obtained 329.1244.

PL3
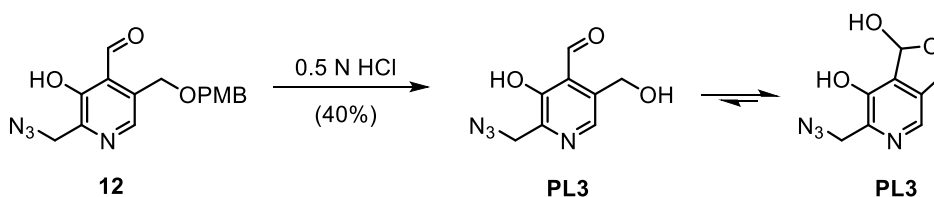
Aldehyde **12** (64 mg, 0.19 mmol) was dissolved in a mixture of 0.5 N HCl and acetone (1:1, 6 ml) and refluxed for 2 h. Upon completion, the reaction was cooled and the acetone was removed under reduced pressure. After washing the remaining solution with Et_2O (3 × 2 ml), the product was purified from the aqueous phase using HPLC method B (t_R = 12.0 min) and was lyophilized to generate the HCl salt of **PL3** (19 mg, 0.078 mmol, 40%) as a fluffy yellow solid. ^{1}H NMR (500 MHz, MeOD) δ 8.03 (s, 1H), 6.53 (d, J = 2.0 Hz, 1H), 5.22 (dd, J = 13.2, 2.0 Hz, 1H), 5.01 (d, J = 13.1 Hz, 1H), 4.53 (d, J = 13.3 Hz, 1H), 4.48 (d, J = 13.3 Hz, 1H). ^{13}C NMR (75 MHz, MeOD) δ 149.0, 144.2, 139.6, 136.0, 133.9, 100.4, 70.8, 51.3. HRMS for $C_8H_8N_4O_3$ $[M+H]^+$ calcd. 209.0675, obtained 209.0668.

Synthesis of PL4

1-methoxy-7-(methoxymethoxy)-6-methyl-1,3-dihydrofuro[3,4-c]pyridine

Pyridoxal hydrochloride

13
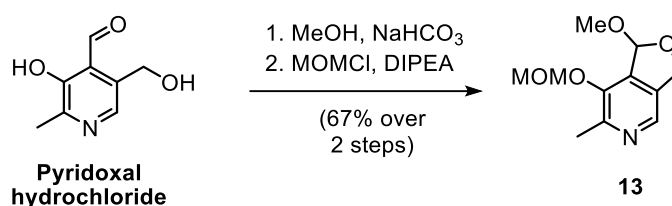
The monomethoxy acetal intermediate was synthesized according to a previously reported protocol by Kim & Jacobson.[194] A suspension of pyridoxal hydrochloride (8 g, 39.3 mmol) in MeOH (50 ml) was refluxed for 1 h. Upon cooling to room temperature, solid NaHCO$_3$ (3.3 g, 39.3 mmol) was added and the reaction was refluxed overnight. The mixture was filtered to remove precipitated NaCl and concentrated to yield crude intermediate acetal, which was used in the following reaction without further purification.

DIPEA (17.0 ml, 98.3 mmol) and MOMCl (4.5 ml, 59.0 mmol) were sequentially added to a solution of crude acetal (39.3 mmol) dissolved in anhydrous DCM (50 ml), and the reaction was stirred for 2 h at room temperature. Upon completion, the solution was concentrated under reduced pressure and the residue was purified by flash chromatography to yield **13** (5.94 g, 26.4 mmol, 67% over 2 steps) as a clear oil. R$_f$ = 0.50 (5% MeOH/DCM). ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.08 (s, 1H), 6.19 (d, J = 1.7 Hz, 1H), 5.36 (d, J = 6.5 Hz, 1H), 5.12 (ddt, J = 12.7, 1.8, 0.9 Hz, 1H), 5.02 (d, J = 6.5 Hz, 1H), 4.95 (dt, J = 12.6, 0.8 Hz, 1H), 3.45 (s, 3H), 3.39 (s, 3H), 2.48 (s, 3H); ^{13}C NMR (75 MHz, CDCl$_3$) δ 150.2, 146.2, 135.6, 135.4, 133.6, 106.1, 96.1, 70.0, 56.7, 54.7, 19.6; HRMS for C$_{11}$H$_{15}$NO$_4$ [M+H]$^+$ calcd. 226.1079, obtained 226.1072.

6-(but-3-yn-1-yl)-1-methoxy-7-(methoxymethoxy)-1,3-dihydrofuro[3,4-c]pyridine

13

14
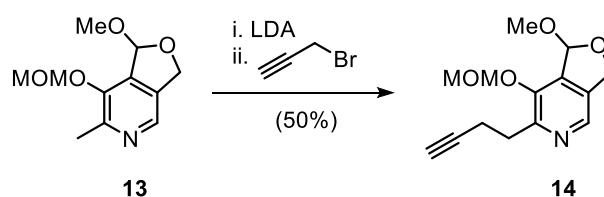
The general alkylation protocol was used to react **13** (1.0 g, 4.4 mmol) with propargyl bromide (80% in toluene; 2.0 ml, 17.8 mmol) in order to synthesize the alkylated product **14** (560 mg, 2.2 mmol, 50%) as a yellow oil. R$_f$ = 0.66 (5% MeOH/DCM). ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.19 (s, 1H), 6.26 (d, J = 1.7 Hz, 1H), 5.45 (d, J = 6.6 Hz, 1H), 5.18 (ddd, J = 12.8, 1.7, 0.9 Hz, 1H), 5.08 (d, J = 6.6 Hz, 1H), 5.02 (d, J = 12.8 Hz, 1H), 3.51 (s, 2H), 3.46 (s, 2H), 3.13 (t, J = 7.7 Hz, 2H), 2.68 – 2.62 (m, 2H), 1.93 (t, J = 2.7 Hz, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 151.2, 146.3, 136.3, 135.6, 133.8, 106.3, 96.3, 84.3, 70.1, 68.5, 57.0, 54.9, 31.5, 17.5; HRMS for C$_{14}$H$_{17}$NO$_4$ [M+H]$^+$ calcd. 264.1236, obtained 264.1228.

PL4
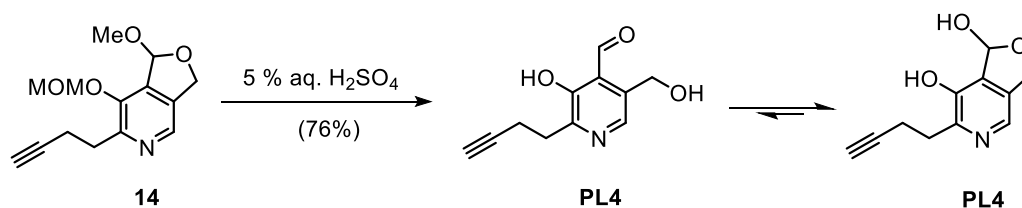
Compound **14** (40 mg, 0.15 mmol) was dissolved in 5% H_2SO_4 (2 ml) in acetone/H_2O (1:1) and heated at 85°C for 1 h. The reaction was allowed to cool and the acetone was removed under reduced pressure. The product was purified by HPLC using method B (t_R = 9.8 min) and lyophilized to generate **PL4** as the H_2SO_4 salt (35 mg, 0.11 mmol, 76%) as a fluffy white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.72 (s, 1H, OH), 7.98 (s, 1H), 6.67 (d, J = 7.7 Hz, 1H, OH), 6.42 (dd, J = 7.5, 1.8 Hz, 1H), 5.05 (ddd, J = 12.9, 1.9, 0.9 Hz, 1H), 4.86 (d, J = 13.1 Hz, 1H), 2.96 – 2.91 (m, 2H), 2.71 (t, J = 2.6 Hz, 1H), 2.57 – 2.50 (m, 1H); 13C NMR (100 MHz, DMSO-d_6) δ 146.6, 145.7, 135.5, 133.8, 132.5, 98.4, 84.5, 71.0, 68.8, 30.5, 16.5; HRMS for $C_{11}H_{11}NO_3$ $[M+H]^+$ calcd. 206.0817, obtained 206.0810.

Synthesis of PL5

1-Methoxy-7-(methoxymethoxy)-6-(7-(trimethylsilyl)hept-6-in-1-yl)-1,3-dihydrofuro[3,4-c]pyridine
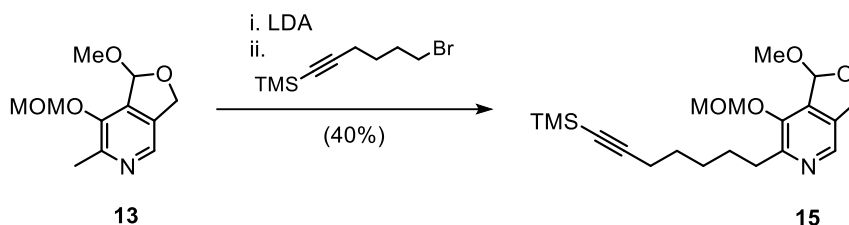
The general alkylation protocol was used to react **13** (200 mg, 0.89 mmol) with TMS-protected 7-bromohex-1-yne (620 mg, 2.67 mmol) in order to synthesize the alkylated product **15** (132 mg, 0.35 mmol, 40%) as a yellow oil. R_f = 0.67 (50% EtOAc/hexanes). ^{1}H NMR (400 MHz, CDCl₃) δ 8.17 (s, 1H), 6.25 (s, 1H), 5.42 (d, *J* = 6.5 Hz, 1H), 5.17 (d, *J* = 12.7 Hz, 1H), 5.06 (d, *J* = 6.5 Hz, 1H), 5.01 (d, *J* = 12.8 Hz, 1H), 3.51 (s, 3H), 3.45 (s, 3H), 2.89 (t, *J* = 7.8 Hz, 2H), 2.21 (t, *J* = 7.0 Hz, 2H), 1.71 (p, *J* = 7.6 Hz, 2H), 1.60-1.52 (m, 2H), 1.50-1.44 (m, 2H), 0.13 (s, 9H). ^{13}C NMR (100 MHz, CDCl₃) δ 153.8, 146.2, 135.8, 135.4, 134.0, 107.7, 106.3, 96.4, 84.4, 70.1, 57.0, 54.9, 32.5, 29.0, 28.7, 28.4, 19.9, 0.3. HRMS for $C_{20}H_{31}NO_4Si$ $[M+H]^+$ calcd. 378.2101, obtained 378.2094.

6-(hept-6-yn-1-yl)-1-methoxy-7-(methoxymethoxy)-1,3-dihydrofuro[3,4-c]-pyridin

15 → **16**
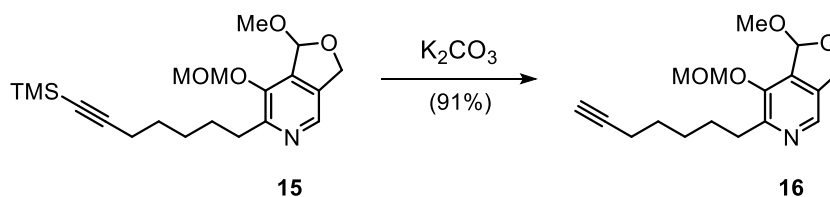
To a solution of **15** (241 mg, 0.63 mmol) dissolved in dry MeOH (10 ml) was added K_2CO_3 (441 mg, 3.2 mmol) and the reaction was stirred overnight at room temperature. Upon completion, Et_2O and H_2O were added (each 20 ml) and the aqueous layer was extracted with Et_2O (3 × 20 ml). The organic phase was concentrated under reduced pressure and the resulting residue was purified by flash chromatography to yield **16** (176 mg, 0.58 mmol, 91%) as a yellow oil. R_f = 0.62 (50% EtOAc/hexane). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.16 (s, 1H), 6.24 (s, 1H), 5.41 (d, J = 6.5 Hz, 1H), 5.16 (d, J = 13.0 Hz, 1H), 5.05 (d, J = 6.5 Hz, 1H), 5.00 (d, J = 13.0 Hz, 1H), 3.50 (s, 3H), 3.45 (s, 3H), 2.87 (t, J = 7.8 Hz, 2H), 2.18 (td, J = 7.0, 2.6 Hz, 2H), 1.91 (t, J = 2.7 Hz, 1H), 1.77-1.65 (m, 2H), 1.63-1.52 (m, 2H), 1.52-1.44 (m, 2H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 153.8, 146.1, 135.7, 135.7, 133.8, 106.3, 96.4, 84.8, 70.1, 68.3, 56.9, 54.9, 32.6, 28.9, 28.5, 28.4, 18.4. HRMS for $C_{17}H_{23}NO_4$ $[M+H]^+$ calcd. 306.1705, obtained 306.1701.

PL5

16 → **PL5** ⇌ **PL5**
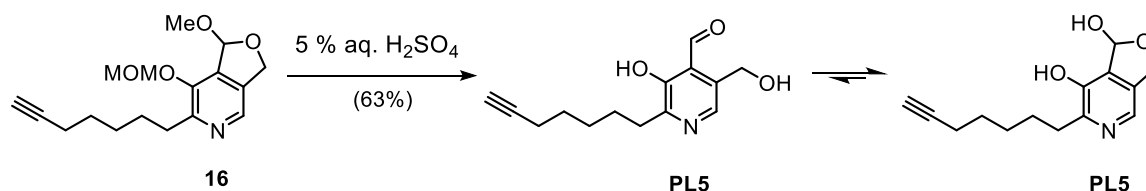
Compound **16** (100 mg, 0.372 mmol) was dissolved in 5% H_2SO_4 (2 ml) in acetone/H_2O (1:1) and heated at 85°C for 1 h. The reaction was allowed to cool and the acetone was removed under reduced pressure. The product was purified by HPLC using method A (t_R = 7.5 min) and lyophilized to generate **PL5** as the H_2SO_4 salt (80 mg, 0.232 mmol, 63%) as a fluffy white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.49 (s, 1H, OH), 7.94 (s, 1H), 6.63 (d, J = 7.1 Hz, 1H, OH), 6.41 (d, J = 7.1 Hz, 1H), 5.04 (d, J = 12.8 Hz, 1H), 4.85 (d, J = 12.8 Hz, 1H), 2.77 − 2.65 (m, 3H), 2.14 (td, J = 6.9, 2.6 Hz, 2H), 1.62 (p, J = 7.6 Hz, 2H), 1.55-1.43 (m, 2H), 1.43-1.34 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 148.9, 145.6, 134.9, 133.7, 132.4, 98.5, 84.6, 71.1, 31.2, 28.1, 27.9, 27.5, 17.6. HRMS for $C_{14}H_{18}NO_3$ $[M+H]^+$ calcd. 248.1287, obtained 248.1281.

Synthesis of PL6

1-Methoxy-7-(methoxymethoxy)-6-(4-((trimethylsilyl)ethinyl)phenethyl)-1,3-dihydrofuro[3,4-c]pyridin
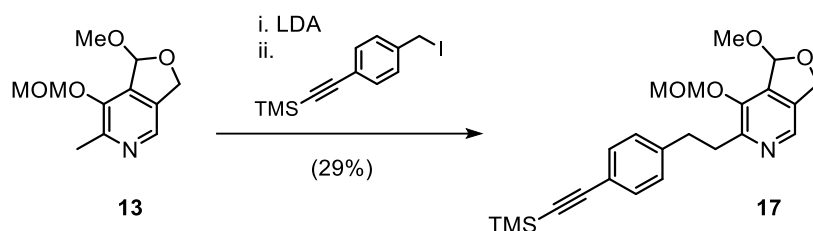
The general alkylation protocol was used to react **13** (200 mg, 0.89 mmol) with TMS-protected 1-ethynyl-4-(iodomethyl)benzene (839 mg, 2.67 mmol) in order to synthesize the alkylated product **17** (106 mg, 0.26 mmol, 29%) as a yellow oil. R_f = 0.63 (50% EtOAc/hexanes). ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.20 (s, 1H), 7.37 (d, J = 8.3 Hz, 2H), 7.16 (d, J = 8.3 Hz, 2H), 6.26 (s, 1H), 5.39 (d, J = 6.6 Hz, 1H), 5.19 (d, J = 13.0 Hz, 1H), 5.03 (d, J = 13.0 Hz, 1H), 5.01 (d, J = 6.6 Hz, 1H), 3.47 (s, 3H), 3.46 (s, 3H), 3.22-3.11 (m, 2H), 3.07-2.99 (m, 2H), 0.23 (s, 9H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 152.3, 146.5, 142.7, 136.4, 132.1, 128.5, 120.8, 106.3, 96.4, 93.7, 70.1, 57.0, 57.0, 55.0, 34.9, 33.9, 0.17. HRMS for C$_{23}$H$_{29}$NO$_4$Si [M+H]$^+$ calcd. 412.1944, obtained 412.1937.

6-(4-ethynylphenethyl)-1-methoxy-7-(methoxymethoxy)-1,3-dihydrofuro-[3,4-c]pyridin
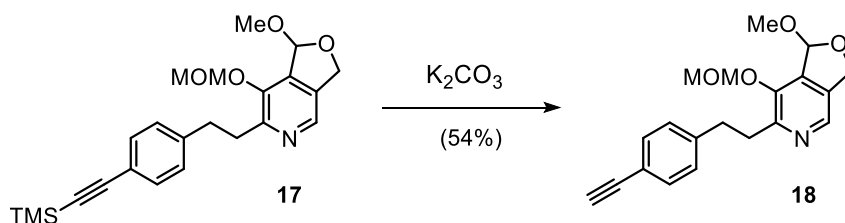
To a solution of **17** (48 mg, 0.17 mmol) dissolved in dry MeOH (10 ml) was added K$_2$CO$_3$ (23 mg, 0.17 mmol) and the reaction was stirred overnight at room temperature. Upon completion, Et$_2$O and H$_2$O were added (each 10 ml) and the aqueous layer was extracted with Et$_2$O (3 × 10 ml). The organic phase was concentrated under reduced pressure and the resulting residue was purified by flash chromatography to yield **18** (30 mg, 0.09 mmol, 54%) as a yellow oil. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.20 (s, 1H), 7.41 (d, J = 8.2 Hz, 2H), 7.19 (d, J = 8.2 Hz, 2H), 6.26 (s, 1H), 5.41 (d, J = 6.6 Hz, 1H), 5.20 (d, J = 12.9 Hz, 1H), 5.05 (d, J = 12.9 Hz, 1H), 5.02 (d, J = 6.6 Hz, 1H), 3.46 (s, 3H), 3.6 (s, 3H), 3.30-3.16 (m, 2H), 3.11-3.03 (m, 3H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 152.3, 146.3, 143.2, 136.1, 135.6, 133.9, 132.2, 128.6, 119.7, 106.3, 96.3, 83.9, 76.7, 70.1, 56.9, 54.9, 34.8, 34.1. HRMS for C$_{20}$H$_{21}$NO$_4$ [M+H]$^+$ calcd. 340.15499, obtained 340.1546.

PL6
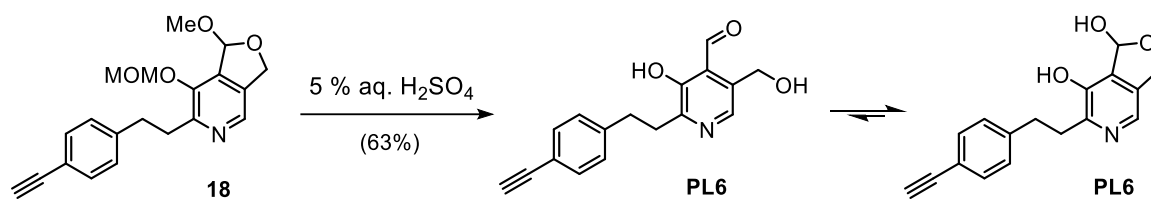
Compound **18** (30 mg, 0.088 mmol) was dissolved in 5% H_2SO_4 (2 ml) in acetone/H_2O (1:1) and heated at 85°C for 1 h. The reaction was allowed to cool and was concentrated under reduced pressure to remove acetone. The product was purified by HPLC using method C (t_R = 8.5 min) and lyophilized to generate **PL6** as the H_2SO_4 salt (21 mg, 0.0554 mmol, 63%) as a fluffy white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.65 (s, 1H, OH), 7.98 (s, 1H), 7.38 (d, J = 8.2 Hz, 2H), 7.26 (d, J = 8.2 Hz, 2H), 6.66 (d, J = 7.6 Hz, 1H, OH), 6.42 (d, J = 7.6 Hz, 1H), 5.05 (d, J = 12.9 Hz, 1H), 4.87 (d, J = 12.9 Hz, 1H), 4.10 (s, 1H), 3.08 – 2.91 (m, 4H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 147.6, 145.7, 143.2, 135.2, 133.8, 132.5, 131.6, 128.6, 119.1, 98.5, 83.6, 68.8, 33.5, 32.7. HRMS for $C_{17}H_{15}NO_3$ [M+H]$^+$ calcd. 282.1130, obtained 282.1123.

8.2   Biological Methods

Bacterial growth media. LB-medium. Peptone (10 g), NaCl (5 g) and yeast extract (5 g) were combined and dissolved in ddH$_2$O to a final volume of 1 L, and adjusted to pH 7.5. **B-medium.** Peptone (10 g), NaCl (5 g), yeast extract (5 g) and K$_2$HPO$_4$ (1 g) were combined and dissolved in ddH$_2$O to a final volume of 1 L, and adjusted to pH 7.5.

Cloning and overexpression of recombinant proteins. N-terminally STREP-II tagged *S. aureus* (Mu50 for Alr and PLK; USA300 for all others) proteins were cloned and expressed using the primers and conditions listed in Table 8.2. Cloning was performed using the Invitrogen Gateway cloning system with pDONR201^{Kan} as the donor vector and pET55dest^{Amp} as the destination vector. Proteins were expressed in *E. coli* BL21(DE3) grown in LB-media containing 0.1 mg/ml Ampicillin at 37°C to an OD$_{600}$ of 0.5-0.6 through the addition of 1 mM isopropyl-1-thio-β-galactopyranoside (IPTG). Bacteria were harvested and washed with PBS prior to cell lysis and protein purification.

Table 8.2 PCR primers and overexpression conditions for recombinant proteins.

No	Protein Name	Uniprot ID, *Gene name*	Primers	Over-expression	MW found (Da)
1.	(Alr) Alanine racemase 1	P63479, *Alr1*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt tcggataaatattatagat (rev) ggggaccactttgtacaagaaagctgggtgtt attttaacaattcatttg	30°C, o/n	m-Met = 45,099.1
2.	(AT1) Aminotransferase, Class I	A0A0H2XFY9, SAUSA300_2497	(fwd) ggggacaagtttgtacaaaaaagcaggcttt atctctaacaaattagcaaat (rev) ggggaccactttgtacaagaaagctgggtgtt attcatataaatatgctagtt	25°C, 6 h	m-Met = 45,024.0
3.	(AT2) Aminotransferase, Class I	A0A0H2XFA8, *SAUSA300_0952*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt aaactttctttaaattctaatt (rev) ggggaccactttgtacaagaaagctgggtgtt atttagtatttaaatatttag	25°C, o/n	m-Met = 45,435.4
4.	(DAT) D-alanine aminotransferase	A0A0H2XHU6, *dat*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt gaaaaaatttttttaaatggtgag (rev) ggggaccactttgtacaagaaagctgggtgtt aaatactgtgtgactctata	25°C, o/n	m-Met = 34,184.5
5.	(CS) Cysteine synthase	A0A0H2XFQ3, *cysK*	(fwd)ggggacaagtttgtacaaaaaagcaggcttt gcacaaaaaccagtagataatatta (rev) ggggaccactttgtacaagaaagctgggtgtt agtcatcgaatgaatataaaggt	37°C, 4 h	m-Met = 35,265.5
6.	(ODC) Ornithine/Lysine/ Arginine decarboxylase	A0A0H2XII6, *SAUSA300_0458*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt aagcaacctattttaaataaattag (rev) ggggaccactttgtacaagaaagctgggtgtt attcatcctcaactaaaatttta	25°C, o/n	m-Met = 53,372.2
7.	(HemQ) coprohaem decarboxylase	Q2FJ56, *SAUSA300_0569*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt agtcaagcagccgaaacatta (rev) ggggaccactttgtacaagaaagctgggtgtt aagaaatcgcaaagaattgatcg	25°C, o/n	m-Met = 31,679.7
8.	(PsuG) pseudouridine-5'-phosphate glycosidase	A0A0H2XJC6, *SAUSA300_0312*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt gcaaatttacaaaagtatattga (rev) ggggaccactttgtacaagaaagctgggtgct acaataatttattaacagcg	25°C, o/n	m-Met = 35,154.5
9.	(HemH) Ferrochelatase	A0A0H2XGS5, *SAUSA300_1782*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt actaaaaaaatgggattattagtta (rev) ggggaccactttgtacaagaaagctgggtgtt aaaatatagacttgatttcatca	25°C, o/n	m-Met = 37,346.6
10.	(Q2FF14) UPF0340 uncharacterized protein	Q2FF14, *SAUSA300_2068*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt aaagatttgacaatgttattag (rev) ggggaccactttgtacaagaaagctgggtgtt attggtatttagcacgtt	25°C, o/n	m-Met = 21,184.9
11.	(GP0) Uncharacterized protein	A0A0H2XGP0, *SAUSA300_2209*	(fwd) ggggacaagtttgtacaaaaaagcaggcttt agcatagttcagttatatga (rev) ggggaccactttgtacaagaaagctgggtgtt attcagggattgtttgat	20°C, o/n	m-Met = 37,052.9
12.	(HH8) Uncharacterized protein	A0A0H2XHH8, *SAUSA300_1082*	(fwd)ggggacaagtttgtacaaaaaagcaggcttt aaagataatttacaacaaatct (rev) ggggaccactttgtacaagaaagctgggtgtc actcctcttctcc	25°C, o/n	m-Met = 27,636.3

Cloning and overexpression of recombinant PLK.   The N-terminally STREP-II tagged *S. aureus* Mu50 pyridoxal kinase (PLK) was prepared as previously described.[75]

Analytical-scale protein purification (AT1, AT2, DAT, CS, OKR-DC, HemQ, PsuG, HemH, Q2FF14, GP0, HH8). Bacterial cell pellets from overexpression cultures (50 ml) were resuspended in 1 ml phosphate buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0). Cells were lysed by sonication and the lysate was clarified by centrifugation (21,000 x g, 30 min, 4°C). Strep-Tactin® spin columns (Iba Lifesciences) were equilibrated with phosphate buffer, according to the user manual (2 × 500 µl, 700 x g, 30 sec, 4°C) and lysate was loaded onto the columns (700 x g, 30 sec, 4°C). Columns were washed with phosphate buffer (4 x 100 µl, 16,200 x g, 30 sec, 4°C) prior to eluting the protein in buffer containing D-biotin (5 mM) (2 × 100 µl, 700 x g, 30 sec, 4°C). Protein concentrations were measured on a Tecan infinite M200Pro plate reader by absorption at 280 nm (Nanoquant plate). A yield of approximately 3 mg/ml (≈ 100 µM) was obtained for each protein and the MW of each protein was confirmed by intact-protein MS. Proteins were found to be sufficiently pure by SDS-PAGE and were stored at -80°C.

Large-scale protein purification (Alr, Apo-Alr). The bacterial cell pellet from the overexpression culture (2 L) was resuspended in phosphate buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0). Cells were lysed by sonication and the lysate was clarified by centrifugation (36,000 x g, 30 min, 4°C). The supernatant was loaded onto a pre-equilibrated 5 ml StrepTrap column (GE Healthcare) using an Äkta purification system (GE Healthcare) and washed with phosphate buffer (50 ml, flow rate = 4 ml/min). Alr was eluted in phosphate buffer containing desthiobiotin (2.5 mM) as the PLP-holoenzyme. In order to generate the apoenzyme (Apo-Alr), column-bound proteins were additionally flushed with phosphate buffer containing 25 mM hydroxylamine (50 ml) prior to washing and elution. Protein-containing fractions were concentrated, desalted into phosphate buffer using a 5 ml HiTrap Desalting column (GE Healthcare) and protein aliquots were flash frozen in liquid nitrogen for storage at -80°C. Alr was found to be sufficiently pure by SDS-PAGE and its MW was confirmed by intact-protein MS. Protein concentrations were measured on a Tecan infinite M200Pro plate reader by absorption at 280 nm (Nanoquant plate).

Large-scale protein purification (Q2FF14, GP0 and HH8). Holoenzymes were overexpressed and purified as described above using the following buffers for both purification and desalting: Q2FF14 (50 mM Tris, 250 mM NaCl, pH 8), GP0 (100 mM Tris, 750 mM NaCH, pH 8) and HH8 (50 mM Tris, 750 mM NaCl, pH 8). Expression culture lysate was incubated with 100 µM PLP for 15 min on ice prior to purification in order to saturate proteins with the cofactor.

PLK kinetics. PLK activity was measured using the previously reported coupled assay.[75]

***In vitro* phosphorylation of PL-probes using PLK.** A solution containing 2 mM PL-probe, 25 µM PLK and 10 mM ATP was prepared in 1 ml kinase buffer (50 mM Tris, 50 mM KCl, 10 mM $MgCl_2$, pH 7.9) and incubated overnight at 4°C with gentle shaking. The solution was filtered through a 10 kDa MWCO centrifugal filter (Sartorius Stedim Biotech) to remove PLK, and phosphorylated probes were used in subsequent experiments without further purification, assuming full conversion (800 µM final concentration). Phosphorylation of PL probes could be confirmed by LC-MS/MS analysis using a H-ESI-II probe and LC conditions as described in the metabolomics section with slight modifications. A positive multiple reaction monitoring (MRM) was used as the detection mode and daughter ions of the respective HRMS probe ions were recorded in a range of 50-2000 *m/z*. Quantification of **PL4P** product samples showed the presence of > 99.5% **PL4P**.

Reconstitution of Apo-Alr with phosphorylated probes. Apo-Alr (50 µM in phosphate buffer) was incubated with phosphorylated probes (5 equivalents) overnight at 4°C with gentle shaking. Residual probe was removed by desalting into phosphate buffer using a 5 ml HiTrap Desalting column (GE Healthcare) on an Äkta purification system.

UV-vis measurements. UV-vis spectra of protein samples (100 µM) were recorded in phosphate buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) at 25°C on a Tecan infinite M200Pro plate reader (300-600 nm, 5 nm increments). For substrate studies, proteins were incubated with 10 equivalents of substrate for 15 min prior to measurement.

Alr kinetics. Racemization of D-Ala to L-Ala was measured using a coupled assay that monitors NADH formation (λ_{abs} =340 nm) by L-alanine dehydrogenase.[195] A premix containing 2 units/ml L-alanine dehydrogenase (Roche), 10 mM NAD^+ and 10 nM Alr was prepared in CHES buffer (100 mM CHES, 100 mM KCl, pH 8.8). 100 µl of the premix was pipetted into flat-bottom 96-well plates (Nunclon™ Delta Surface, Thermo Scientific) containing 5 µl D-Ala (1, 2, 10, 20, 50, 100, 200 and 300 mM). The reactions were incubated at 37°C and NADH formation was measured at 340 nm in 1 min kinetic intervals using a Tecan infinite M200Pro plate reader. Enzymatic activity was assayed in three replicates. Initial rates of NADH formation (µM/min) were determined for the different substrate concentrations and the data (mean ± standard deviation) was fitted to the Michaelis-Menten equation (non-linear regression) using GraphPad Prism-v5.0 software.

$NaBH_4$-reduction of proteins for intact-protein MS. 10 µM protein samples (25 µl) were incubated with 4 equivalents of cofactor derivative (1 µl of 500 µM stock PL, PLP, **PL4** or **PL4P**) in PBS for 30 min and subsequently treated with 10 mM $NaBH_4$ (2 µl of 250 mM stock prepared fresh in 0.1 M NaOH) at room temperature for 30 min. For Alr holoenzymes substituted with phosphorylated cofactor probes, samples were directly reduced without prior incubation. The $NaBH_4$ was quenched by acidification to pH 5-6 with HCl (5-10 µl of 0.5% FA) and neutralized to pH 7 with NaOH (5-10 µl of 0.1 M NaOH). Samples were diluted to 50 µl with PBS (5 µM final enzyme concentration) for measurement by intact-protein MS.

Intact-protein mass spectrometry. Samples were desalted and measured using a MassPREP On-Line Desalting Cartridge (Waters) on an Ultimate 3000 HPLC system (Dionex) coupled to a Finnigan LTQ-FT Ultra mass spectrometer (Thermo Scientific) with electrospray ionization (spray voltage 4.0 kV, tube lens 110 V, capillary voltage 48 V, sheath gas 60 arb, aux gas 10 arb, sweep gas 0.2 arb). Xcalibur Xtract Software (Thermo Scientific) was used for data analysis and deconvolution.

***In vitro* labeling of recombinant proteins.** 10 µM protein samples (50 µl in PBS) were labeled with PL(P)-probes (2 µl of 25× stock) at room temperature for 30 min. Upon reduction with 10 mM NaBH$_4$ (2 µl of 250 mM prepared fresh in 0.1 M NaOH) at room temperature for 30 min, proteins were precipitated by adding ice-cold acetone (4× volume) and incubating at -20°C for at least 1 h. Precipitated proteins were pelleted by centrifugation (18,000 x *g*, 15 min, 4°C) and washed with ice-cold MeOH (2 × 0.2 ml), using sonication to resuspend the pellets between washes. Pelletized proteins were resuspended in 50 µl PBS containing 0.4% (w/v) SDS and CuAAC to rhodamine-azide was performed by adding 0.1 mM Rh-N$_3$ (1 µl of 5 mM stock in DMSO), 0.5 mM BTTAA ligand[196] (2.5 µl of 10 mM stock in ddH$_2$O), 1 mM CuSO$_4$ (1 µl of 50 mM stock in ddH$_2$O), 2 mM NaAsc (1 µl of 100 mM prepared fresh in ddH$_2$O) and incubating at room temperature for 1 h in the dark. For **PL3**(P)-labeled proteins, CuAAC was performed using 0.1 mM Rh-alkyne (1 µl of 5 mM stock in DMSO) instead of Rh-N$_3$. Strain-promoted click chemistry was performed using 0.02 mM rhodamine-DBCO reagent (1 µl of 1 mM stock in DMSO). Samples were quenched with 2× gel loading buffer and analyzed by SDS-PAGE (12.5% polyacrylamide gels) with fluorescence scanning.

Analytical size-exclusion chromatography (SEC). Sample protein (100 µl, 50 µM) was injected onto a Superdex 200 10/300 gel filtration column (GE Healthcare) connected to an Äkta purification system 10 (GE Healthcare), which was pre-equilibrated with respective protein buffers used for purification. UV absorption was recorded at 280 nm. The oligomerization state was calculated from a calibration curve determined for the column using a protein standard mix (Gel Filtration Calibration Kit, GE Healthcare).

Enzyme substrate screens by MS. 10 µM enzyme was incubated with a mixture of amino acids (0.5 mM of each proteinogenic L-amino acid except Tyrosine, including Ornithine) and cofactor mix (100 µM of each NAD, NADH, NADP, NADPH, FAD, FMN, ATP, SAM, MgCl$_2$, MnCl$_2$, AcCoA, PLP) in ddH$_2$O for 30 min at 37°C. Reactions were stopped by precipitating protein using 400 µl ice-cold MeOH and incubating at -20°C for 1 h. Samples were subsequently centrifuged (18,000 x *g*, 15 min, 4°C) and the supernatant was evaporated in a speedvac centrifuge. The resulting residue was dissolved 150 µl H$_2$O/ACN (50:50, incl. 1% FA) and centrifuged (18,000 x *g*, 15 min, 4°C) to remove undissolved particles prior to measurement on an Ultimate 3000 RSLC system (Thermo Scientific) coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Scientific). Chromatographic separation was performed using a SeQuant® ZIC®-pHILIC (250 x 2.1 mm, 5 µm, 100 Å) (Merck Millipore).

Assay for *N*-Acetyltransferase activity. A previously described broad-substrate screening assay by Kuhn *et al.*[122] was used to assess *N*-acetyltransferase activity of GP0. Therefore, 50 µl of a mixture containing 2.5 mM substrate (Table 8.3), 0.5 mM AcCoA and 1 µM enzyme in 50 mM Tris-HCl, pH 8 were added to each well of a 96-well plate (Nunclon™ Delta Surface, Thermo Scientific) and incubated at 33°C for 15 min. Next, 50 µl of 6M guanidine-HCl buffer were added, followed by 100 µl of a solution containing Ellman's reagent (0.4 mM 5,5'-dithio-bis-(2-nitrobenzoic acid) in 100 mM Tris-HCl pH 8). The plate was incubated at room temperature for 10 min and absorption at a wavelength of 415 nm was measured using a Tecan infinite M200Pro plate reader. Samples were measured in duplicate.

Table 8.3 Substrates used for N-acetyltransferase activity assay.

No.	Substrate	No.	Substrate
1.	Spermine	24.	Erythromycin
2.	Spermidine	25.	Kanamycin
3.	L-Lys	26.	Streptomycin
4.	L-Gln	27.	Tobramycin
5.	L-Asn	28.	Ampicillin
6.	L-Arg	29.	Puromycin
7.	L-Asp	30.	Chloramphenicol
8.	L-Glu	31.	Bacitracin
9.	Gly	32.	Colistin
10.	L-Ala	33.	Adenine
11.	L-Ser	34.	Guanosine
12.	L-Thr	35.	Cytidine
13.	L-Met	36.	NAD
14.	L-Trp	37.	AMP
15.	L-Phe	38.	Thiamine pyrophosphate
16.	L-Tyr	39.	Urea
17.	L-Val	40.	Guanidine
18.	L-Ile	41.	Pyridoxamine
19.	L-Leu	42.	Thiamine
20.	L-homoserine	43.	Folic acid
21.	L-Citrulline	44.	4-aminobenzoic acid
22.	L-Orn	45.	D-glucosamine
23.	Gentamicin	46.	Tyramine

***S. aureus* USA300 TnPdxS growth curves.** The *S. aureus* USA300 transposon mutant TnPdxS (gene: SAUSA300_0504; pyridoxal biosynthesis lyase PdxS) was obtained from the Nebraska Transposon Mutant Library within the Network on Antimicrobial Resistance in Staphylococcus aureus (NARSA).[197] Bacteria were kept under antibiotic pressure (erythromycin) for all stages of the experiment to select for the transposon phenotype. Chemically-defined media (CDM) free of PL was prepared according to *Liebeke et al.*[198] and sterilized by filtration through a 0.2 µM PES filter (VWR). Overnight cultures of *S. aureus* USA300 TnPdxS were grown in B-medium containing 5 µg/ml erythromycin (Erm). Cultures were harvested, washed twice with 5 ml CDM-Erm (5 µg/ml)

and resuspended in CDM-Erm. 150 µl of bacterial culture (initial OD_{600} = 0.08) was added to flat-bottomed 96-well plates (Nunclon Delta surface, Thermo Scientific) containing 1.5 µl PL or PL-probe (100× stock in DMSO), and the plates were incubated at 37°C for 24 h with shaking (220 rpm). Endpoint measurements of bacterial growth were recorded at OD_{600} on a Tecan Infinite Pro platereader. Concentration of probe was plotted versus percent of maximum growth with respect to bacteria grown on PL-substituted media. Measurements were conducted in triplicate and were obtained from three individual biological experiments. Wells containing media substituted with DMSO or the **PL2** control probe were used as negative controls and served to validate the transposon mutant phenotype as no bacterial growth could be detected (OD_{600} < 0.1).

Labeling of recombinant holoenzymes (Alr-PL1P and Alr-PL4P) spiked into cell lysate. Cell lysate was prepared from stationary phase *S. aureus* USA300 TnPdxS (OD_{600} = 2.2-2.5) grown in B-medium containing 5 µg/ml Erm. Bacteria were harvested (6,000 rpm, 5 min, 4°C), washed twice with cold PBS and resuspended in PBS to an OD_{600} = 40. A 1 ml portion of this suspension was transferred to a 2 ml glass-bead lysis tube (Precellys® Lysing Kit, soil-grinding SK38) and bacteria were lysed using a Precellys® 24 Homogenizer (6 × 15 sec shaking at 5400 rpm, with cooling breaks on ice between cycles). The lysate was clarified by centrifugation (18,000 x *g*, 30 min, 4°C), and to 45 µl of the clarified lysate were added 5 µM holoenzyme (5 µl of 50 µM stock in PBS). Samples were subsequently subjected to $NaBH_4$ reduction, protein precipitation, CuAAC and SDS-PAGE analysis as described above for labeling of recombinant proteins. Proteins were treated with IAA (2 µl of 25× stock, 30 min, rt) prior to CuAAC as described below.

Analytical Labeling in *S. aureus*. *S. aureus* USA300 TnPdxS was grown to early stationary phase (5 h, OD_{600} = 2.2 - 2.5) in B-medium containing 5 µg/ml Erm. Bacteria were harvested (6,000 rpm, 5 min, 4°C), washed twice with CDM-Erm (5 µg/ml) and resuspended to an OD_{600} = 40. Per sample, 0.2 ml of this suspension was incubated with 2 µl probe or DMSO (100× stock in DMSO) for 2 h at 37°C, with shaking. Bacteria were subsequently harvested, washed with PBS (2 × 0.5 ml, 4°C), resuspended in 0.2 ml PBS and transferred to 0.5 ml glass-bead lysis tubes (0.2 g of 0.5 mm glass beads self-packed into 0.5 ml bulk tubes with caps, MoBio Laboratories). Cells were lysed mechanically using a Precellys® 24 Homogenizer (6 × 15 sec shaking at 5400 rpm) with cooling breaks on ice between cycles. The lysates were clarified by centrifugation (18,000 x *g*, 30 min, 4°C) and were reduced with 20 mM $NaBH_4$, (8 µl of 500 mM stock prepared fresh in 0.1 M NaOH) at room temperature for 30 min. Proteins were subsequently precipitated by adding ice-cold acetone (4× volume) and storage at -20°C overnight. Precipitated proteins were pelleted by centrifugation (18,000 x *g*, 15 min, 4°C) and washed with ice-cold MeOH (2 × 0.5 ml), using sonication to resuspend the pellets between washes. Proteins were solubilized in 50 µl PBS containing 0.4% (w/v) SDS and subjected to CuAAC by adding 0.1 mM $Rh-N_3$ (1 µl of 5 mM stock in DMSO), 0.5 mM BTTAA[196] ligand (2.5 µl of 10 mM stock in ddH₂O), 1 mM $CuSO_4$ (1 µl of 50 mM stock in ddH₂O), 2 mM NaAsc (1 µl of 100 mM prepared fresh in ddH₂O) and incubation at room

temperature for 1 h. Samples were quenched with 2× gel loading buffer and analyzed by SDS-PAGE (12.5% polyacrylamide gels) with fluorescence scanning.

Selection and validation of *S. aureus* USA300 TnQ2FF14 and TnGP0 transposon mutants. *S. aureus* USA300 strains with insertional disruption of genes encoding the proteins Q2FF14 (SAUSA300_2068) and GP0 (SAUSA300_2209) were obtained from the Nebraska transposon mutant library.[84] Cultures were streaked onto B-medium agar plates containing Erm (5 µg/ml) and colonies were selected for inoculation of overnight cultures in B-medium containing Erm (5 µg/ml). The transposon insertion was confirmed by PCR using the corresponding primers listed in Table 8.2 (Figure 8.1).
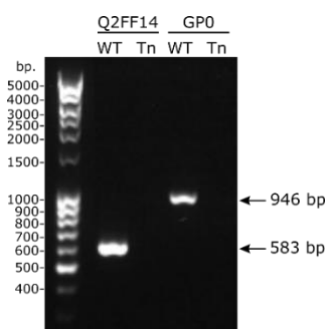
Figure 8.1 Validation of *S. aureus* TnQ2FF14 and TnGP0 transposon insertion. Transposon insertion into the genes encoding Q2FF14 and GP0 was confirmed by PCR using respective primer sequences (Table 8.2). PCR using wild-type *S. aureus* USA300 DNA resulted in bands corresponding to the respective genes (583 bp and 946 bp, respectively; 1% Agarose gel), while the bands could not be detected in the transposon mutant strains.

