Selective protein functionalisation via enzymatic phosphocholination

Philipp Ochtrop

Doctoral Thesis, Department of Chemistry
Umeå University, 2017
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>iv</td>
</tr>
<tr>
<td>List of Publications</td>
<td>viii</td>
</tr>
<tr>
<td>Enkel sammanfattning på svenska</td>
<td>x</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Protein labeling</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Fluorescent proteins</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Incorporation of bioorthogonal handles for protein labeling</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Tag-based protein labeling</td>
<td>7</td>
</tr>
<tr>
<td>1.5 <em>Legionella</em> enzymes AnkX and Lem3</td>
<td>11</td>
</tr>
<tr>
<td>2. Objectives</td>
<td>14</td>
</tr>
<tr>
<td>3. Covalent protein labeling by reversible enzymatic phosphocholination (Paper I)</td>
<td>16</td>
</tr>
<tr>
<td>3.1 Design and synthesis of functional CDP-choline derivatives</td>
<td>17</td>
</tr>
<tr>
<td>3.2 Evaluating CDP-derivatives as AnkX substrates</td>
<td>23</td>
</tr>
<tr>
<td>3.3 Identification of a minimal peptide recognition motif for the AnkX catalysed phosphocholination</td>
<td>26</td>
</tr>
<tr>
<td>3.4 Evaluation of model proteins for the AnkX catalysed phosphocholination</td>
<td>29</td>
</tr>
<tr>
<td>3.5 Fluorescent labeling of model proteins</td>
<td>31</td>
</tr>
<tr>
<td>4. Exploring the substrate scope for the AnkX catalysed phosphocholination (Paper II)</td>
<td>34</td>
</tr>
<tr>
<td>4.1 Synthesis of CDP-choline derivatives</td>
<td>35</td>
</tr>
<tr>
<td>4.2 Versatile functionalisation of Rab1b by AnkX phosphocholination</td>
<td>41</td>
</tr>
<tr>
<td>4.3 Dual labeling of proteins using AnkX and Lem3</td>
<td>44</td>
</tr>
<tr>
<td>5. Advancing AnkX mediated phosphocholination - ongoing work</td>
<td>48</td>
</tr>
<tr>
<td>5.1 Towards cell surface protein labeling</td>
<td>48</td>
</tr>
<tr>
<td>5.2 Covalent trapping of phosphocholine transfer</td>
<td>50</td>
</tr>
<tr>
<td>5.3 Development of cell-permeable CDP-choline analogues</td>
<td>56</td>
</tr>
<tr>
<td>6. Identification of cellular protein targets of the AnkX catalysed phosphocholination (Paper III)</td>
<td>61</td>
</tr>
<tr>
<td>6.1 Isolation and identification of AnkX substrate proteins from cell lysate</td>
<td>61</td>
</tr>
<tr>
<td>6.2 Investigation of IMPDH2 as a potential target of the AnkX catalysed phosphocholination</td>
<td>65</td>
</tr>
<tr>
<td>7. Summary and concluding remarks</td>
<td>69</td>
</tr>
<tr>
<td>8. Acknowledgements</td>
<td>73</td>
</tr>
<tr>
<td>References</td>
<td>75</td>
</tr>
<tr>
<td>Appendix</td>
<td>86</td>
</tr>
</tbody>
</table>
Abstract

Proteins are the most abundant biomolecules within a cell and are involved in all biochemical cellular processes ultimately determining cellular function. Therefore, to develop a complete understanding of cellular processes, obtaining knowledge about protein function and interaction at a molecular level is critical. Consequently, the investigation of proteins in their native environment or in partially purified mixtures is a major endeavour in modern life sciences. Due to their high chemical similarity, the inherent problem of studying proteins in complex mixtures is to specifically differentiate one protein of interest from the bulk of other proteins. Site-specific protein functionalisation strategies have become an indispensable tool in biochemical- and cell biology studies.

This thesis presents the development of a new enzymatic site-specific protein functionalisation strategy that is based on the reversible covalent phosphocholination of short amino acid sequences in intact proteins. A synthetic strategy has been established that allows access to functionalised CDP-choline derivatives carrying fluorescent reporter groups, affinity tags or bioorthogonal handles. These CDP-choline derivatives serve as co-substrates for the bacterial phosphocholinating enzyme AnkX from *Legionella pneumophila*, which transfers a phosphocholine moiety to the switch II region of its native target protein Rab1b during infection.

We identified the octapeptide sequence TITSSYYR as the minimum recognition sequence required to direct the AnkX catalysed phosphocholination and demonstrated the functionalisation of proteins of interest carrying this recognition tag at the N- or C-terminus as well as in internal loop regions.

Moreover, this covalent modification can be hydrolytically reversed by the action of the *Legionella* enzyme Lem3, which makes the labeling strategy the first example of a covalent and reversible approach that is fully orthogonal to current existing methodologies. Thus, the here presented protein functionalisation approach holds the potential to increase the scope of possible labeling strategies in complex biological systems.

In addition to the labeling of tagged target proteins, a CDP-choline derivative equipped with a biotin affinity-tag was synthesised and used in pull-down experiments to investigate the substrate scope of AnkX and to elucidate the role of protein phosphocholination during *Legionella pneumophila* infection.
List of Abbreviations

A, Ala    alanine
ABC       ammonium bicarbonate
AGT       alkylguanine-DNA transferase
AMP       adenosine monophosphate
anhyd.    anhydrous
AnkX      ankeryn repeat containing protein X
aq.       aqueous
ATP       adenosine triphosphate
BARAC     biarylazacyclooctynes
BirA      biotin ligase
Boc       tert-butylxcarbonyl
C, Cys    cysteine
cAMP      cyclic adenosine monophosphate
CBS       cystathionine beta synthase / Bateman domain
Cbz       carboxybenzyl
CDP       cytidine diphosphate
CFP       cyan fluorescent protein
CMP       cytidine monophosphate
CoA       coenzyme A
CuCAAC    copper (I) catalysed alkyne-azide cycloaddition
D, Asp    aspartic acid
Da        dalton
DBCO      dibenzocyclooctynes
DCC       dicyclohexylcarbodiimid
DCM       dichloromethane
DIC       differential interference contrast
DIFO      difluorocyclooctyne
DMEM      Dulbecco’s modified eagle medium
DMF       dimethyl formamide
DNA       deoxyribonucleic acid
DTT       dithiothreitol
E, Glu    glutamic acid
E. coli   Escherichia coli
EDTA      Ethylenediaminetetraacetic acid
EGTA      ethylene glycol-bis(β-aminoethyl ether)-N,N',N'-tetraacetic acid
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ESI-MS</td>
<td>electrospray-ionisation mass spectrometry</td>
</tr>
<tr>
<td>F, Phe</td>
<td>phenylalaine</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FGE</td>
<td>formylglycine-generating enzyme</td>
</tr>
<tr>
<td>Fic</td>
<td>filamentation induced by cAMP</td>
</tr>
<tr>
<td>FL</td>
<td>fluorescein</td>
</tr>
<tr>
<td>Fmoc</td>
<td>fluorenlymethoxycarbonyl</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>G, Gly</td>
<td>glycine</td>
</tr>
<tr>
<td>GDP</td>
<td>guanosine diphosphate</td>
</tr>
<tr>
<td>GEF</td>
<td>guanosine exchange factor</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>GTPase</td>
<td>guanosine triphosphatase</td>
</tr>
<tr>
<td>H, His</td>
<td>histidine</td>
</tr>
<tr>
<td>HBI</td>
<td>4-(p-hydroxybenzylidene)-imidazolidin-5-one</td>
</tr>
<tr>
<td>HEK cells</td>
<td>human embryonic kidney cells</td>
</tr>
<tr>
<td>HPLC</td>
<td>high-performance liquid chromatography</td>
</tr>
<tr>
<td>HRMS</td>
<td>high-resolution mass spectrometry</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>hY2R</td>
<td>human neuropeptide Y2 receptor</td>
</tr>
<tr>
<td>I, Ile</td>
<td>isoleucine</td>
</tr>
<tr>
<td>IMP</td>
<td>inosine-5'-monophosphate</td>
</tr>
<tr>
<td>IMPDH2</td>
<td>inosine-5'-monophosphate dehydrogenase typ II</td>
</tr>
<tr>
<td>K, Lys</td>
<td>lysine</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo dalton</td>
</tr>
<tr>
<td>L, Leu</td>
<td>leucine</td>
</tr>
<tr>
<td>LC/MS</td>
<td>liquid chromatography/mass spectrometry</td>
</tr>
<tr>
<td>LCV</td>
<td><em>Legionella</em> containing vacuole</td>
</tr>
<tr>
<td>LplA</td>
<td>lipoic acid ligase</td>
</tr>
<tr>
<td>m/z</td>
<td>mass to charge ratio</td>
</tr>
<tr>
<td>MBP</td>
<td>maltose binding protein</td>
</tr>
<tr>
<td>MeCN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>N, Asn</td>
<td>asparagine</td>
</tr>
<tr>
<td>NAD+</td>
<td>nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NADH</td>
<td>nicotinamide adenine dinucleotide, reduced form</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>NBD</td>
<td>4-chloro-7-nitrobenzofurazan</td>
</tr>
<tr>
<td>NHS</td>
<td>N-hydroxysuccinimid</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>o.n.</td>
<td>over night</td>
</tr>
<tr>
<td>P, Pro</td>
<td>proline</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PC</td>
<td>phosphocholine</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PIPES</td>
<td>1,4-Piperazinediethanesulfonic acid</td>
</tr>
<tr>
<td>POCl₃</td>
<td>phosphorus oxychloride</td>
</tr>
<tr>
<td>POI</td>
<td>protein of interest</td>
</tr>
<tr>
<td>PPh₃</td>
<td>triphenylphosphine</td>
</tr>
<tr>
<td>PPTase</td>
<td>phosphopantetheiny-transferase</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>Q, Gln</td>
<td>glutamine</td>
</tr>
<tr>
<td>R, Arg</td>
<td>arginine</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>ROS</td>
<td>reactive oxygen species</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>S, Ser</td>
<td>serine</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SPAAC</td>
<td>strain-promoted alkyne-azide cycloadditions</td>
</tr>
<tr>
<td>SPPS</td>
<td>solid phase peptide synthesis</td>
</tr>
<tr>
<td>SrtA</td>
<td>sortase</td>
</tr>
<tr>
<td>STV</td>
<td>streptavidin</td>
</tr>
<tr>
<td>SUMO</td>
<td>small ubiquitin-related modifier</td>
</tr>
<tr>
<td>t-BuOH</td>
<td>tert-butanol</td>
</tr>
<tr>
<td>T, Thr</td>
<td>threonine</td>
</tr>
<tr>
<td>TAMRA</td>
<td>tetramethylrhodamine</td>
</tr>
<tr>
<td>TBS</td>
<td>tris buffered saline</td>
</tr>
<tr>
<td>TCO</td>
<td>trans-cyclooctene</td>
</tr>
<tr>
<td>TEV</td>
<td>tobacco etch virus</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>TLC</td>
<td>thin-layer chromatography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>TOF</td>
<td>time of flight</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer ribonucleic acid</td>
</tr>
<tr>
<td>UAA</td>
<td>unnatural amino acid</td>
</tr>
<tr>
<td>V, Val</td>
<td>valine</td>
</tr>
<tr>
<td>W, Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>XMP</td>
<td>xanthosine-5'-monophosphate</td>
</tr>
<tr>
<td>Y, Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>YFP</td>
<td>yellow fluorescent protein</td>
</tr>
</tbody>
</table>
List of Publications

This thesis is based on the following papers and manuscripts, which are attached in the end and referred to in the text by the corresponding Roman numerals.


*These authors contributed equally to this work.

Papers have been reprinted with the permission from the publishers.
**Author's contributions**

Paper I: All of the organic synthesis of CDP-choline derivatives, all synthesis of the phosphocholinated and non-phosphocholinated peptides, minor writing of the manuscript, translation of the paper for *Angewandte Chemie*, German edition.

Paper II: Major planning, major part of the organic synthesis of CDP-choline derivatives and dyes, minor contribution to the analysis of protein functionalisation, writing the manuscript.

Paper III: Minor planning, organic synthesis of the CDP-choline biotin probe, all protein labeling and protein isolation experiments, analysis of the proteomic data, planning and design of phosphocholination- and activity assay of IMPDH2, writing the manuscript.


Vi har vidare undersökt vilka proteiner AnkX reagerar med inuti celler, vi använde oss av ett CDP-kolinderivat funktionaliserat med biotin, vilket har tillåtit oss att fiska ut alla de proteiner som fosfokolineras av AnkX. Förutom de små GTPaserna i Rab-familjen så identifierade vi även IMPDH2, ett enzym som reglerar det hastighetsbestämmande steget i syntesen av guanosin-nukleotider. Detta är mycket intressant, eftersom det leder till frågan ifall *Legionella pneumophila* manipulerar sin värdcell genom att förändra mängden GTP i förhållande till ATP.
1. Introduction

1.1 Protein labeling

An understanding of protein function at a molecular level is critical for elucidating biochemical cellular processes and how they influence diseases. To study the function of a specific protein of interest (POI) it is often necessary to modify the heterologously expressed and purified POI with a unique functionality that is not defined by the genetic code. These functionalities can range from natural post-translational modifications (PTMs) to biophysical probes or affinity handles. PTMs, for example, play an important role in tuning the physiochemical properties of proteins by modulating enzymatic activity or controlling protein-protein recognition. The installation of a PTM onto the POI is therefore essential to generate the physiological relevant and active form of the protein, which is necessary to study and determine its exact biological function. The incorporation of a physical probe on the other hand adds a new and unique functionality to the POI that is not found elsewhere in a biological environment. For example labeling a protein with a fluorescent dye allows tracking of its localisation or trafficking within the complex environment of living cells.

As such synthetic proteins have proved to be valuable tools across a range of scientific disciplines, several techniques have been developed to equip proteins with various non-natural functionalities. Initial experiments aimed at the direct functionalisation of nucleophilic amino acid residues on the protein surface with suitable electrophiles like N-hydroxysuccinimidy (NHS) esters. Given the high abundance of nucleophilic lysine (Lys) side chains on the surface of most proteins, this reaction offers poor control over the position of modification and results in non-homogenous protein labeling. This unspecific labeling may disturb the structure and function of the targeted protein. Furthermore, this approach does not discriminate between different proteins in a mixture, which makes it only applicable for purified proteins. Another approach to directly label surface exposed amino acid residues is to address the naturally less abundant nucleophilic amino acid cysteine (Cys). The thiol group of Cys can be targeted with maleimide-functionalised probes in Michael addition reactions. However, this approach is also limited to purified proteins or other cysteine free environments, excluding any in cell or in vivo labeling.

While these direct chemical functionalisation methodologies may be sufficient for in vitro applications with purified proteins, other more demanding studies require more advanced labeling techniques that result in a defined modification of the target protein. In particular investigations of protein localisation and dynamics in their natural environment require
fluorescent labeling techniques that are protein-specific and applicable in protein mixtures or even in live cells or animals. The most commonly utilised strategy is to employ intrinsically fluorescent proteins that can be fused to the protein of interest.

1.2 Fluorescent proteins

Established almost 20 years ago, fluorescent proteins have become indispensable tools in cell biology to study the localisation and interaction of proteins inside living cells and animals. In principle, the genetic information of a fluorescent protein can be fused to any gene of interest. The translation of this genetic fusion results in the POI connected to the fluorescent protein via its N- or C-terminus and allows the visualisation of its spatial and temporal distribution by fluorescence microscopy (Figure 1a). The most prominent and first isolated example of such fluorescent proteins is the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*. GFP is a 27 kDa protein that folds into a 11-β-sheet barrel and gains its green fluorescence from a spontaneous cyclisation of the tripeptide Ser65-Tyr66-Gly67 to the chromophore 4-(p-hydroxybenzylidene)-imidazolidin-5-one (HBI) (Figure 1c). HBI exhibits an excitation maximum at 488 nm and an emission maximum at 510 nm. Genetic engineering of the amino acids that contribute to the chromophore, as well as isolation of other fluorescent proteins from different jellyfish and corals, has produced an array of available fluorescent proteins with colors ranging from cyan (CFP), yellow (YFP), orange (mOrange) and red (mCherry) to far-red (mPlum). This large selection of fluorescent proteins allows the convenient multicolor labeling of various proteins in different tissues and organisms, making fluorescent protein labeling the method of choice for many applications. However, there are also disadvantages in creating fluorescent protein fusions. As GFP is relatively large itself, its fusion to POIs always adds a substantially sized protein of around 27 kDa to the POI. This increase in size may perturb the structure, function, or cellular localisation of the investigated protein. This can be particularly detrimental when studying proteins that are part of protein-protein interaction networks like in signal transduction. Furthermore, the photostability and photobleaching properties of the fluorescent proteins are often inferior to those of small molecule fluorophores.
Figure 1: The green fluorescent protein (GFP). a) A gene of interest is genetically fused to the GFP gene. The transcription and translation into a protein results in the protein of interest connected to GFP that is excited with blue light (488 nm) and emits green light (510 nm). b) Crystal structure of GFP with highlighted chromophore HBI. c) Cyclisation of amino acids Ser65, Tyr66 und Gly67 to form the chromophore HBI: i) cyclisation, ii) dehydration, iii) oxidation.