Selection and validation of *Pseudomonas aeruginosa* PAO1 TnPdxJ transposon mutant. The *Pseudomonas aeruginosa* PAO1 TnPdxJ transposon mutant (PW2400, PA0773, genotype pdxJ-C07:ISphoA/hah) was obtained from the University of Washington PAO1 transposon mutant library.[173,199] Cultures were streaked onto LB-agar plates and colonies were selected for inoculation of overnight cultures in LB medium containing 30 µg/ml tetracycline. The transposon insertion was confirmed by PCR as described on the two-allele library website, hosted by the Manoil lab (http://www.gs.washington.edu/labs/ manoil/libraryindex.htm) using the recommended primer sequences (Figure 8.2).

Table of Primers

Name	Sequence
33983R.f	ATGAACTTCGAGATGGGGGT
33983R.r	GCTTCGACACGGTATTCTCC
Hah-138	CGGGTGCAGTAATATCGCCCT
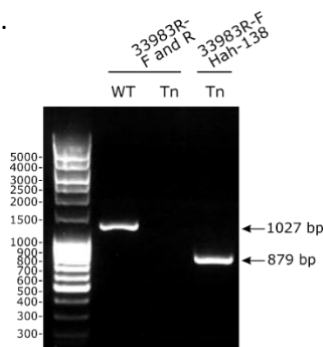
Figure 8.2 Validation of *P. aeruginosa* TnPdxJ transposon insertion. Transposon insertion into the gene *pdxj* was confirmed by PCR using recommended primer sequences (Washington University, Manoil lab). PCR using primers 33983R.f and 33983R.r with wild-type *P. aeruginosa* PAO1 DNA resulted in a band at 1027 bp (1% Agarose gel) corresponding to the *pdxj* gene. The band could not be detected in the *P. aeruginosa* PAO1 TnPdxJ strain. PCR using the primer 33983R.r and Hah-138 (site within transposon) produced the expected band at 879 bp.

8.3 Crystallography

Crystallization and structure determination for Alr. High-throughput crystallization screening was carried out (Phoenix, Art Robbins) and all crystals were obtained by the sitting-drop vapour diffusion method at 4 °C. Apo-Alr was crystallized with 0.1 M citric acid pH 4, 1.6 M ammonium sulphate as the precipitant. Reconstituted Alr-**PL1P** and Alr-**PL4P** holoenzymes were crystallized in 0.1 M Hepes pH 7 and 2 M sodium chloride and 0.1 M sodium acetate pH 5 containing 10% 2-methyl-1,3,-propanediol (MPD), respectively. Prior to cryo-cooling and storage in liquid nitrogen, either lithium sulphate (Apo-Alr, Alr-**PL1P** complex) or MPD (Alr-**PL4P** complex) was added to the crystallization drop for cryo-protection. Crystals were screened and diffraction data was collected at the beamlines of the European Synchrotron Radiation Facility (ESRF, ID23-2 and ID29) and the Swiss Light Source (SLS, PX I). The data were processed with XDS[200] and the crystals belonged to the space group $P2_12_12_1$. The resolution cut-offs were chosen according to the correlation coefficient of random half-data sets (1/2 CC) at about 50%.[201-203] The structures were solved using the coordinates of the previously reported structure (PDB code 4A3Q)[83] by molecular replacement in PHASER.[204] This was followed by simulated annealing in PHENIX,[205,206] prior to iterative cycles of manual model building in COOT[207] and restraint and TLS refinement in REFMAC5.[208] Structure optimization was carried out using the PDBredo server.[209] For data processing and structure refinement statistics see Table 8.4. All structural figures were prepared with PyMol (Delano Scientific).

Crystallization and structure determination for Q2FF14. Crystals of *S. aureus* Q2FF14 were obtained in 0.1 M bicine, pH 9.0, 1.6 M ammonium sulphate. Prior to cryo-cooling and storage in liquid nitrogen, 2 M lithium sulphate was added to the crystallisation drop. Data collection was carried out at the ESRF beamline MASSIF-1 automatically[210,211] and the data were processed with XDS to 1.75 Å (resolution cut-off $CC_{1/2}$ ~ 50%, symmetry $P2_1$). The structure was solved by molecular replacement with PHASER using the coordinates of the homologous, hypothetical protein TT1679 from *Thermus thermophilus* (PDB code 1V8D, 40% sequence identity). Model building, refinement and validation were carried out as for Alr (see above). For data processing and structure refinement statistics see Supporting Table 8.4.

All structural figures were prepared with PyMol (Delano Scientific) or UCSF Chimera. Sequence alignments were done with ClustalX.[116]

Table 8.4 Crystallographic data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parantheses.

	Apo	PL1P-complex	PL4P-complex	Q2FF14
Wavelength (Å)	0.976	0.873	0.979	0.966
Resolution range	45.2-2.15 (2.2-2.15)	49.5-2.45 (2.54-2.45)	49.4-1.9 (1.96-1.9)	45.3-1.75 (1.81-1.75)
Space group	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1$
Unit cell	64.7 114.5 126.1	85.9 106.9 131.0	85.5 106.8 130.8	54.7 136.9 79.2 $\beta = 98.6$
Total reflections	267,266 (22,547)	224,459 (21,447)	637,180 (55,865)	356,535 (35,255)
Unique reflections	51,458 (4,390)	44,955 (4,399)	95,389 (8,791)	114,397 (11,371)
Multiplicity	5.2 (5.1)	5.0 (4.9)	6.7 (6.4)	3.1 (3.1)
Completeness (%)	1.00 (0.99)	1.00 (0.99)	0.99 (0.92)	0.99 (0.99)
Mean I/sigma(I)	6.0 (0.4)	6.6 (1.4)	10.7 (1.2)	10.4 (0.90)
Wilson B-factor	34.4	32.3	28.7	30.3
R-merge	0.166 (1.73)	0.213 (1.02)	0.13 (1.48)	0.056 (1.19)
CC1/2	0.99 (0.85)	0.983 (0.5)	1.0 (0.45)	0.998 (0.534)
Reflections used in refinement	50,710 (4,797)	44,949 (4,399)	95,378 (8,791)	114,296 (11,325)
Reflections used for R-free	2,534 (237)	2,247 (219)	4,786 (462)	5,713 (563)
R-work	0.2041 (0.337)	0.195 (0.297)	0.199 (0.346)	0.194 (0.398)
R-free	0.234 (0.37)	0.215 (0.33)	0.21 (0.359)	0.220 (0.396)
Number of non-hydrogen atoms	6,475	6,341	6,712	8,510
macromolecules	6,040	6,102	6,104	8,006
ligands	138	68	125	85
Protein residues	763	771	764	1,058
RMS(bonds)	0.011	0.009	0.012	0.017
RMS(angles)	1.42	1.4	1.5	1.9
Ramachandran favoured (%)	97	97	97	96
Ramachandran allowed (%)	2.9	3	2.7	3
Ramachandran outliers (%)	0.3	0	0	0.57
Rotamer outliers (%)	1.7	1.2	0.9	1.6
Clashscore	1.5	2.3	1.1	3.2
Average B-factor	34.0	35.5	22.0	42.16
macromolecules	33.7	35.5	20.2	42.12
ligands	56.0	50.9	46.5	56.3
solvent	29.0	27.7	39.1	40.2
Number of TLS groups	14	2	14	42

8.4   Proteomics

***S. aureus* growth and labeling for chemical proteomics (standard protocol).** Overnight cultures of *S. aureus* USA300 TnPdxS grown in B-medium containing 5 µg/ml Erm were diluted 1:100 into fresh media (~ 100 ml for 3 samples) and grown for 5 h until early stationary phase (OD_{600} = 2.2 - 2.5). Bacteria were harvested (6,000 rpm, 5 min, 4°C), washed twice with 10 ml CDM-Erm (5 µg/ml) and resuspended to an OD_{600} = 40. For each sample, 1 ml of bacterial suspension (OD_{600} = 40) was incubated with 10 µl probe (100× stock in DMSO) or DMSO in culture tubes for 2 h at 37°C, with shaking. For samples treated with DCS, bacteria were incubated with 50 µl DCS (20× stock prepared fresh in 0.1 M sodium phosphate buffer, pH 8) for 30 min at 37°C, with shaking, prior to probe labeling. Bacteria were subsequently harvested, washed with PBS (2 × 1 ml, 4°C) and resuspended in 1 ml PBS for cell lysis.

***S. aureus* growth and labeling for chemical proteomics (PL1/PL3 growth supplementation).** Overnight cultures of *S. aureus* USA300 TnPdxS grown in B-medium (20 ml) containing 5 µg/ml Erm were harvested (6,000 rpm, 5 min, 4°C), washed twice with 5 ml CDM-Erm (5 µg/ml) and resuspended in 5 ml. This suspension was used to inoculate bacterial cultures to an OD_{600} = 0.1 in 100 ml CDM-Erm supplemented with either 25 µM probe (25 µl of 100 mM stock in DMSO) or 0.5 µM PL (25 µl of 2 mM stock in DMSO) in baffled flasks. The cultures were grown for 30 h at 37°C with shaking, at which point they were harvested (6,000 rpm, 5 min, 4°C), washed twice with 1 ml PBS, resuspended and adjusted to 1 ml of OD_{600} = 60 prior to cell lysis.

***P. aeruginosa* growth and labeling for chemical proteomics (standard protocol).** Overnight cultures of *P. aeruginosa* PAO1 TnPdxJ (obtained as described in biological methods) grown in LB-medium were diluted 1:100 into fresh media and grown for 5 h until early stationary phase. Bacteria were harvested (6,000 rpm, 5 min, 4°C), washed twice with 10 ml MOPS-succinate medium (50 mM MOPS pH7.2, 20 mM sodium succinate, 20 mM NH4Cl, 1 mM $MgSO_4$, 10 mM KCl, 3.5 µM $FeSO_4$) and resuspended to an OD_{600} = 40. For each sample, 1 ml of bacterial suspension (OD_{600} = 40) was incubated with 10 µl probe (100× stock in DMSO) or DMSO in culture tubes for 2 h at 37°C with shaking. Bacteria were subsequently harvested, washed with PBS (2 × 1 ml, 4°C) and resuspended in 1 ml PBS for cell lysis.

Bacterial cell lysis and $NaBH_4$ reduction. Bacterial suspensions prepared above were transferred to 2 ml glass-bead lysis tubes (Precellys® Lysing Kit, soil-grinding SK38). Cells were lysed mechanically using a Precellys® 24 Homogenizer (6 × 15 sec shaking at 5400 rpm) with 2-min cooling breaks on ice between cycles. Upon centrifugation (18,000 x *g*, 30 min, 4°C), the clarified lysates were reduced with 20 mM $NaBH_4$, (500 mM stock prepared fresh in 0.1 M NaOH) at room temperature for 30 min. Proteins were subsequently precipitated by adding ice-cold acetone (4× volume) and stored at -20°C overnight. Precipitated proteins were pelleted by centrifugation

(18,000 x *g*, 15 min, 4°C) and washed with ice-cold MeOH (2 × 1 ml), using sonication to resuspend the pellets between washes. Proteins were solubilized in 0.5 ml PBS containing 0.4% (w/v) SDS and protein concentrations were adjusted to 2 mg/ml (= 1 mg total per sample). Samples from chemical supplementation experiments were instead solubilized in 1 ml PBS containing 0.4% SDS (w/v) and adjusted to 2 mg/ml (= 2 mg total per sample).

***P. aeruginosa* labeling in cell lysate.** Overnight cultures of *P. aeruginosa* PAO1 grown in LB-medium were diluted 1:100 into fresh media and grown for 5 h until early stationary phase. Bacteria were harvested (6,000 rpm, 5 min, 4°C), washed twice with 10 ml PBS buffer and resuspended to an OD_{600} = 40 in PBS containing protease inhibitors (Complete Mini Protease Inhibitor Cocktail, 1 tablet per 10 ml PBS, Roche). Bacteria were lysed by sonication and centrifuged (18,000 x *g*, 30 min, 4°C). For each labeling reaction, 750 µl clarified lysate was incubated with 5 µl **PL4P** (300× stock in kinase buffer: 50 mM Tris, 50 mM KCl, 10 mM $MgCl_2$, pH 7.9) or kinase buffer as a control, for 45 min on ice. Samples were reduced with 20 mM $NaBH_4$, (500 mM stock prepared fresh in 0.1 M NaOH) at room temperature for 30 min and proteins were subsequently precipitated by adding ice-cold acetone (4× volume) and storing at -20°C overnight. Precipitated proteins were pelleted by centrifugation (18,000 x *g*, 15 min, 4°C) and washed with ice-cold MeOH (2 × 1 ml), using sonication to resuspend the pellets between washes. Proteins were solubilized in 0.5 ml PBS containing 0.4% (w/v) SDS and protein concentrations were adjusted to 2 mg/ml (= 1 mg total per sample) prior to click chemistry and enrichment.

HeLa cell growth and labeling for chemical proteomics. HeLa cells were grown to 80-90% confluency in a 15 cm dish in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 5 mM glutamine and 10% FCS. Cells were incubated with 100 or 10 µM **PL4** probe (100 mM stock) or DMSO for 1 h at 37°C, upon which they were detached with a cell scraper, centrifuged (600 x *g*, 5 min, 4°C) and washed with 1 ml PBS. Cell pellets were resuspended in 1 ml lysis buffer (PBS + 1% v/v NP-40 and 1% w/v sodium deoxycholate) and incubated on ice for 20 min. The lysate was clarified by centrifugation (18,000 x *g*, 30 min, 4°C) and samples were then reduced with 20 mM $NaBH_4$ (500 mM stock prepared fresh in 0.1 M NaOH) at room temperature for 30 min. Proteins were subsequently precipitated by adding ice-cold acetone (4× volume) and stored at -20°C overnight. Precipitated proteins were pelleted by centrifugation (18,000 x *g*, 15 min, 4°C) and washed with ice-cold MeOH (2 × 1 ml), using sonication to resuspend the pellets between washes. Proteins were solubilized in 0.5 ml PBS containing 0.4% (w/v) SDS and protein concentrations were adjusted to 2 mg/ml (= 1 mg total per sample) prior to click chemistry and enrichment.

Click chemistry. Samples were subjected to CuAAC by adding 0.1 mM tri-functional rhodamine-biotin-azide[96] linker (10 mM stock in DMSO), 0.5 mM BTTAA ligand[196] (10 mM stock in ddH_2O), 1 mM $CuSO_4$ (50 mM stock in ddH_2O) and 2 mM NaAsc (100 mM prepared fresh in ddH_2O) to each sample and incubating for 1 h at room temperature in the dark. **PL1** and **PL2**-labeled samples were

incubated with 20 mM IAA (500 mM prepared fresh in ddH$_2$O) for 30 min at rt prior to click chemistry. **PL3**-labeled samples were incubated with 0.2 mM EZ-Link^{TM} Phosphine-PEG$_3$-Biotin (Sigma Aldrich) (10 mM stock in DMSO) for 4 h at 37°C, followed by 20 h at 25°C. Upon acetone precipitation and washing (as described above), the protein pellets were re-solubilized in 0.5 ml PBS containing 0.4% SDS (w/v) and centrifuged (18,000 x g, 5 min, rt) prior to avidin bead enrichment. Samples from chemical supplementation experiments were instead re-solubilized in 1 ml PBS containing 0.4% (w/v) SDS.

Enrichment and sample preparation for MS. Solubilized, labeled protein (0.5-1 ml, 2 mg/ml in 0.4% (w/v) SDS in PBS) was transferred to Lo-bind Eppendorf tubes containing 50 μl of avidin-agarose bead slurry (Sigma Aldrich) pre-equilibrated with 0.4% (w/v) SDS in PBS (3 × 1 ml, 400 x g, 5 min, rt), and samples were rotated for 1 h at room temperature. The beads were then washed with 0.4% (w/v) SDS in PBS (3 × 1 ml), 6 M urea (2 × 1 ml) and finally PBS (3 × 1 ml). During the last washing step, 10% of the sample was removed for gel-based analysis of enrichment. This fraction was heated (10 min, 95°C) in 25 μl gel loading buffer, centrifuged (13,000 rpm, 3 min) and analyzed by fluorescence SDS-PAGE (NuPAGE 4-12% Bis-Tris gel, Thermo Fisher Scientific). The remaining beads were resuspended in 200 μl X-buffer (7 M urea, 2 M thiourea in 20mM HEPES buffer, pH 7.5). Upon reduction with 5 mM TCEP (2 μl of 500 mM stock in ddH$_2$O) for 1 h at 37°C, proteins were alkylated using 10 mM IAA (4 μl of 500 mM stock in ddH$_2$O) for 30 min at 25°C and samples were quenched with 10 mM DTT (4 μl of 500 mM stock in ddH$_2$O) for 30 min at 25°C. Enzymatic digestion using LysC (1 μl of 0.5 μg/μl, Wako, MS-grade) was first carried out for 2 h at 25°C, upon which samples were diluted with triethylammonium bicarbonate (TEAB) buffer (600 μl of 50 mM stock in ddH$_2$O) and digested with trypsin (1.5 μl of 0.5 μg/μl in 50 mM acetic acid, Promega, sequencing grade) for an additional 16 h at 37°C. Samples were acidified to 1% (v/v) FA and desalted using SepPak® C18 cartridges (50 mg, Waters) with a vacuum manifold.[212] The cartridges were first washed with ACN (2 × 1 ml) and equilibrated with 0.1% (v/v) TFA (3 × 1 ml) prior to sample loading. After washing with 0.1% (v/v) TFA (3 × 1 ml) and 0.5% (v/v) FA (1 × 0.5 ml), peptides were eluted in 80% (v/v) ACN containing 0.5% FA (3 × 0.25 ml) and freeze-dried using a speedvac centrifuge. Samples were prepared for MS-analysis by dissolving in 25 μl of 1% (v/v) FA and filtering through 0.22 μm PVDF filters (Millipore). All samples were prepared in three biological replicates from individual cultures and were processed in parallel.

Dimethyl labeling. Labeled proteins were enriched, reduced, alkylated and digested as described above. Samples were acidified to 1% (v/v) FA and peptides were dimethyl labeled on column using SepPak® C18 cartridges (50 mg, Waters) with a vacuum manifold.[212] The cartridges were first washed with ACN (2 × 1 ml) and equilibrated with 0.5% (v/v) FA (3 × 1 ml) prior to loading the samples. After washing with 0.5% (v/v) FA (4 × 1 ml), each sample was flushed with respective labeling reagents (5 × 1 ml Light: 30 mM NaBH$_3$CN, 0.2% formaldehyde; Medium: 30 mM NaBH$_3$CN, 0.2% formaldehyde-d_2; Heavy: 30 mM NaBD$_3$CN, 0.2% ^{13}C-formaldehyde-d_2, each in 45

mM sodium phosphate buffer pH 7.5). The columns were subsequently washed with 0.5% (v/v) FA (2 × 1 ml) and eluted in 80% (v/v) ACN containing 0.5% FA (3 × 0.25 ml). Labels were mixed and freeze-dried using a speedvac centrifuge. Samples were prepared for MS-analysis by dissolving in 20 µl of 1% (v/v) FA and filtering through 0.22 µm PVDF filters (Millipore). All samples were prepared in three biological replicates from individual cultures and were processed in parallel.

Whole-proteome samples. Overnight cultures of wild-type *S. aureus* USA300 and TnQ2FF14 or TnGP0 (obtained as described in Biological Methods) were grown in B-medium containing 5 µg/ml Erm and diluted 1:100 into fresh media for the day culture. Bacteria were grown for 5 h until early stationary phase (OD_{600} = 5), at which point 2 ml of the culture was harvested (6,000 rpm, 5 min, 4°C). The cell pellet was washed with PBS (2 × 1 ml, 4°C) and resuspended in 0.5 ml lysis buffer (50 mM Tris, 75 mM NaCl pH 8, containing 8 M urea, 1 mM EDTA and Complete Mini Protease Inhibitor Cocktail, Roche (1 tablet per 10 ml)). Samples were lysed in 2 ml glass-bead lysis tubes (Precellys® Lysing Kit, soil-grinding SK38) using a Precellys® 24 Homogenizer (6 × 15 sec shaking at 5400 rpm) with 2-min cooling breaks on ice between cycles. Cell lysates were clarified by centrifugation (18,000 x *g*, 30 min, 4°C) and protein concentration was adjusted to 0.2 ml of 0.25 mg/ml (= 0.25 mg). Samples were reduced with 10 mM DTT (1 M stock in ddH$_2$O) for 1 h at 37°C, alkylated with 55 mM IAA (500 mM stock in ddH$_2$O) for 30 min at 25°C and quenched with 40 mM DTT (1 M stock in ddH$_2$O) for 30 min at 25°C. Proteins were first digested using LysC (2 µl of 0.5 µg/µl, Wako, MS-grade) for 2 h at 25°C, upon which samples were diluted with triethylammonium bicarbonate (TEAB) buffer (600 µl of 50 mM stock in ddH$_2$O) and digested with trypsin (4 µl of 0.5 µg/µl in 50 mM acetic acid, Promega, sequencing grade) for an additional 16 h at 37°C. Samples were acidified to 1% (v/v) FA and desalted using SepPak® C18 cartridges (50 mg, Waters) with a vacuum manifold.[212] The cartridges were first washed with ACN (2 × 1 ml) and equilibrated with 0.1% (v/v) TFA (3 × 1 ml) prior to loading the samples. After washing with 0.1% (v/v) TFA (3 × 1 ml) and 0.5% (v/v) FA (1 × 0.5 ml), peptides were eluted in 80% (v/v) ACN containing 0.5% FA (3 × 0.25 ml) and freeze-dried using a speedvac centrifuge. Samples were prepared for MS-analysis by dissolving in 100 µl of 1% (v/v) FA and filtering through 0.22 µm PVDF filters (Millipore). All samples were prepared in three biological replicates from individual cultures and were processed in parallel.

MS measurement and analysis. LC-MS/MS analysis was performed with an Ultimate3000 Nano-HPLC system (Thermo Scientific) coupled to an Orbitrap Fusion instrument (Thermo Scientific). Samples were loaded on a 2 cm PepMap RSLC C18 trap column (particles 2 µm, 100 Å, inner diameter 75 µm, Thermo Scientific) with 0.1% TFA and separated on a 50 cm PepMap RSLC C18 column (particles 2 µm, 100 Å, inner diameter 75 µm, Thermo Scientific) heated at 50 °C. The gradient was run from 5-32% ACN containing 0.1% formic acid during a 152 min method (7 min 5%, 105 min to 22%, 10 min to 32%, 10 min to 90%, 10 min wash at 90%, 10 min equilibration at 5%) at a flow rate of 300 nl/min. Survey scans (m/z 300-1500) were acquired in the orbitrap with a resolution of 120,000 at m/z 200 and the maximum injection time set to 50 ms (target value

2e5). The most intense ions of charge states 2-7 were selected for fragmentation with high-energy collisional dissociation at a normalized collision energy of 30%. The instrument was operated in top speed mode and spectra were acquired in the ion trap with the maximum injection time set to 50 ms (target value 1e4). The option to inject ions for all available parallelizable time was enabled. Dynamic exclusion of sequenced peptides was set to 60 s. Data were acquired using Xcalibur software version 3.0sp2 (Thermo Scientific).

MS raw files were analyzed with MaxQuant[213] software (version 1.5.3.8). MS/MS-based peptide identification was carried out using the Andromeda[214] search engine with the *S. aureus* USA300 UniProtKB database (May 2016, 2607 accessions). For other experiments, the *S. aureus* NCTC8325 UniProtKB database (October 2017, 2889 accessions), the *P. aeruginosa* PAO1 UniProtKB database (September 2017, 5563 accessions) or human UniprotKB database (May 2016, 70625 accessions) were used. The following parameter settings were used: peptide and protein FDR 1%, enzyme specificity trypsin, minimal number of amino acids required for peptide identification 7, variable modification methionine oxidation, fixed modification carbamidomethylation. At least two unique peptides were required for the identification of proteins. All other parameters were used according to the default settings. For label-free protein quantification, the MaxLFQ[94] algorithm was used as part of the MaxQuant environment. The 'match between runs' option was enabled to maximize the number of quantification events across all replicates. Statistical analysis was performed in Perseus[215] (version 1.5.5.3). Proteins identified only by site modification, reverse hits or contaminants were removed. Data were filtered to retain only those proteins with > 2 MS/MS counts and quantified in at least 2 out of 3 replicates (alternatively: 5/6 replicates or 8/9 replicates). Missing values were then imputed on the basis of a normal distribution (width=0.3, down-shift=1.8). Volcano plots were generated by performing a two-sample *t*-test with permutation-based statistics (FDR 0.05, s0 = 1, or other as specified). The complete lists of PLP-DEs (PLPomes) for the selected species were identified by gene ontology (GO-term: 0030170; pyridoxal-phosphate binding), EC classification (enzyme.expasy.org; cofactor search: pyridoxal 5'-phosphate), and Uniprot (www.uniprot.org, cofactor search: pyridoxal 5'-phosphate).

Overrepresentation analysis is based on gene ontology annotations and was performed with the Bingo App in the Cytoscape environment.[216] Statistically significant proteins from the volcano plot were compared to all proteins present in the plot for the annotation pyridoxal phosphate binding (GO:0030170) in the category of molecular process. According to literature research (see searches described above), we additionally annotated the following proteins as pyridoxal phosphate binding and added them to the GO list: A0A0H2XFQ3, A0A0H2XG73, A0A0H2XGD2, A0A0H2XGZ7, A0A0H2XHH8, A0A0H2XII6, A0A0H2XIJ0, A0A0H2XIS2, A0A0H2XJX6, A0A0H2XKI8, Q2FF55, Q2FGI6, Q2FGI7, Q2FH63, Q2FH64, Q2FHT1 and Q2FKE3. Analysis was based on a hypergeometrical test with the multiple testing correction according to Benjamini-Hochberg and a significance level of 0.05.

Confidence plot analysis. Only proteins significantly enriched (104 proteins, volcano plot Figure 4.3b, Table 4.1) were selected for analysis. Profile plots of $\log_2$ LFQ intensities of individual proteins across proteomic samples were generated using Perseus. Each of the 18 PLP-DEs was individually selected as a reference profile and the 10 or 20 most similar profiles were identified by Pearson correlation. Corresponding proteins were grouped into confidence classes as illustrated in Figure 4.3.

Heatmap and clustering analysis for concentration-dependent probe enrichment. The mean raw LFQ intensities from triplicate measurements were normalized to DMSO (= 0) and highest LFQ intensity (= 1) for individual proteins across eight probe concentrations. Samples negatively enriched compared to DMSO were set equal to 0 for analysis. Profile plots of the normalized mean LFQ intensities were generated across different probe concentrations using Perseus and hierarchical clustering (Pearson, max number of clusters = 5) was performed to group similarly behaving proteins. The heatmap was prepared using OriginPro (2017) graphing and analysis software.

Heatmap analysis of PL4, PL5 and PL6. The mean raw LFQ intensities from triplicate measurements were normalized to DMSO (= 0) and highest LFQ intensity (= 1) for individual PLP-DE within **PL4**-, **PL5**- and **PL6**-enriched samples. The heatmap was prepared using OriginPro (2017) graphing and analysis software.

8.5 Metabolomics

Targeted PL4P metabolomics. Bacteria were grown and labeled as described for the proteomic workflow (see Proteomics section). For each sample, 2.5 ml of bacterial suspension (OD_{600} = 40) was incubated with 25 µl probe (1000× stock in DMSO) in culture tubes for 2 h at 37°C with shaking. Samples were then shock-cooled to 4°C in a dry ice/ethanol bath while agitating to avoid local freezing. Bacteria were harvested (6000 rpm, 5 min, 4°C), washed with 0.6% (w/v) NaCl solution (2 × 5 ml) and resuspended in 5 ml extraction buffer (100% MeOH containing 0.5 µM PL as an internal standard). Cells were lysed by sonication (3 x 20 sec, with cooling breaks on ice) and samples were stored overnight at -80°C. Cell debris and precipitated protein were pelleted by centrifugation (18,000 x g, 30 min, 4°C) and the clarified extract was evaporated under a stream of nitrogen. Metabolome samples were dissolved in 150 µl H_2O/ACN (50:50, incl. 1% FA), centrifuged (18,000 x g, 15 min, 4°C) and the supernatant was stored at -80 °C until LC-MS analysis. Amounts of **PL4** and **PL4P** were quantified using calibration curves. Samples were measured on an Ultimate 3000 RSLC system (Thermo Scientific) coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Scientific). Chromatographic separation was performed using a SeQuant® ZIC®-pHILIC (250 x 2.1 mm, 5 µm, 100 Å) (Merck Millipore). Data was processed using Xcalibur 2.2 SP1.48 Quan Browser (Thermo Scientific).

Whole-metabolome profiling of *S. aureus* USA300 TnQ2FF14 and TnGP0 vs WT. Overnight cultures of wild-type *S. aureus* USA300 and TnQ2FF14 or TnGP0 (obtained as described in Biological Methods) were grown in B-medium containing 5 µg/ml Erm and diluted 1:100 into fresh B-medium for the day culture. Bacteria were grown for 5 h until early stationary phase (OD_{600} = 5), at which point 15 ml of the culture was harvested (6,000 rpm, 5 min, 4°C) and the bacteria were washed with 0.9% NaCl (2 × 5 ml, 4°C). Cells were lysed by sonication (3 x 20 sec, with cooling breaks on ice) and samples were stored overnight at -80°C. Cell debris and precipitated protein were pelleted by centrifugation (18,000 x g, 30 min, 4°C) and the clarified extract was evaporated in a speedvac centrifuge. Metabolome samples were dissolved in 150 µl H_2O/ACN (50:50, incl. 1% FA), centrifuged (18,000 x g, 15 min, 4°C) and the supernatant was stored at -80 °C until LC-MS analysis. Samples were measured on an Ultimate 3000 RSLC system (Thermo Scientific) coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Scientific). Chromatographic separation was performed using a SeQuant® ZIC®-pHILIC (250 x 2.1 mm, 5 µm, 100 Å) (Merck Millipore). Data was processed using Xcalibur 2.2 SP1.48 Quan Browser (Thermo Scientific) and analyzed using XCMS.[217]

8.6   Binding-site Identification

Sample preparation. Recombinant protein samples (2 µg, 50 µl) were incubated with 5 equivalents of PLP for 30 min at rt, and were subsequently reduced with 10 mM NaBH$_4$ (2 µl of 250 mM stock prepared fresh in 0.1 M NaOH) for 30 min at rt. Proteins were precipitated by adding ice-cold acetone (4× volume) and incubating at -20°C for at least 1 h. Precipitated proteins were pelleted by centrifugation (18,000 x *g*, 15 min, 4°C) and washed with ice-cold MeOH (2 × 0.2 ml), using sonication to resuspend the pellets between washes. Samples were resuspended in 100 µl X-buffer (7 M urea, 2 M thiourea in 20mM HEPES buffer, pH 7.5) and reduced with 5 mM TCEP (1 µl of 500 mM stock in ddH$_2$O) for 1 h at 37°C. Proteins were alkylated using 10 mM IAA (2 µl of 500 mM stock in ddH$_2$O) for 30 min at 25°C, followed by quenching with 10 mM DTT (2 µl of 500 mM stock in ddH$_2$O) for 30 min at 25°C. Samples were diluted with 600 µl chymotrypsin buffer (100 mM Tris-HCl, 10 mM CaCl$_2$, pH 8) and digested by incubating with chymotrypsin (1 µl of 0.5 µg/µl in 1 mM HCl, Promega, sequencing grade) for 16 h at 25°C. Double-digested samples (Q2FF14) were first incubated with trypsin (1 µl of 0.5 µg/µl in 50 mM acetic acid, Promega, sequencing grade) for 30 min at 37°C and were then digested with chymotrypsin as described above. Samples digested with GluC (Q2FF14) were incubated with the enzyme (5 µl of 0.1 µg/µl in ddH$_2$O, Roche, sequencing grade) for 16 h at 25°C. Reactions were quenched by acidification to 1% (v/v) FA and peptides were desalted using SepPak® C18 cartridges (50 mg, Waters) with a vacuum manifold.[212] The cartridges were washed with ACN (2 × 1 ml) and equilibrated with 0.1% (v/v) TFA (3 × 1 ml) prior to loading the samples. After washing with 0.1% (v/v) TFA (3 × 1 ml) and 0.5% (v/v) FA (1 × 0.5 ml), peptides were eluted in 80% (v/v) ACN containing 0.5% FA (3 × 0.25 ml) and freeze-dried using a speedvac centrifuge. Peptides were prepared for MS-analysis by dissolving in 25 µl of 1% (v/v) FA and filtering through 0.22 µm PVDF filters (Millipore).

MS measurement and analysis. LC-MS/MS analysis was performed with an Ultimate3000 Nano-HPLC system (Thermo Scientific) coupled to an Orbitrap Fusion instrument (Thermo Scientific) as described in the proteomics section, with some modifications. For PLP binding site identification, chymotrypsin digested samples were measured with a shorter gradient during a 62 min method. Charge states 1-7 were considered for measurement and depend scans were only performed on single charge state per precursor and measured in the orbitrap (AGC target 5e4, injection time 50 ms). MS raw files were analyzed with MaxQuant[213] software (version 1.5.3.8) as described in the proteomics section with some modifications. MS/MS-based peptide identification was carried out using the Andromeda[214] search engine. For PLP binding site identification, the enzyme specificity was set to chymotrypsin (with additional cleavage after leucine and methionine) for all enzymes except for Q2FF14 which was double-digested with trypsin and chymotrypsin or digested with GluC. As a variable modification, the PLP moiety (+ 231.02966) at lysine residues was added with neutral losses of H$_3$PO$_4$ and the PLP moiety implemented. As we expect one PLP modification site

per protein, we selected the site with the highest confidence based on 100% localization probability, best PEP (posterior error probability), score, score difference, delta score and manual evaluation of MS/MS spectra from MaxQuant (see Appendix).

References

1   Di Salvo, M. L., Budisa, N. & Contestabile, R. in *Beilstein Bozen Symposium on Molecular Engineering and Control* 27-66 (Beilstein-Institut, Prien (Chiemsee), Germany, 2013).

2   Eliot, A. C. & Kirsch, J. F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **73**, 383-415 (2004).

3   Schneider, G., Kack, H. & Lindqvist, Y. The manifold of vitamin B6 dependent enzymes. *Structure* **8**, R1-6 (2000).

4   Percudani, R. & Peracchi, A. A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.* **4**, 850-854 (2003).

5   Toney, M. D. Reaction specificity in pyridoxal phosphate enzymes. *Arch. Biochem. Biophys.* **433**, 279-287 (2005).

6   Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim. Biophys. Acta* **1814**, 1407-1418 (2011).

7   Keller, J. W. *et al.* Pseudomonas cepacia 2,2-dialkylglycine decarboxylase. Sequence and expression in Escherichia coli of structural and repressor genes. *J. Biol. Chem.* **265**, 5531-5539 (1990).

8   Dunathan, H. C. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **55**, 712-716 (1966).

9   Chan-Huot, M., Sharif, S., Tolstoy, P. M., Toney, M. D. & Limbach, H. H. NMR studies of the stability, protonation States, and tautomerism of (13)C- AND (15)N-labeled aldimines of the coenzyme pyridoxal 5'-phosphate in water. *Biochemistry* **49**, 10818-10830 (2010).

10   Graber, R. *et al.* Conversion of aspartate aminotransferase into an L-aspartate beta-decarboxylase by a triple active-site mutation. *J. Biol. Chem.* **274**, 31203-31208 (1999).

11   Seebeck, F. P. & Hilvert, D. Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation. *J. Am. Chem. Soc.* **125**, 10158-10159 (2003).

12   Metzler, D. E. & Snell, E. E. Deamination of serine. I. Catalytic deamination of serine and cysteine by pyridoxal and metal salts. *J. Biol. Chem.* **198**, 353-361 (1952).

13   Olivard, J., Metzler, D. E. & Snell, E. E. Catalytic racemization of amino acids by pyridoxal and metal salts. *J. Biol. Chem.* **199**, 669-674 (1952).

14   Christen, P. & Mehta, P. K. From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Chem. Rec.* **1**, 436-447 (2001).

15   Dunathan, H. C. & Voet, J. G. Stereochemical evidence for the evolution of pyridoxal-phosphate enzymes of various function from a common ancestor. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3888-3891 (1974).

16   Grishin, N. V., Phillips, M. A. & Goldsmith, E. J. Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Protein Sci.* **4**, 1291-1304 (1995).

17   Jansonius, J. N. Structure, evolution and action of vitamin B6-dependent enzymes. *Curr. Opin. Struct. Biol.* **8**, 759-769 (1998).

18   Mehta, P. K. & Christen, P. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* **74**, 129-184 (2000).

19   Du, Y. L. *et al.* A pyridoxal phosphate-dependent enzyme that oxidizes an unactivated carbon-carbon bond. *Nat. Chem. Biol.* **12**, 194-199 (2016).

20   Berkovitch, F. *et al.* A locking mechanism preventing radical damage in the absence of substrate, as revealed by the x-ray structure of lysine 5,6-aminomutase. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15870-15875 (2004).

21   Lepore, B. W., Ruzicka, F. J., Frey, P. A. & Ringe, D. The x-ray crystal structure of lysine-2,3-aminomutase from Clostridium subterminale. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13819-13824 (2005).

22   di Salvo, M. L. *et al.* On the mechanism of Escherichia coli pyridoxal kinase inhibition by pyridoxal and pyridoxal 5'-phosphate. *Biochim. Biophys. Acta.* **1854**, 1160-1166 (2015).

23   Ghatge, M. S. *et al.* Pyridoxal 5'-phosphate is a slow tight binding inhibitor of E. coli pyridoxal kinase. *PLoS One* **7**, e41680, doi:10.1371/journal.pone.0041680 (2012).

24   Ronchi, S., Zapponi, M. C. & Ferri, G. Inhibition by pyridoxal-phosphate of glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* **8**, 325-331 (1969).

25   Soler, M., Jabalquinto, A. M. & Beytia, E. Hog liver mevalonate kinase: inactivation by pyridoxal-5'-phosphate and evidence of dead-end inhibition by one of the substrates. *Int. J. Biochem.* **10**, 931-935 (1979).

26   Yang, H. & Abeles, R. H. Purification and properties of Escherichia coli 4'-phosphopantothenoylcysteine decarboxylase: presence of covalently bound pyruvate. *Biochemistry* **26**, 4076-4081 (1987).

27   Bramucci, E., Milano, T. & Pascarella, S. Genomic distribution and heterogeneity of MocR-like transcriptional factors containing a domain belonging to the superfamily of the pyridoxal-5'-phosphate dependent enzymes of fold type I. *Biochem. Biophys. Res. Commun.* **415**, 88-93 (2011).

28   Suvorova, I. A. & Rodionov, D. A. Comparative genomics of pyridoxal 5'-phosphate-dependent transcription factor regulons in Bacteria. *Microb. Genom.* **2**, e000047, doi:10.1099/mgen.0.000047 (2016).

29   Takenaka, T., Ito, T., Miyahara, I., Hemmi, H. & Yoshimura, T. A new member of MocR/GabR-type PLP-binding regulator of D-alanyl-D-alanine ligase in Brevibacillus brevis. *FEBS J.* **282**, 4201-4217 (2015).

30   Cellini, B., Montioli, R., Oppici, E., Astegno, A. & Voltattorni, C. B. The chaperone role of the pyridoxal 5'-phosphate and its implications for rare diseases involving B6-dependent enzymes. *Clin. Biochem.* **47**, 158-165 (2014).

31   Amadasi, A. *et al.* Pyridoxal 5'-phosphate enzymes as targets for therapeutic agents. *Curr. Med. Chem.* **14**, 1291-1324 (2007).

32   Kappes, B., Tews, I., Binter, A. & Macheroux, P. PLP-dependent enzymes as potential drug targets for protozoan diseases. *Biochim. Biophys. Acta* **1814**, 1567-1576 (2011).

33   Percudani, R. & Peracchi, A. The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. *BMC Bioinformatics* **10**, 273 (2009).

34   Catazaro, J., Caprez, A., Guru, A., Swanson, D. & Powers, R. Functional evolution of PLP-dependent enzymes based on active-site structural similarities. *Proteins* **82**, 2597-2608 (2014).

35   Whittaker, M. M., Penmatsa, A. & Whittaker, J. W. The Mtm1p carrier and pyridoxal 5'-phosphate cofactor trafficking in yeast mitochondria. *Arch. Biochem. Biophys.* **568**, 64-70 (2015).

36   Simon, E. S. & Allison, J. Determination of pyridoxal-5'-phosphate (PLP)-bonding sites in proteins: a peptide mass fingerprinting approach based on diagnostic tandem mass spectral features of PLP-modified peptides. *Rapid Commun. Mass Spectrom.* **23**, 3401-3408 (2009).

37   Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **77**, 383-414 (2008).

38   Geurink, P. P., Prely, L. M., van der Marel, G. A., Bischoff, R. & Overkleeft, H. S. Photoaffinity labeling in activity-based protein profiling. *Top Curr. Chem.* **324**, 85-113 (2012).

39   Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686-4687 (2003).

40   Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046-15047 (2004).

41   Saxon, E. & Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007-2010 (2000).

42   Weerapana, E., Speers, A. E. & Cravatt, B. F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)--a general method for mapping sites of probe modification in proteomes. *Nat. Protoc.* **2**, 1414-1425 (2007).

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

44   Parker, C. G. *et al.* Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527-541 e529, doi:10.1016/j.cell.2016.12.029 (2017).

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

46   S.A., G. M. K. J. S. Electrophilic natural products and their biological targets. *Nat. Prod. Rep.* **9**, 659-682 (2012).

47   Grammel, M. & Hang, H. C. Chemical reporters for biological discovery. *Nat. Chem. Biol.* **9**, 475-484 (2013).

48   Liu, T. W. *et al.* Genome-wide chemical mapping of O-GlcNAcylated proteins in Drosophila melanogaster. *Nat. Chem. Biol.* **13**, 161-167 (2017).

49   Laughlin, S. T. & Bertozzi, C. R. Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation. *Nat. Protoc.* **2**, 2930-2944 (2007).

50   Tate, E. W., Kalesh, K. A., Lanyon-Hogg, T., Storck, E. M. & Thinon, E. Global profiling of protein lipidation using chemical proteomic technologies. *Curr. Opin. Chem. Biol.* **24**, 48-57 (2015).

51   Anderson, L. N. *et al.* Live Cell Discovery of Microbial Vitamin Transport and Enzyme-Cofactor Interactions. *ACS Chem. Biol.* **11**, 345-354 (2016).

52   Romine, M. F. *et al.* Elucidation of roles for vitamin B12 in regulation of folate, ubiquinone, and methionine metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1205-E1214 (2017).

53   Korytnyk, W. & Ahrens, H. 5-homopyridoxals, 5-thiopyridoxal, and related compounds. Synthesis, tautomerism, and biological properties. *J. Med. Chem.* **14**, 947-952 (1971).

54   Korytnyk, W., Srivastava, S. C., Angelino, N., Potti, P. G. & Paul, B. A general method for modifying the 2-methyl group of pyridoxol. Synthesis and biological activity of 2-vinyl- and 2-ethynylpyridoxols and related compounds. *J. Med. Chem.* **16**, 1096-1101 (1973).