1.3 Incorporation of bioorthogonal handles for protein labeling

To overcome the drawbacks of fluorescent protein fusions, significant effort has been spent to develop techniques for protein functionalisation that work with smaller protein fusions or do not need any further amino acid addition. The smallest possible modification of a POI is the incorporation of an unnatural amino acid (UAA) into the polypeptide chain of a POI (Figure 2).12,13 One of the most powerful approaches to do this is genetic code expansion. Here, an orthogonal aminoacyl-tRNA synthetase / tRNA pair is used to specifically insert an unnatural amino acid during mRNA translation in response to an amber stop codon (UAG) (Figure 2b). This codon can be placed at a defined position in the gene of interest and thereby allows the exact positioning of the new functionality into the POI.14 The orthogonal
aminoacyl-tRNA and the corresponding tRNA are often imported from other organisms and introduced to the heterologous host used for the protein expression. An orthogonal aminoacyl-tRNA synthetase does not load any endogenous tRNAs in the host cell, but specifically aminoacylates its cognate orthogonal tRNA with the UAA (Figure 2a). Furthermore, an orthogonal tRNA is not a substrate for endogenous aminoacyl-tRNA synthetases, but is a specific substrate for the orthogonal synthetase. This approach allows rapid and chemoselective labeling at a single site in the protein when the UAA that is incorporated into the POI contains an appropriate bioorthogonal handle for functionalisation.

![Figure 2: Incorporation of unnatural amino acids (UAA). a) Orthogonal aminoacyl-tRNA synthetase loads its partner tRNA with the UAA. b) The loaded orthogonal tRNA recognizes the stop codon UAG in the mRNA and the UAA is incorporated by the ribosome into the growing peptide chain.](image)

Suitable chemical reactions for such selective labeling in biological environments have to fulfill several criteria. Bioorthogonal reactions have to be selective for the incorporated functional group and the labeling probe. Also, reactions have to proceed under physiological conditions in aqueous systems at pH around 6-8 and ambient temperatures. Most importantly, bioorthogonal reactions should not cross-react with functional groups that are present in proteins or other biomolecules like DNA, lipids or other metabolites. Furthermore, it is desirable that bioorthogonal reactions produce stable covalent bonds and that the incorporated functionality, as well as the reactive probe, are stable before the reaction and non-toxic to living systems.

In recent years considerable effort has driven the development of bioorthogonal reactions that fulfill most of these criteria (Figure 3, page 5). Among the first functionalities that have been explored for bioorthogonal probes are aldehydes and ketones. Ketones are small in size and easily accessible by organic synthesis. Furthermore, they are conveniently installed into biomolecules by metabolic incorporation in the form of ketone
containing analogues of natural substrates.\textsuperscript{19,20} Once installed in biomolecules, the carbonyl group of ketones and aldehydes can be addressed with strong α-effect nucleophiles, such as hydrazines, hydrazides or alkoxyamines, to form hydrazone and oxime products respectively under slightly acidic conditions (pH 4-6) (Figure 3a).\textsuperscript{21,22} However, due to the required acidic reaction conditions and competition with other carbonyl-bearing metabolites like pyruvate, sugars and various co-factors, these chemo-selective reactions are not fully bioorthogonal and are best suited for \textit{in vitro} or cell surface labeling.\textsuperscript{23,24}

In contrast to aldehydes and ketones, the azide group is truly absent in biological systems and completely orthogonal to other biological functionalities.\textsuperscript{25} Further advantages of the azide group include its small size, relative stability and its easy incorporation into glycans,\textsuperscript{26-28} proteins,\textsuperscript{29} lipids\textsuperscript{30,31} and nucleic acids.\textsuperscript{32-33} First described 50 years ago, azides were found to act as 1,3-dipoles in [3+2] cycloadditions with alkynes to form stable triazoles under high temperatures and pressure.\textsuperscript{34} The cycloaddition of azides and alkynes became available for application in biological systems when copper (I) salts were found to catalyse triazole formation under milder reaction conditions (Figure 3b).\textsuperscript{35,36} This copper (I) catalysed alkyne-azide cycloaddition (CuCAAC), now widely called "click reaction", proceeds faster in water than in organic solvents and the reactants, azide and alkyne, are absent from biological systems, which is the reason for the popularity of this reaction in biological studies.\textsuperscript{37,38} The main disadvantage of the CuCAAC is the toxicity of Cu(I)-salts to biological systems at concentrations around 500 µM.\textsuperscript{37} The toxic effect of copper salts originates from its catalytic generation of reactive oxygen species (ROS) that cause cellular stress. The ROS formation can be circumvented by the application of Cu ligands\textsuperscript{39,40} or by avoiding the use of copper and introducing ring strain into the reacting alkyne, thus making the Cu(I) catalyst unnecessary.\textsuperscript{41,42}

The reaction of strained alkynes with azides is commonly termed "strain-promoted alkyne-azide cycloadditions" (SPAAC) and typically uses cyclooctyne derived probes to label azide bearing biomolecules without the assistance of toxic Cu(I) (Figure 3c).\textsuperscript{43} The development of more reactive probes like α-difluorocyclooctyne (DIFO),\textsuperscript{44} dibenzocyclooctynes (DBCO)\textsuperscript{45} and biarylazacycloclooctynes (BARAC)\textsuperscript{46} derivatives improved the reaction kinetics of the SPAAC (1 M⁻¹ s⁻¹) into ranges comparable the CuCAAC (10⁻²00 M⁻¹ s⁻¹).

Another strain promoted cycloaddition that is often used for bioconjugation reactions is the reaction of strained alkenes with tetrazines in an inverse-electron-demand Diels-Alder cycloaddition (Figure 3d).\textsuperscript{47} The exploration of this reaction in aqueous media demonstrated that highly strained alkenes, like \textit{trans}-cyclooctene (TCO), react extraordinary fast with mono-aryl tetrazines and can be used to modify proteins \textit{in vitro} and \textit{in vivo}.
Currently this reaction exhibits the fastest reaction kinetics for available bioorthogonal reactions \(10^3\text{ to }10^4 \text{ M}^{-1}\text{s}^{-1}\). Furthermore, if red fluorophores are connected to the tetrazine, their fluorescence is quenched and becomes active upon conjugation with the strained alkene that is installed in the biomolecule. This fluorogenic property reduces the background for labeling reactions and makes the signal-to-noise of this labeling reaction superior.

**Figure 3:** Biocompatible and bioorthogonal reactions for selective bioconjugations. 

- **a)** Aldehydes or ketones react with hydrazines or alkoxyamines to form hydrazones or oximes.
- **b)** Azides and alkynes form triazoles in copper catalysed cycloaddition reactions.
- **c)** Azides react with strained alkynes in strain-promoted cycloaddition reactions to form bicyclic triazoles.
- **d)** Reaction of trans-cyclooctenes and tetrazines to form dihydropyridazines.
However, the incorporation of an UAA with bioorthogonal functionalities into proteins often suffers from low efficacies and is very challenging as strong expertise in engineering of suitable orthogonal aminoacyl-tRNA synthetases is needed. Therefore other methods are required that allow a more convenient installation of bioorthogonal groups or biological probes into proteins of interest.

1.4 Tag-based protein labeling

More straightforward techniques for the site-specific protein labeling can be summarised by the term tag-based protein labeling. These techniques are similar to the approach of fluorescent protein fusion as they are also based on genetic fusion of additional amino acids to the POI. In contrast to fluorescent proteins, these techniques utilise smaller sized amino acid sequences as recognition motifs. These tags are able to recognise and bind a small-molecule probe or are recognised by enzymes that transfer the desired label onto a specific amino acid residue within the sequence. Strategies for such tag-based protein functionalisation can be further grouped into three different approaches; a) self-labeling tags, b) self-labeling proteins and c) tags for enzyme mediated labeling.

Figure 4: The fluorogenic dye ReAsH is recognised by the characteristic hexapeptide motif Cys-Cys-Pro-Gly-Cys-Cys of the tetra-cysteine tag and is "turned on" upon binding.

1.4.1 a) Self-labeling tags are very small and contain only a few amino acids that are sufficient to bind small organic molecules. For example, the tetra-cysteine tag method uses a short hexapeptide sequence that is genetically encoded into POIs. This tag exploits the large affinity of the Cys-Cys-Pro-Gly-Cys-Cys sequence towards bisarsenical-functionalised fluorescent dyes such as the fluorescein derivative FlAsH and the resorufin derivative ReAsH (Figure 4). The arsenic functions of these cell permeable dyes are masked with 1,2-ethanediethiol and undergo a thiol exchange reaction with the thiol groups of the tetra-cysteine motive. An advantage of
the bisarsenical dyes is their fluorogenic properties. FlAsH and ReAsH are essentially nonfluorescent in solution and are switched on upon binding to the tag, which enhances the signal-to-noise ratio in labeling reactions. Nevertheless, the bisarsenical dyes also come with limitations. Arsenic containing compounds are generally cytotoxic, which limits their utilisation in live cell or animal experiments. Furthermore, these hydrophobic dyes are known to bind nonspecifically to membranes or off-target cellular proteins, demanding extra washing steps in the presence of competing thiols, resulting in decreased selectivity.

1.4.2 b) Self-labeling proteins like the SNAP-, CLIP- or Halo-tag offer a much higher selectivity than the cysteine based self-labeling approaches. These techniques are based on the fusion of engineered proteins with enzymatic activity to the POIs. The introduced enzyme recognises and binds a small molecule to subsequently attach a functional probe to the fusion protein (Figure 5). The SNAP-tag for example is derived from the DNA repair enzyme O-6-alkylguanine-DNA alkyltransferase (AGT) and binds to O-6-benzylguanines to transfers the benzyl group, along with the desired label, to a cysteine in the active site.

![Figure 5](image)

**Figure 5:** Self-labeling proteins. a) The SNAP-tag recognises O-6-benzylguanine derivatives and transfers the functionalised benzyl group covalently to a cysteine in the active site. b) The engineered CLIP-tag is a variant of the SNAP-tag and uses O-6-benzylcytosine derivatives to transfer the functionalised benzyl group. c) The Halo-tag is a modified variant of haloalkane dehalogenase and covalently reacts with 1-chlorohexyl derivatives.
The CLIP-tag is a similar technique, which utilises another engineered AGT and was developed as an alternative and orthogonal approach to the SNAP-tag. The CLIP-tag selectively reacts with O-6-benzylcytosine derivatives instead and can be used in parallel with the SNAP-tag, offering the possibility to directly label two different proteins in the same cell. The Halo-tag is an engineered variant of the bacterial enzyme haloalkane dehalogenase. The Halo-tag recognises functionalised alkyl halides as substrates and covalently traps the alkyl residues in its active site. Given their fast kinetics, specificity and orthogonality, all of these tags are well suited for the protein labeling within cells. Nevertheless, since these tags are engineered enzymes they are much larger then the tetra cysteine tag, and may therefore interfere with the function of proteins.

1.4.3 c) Tags for enzyme-mediated labeling combine the small tag size provided by self-labeling tags and the specificity shown by the self-labeling proteins. These approaches make use of engineered enzymes that recognise a peptide sequence fused to a protein of interest and covalently attach a labeled substrate to it. The labeling probes are usually analogues of enzyme substrates that contain either a fluorophore or a bioorthogonal handle that can direct the subsequent labeling reaction (Figure 6).

![Figure 6: Enzyme mediated labeling. a) In one-step labeling approaches, engineered enzymes use functionalised co-substrates to site-specifically label small amino acid tags in a protein of interest. b) Two-step labeling approaches use enzymes to covalently transfer a specific functionality to protein tags that can subsequently be addressed with complementary probes.](image)

Currently, the shortest recognition tags available are used in combination with the enzymes sortase (SrtA) and formylglycine-generating enzyme (FGE). Both enzymes recognise a five amino acid sequence at the N- or C-terminus of POIs. While SrtA can only be used to attach a peptide or protein containing multiple glycine, to its LPXTG recognition tag, the
formylglycine generating enzyme converts a cysteine within the motif CXPXK to the aldehyde-bearing residue formylglycine. As the aldehyde function is an otherwise naturally not occurring functionality, it can be selectively addressed with exogenous groups that react specifically with aldehydes, like functionalised hydrazine- or aminooxy-containing small molecule probes in a second labeling step. Another two-step labeling approach uses the enzyme biotin ligase (BirA) from Escherichia coli (E. coli). BirA recognises a 15 amino acid long peptide that is fused to the POI and transfers a biotin moiety, which can be subsequently visualised with streptavidin-fluorophore conjugates. Variants of biotin ligase from other species can also be used to label proteins with biotin derivatives that are functionalised with alkyne or azide groups. Once attached to POIs, these functionalities can be addressed by CuCAACA-chemistry with a variety of small-molecule probes. However, the presence of biotin and endogenous biotinylated proteins limits the application of BirA in cellular imaging.

These previously introduced two-step labeling approaches benefit from their high flexibility and the variety of labels that can be attached to the modified POI in the second labeling step. Some applications however require fast labeling of POIs in a single step. Here the modifying enzyme needs to accept small molecules that already carry the desired probe and transfer these to the POI. A well-established example of one-step enzyme mediated labeling is the phosphopantetheinyl-transferase (PPTase)-mediated labeling, using the PPTases Sfp or AcpS. Both of these enzymes transfer modified phosphopantetheine derivatives to a serine in octa- to undecapeptide recognition sequences. PPTase-mediated labeling is a good choice for labeling extracellular small tags with arbitrary substrates. However, the method is restricted to the labeling of cell surface proteins because intracellular coenzyme A (CoA) would interfere with the labeling and the CoA-derivatives used for labeling would interfere with other CoA dependent cellular processes. Another enzyme that has a high tolerance for various substrates is the lipoic acid ligase (LplA). Lipoic acid ligase is used to transfer functional labels containing reactive bromides, azides and photocrosslinkers to proteins tagged with a 13 to 22 amino acid target sequence. Engineered variants of this enzyme have been shown to accept fluorescent coumarin as a substrate, demonstrating the versatility of this approach. Nevertheless, each new substrate demands a new round of protein engineering to adapt the lipoic acid ligase to the novel functionality. Further examples of recently developed tag based labeling techniques include tub-tag labeling, π-clamp labeling, or the use of coiled-coil templated acyl transfer. The recent progress in the field demonstrates the high research interest in new labeling techniques that are orthogonal to existing methods and provide new tools for studying proteins and their function (Table 1).
Table 1: Enzymatic tag-based labeling approaches. Targeted residues are underlined.

<table>
<thead>
<tr>
<th>enzyme</th>
<th>tag</th>
<th>substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>formylglycine-generating enzyme</td>
<td>CXPXR</td>
<td>HN...X</td>
</tr>
<tr>
<td>sortase A (SrtA)</td>
<td>LPXTGXX</td>
<td></td>
</tr>
<tr>
<td>biotin ligase (BirA)</td>
<td>GLNDIFEAQKIEWHE</td>
<td></td>
</tr>
<tr>
<td>phosphopantetheinylltransferase (Sfp, AcpS)</td>
<td>D$\text{S}$LEFIASKLA</td>
<td></td>
</tr>
<tr>
<td>lipoic acid ligase (LplA)</td>
<td>GFEIDKVVYDLDA</td>
<td></td>
</tr>
</tbody>
</table>

a suitable for one-step labeling, b suitable for two-step labeling

1.5 Legionella enzymes AnkX and Lem3

The labeling strategies introduced in section 1.4.2 and 1.4.3 gain their specificity from using engineered enzymes whose natural function is the catalysis of post-translational modifications. Post-translational modifications are necessary to add functionalities to a protein after or during its translation from RNA to extend the limited repertoire of functional groups provided by the 20 natural amino acids. These new functionalities are often essential for the protein to fulfill its physiological function. The installed chemical groups can range from small chemical moieties like phosphates or acetyl groups to more complex molecules like sugars and large hydrophobic groups like lipids. The usage of such post-translational modifications in the development of new techniques for protein modification
requires the creative manipulation of known enzymes or the discovery of new suitable enzymes with novel properties and reactivity. Bacteria are often valuable sources for such new enzymes with interesting functions.

For example, the intracellular pathogen *Legionella pneumophila* possesses around 330 effector proteins that have evolved to specifically modify proteins from its host cell. After entering the host cell, *L. pneumophila* injects its effector proteins into the host cell cytosol and hijacks cellular communication and transportation pathways to form a replicative permissive niche, called the Legionella-containing vacuole (LCV).^{86–95}

One of the host cell proteins that is modified and repurposed throughout these events is the small guanosine triphosphatase (GTPase) Rab1b.^{96,97} Rab1b is addressed by seven different effector proteins and consequently recruited to the cytosolic face of the LCV where it triggers the accumulation of vesicles to supply the LCV with nutrients and membrane material.^{98} One of the translocated effectors that targets Rab1b was identified as the previously unknown phosphocholine transferase AnkX.^{99,100}

AnkX is an ankyrin repeat containing protein and belongs to the family of Fic (filamentation induced by cAMP) proteins. Fic-proteins usually introduce a variety of modifications, such as adenylylation, uridylylation, guanylylation, and phosphorylation to tyrosine or threonine residues of target proteins.^{101,102} In contrast to this modification pattern, AnkX uses cytidine 5'-diphosphocholine (CDP-choline) as a co-substrate to transfer a phosphocholine moiety to a specific serine or threonine residue in the switch II region of its substrate proteins Rab1b and Rab35 respectively. The switch II region of small GTPases can adopt two distinct conformations depending on which cofactor the protein is bound to (Figure 7).^{103} When bound to GTP, they are considered to be in their active state where the switch II region is folded into a defined and well-ordered tertiary structure.^{104} When bound to GDP, the switch II region is disordered and therefore has a highly flexible conformation. While most effector proteins targeting the switch II region of Rab1b have a clear preference for the active conformation,^{105} AnkX interestingly phosphocholinates both conformational states of Rab1b with similar catalytic efficiency, showing a 2.5 fold higher preference for the disordered inactive form.^{106} This preference for the unstructured peptide sequence of the switch II region suggests that AnkX recognises a relatively small amino-acid sequence and does not need a discrete ordered three-dimensional structure to bind its substrate proteins.

Another enzyme was found among the 330 effector proteins from *L. pneumophila* that displays the reversed enzymatic activity of AnkX. The dephosphocholinase is denoted Lem3 and it is active at a later time point of infection where it hydrolytically cleaves the phosphocholine modification from Rab1b and restores the unmodified wild type protein (Figure 7).^{106}
Figure 7: Phosphocholination of Rab1b. a) The effector protein AnkX from *Legionella pneumophila* transfers a phosphocholine moiety derived from CDP-choline to serine 76 in the switch II loop of the small GTPase Rab1b. This post-translational modification is hydrolytically cleaved by another *Legionella* effector protein Lem3. b) The crystal structure of the small GTPase Rab1b shows Rab1b bound to guanosine diphosphate (GDP) adopting the inactive conformation (left). The switch II loop is structurally disordered and not visible in the crystal structure (scattered line). The right structure displays Rab1b in its active conformation, bound to guanosine triphosphate. Here, the switch II loop adopts a well-defined helix structure.
2. Objectives

The overall objective of the work that forms the basis of this thesis is to develop and establish a new method for protein functionalisation. The development of such a new protein labeling strategy has to meet several challenges. An ideal method should be specific for the target protein and site-specific within the protein. Furthermore, it should make use of a small tag and a small labeling probe to minimise the impact on the structure and function of the investigated protein. As cellular processes often occur on a very short time scale, labeling reactions to visualise these processes have to be fast, quantitative and result in a stable covalent bond between protein and label. Preferable, a labeling technique allows the attachment of a wide range of functional modifications, including conditional cleavage of the modification.\textsuperscript{107,108}

Many of the established approaches do not satisfy all of the above-mentioned requirements. Some methodologies like the PPTase or lipoic acid mediated labeling need large tags to recognise the substrate proteins and transfer sterically cumbersome chemical groups to the POI. Other techniques like the FGE- or BirA-mediated labeling display a narrow substrate range and are dependent on a second labeling step. To overcome these limitations we aim to exploit the unique reactivities displayed by the \textit{Legionella} effector enzymes AnkX and Lem3 to establish a novel protein functionalisation strategy. The beneficial characteristics of these two enzymes that form the basis of the here presented work are:

- AnkX is specific for Rab1b and Rab35 and does not modify other proteins from the large Rab-family which contains more than 60 members.\textsuperscript{109,110}
- AnkX modifies Rab1b and Rab35 site-specifically on Ser76 and Thr76 in their switch II region.
- AnkX prefers the unstructured peptide sequence of the switch II region, turning the amino acid sequence of this region into a promising template for a short recognition tag.
- AnkX uses CDP-choline as a co-substrate to transfer the relatively small phosphocholine moiety (MW = 165 Da) to the target protein.
- The catalytic efficiency of \(1.0 \times 10^5\) M\(^{-1}\)s\(^{-1}\) of the AnkX catalysed phosphocholinlation of Rab1b shows that this reaction is potentially as fast as the tetra-cysteine tag and the SNAP-tag.\textsuperscript{107}
- The modification of target proteins with phosphocholine results in a stable covalent phosphodiester bond between phosphocholine and a serine or threonine residue.
• AnkX preferentially accepts CDP-choline as co-substrate but \textit{in vitro} experiments demonstrate that also CDP serves as a substrate, resulting in the modification of Rab1b with a phosphate group. This suggests that derivatisation of CDP or CDP-choline can yield functionalised CDP-derivatives that are accepted by AnkX and transferred to target proteins.

• Lem3 exhibits catalytic efficiencies of $2 \times 10^5 \text{M}^{-1}\text{s}^{-1}$ for the dephosphocholination of Rab1b, suggesting that Lem3 can be a useful tool to reverse the envisioned labeling and restore unmodified POI.

In order to capitalise on these enzymatic characteristics and develop them into a general applicable technique for site-specific protein functionalisation, several challenges have to be addressed. First of all, we need to develop chemically derivatised CDP-choline analogues that serve as co-substrates for AnkX and are transferred to its target proteins. Preferably, these CDP-choline analogues will allow the attachment of a wide range of functionalities to modify proteins in a single step. In addition, the design of CDP-choline derivatives that carry bioorthogonal functionalities can be applicable for modular and flexible two-step labeling approaches similar to those discussed in section 1.3.

Next we will have to identify the minimal peptide sequence from the switch II region of Rab1b and Rab35 that is necessary for the AnkX recognition. This sequence should be as short as possible to minimise unwanted interference with the structure and function of the tagged protein. Furthermore a short amino-acid tag will allow positioning of the tag at the C- or N-termini of POIs and even in flexible internal loop regions.

With respect to the enzymes AnkX and Lem3, we will investigate their substrate scope towards synthetic CDP-choline and phosphocholine derivatives and explore their viability for labeling and delabeling reactions of model proteins that are equipped with the suitable recognition sequence.

Furthermore, we aim at advancing the labeling technique from pure \textit{in vitro} applications towards more demanding experimental settings in the field of cell-surface- and live cell imaging. Protein functionalisation inside living cells requires the development of cell-permeable CDP-choline derivatives and AnkX variants that do not target endogenous proteins.

Finally, we want to develop functional CDP-choline derivatives that can be used as molecular tools to further investigate the mechanism of the AnkX catalysed phosphocholination of Rab1b and its role during bacterial infections with \textit{Legionella pneumophila}. Accordingly, suitable CDP-choline derivatives shall be synthesised and used to identify new target proteins for the enzymatic phosphocholination and enable screening for other phosphocholine transferase than AnkX.
3. Covalent protein labeling by reversible enzymatic phosphocholination (Paper I)

The site-specific modification of proteins with defined functionalities is often crucial for investigating protein function and protein-protein interactions *in vitro* and *in vivo*. Hence, there is an ever-growing demand for site-directed labeling strategies that allow the convenient modification of proteins with functionalities that are not defined by the genetic code. A popular approach to achieve this goal is the utilisation of enzymes that specifically recognise small amino acid sequences in a protein of interest and transfer an enzyme-specific small organic molecule to a defined nucleophilic residue within that sequence. Examples of such tag-based enzymatic labeling strategies are the PPTase or LpIA mediated labeling (see section 1.4.3).

To develop and establish an alternative strategy for the regioselective labeling and delabeling of proteins, we envisioned that the recently discovered unique catalytic phosphocholination and dephosphocholination activities of the enzymes AnkX and Lem3 from *Legionella pneumophila* could be exploited.99,100 AnkX uses the endogenous nucleotide CDP-choline and transfers a phosphocholine moiety to Ser76 in the switch II loop of the small GTPase Rab1b. Lem3 on the other hand cleaves this specific modification from Rab1b and restores the original unmodified protein (Figure 7).

**Figure 8:** Envisioned labeling strategy based on the enzymatic activities of AnkX and Lem3. AnkX covalently modifies a recognition tag from Rab1b with functionalised phosphocholine analogues derived from synthetically prepared CDP-choline derivatives. The tag will be placed at the termini or in loop regions of protein of interest (POI). Furthermore, Lem3 will be able to remove this modification and restore the unmodified POI.
We envisioned to derivatise AnkX’s co-substrate, CDP-choline, to create functional analogues that can carry any label of choice and still serve as substrates for the phosphocholine transfer. These derivatives could then be transferred to proteins of interest that are equipped with a recognition sequence derived from the switch II loop of Rab1b. It is intended that this small peptide tag could be placed both at the N- or C-terminus and in internal loop regions of POIs. In addition, we aimed at applying the dephosphocholinating activity of Lem3 to hydrolytically reverse the introduced modification from POIs; turning this approach into a covalent yet reversible strategy for protein labeling (Figure 8).

3.1 Design and synthesis of functional CDP-choline derivatives

The development of a protein functionalisation strategy based on AnkX catalysed phosphocholination strongly depends on the synthesis and applicability of functional CDP-choline derivatives. These derivatives have to be substrates for AnkX and also permit a label transfer to POIs along with the phosphocholine group. The design of such a functional CDP-choline analogue requires consideration of its binding mode to AnkX and the catalytic mechanism of the phosphocholine transfer to target proteins.

**Figure 9:** Crystal structure of AnkX’s CMP domain (green) and its Fic domain (orange) in complex with CDP-choline. The interaction of the CMP-domain with the cytosine base of CDP-choline positions the nucleotide in the active site. The choline head group is recognised by a combination of electrostatic interactions with Glu226 and Asp265 and a cation–π interaction with Phe107. This defined binding mode allows the attack from His229 on the beta phosphate of CDP-choline and the subsequent transfer of phosphocholine to target proteins.
As introduced in section 1.5, AnkX belongs to the Fic protein family. As typical for Fic proteins, AnkX possesses a characteristic and structurally conserved Fic domain. This Fic domain consists of three alpha helices of which two are connected with a loop that contains the conserved Fic motif $\text{H}x\text{Fx}[\text{D/E}][\text{G/A}]\text{N}[\text{G/K}]\text{R}$ (Figure 9). As discussed in 1.5 Fic-proteins usually catalyse nucleotidylylations, such as adenylylation, to modify target proteins. To induce these reactions, the Fic domain binds and positions ATP in a conserved way that allows the attack of a nucleophilic amino acid residue on the $\alpha$ phosphate and the subsequent transfer of AMP to the target protein.

In contrast to this conserved binding mode, a series of interactions between AnkX’s CMP-domain (residues 1-48 and 293-331) and the cytosine base of CDP-choline lead to a reversed binding mode of the nucleotide. This positioning is supported by a combination of electrostatic and cation-\(\pi\) interactions of the positively charged choline head group with negatively charged residues Glu226, Asp265 and aromatic Phe107. The distinct nucleotide binding leads to the attack of the catalytic His229 on the $\beta$ phosphate of CDP-choline and the subsequent transfer of the phosphocholine moiety to target proteins.

**Figure 10:** Design of a functional CDP-choline analogue. The surface presentation of AnkX in complex with CDP-choline displays the accessibility of the choline head group in the binding site of AnkX. The functionalised CDP-choline analogue is derivatised with a flexible PEG spacer (green) at the choline head group, allowing easy cargo loading via the installed terminal amine prior to AnkX mediated protein ligation. The designed CDP-choline analogue is synthetically accessible from a derivatised phosphocholine spacer and an activated CMP building block.
This structural analysis of the CDP-choline binding site suggests that modifications at the cytosine moiety would interfere with various interactions between AnkX’s CDP-domain and the nucleotide. This could lead to a slower reaction rate or even prevent the phosphocholine transfer. Therefore, we rationalised that modification of CDP-choline should be restricted to the phosphocholine motif. Fortunately, we had access to the crystal structure of AnkX in complex with CDP-choline to support our probe design. The surface presentation of AnkX in complex with CDP-choline shows that the quaternary ammonium head group is positioned in a deep but solvent accessible pocket, suggesting that a label could be attached here via a flexible hydrophilic linker of sufficient length (Figure 10). A terminal amino functionality could further serve as a convenient attachment point for a wide range of NHS-activated carboxy functionalised labels.

The designed CDP-choline construct can be retrosynthetically disconnected across the pyrophosphate moiety to an activated CMP building block and the corresponding phosphocholine derivative. The phosphocholine analogues are accessible from Boc-protected amino polyethylene glycol (PEG) spacer in three steps and the CMP can be activated with morpholine/dicyclohexylcarbodiimid.

Given the synthetic difficulties connected to preparative CDP-choline chemistry, especially the handling of polar intermediates and required nucleotide protection, we developed a new synthetic strategy that is based on the use of unprotected cytidine monophosphate (CMP) to access the functional CDP-choline derivatives. In order to find the proper conditions for the coupling of CMP to the phosphocholine analogues, Boc-protected phosphoryl-ethanolamine was used as a model substrate. Following a procedure from Agranoff et al. CMP 1 was first activated with morpholine and dicyclohexylcarbodiimid (DCC) yielding cytidine 5’-monophospho-morpholidate 2 as the 4-morpholine-N,N’-dicyclohexylcarboxamidine salt.111 The reaction between 2 and N-Boc-phosphorylethanolamine was particularly moisture sensitive and extensive experimentation revealed that combining these two reagents in benzene, followed by lyophilisation prior to the reaction was crucial for its success. Pyridine was used as solvent for the 1H-tetrazole induced coupling, which afforded 3 in 68% yield.112 Deprotection of the terminal amine under acidic conditions yielded CDP-ethanolamine 4 (Scheme 1).
Scheme 1: Synthesis of CDP-ethanolamine 4.

Reagents and conditions: (i) morpholine, DCC, t-BuOH/H$_2$O 1:1, 95°C, 5h; (ii) 1H-tetrazole, anhyd. pyridine, 60°C, o.n.; (iii) aq. HCl, RT, 2h.

To validate our main hypothesis, we needed to assess the usability of functionalised CDP-choline derivatives as substrates for AnkX and subsequent protein modification. To that end, AnkX’s natural substrate Rab1b (50 µM) was incubated with the respective nucleotide 3 or 4 (1 mM) and a catalytic amount of AnkX (0.1 µM) was added to initiate the reaction. After incubation over night the proteins were analysed by ESI-MS. The spectrometric data demonstrated that AnkX is able to recognise and transfer the CDP-ethanolamine 4, leading to a mass increase of Rab1b of 125 Da, which correlates with the calculated mass of phosphorylethanolamine incorporation. In contrast, the transfer of Boc-protected phosphoryl-ethanolamine analogue 3 could not be observed. These results suggest that AnkX accepts analogues that lack the quaternary amine, but that a protection of the amine prevents a transfer reaction. This could be due to depletion of the positive charge at the amine, which might be necessary for electrostatic interactions with AnkX, or due to steric clashes of the bulky tert-butyloxy carbonyl within the binding site.