55   Muller, I. B. *et al.* Poisoning pyridoxal 5-phosphate-dependent enzymes: a new strategy to target the malaria parasite Plasmodium falciparum. *PLoS One* **4**, e4406, doi:10.1371/journal.pone.0004406 (2009).

56   Fonda, M. L. Interaction of pyridoxal analogues with glutamate apodecarboxylase and aspartate apoaminotransferase. *J. Biol. Chem.* **246**, 2230-2240 (1971).

57   Mechanik, M. L., Torchinsky, Y. M., Florentiev, V. L. & Karpeisky, M. Y. Interaction of the apoenzyme of L-glutamate decarboxylase with pyridoxal phosphate analogues. *FEBS Lett.* **13**, 177-180 (1971).

58   Denesyuk, A. I., Denessiouk, K. A., Korpela, T. & Johnson, M. S. Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. *J. Mol. Biol.* **316**, 155-172 (2002).

59   Schnackerz, K. D. & Cook, P. F. Resolution of pyridoxal 5'-phosphate from O-acetylserine sulfhydrylase from Salmonella typhimurium and reconstitution of apoenzyme with cofactor and cofactor analogues as a probe of the cofactor binding site. *Arch. Biochem. Biophys.* **324**, 71-77 (1995).

60   Griswold, W. R. & Toney, M. D. Role of the pyridine nitrogen in pyridoxal 5'-phosphate catalysis: activity of three classes of PLP enzymes reconstituted with deazapyridoxal 5'-phosphate. *J. Am. Chem. Soc.* **133**, 14823-14830 (2011).

61   Blethen, S. L., Boeker, E. A. & Snell, E. E. Argenine decarboxylase from Escherichia coli. I. Purification and specificity for substrates and coenzyme. *J. Biol. Chem.* **243**, 1671-1677 (1968).

62   Muhlradt, P. F., Morino, Y. & Snell, E. E. Vitamin b(2) analogues. Synthesis and biological activity of homologues of pyridoxal 5'-phosphate. *J. Med. Chem.* **10**, 341-344 (1967).

63   Morino, Y. & Snell, E. E. Coenzyme activity of homologues of pyridoxal phosphate. *Biochemistry* **57**, 1692-1699 (1967).

64   Boekelheide, V. & Linn, W. J. Rearrangements of N-Oxides. A Novel Synthesis of Pyridyl Carbinols and Aldehydes. *J. Am. Chem. Soc.* **76**, 1286-1291 (1954).

65   Roth, G. J., Liepold, B., Müller, S. G. & Bestmann, H. J. Further Improvements of the Synthesis of Alkynes from Aldehydes. *Synthesis* **1**, 59-62 (2004).

66   Yang, D., Shih, Y. & Liu, H. Chemical synthesis of stereospecifically labeled pyridoxamine 5'-phosphate. *J. Org. Chem.* **56**, 2940-2946 (1991).

67   Kaiser, E. M. *et al.* Regiointegrity of carbanions derived by selective metalations of dimethylpyridines and -quinolines. *J. Organomet. Chem.* **213**, 405-417 (1981).

68   Kim, Y.-C. & Jacobson, K. A. Versatile synthesis of 6-alkyl and aryl substituted pyridoxal derivatives. *Synthesis* **1**, 119-122 (2000).

69   Mukherjee, T., Hanes, J., Tews, I., Ealick, S. E. & Begley, T. P. Pyridoxal phosphate: biosynthesis and catabolism. *Biochim. Biophys. Acta* **1814**, 1585-1596 (2011).

70   Mulligan, J. H. & Snell, E. E. Transport and metabolism of vitamin B6 in lactic acid bacteria. *J. Biol. Chem.* **252**, 835-839 (1977).

71   Stolz, J. & Vielreicher, M. Tpn1p, the plasma membrane vitamin B6 transporter of Saccharomyces cerevisiae. *J. Biol. Chem.* **278**, 18990-18996 (2003).

72   di Salvo, M. L., Contestabile, R. & Safo, M. K. Vitamin B(6) salvage enzymes: mechanism, structure and regulation. *Biochim. Biophys. Acta.* **1814**, 1597-1608 (2011).

73   Ahrens, M. L., Maass, G., Schuster, P. & Winkler, H. Kinetic study of the hydration mechanism of vitamin B6 and related compounds. *J. Am. Chem. Soc.* **92**, 6134-6139 (1970).

74   di Salvo, M. L., Safo, M. K. & Contestabile, R. Biomedical aspects of pyridoxal 5'-phosphate availability. *Front Biosci (Elite Ed)*, 897-913 (2012).

75   Nodwell, M. B., Koch, M. F., Alte, F., Schneider, S. & Sieber, S. A. A subfamily of bacterial ribokinases utilizes a hemithioacetal for pyridoxal phosphate salvage. *J. Am. Chem. Soc.* **136**, 4992-4999 (2014).

76   Strominger, J. L., Ito, I. & Threnn, R. H. Competitive inhibition of enzymatic reactions by oxamycin. *J. Am. Chem. Soc.* **82**, 998-999 (1960).

77   Watanabe, A. *et al.* Reaction mechanism of alanine racemase from Bacillus stearothermophilus: x-ray crystallographic studies of the enzyme bound with N-(5'-phosphopyridoxyl)alanine. *J. Biol. Chem.* **277**, 19166-19172 (2002).

78   Toney, M. D. & Kirsch, J. F. Lysine 258 in aspartate aminotransferase: enforcer of the Circe effect for amino acid substrates and general-base catalyst for the 1,3-prototropic shift. *Biochemistry* **32**, 1471-1479 (1993).

79   Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* **41**, 2596-2599 (2002).

80   Speers, A. E. & Cravatt, B. F. Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.* **11**, 535-546 (2004).

81   Tornoe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057-3064 (2002).

82   van Geel, R., Pruijn, G. J., van Delft, F. L. & Boelens, W. C. Preventing thiol-yne addition improves the specificity of strain-promoted azide-alkyne cycloaddition. *Bioconjug. Chem.* **23**, 392-398 (2012).

83   Scaletti, E. R., Luckner, S. R. & Krause, K. L. Structural features and kinetic characterization of alanine racemase from Staphylococcus aureus (Mu50). *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 82-92 (2012).

84   Fey, P. D. *et al.* A genetic resource for rapid and comprehensive phenotype screening of nonessential Staphylococcus aureus genes. *MBio.* **4**, e00537-00512, doi:10.1128/mBio.00537-12 (2013).

85   Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017-1031 (2007).

86   Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17** (1999).

87   Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6591-6596 (1999).

88   Paša-Tolić, L. *et al.* High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry. *J. Am. Chem. Soc.* **121**, 7949-7950 (1999).

89   Ong, S. E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252-262 (2005).

90   Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics* **1**, 376-386 (2002).

91   Hsu, J.-L., Huang, S.-Y., Chow, N.-H. & Chen, S.-H. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Anal. Chem.* **75**, 6843-6852 (2003).

92   Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154-1169 (2004).

93   Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895-1904 (2003).

94   Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell Proteomics* **13**, 2513-2526 (2014).

95   Wu, Y., Chen, J., Liu, Z. & Wang, F. Identification of pyridoxal phosphate modified proteins using mass spectrometry. *Rapid Commun. Mass Spectrom.* (2017).

96   Eirich, J. *et al.* Pretubulysin derived probes as novel tools for monitoring the microtubule network via activity-based protein profiling and fluorescence microscopy. *Mol. Biosyst.* **8**, 2067-2075 (2012).

97   Huisgen, R., Szeimies, G. & Möbius, L. 1.3-Dipolare Cycloadditionen, XXXII. Kinetik der Additionen organischer Azide an CC-Mehrfachbindungen. *Chemische Berichte* **100**, 2494-2507 (1967).

98   Rudolf, G. C., Heydenreuter, W. & Sieber, S. A. Chemical proteomics: ligation and cleavage of protein modifications. *Curr. Opin. Chem. Biol.* **17**, 110-117 (2013).

99   Ning, X., Guo, J., Wolfert, M. A. & Boons, G. J. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. *Angew. Chem. Int. Ed. Engl.* **47**, 2253-2255 (2008).

100   Jewett, J. C., Sletten, E. M. & Bertozzi, C. R. Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. *J. Am. Chem. Soc.* **132**, 3688-3690 (2010).

101   Laughlin, S. T., Baskin, J. M., Amacher, S. L. & Bertozzi, C. R. In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **320**, 664-667 (2008).

102   Staudinger, H. & Meyer, J. Über neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **2**, 635-646 (1919).

103   Nilsson, B. L., Kiessling, L. L. & Raines, R. T. Staudinger ligation: a peptide from a thioester and azide. *Org. Lett.* **2**, 1939-1941 (2000).

104   Galperin, M. Y. & Koonin, E. V. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* **32**, 5452-5463 (2004).

105   Galperin, M. Y. Conserved 'hypothetical' proteins: new hints and new puzzles. *Comp. Funct. Genomics* **2**, 14-18 (2001).

106   Jaroszewski, L. *et al.* Exploration of uncharted regions of the protein universe. *PLoS Biol.* **7**, e1000205, doi:10.1371/journal.pbio.1000205 (2009).

107   Fleischman, N. M. *et al.* Molecular characterization of novel pyridoxal-5'-phosphate-dependent enzymes from the human microbiome. *Protein Sci.* **23**, 1060-1076 (2014).

108   Knight, A. M. *et al.* Bioinformatic analysis of fold-type III PLP-dependent enzymes discovers multimeric racemases. *Appl. Microbiol. Biotechnol.* **101**, 1499-1507 (2017).

109   Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-285 (2016).

110   Mozzarelli, A. & Bettati, S. Exploring the pyridoxal 5'-phosphate-dependent enzymes. *Chem. Rec.* **6**, 275-287 (2006).

111   Prosser, G. A. & de Carvalho, L. P. Metabolomics Reveal d-Alanine:d-Alanine Ligase As the Target of d-Cycloserine in Mycobacterium tuberculosis. *ACS Med. Chem. Lett.* **4**, 1233-1237 (2013).

112   de Carvalho, L. P. *et al.* Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase. *Chem. Biol.* **17**, 323-332 (2010).

113   Saito, N. *et al.* Metabolite profiling reveals YihU as a novel hydroxybutyrate dehydrogenase for alternative succinic semialdehyde metabolism in Escherichia coli. *J. Biol. Chem.* **284**, 16442-16451 (2009).

114   Eswaramoorthy, S. *et al.* Structure of a yeast hypothetical protein selected by a structural genomics approach. *Acta. Crystallogr. D. Biol. Crystallogr.* **59**, 127-135 (2003).

115   Ito, T. *et al.* Conserved pyridoxal protein that regulates Ile and Val metabolism. *J. Bacteriol.* **195**, 5439-5449 (2013).

116   Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).

117   Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324, doi:10.1093/nar/gku316 (2014).

118   Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-549 (2010).

119   Kishishita, S., Terada, T., Shirouzu, M., Kuramitsu, S., Yokoyama, S. Crystal structure of the conserved hypothetical protein TT1679 from Thermus thermophilus. *Protein Data Bank (PDB), unpublished* (2004).

120   Bhuiya, M. W. *et al.* The first crystal structure of hyperthermostable NAD-dependent glutamate dehydrogenase from Pyrobaculum islandicum. *J. Mol. Biol.* **345**, 325-337 (2005).

121   Srivastava, P. *et al.* Structural characterization of a Gcn5-related N-acetyltransferase from Staphylococcus aureus. *PLoS One* **9**, e102348, doi:10.1371/journal.pone.0102348 (2014).

122   Kuhn, M. L., Majorek, K. A., Minor, W. & Anderson, W. F. Broad-substrate screen as a tool to identify substrates for bacterial Gcn5-related N-acetyltransferases with unknown substrate specificity. *Protein Sci.* **22**, 222-230 (2013).

123   Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747-751 (2005).

124   Sulzenbacher, G. *et al. Crystal structure of E. coli K-12 YggS* (Protein Data Bank, 2006).

125   Prunetti, L. *et al.* Evidence That COG0325 Proteins are involved in PLP Homeostasis. *Microbiology* (2016).

126   Darin, N. *et al.* Mutations in PROSC Disrupt Cellular Pyridoxal Phosphate Homeostasis and Cause Vitamin-B6-Dependent Epilepsy. *Am. J. Hum. Genet.* **99**, 1325-1337 (2016).

127   Ebihara, A. *et al.* Structure-based functional identification of a novel heme-binding protein from Thermus thermophilus HB8. *J. Struct. Funct. Genomics* **6**, 21-32 (2005).

128   Dailey, H. A., Gerdes, S., Dailey, T. A., Burch, J. S. & Phillips, J. D. Noncanonical coproporphyrin-dependent bacterial heme biosynthesis pathway that does not use protoporphyrin. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2210-2215 (2015).

129   Lobo, S. A. *et al.* Staphylococcus aureus haem biosynthesis: characterisation of the enzymes involved in final steps of the pathway. *Mol. Microbiol.* **97**, 472-487 (2015).

130   Labbé, R. F. & Nielson, L. in *Porphyrins in human diseases, 1st International Porphyrin Meeting, Freiburg i.Br., May 1975.* Vol. (ed M. Doss) 141-147 (Karger, Basel, Switzerland, 1976).

131   Charette, M. & Gray, M. W. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49**, 341-351 (2000).

132   Thapa, K., Oja, T. & Metsa-Ketela, M. Molecular evolution of the bacterial pseudouridine-5'-phosphate glycosidase protein family. *FEBS J* **281**, 4439-4449 (2014).

133   Cho, S. W., Yoon, H. Y., Ahn, J. Y., Lee, E. Y. & Lee, J. Cassette mutagenesis of lysine 130 of human glutamate dehydrogenase. An essential residue in catalysis. *Eur. J. Biochem.* **268**, 3205-3213 (2001).

134   Liu, J. Q., Dairi, T., Kataoka, M., Shimizu, S. & Yamada, H. L-allo-threonine aldolase from Aeromonas jandaei DK-39: gene cloning, nucleotide sequencing, and identification of the pyridoxal 5'-phosphate-binding lysine residue by site-directed mutagenesis. *J. Bacteriol.* **179**, 3555-3560 (1997).

135   Chen, D. & Frey, P. A. Identification of lysine 346 as a functionally important residue for pyridoxal 5'-phosphate binding and catalysis in lysine 2, 3-aminomutase from Bacillus subtilis. *Biochemistry* **40**, 596-602 (2001).

136   Tang, K. H., Harms, A. & Frey, P. A. Identification of a novel pyridoxal 5'-phosphate binding site in adenosylcobalamin-dependent lysine 5,6-aminomutase from Porphyromonas gingivalis. *Biochemistry* **41**, 8767-8776 (2002).

137   Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190-D199 (2017).

138   Lambert, M. P. & Neuhaus, F. C. Mechanism of D-cycloserine action: alanine racemase from Escherichia coli W. *J. Bacteriol.* **110**, 978-987 (1972).

139   Meroueh, S. O. *et al.* Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4404-4409 (2006).

140   Lobocka, M., Hennig, J., Wild, J. & Klopotowski, T. Organization and expression of the Escherichia coli K-12 dad operon encoding the smaller subunit of D-amino acid dehydrogenase and the catabolic alanine racemase. *J. Bacteriol.* **176**, 1500-1510 (1994).

141   Strych, U., Huang, H. C., Krause, K. L. & Benedik, M. J. Characterization of the alanine racemases from Pseudomonas aeruginosa PAO1. *Curr. Microbiol.* **41**, 290-294 (2000).

142   Strych, U. & Benedik, M. J. Mutant analysis shows that alanine racemases from Pseudomonas aeruginosa and Escherichia coli are dimeric. *J. Bacteriol.* **184**, 4321-4325 (2002).

143   Mitnick, C. D. *et al.* Comprehensive treatment of extensively drug-resistant tuberculosis. *N. Engl. J. Med.* **359**, 563-574 (2008).

144   Torun, T. *et al.* Side effects associated with the treatment of multidrug-resistant tuberculosis. *Int. J. Tuberc. Lung. Dis.* **9**, 1373-1377 (2005).

145   Caminero, J. A., Sotgiu, G., Zumla, A. & Migliori, G. B. Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *Lancet Infect. Dis.* **10**, 621-629 (2010).

146   LeMagueres, P. *et al.* The 1.9 A crystal structure of alanine racemase from Mycobacterium tuberculosis contains a conserved entryway into the active site. *Biochemistry* **44**, 1471-1481 (2005).

147   Fenn, T. D., Stamper, G. F., Morollo, A. A. & Ringe, D. A side reaction of alanine racemase: transamination of cycloserine. *Biochemistry* **42**, 5775-5783 (2003).

148   Peisach, D., Chipman, D. M., Van Ophem, P. W., Manning, J. M. & Ringe, D. D-cycloserine inactivation of D-amino acid aminotransferase leads to a stable noncovalent protein complex with an aromatic cycloserine-PLP derivative. *J. Am. Chem. Soc.* **120**, 2268-2274 (1998).

149   Soper, T. S. & Manning, J. M. Different modes of action of inhibitors of bacterial D-amino acid transaminase. A target enzyme for the design of new antibacterial agents. *J. Biol. Chem.* **256**, 4263-4268 (1981).

150   David, S. Synergic activity of D-cycloserine and beta-chloro-D-alanine against Mycobacterium tuberculosis. *J. Antimicrob. Chemother.* **47**, 203-206 (2001).

151   Neuhaus, F. C. Selective inhibition of enzymes utilizing alanine in the biosynthesis of peptidoglycan. *Antimicrob. Agents. Chemother. (Bethesda)* **7**, 304-313 (1967).

152   Malashkevich, V. N., Strop, P., Keller, J. W., Jansonius, J. N. & Toney, M. D. Crystal structures of dialkylglycine decarboxylase inhibitor complexes. *J. Mol. Biol.* **294**, 193-200 (1999).

153   Lowther, J. *et al.* Inhibition of the PLP-dependent enzyme serine palmitoyltransferase by cycloserine: evidence for a novel decarboxylative mechanism of inactivation. *Mol. Biosyst.* **6**, 1682-1693 (2010).

154   Sieradzki, K. & Tomasz, A. Suppression of beta-lactam antibiotic resistance in a methicillin-resistant Staphylococcus aureus through synergic action of early cell wall inhibitors and some other antibiotics. *J. Antimicrob. Chemother.* **39**, 47-51 (1997).

155   Roze, U. & Strominger, J. L. Alanine racemase from Staphylococcus aureus: conformation of its substrates and its inhibitor, D-cycloserine. *Mol. Pharm.* **2**, 92-94 (1966).

156   Shah, P. & Swiatlo, E. A multifaceted role for polyamines in bacterial pathogens. *Mol. Microbiol.* **68**, 4-16 (2008).

157   Hutton, C. A., Perugini, M. A. & Gerrard, J. A. Inhibition of lysine biosynthesis: an evolving antibiotic strategy. *Mol. Biosyst.* **3**, 458-465 (2007).

158   Contestabile, R. *et al.* l-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase. A subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem.* **268**, 6508-6525 (2001).

159   di Salvo, M. L. *et al.* Alanine racemase from Tolypocladium inflatum: a key PLP-dependent enzyme in cyclosporin biosynthesis and a model of catalytic promiscuity. *Arch. Biochem. Biophys.* **529**, 55-65 (2013).

160   Halouska, S. *et al.* Metabolomics analysis identifies d-Alanine-d-Alanine ligase as the primary lethal target of d-Cycloserine in mycobacteria. *J. Proteome Res.* **13**, 1065-1076 (2014).

161   Neuhaus, F. C. The enzymatic synthesis of D-alanyl-D-alanine. II. Kinetic studies on D-alanyl-D-alanine synthetase. *J. Biol. Chem.* **237**, 3128-3135 (1962).

162   Neuhaus, F. C. & Lynch, J. L. The Enzymatic Synthesis of D-Alanyl-D-Alanine. 3. On the Inhibition of D-Alanyl-D-Alanine Synthetase by the Antibiotic D-Cycloserine. *Biochemistry* **3**, 471-480 (1964).

163   Zawadzke, L. E., Bugg, T. D. & Walsh, C. T. Existence of two D-alanine:D-alanine ligases in Escherichia coli: cloning and sequencing of the ddlA gene and purification and characterization of the DdlA and DdlB enzymes. *Biochemistry* **30**, 1673-1682 (1991).

164   Boucher, H. W. *et al.* Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **48**, 1-12 (2009).

165   Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12**, 371-387 (2013).

166   Fischbach, M. A. & Walsh, C. T. Antibiotics for emerging pathogens. *Science* **325**, 1089-1093 (2009).

167   Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71-109 (2011).

168   Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* **5**, 175-186 (2007).

169   Brown, L., Wolf, J. M., Prados-Rosales, R. & Casadevall, A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nat. Rev. Microbiol.* **13**, 620-630 (2015).

170   Li, X. Z. & Nikaido, H. Efflux-mediated drug resistance in bacteria. *Drugs* **64**, 159-204 (2004).

171   LaBauve, A. E. & Wargo, M. J. Growth and laboratory maintenance of Pseudomonas aeruginosa. *Curr. Protoc. Microbiol.* **Chapter 6**, Unit 6E 1, doi:10.1002/9780471729259.mc06e01s25 (2012).

172   Mittenhuber, G. Phylogenetic analyses and comparative genomics of vitamin B6 (pyridoxine) and pyridoxal phosphate biosynthesis pathways. *J. Mol. Microbiol. Biotechnol.* **3**, 1-20 (2001).

173   Held, K., Ramage, E., Jacobs, M., Gallagher, L. & Manoil, C. Sequence-verified two-allele transposon mutant library for Pseudomonas aeruginosa PAO1. *J. Bacteriol.* **194**, 6387-6389 (2012).

174   Fedde, K. N. & Whyte, M. P. Alkaline phosphatase (tissue-nonspecific isoenzyme) is a phosphoethanolamine and pyridoxal-5'-phosphate ectophosphatase: normal and hypophosphatasia fibroblast study. *Am. J. Hum. Genet.* **47**, 767-775 (1990).

175   Clayton, P. T. B6-responsive disorders: a model of vitamin dependency. *J. Inherit. Metab. Dis.* **29**, 317-326 (2006).

176   Gdynia, H. J. *et al.* Severe sensorimotor neuropathy after intake of highest dosages of vitamin B6. *Neuromuscul. Disord.* **18**, 156-158 (2008).

177   Schaumburg, H. *et al.* Sensory neuropathy from pyridoxine abuse. A new megavitamin syndrome. *N. Engl. J. Med.* **309**, 445-448 (1983).

178   Karlsson, A., Fonnum, F., Malthe-Sorenssen, D. & Storm-Mathisen, J. Effect of the convulsive agent 3-mercaptopropionic acid on the levels of GABA, other amino acids and glutamate decarboxylase in different regions of the rat brain. *Biochem. Pharmacol.* **23**, 3053-3061 (1974).

179   Jung, M. J., Lippert, B., Metcalf, B. W., Bohlen, P. & Schechter, P. J. gamma-Vinyl GABA (4-amino-hex-5-enoic acid), a new selective irreversible inhibitor of GABA-T: effects on brain GABA metabolism in mice. *J. Neurochem.* **29**, 797-802 (1977).

180   Lippert, B., Metcalf, B. W., Jung, M. J. & Casara, P. 4-amino-hex-5-enoic acid, a selective catalytic inhibitor of 4-aminobutyric-acid aminotransferase in mammalian brain. *Eur. J. Biochem.* **74**, 441-445 (1977).

181   Kalviainen, R. *et al.* Vigabatrin, a gabaergic antiepileptic drug, causes concentric visual field defects. *Neurology* **53**, 922-926 (1999).

182   Galluzzi, L. *et al.* Effects of vitamin B6 metabolism on oncogenesis, tumor progression and therapeutic responses. *Oncogene* **32**, 4995-5004 (2013).

183   Potera, C., Rose, D. P. & Brown, R. R. Vitamin B6 deficiency in cancer patients. *Am. J. Clin. Nutr.* **30**, 1677-1679 (1977).

184   Tryfiates, G. P. & Morris, H. P. Effect of pyridoxine deficiency on tyrosine transaminase activity and growth of four Morris hepatomas. *J. Natl. Cancer Inst.* **52**, 1259-1262 (1974).

185   Snell, K., Natsumeda, Y., Eble, J. N., Glover, J. L. & Weber, G. Enzymic imbalance in serine metabolism in human colon carcinoma and rat sarcoma. *Br. J. Cancer.* **57**, 87-90 (1988).

186   Thorndike, J., Pelliniemi, T. T. & Beck, W. S. Serine hydroxymethyltransferase activity and serine incorporation in leukocytes. *Cancer Res.* **39**, 3435-3440 (1979).

187   Lin, H. B. *et al.* Mimosine targets serine hydroxymethyltransferase. *J. Biol. Chem.* **271**, 2548-2556 (1996).

188   DiSorbo, D. M. & Litwack, G. Vitamin B6 kills hepatoma cells in culture. *Nutr. Cancer* **3**, 216-222 (1982).

189   Shultz, T. D., Santamaria, A. G., Gridley, D. S., Stickney, D. R. & Slater, J. M. Effect of pyridoxine and pyridoxal on the in vitro growth of human malignant melanoma. *Anticancer Res.* **8**, 1313-1318 (1988).

190   Galluzzi, L. *et al.* Prognostic impact of vitamin B6 metabolism in lung cancer. *Cell Rep.* **2**, 257-269 (2012).

191   Broncel, M., Serwa, R. A., Bunney, T. D., Katan, M. & Tate, E. W. Global Profiling of Huntingtin-associated protein E (HYPE)-Mediated AMPylation through a Chemical Proteomic Approach. *Mol. Cell Proteomics* **15**, 715-725 (2016).

192   Westcott, N. P., Fernandez, J. P., Molina, H. & Hang, H. C. Chemical proteomics reveals ADP-ribosylation of small GTPases during oxidative stress. *Nat. Chem. Biol.* **13**, 302-308 (2017).

193   Wright, M. H. *et al.* Validation of N-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach. *Nat. Chem.* **6**, 112-121 (2014).

194   Kim, Y. C. & Jacobson, K. A. Versatile synthesis of 6-alkyl and aryl substituted pyridoxal derivatives. *Synthesis-Stuttgart*, 119-122 (2000).

195   Sun, S. & Toney, M. D. Evidence for a two-base mechanism involving tyrosine-265 from arginine-219 mutants of alanine racemase. *Biochemistry* **38**, 4058-4065 (1999).

196   Besanceney-Webler, C. *et al.* Increasing the efficacy of bioorthogonal click reactions for bioconjugation: a comparative study. *Angew. Chem. Int. Ed. Engl.* **50**, 8051-8056 (2011).

197   Fey, P. D. *et al.* A Genetic Resource for Rapid and Comprehensive Phenotype Screening of Nonessential Staphylococcus aureus Genes. *Mbio* **4**, doi:ARTN e00537-12.10.1128/mBio.00537-12 (2013).

198   Liebeke, M. *et al.* A metabolomics and proteomics study of the adaptation of Staphylococcus aureus to glucose starvation. *Mol. Biosyst.* **7**, 1241-1253 (2011).

199   Jacobs, M. A. *et al.* Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14339-14344 (2003).

200   Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 133-144 (2010).

201   Diederichs, K. & Karplus, P. A. Better models by discarding data? *Acta Crystallogr. D. Biol. Crystallogr* **69**, 1215-1222 (2013).

202   Evans, P. Biochemistry. Resolving some old problems in protein crystallography. *Science* **336**, 986-987 (2012).

203   Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).

204   McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).

205   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213-221 (2010).

206   Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352-367 (2012).

207   Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486-501 (2010).

208   Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 355-367 (2011).

209   Stamper, G. F., Morollo, A. A. & Ringe, D. Reaction of alanine racemase with 1-aminoethylphosphonic acid forms a stable external aldimine. *Biochemistry* **37**, 10438-10445 (1998).

210   Bowler, M. W. *et al.* MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *J. Synchrotron Radiat.* **22**, 1540-1547 (2015).

211   Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallogr. D. Biol. Crystallogr.* **71**, 1757-1767 (2015).

212   Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4**, 484-494 (2009).

213   Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301-2319 (2016).

214   Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794-1805 (2011).

215   Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731-740 (2016).

216   Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449 (2005).

217   Huan, T. *et al.* Systems biology guided by XCMS Online metabolomics. *Nat. Methods* **14**, 461 (2017).

Appendix

Abbreviations

ABMP	activity-based metabolomic profiling		DMSO	dimethyl sulfoxide
ABP	activity-based probes		DOPA	L-3,4-dihydroxyphenylalanine
ABPP	activity-based protein profiling		DTNB	5,5'-dithiobis(2-nitrobenzoic acid)
AcCoA	acetyl coenzyme A		DTT	dithiothreitol
ACN	acetonitrile		DUF	domain of unknown function
ADP	adenosine diphosphate		DXP	de-oxyxylulose 5-phosphate
ALP	alkaline phosphatase		ECD	electron capture dissociation
Alr	alanine racemase		EDTA	ethylenediaminetetraacetic acid
AMP	adenosine monophosphate		Erm	erythromycin
AT	aminotransferase		ESI	electrospray ionization
ATP	adenosine 5'-triphosphate		ETD	electron transfer dissociation
BARAC	biarylazacyclooctyne		FA	formic acid
bp	basepairs		FAD	flavin adenine dinucleotide
BTTAA	2-[4-((bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino)methyl)-1H-1,2,3-triazol-1-yl]acetic acid		FDR	false discovery rate
			FMN	flavin mononucleotide
CAM	ceric ammonium molybdate		GABA	gamma-aminobutyric acid
CDM	chemically-defined media		GlcNAc	N-acetylglucosamine
CHES	N-cyclohexyl-2-aminoethanesulfonic acid		GNAT	Gcn5-related N-acetyltransferases
CID	collision-induced dissociation		GO	gene ontology
CNS	central nervous system		h	hour
CPASS	comparison of active site structures		HCD	higher energy collisional dissociation
CS	cysteine synthase		HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
CuAAC	copper-catalyzed azide-alkyne 1,3-dipolar cycloaddition		HILIC	hydrophilic interaction chromatography
Da	Dalton		HPLC	high-performance liquid chromatography
DAP	diaminopimelate		HRMS	high-resolution mass spectrometry
DAP-DC	diaminopimelate decarboxylase		IAA	iodoacetamide
DAT	D-amino acid aminotransferase		IMAC	immobilized metal ion affinity chromatography
DBCO	dibenzocyclooctyne		IPTG	isopropyl-1-thio-β-galactopyranoside
DCM	Dichloromethane		iTRAQ	isotope tags for relative and absolute quantification
DCS	D-cycloserine			
ddH$_2$O	double-distilled H2O		kDa	kilo Dalton
Ddl	D-ala-D-ala ligase		LB	lysogeny broth
DIFO	Difluorocyclooctyne		LC	liquid chromatography
DIPEA	N,N-diisopropylethylamine		LC-MS	liquid chromatography - mass spectrometry
DMEM	Dulbecco's Modified Eagle Medium		LDA	lithium diisopropylamide
DMF	dimethylformamide		LFQ	label-free quantification
LTQ	linear trap quadrupole		SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

m/z	mass to charge ratio		sec	second
*m*CPBA	meta-chloroperoxybenzoic acid		SEC	size-exclusion chromatography
min	minute		SHMT	serine hydroxymethyltransferase
MOM	methoxymethyl		SILAC	stable isotope labeling by amino acids in cell culture
MS	mass spectrometry		SPAAC	strain-promoted azide-alkyne cycloaddition
MS/MS	tandem mass spectrometry		SSM	secondary structure matching
MW	molecular weight		TBS, TBDMS	tert-butyldimethylsilyl
MWCO	molecular weight cut-off		TCEP	tris(2-carboxyethyl)phosphine
NaAsc	sodium ascorbate		TEAB	triethylammonium bicarbonate
NaBH$_4$	sodium borohydride		TEV	tobacco etch virus
NAD, NADH	nicotinamide adenine dinucleotide		TFA	trifluoroacetic acid
NADP, NADPH	Nicotinamide adenine dinucleotide phosphate		TFAA	trifluoroacetic anhydride
NMR	Nuclear magnetic resonance		THF	tetrahydrofuran
OAS	*O*-acetylserine		TLC	thin-layer chromatography
OD	optical density		TMS	trimethylsilyl
OKR-DC	ornithine/lysine/arginine decarboxylase		TMT	tandem mass tag
PBS	phosphate-buffered saline		Tn	transposon
PCR	polymerase chain reaction		TOP-ABPP	tandem-orthogonal proteolysis
PDB	protein data bank		Tris	tris(hydroxymethyl)aminomethane
PEG	polyethyleneglycol		UDP	uridine diphosphate
PEP	posterior error probability		UPF	uncharacterized protein family
PLK	pyridoxal kinase		UV	ultraviolet
PLP	pyridoxal phosphate			
PLP-DE	pyridoxal phosphate-dependent enzyme			
PMA	phosphomolybdic acid			
PMB	*para*-methoxybenzyl			
PMP	pyridoxamine phosphate			
PNP	pyridoxine phosphate			
PNPOx	pyridoxine-5'-phosphate oxidase			
Ppase	phosphatase			
PPAT	phosphopantothenoylcysteine			
ppm	parts per million			
PPTS	pyridinium *p*-toluenesulfonate			
PROSC	proline synthetase co-transcribed protein			
PTMs	post-translational modifications			
R5P	D-ribose 5-phosphate			
Rh	rhodamine			
rpm	revolutions per minute			
rt	room temperature			
SAM	S-adenosyl methionine			
SAR	structure-activity relationships			
SDS	sodium dodecyl sulfate			

Supplementary gel images

Chapter 3

Supplementary gels for Figure 3.3e
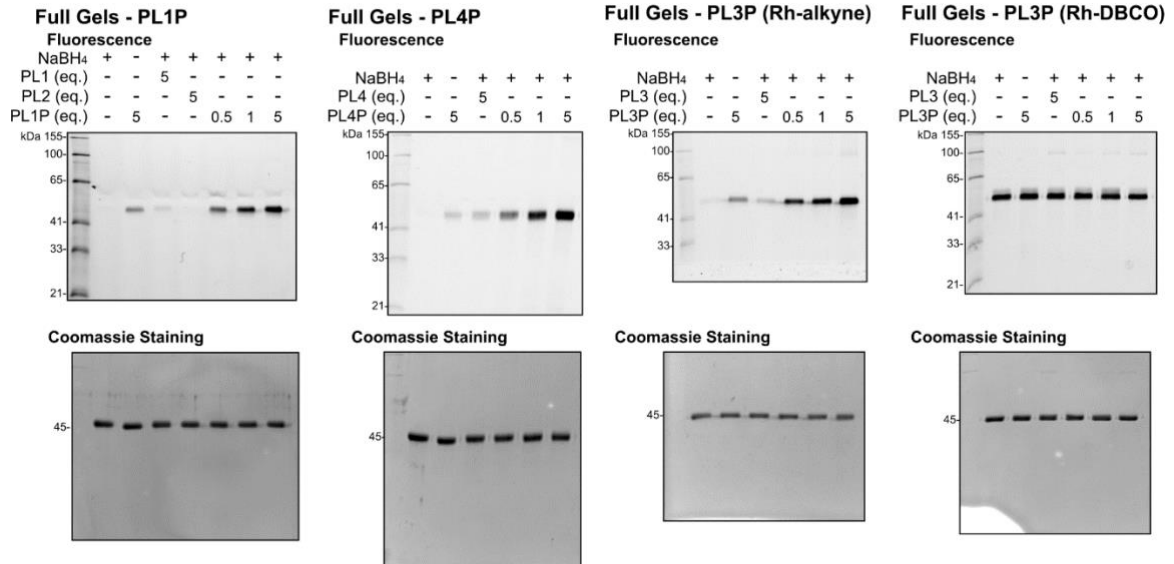
Supplementary gels for Figure 3.6a
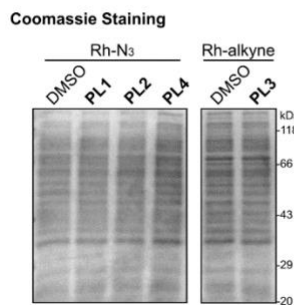
Supplementary gels for Figure 3.7
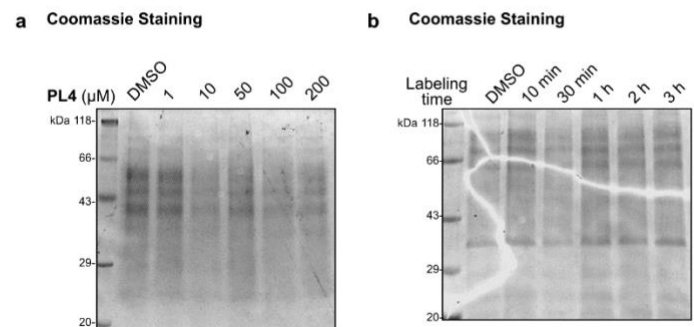
Chapter 4

Supplementary gels for Figures 5.1 and 5.2 (known PLP-DEs)

Full Gels - DAT

Full Gels - CS
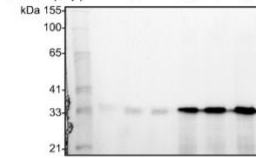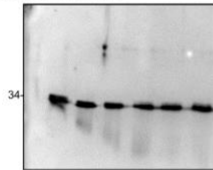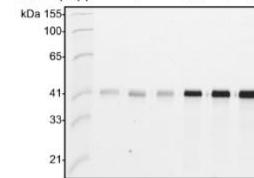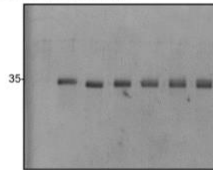
Full Gels - OKR-DC

Supplementary gels for Figure 5.3 (uncharacterized proteins)

Full Gels - Q2FF14

Full Gels - GP0
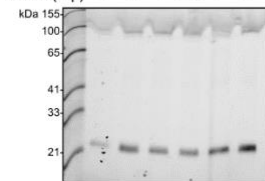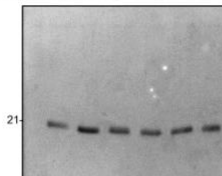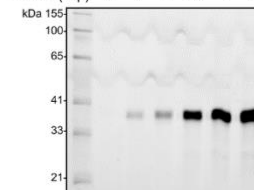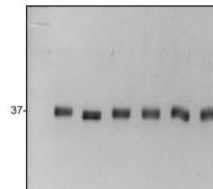
Full Gels - HH8

Supplementary gels for Figure 5.9 (alternative-function proteins)

Full Gels - HemQ

Full Gels - PsuG
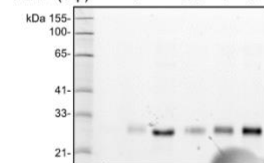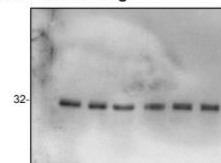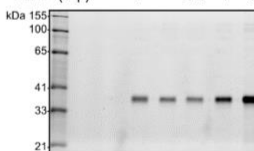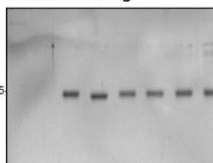
Full Gels - HemH

MS/MS spectra for PLP binding site identification

Supplementary information for PLP binding-site identification. MS/MS spectra of binding site peptides as exported from MaxQuant Viewer.