To determine the distance required between CDP-choline and its bulky cargo such that the AnkX catalysed transfer of a modified phosphocholine is viable, we synthesised phosphocholine analogues with PEG spacers of varying length (Scheme 2).
Reagents and conditions: (i) iodine, imidazole, PPh₃, anhyd. THF, 0°C-RT, 15 min., then 5 or 6, 2h; (ii) N,N-dimethylaminoethanol, RT, o.n.; (iii) POCl₃, 2,6-lutidine, anhyd. MeCN, 0°C, 30 min. then H₂O, 15 min.; (iv) cytidine-5′-monophosphomorpholidate, 1H-tetrazole, anhyd. pyridine, 60°C, o.n., then TFA/DCM (1:1), 0°C-RT, 30 min.

The syntheses commenced with an iodination of the corresponding PEG-spacers 5 and 6 using the iodine-triphenylphosphine-imidazole reagent. The iodinated spacers 7 and 8 were used to alkylate N,N-dimethyl ethanolamine to form the quaternary amine spacers 9 and 10. For the following phosphorylation several procedures were evaluated. During initial attempts, spacer 9 was phosphorylated by the tetrazole-promoted phosphorimidate method and converted to the corresponding bis-benzyl phosphate esters that were purified on reverse phase C18 sep pak column. Debenzylation in the presence of diisopropylethylamine yielded phosphorylated spacer 11. Due to the instability of bis-benzyl phosphates in the presence of the nucleophilic tetrazole, this route lead to various products and low yields of isolated benzyl phosphate intermediates. As an alternative phosphorylation method, spacer 9 was treated with phosphorus oxychloride and 2,6-lutidine in trimethylphosphate. The in situ formed intermediate was directly hydrolyzed in aqueous triethylammonium bicarbonate buffer to give 11. This method increased the yield and simplified the synthesis, but the use of trimethylphosphate and triethylammonium bicarbonate buffer required time-consuming ion exchange purification. Replacing trimethylphosphate with dry acetonitrile and hydrolyzing the intermediate in aqueous solution of 2,6-lutidine gave phosphorylated spacer 11 and in good yields of 40-60% after sep pak purification. Finally, the diphosphate was formed via the tetrazole promoted coupling of phosphorylated spacer 11 or 12 to cytidine 5′-monophosphomorpholidate 4-morpholine-N,N′-dicyclohexylcarboxamidine salt. Deprotection of the terminal amine gave the CDP-choline analogues 13 and 14 (Scheme 2).
**Scheme 3:** Synthesis of CDP-analogues 23-25 with different functionalised PEG-spacer.

Reagents and conditions: (i) PPh₃, imidazole, I₂, THF, 0°C-RT, 16h; (ii) 17: N,N-dimethylaminoethanol, RT, o.n., 18: 2-(Methylamino)ethanol, RT, 16h; (iii) POClₓ, 2,6-lutidine, dry MeCN, 0°C, 30 min, then H₂O, 15 min.; (iv) cytidine-5’-monophosphomorpholidate, 1H-tetrazole, anhyd. pyridine, 60°C, o.n., then TFA, DCM, 0°C-RT, 30 min.

After synthesis of the CDP-choline derivatives 13 and 14, we explored if the cationic quaternary amine is necessary for the substrate recognition, or if this delicate position can be changed to an easier to handle amine- or ether linkage. To investigate the importance of the charged quaternary amine in the γ-position of the pyrophosphate, spacer analogues with variations at this position were synthesised (Scheme 3). To ensure that the chosen linker length is long enough to not affect the transfer reaction, the longer double PEG spacer 15 was used as starting material. After iodination of 15 with triphenylphosphine and imidazole, the iodinated spacer 16 was used to alkylate N,N-dimethyl ethanolamine to prepare the choline analogue 17 or 2-(N-methylamino)ethanol to prepare mono-methylated intermediate 18. After phosphorylation with phosphorous oxychloride in presence of 2,6-lutidine, the resulting spacers 20 and 21 were coupled to the activated cytidine 5’-monophosphomorpholidate. Final deprotection of the terminal amine yielded in a PEG₆ CDP-choline derivative 23 and a PEG₆ methylamine.
analogue 24. Additionally, the direct phosphorylation of spacer 19 followed by CDP-coupling and deprotection produced the corresponding PEG₇ ether analogue 25.

In summary, we established a robust synthetic route to access PEGₙ functionalised CDP-choline analogues that can be conveniently functionalised with any amine-reactive probe via a terminal primary amine function. The final analogues are accessible in a 5-step synthesis starting with readily available PEG derivatives.

3.2 Evaluating CDP-derivatives as AnkX substrates

Once the previously described synthesis strategy to access the designed CDP-choline analogues was established, the viability of these derivatives in the envisioned protein labeling strategy was evaluated. To analyse the transfer efficiencies of each phosphocholine analogue, a fluorescent-based read-out was established that enables the straightforward quantification of modified target proteins. AnkX’s natural substrate protein Rab₁b was chosen to serve as the acceptor protein and the CDP-choline analogues were equipped with a fluorescent label.

Demonstrating the operationally simple modification of the designed CDP-choline derivatives, the primary amines of 13 and 14, as well as 23-25, were reacted with NHS-activated fluorescein-5(6) carboxylate and purified over a short C₁₈ sep pak column to yield fluorescein analogues 26, 27 and 28-30 (Scheme 4 and Table 2).

**Scheme 4**: Representative synthesis of fluorescein derivatised CDP-choline derivatives.

**Reagents and conditions**: (i) NHS-5(6)carboxyfluorescein, sat. NaHCO₃ (aq.), pH 8, RT, 30 min.
For the labeling reactions, Rab1b (50 µM) was incubated with each of the derivatives (1 mM) in presence of catalytic amounts of AnkX (0.5 µM) for two hours at room temperature. Samples were taken at different time points and quenched by heat-induced denaturation in SDS buffer. The samples from different time points were analysed by in-gel fluorescence after SDS-PAGE and normalised against subsequent coomassie or silver stain. The catalytic efficiencies were calculated from the observed increase of fluorescence intensity as a function of time and listed below (Table 2).

Table 2 AnkX and Lm3 catalyzed reversible fluorescent labeling of Rab1b with synthetic CDP-choline analogues 26-30.1

<table>
<thead>
<tr>
<th>Entry</th>
<th>ID</th>
<th>R</th>
<th>AnkX $k_{cat}/K_m$ [M$^{-1}$s$^{-1}$]</th>
<th>Len3 $k_{cat}/K_m$ [M$^{-1}$s$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td></td>
<td>3.8 x 10$^3$</td>
<td>7.4 x 10$^1$</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td></td>
<td>5.0 x 10$^2$</td>
<td>3.9 x 10$^1$</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td></td>
<td>no reaction</td>
<td>n.a.</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td></td>
<td>0.9 x 10$^2$</td>
<td>6.6 x 10$^1$</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td></td>
<td>no reaction</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

1 catalytic efficiencies were recorded by Katharina Heller in the laboratory of Prof. Aymelt Itzen, AG Proteinchemie, Technical University Munich (TUM), Germany.
We detected fluorescent Rab1b after incubation with three of the five tested derivatives. These results demonstrate the general applicability of the synthetically prepared CDP-choline analogues in AnkX mediated labeling of substrate proteins like Rab1b. Comparison of the first three entries in table 2 allows the evaluation of different linker length and how this factor affects the performance of each analogue. Here, the medium sized phosphocholine analogue 26 is transferred with the highest catalytic efficiency while a longer spacer between fluorescein label and the quaternary ammonium functionality, as present in compound 28, slows down the transfer. A decrease in linker length to a diethylene glycol spacer (derivative 27) prevents transfer of the fluorescein. In this case the spacer is clearly too short to reach out of the nucleotide-binding site, leading to sterical clashes between the bulky fluorescein group and surface amino acids of AnkX’s active site. Comparison between derivatives 26-28 demonstrates that the AnkX mediated labeling of proteins with bulky cargo like fluorescein or other fluorophores is feasible when CDP-choline analogues with spacers of at least four PEG units are used. Furthermore, the recorded catalytic efficiency of 26 suggest that smaller CDP-choline analogues perform slightly better in the transfer reaction than larger constructs.

To further analyse the effect of the quaternary ammonium function on the nucleotide recognition and phosphocholine transfer, entries 1, 4 and 5 can be compared. While the amine derivatives 28 and 29 where both transferred to Rab1b, analogue 30, which has an ether linkage in place of the amine/ammonium function, was not accepted as a substrate for AnkX. These results are consistent with expectations derived from the crystal structure of AnkX in complex with CDP-choline (Figure 9, page 17), showing that the choline group forms salt bridges with Glu226 and Asp265, as well as a cation–π interaction with Phe107 from AnkX. Removing one methyl group, thereby eliminating the permanent cationic charge, leads to a four-fold drop in catalytic efficiency. Although analogue 29 will be partially protonated under physiological pH, the significant decrease in transfer efficiency demonstrates the importance of the choline head group for the synthetic CDP-choline analogues.

Next, we were curious if the installed fluorescent phosphocholine analogues can be removed from Rab1b by applying the phosphocholine hydrolase Lem3. To that end, quantitatively phosphocholinated Rab1b (50 µM) modified with analogues 26, 28 and 29 were incubated with catalytic amounts of Lem3 (2 µM) and the decrease of in-gel fluorescence was analysed after SDS-PAGE at different time points. The results shown in table 2 display that Lem3 is able to delabel all modified Rab1b substrates. Furthermore, it appears that Lem3 does not have a distinct preference for any phosphocholine derivative, which is reflected by similar catalytic efficiencies for all three analogues.
In summary, it was determined that appropriately functionalised CDP-choline derivatives can be used to fluorescently label the AnkX target protein Rab1b via catalytic phosphocholination. The described results of the labeling and delabeling experiments form a solid basis for the further development of a generally applicable protein labeling strategy.

3.3 Identification of a minimal peptide recognition motif for the AnkX catalysed phosphocholination

In order to turn proteins other than Rab1b and Rab35 into substrates for the labeling approach, these proteins have to be equipped with a peptide sequence that is recognised by AnkX. As introduced in section 1.5, AnkX naturally phosphocholinates Ser76 from the switch II loop of Rab1b and the corresponding Thr76 from Rab35. Consequently, this sequence was used as a template to identify the smallest possible sequence that is recognised by AnkX and serves as an acceptor for the phosphocholine transfer. As a starting point, a fragment of Rab1b's switch II region was synthesised by solid phase peptide synthesis (SPPS) and the resulting peptide ERFRTITSSYYRGG was investigated for phosphocholination by AnkX. After incubation of the peptide (0.5 mM) with CDP-choline (5 mM) in presence of AnkX (2 µM) for 4h the peptide sample was analysed by LC/ESI-MS (Figure 11).

**Figure 11:** Identification of the recognition sequence for the AnkX catalysed phosphocholination. Left: Mass spectra of tetradecapeptide from the switch II loop of Rab1b containing AnkX’s natural substrate amino acid Ser76 (a) before and (b) after overnight incubation with AnkX and CDP-choline. The mass increase of 165.11 m/z indicates a successful phosphocholination of the peptide. c) Schematic presentation of alanine scanning to identify necessary amino acids for AnkX recognition.
The initial peptide peak of 1692.35 m/z disappeared and a new peak of 1857.46 m/z was detected. The difference of 165.1 m/z corresponds to the expected gain in molecular weight caused by the addition of a phosphocholine moiety to the peptide. This result confirmed the initial assumption that AnkX recognises its substrates based on their primary amino-acid sequence and does not need the defined three-dimensional secondary or tertiary structure of a folded domain to successfully transfer the phosphocholine moiety. The detection of a single product with an increase of 165.1 m/z in mass further demonstrates that AnkX specifically modifies only one amino acid in presence of several other nucleophilc residues. The modified residue is likely the second serine in the sequence, which represents Ser76 from the switch II region.

With the intention of minimising interference with the function of tagged model proteins, the recognition tag size has to be minimised. Furthermore, a small size would allow the free positioning of the tag in loop regions as well as on the termini of POI. Therefore we investigated if the 14 amino-acid sequence could be truncated further without preventing the phosphocholine transfer. Alanine scanning was used to determine the contribution of each amino acid to the AnkX recognition of the previously synthesised fragment. In detail, 12 mutants of the native peptide were synthesesed by replacing each residue with alanine. The resulting peptides were tested as substrates for the phosphocholination under the conditions as described for the native peptide. The LC/ESI-MS based analysis revealed that a residue exchange to alanine was fully accepted up- and downstream of octapeptide TITSSYYR. Replacement of the first threonine, isoleucine and the terminal arginine within TITSSYYR leads to incomplete phosphocholination under the chosen reaction conditions, while changes within the core sequence TSSYY completely abolished phosphocholination. From this it could be concluded that the octapeptide sequence TITSSYYR is the minimum conserved recognition motif for efficient phosphocholination by AnkX.

As the long-term goal was to achieve a fully reversible labeling strategy, the identified peptide tag also has to be recognised by Lem3. In order to investigate if Lem3 is able to hydrolyse a phosphocholine moiety from short peptide sequences, we had to synthetically prepare defined phosphocholinated peptides from the switch II region by SPPS. There are two strategies to access peptides that are modified with post-translational modifications like the required phosphocholine. In the first strategy, the unmodified peptide is synthesesed on solid support by standard Fmoc synthesis and the modification of the desired residue is introduced after completion. Here the selectivity for a specific residue is achieved by the use of orthogonal protecting groups for the different nucleophilc residues. In a second approach, the desired modification is already present on the amino acid that is used in SPPS. Here a custom designed amino acid
building block has to be synthesised beforehand, which is than incorporated into the growing peptide chain during SPPS, yielding the desired peptide after global deprotection and cleavage from the resin.\(^{27} - ^{119}\)

We synthesised phosphocholinated serine and threonine building blocks suitable for Fmoc based SPPS in order to prepare several potential substrate peptides for Lem3.\(^{120}\) As it was discovered that the TITSSYYR is sufficient to serve as substrate for AnkX, the investigation started by preparing the octapeptide bearing a phosphocholinated serine in position five. This peptide (50 \(\mu\)M) was incubated with Lem3 (1 \(\mu\)M) and the reaction was monitored by LC/ESI-MS after over night reaction (Figure 12). The initial peak of 1155.4 m/z corresponding to the expected mass of the peptide TITSS(PC)YYR disappeared after over night incubation and a new peak appeared at 989.9 m/z. The detected mass decrease of 165.5 m/z matches the calculated size of the phosphocholine moiety and confirms that Lem3 recognises the octapeptide and is able to hydrolyse the installed PC-group. Interestingly, when the phosphocholinated amino acid is changed from serine to threonine to mimic the switch II sequence of Rab35, dephosphocholination by Lem3 was not detected. Also other phosphocholinated serine and threonine residues in the octapeptide motif are no substrate for Lem3. This observation indicates that Lem3 is very specific for phosphocholinated Ser76 of Rab1b.

![Figure 12: Identification of a sequence for the Lem3 catalysed dephosphocholination. Left: Mass spectra of octapeptide TITSSYYR from the switch II loop of Rab1b containing phosphocholinated Ser76 (a) before and (b) after incubation with Lem3. The mass decrease of 165.5 m/z indicates a successful dephosphocholination of the peptide. c) Schematic presentation of alanine scanning to identify necessary amino acids for Lem3 recognition.](image-url)
To further investigate the effect of amino acids adjacent to the octapeptide TITSSYYR on the Lem\textsubscript{3} efficiency, a similar alanine scanning approach as for the phosphocholination reaction was employed (Figure 12c). We synthesised 12 mutants of the native peptide with conserved phosphocholinated Ser\textsubscript{76} and sequentially replaced each residue with alanine. These peptides were then investigated as substrates for the Lem\textsubscript{3} catalysed dephosphocholination under the analogous conditions as described for the native peptide. Interestingly, all peptides were dephosphocholinated to some extent but the degree of demodification varied significantly. In particular, residues positioned N-terminal of the modified serine seemed to have a detrimental impact on Lem\textsubscript{3} activity.

In summary, we identified the short octapeptide TITSSYYR from the switch II region of Rab\textsubscript{1b} as the minimum recognition sequence for the efficient Ank\textsubscript{X} catalysed phosphocholination. Furthermore, it was demonstrated that the phosphocholinated variant of this small peptide is recognised and dephosphocholinated by Lem\textsubscript{3}. Intriguingly, the corresponding sequence TITSTYYR from Rab\textsubscript{35} also serves as a substrate for Ank\textsubscript{X} but not for Lem\textsubscript{3}. In section 4, these divergent substrate specificities of Ank\textsubscript{X} and Lem\textsubscript{3} for different peptide sequences will be exploited for a double labeling strategy.