PLP binding site for AT1 (A0A0H2XFY9), K233

PLP binding site for AT2 (A0A0H2XFA8), K232
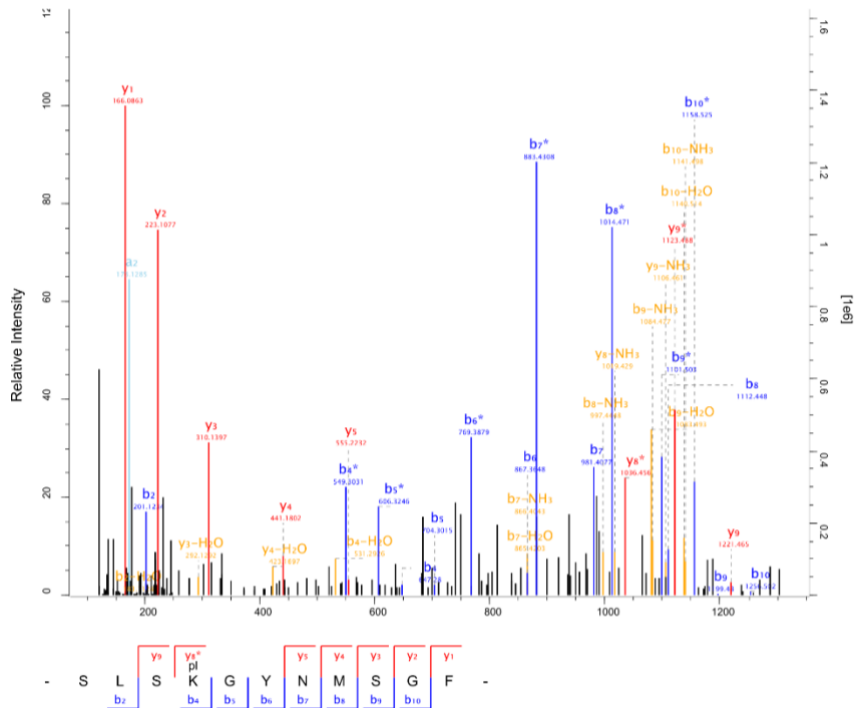
PLP binding site for CS (A0A0H2XFQ3), K46
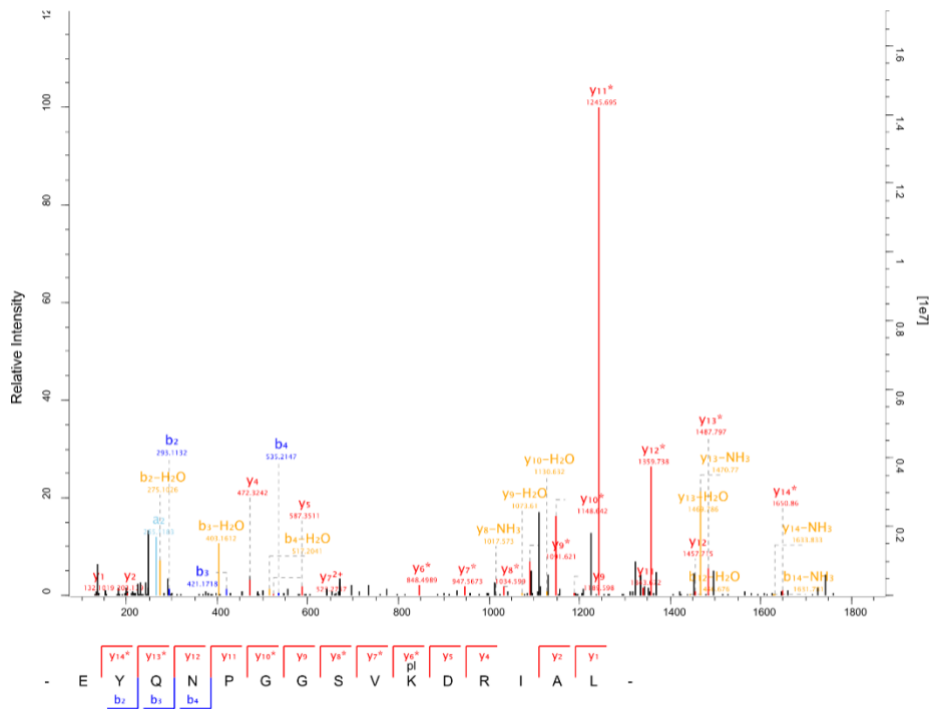
PLP binding site for DAT (A0A0H2XHU6), K157
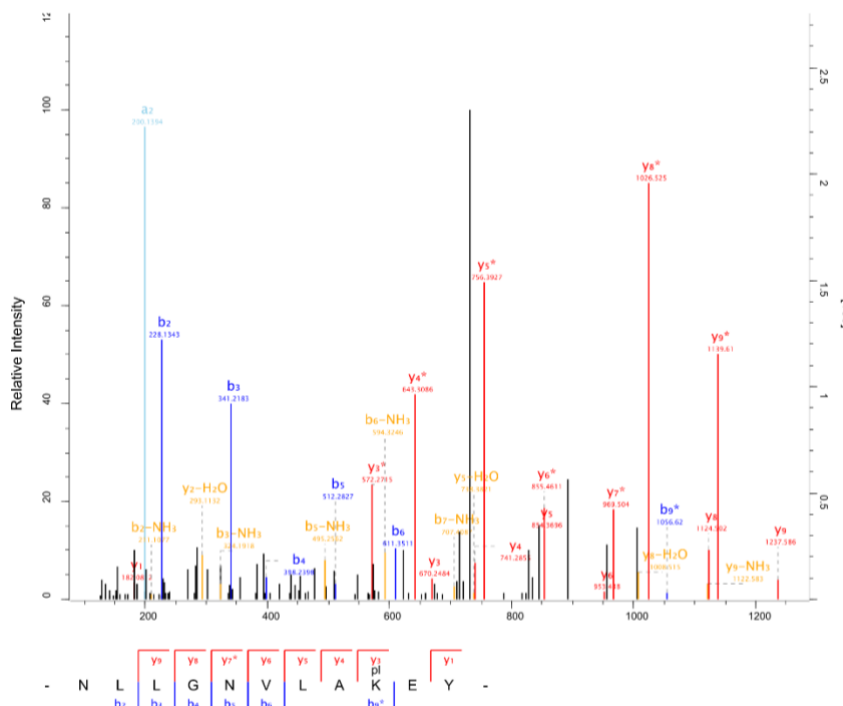
PLP binding site for GP0 (A0A0H2XGP0), K172
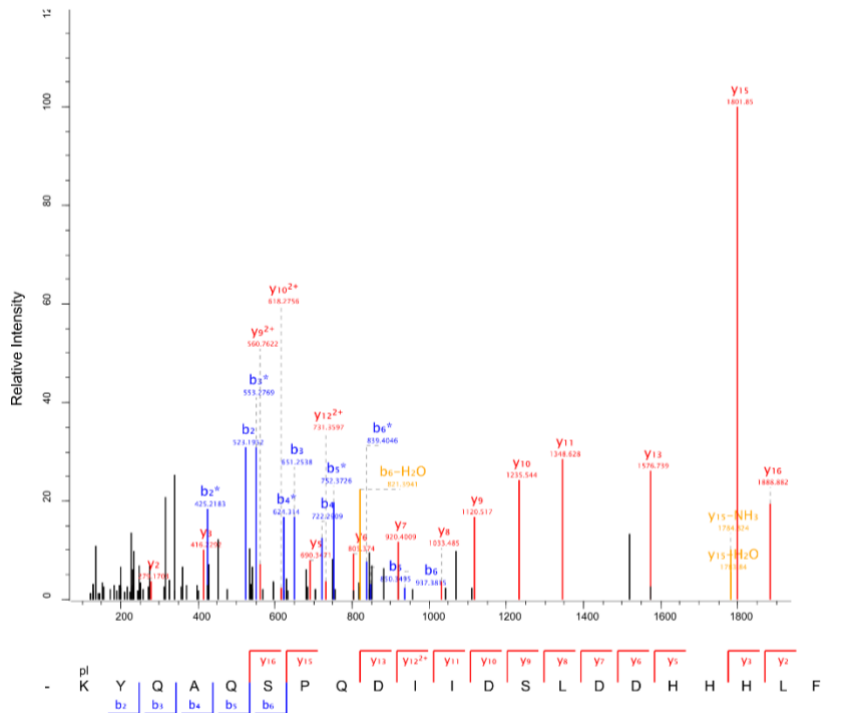
PLP binding site for HH8 (A0A0H2XHH8), K32
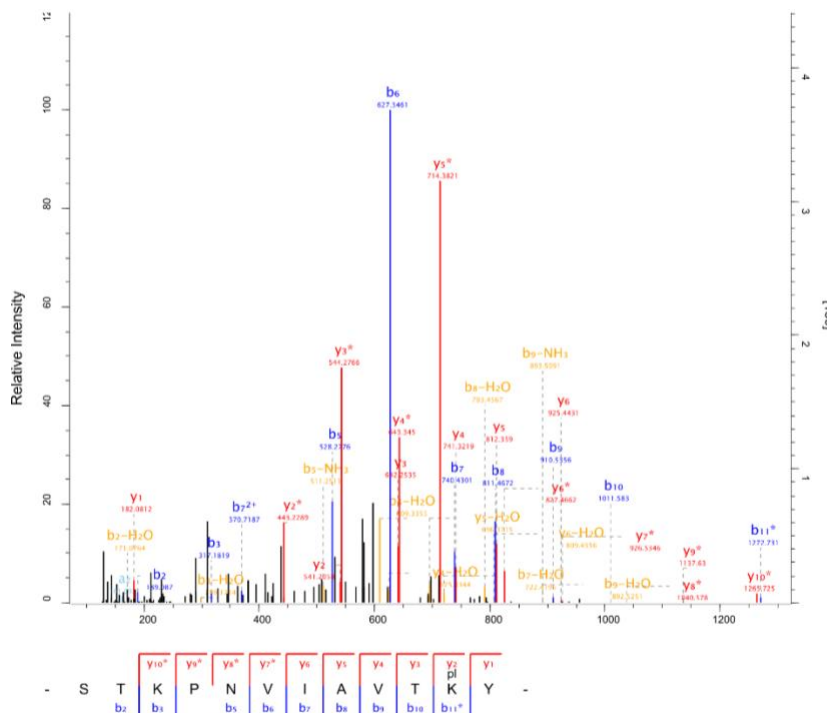
PLP binding site for OKR-DC (A0A0H2XII6), K210
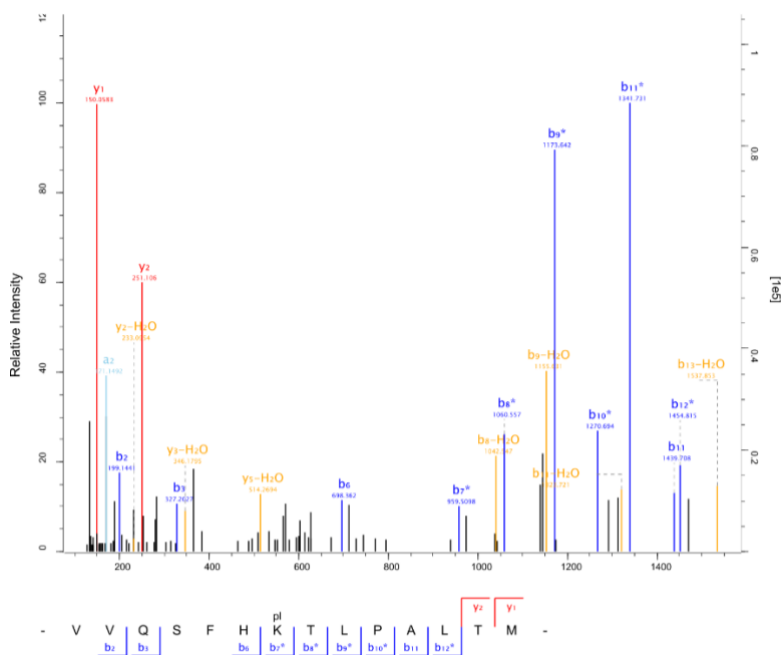
PLP binding site for Q2FF14, K58
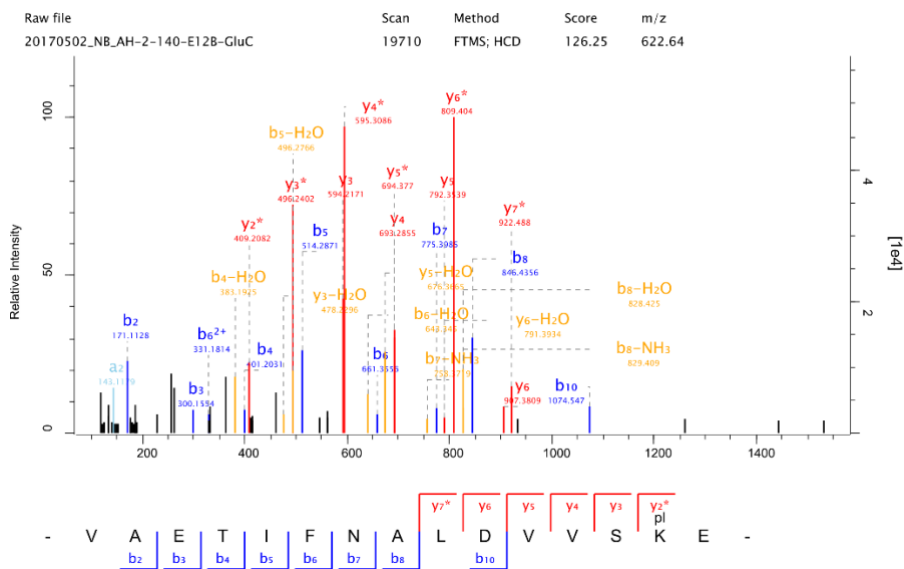
NMR Spectra

^{1}H NMR of compound **1** (360 MHz, CDCl$_3$)
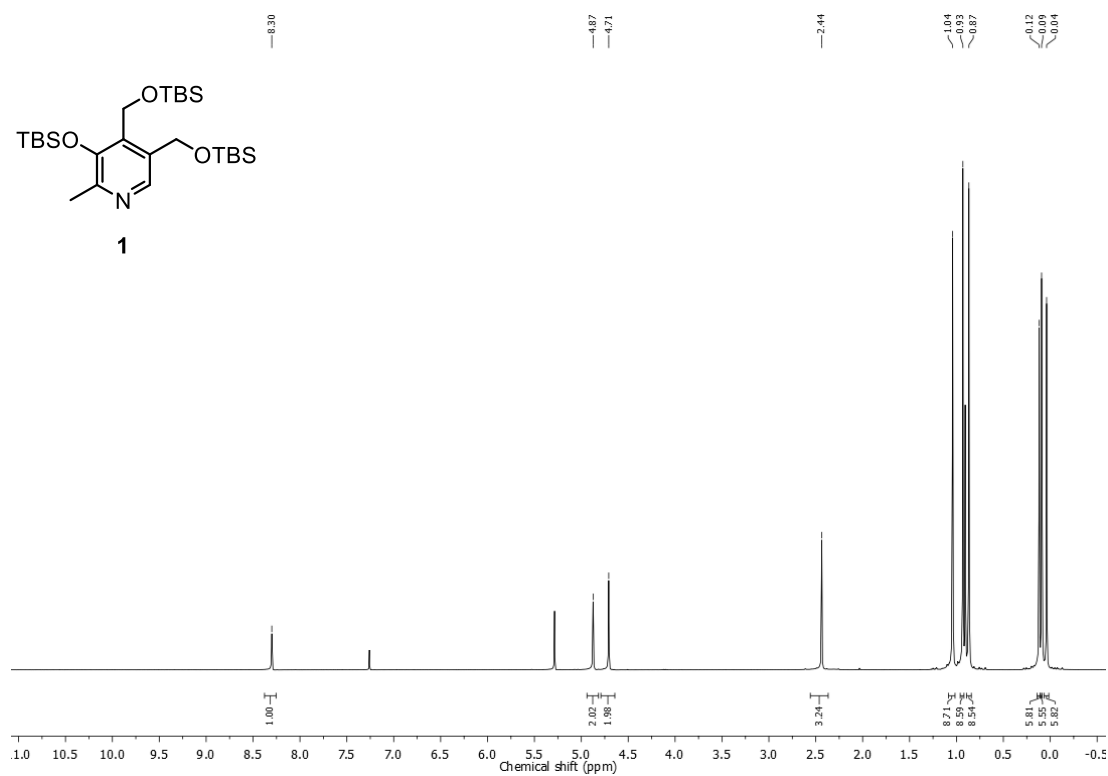
^{13}C NMR of compound **1** (90 MHz, CDCl$_3$)
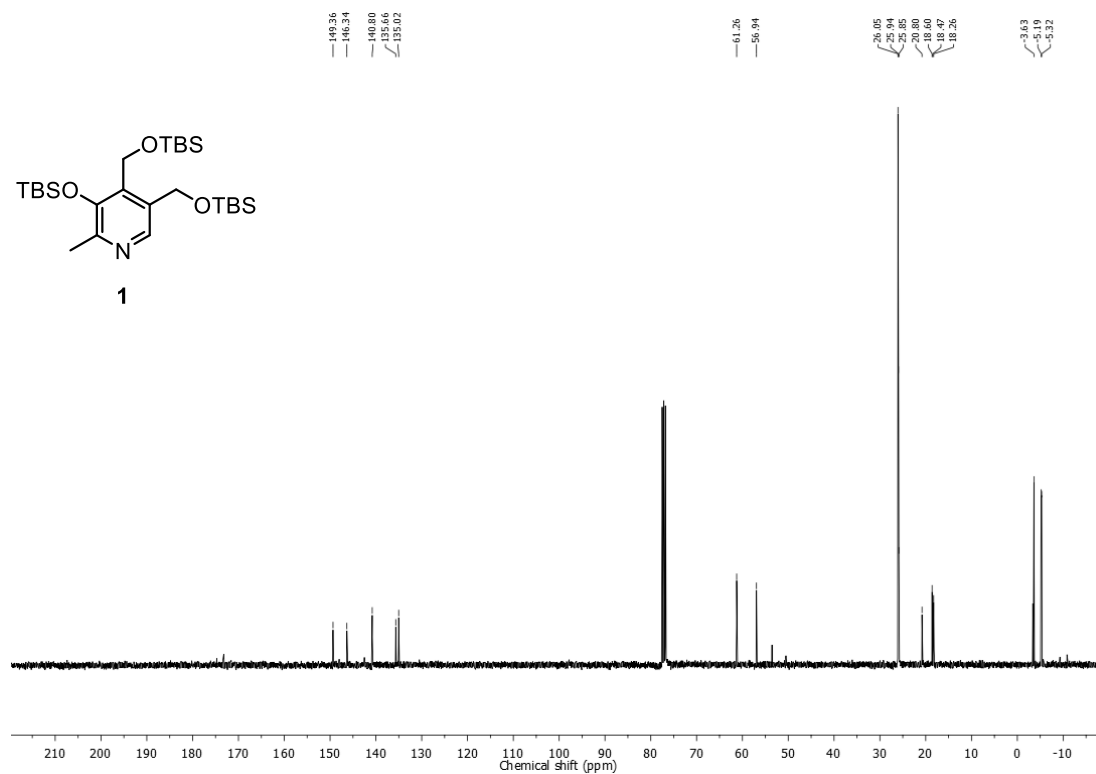
¹H NMR of compound **2** (360 MHz, CDCl₃)

9.16

7.93

5.12
4.75
4.59

0.95
0.89

0.18
0.07

0.88

1.00

1.92
1.86
1.85

8.56
8.99

5.51
5.72

1.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.

Chemical shift (ppm)
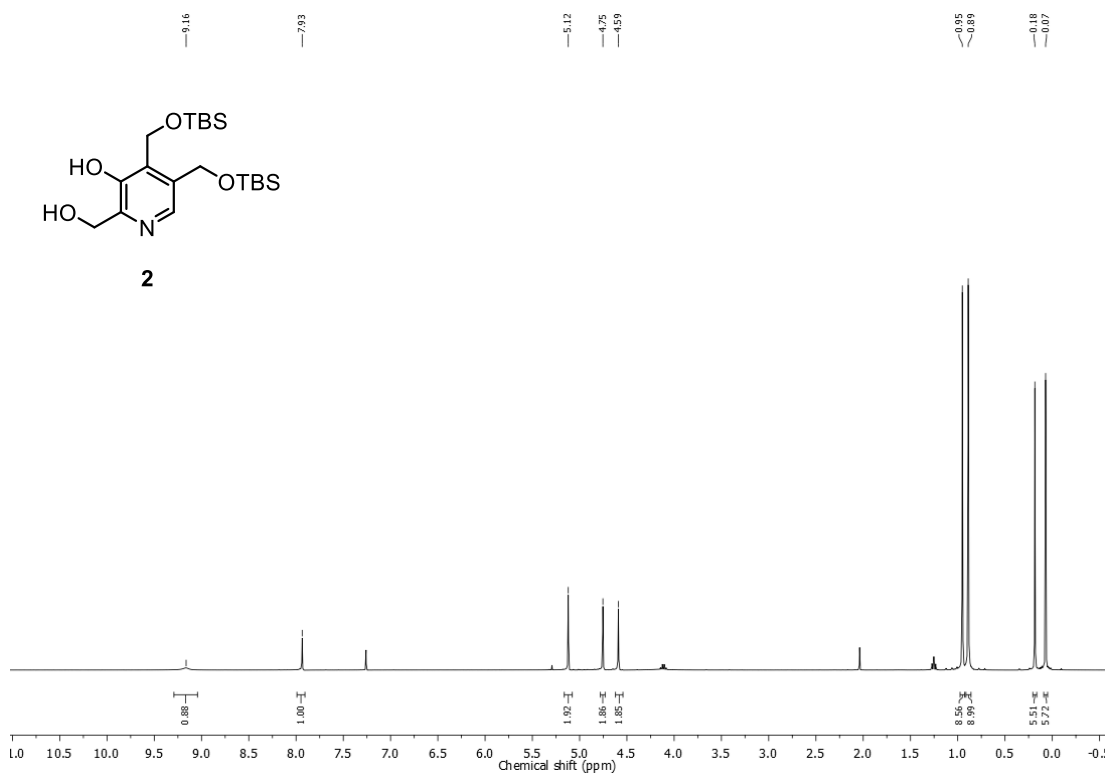
¹³C NMR of compound **2** (90 MHz, CDCl₃)

150.14
147.79

138.26

131.33
129.23

63.05
61.70
60.53

25.95
25.84

18.34
18.27

5.15
5.51

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10

Chemical shift (ppm)
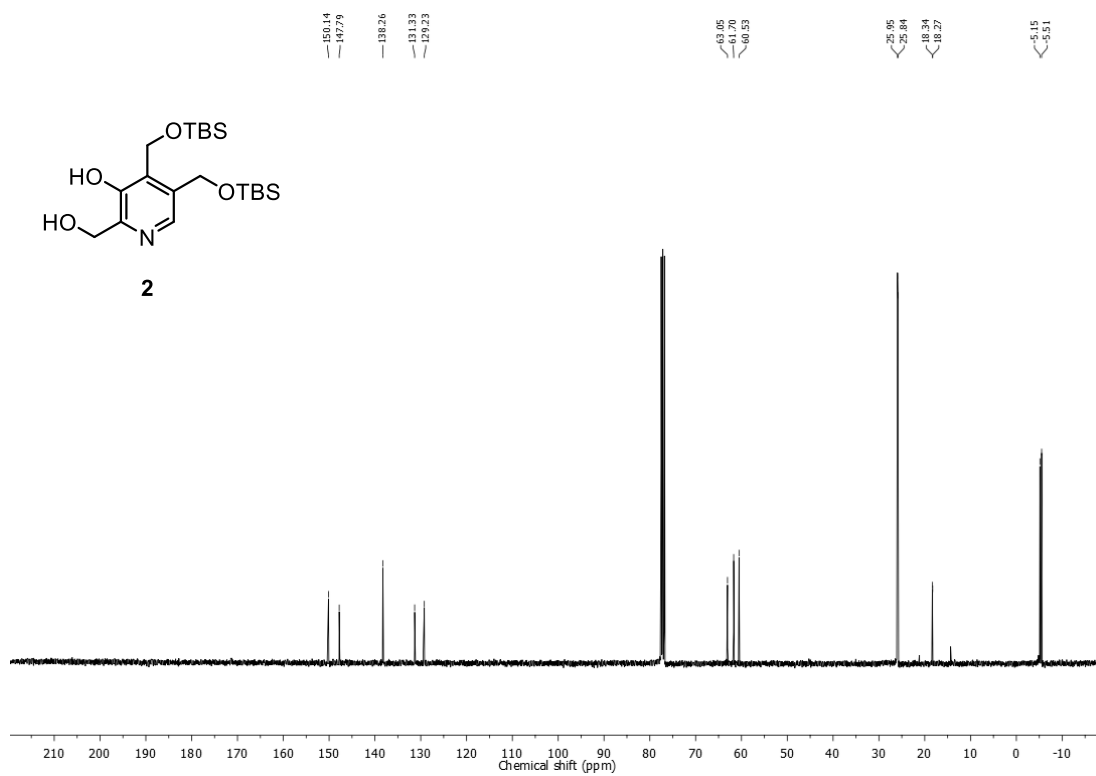
^{1}H NMR of compound **3** (360 MHz, CDCl$_3$)

11.03
10.04
8.59
5.00
4.81
0.96
0.89
0.13
0.09

1.00
0.98
0.97
2.09
2.00
9.31
9.50
5.71
6.29

Chemical shift (ppm)
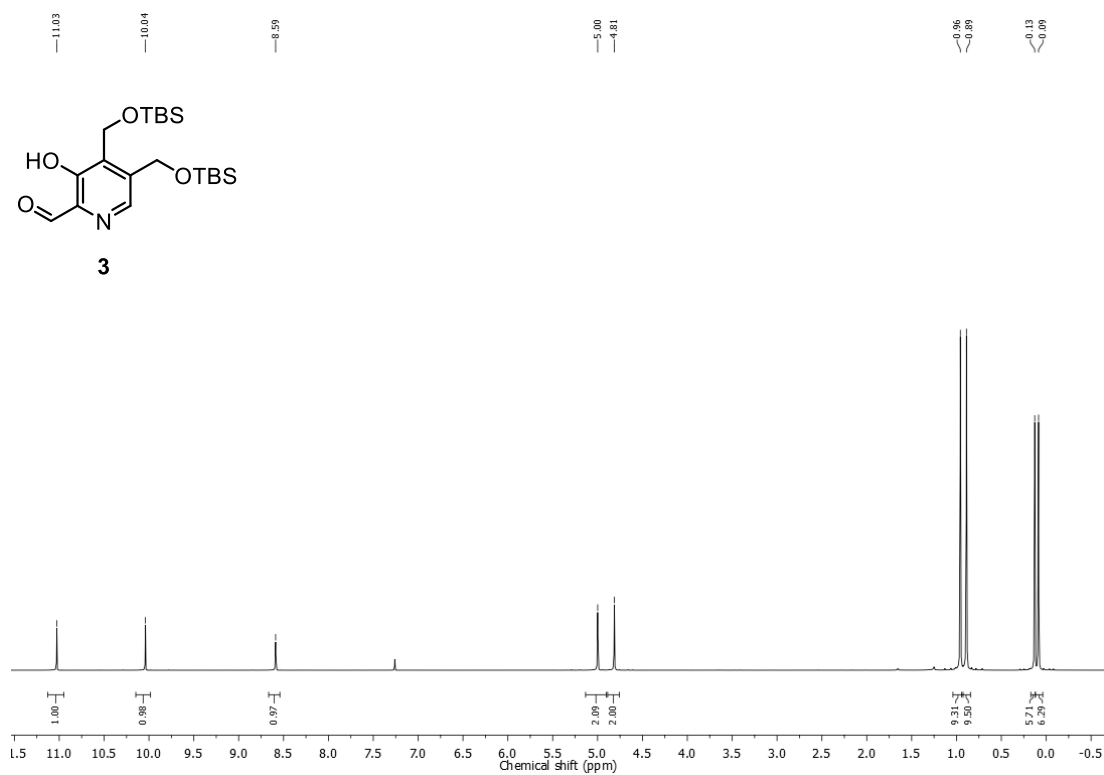
^{13}C NMR of compound **3** (90 MHz, CDCl$_3$)

198.69
155.63
143.89
141.55
135.53
133.62
61.04
54.65
26.01
25.95
18.49
18.37
-5.26
-5.33

Chemical shift (ppm)
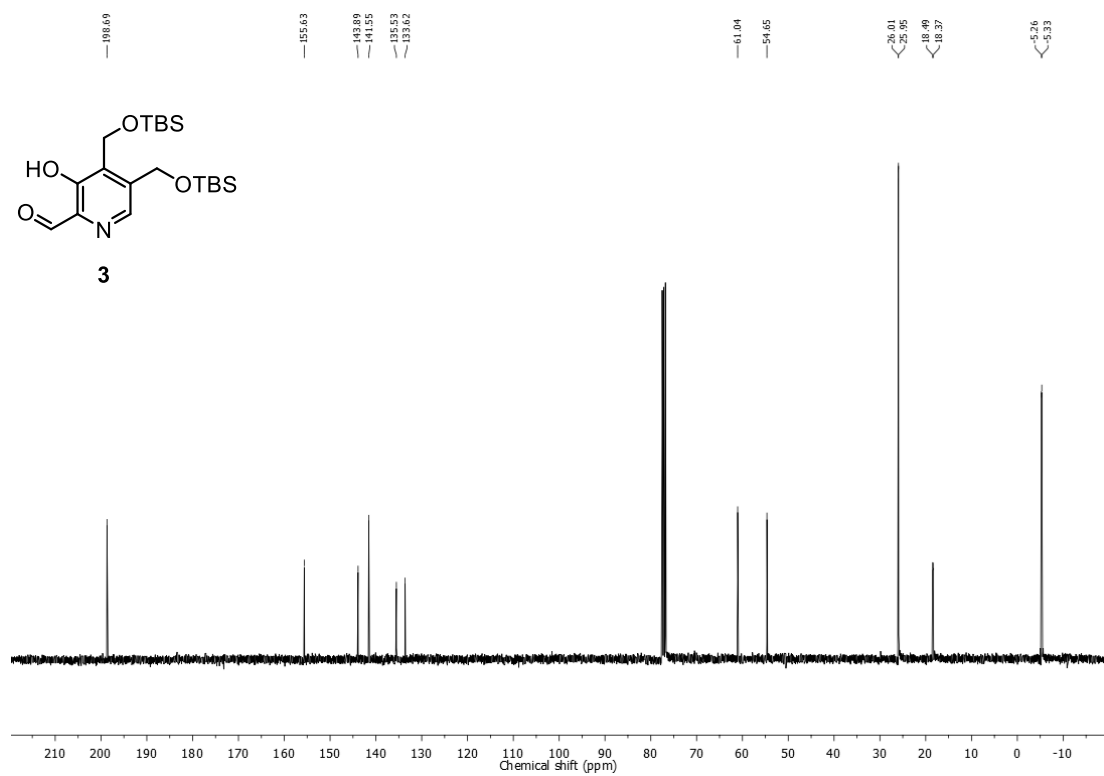
¹H NMR of compound **4** (360 MHz, CDCl₃)
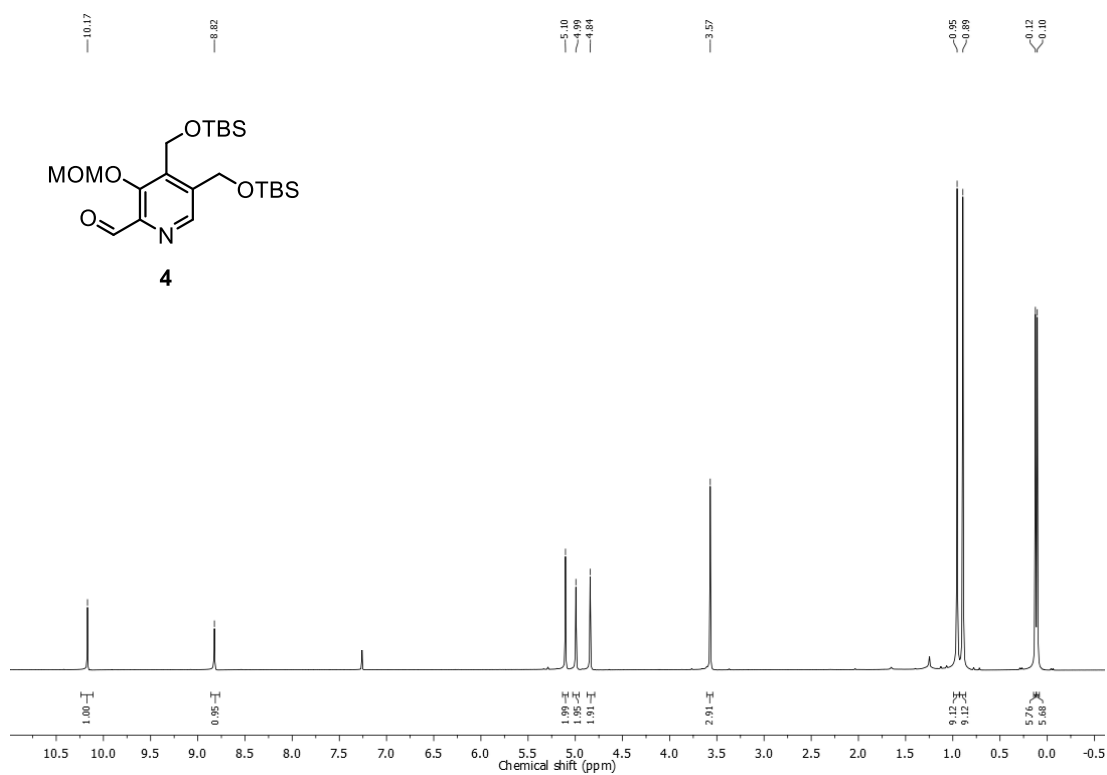
¹³C NMR of compound **4** (90 MHz, CDCl₃)
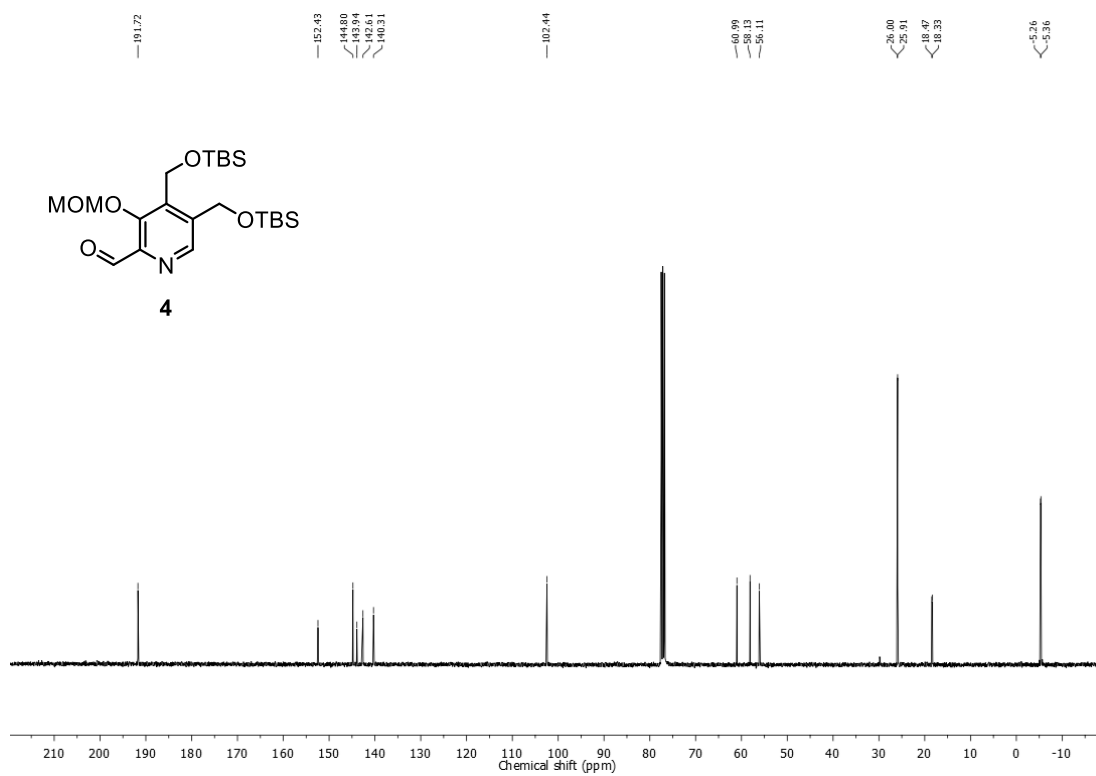
^{1}H NMR of compound **5** (360 MHz, CDCl$_3$)

8.55

5.20
4.92
4.79

3.58
3.31

0.93
0.88

0.11
0.09

1.10

2.17
2.10
2.02

3.05
1.00

11.09
11.23

15.24

11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
Chemical shift (ppm)
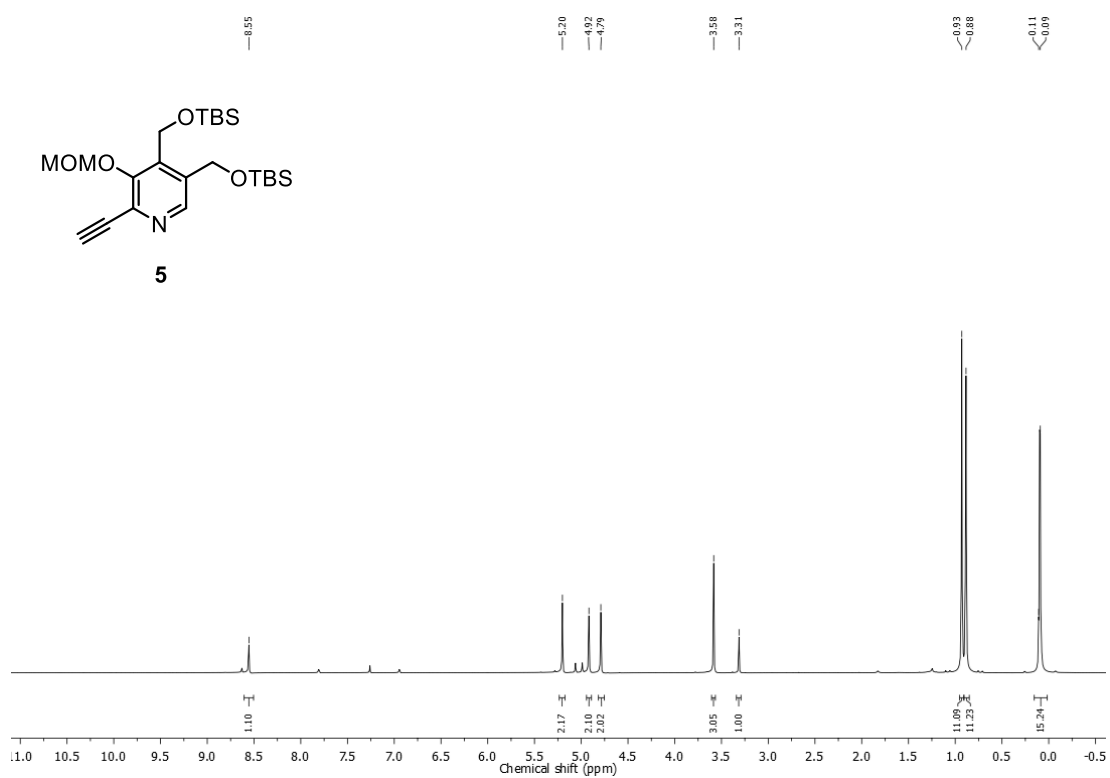
^{13}C NMR of compound **5** (90 MHz, CDCl$_3$)

153.08

144.69

138.82
137.94
135.26

100.72

80.79

60.85
58.08
56.32

26.04
25.99
25.91

18.45
18.32

-5.25
-5.37

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
Chemical shift (ppm)
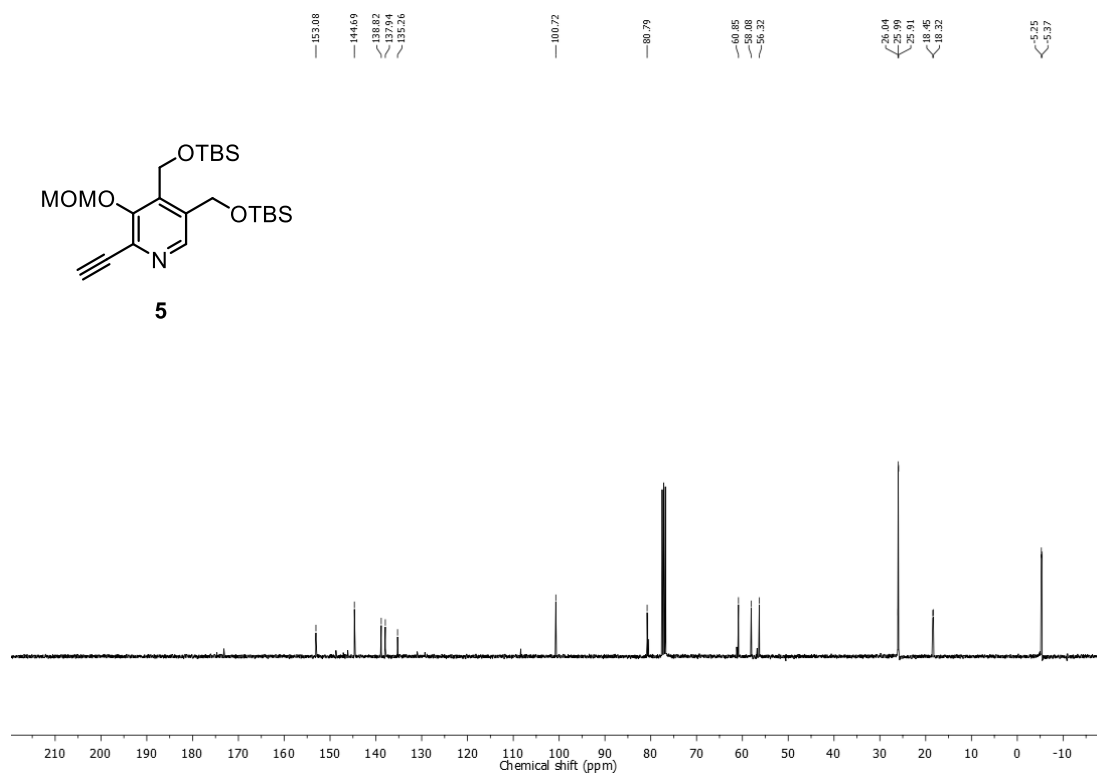
¹H NMR of compound **6** (360 MHz, MeOD)

8.38

5.28
4.85
4.80
4.01
3.63

1.00   2.02   1.94   0.97   2.85
1.79

| 1.0 | 10.5 | 10.0 | 9.5 | 9.0 | 8.5 | 8.0 | 7.5 | 7.0 | 6.5 | 6.0 | 5.5 | 5.0 | 4.5 | 4.0 | 3.5 | 3.0 | 2.5 | 2.0 | 1.5 | 1.0 | 0.5 | 0.0 | -0.5 |

Chemical shift (ppm)
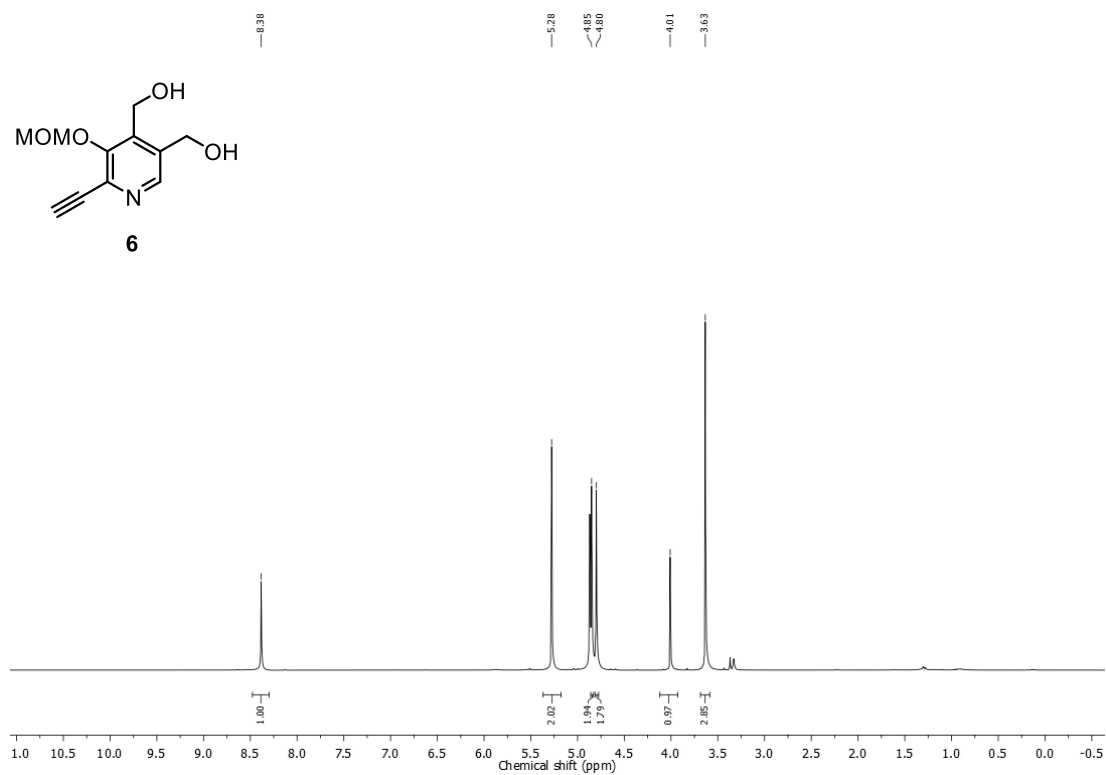
¹³C NMR of compound **6** (90 MHz, MeOD)

155.51

145.81
142.81
139.36
136.95

101.70

84.09
80.62

60.39
58.34
55.66

| 20 | 210 | 200 | 190 | 180 | 170 | 160 | 150 | 140 | 130 | 120 | 110 | 100 | 90 | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 | 0 | -10 |

Chemical shift (ppm)
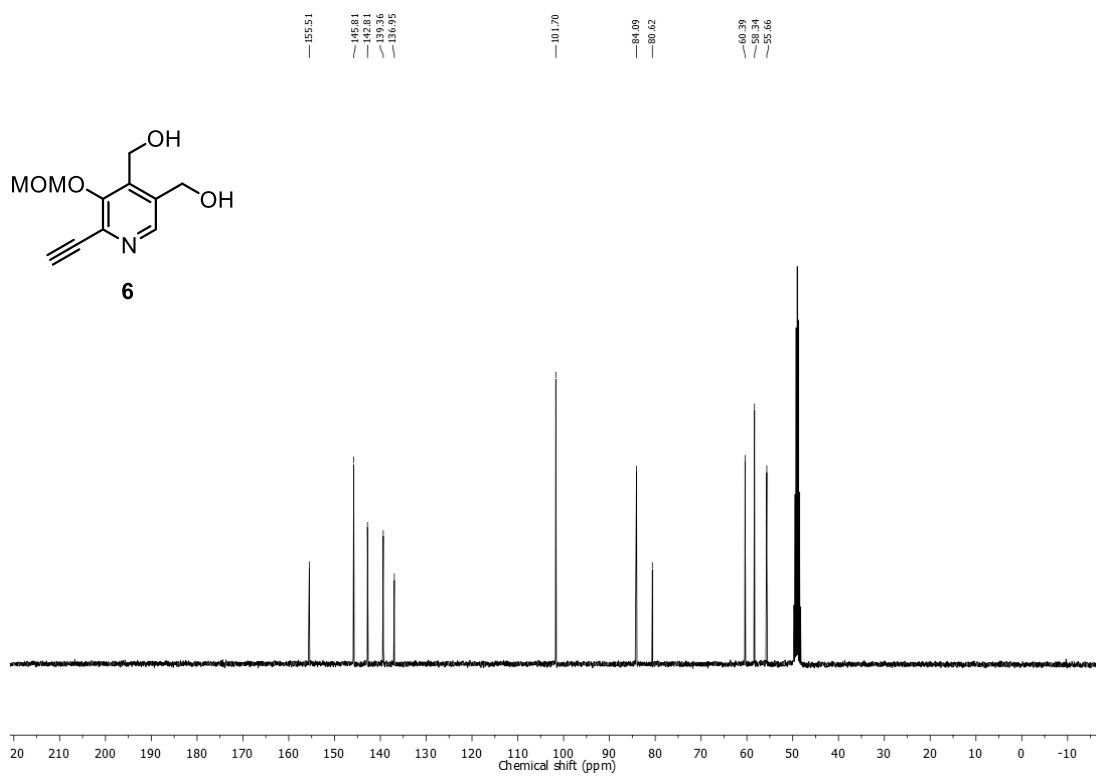
1H NMR of compound **7a** (500 MHz, MeOD)
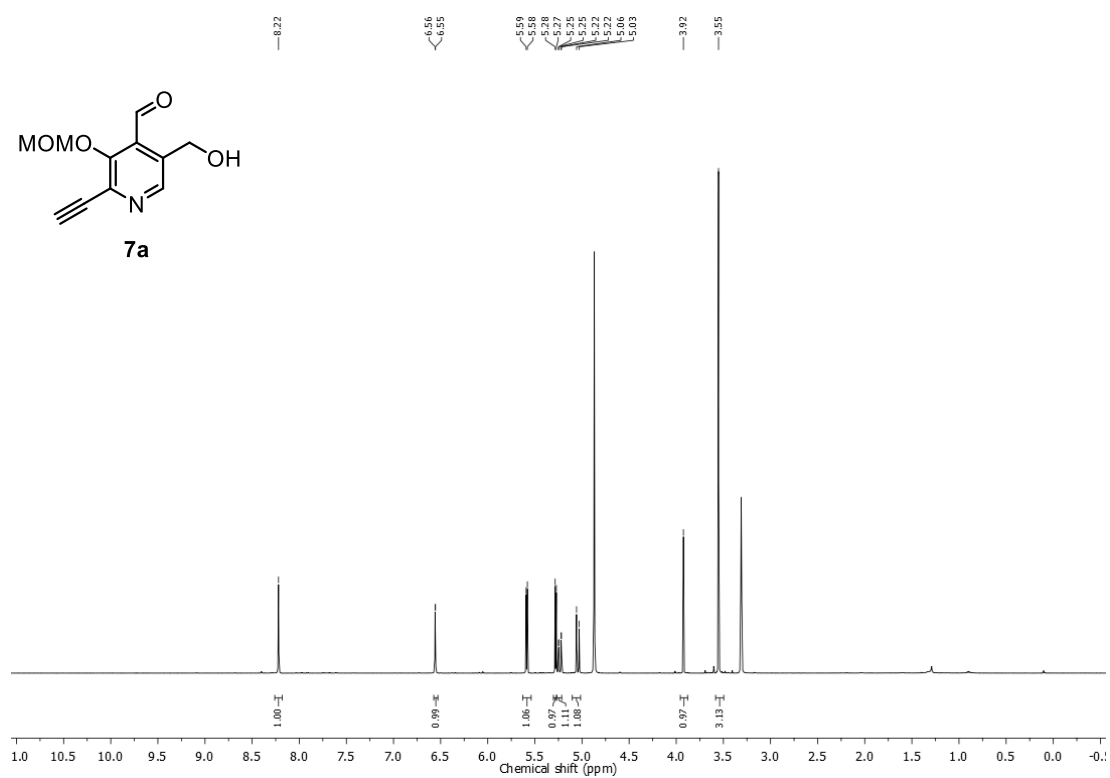
13C NMR of compound **7a** (120 MHz, MeOD)
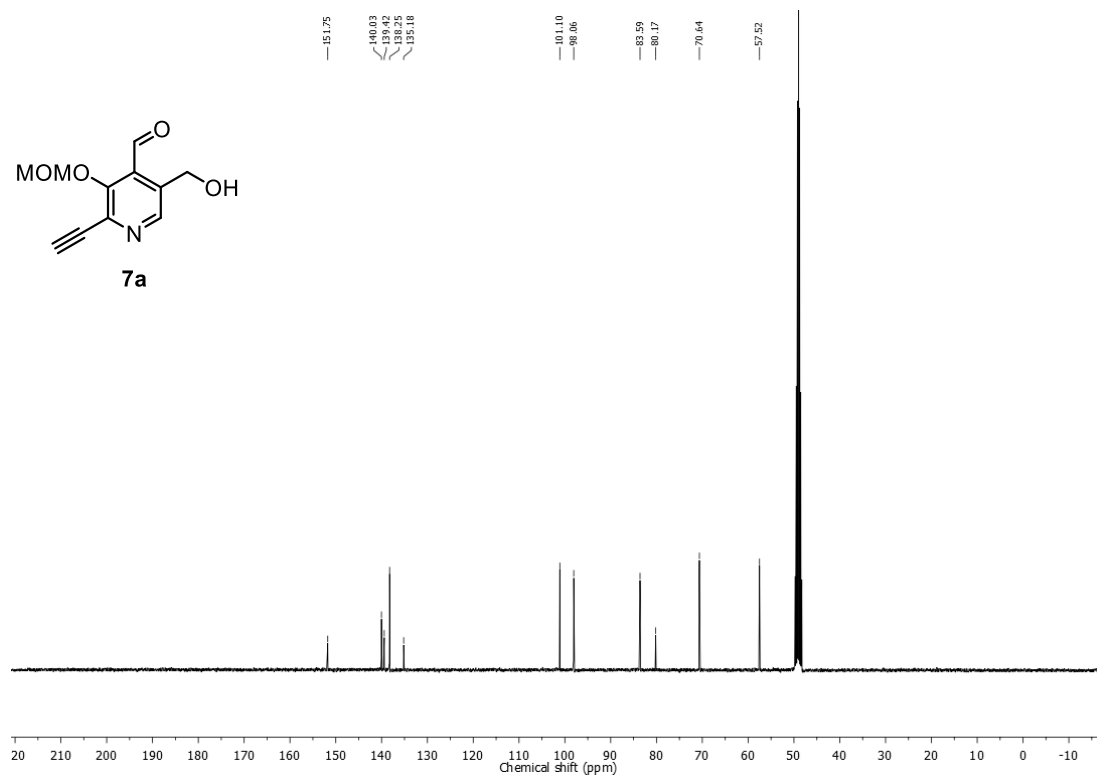
HSQC spectrum of compound **7a**
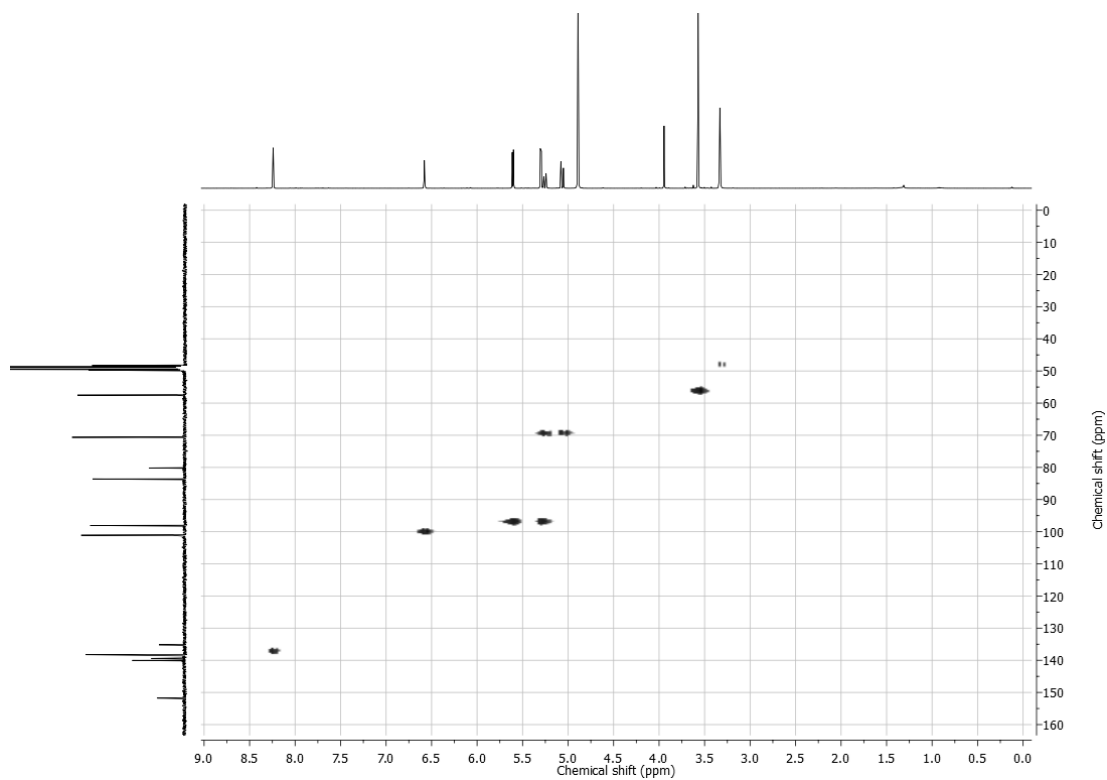
HMBC spectrum of compound **7a**
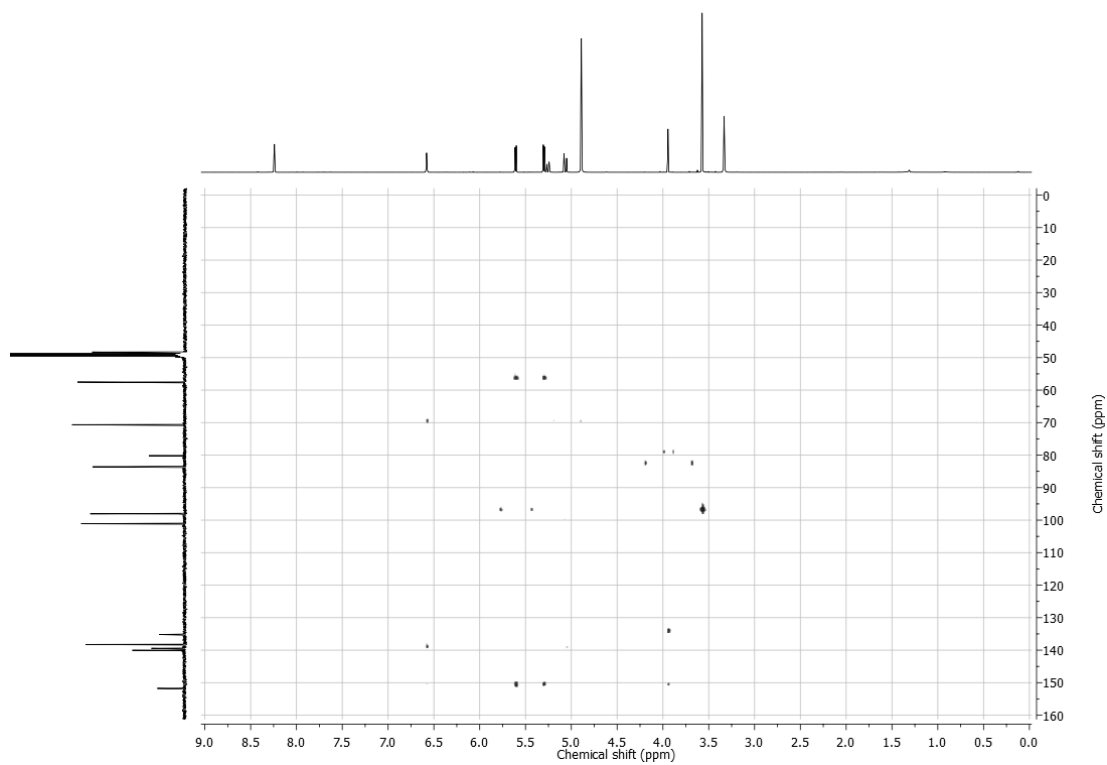
NOESY spectrum of compound **7a**
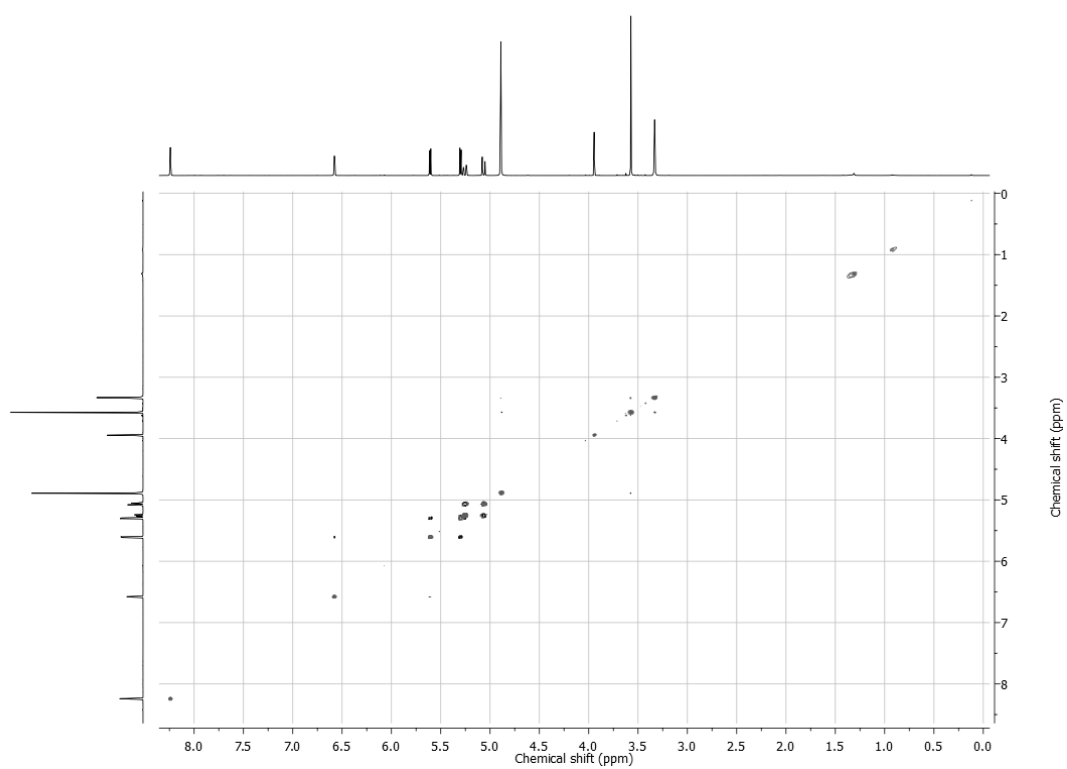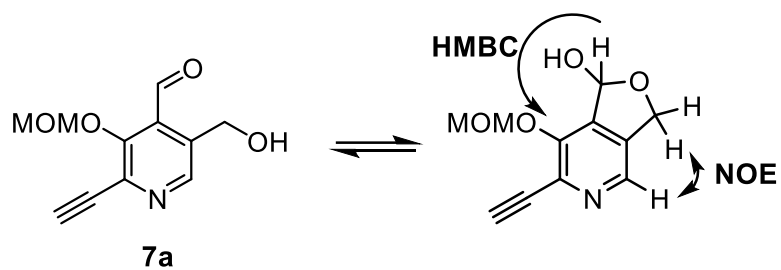
¹H NMR of compound **7b** (500 MHz, MeOD)
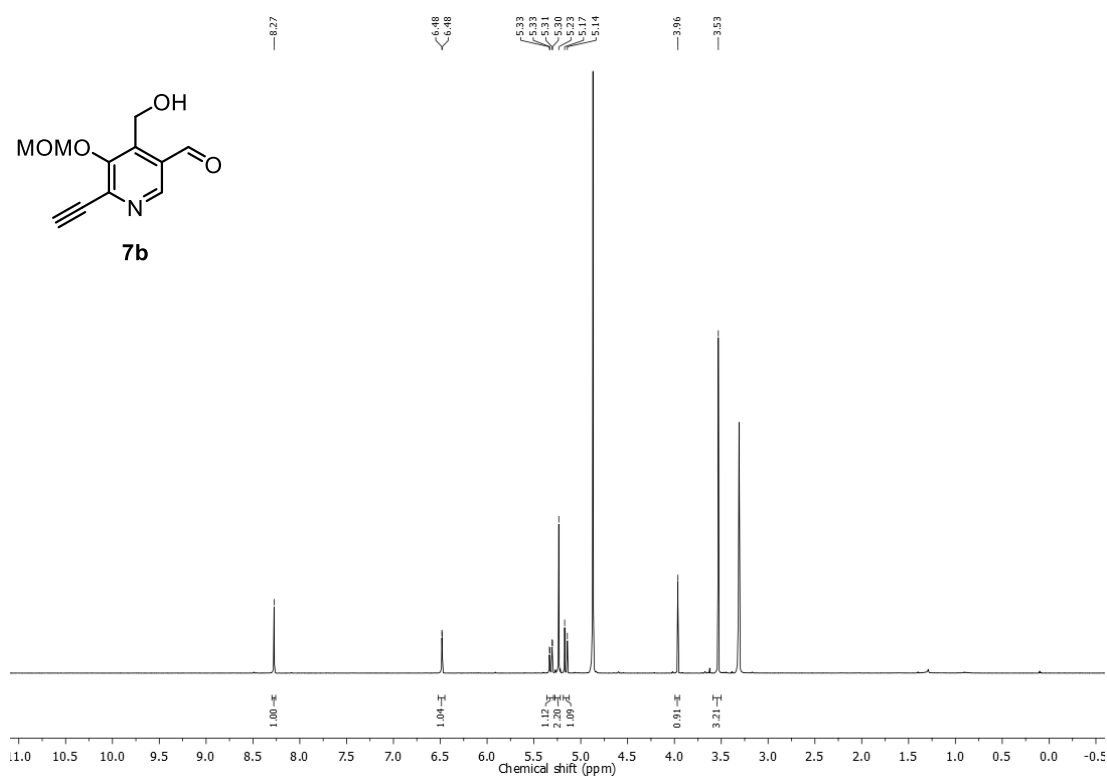
¹³C NMR of compound **7b** (120 MHz, MeOD)
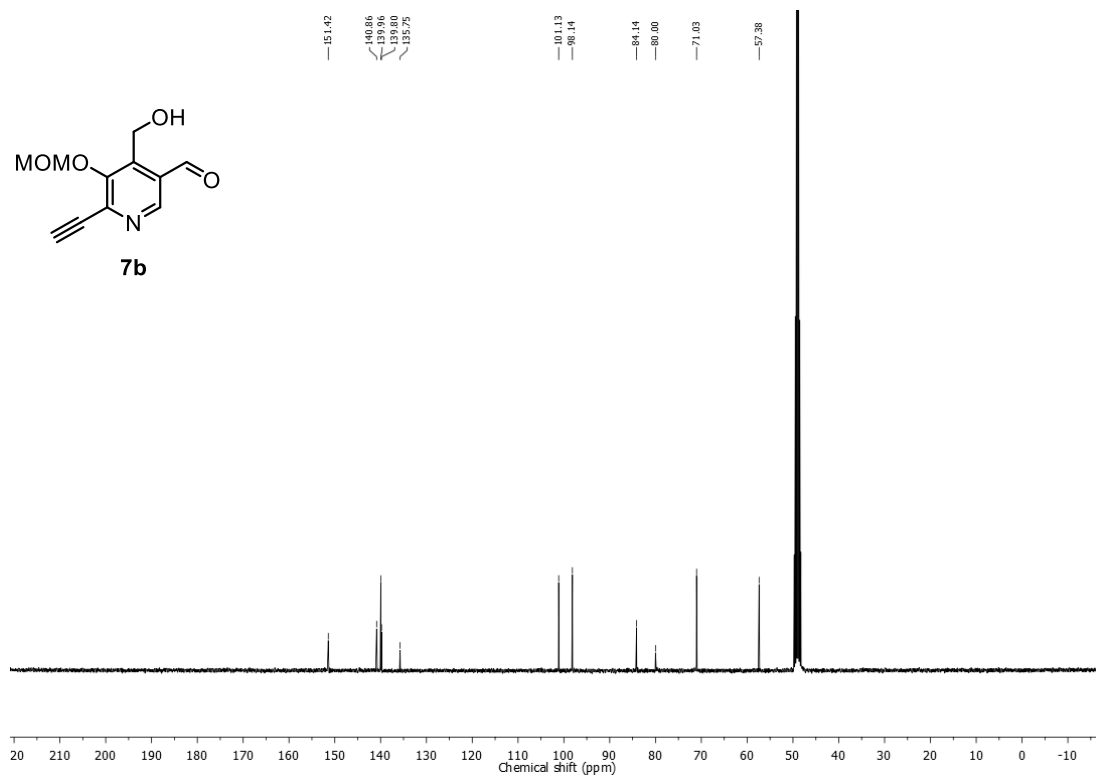
HSQC spectrum of compound **7b**
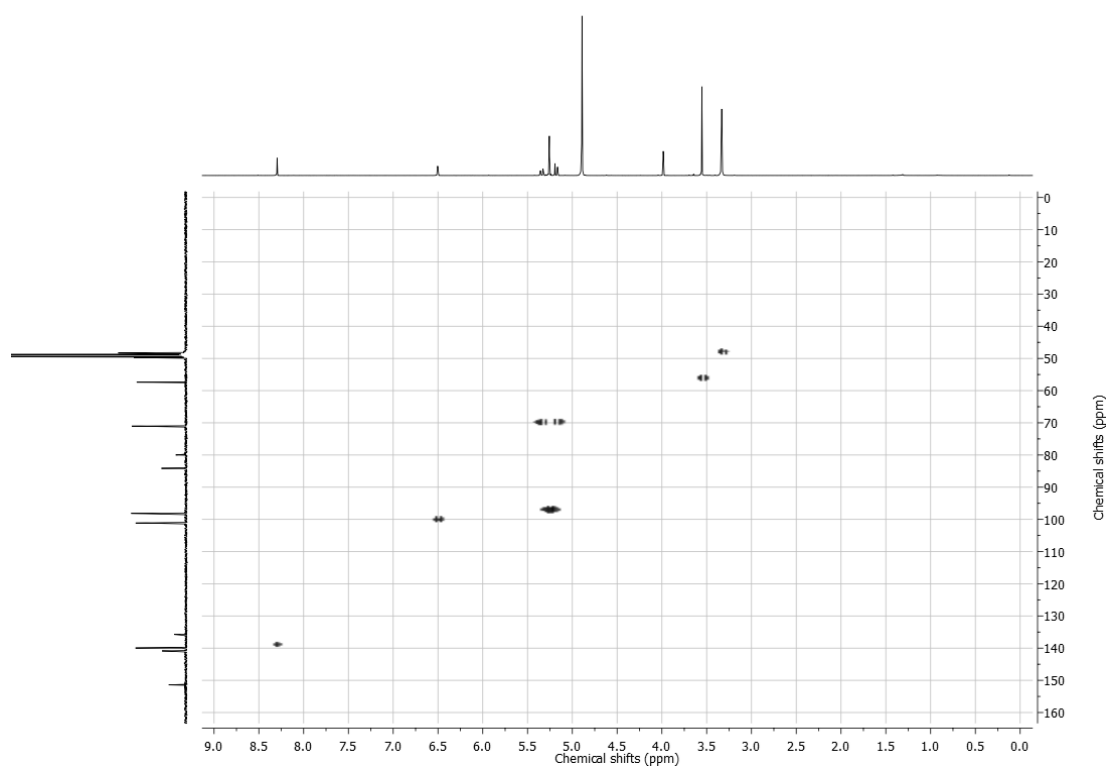
HMBC spectrum of compound **7b**
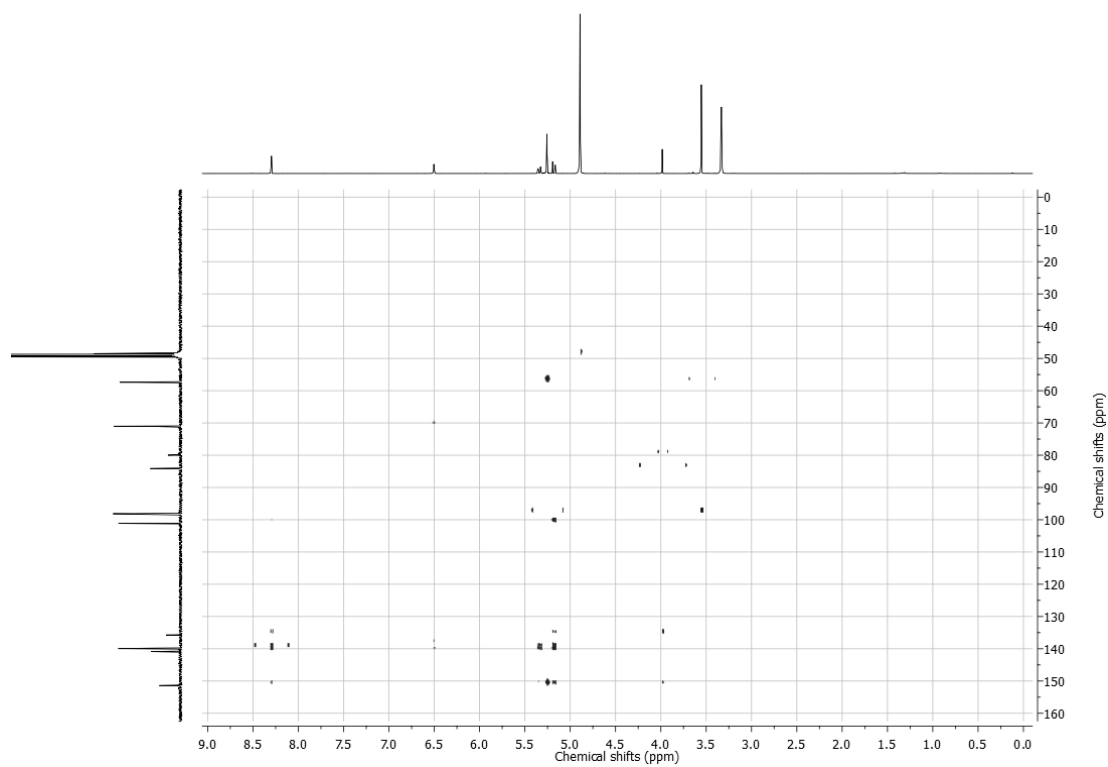
NOESY spectrum of compound **7b**
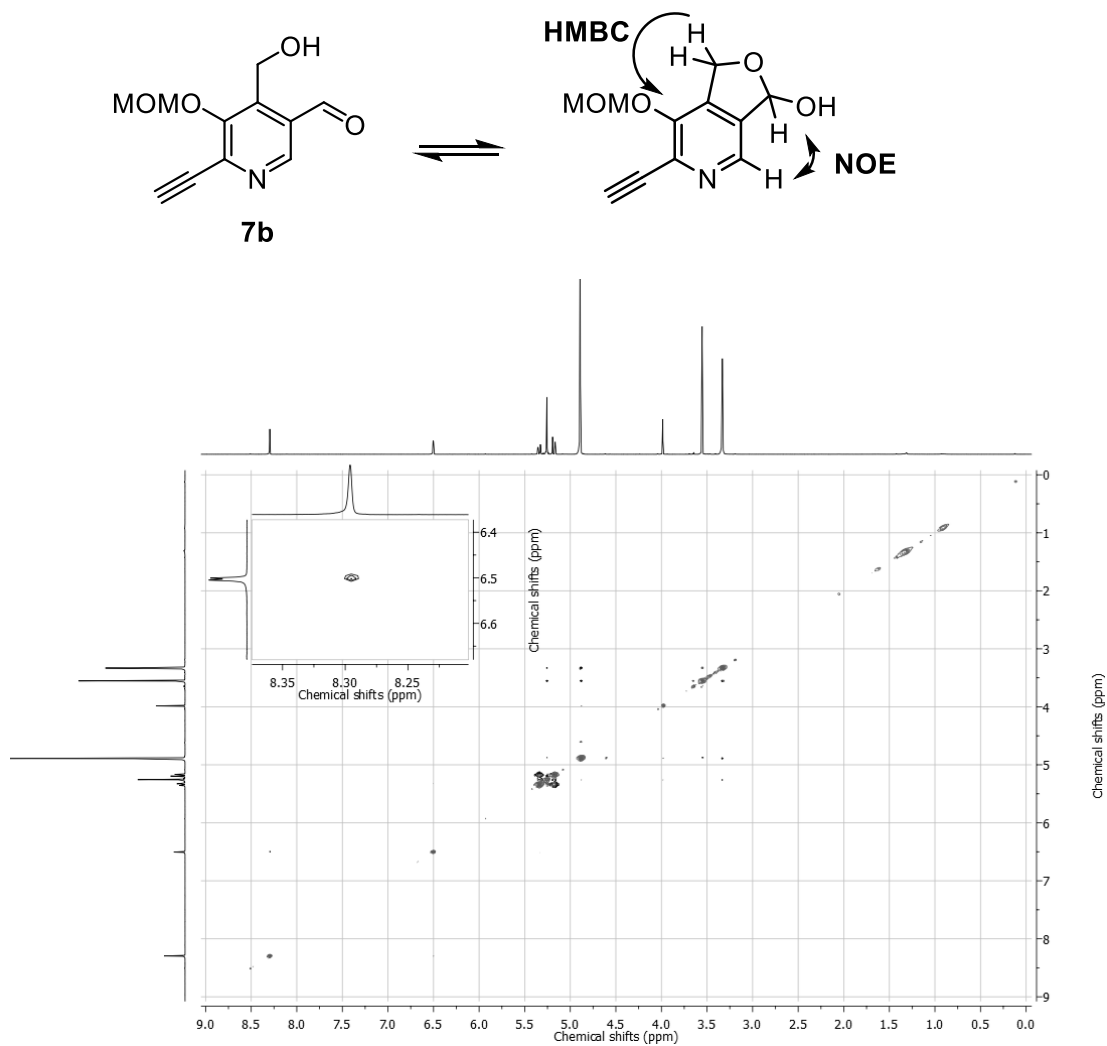
¹H NMR of compound **PL1** (500 MHz, DMSO-d₆)
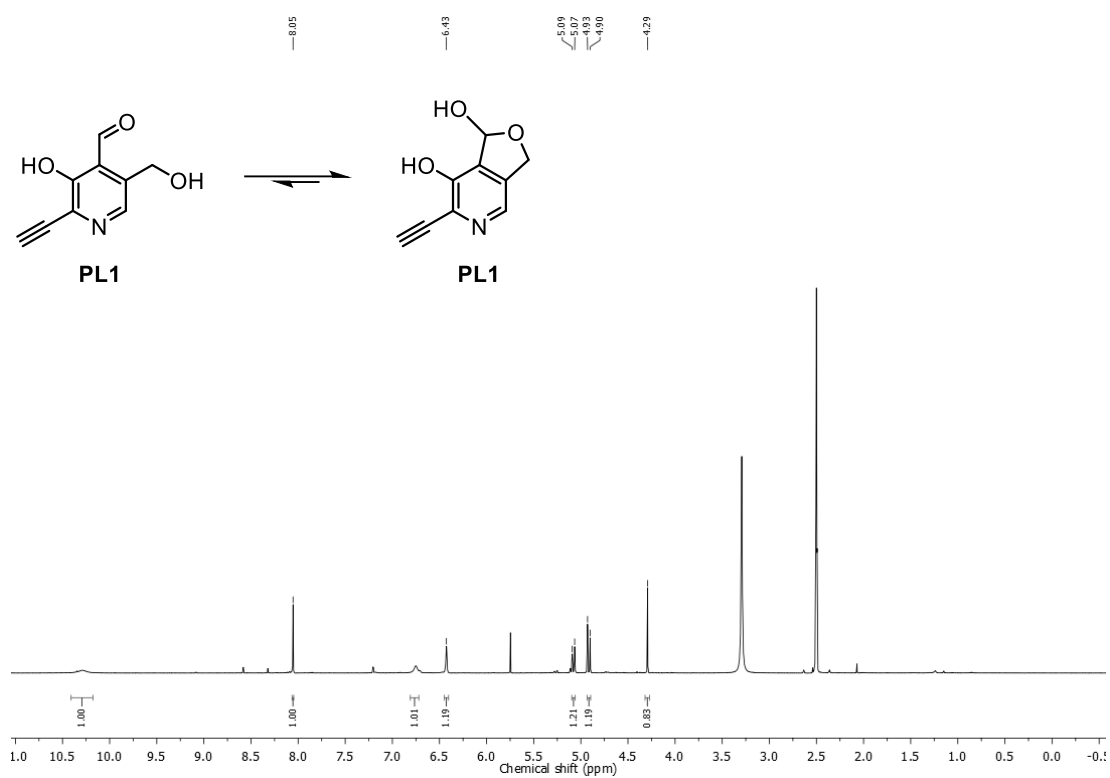
¹³C NMR of compound **PL1** (125 MHz, DMSO-d₆)
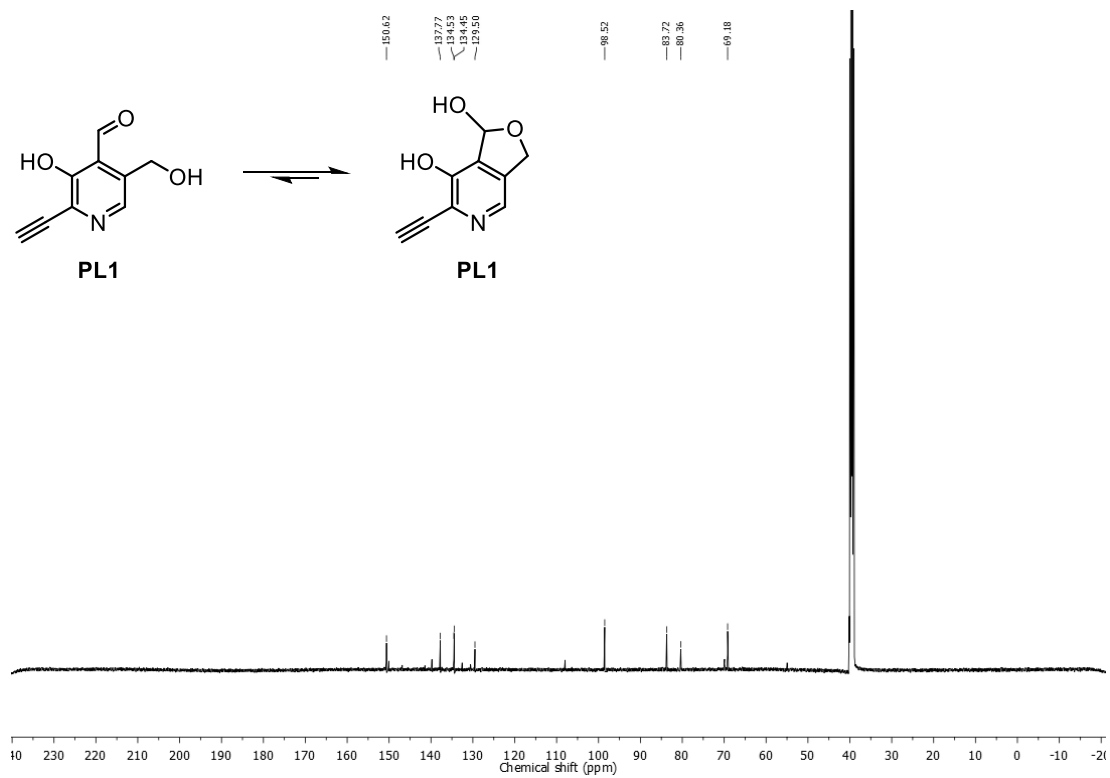
HSQC spectrum of compound **PL1**
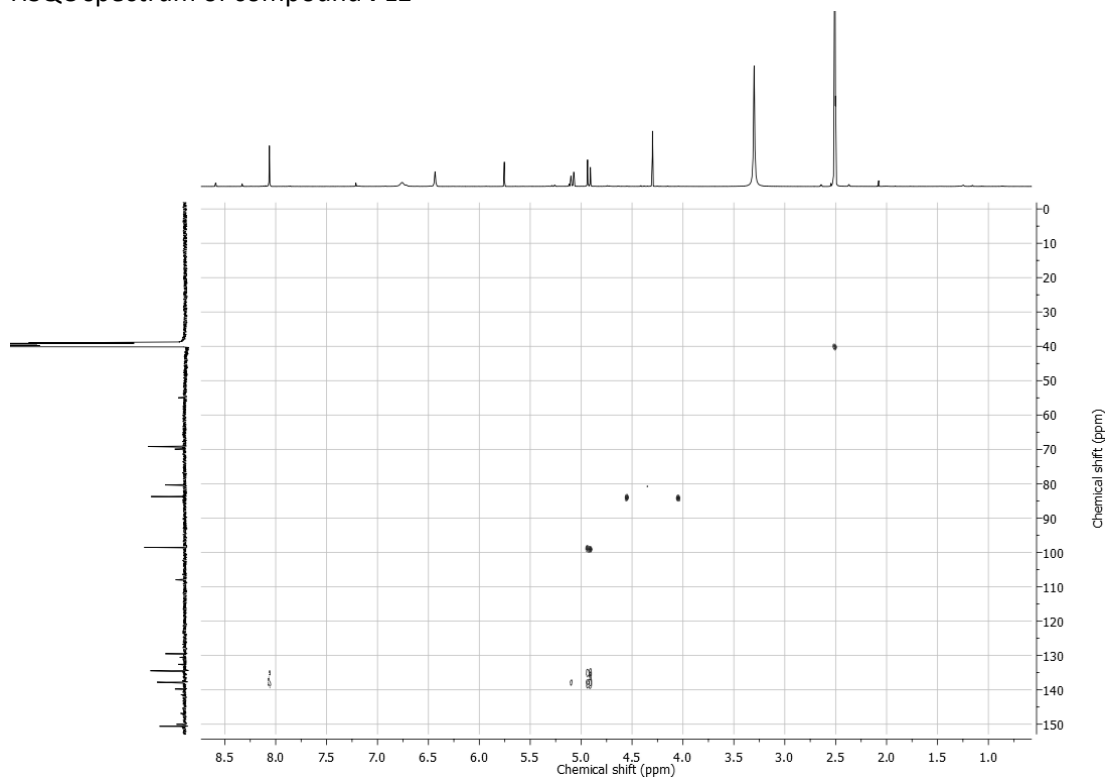
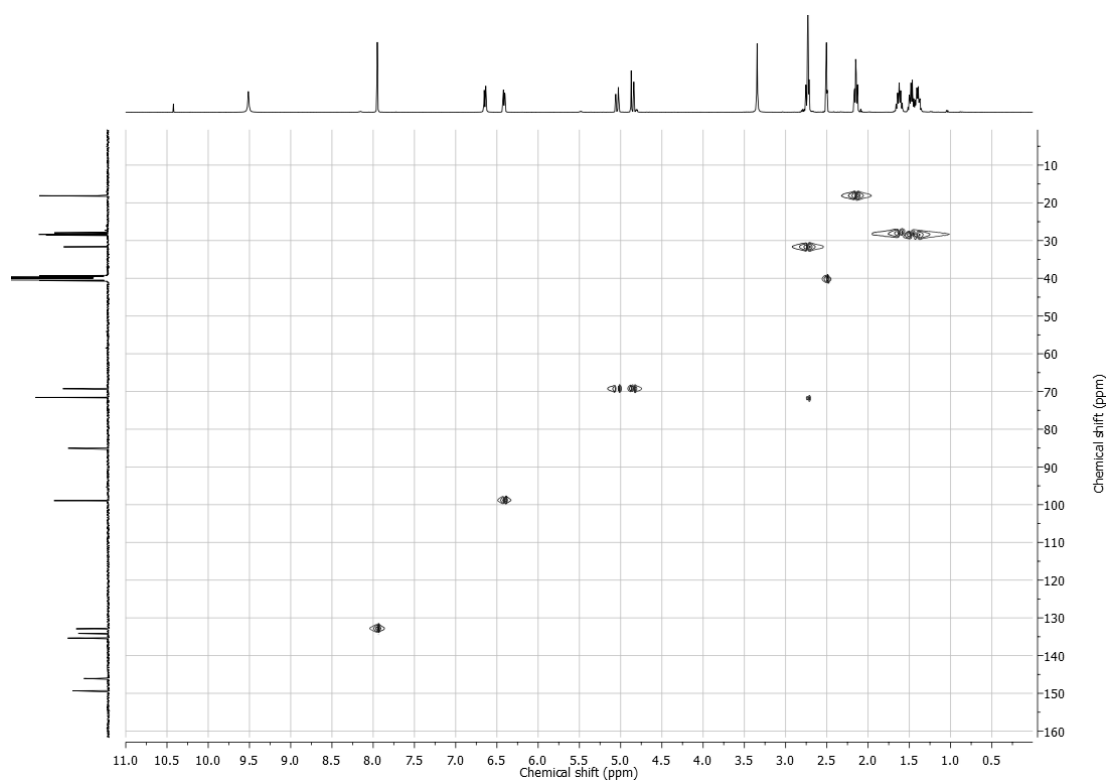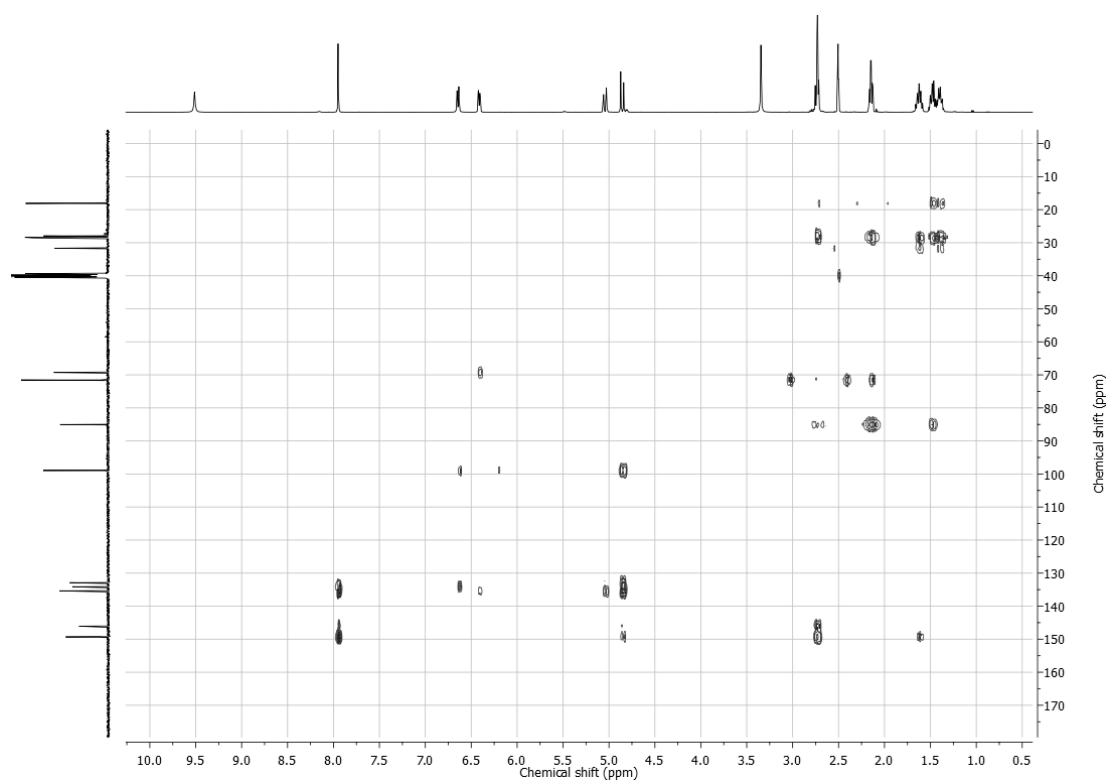HMBC spectrum of compound **PL1**
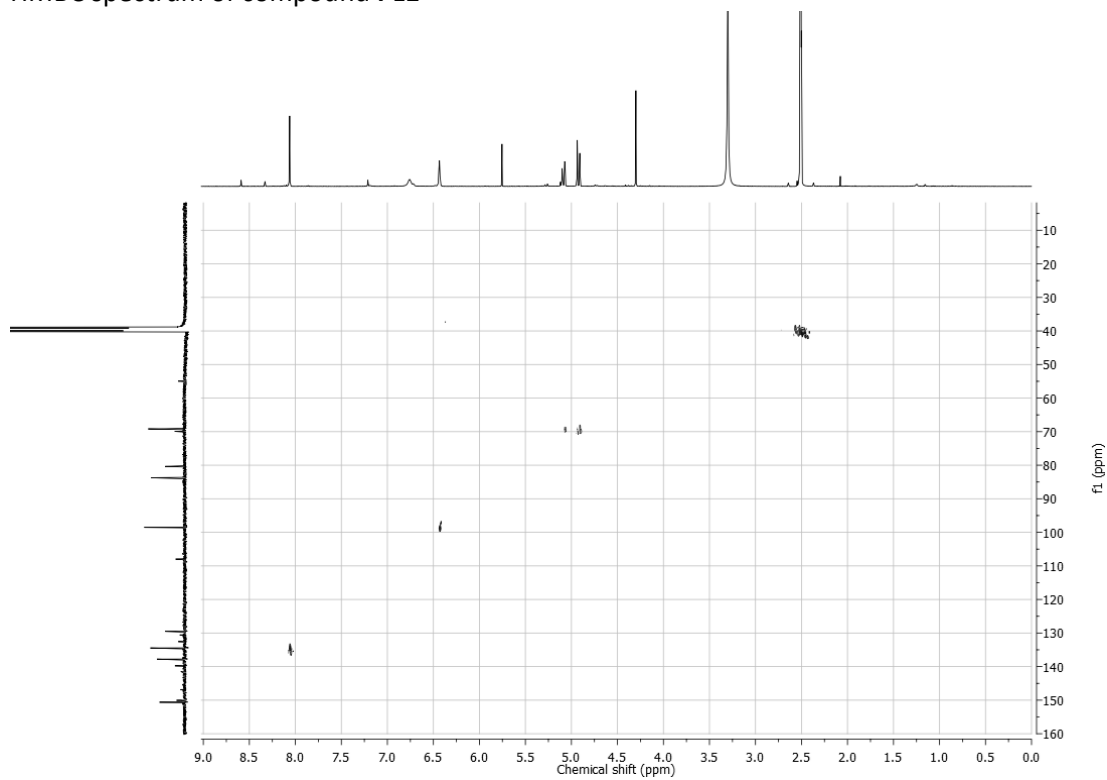
COSY spectrum of compound **PL1**
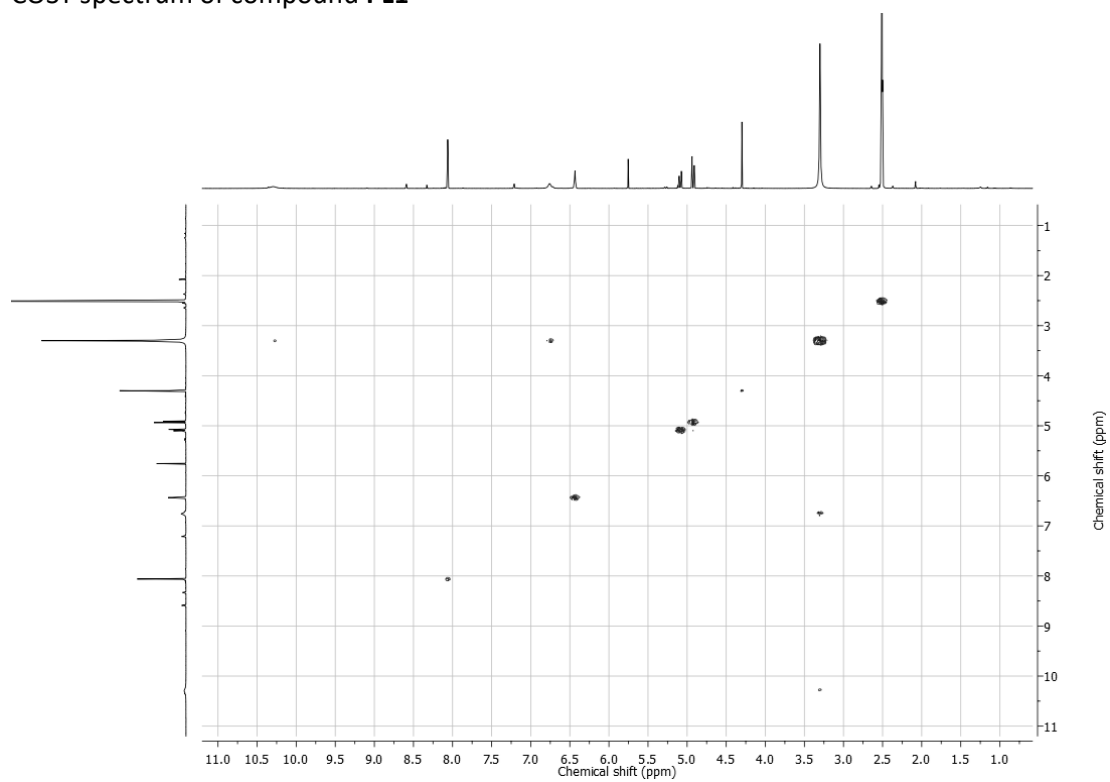
^{1}H NMR of compound **PL2** (500 MHz, MeOD)
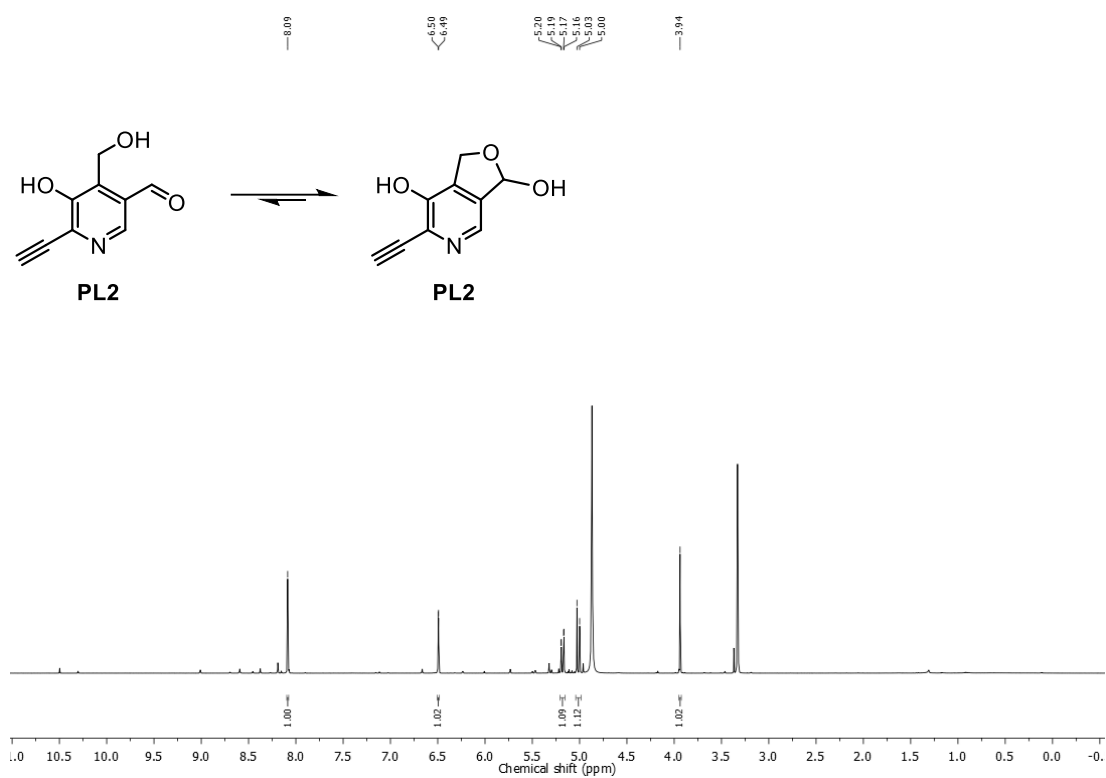
^{13}C NMR of compound **PL2** (125 MHz, MeOD)
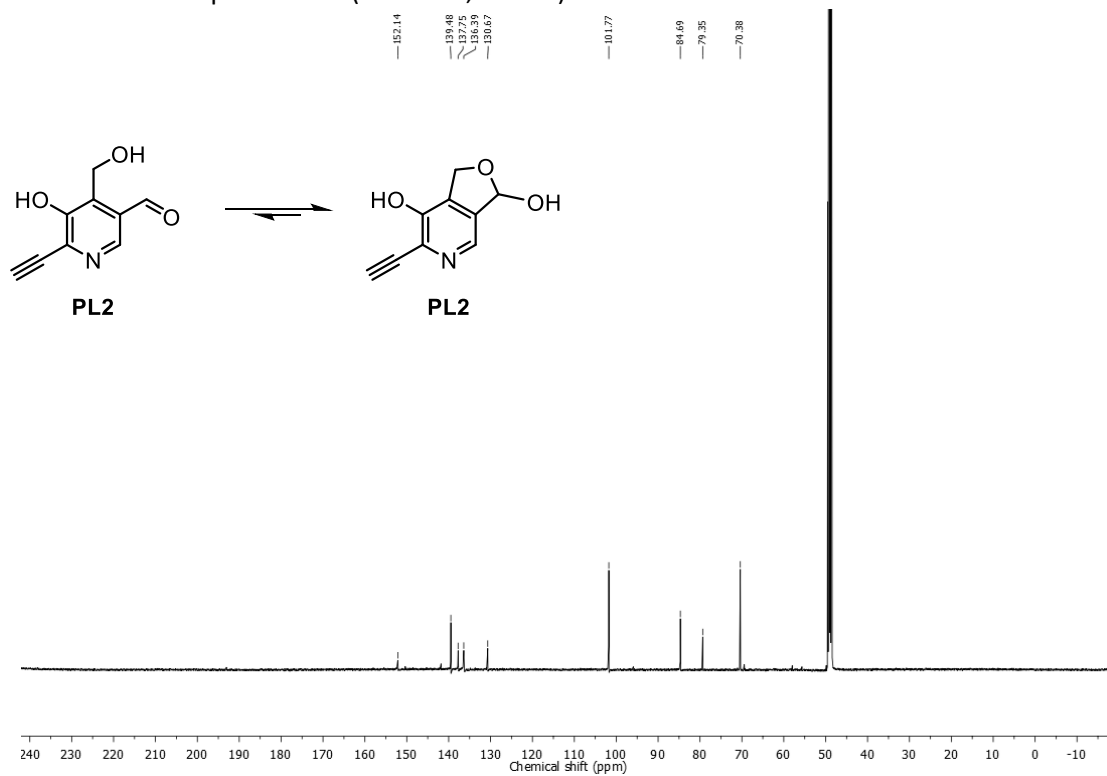
HSQC spectrum of compound **PL2**
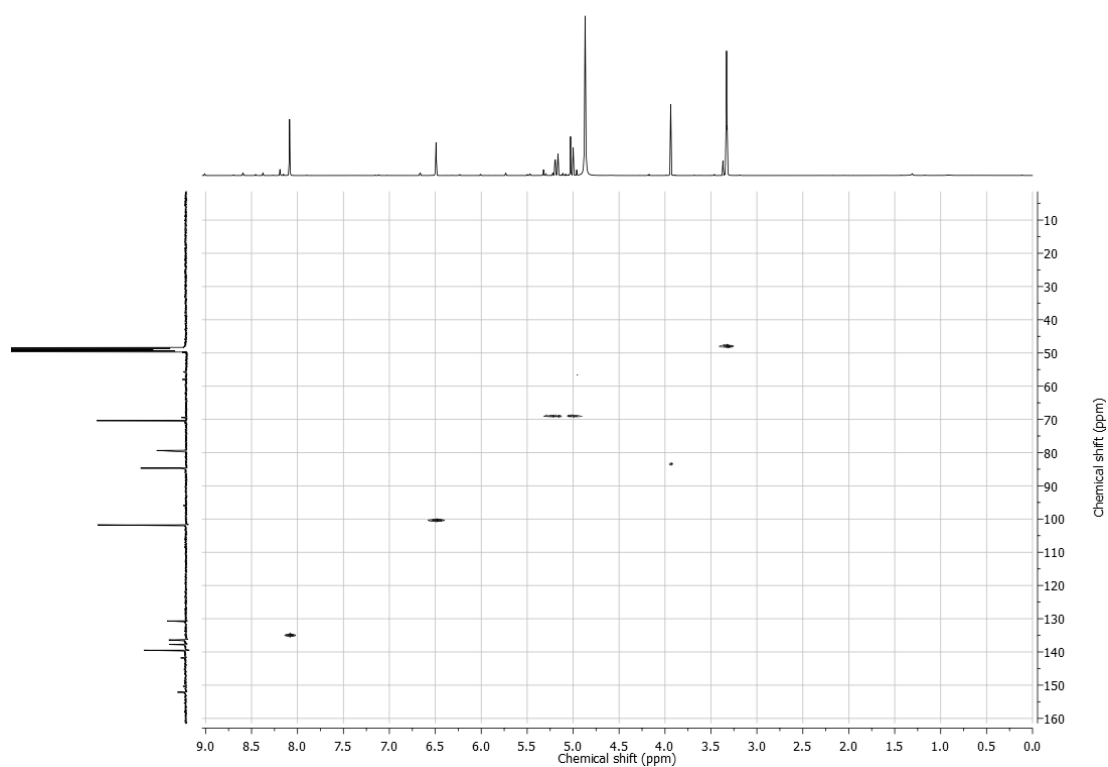
HMBC spectrum of compound **PL2**
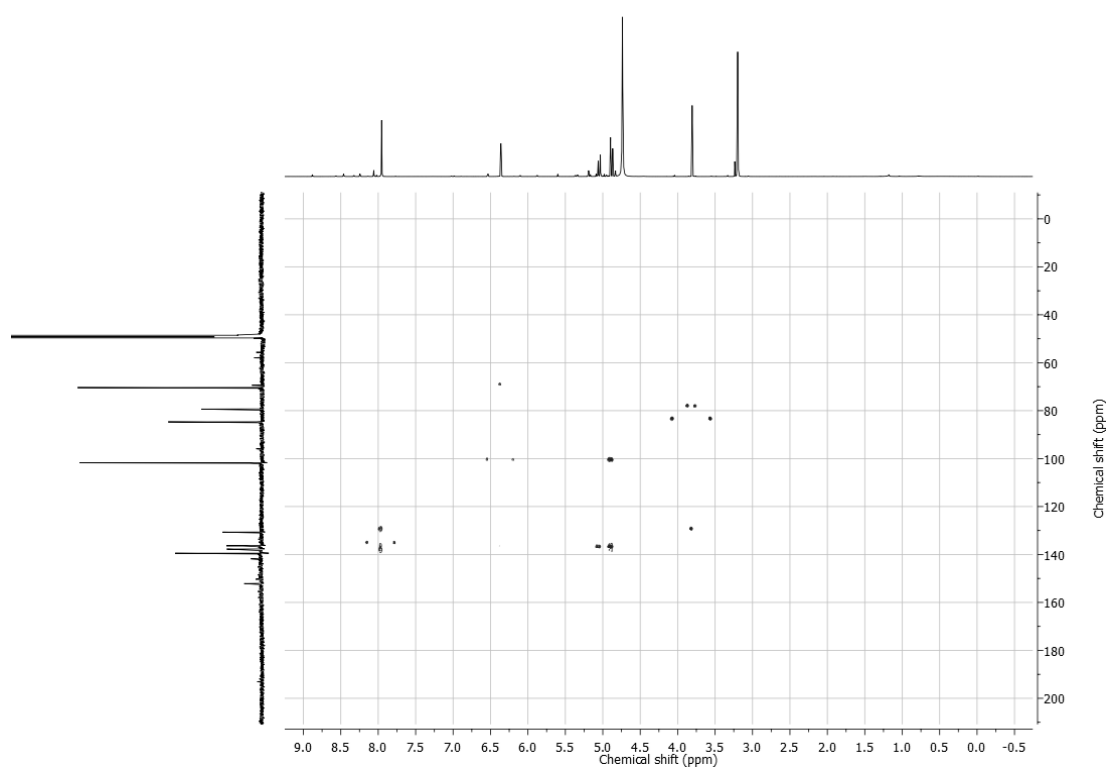
COSY spectrum of compound **PL2**
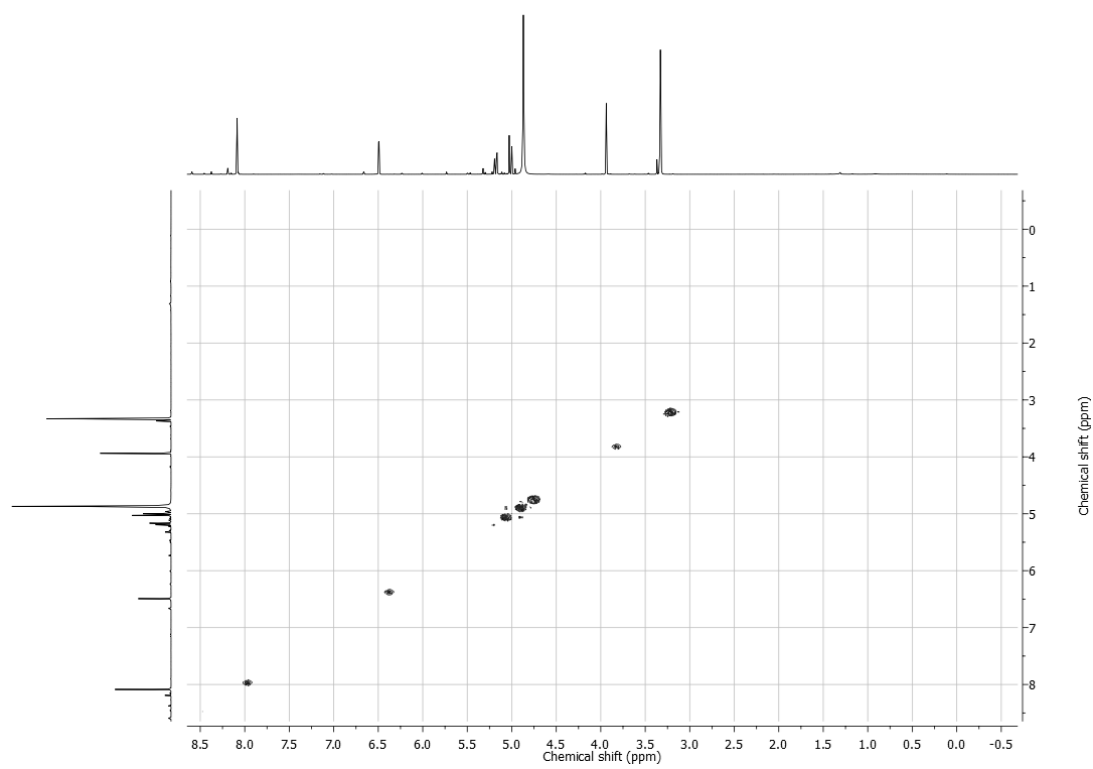
1H NMR of compound **8** (400 MHz, CDCl₃)
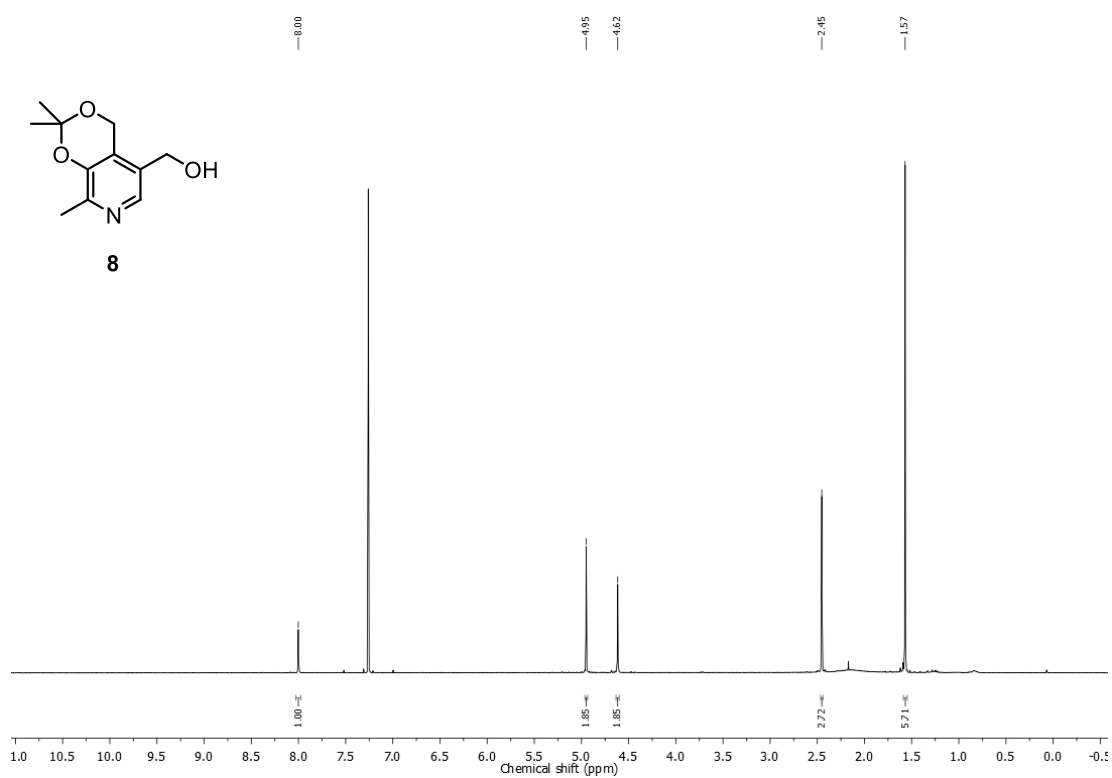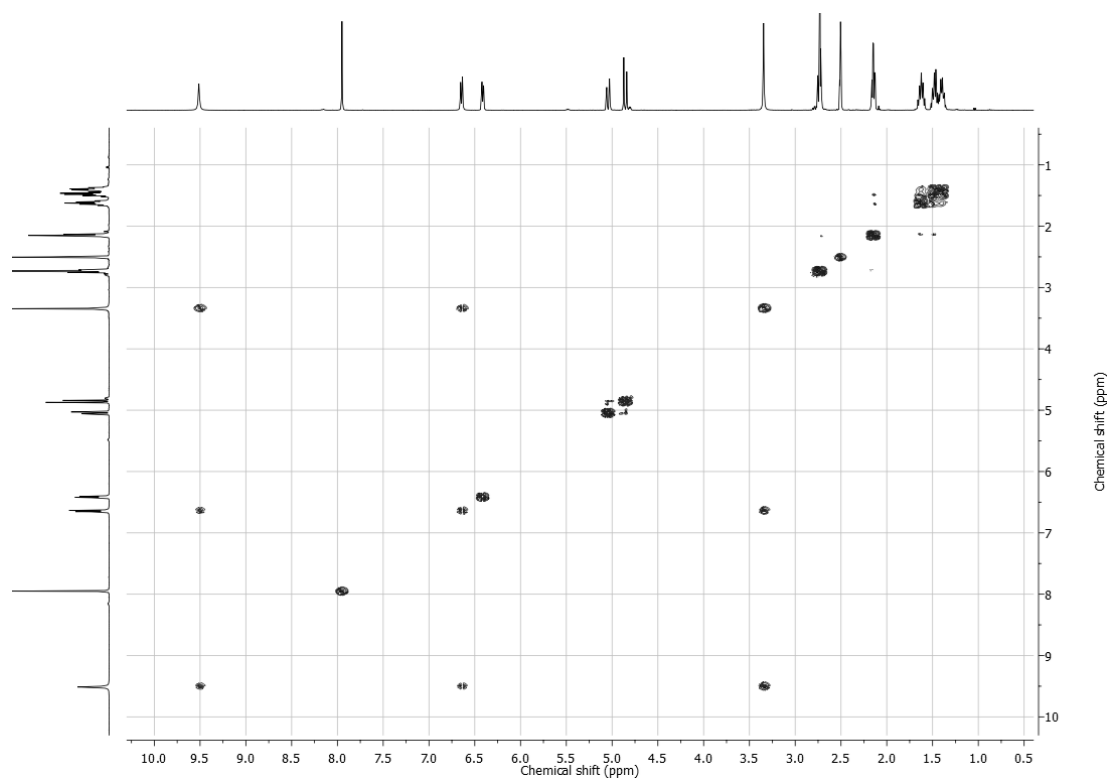