### 3.4 Evaluation of model proteins for the Ank\textsubscript{X} catalysed phosphocholination

As the next step towards a generally applicable protein labeling strategy, we had to evaluate if the sequence TITSSYYR from 3.3 could be phosphocholinated by Ank\textsubscript{X} when fused to a protein of interest. Three model proteins were chosen and the potential recognition sequence was introduced at different positions within the proteins. In order to have a diverse set of model proteins, the small ubiquitin-related modifier (SUMO) was chosen as a representative of small proteins (14.4 kDa), the guanosine nucleotide exchange factor (GEF) domain of Drr\textsubscript{A} as a medium sized enzyme (21.6 kDa), and the maltose binding protein (MBP) as an example for larger proteins (45 kDa) (Figure 13, page 30). Furthermore, measuring the exchange rate of Drr\textsubscript{A} variants that are equipped with the recognition tag, would allow us to investigate the impact of the tag or its phosphocholination on enzyme activity.

The octapeptide sequence was recombinantly positioned at the N- and C-termini of SUMO and MBP, as well as in selected loop regions of Drr\textsubscript{A}. These model proteins were heterologously expressed in E. coli and submitted to the phosphocholination assay. Each of the heterologously expressed proteins (50 µM) was incubated with a catalytic amount of Ank\textsubscript{X} (0.5 µM) in the presence of CDP–choline (1 mM) for 8h. Samples were taken at different
time points to monitor the phosphocholination progress by LC/ESI-TOF MS and western blot analysis using a monoclonal anti-phosphocholine antibody.

![Image of model proteins](image)

**Figure 13:** Model proteins for the AnkX catalysed phosphocholination. The recognition sequence TITSSYYR was fused to the N- (blue) or C-terminus (red). a) Small ubiquitin-related modifier (SUMO) with highlighted termini. b) GEF domain of the guanine exchange factor DrrA<sub>340-533</sub>. Additionally to N- and C-terminal fusions, DrrA variants were produced where the recognition sequence was placed in four different loop regions that are indicated by colored spheres. c) Maltose binding protein (MBP) with highlighted termini.

The qualitative mass spectrometric analysis of phosphocholination revealed that both SUMO proteins, the N-terminal MBP and all DrrA variants were successfully phosphocholinated after 8 hours. The C-terminal TITSSYYR-tagged MBP showed trace amount of the phosphocholinated product. These initial results demonstrate that the identified tag can generally be recognised by AnkX as a protein fusion and can be phosphocholinated if placed in an accessible position of the protein of interest. Here the quantitative phosphocholination of all DrrA variants, where the TITSSYYR recognition tag is inserted in different loop regions has to be highlighted. The inclusion of recognition tags into loop regions of proteins significantly expands the applicability of tag based protein functionalisation techniques as most of the available tags can only be used at terminal positions of protein of interest. Furthermore, it is noteworthy that DrrA variants carrying the recognition tag at different positions were still active and also phosphocholination of the tag did not lead to loss of activity. This demonstrates that the AnkX labeling does not interfere with protein function, when the tag-position is evaluated appropriately.

In order to exclude the possibility that AnkX unspecifically phosphocholinates one or more of the native amino acids in the model proteins, SUMO was fused to the alternative C-terminal recognition

---

<sup>2</sup> Cloning, expression, and phosphocholination experiments of the model proteins were performed by Dr. Katharina Heller and Florian Zauner in the laboratory of Prof. Aymelt Itzen, AG Proteinchemie, Technical University Munich (TUM), Germany.
sequence TITSSAYR, lacking the serine that is targeted by AnkX. Using the phosphocholination assay described above, we could show that this potential substrate failed to undergo the AnkX-mediated phosphocholine transfer. This experiment once more demonstrates the residue specificity of the AnkX catalysed phosphocholination and the suitability of this reaction for site-specific protein labeling.

Quantitative densiometric analysis of the western-blot data was used to calculate the apparent catalytic efficiencies ($k_{cat}/K_M$) for each substrate protein in order to compare their performance as substrates for AnkX. Not surprisingly, the artificial substrates perform less well than the native substrate protein Rab1b, which is phosphocholinated with a $k_{cat}/K_M$ of $6.7 \times 10^3$ M$^{-1}$s$^{-1}$. Among the tested model proteins, the N-terminal tagged SUMO appeared to be the best substrate with a 55 fold lower efficiency of $1.2 \times 10^2$ M$^{-1}$s$^{-1}$. The other substrate proteins all performed equally, with $k_{cat}/K_M$ values of around 10 M$^{-1}$s$^{-1}$. Although the evaluated model proteins are phosphocholinated slower than Rab1b, the use of TITSSYYR as a protein tag still allows for efficient phosphocholination by AnkX on a reasonable timescale.

### 3.5 Fluorescent labeling of model proteins

Building upon the described exploratory experiments, it had to be determined whether AnkX catalysed phosphocholination is applicable as a general labeling technique. To verify this, we examined if the synthetically prepared functionalised CDP-choline analogues from 3.3 could be used to site-specifically label the evaluated model proteins from 3.5. Recapitulating, we determined that the CDP-choline analogues 26, 28 and 29 were accepted by AnkX and could be used to fluorescently label Rab1b via the phosphocholination reaction. Furthermore, the N-terminal tagged SUMO was identified as the best substrate protein for the AnkX catalysed phosphocholine transfer. We chose TITSSYYR-SUMO as a model substrate to investigate if the CDP-choline derivatives 26, 28 and 29 can be used to site-selectively functionalise SUMO with a fluorescein group.

In a series of experiments, TITSSYYR-SUMO (50 µM) was incubated with the different CDP-choline analogues (1 mM) in the presence of catalytic amounts of AnkX (0.5 µM) and the reaction progress was analysed with in-gel fluorescence and normalised against subsequent coomassie or silver stain. The time-dependent increase in fluorescence intensity was used to calculate $k_{cat}/K_M$ values for each CDP-choline derivative (Table 3, page 32).

As observed in the labeling of Rab1b, all CDP-choline analogues tested (26, 28 and 29) resulted in fluorescently labeled SUMO proteins. Furthermore, the substrate profile for this reaction was identical to that of Rab1b. This confirms that the AnkX-recognition sequence TITSSYYR is
sufficient to turn proteins of interest into substrates for phosphocholine transfer.

**Table 3:** AnkX and Lem3 catalysed fluorescent labeling and delabeling of the model protein TITSSYYR-SUMO with synthetic CDP-choline analogues 26, 28 and 29.

<table>
<thead>
<tr>
<th>Entry</th>
<th>ID</th>
<th>R</th>
<th>AnkX $k_{cat}/K_M$ [M⁻¹s⁻¹]</th>
<th>Lem3 $k_{cat}/K_M$ [M⁻¹s⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td><img src="image1" alt="Chemical Structure" /></td>
<td>0.4×10²</td>
<td>5.6×10³</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td><img src="image2" alt="Chemical Structure" /></td>
<td>1.4×10²</td>
<td>6×10³</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td><img src="image3" alt="Chemical Structure" /></td>
<td>0.1×10²</td>
<td>16×10³</td>
</tr>
</tbody>
</table>

A more detailed comparison of the different CDP-choline analogues showed that derivative 26 is the most efficient substrate for TITSSYYR-SUMO. The PEG₄-analogue is transferred with a $k_{cat}/K_M$ of 1.4×10² M⁻¹s⁻¹ and modified approximately 70% of TITSSYYR-SUMO in 180 min. This reaction rate is faster than the sortase mediated labeling (50 M⁻¹s⁻¹) and just one order of magnitude slower than the Clip-tag (10³ M⁻¹s⁻¹). In contrast, the use of analogues 28 and 29 resulted in only trace incorporation of fluorescein-derivatised phosphocholine to TITSSYYR-SUMO. The slower reaction rates of the PEG₆ analogue 28 and the PEG₆ amino-methyl analogue 29 suggest

---

3 catalytic efficiencies were recorded by Dr. Katharina Heller in the laboratory of Prof. Aymelt Itzen, AG Proteinchemie, Technical University Munich (TUM), Germany.
that the longer spacer or variations at the choline head group reduce the affinity for AnkX and impairs the transfer of the derivatised phosphocholine group. This is consistent with the observations from experiments in 3.3. Cumulatively, these results demonstrate that the AnkX catalysed phosphocholination can be exploited to site-specifically label proteins of interest with synthetically prepared CDP-choline analogues equipped with unnatural functionalities, like the fluorophore fluorescein.

In order to further expand the applicability of this protein labeling strategy and turn this approach into a covalent and reversible strategy for protein labeling, we asked whether Lem3 could be utilised to detach the incorporated functional phosphocholine derivatives from the modified target proteins. To determine the feasibility of this, we subjected fluorescently phosphocholinated TITSSYR-SUMO (50 µM) to Lem3-mediated (2 µM) enzymatic dephosphocholination and analysed the protein samples by in-gel fluorescence after SDS-PAGE. The time dependent decrease in fluorescence, resulting in the complete loss of fluorescence, revealed that TITSSYR-SUMO could be effectively dephosphocholinated by Lem3. As expected, the calculated $k_{cat}/K_M$ values for the delabeling reactions further suggest that the internal loop region of Rab1b is the most efficient substrate for Lem3. Nevertheless, the different labeled SUMO proteins also react on comparable timescales. Strikingly, Lem3 seems to be less demanding for the particular phosphocholine derivative than for the tag sequence.
4. Exploring the substrate scope for the AnkX catalysed phosphocholination (Paper II)

In recent years many strategies for site-specific protein functionalisation have been developed that utilise highly specific enzymes and their corresponding enzyme tags (see section 1.4.3). The specificity of these exploited enzymes for their respective recognition tags is often accompanied with a rather narrow tolerance for suitably derivatised co-substrates. This limitation reduces both the applicability and versatility of these approaches where protein labeling with a wide range of different functional probes is required. One strategy to expand the scope of individual enzymes to accept more variable substrates is to genetically engineer their active sites to accept the desired functional probe. However, this directed evolution approach often goes in hand with a drop in enzymatic efficiency and/or results in a number of enzyme mutants that are each specialised for only one specific substrate. In contrast to such a labour intensive strategy it would be beneficial to have access to a labeling strategy that uses only one enzyme with a broad substrate scope, allowing the modification of proteins with any desired functional probe.

In section 3 of this thesis, a new site-specific strategy for the covalent functionalisation of tagged proteins was introduced. It was demonstrated that the bacterial enzymes AnkX and Lem3 could be exploited to fluorescently label and delabel substrate proteins that are equipped with a short recognition tag. In this approach, synthetically prepared CDP-choline analogues, functionalised with fluorescein via a flexible PEG-spacer, were used as co-substrates for AnkX. Ideally, the fluorescein group could be replaced with any conceivable functionality without hampering the transfer to target proteins. This procedure would ultimately allow convenient labeling of target proteins with an unlimited variety of functional groups. To explore the substrate scope of AnkX and the generality of the outlined approach, we synthesised a range of different CDP-choline analogues and investigated their applicability for the labeling of substrate proteins.

To cover a wide range of functionalities, CDP-choline analogues were prepared that are equipped with different fluorophores and affinity handles. As demonstrated in section 3.3 and 3.6, the probe design from 3.2 should be suitable for such CDP-choline analogues to allow convenient one-step labeling of substrate proteins.

Furthermore, we envisioned expanding the toolbox for AnkX-mediated labeling by incorporating small bioorthogonal handles into proteins of interest using appropriately functionalised CDP-choline analogues. These
unique handles could subsequently be labeled with complementary functionalised probes in highly specific bioorthogonal ligation reactions. To that end, CDP-choline analogues were designed that are compatible with commonly used bioorthogonal reactions as Cu(I)-catalysed azide–alkyne cycloadditions (CuAAC), strain-promoted azide–alkyne cycloadditions (SPAAC), and inverse-electron-demand Diels–Alder reactions between strained alkenes and tetrazines. In combination with the small recognition tag that could be inserted anywhere in the polypeptide chain of proteins, this approach would allow a flexible and modular two-step labeling approach and significantly expand the versatility of the AnkX catalysed protein functionalisation.

4.1 Synthesis of CDP-choline derivatives

In order to explore AnkX’s substrate scope for derivatised CDP-choline, we started with the preparation of analogues functionalised with different fluorophores. This should demonstrate the applicability of AnkX mediated phosphocholination for convenient one-step fluorescent labeling. As the highest functionalisation rates in 3.3 and 3.6 were achieved when the fluorescent probe was attached via a PEG₄ spacer to the quaternary amine of CDP-choline, this design was chosen to prepare p-aminobenzyl-4-chloro-7-nitrobenzofurazan (aminobenzyl-NBD) and coumarin functionalised CDP-choline derivatives (31 respectively 32) (Scheme 5). The aminobenzyl-NBD is a commonly used dark quencher and can be used in combination with fluorescent coumarin to evaluate distances in proteins via Förster resonance energy transfer (FRET) measurements.¹²¹⁻¹²⁴

**Scheme 5:** Synthesis of aminobenzyl-NBD 31 and coumarin 32 functionalised CDP-choline derivatives.

**Reagents and conditions:** (i) NBD-4-aminophenylacetic acid-NHS ester or 7-(dimethylamino)coumarin-4-acetic acid-NHS ester, sat. NaHCO₃ (aq.), pH 8, RT, 30 min.
The synthesis of compounds 31-32 followed a similar route as described previously in section 3.2. The terminal amine of compound 13 was treated with the NHS-esters of NBD-4-aminophenylacetic acid and 7-(dimethylamino)coumarin-4-acetic acid in aqueous bicarbonate buffer to give the labeled CDP-choline analogues 31 and 32, respectively.

In addition to the amino-terminated CDP-choline derivative 13, a CDP-choline analogue with a terminal alkyne 33 was prepared that allows the functionalisation with azide derivatives in a CuCAACA (Scheme 6). We planned that this alkyne analogue could be used for bioorthogonal ligation reactions after incorporation into proteins or as a precursor for variable functionalisation with azide bearing probes. CDP-choline 33 was prepared by derivatisation of 13 with NHS-activated 4-pentynoic acid in aqueous bicarbonate buffer at pH 8. The terminal alkyne of 33 was then functionalised via a click reaction with 5-carboxytetramethylrhodamine (5-TAMRA) PEG₃-azide in aqueous copper sulfate solution in presence of sodium ascorbate to prepare 5-TAMRA labeled CDP-choline analogue 34.

**Scheme 6:** Synthesis of alkyne- and TAMRA-functionalised CDP-Choline 33 and 34.

Reagents and conditions: (i) 4-pentynoic acid NHS ester, sat. NaHCO₃ (aq.), pH 8, RT, 30 min.; (ii) 5-carboxytetramethylrhodamine PEG₃ azide, CuSO₄, ascorbate, H₂O, RT, 8h.
Another commonly used moiety for the functionalisation of proteins is the affinity handle biotin. Biotin binds strongly to the tetrameric protein streptavidin (STV) and can be used to selectively enrich labelled proteins from complex mixtures in pull-down experiments. The synthesis of the biotin functionalised CDP-choline was achieved in a similar fashion to the previously described procedures by derivatisation of 13. Instead of activated NHS-ester, a biotin-4-nitrophenyl ester was used to furnish CDP-analogue 35 (Scheme 7).

**Scheme 7**: Synthesis of biotinylated CDP-choline analogue 35.

![Scheme 7](image)

**Reagents and conditions**: (i) biotin-4-nitrophenyl ester, H$_2$O/MeCN, RT, 3h.

Results presented in 3.3 and 3.6 suggested that the optimal co-substrates for the AnkX catalysed phosphocholination are small and structurally close to the native CDP-choline. To further improve the CDP-choline co-substrates and to consequently increase their transfer rate in the AnkX catalysed labeling reaction, functional CDP-choline analogues were designed with as minimal modification as possible. Here, the choline head group was once derivatised with an alkyne 39 and also with an azide moiety 43, only connected via a small aliphatic carbon chain (Scheme 8). These groups were chosen as they represent the smallest available functionalities that can be addressed in bioorthogonal labeling reactions after incorporation into proteins of interest. For the synthesis of the alkyne bearing CDP-choline analogue, we chose to start the reaction sequence with N,N-dimethyl ethanolamine 36. In the first step the alkyne moiety was introduced by alkylation of 36 with propargyl bromide in THF to furnish N-propargylcholine 37 as bromide salt. Subsequently, 37 was phosphorylated with POCl$_3$ in dry acetonitrile as described in 3.2. The resulting phosphocholine analogue 38 was submitted to the tetrazole promoted pyrophosphate coupling with cytidine 5'-monophosphomorpholidate to yield propargyl CDP-choline 39.
Scheme 8: Synthesis of propargyl CDP-choline 39.

Reagents and conditions: (i) propargyl bromide, anhyd. THF, 0°C - RT, o.n.; (ii) POCl₃, 2,6-lutidine, dry MeCN, 0°C, 30 min. then H₂O, 15 min.; (iii) cytidine-5'-monophosphoromorpholidate, 1H-tetrazole, anhyd. pyridine, 60°C, o.n.

To access the azido derivatised CDP-choline analogue 43, 2-dimethylamino-ethanol 36 was alkylated with an excess of 1,3-dibromopropane to give the choline derivative 40 as bromide salt. The remaining bromine of 40 was substituted with sodium azide to install the desired azide functionality of compound 41. Following the procedures described earlier, 41 was phosphorylated to provide intermediate 42 and subsequently coupled with activated cytidine monophosphate to form the azide functionalised CDP-choline derivative 43 (Scheme 9).

Scheme 9: Synthesis of azide derivatised CDP-choline analogue 43.

Reagents and conditions: (i) 1,3 dibromopropane, neat, RT, o.n.; (ii) NaN₃, DMF, 70°C, o.n.; (iii) POCl₃, 2,6-lutidine, dry MeCN, 0°C, 30 min. then H₂O, 15 min.; (iv) cytidine-5'-monophosphomorpholidate, 1H-tetrazole, anhyd. pyridine, 60°C, o.n.
As discussed previously in 1.3, the use of alkyne and azide groups in bioconjugation reactions is superior due to their small size and reactivity towards each other. However, their application in bioorthogonal chemistry is limited due to the dependence on cell-toxic Cu(I) for their ligation in CuCAACAs. While the azide derivative 43 could also be addressed with strained alkynes and alkenes, conjugation of the propargyl derivative 39 is typically dependent on Cu(I). In order to circumvent these conditions, CDP-choline analogues with advantageous bioorthogonal handles were designed.

We decided to prepare CDP-choline analogues functionalised with dibenzocyclooctynes (DBCO), trans-cyclooctene (TCO) and tetrazine handles, as these functionalities can be chemoselectively addressed with their complementary reaction partners in cycloadditions exhibiting reasonably fast reaction kinetics. The DBCO analogue serves as an alternative to the alkyne derivative and can be addressed with azide containing probes in strain promoted cycloadditions after incorporation into proteins. Furthermore, we desired to prepare TCO and tetrazine derivatives, as these functionalities react with superior reaction rates in an inverse-electron-demand Diels-Alder cycloaddition, thus also allowing the labeling and capturing of fast processes occurring inside living cells. This combination of functionalities will give access to a versatile toolbox to address the most common classes of bioorthogonal chemistries.

As observed in 3.3, AnkX does not transfer bulky cargo like the fluorescein group to its target proteins when the functionality is connected with a short diethylene glycol spacer. Therefore, as some of the envisioned bioorthogonal functionalities are rather bulky, the optimal linker length was probed for each bioorthogonal handle. We synthesised different variants of the bioorthogonal CDP-choline analogues with C3-spacer as well as tetraethylene glycol spacer to identify the optimal linker length that allows the fastest transfer of each bioorthogonal handle (Scheme 10 and 11).
Scheme 10: Synthesis of C3-linked bioorthogonal CDP-choline analogues.

Reagents and conditions: (i) Pd/C, H₂, MeOH/H₂O, RT, 4h; (ii) DBCO-, TCO- or tetrazine NHS ester, sat. NaHCO₃ (aq.), pH 8, RT, 30 min.

For the synthesis of bioorthogonal CDP-choline analogues, which contain the short C₃ linker, a synthetic route analogous to that used for the azido derivative 43 was employed. The azide group of 43 represents a masked amine that was liberated by hydrogenation to yield compound 44. Subsequently, the terminal amine of 44 was treated with the corresponding NHS-activated bioorthogonal groups to access derivatives 45-47.

To access CDP-choline derivatives with PEG₄ spacers more appropriate for sterically more demanding functional groups like DBCO or tetrazine, the synthetic strategy utilised in section 3.2 was adapted. Instead of starting with Boc-protected amino-spacer 5, the synthesis commenced with a bifunctional PEG₄ spacer 48 terminated with a bromine and a carboxybenzyl (Cbz) protected amine. The brominated spacer 48 was first treated with neat 2-dimethylaminoethanol to yield the choline like spacer 49. Spacer 49 was phosphorylated and the resulting phosphorylated spacer 50 was used to furnish CDP-choline analogue 51 via a tetrazole promoted coupling with activated cytidine 5′-monophosphoromorpholidate. The terminal amine functionality of analogue 51 was liberated from the Cbz protecting group by catalytic hydrogenation and the final derivatives 53-55 were furnished by coupling 52 to the appropriate NHS-ester under basic aqueous reaction conditions (Scheme 11).
**Scheme 11** Synthesis of PEG4-linked bioorthogonal CDP-choline analogues.

Reagents and conditions: (i) N,N-dimethylaminoethanol, neat, 70°C, o.n.; (ii) POCl₃, 2,6-lutidine, anhyd. MeCN, 0°C, 30 min. then H₂O, 15 min; (iii) cytidine-5'-monophosphomorpholidate, 1H-tetrazole, anhyd. pyridine, 60°C, o.n.; (iv) Pd/C, H₂, MeOH/H₂O, RT, 8h; (v) DBCO-, TCO- or tetrazine NHS ester, sat. NaHCO₃ (aq.), pH 8, RT, 30 min.

The straightforward synthesis of the here presented 12 different CDP-choline analogues, each accessed in three to six steps; demonstrate the high flexibility of the established synthesis strategies. Avoiding chromatographic purification of the highly polar intermediates simplifies the handling and allows reasonable overall yields.

4.2 Versatile functionalisation of Rab₁b by AnkX phosphocholination

After successful synthesis of various CDP-choline analogues displaying a wide range of useful functionalities, the incorporation of these derivatives into AnkX’s natural substrate protein Rab₁b was investigated. Rab₁b (50 µM) was incubated with the corresponding CDP-choline derivative (200 µM) and catalytic amount of AnkX (1 µM) was added to start the reaction. The transfer of the corresponding phosphocholine analogues was followed by LC/ESI-TOF MS. The mass spectrometric analysis of the protein
samples revealed that all derivatives, independent of functionality or linker length, were accepted by AnkX and lead to fully modified Rab1b after overnight incubation. While these results were expected for the PEG₄ functionalised CDP-choline analogues, it was surprising that also bulky cargo like the DBCO was transferred to Rab1b when only connected with the short C₃ spacer.

To gain insight into the preferences of AnkX for the different derivatives and to quantify the transfer rate of each compound, the labeling reactions were sampled after different time-points and the reaction progress was analysed by in-gel fluorescence or western blot after SDS-PAGE and normalisation against subsequent coomassie stain. The incorporation of bioorthogonal analogues was visualised by labeling the functionalised protein with compatible Cy5-dyes and the incorporation of biotin was detected with a horseradish peroxidase streptavidin (HRP-STV) conjugate. The time-coursed data was then used to calculate the apparent catalytic efficiencies ($k_{cat}/K_M$) for each derivative (Figure 14 and 15).

Figure 14: Representative presentation of time coursed Rab1b labeling with CDP-choline analogues 39, 45, 46 and 55. To visualise the degree of modification, modified proteins were incubated with accordingly functionalised Cy5 dyes at each timepoint (39: azide-Cy5, CuSO₄, sodium ascorbate; 45: azide-Cy5, 46: tetrazine-Cy5, 55: TCO-Cy5) and the fluorescence intensity was plotted against time. Normalisation against coomassie stain determined quantitative labeling at the endpoint of the reaction.

---

4 catalytic efficiencies were recorded in cooperation with M.Sc. Stefan Ernst in Umeå and the laboratory of Prof. Aymelt Itzen, AG Proteinchemie, Technical University Munich (TUM), Germany.
Figure 15 summarises the catalytic efficiencies of all investigated CDP-choline analogues. Derivative 26, which gave the fastest reaction kinetics in the previous labeling experiments (see section 3.3), was included as a reference. Interestingly all of the tested derivatives performed as well or significantly better than compound 26. As expected, the incorporation efficiencies of the PEG₄ linked fluorescent analogues 31 (2.5 × 10⁻³ M⁻¹s⁻¹), 32 (6.9 × 10⁻² M⁻¹s⁻¹) and 34 (7.6 × 10⁻² M⁻¹s⁻¹) are all on a similar timescale comparable to the transfer rate of 26.

Also the biotin analogue 35 is transferred to Rab1b with reasonable fast kinetics (10⁻³ M⁻¹s⁻¹) that are comparable to efficacies of the CLIP-tag. This proves that the AnkX mediated one-step functionalisation with a range of fluorophores and other small molecules like biotin is feasible on a practical time-scale. However, it has to be noted that the here tested acceptor protein Rab1b is AnkX's native substrate and that the functionalisation of other substrate proteins probably will exhibit slower transfer rates.

![Figure 15: Transfer rates of CDP-choline analogues to Rab1b.](image-url)
Next, we compared the incorporation rates of the CDP-choline derivatives that were designed for the bioorthogonal two-step protein labeling. As expected the highest catalytic efficiency of AnkX was observed in combination with propargyl CDP-choline 39, which bears the smallest modification among the tested substrates. The propargyl CDP-choline 39 was transferred to Rab1b with a high rate \(1.8 \times 10^4 \text{ M}^{-1} \text{s}^{-1}\), which is one order of magnitude faster than the previously tested analogues for the one-step labeling and are in the range of the SNAP- and Flash-tag \(3 \times 10^4 \text{ M}^{-1} \text{s}^{-1}, 1 \times 10^4 \text{ M}^{-1} \text{s}^{-1}\), respectively. The azide analogue 43 performed slightly worse \(1.2 \times 10^4 \text{ M}^{-1} \text{s}^{-1}\), but is still in the same range as the before mentioned functionalisation techniques. Considering the straightforward synthesis of 39 and 43, these two CDP-choline analogues represent a good choice to install bioorthogonal functionalities into protein of interest.

Surprisingly, the DBCO-C3 analogue 45 was incorporated with a higher transfer rate \(8.4 \times 10^3 \text{ M}^{-1} \text{s}^{-1}\) than the DBCO-PEG 4 \(7.4 \times 10^2 \text{ M}^{-1} \text{s}^{-1}\) and the remaining trans-cyclooctenes 46, 54 \(1.0 \times 10^3 \text{ M}^{-1} \text{s}^{-1}, 1.2 \times 10^3 \text{ M}^{-1} \text{s}^{-1}\) respectively) and tetrazines 47, 55 \(1.1 \times 10^3 \text{ M}^{-1} \text{s}^{-1}, 9.9 \times 10^2 \text{ M}^{-1} \text{s}^{-1}\) respectively). This was an unexpected result, as the incorporation of DBCO into proteins is usually difficult due to its bulkiness and lipophilicity. When the transfer rates of the remaining TCO and tetrazine functionalised analogues were compared, it shows that AnkX does not show a clear preference for any functionality nor for PEG or C3 connected derivatives. All of the investigated derivatives were transferred equally well to Rab1b with catalytic efficiencies of around \(10^3 \text{ M}^{-1} \text{s}^{-1}\).

In conclusion, AnkX demonstrates a high promiscuity towards a wide range of functional CDP-choline. The ability of AnkX to accept a variety of substrates and to transfer them to substrate proteins at practical rates makes this approach superior to any of the existing methodologies for in vitro protein labeling.

### 4.3 Dual labeling of proteins using AnkX and Lem3

So far we demonstrated that the catalytic activity of AnkX could be exploited for site-specific protein labeling with a broad array of functionalities. In combination with a short peptide tag, this enzyme-mediated functionalisation strategy should be suitable for the observation and study of protein expression, protein localisation and trafficking without perturbations of the native protein function. However, protein modification on a single site in the peptide chain also has its limitations. For example, monitoring conformational changes within a protein or studying protein-protein interactions requires dual site-specific labeling that permits the attachment of two different fluorescent molecules onto two distinct sites in the same POL. When the fluorescent molecules are positioned in close proximity and
possess overlapping emission (donor) and excitation (acceptor) spectra, conformational changes or interactions with other proteins can be studied by measuring the FRET between the two incorporated dyes.\textsuperscript{130,131} One way to achieve the required site-specific dual labeling of POIs is using the genetic code expansion strategy to incorporate bioorthogonal handles as described in section 1.3.\textsuperscript{132} Although this technique is superior as it only affects two residues in the POI, and the installed UAAs serve as excellent site-specific handles for protein labeling, the direct incorporation of two different UAAs into a single POI remains a challenging task well beyond of current state-of-the-art techniques. This is due to the lower expression levels of proteins with dual-site incorporated UAAs in comparison to single-site incorporated proteins.\textsuperscript{133}

Another more readily accessible method to introduce two distinct functionalities into POI is the use of two orthogonal peptide-tags for enzyme-mediated labeling.\textsuperscript{134} Here two unique peptide sequences, each directing an enzyme-mediated labeling reaction, can be site-specifically fused into target proteins. Due to the high specificity and low cross-activity among different enzyme-substrate pairs, this method is a straightforward approach for dual site-specific protein labeling.\textsuperscript{135} However a major drawback of this strategy remains that the positioning of many tags is restricted to the N- or C-termini of POIs\textsuperscript{136} and only a few examples make use of strategies suitable for internal protein labeling.\textsuperscript{137,138}

To remedy this, and to extend the AnkX mediated labeling approach, we have developed a strategy that allows the dual labeling of proteins with the two enzymes AnkX and Lem3. This strategy is proposed to make use of two slightly different peptide tags that can be installed at two different positions in a protein of interest. So far the AnkX mediated labeling strategy made use of the octapeptide tag TITSSYYR that is derived from the switch II region of the small GTPase Rab1b. However, as introduced in 1.5, AnkX also phosphocholinates the switch II regions of the two structurally related GTPases Rab1a and Rab35. While the amino acid sequence of Rab1b and Rab1a are identical, Rab35 possesses a threonine instead of a serine at the modified position. Interestingly, AnkX modifies Ser76 from Rab1b, Ser79 from Rab1a and Thr76 from Rab35, while Lem3 is only able to cleave the installed modification from serine residues, as was already elucidated in 3.3. This residue-specific reactivity displayed by Lem3 allows differentiation of the two sequences TITSSYYR (S-Tag) and TTSTYYR (T-Tag), enabling the double labeling of POIs (Figure 16, page 46). Positioning the S-Tag and the T-Tag at distinct sides of a protein of interest would allow for functionalisation of both positions in a first labeling step with one of the introduced CDP-choline derivatives from section 4.2. Subsequent application of Lem3 will lead to a site-specific demodification of the S-Tag while the modification of the T-Tag persists. A second labeling step with a
different CDP-choline analogue would finally yield in a POI that is functionalised with two distinct moieties on two defined positions.

**Figure 16:** a) Schematic representation of site-specific dual labeling using AnkX and Lem3 in combination with two distinct recognition tags and CDP-choline analogues 26 and 34. b) Fluorescent scan displaying the progress of dual labeling. Green signals correspond to fluorescein fluorescence while red signals belong to TAMRA fluorescence.5

As an initial model substrate for a proof-of-principle experiment, the maltose binding protein was equipped with a N-terminal T-tag and fused to Rab1b with its native internal recognition sequence TITSSYYR (S-Tag). Additionally, a peptide cleavage site for the tobacco etch virus endopeptidase (TEV-site) was introduced in between the two proteins to allow the separate analysis of each protein modification by in-gel fluorescence after enzymatic cleavage (Figure 15). In a first labeling step MBP-Rab1b (40 µM) was incubated with the CDP-choline derivative 26 (5 mM) and catalytic amount of AnkX (4 µM). After incubation with Lem3 (0.5 µM) MBP-Rab1b was purified by affinity chromatography using amylose resin. For a second labeling the purified MBP-Rab1b (5 µM) was incubated with CDP-choline derivative 34 and AnkX (0.25 µM). Samples were taken after each step and digested with TEV. Subsequently the labeling was analysed by In-Gel fluorescence after SDS-PAGE (Figure 16).