13C NMR of compound **8** (100 MHz, CDCl₃)
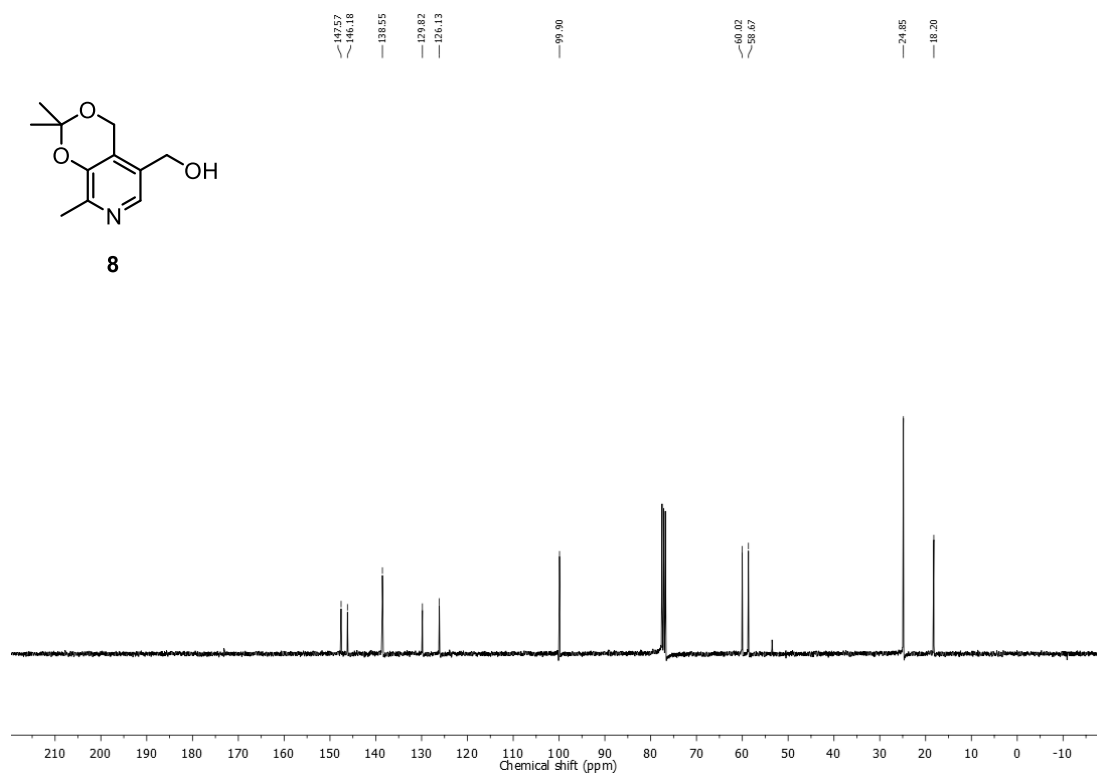
^{1}H NMR of compound **9** (300 MHz, CDCl$_3$)
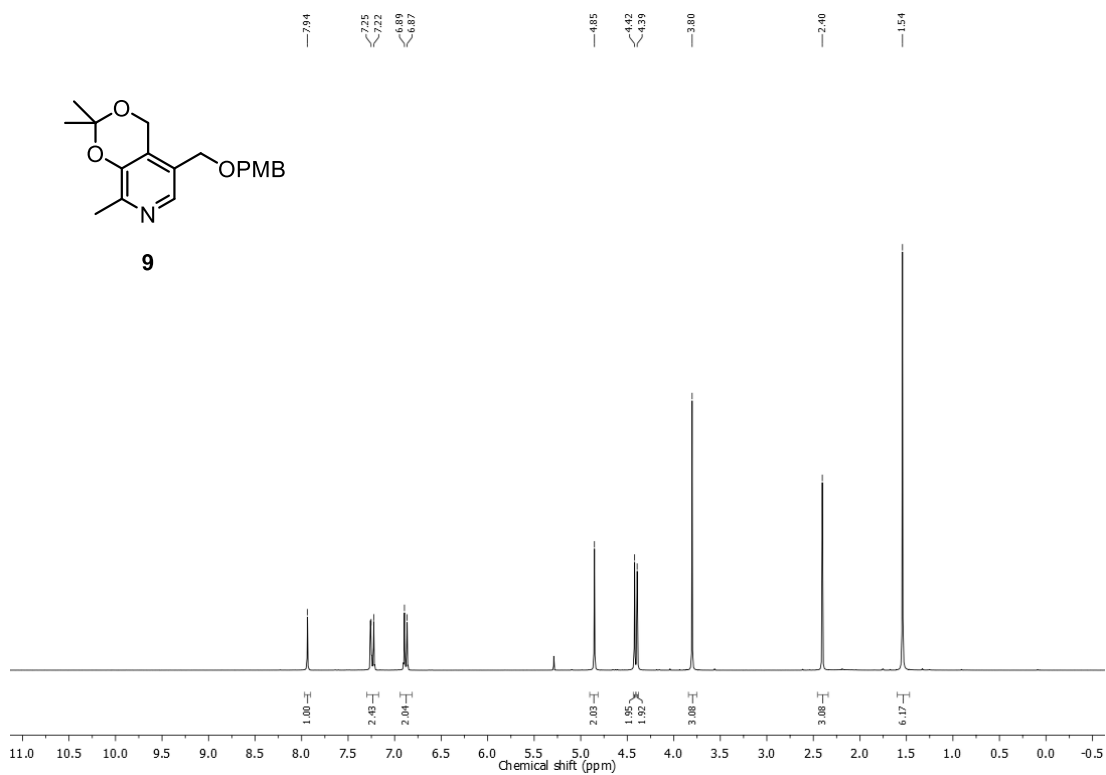
^{13}C NMR of compound **9** (75 MHz, CDCl$_3$)
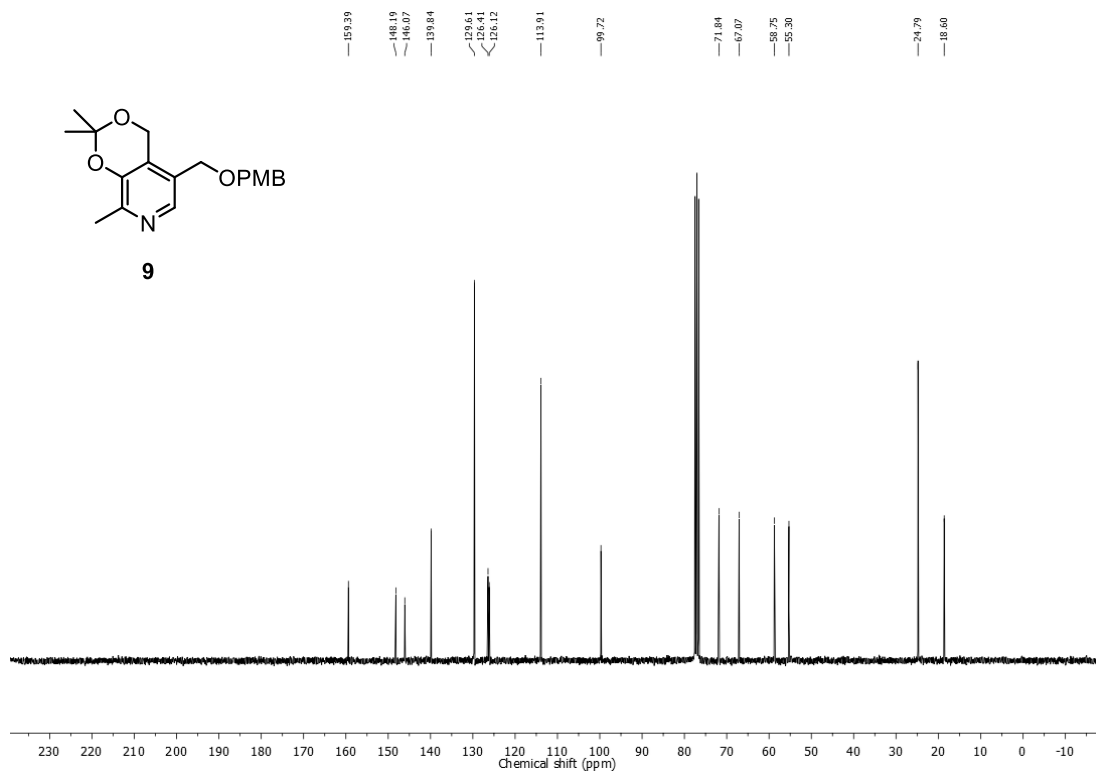
^{1}H NMR of compound **10** (300 MHz, CDCl$_3$)
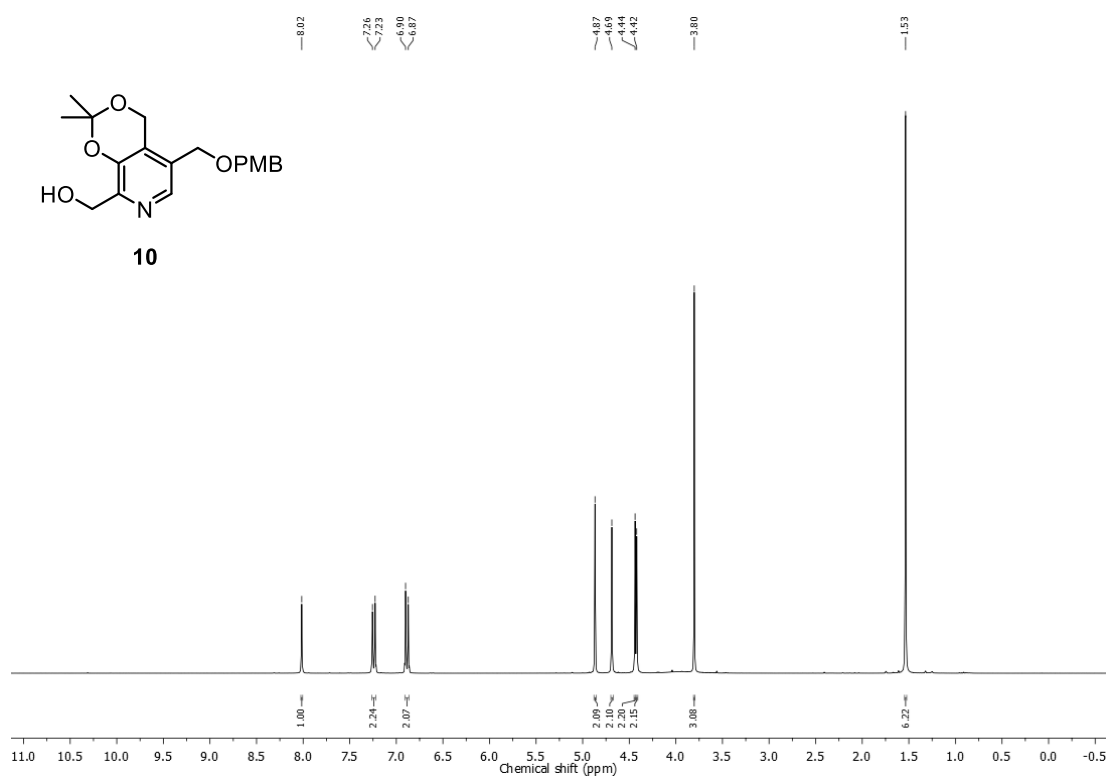
^{13}C NMR of compound **10** (75 MHz, CDCl$_3$)
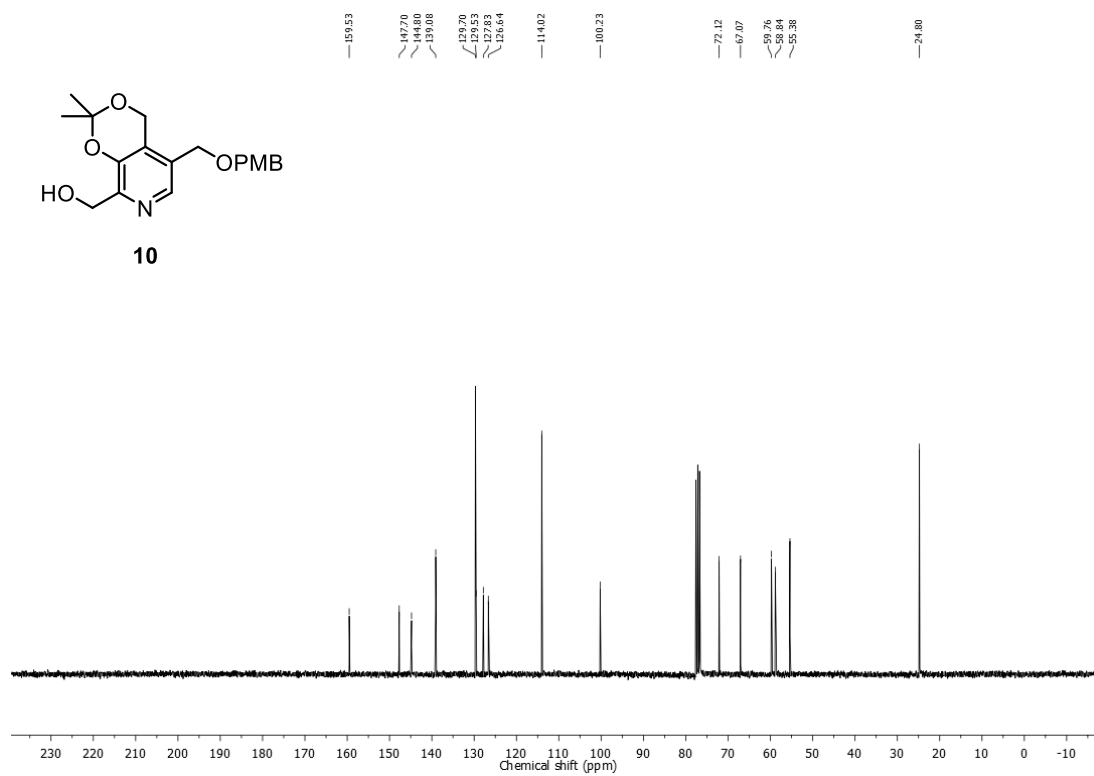
1H NMR of compound 11 (300 MHz, CDCl₃)

8.05
7.27
7.24
6.91
6.88
4.88
4.46
4.43
4.42
3.80
1.57

0.92 2.06 2.32 2.00 1.75 1.84 1.91 3.34 6.06

Chemical shift (ppm)
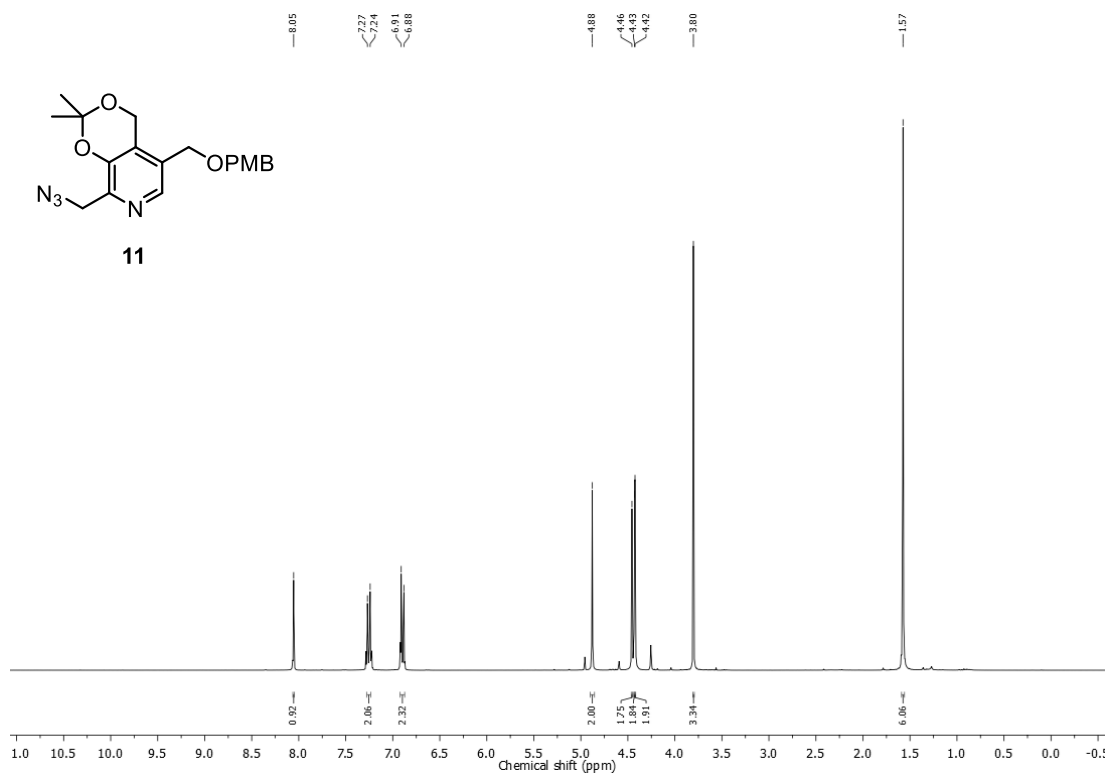
13C NMR of compound 11 (75 MHz, CDCl₃)

159.44
146.40
144.30
140.24
129.61
129.36
129.03
127.39
114.14
113.92
100.46
72.17
66.86
58.66
55.25
50.80
24.68

Chemical shift (ppm)
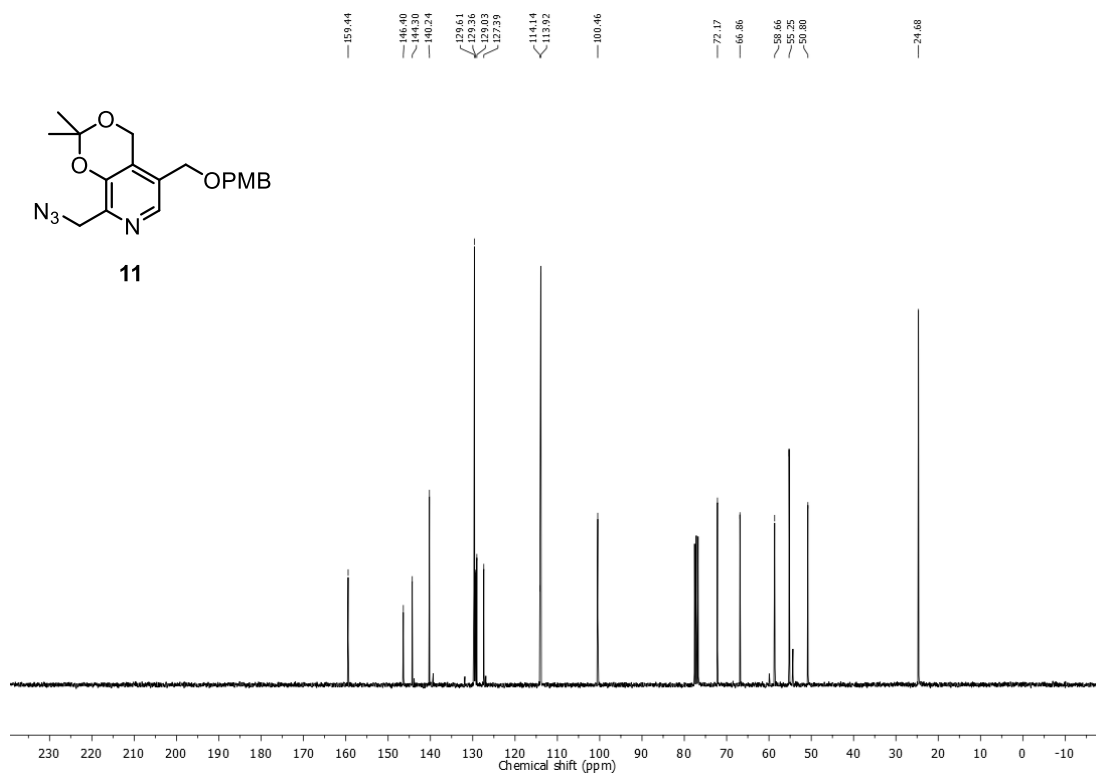
¹H NMR of compound **12** (400 MHz, CDCl₃)
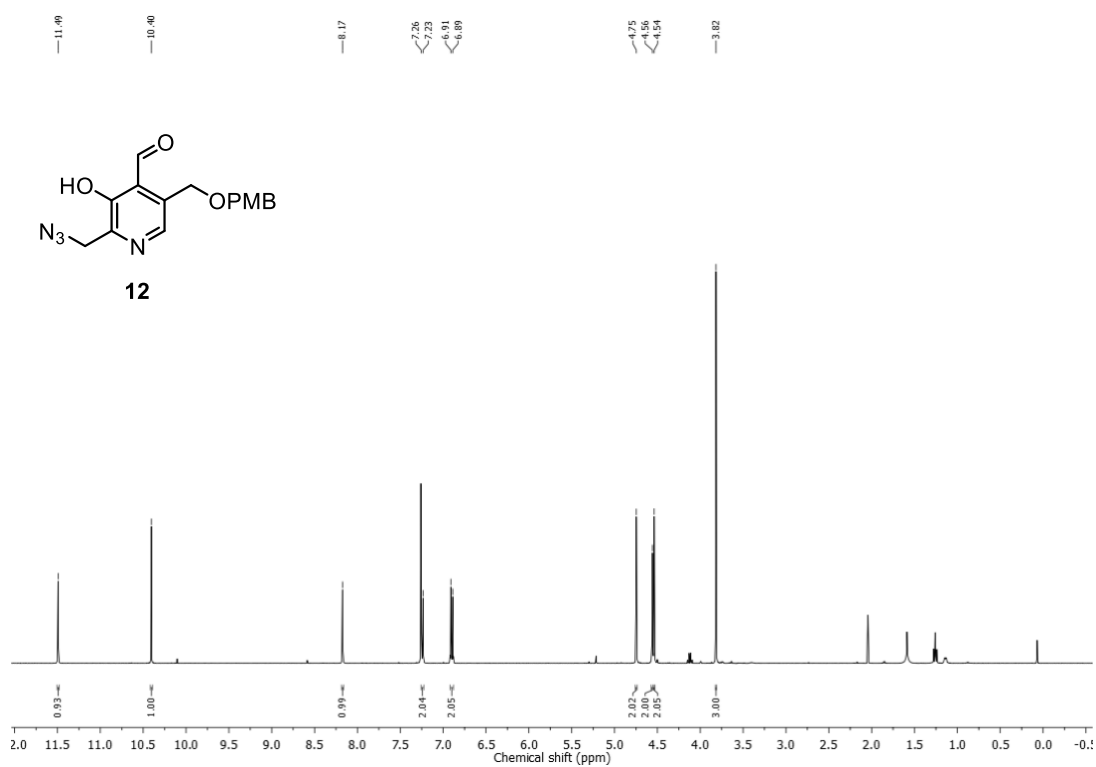
¹³C NMR of compound **12** (100 MHz, CDCl₃)
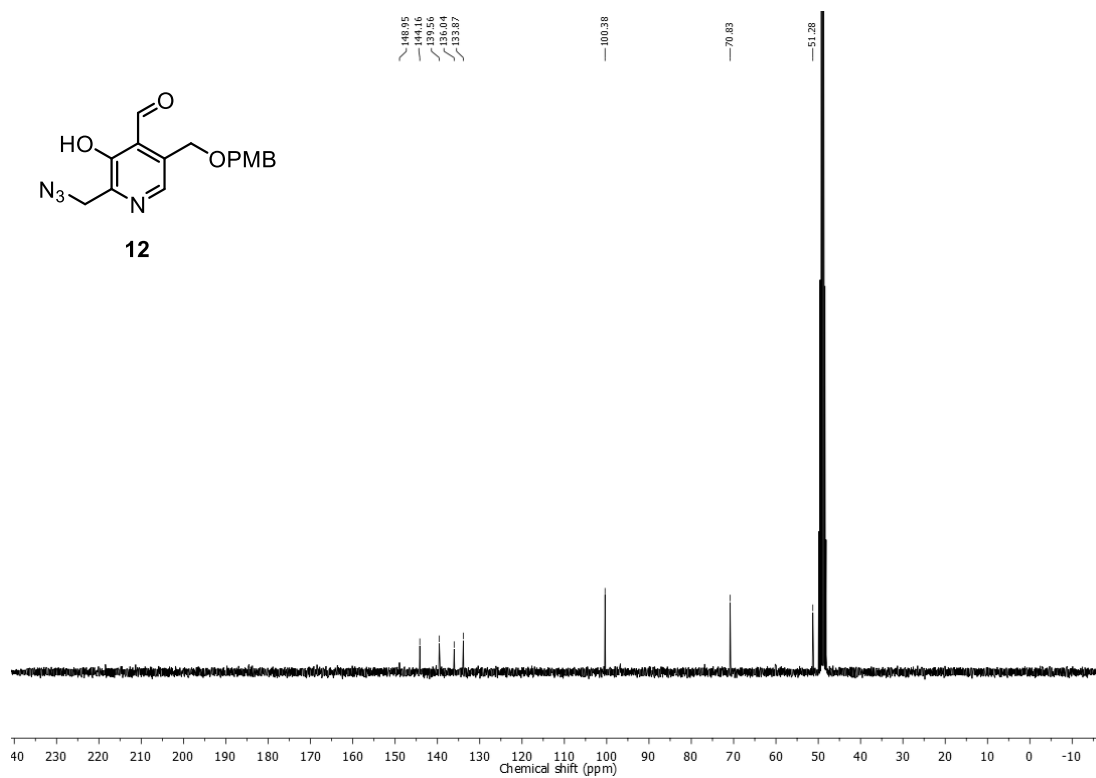
¹H NMR of compound **PL3** (500 MHz, MeOD)
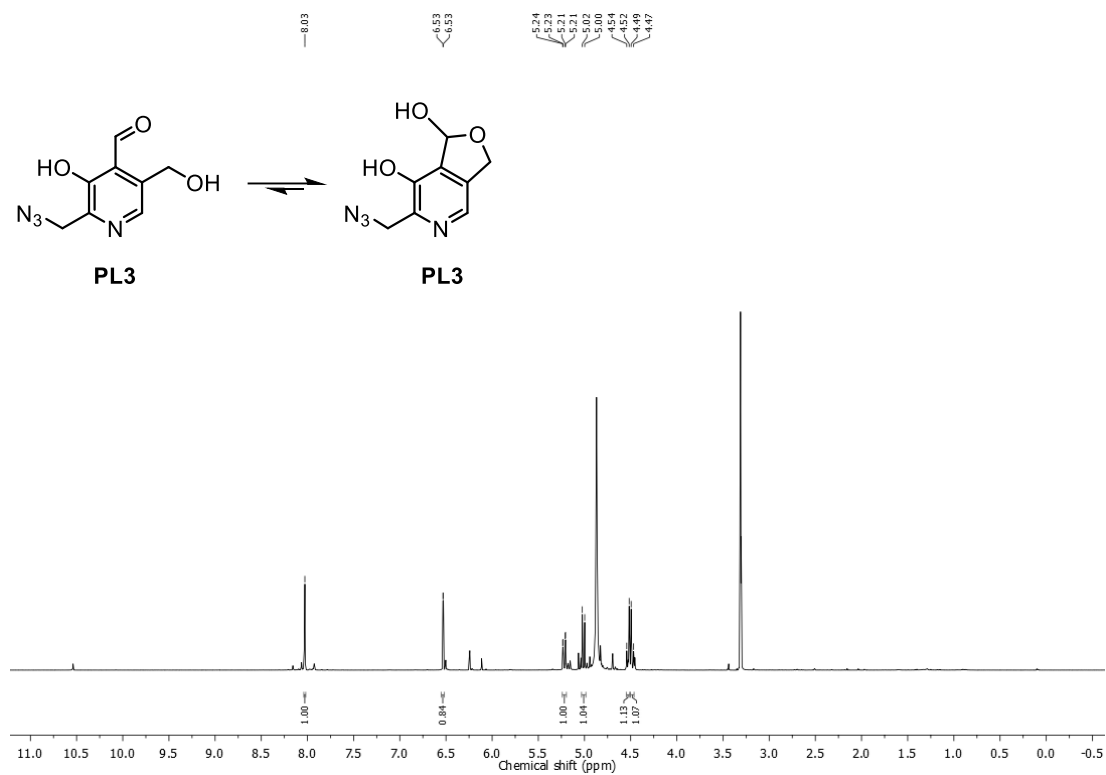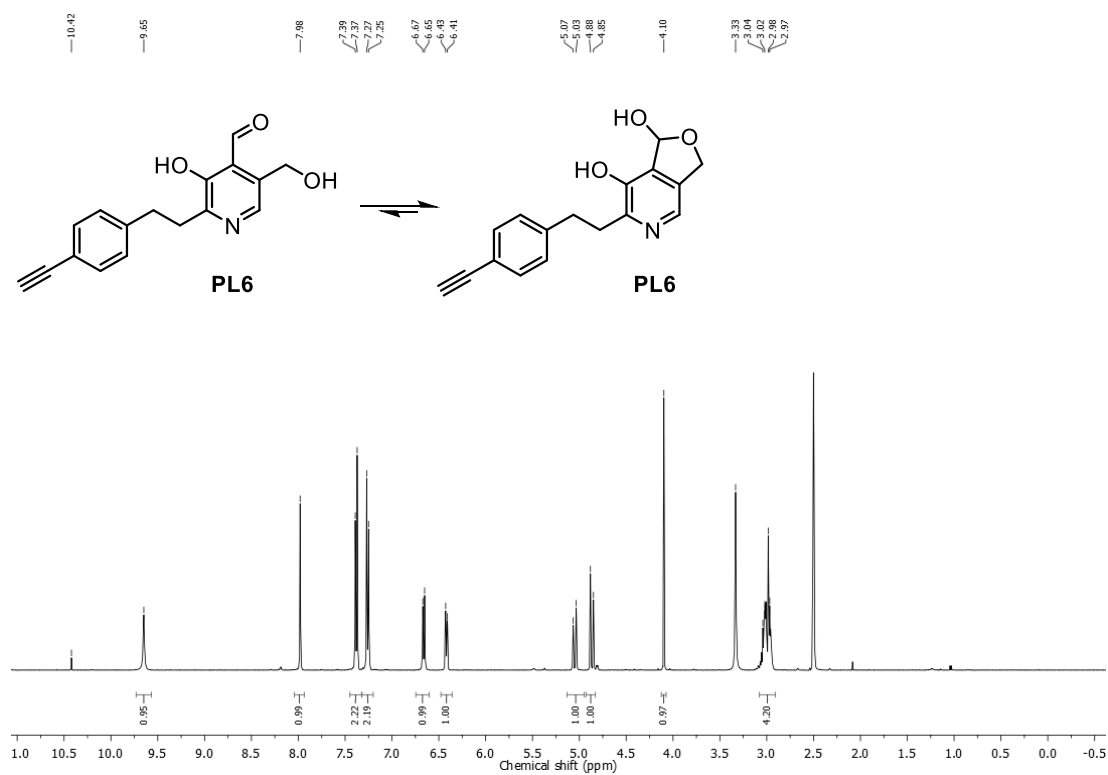

¹³C NMR of compound **PL3** (75 MHz, MeOD)
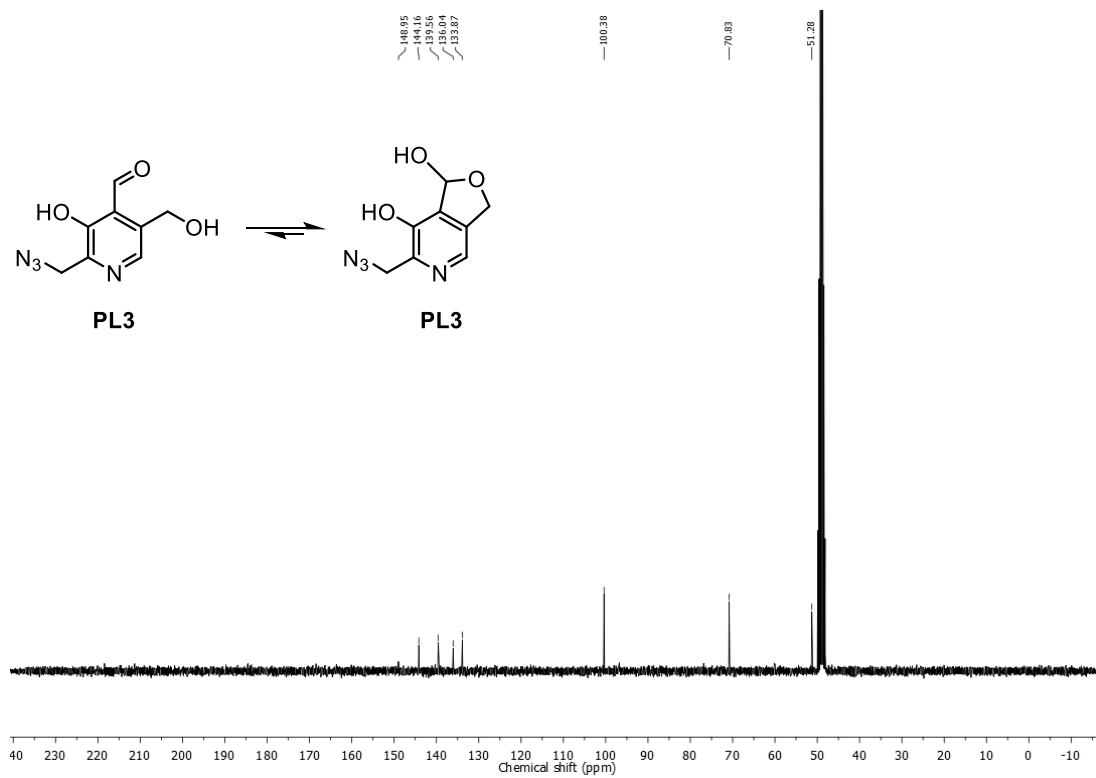
HSQC spectrum of compound **PL3**
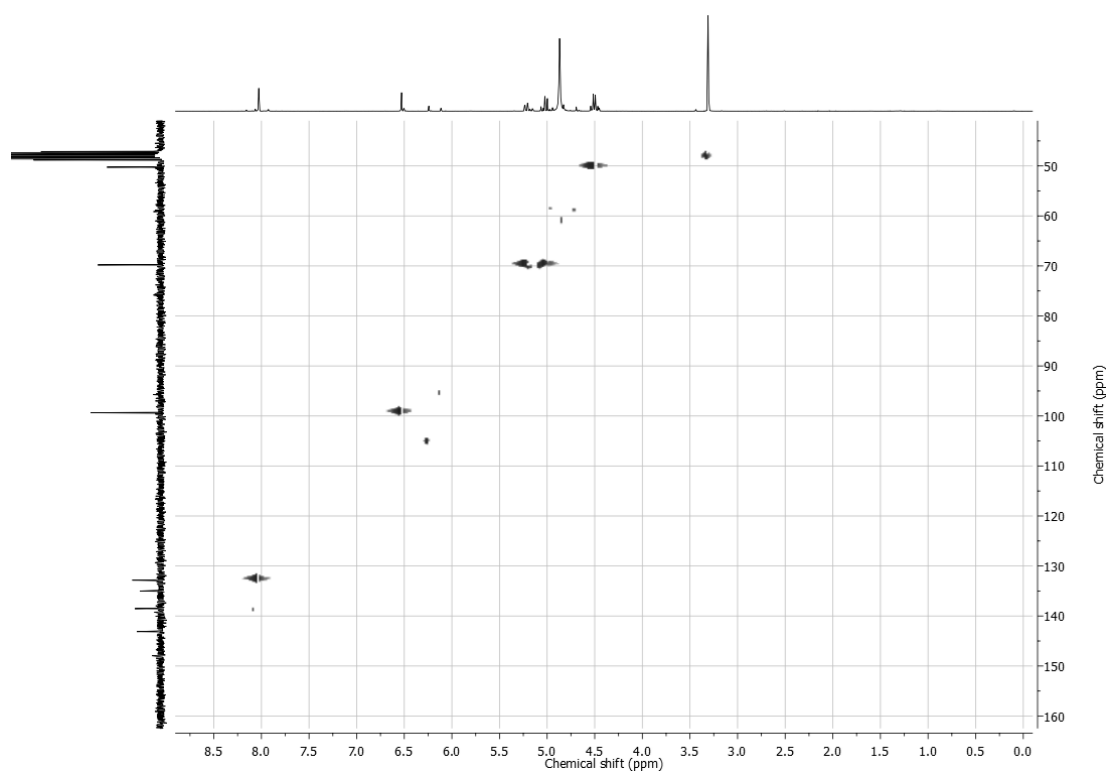
HMBC spectrum of compound **PL3**
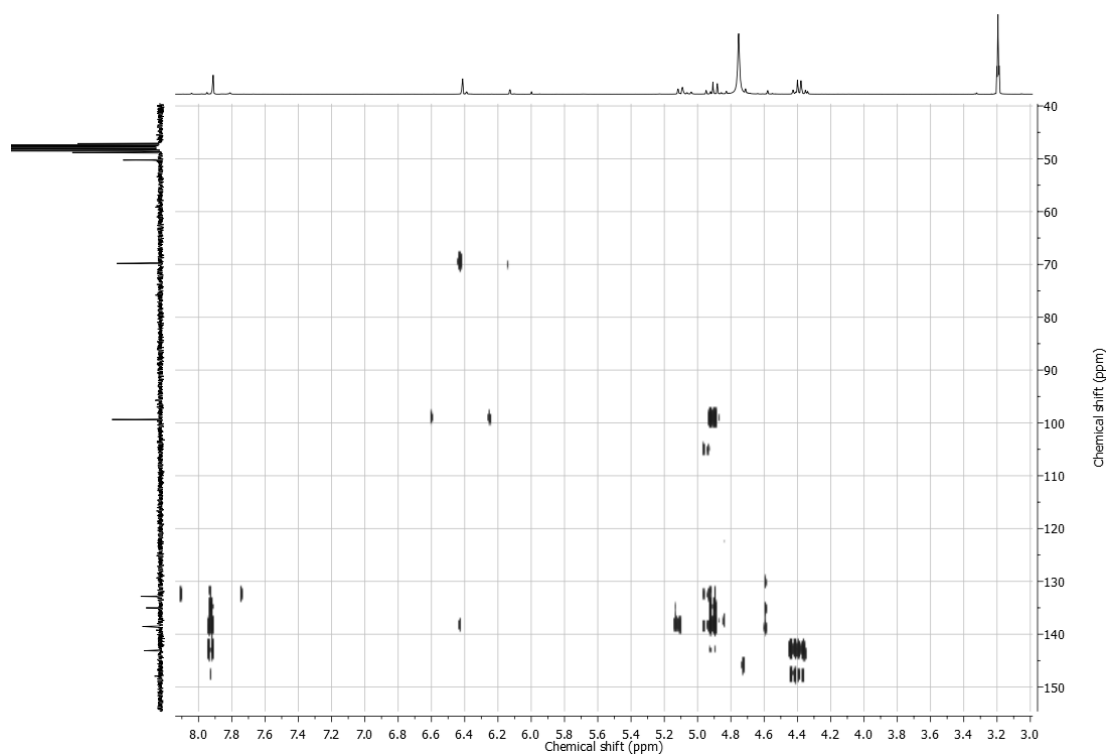
COSY spectrum of compound **PL3**
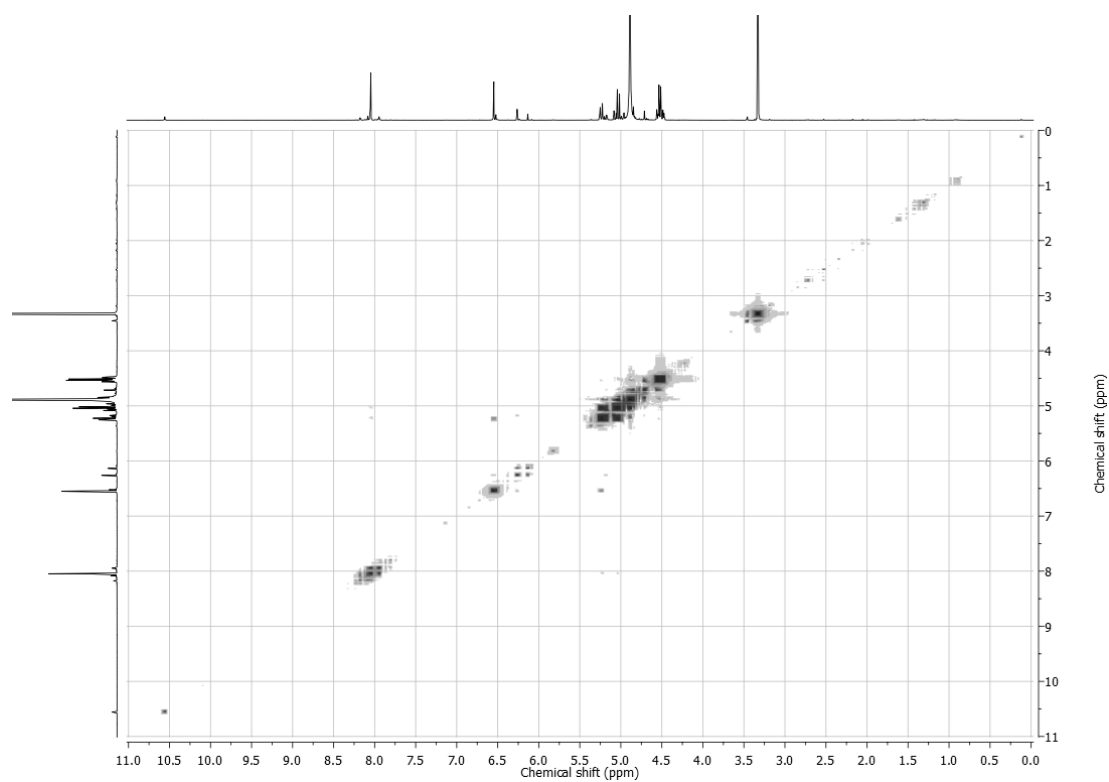
^{1}H NMR of compound **13** (300 MHz, CDCl$_3$)
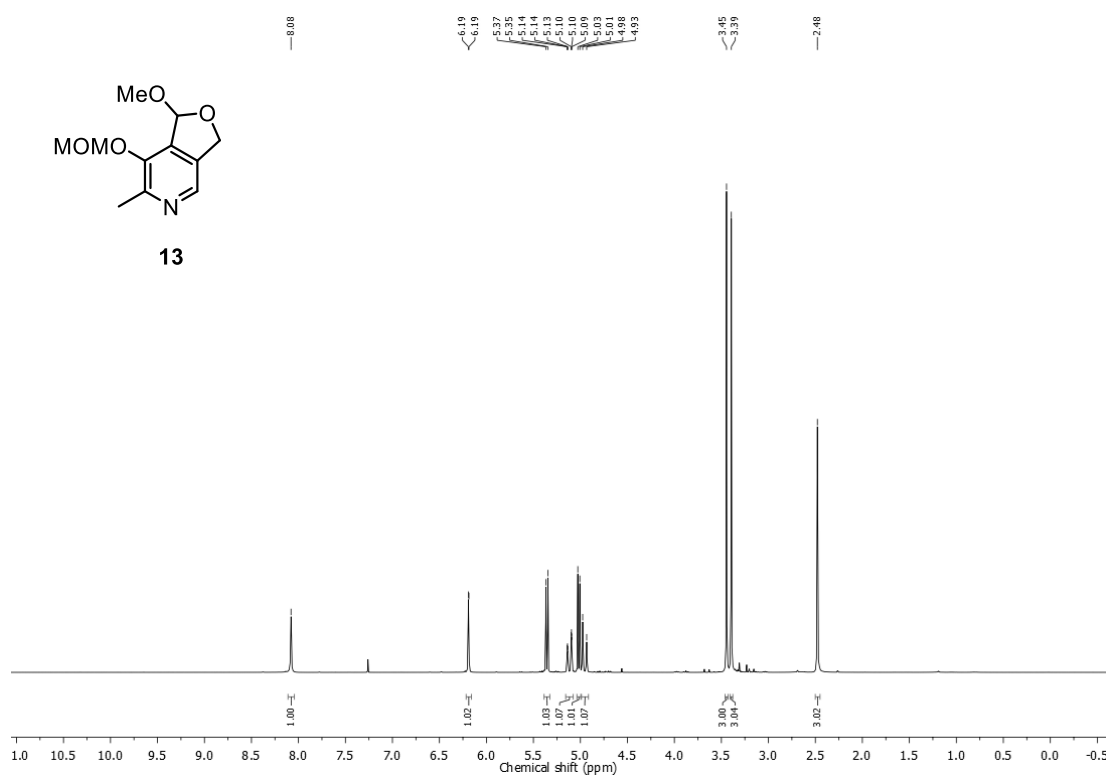
^{13}C NMR of compound **13** (75 MHz, CDCl$_3$)
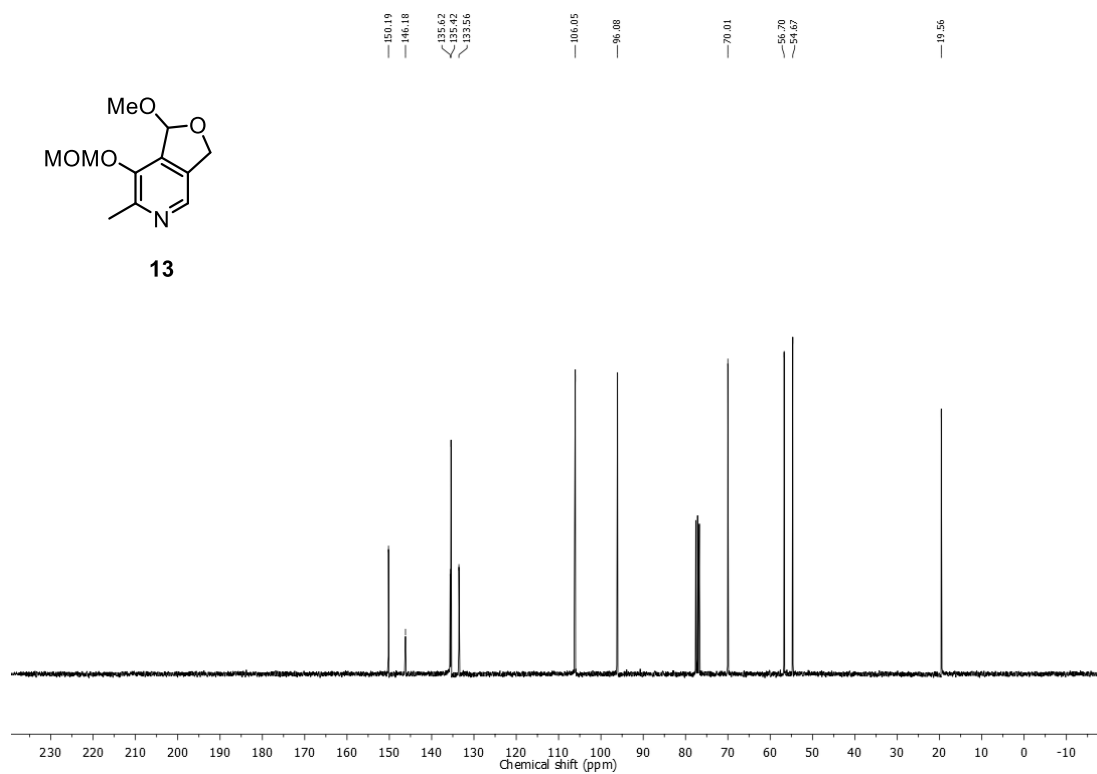
¹H NMR of compound **14** (400 MHz, CDCl₃)
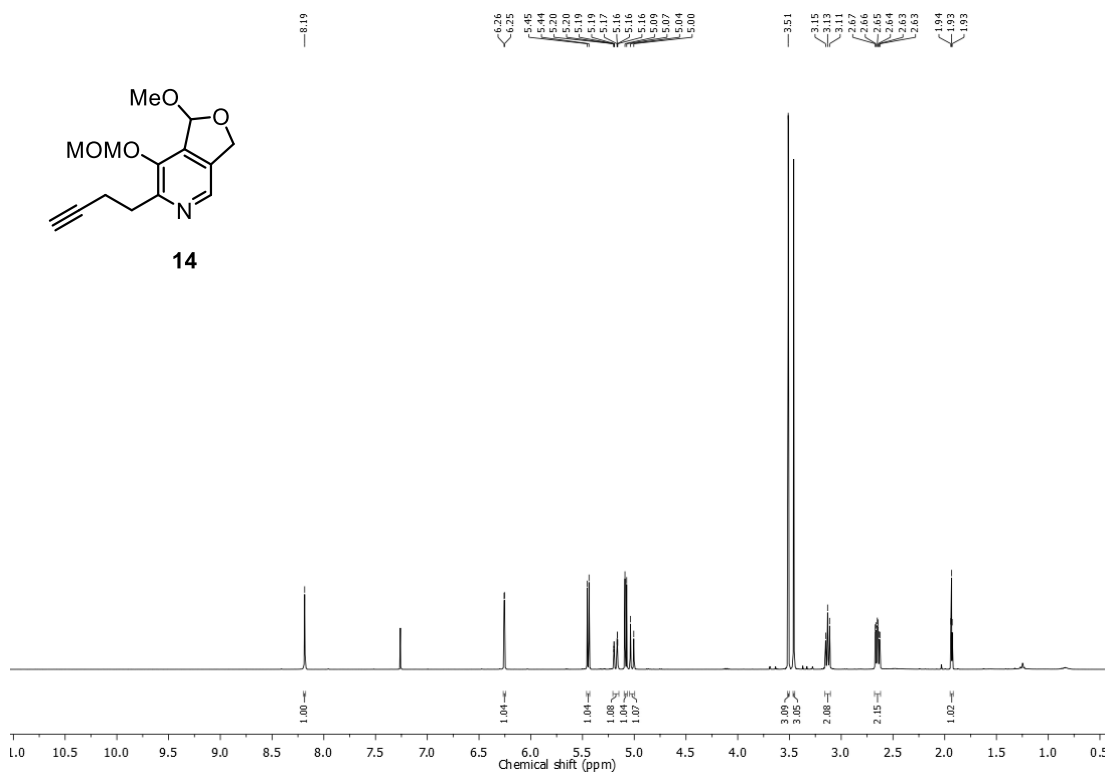
¹³C NMR of compound **14** (100 MHz, CDCl₃)
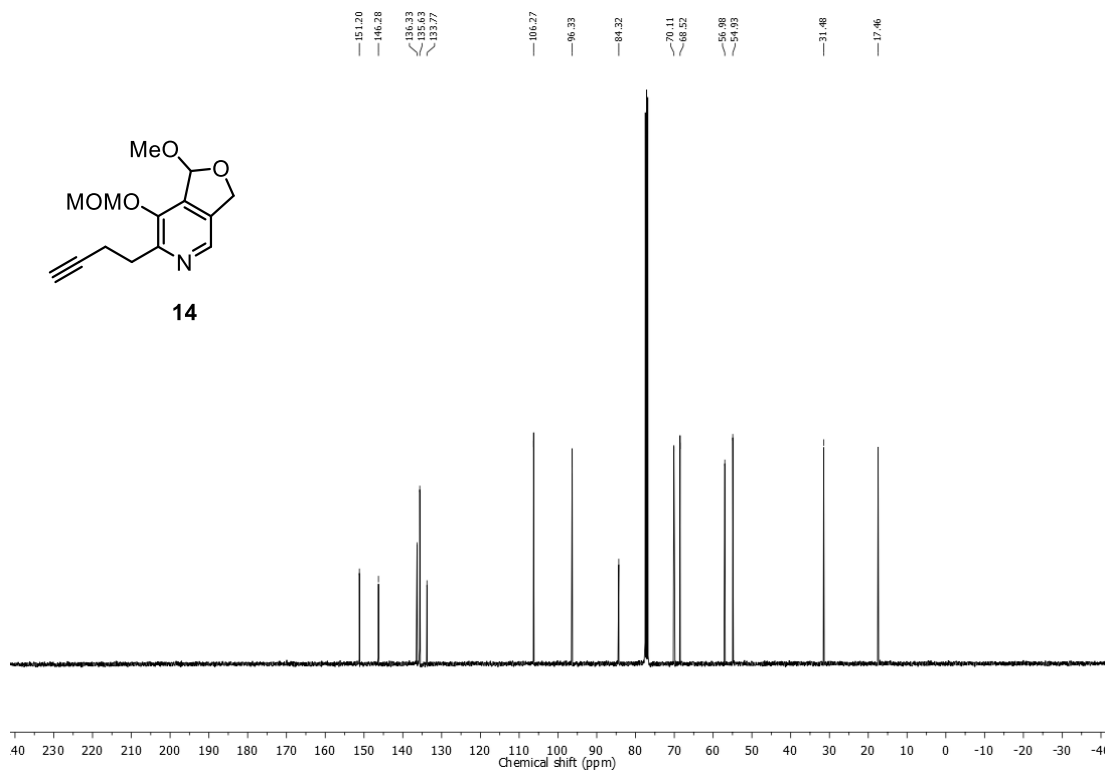
^{1}H NMR of compound **PL4** (400 MHz, DMSO-d_6)
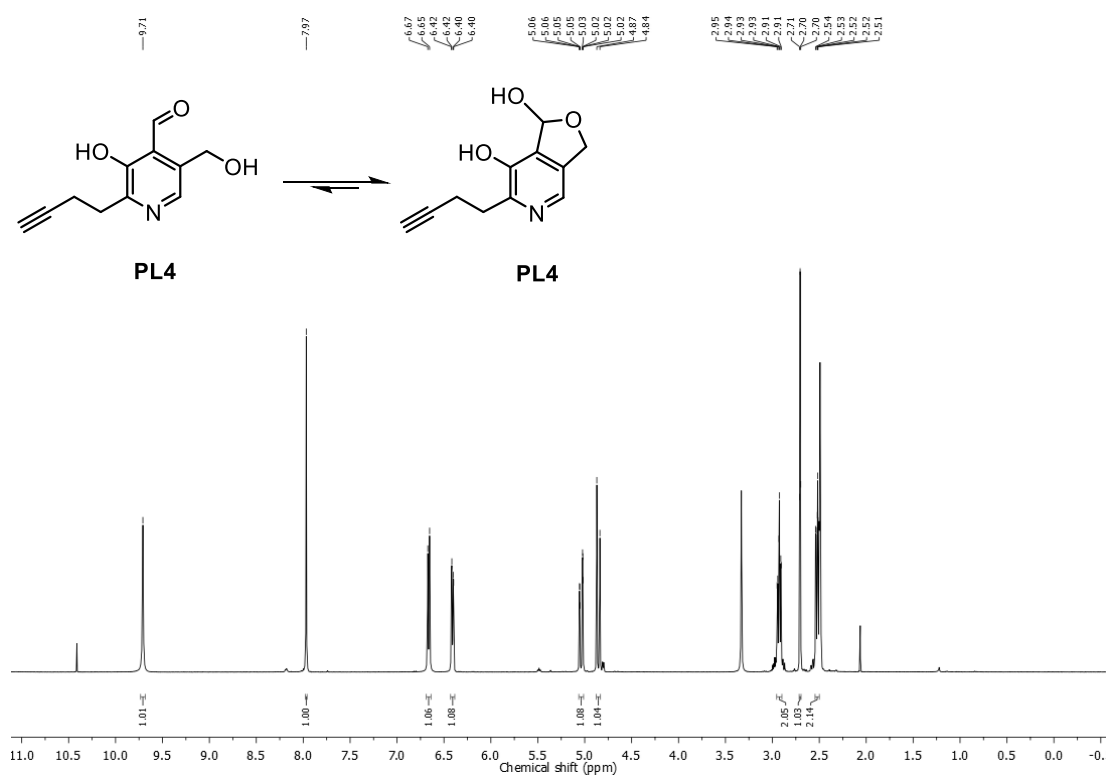
## ^{13}C NMR of compound **PL4** (100 MHz, DMSO-d_6)
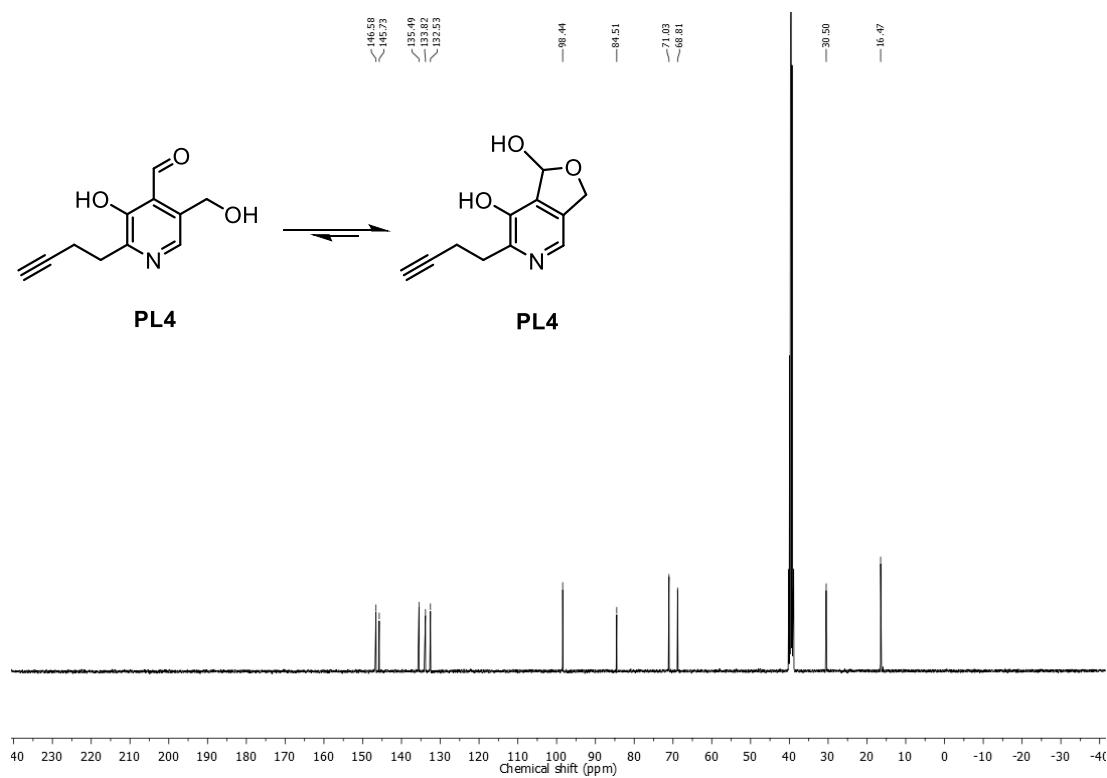
HSQC spectrum of compound **PL4**
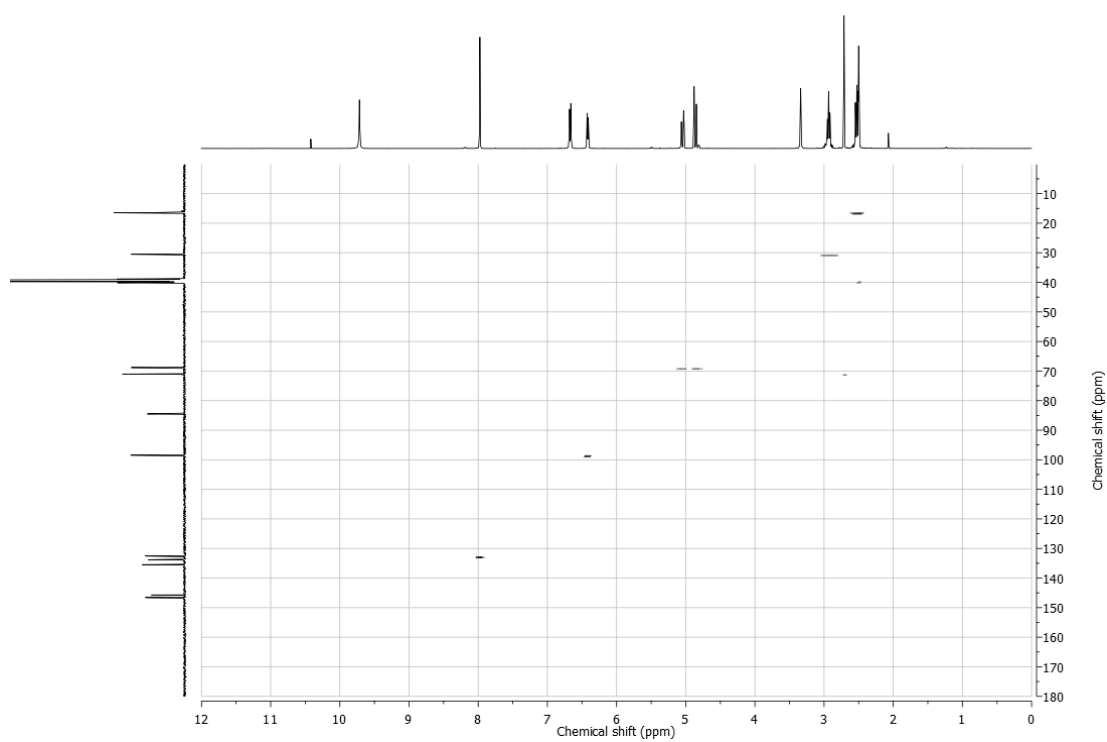
HMBC spectrum of compound **PL4**
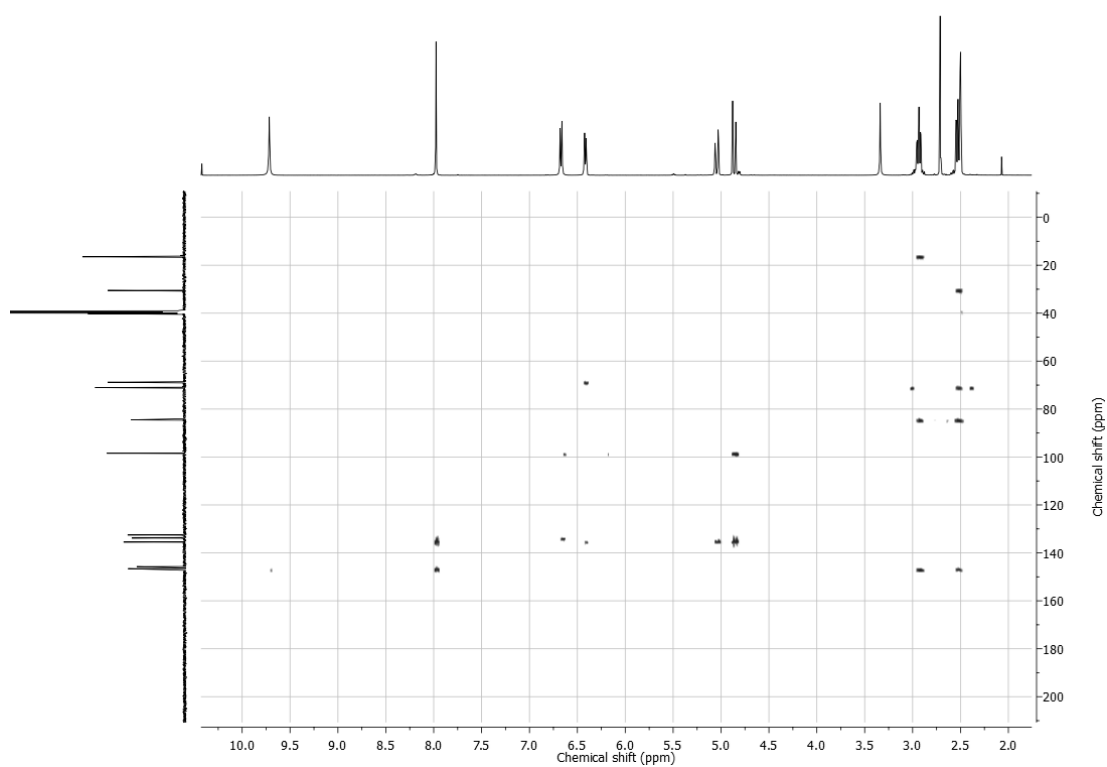
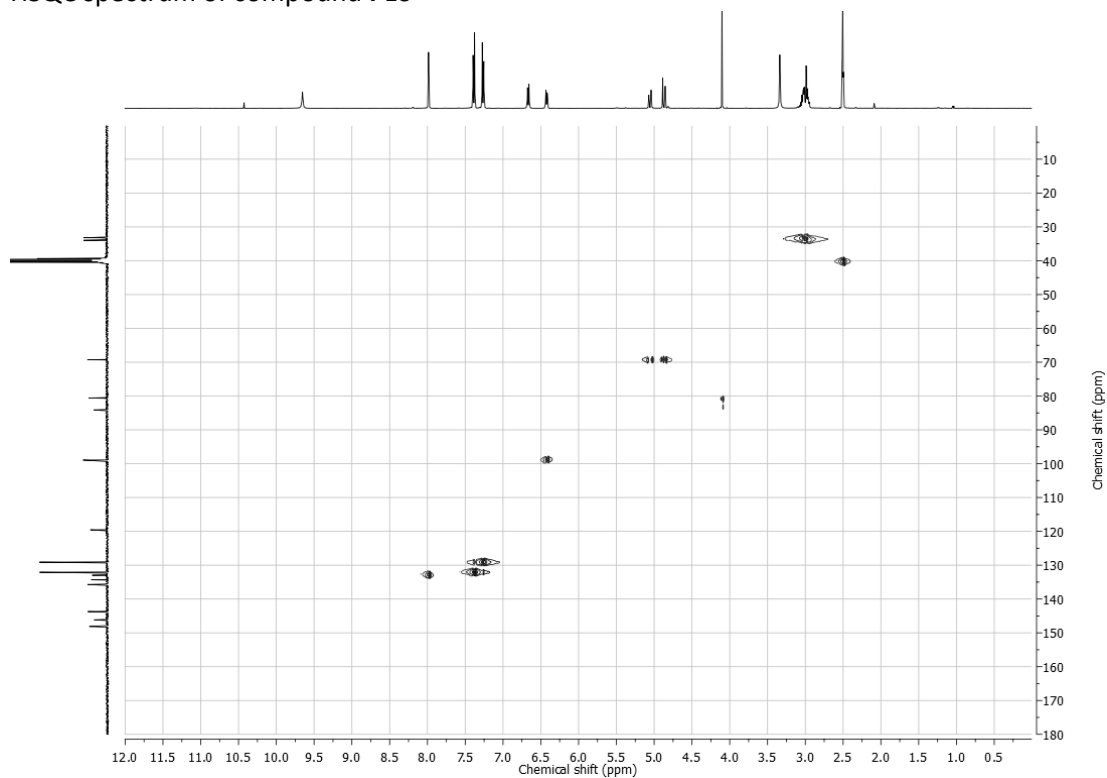COSY spectrum of compound **PL4**
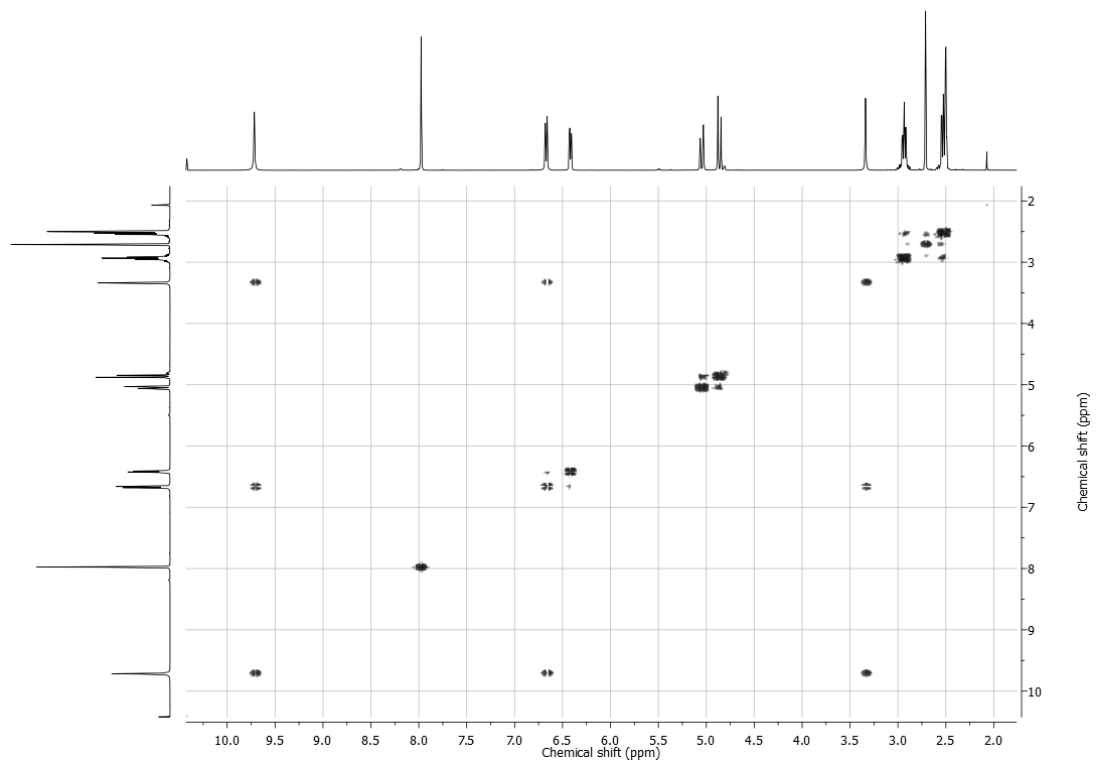
^{1}H NMR of compound **15** (400 MHz, CDCl$_3$)
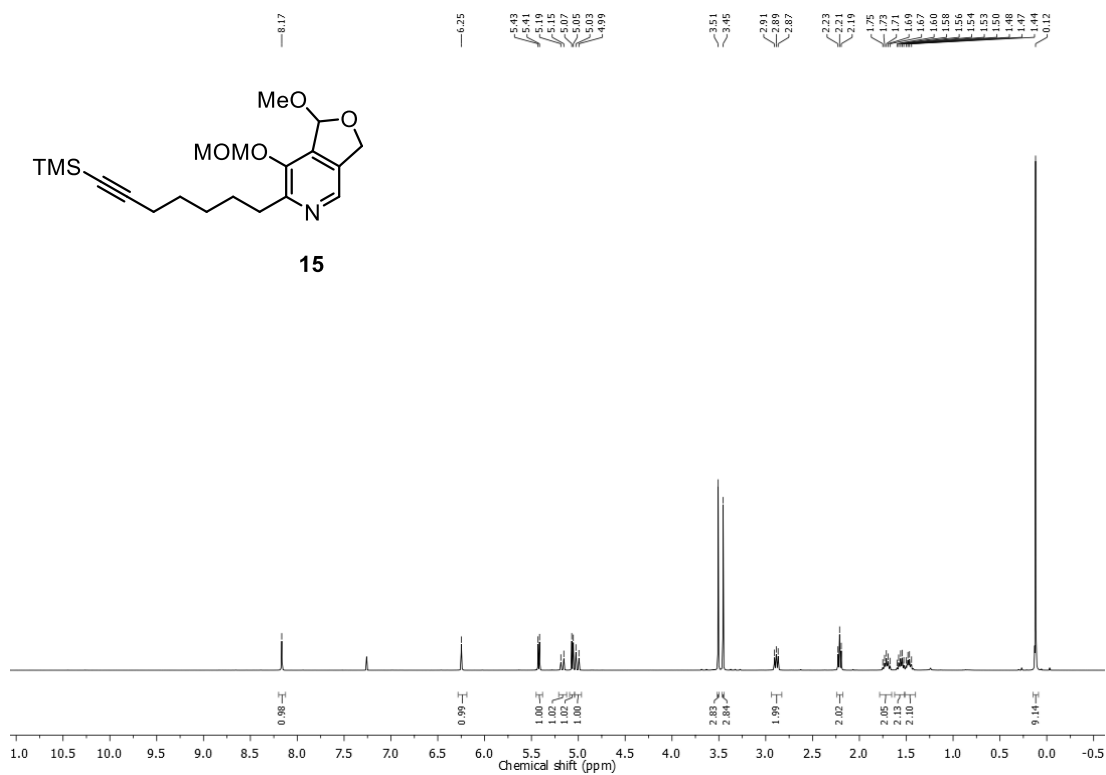
^{13}C NMR of compound **15** (100 MHz, CDCl$_3$)
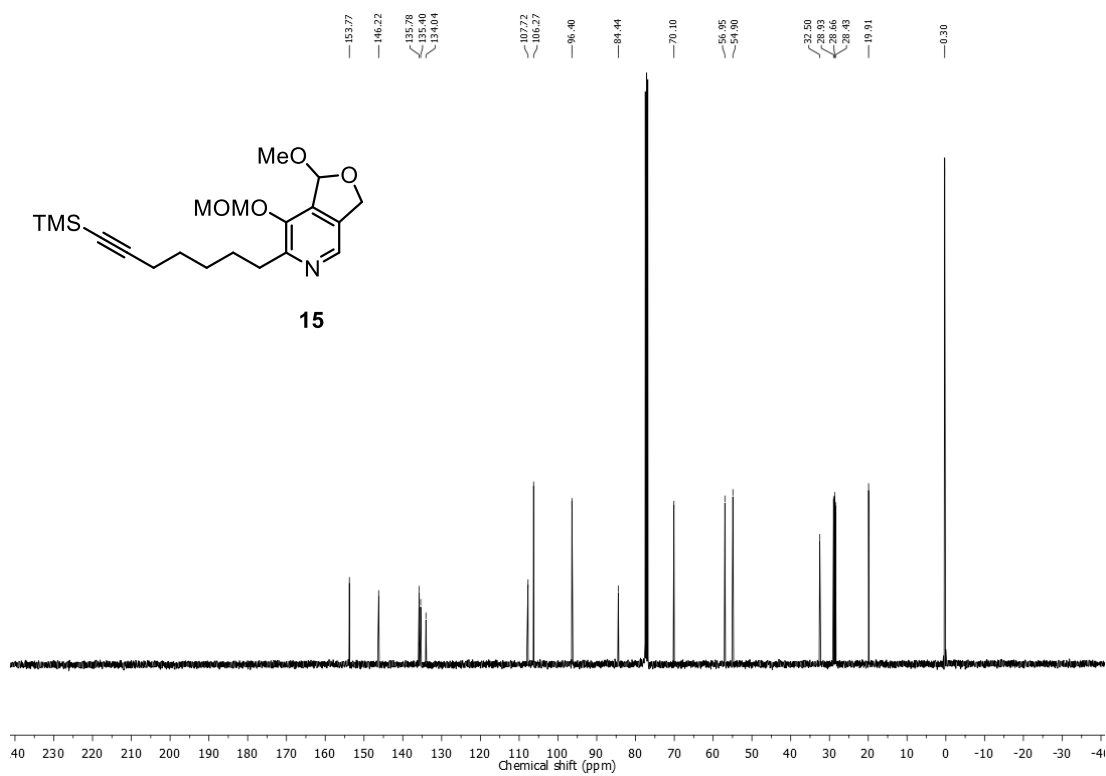
^{1}H NMR of compound **16** (400 MHz, CDCl$_3$)
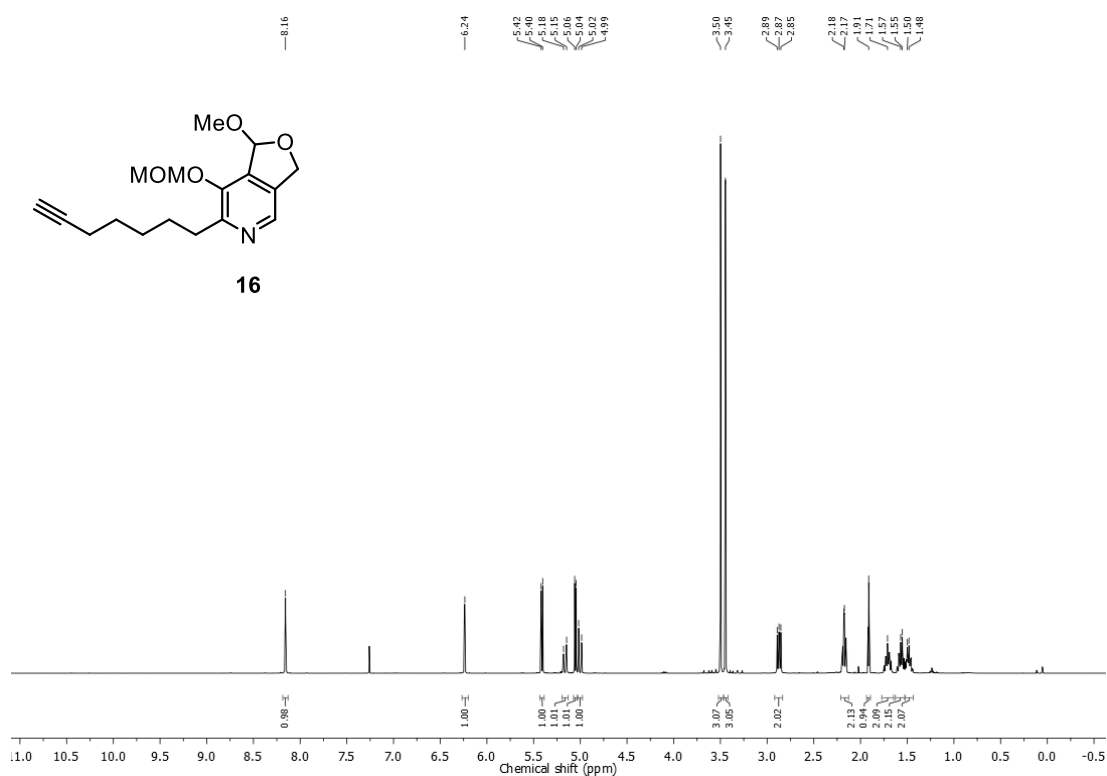
^{13}C NMR of compound **16** (100 MHz, CDCl$_3$)
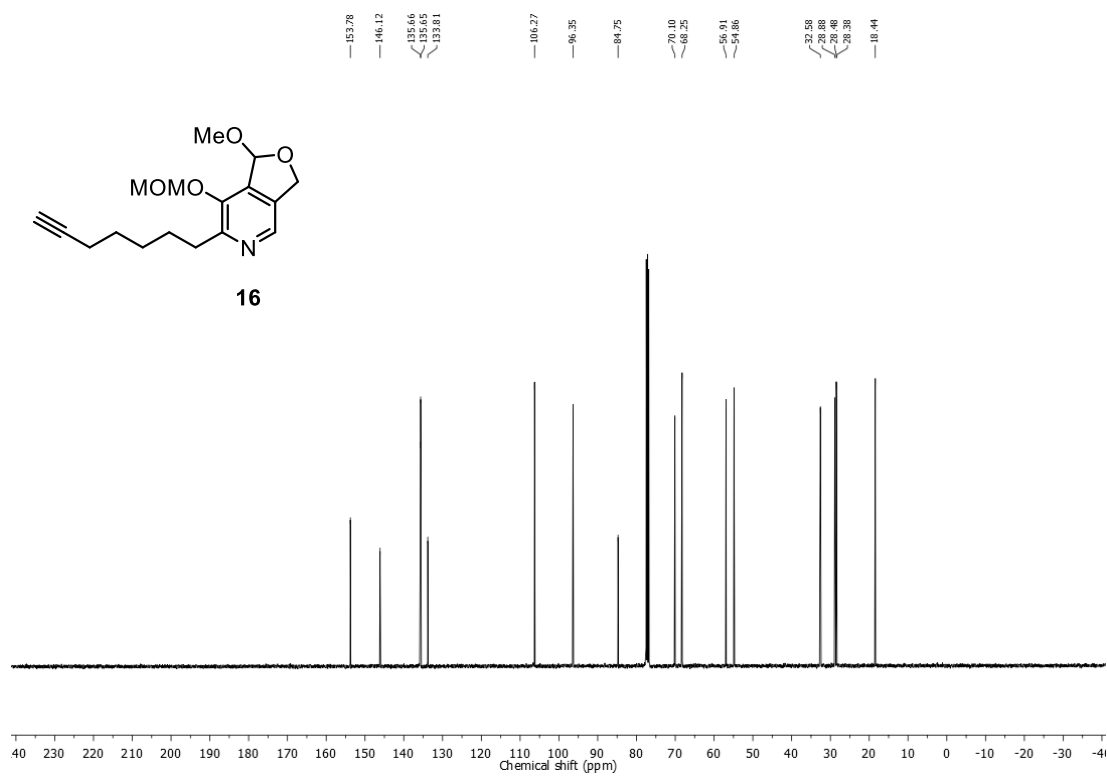
¹H NMR of compound **PL5** (400 MHz, DMSO-*d*₆)
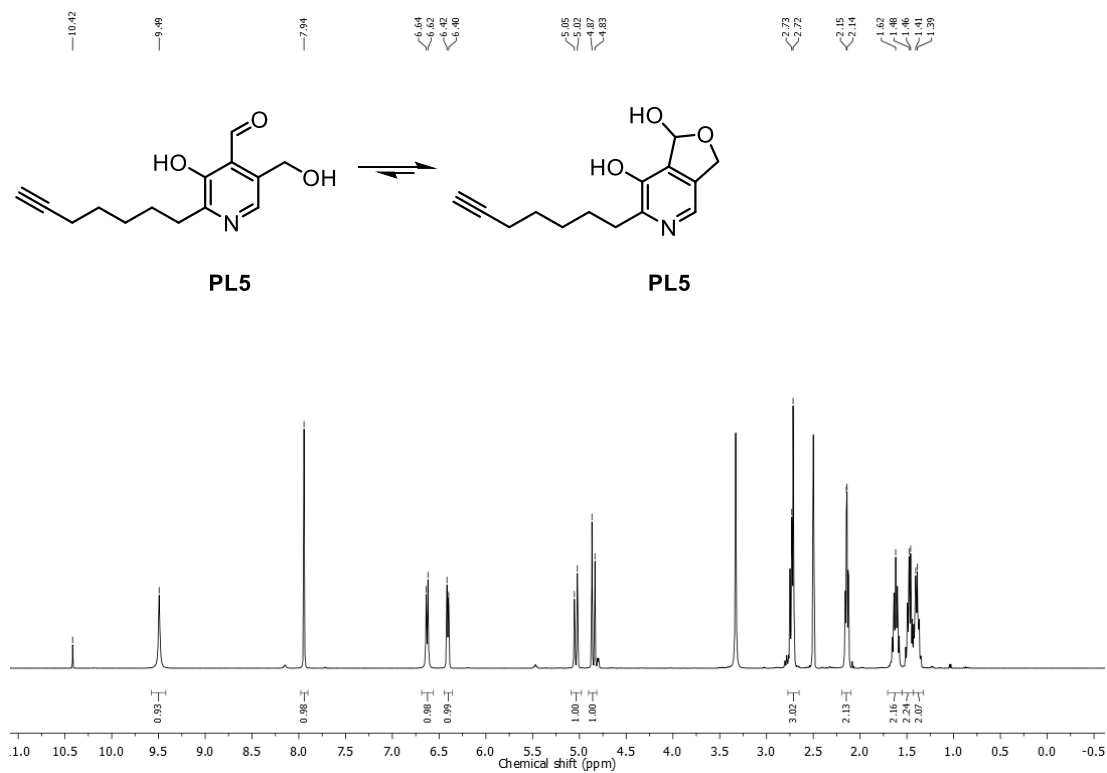
¹³C NMR of compound **PL5** (100 MHz, DMSO-*d*₆)
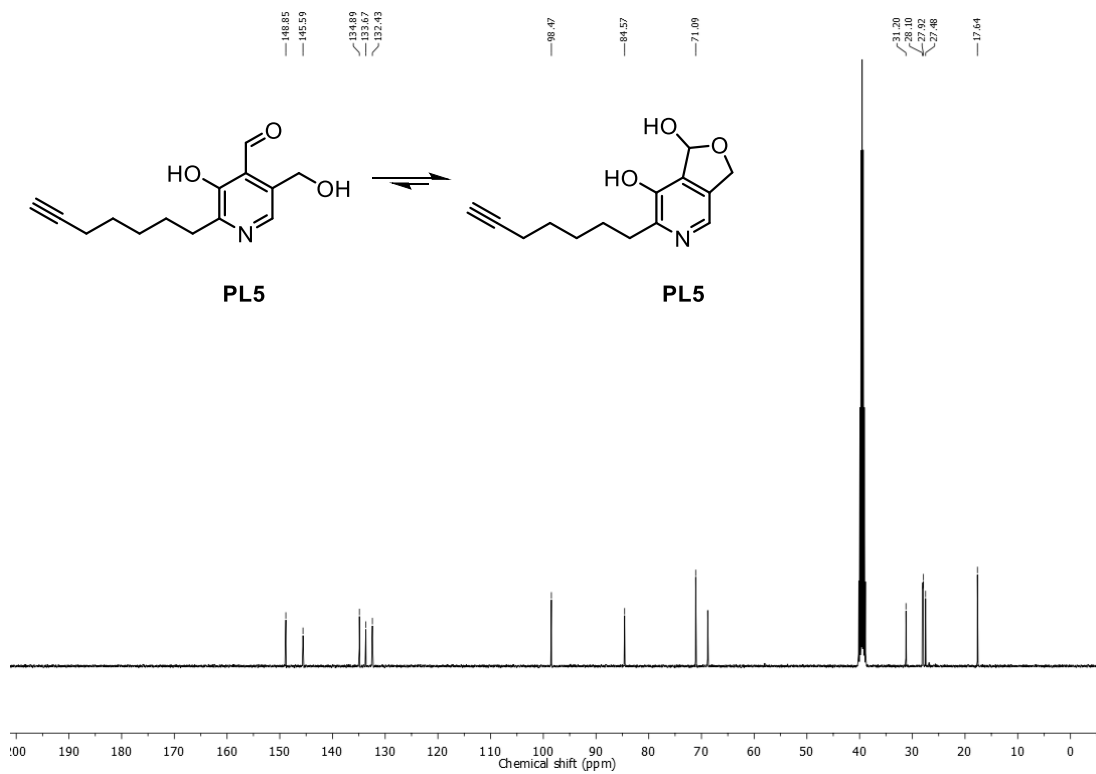
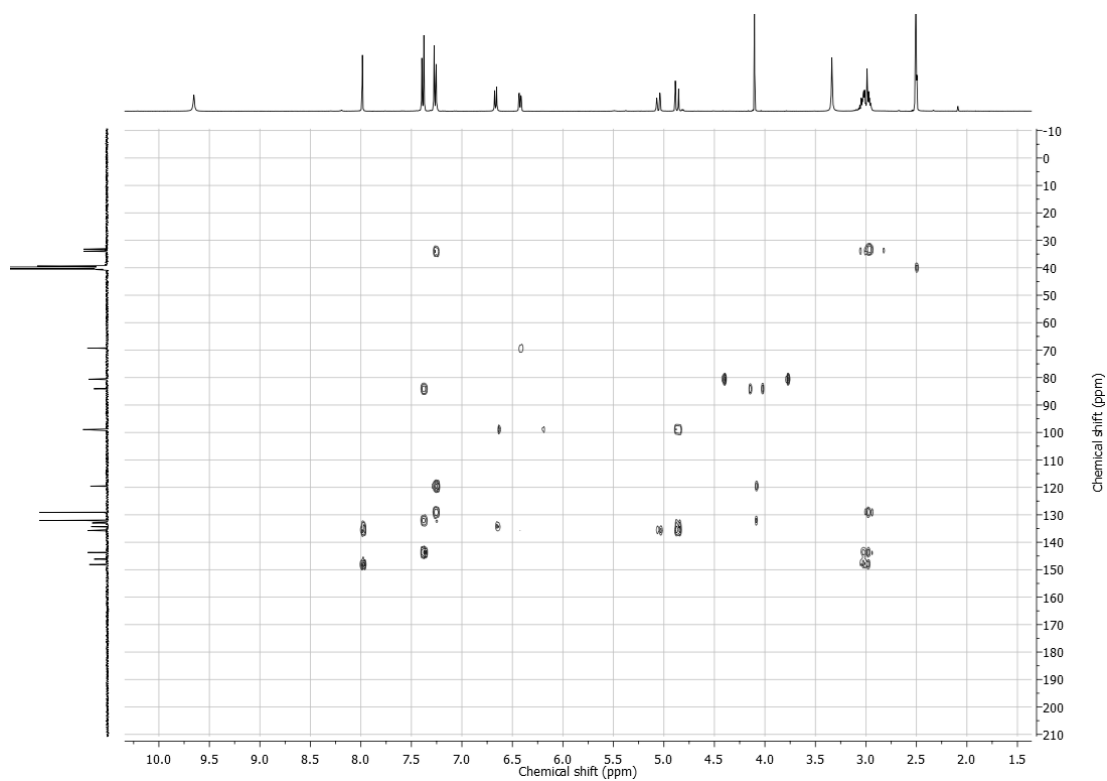HSQC spectrum of compound **PL5**