---

5 In-Gel fluorescence data was recorded by M.Sc. Stefan Ernst in the laboratory of Prof. Aymelt Itzen, AG Proteinchemie, Technical University Munich (TUM), Germany.
The fluorescent scan in figure 16b documents the progress of the dual labeling. After the first labeling, fluorescent signals that correspond to a functionalisation with fluorescein carrying phosphocholine analogue are detected before and after TEV cleavage of the fusion protein. After delabeling with Lem3 the fluorescein signal of the uncleaved reaction decreases. This decrease is due to total loss of the fluoresceine signal in Rab1b as can be seen in the TEV cleaved reaction. Moreover, the fluorescent signal for MBP in the same lane verifies that the Lem3 activity is site-specific and the fluorescein modification of the T-tag persists. After the second labeling with TAMRA functionalised CDP-choline analogue 34, the uncleaved reaction displays signals in both the fluorescein and TAMRA channel for the MBP-Rab fusion protein, which results in an orange merged signal. These signals indicate a dual labeling of the fusion protein with both fluorophores. Furthermore, the TEV cleaved sample displays the expected signals for a successful dual labeling. The MBP protein signals show that it is exclusively modified with fluorescein while the Rab1b only shows signals for TAMRA. The unexpected Rab1b band in the uncleaved reaction can be explained with TEV contamination and little cleavage of the fusion protein. Thus, this proof-of-principle experiment demonstrates that the alternate use of AnkX and Lem3 in combination with the two different recognition tags (S-tag and T-Tag) can be exploited to site-selectively label proteins of interest with two distinct functionalities.
5. Advancing AnkX-mediated phosphocholination - ongoing work

In the previous chapters it was demonstrated that AnkX/Lem3 controlled phosphocholination could be exploited to establish a general protein labeling strategy based on the reversible transfer of functionalised phosphocholine. This approach proved to be applicable for in vitro labeling of purified substrate proteins (section 3.5) with a wide range of valuable functional probes (section 4.2). However, understanding protein function and its impact on molecular processes ultimately requires the investigation of proteins in their native environment. Hence, the final goal is to combine AnkX, Lem3 and the derivatised nucleotides for protein functionalisation and manipulation inside living cells. With this intention, the application of the AnkX mediated phosphocholination needed to be evaluated in more complex biological settings, like on cell-surfaces or in cell lysates. Furthermore, successful protein functionalisation inside living cells required the development of cell-permeable CDP-choline derivatives and engineering of AnkX variants that do not target endogenous proteins.

5.1 Towards cell surface protein labeling

The application of AnkX mediated phosphocholination for the labeling of cell surface proteins would significantly expand the versatility of the described labeling strategy. The imaging of membrane receptors and monitoring their localisation and interaction with other proteins is of particular interest. The labeling of cell surface receptors however is more challenging than protein labeling in solution. As cell surface receptors are embedded in the lipid bilayer of the cell membrane they are usually only limited accessible at extracellular loop regions and termini. Furthermore, the cell surface is crowded with various proteins that are often modified with complex carbohydrates, reducing the accessibility of distinct receptors even further.

To evaluate membrane protein functionalisation via the AnkX catalysed phosphocholination we chose the human neuropeptide Y2 receptor (hY2R) as an initial model receptor. HY2R is a G-protein coupled receptor with seven trans membrane helices displaying extracellular N- and intracellular C-termi. Consequently, the recognition sequence TTTSSYR was fused to the extracellular N-terminus in order to turn hY2R into a potential substrate for AnkX that is accessible for phosphocholination from the outside of a cell (Figure 17).
Figure 17: Labeling of cell surface human neuropeptide Y₂ receptors (hY₂R) using AnkX catalysed phosphocholinlation. a) Schematic outline of the receptor labeling. hY₂R (blue) is embedded in the cell membrane and C-terminal fused to the fluorescent protein mCherry (red) on the intracellular side of the membrane. The N-terminus of hY₂R is fused to the octapeptide recognition tag TITSSYYR enabling the AnkX catalysed labeling with fluorescein functionalised CDP-choline 26. b) HEK293T cells, expressing TITSSYYR-HA-hY₂R-mCherry, treated with CDP-choline 26 (100 µM) in presence (i) and absence (ii) of AnkX for 3 hours. Fluorescence microscopic images were acquired in differential interference contrast channel (DIC, grey), in the mCherry channel (red) and in the fluorescein channel (green). The merge shows all three channels superimposed and indicates co-localisation when yellow.⁶

⁶ Cell labeling experiments were performed by Dr. Katharina Heller in the laboratory of Prof. Aymelt Itzen, AG Proteinchemie, Technical University Munich (TUM), Germany.
The red fluorescent protein m-Cherry was fused to the C-terminus of the receptor to follow its expression by fluorescence microscopy. HEK293T cells were transfected with the plasmid containing the TITSSYYR-HA-hY2R-mCherry construct and cultivated in DMEM and 10% FBS for 24h to allow expression of the receptor. For initial labeling, the cells were incubated with AnkX (2 µM) and fluorescein functionalised CDP-choline analogue 26 (100 µM) for 3 hours, washed and analysed by fluorescence microscopy. Figure 17b i shows the results of the labeling reaction exhibiting a strong fluorescence signal in the fluorescein channel (green) that co-localises with the signal of mCherry (red). The merge of these two channels (yellow) shows co-localisation of fluorescein and mCherry signals, suggesting successful labeling of the receptor-fusion. However, the control reaction (figure 17b ii) exhibits similar signals. Here the transfected cells were incubated with CDP-choline 26 without AnkX. Although the overlay indicates a lower degree of co-localisation as the labeling experiment, the high fluorescein signals suggests an unspecific interaction of the labeling probe 26 with the receptor and/or the cell surface. Therefore, when considering the labeling experiment and the control experiment, it cannot be concluded that the hY2R was selectively labeled with CDP-choline 26.

Consequently, the labeling of cell-surface receptors using the AnkX catalysed phosphocholination requires further optimisation. In future experiments the labeling will be investigated with other CDP-choline derivatives with the aim of reducing unspecific background labeling of the cell surface. To that end the functionalisation of the receptor with TCO-functionalised CDP-choline analogues 46 or 54 and subsequent labeling with fluorogenic tetrazine dyes could significantly improve the signal to noise ratio. Nevertheless, the experiments demonstrated that the addition of the AnkX recognition sequence into a protein expressed in mammalian cell culture is possible and does not lead to cell-toxic effects. Furthermore, the exposure of HEK cells to AnkX and CDP-choline derivatives did not affect the cell viability under the examined conditions.

### 5.2 Covalent trapping of phosphocholine transfer

A general prerequisite for the application of the AnkX mediated phosphocholination inside living cells is the efficient expression of tagged target proteins and, most importantly, of AnkX in mammalian cell cultures without affecting natural cellular processes that could potentially cause cell toxicity. Furthermore, the prospective use of AnkX inside living cells requires a high specificity of AnkX for its tagged target protein without modifying any other endogenous proteins. Experiments transiently expressing AnkX in mammalian cell cultures indicated that AnkX interferes with the membrane trafficking from the endoplasmic reticulum (ER) to the Golgi apparatus,
consequently leading to Golgi fragmentation and cell death.\textsuperscript{141} This disruption of membrane traffic can be attributed to the phosphocholination of AnkX’s native target proteins Rab1b and Rab35. Thus, protein engineering is required for future applications to create AnkX variants that do not modify endogenous proteins but only recognise and phosphocholinate proteins tagged with a suitable recognition sequence. Such a rational mutation of AnkX requires detailed structural information of AnkX’s active site and its interaction with the switch II loop of Rab1b and Rab35 or the peptide sequences TITSSYYR and TITSTYYR respectively. To date, no structural information is available that displays any of these interactions. Moreover the existing crystal structures of AnkX were obtained from truncated proteins lacking 4 of its 10 ankeryn repeats that presumably participate in the protein-protein recognition of substrates.\textsuperscript{142} Attempts to co-crystallise AnkX with Rab1b have failed so far. With the intention of stabilising the complex between AnkX and target proteins, thereby increasing chances of crystallisation, we combined electrophilic CDP-choline derivatives and single cysteine mutants of AnkX to covalently trap the phosphocholine transfer into a ternary covalent complex (Figure 18d, page 52).

The latest proposed mechanism for the AnkX catalysed phosphocholine transfer suggests a ping-pong mechanism in which the transferred phosphocholine is intermediately bound to the catalytic His\textsubscript{229} of AnkX’s Fic domain and subsequently transferred to the substrate protein.\textsuperscript{143} To trap the transfer we intended to exploit the proximity-induced reactivity between an introduced cysteine in the active side of AnkX and a weak electrophile that is connected to the choline head group of CDP-choline. We chose a chloroacetamide group as the electrophile due to its attenuated reactivity and created five different cysteine mutants of AnkX (V105C, F107C, P110C, R113C, N262C). The position of each cysteine was chosen due to their expected close proximity to the functionalised choline head group, allowing a nucleophilic attack on the alpha carbon of the chloroacetamide to establish a covalent bond between AnkX and phosphocholine. Subsequently Rab1b can bind to the binary complex and attack the phosphocholine with Ser76, resulting in a covalent ternary complex between AnkX, a phosphocholine linker and Rab1b. To elucidate the optimal distances between the quaternary amine of CDP-choline, the chloroacetamide and each of the introduced cysteines, CDP-choline analogues with C3 and C4 linkers were prepared. The C3-chloroacetamide functionalised CDP-choline 56 was accessible via a similar strategy as described earlier for the preparation of bioorthogonal functionalised CDP choline analogues 45-47 in section 4.1. In the final step intermediate 44 was treated with chloroacetic acid NHS-ester to give chloroacetamide 56. To prepare the C4 analogue 57 the synthesis strategy was adjusted by choosing 1,4-dibromobutane instead of 1,3-dibromopropane in the initial alkylation step.
Figure 18: Concept for covalent trapping the phosphocholine transfer from AnkX to Rab1b. 
a) Proposed mechanism for the phosphocholine transfer to Rab1b. His229 attacks the β-phosphate of CDP-choline releasing CMP. When AnkX binds Rab1b, Ser76 can attack the activated phosphoramide resulting in phosphocholination of Rab1b. 
b) Crystal structure of AnkX in complex with CDP-choline. Five amino acids are shown that are in close proximity to the choline head group, suitable for cysteine point mutations. 
c) Design of CDP-choline derivatives bearing electrophilic chloroacetamides (blue). 
d) A cysteine mutant of AnkX binding the chloroacetamide functionalised CDP-choline. The attack of His229 on the β-phosphate of CDP-choline and the attack by the introduced cysteine on the alpha carbon of the chloroacetamide establishes two covalent bonds between AnkX and phosphocholine. The attack of Ser76 from Rab1b would lead to a covalent complex between AnkX and Rab1b.
Next, CDP-choline analogue 56 was used to evaluate if the AnkX cysteine mutants accept the derivatised nucleotide and form a covalent bond via the chloroacetamide. We treated AnkX_F107C (50 µM) with nucleotide 56 (500 µM) and analysed the reaction by LC/ESI-MS after overnight incubation (Figure 19a,b). The mass spectra display three major peaks at 92725.94 m/z (unmodified protein), 92991.66 m/z (+264.72 m/z) and at 93297.93 m/z (+571.99 m/z). The later peak corresponds to the expected mass of AnkX_F107C covalently bound to 56 after nucleophilic attack on the chloroacetamide (Figure 19c), whilst the smaller mass increase observed for the second peak can be explained by hydrolysis of the covalently bound nucleotide along the pyrophosphate moiety (Figure 19d). Therefore, this initial experiment verifies that the introduction of a nucleophilic cysteine residue at an appropriate position in the active site of AnkX allows the trapping of a chloroacetamide derivatised CDP-choline derivative 56. Furthermore, it was observed that the bound nucleotide is hydrolysed over time when no acceptor protein for phosphocholine transfer is present.

Figure 19: Mass spectrometric analysis of AnkX_F107C before (a) and after (b) incubation with chloroacetamide CDP-choline analogue 56. a) One peak is displayed corresponding to the expected mass of AnkX_F107C (calc.: 92726 Da) b) Next to the mass peak corresponding to the unmodified protein, two more peaks are detected with increased mass of +265.72 m/z and +571.99 m/z. c) Cysteine modification with 56 corresponds to a mass increase of +572.15 m/z d) Cysteine modification with a phosphocholine fragment from 56 corresponds to a mass increase of 267.11 m/z.
Considering this information, we conceived two alternate approaches to evaluate the applicability of the five different AnkX mutants and the electrophilic chloroacetamide CDP-choline to form a covalent complex with Rab1b. In the first approach, each of the AnkX cysteine mutants was incubated with CDP-choline and Rab1b allowing the one-step formation of the desired ternary complex (Figure 20a). However, as the mass spectrometric results suggest the partial hydrolysis of CDP-choline, this approach could lead to low yields of the desired complex. Therefore, in the second approach Rab1b was first phosphocholinated with CDP-choline, catalysed by AnkX_wt, purified and subsequently incubated with each of the cysteine mutants (Figure 20b). This approach is unaffected by possible CDP-choline hydrolysis and can still lead to a ternary complex when AnkX binds phosphocholinated Rab1b (Figure 20).

Initial experiments to evaluate these two different approaches under several conditions were performed and analysed by Western Blot after SDS-PAGE. Rab1b containing protein bands were detected using an anti-Rab antibody. Figure 20c shows a Western blot after incubation of AnkX_F107C and AnkX_N262C with CDP-choline and Rab1b applying method A. Lanes 4-7 show Rab1b containing protein bands that could correspond to the desired complex between AnkX and Rab1b. Figure 20d displays the results of experiments with AnkX mutants V105C, F107C, P110C, R113C and N262C using Method B. Here, all tested AnkX mutants form a complex with Rab1b. However, especially at elevated temperatures, additional bands that correspond to higher aggregates were detected. These initial results generally show that the chosen AnkX mutants are suitable to form covalent complexes with Rab1b in combination with chloroacetamide CDP-choline. In future experiments, the exact composition of these complexes have to be determined, and conditions for the complex formation must be optimised.
Figure 20: Schematic representation of covalent complex formation between AnkX and Rab1b. 

a) Method A: AnkX cysteine mutant and Rab1b are incubated with CDP-choline derivative 56 to form a ternary complex. 
b) Method B: Rab1b is functionalised with CDP-choline derivative 56 catalysed by AnkX_wt. The modified Rab1b is then incubated with AnkX cysteine mutants to form the ternary complex. 
c) Western blot against Rab1b after incubation of indicated AnkX mutants (25 µM) with CDP-choline 56 (500 µM) and the following Rab1b concentrations (approach A). 1: 50 µM Rab1b, 25°C; 2: 50 µM Rab1b, 37°C; 3: 50 µM Rab1b, MgCl₂, 25°C; 4: 300 µM Rab1b, MgCl₂, 37°C; 5: 50 µM Rab1b, 25°C; 6: 50 µM Rab1b, MgCl₂; 7: 300 µM Rab1b, MgCl₂; 
d) Western blot against Rab1b after incubation of indicated AnkX mutants (10 µM) with modified Rab1b (50 µM) at different temperatures (approach b). i: 25°C, ii: 37°C.

7 Experiments were performed by M.Sc. Stefan Ernst in the laboratory of Prof. Aymelt Itzen, AG Proteinchemie, Technical University Munich (TUM), Germany.
5.3 Development of cell-permeable CDP-choline analogues

The application of AnkX catalysed phosphocholine as a protein functionalisation strategy inside living cells strongly depends on the delivery of synthetic CDP-choline derivatives into the cytosol of investigated cells.

The most straightforward approach to deliver labeling probes into the cytosol of cultured cells is by passive diffusion over the cell membrane. However, nucleotides like ATP, GTP and also CDP-choline are not cell-permeable due to their high polarity and negatively charged phosphates that lead to repulsion with phospholipids in the cell membrane. Alternatively, these impermeable labeling probes can be delivered into the cytosol by microinjection or electroporation of the membrane. Such approaches are prone to membrane disruption and might interfere with signaling pathways or other membrane related cellular processes.