HMBC spectrum of compound **PL5**

COSY spectrum of compound **PL5**

^{1}H NMR of compound **17** (400 MHz, CDCl$_3$)
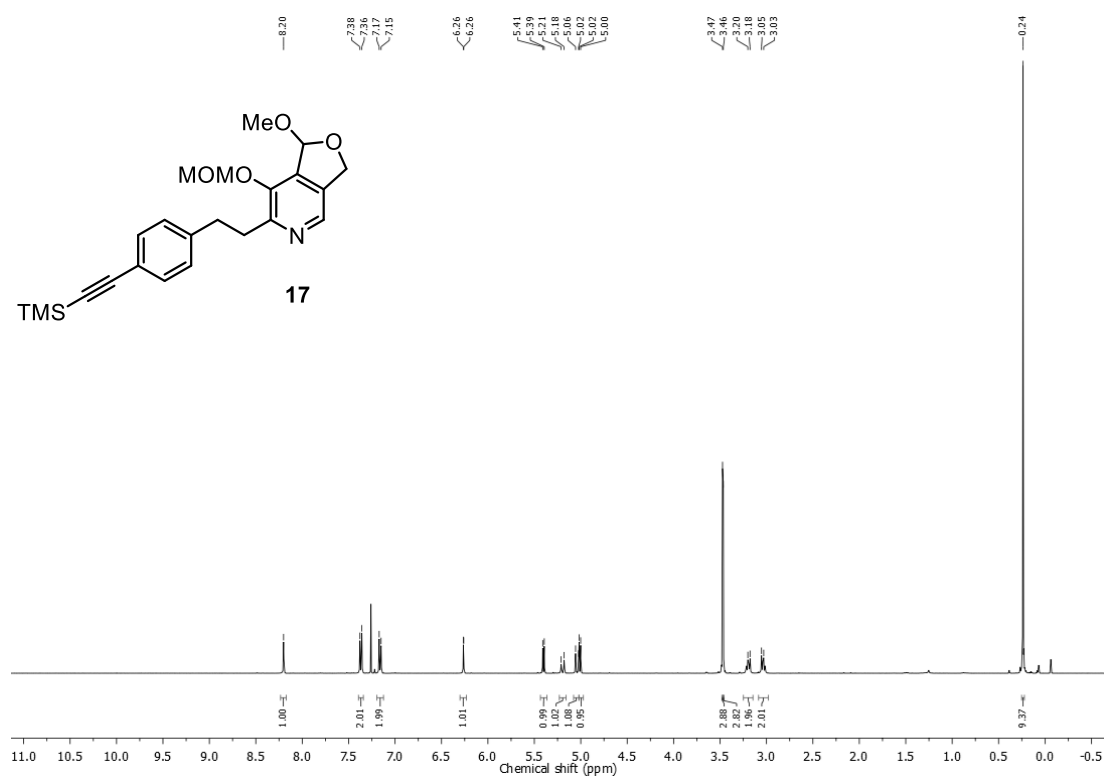
^{13}C NMR of compound **17** (100 MHz, CDCl$_3$)
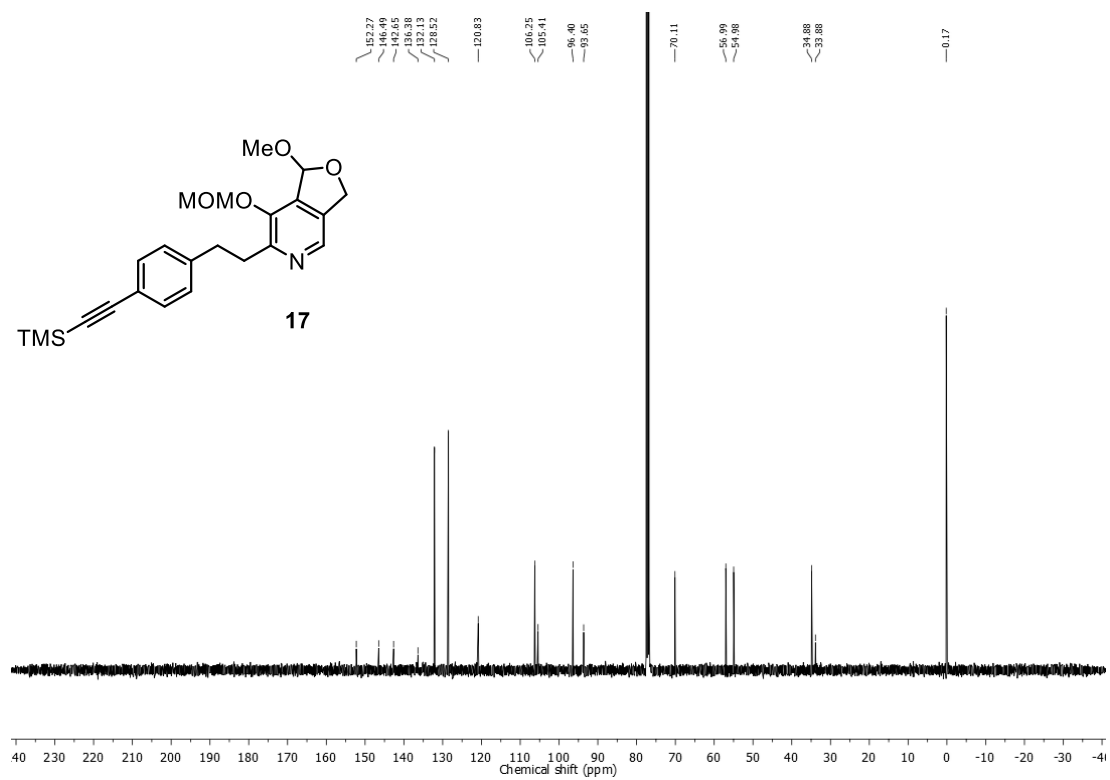
¹H NMR of compound **18** (400 MHz, CDCl₃)
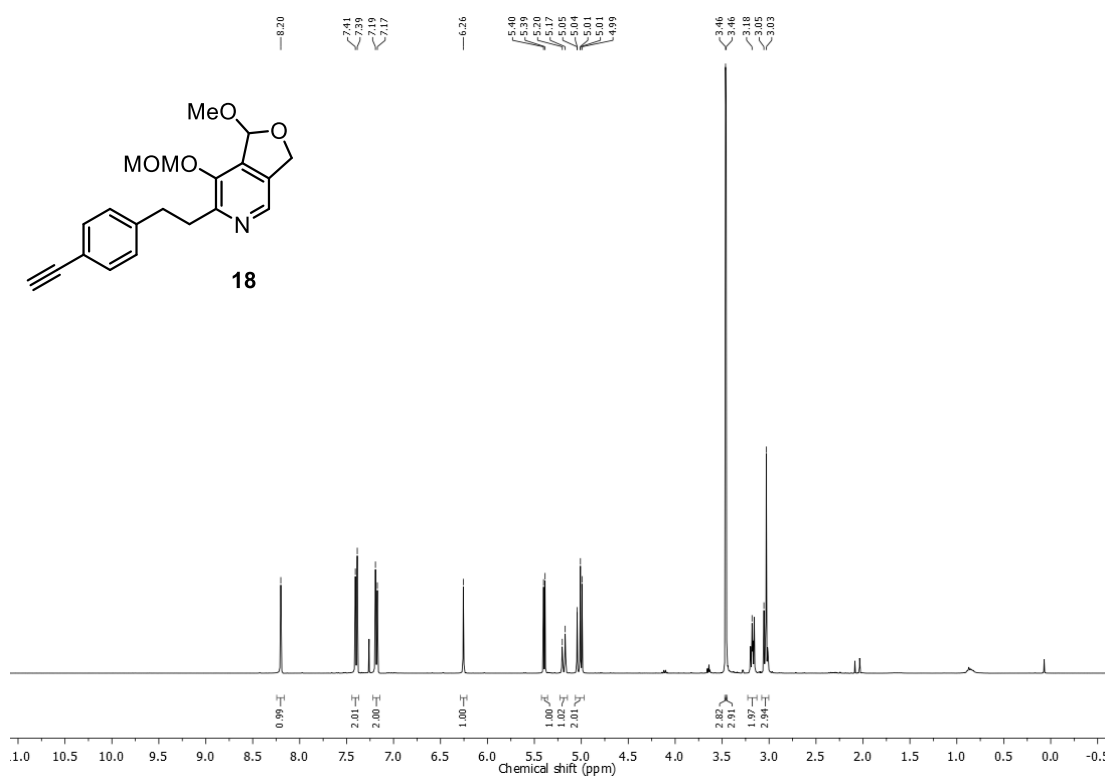
¹³C NMR of compound **18** (100 MHz, CDCl₃)
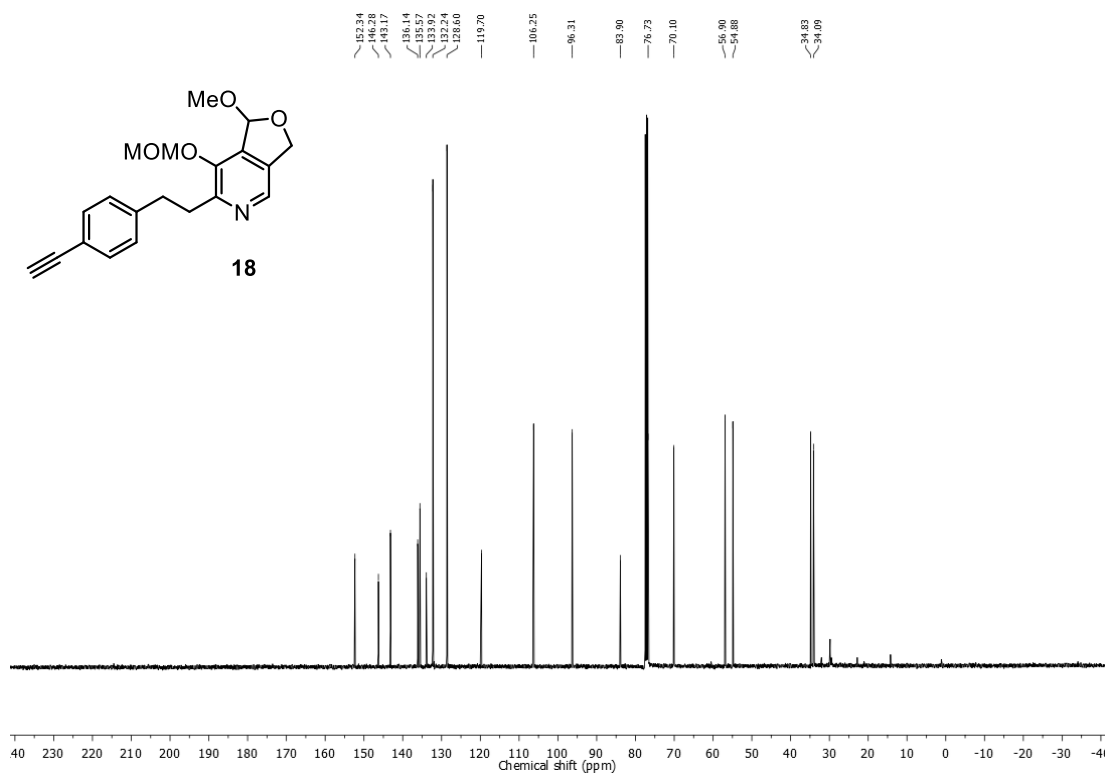
^{1}H NMR of compound **PL6** (400 MHz, DMSO-d_6)

^{13}C NMR of compound **PL6** (100 MHz, DMSO-d_6)
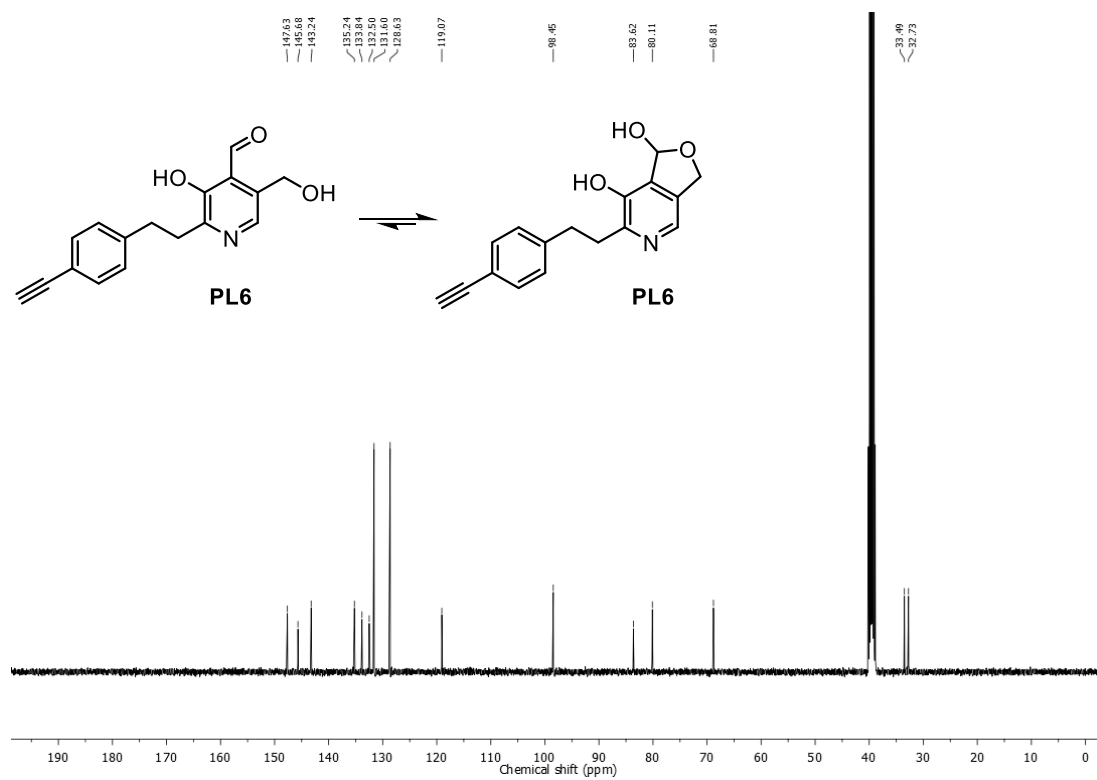
HSQC spectrum of compound **PL6**

HMBC spectrum of compound **PL6**

COSY spectrum of compound **PL6**
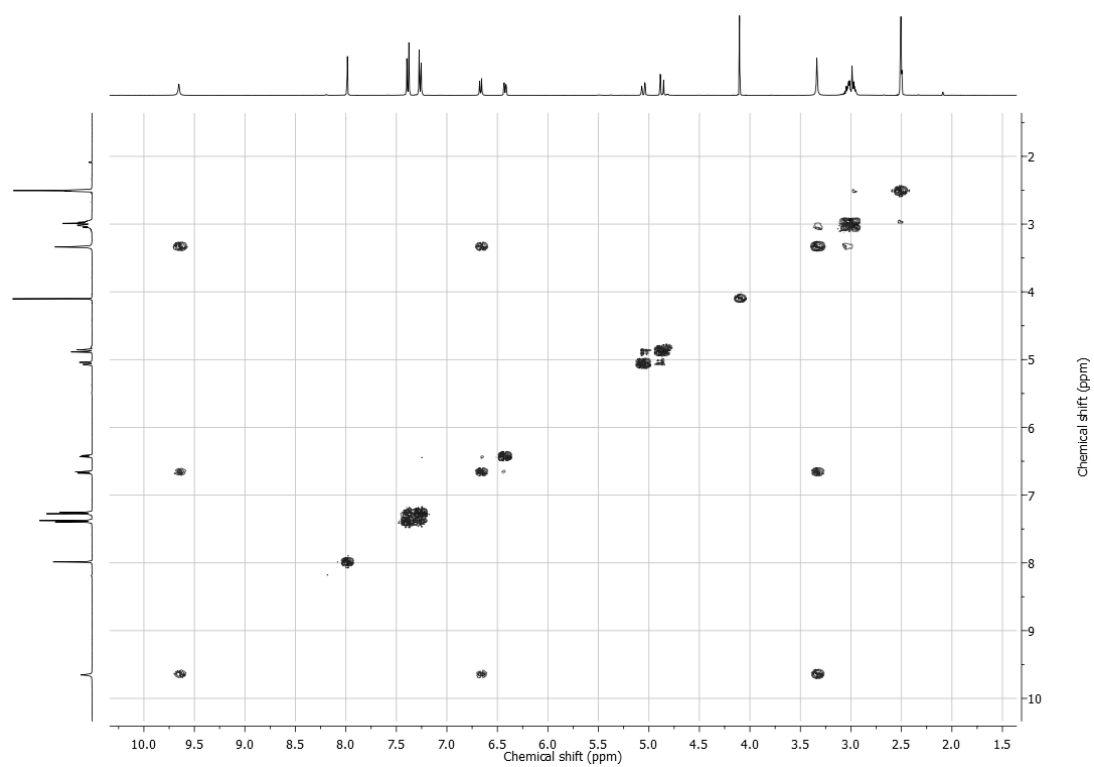
Curriculum Vitae

Annabelle Dorothée Hoegl

Date of birth   January 4, 1988
Wolfratshausen, Germany

Education

02.2014 – 03.2018   Ph.D. Chemistry
Technische Universität München, Germany

09.2011 – 10.2013   M.Sc. Chemistry
McGill University, Montréal, Canada

09.2006 – 05.2010   B.Sc. Biochemistry
McGill University, Montréal, Canada