**Figure 21:** Design of cell permeable nucleotide analogues. a) Structures of impermeable PEG containing ATP analogue 58 in comparison with cell permeable analogue 59. b) Intracellular kinase catalysed biotinylation by phosphorylation using cell permeable ATP analogue 59. c) Design of cell-permeable polyamine CDP-choline analogue 68 functionalised with a terminal alkyne allowing two step protein labeling via 'click' reactions.
To avoid such harsh conditions, Fouda et al. used an alternative approach to turn an impermeable ATP analogue 58 into a cell-permeable probe 59 that was used for intracellular kinase catalysed biotinylation (Figure 21a,b). Instead of using a hydrophilic PEG spacer to connect a desired label to the nucleotide, this approach uses a polyamine linker that is positively charged under physiological conditions and is thought to partially neutralise the triphosphate charge, consequently enhancing cell permeability. As the impermeable CDP-choline analogue 26 exhibits structural similarities to the ATP analogue 58 we chose to apply this method and designed a polyamine containing CDP-choline analogue 68 (Figure 21c).

**Scheme 12:** Synthesis of polyamine linked CDP-choline 68.

Reagents and conditions: (i) ethyl trifluoroacetate, MeCN, -78°C-RT, 3h, then Boc,O, Et₃N, 0°C-RT, o.n.; (ii) NH₄OH (aq), MeOH, 60°C, 4h, then evaporate, Cbz-Cl, Et₃N, 0°C-RT, o.n.; (iii) NaBH₄CN, formaldehyde (37%), acetic acid, EtOH, 0°C-RT, o.n.; (iv) iodo-ethanol, Et₂O, RT, 24h; (v) TFA/DCM (1:1, v/v), 0°C-RT, 30 min, then evaporate, NaBH₃CN, formaldehyde (37%), acetic acid, EtOH, 0°C-RT, o.n. then 5 M NaOH (aq.); (vi) POCl₃, 2,6-lutidine, anhyd. MeCN, 0°C, 30 min. then H₂O, 15 min.; (vii) Pd(OH)₂/C, H₂, MeOH, acetic acid, RT, 1.5h, then 4-pentynoic acid NHS ester, sat. NaHCO₃ (aq.), pH 8, RT, 40 min; (viii) cytidine-5'-monophosphomorpholidate, 1H-tetrazole, anhyd. pyridine, 60°C, o.n.
To prepare the proposed polyamine functionalised CDP-choline analogue 68, spermine was chosen as a precursor that mimics the original size of the PEG linker in 26 (Scheme 12). The synthesis commenced by selectively masking the terminal primary amines as trifluoroacetamides followed by protection of internal secondary amines with Boc groups to yield fully protected 61,145,146. The terminal amines were liberated by cleavage of the trifluoroacetamides with aqueous NH₄OH in methanol and treatment with one equivalent of Cbz-chloride, furnishing mono protected 62. The remaining unprotected amine was di-methylated by reductive amination with formaldehyde yielding 63 followed by alkylation with iodoethanol, resulting in the choline-like polyamine 64. Deprotection under acidic conditions of the internal Boc-amines and subsequent methylation by reductive amination yielded in polyamine 65. Next, 65 was phosphorylated following the established procedures as described earlier to give the phosphocholine derivative 66. To allow the convenient functionalisation of the final CDP-choline analogue with azide dyes via 'click' chemistry, an alkyne was introduced after deprotection of the terminal amine by catalytic hydrogenation. Finally, 67 was coupled with cytidine 5'-monophosphomorpholidate to afford the desired polyamine functionalised CDP-choline analogue 68 (Scheme 12).

Next, we investigated the cell permeability of CDP-choline 68 by fluorescence microscopy. For visualisation, 68 was labeled via a click reaction with 5-TAMRA-PEG₃-azide to afford 69 in 20% yield. 69 was added to FlpIn TRex HeLa cells stably expressing GFP-fused SNX-9, a membrane-associated protein that enables the visualisation and localisation of the cell membrane.147,148 Cells were incubated with 69 (50 µM) for 20 minutes at 37°C and 5% CO₂. As a control experiment, SNX-9 FlpIn cells were treated with the supposed impermeable TAMRA-PEG₄ CDP-choline 35 (50 µM) under analogous conditions. To our surprise the fluorescence microscopy images show intracellular TAMRA fluorescence in both experiments, with even higher intensities observed in the control experiment (Figure 22c). This observation suggests that both CDP-choline derivatives 69 and 34 are cell permeable. The unexpected permeability of CDP-choline 34 could be explained by the nature of the attached fluorescent dye TAMRA. Under physiological conditions TAMRA is cationic and acquires a net positive charge. In combination with the positive charged choline head group this could be sufficient to neutralise the negative charged diphosphate of CDP-choline. The acquired fluorescent images show that the TAMRA fluorescence is not evenly distributed over the cytosol but is concentrated in several circular areas. This local could indicate endosomal uptake of the polyamine functionalised CDP-choline derivative. Furthermore, incubation with 69 lead to higher TAMRA background all over the cell chamber resulting from unspecific accumulation of 69 on the glass surface, as can be seen in the
Figure 22: Cell-permeability investigation of CDP-choline analogues. a) Structures of 5-TAMRA functionalised CDP-choline analogues 35 and 69. 69 is equipped with a polyamine spacer (blue) while 35 bears a PEG spacer (red) to connect CDP-choline and dye. b) Flp-In TRex HeLa cells treated with CDP-choline 69 for 20 minutes. Fluorescent images were acquired in the TAMRA channel (left) to visualise compound 69 and in the GFP channel (center) to visualise the cell membrane. The merged image (right) shows the membrane in green and localisation of 69 in red. The cross-section view along the yellow line suggests that 69 is located inside the cell and accumulated on the bottom of the cell chamber indicated by the sharp red line. c) Flp-In TRex HeLa cells treated with CDP-choline 34 for 20 minutes. Fluorescent images were acquired in the TAMRA channel (left) to visualise compound 34 and in the GFP channel (center) to visualise the cell membrane. The merged image (right) shows the membrane in green and localisation of 34 in red. The cross-section view along the yellow lines shows that CDP-choline 34 is located inside the cell with accumulation in the nucleus.

8 Cell pictures were recorded by Dr. Madlen Hubert in the laboratory of Richard Lundmark, Department of Integrative Medical Biology, Umeå University Sweden.
cross-section images. To eliminate this unspecific background and exclude other effects of the TAMRA dye on the permeability of CDP-choline analogues, we would need to investigate if unlabeled derivatives 33 and 68 are able to enter cells. A suitable experiment would be the treatment of cells transiently expressing AnkX with unlabeled compounds 68 and 33. Subsequent cell lysis and labeling of the cellular proteins with azide-functionalised dyes would allow visualisation of AnkX substrate proteins like Rab1b and Rab35 if the probes were cell permeable and functional within the cells.
6. Identification of cellular protein targets of the AnkX catalysed phosphocholination (Paper III)

To enable the use of AnkX in combination with synthetic CDP-choline analogues for protein labeling inside living cells, the phosphocholination reaction has to be specific for the tagged proteins of interest without modifying endogenous proteins. To date, the only known cellular target proteins of the AnkX catalysed phosphocholination that might interfere with labeling are the small GTPases Rab1b and Rab35 from the Rab1 subfamily. As addressed in section 5.2, AnkX variants have to be created that do not recognise Rab proteins as substrates to avoid their modification in live cell imaging experiments. Furthermore, investigations are required to identify other possible intracellular target proteins that could be substrates for the phosphocholination and might cause background labeling or even interfere with the normal cellular function when AnkX is introduced to certain cell lines. Isolating and identifying possible target proteins for the AnkX catalysed phosphocholination will also further elucidate the role and function of this intriguing and unique posttranslational modification during host-pathogen interactions.

6.1 Isolation and identification of AnkX substrate proteins from cell lysate

In chapters 3-5 synthetically prepared CDP-choline derivatives were used to functionalise model proteins with fluorescent dyes in one-step and two-step labeling experiments. These applications proved that AnkX exhibits a rather broad substrate tolerance for CDP-choline analogues as well as for different substrate proteins (SUMO, DrrA and MBP) when these proteins are fused to the recognition sequence from Rab1b. As a short recognition tag seems to be sufficient for AnkX to recognise its target proteins, other proteins with similar peptide sequences might also serve as target proteins for the posttranslational phosphocholination. To identify possible substrate proteins for the AnkX catalysed phosphocholination, we envisioned to label cell lysate with biotin functionalised CDP-choline analogue 35 that is able to biotinylate Rab1b as demonstrated in chapter 4.2. The biotinylation via phosphocholination of substrate proteins would both allow their visualisation with suitable streptavidin (STV) conjugates after western blot and also facilitate their isolation from the cell lysate for further analysis via proteomic techniques (Figures 23,24).
To screen a full set of soluble cellular proteins as relevant substrates for the AnkX catalysed phosphocholination, we chose HeLa cell lysate and incubated the cell lysate (500 µg total protein in 250 µL reaction volume) with CDP-choline 35 (250 µM) in the presence of AnkX (1 µM). Samples were taken at different time points and biotinylated proteins were detected after SDS/PAGE and western blot with a fluorescent STV conjugate. As a negative control experiment, cell lysate was incubated with CDP-choline 35 without adding AnkX to exclude endogenous biotinylation. Furthermore, a positive control experiment was performed with Rab1b spiked cell lysate as an internal control for phosphocholination activity (Figure 23).

The fluorescent read out of the western blots shown in Figure 23b displayed three prominent bands corresponding to biotinylated proteins at 26, around 55-70 and 95 kDa. The detection of these distinct bands could suggest selective labeling of only few substrate proteins of the AnkX catalysed phosphocholination. Prolonged incubation times led to an increase of less defined signals, indicating that unspecific labeling increases over time. The positive control experiment, which contains additional Rab1b, exhibited higher density for the lower band at around 26 kDa than detected in the labeling experiment. This indicates that these bands result from phosphocholinated Rab proteins. Furthermore, the high intensity protein band at 95 kDa likely corresponds to AnkX caused by autophosphocholination. While these two signals were expected, the third distinct signal at 55 kDa might represent a yet unknown target for the AnkX catalysed phosphocholination. Thus to further analyse and identify the detected proteins, we repeated the labeling experiment and enriched biotinylated proteins from the cell lysate with STV-coated magnetic beads (Figure 24). The isolated proteins were separated by SDS-PAGE, and the previously identified protein bands were cut from the gel. The proteins were extracted from the gel pieces, digested into tryptic peptide fragments and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) to identify the peptide sequences.
Figure 23: AnkX catalysed labeling of cell lysate with biotin functionalised CDP-choline 35.

a) Workflow of western blot experiment. Cultivated cells are lysed and soluble proteins are separated from insoluble cell components. The lysate is incubated with AnkX and CDP-choline 35 to modify substrate proteins of AnkX with a biotin handle. The proteins are separated by SDS-PAGE and subsequently transferred to a nitrocellulose membrane. Finally, the biotinylated proteins are visualised and detected with a fluorescein-streptavidin (FL-STV) conjugate that selectively binds to biotin.

b) Fluorescein read out of the western blots showing the labeling of cell lysate (left), a negative control lacking AnkX (center) and a positive control containing Rab1b (right) after 6, 12 and 24h of incubation. The most prominent protein bands are detected at 95, 55 and 26 kDa in both the labeling experiment and the positive control. Strong signals in the positive control at 26 kDa suggest that the corresponding bands in the labeling experiment result from labeled Rab proteins. The signals at 95 kDa are likely to result from AnkX auto labeling. Low signal intensities in the negative control furthermore suggest that the labeling is specific.
Figure 24: Enrichment and identification of biotinylated proteins. a) Workflow of protein identification after lysate biotinylation. Modified proteins are enriched with streptavidin (STV) coated magnetic beads and separated by SDS-PAGE. Proteins are isolated from the gel and digested. Liquid chromatography tandem mass spectrometric analysis of isolated peptides allows identification of corresponding proteins. b) List of most probable identified proteins after statistical analysis. c) Western blot of isolated proteins from HeLa cell lysate separately detected with streptavidin (STV) and anti IMPDH2 antibody to verify the enrichment of IMPDH2.

The identification of peptide sequences and their comparison with the human protein database from Uniprot and the Legionella pneumophila database from SwissProt allowed the assignment of potential protein targets. To verify specific hits all peptides found in the negative control experiment were removed and the remaining proteins were ranked by the number of identified peptides and unique peptides. Here, the value for unique peptides is more significant as it gives the number of peptides that are specific for a distinct protein. As listed in figure 24b, 71 unique peptides of AnkX were

---

9 LC-MS/MS was performed by Dr. Petra Janning, Department of Chemical Biology, Max Planck Institute of Molecular Physiology Dortmund, Germany.
identified from the protein band at 95 kDa, which is equivalent to 75.8% of AnkX's sequence. Furthermore, 52 unique peptides from AnkX were also identified from the sample taken at around 55 kDa. These results confirm the autophosphocholinlation of AnkX and suggest proteolytic fragmentation of AnkX in the cell lysate, resulting in at least two fragments that run separately on a SDS-PAGE. The analysis of the protein band at around 26 kDa confirmed the expected modification of Rab proteins. Several peptides were identified that originate from different Rab protein family members, including Rab2a, Rab1a, Rab1b and Rab35. Of particular interest, 26 unique peptides belonging to inosine-5'-monophosphate dehydrogenase type II (IMPDH2) were identified in the sample taken at around 55 kDa. This corresponds to 69.5 % sequence coverage of the 55 kDa protein, strongly suggesting relevance of this identification. To further validate the identification of this previously unknown substrate protein of the AnkX catalysed phosphocholinlation, the pull-down was repeated and the isolated proteins were detected separately with a STV conjugate and an anti IMPDH2 antibody. As shown in figure 24c, the strong signal of the anti-IMPDH2 antibody verifies that IMPDH2 is isolated from HeLa cell lysate after biotinylation via AnkX catalysed phosphocholinlation.

This enrichment of known (Rab proteins and AnkX) and yet unknown protein substrates for the AnkX catalysed phosphocholinlation demonstrates the value of synthetic CDP-choline analogues like for the investigation of this rather unexplored posttranslational modification. Keeping in mind that Fic-proteins like AnkX are widespread across all kingdoms of life, this probe could be a valuable molecular tool to screen for phosphocholinlation activity in many organisms and may facilitate the identification of phosphocholinlated proteins.

6.2 Investigation of IMPDH2 as a potential target of the AnkX catalysed phosphocholinlation

The novel identified substrate protein Inosine-5'-monophosphate dehydrogenase type II is an intriguing and well studied enzyme as it is linked to several cellular processes like nucleotide pool balancing and cell proliferation. IMPDH2 catalyses the oxidation of inosine-5'-monophosphate (IMP) to xanthosine-5'-monophosphate (XMP) (Figure 26, page 67), which is the first committed and rate limiting step in guanine nucleotide biosynthesis. As IMP is also precursor to adenine nucleotides, IMPDH2 activity affects not only the guanine but the whole purine nucleotide metabolism and is ultimately linked to ATP and GTP pool sizes. Inhibition of IMPDH2 depletes guanine nucleotides and increases adenine nucleotide pools in a variety of investigated cell lines, consequently affecting several processes like RNA and DNA synthesis as well as cell growth.
Targeting IMPDH2 by posttranslational phosphocholination might enable *Legionella pneumophila* to manipulate regulation mechanisms; thus having major cellular consequences.

To further explore the AnkX catalysed phosphocholination of IMPDH2 we recombinantly expressed the protein and investigated it in a similar phosphocholination assay that was previously used to evaluate model proteins in section 3.4. The purified IMPDH2 (5 µM), CDP-choline (100 µM) and AnkX (0.5 µM) were incubated at room temperature and samples were taken at different time points. The reaction was quenched in boiling laemmli buffer and the proteins were analysed for phosphocholination after SDS-PAGE and Western-Blot using an anti-phosphocholine antibody. As a control reaction, IMPDH2 and CDP-choline were incubated without AnkX (Figure 25). Western blot analysis displayed a time dependent phosphocholination of IMPDH2 in presence of AnkX reaching a state of maximum phosphocholination after 3 h. The comparison between the phosphocholination reaction and the control after over night incubation, demonstrates that the observed phosphocholination of IMPDH2 is AnkX dependent (Figure 25b).

To further determine which amino acid residue of IMPDH2 is targeted by AnkX and modified with a phosphocholine group, fully phosphocholinat

![Figure 25](image_url)
additional mass of a phosphocholine moiety (Figure 26). Fragmentation as neutral loss identified threonine 147 and serine 160 as sites of modification. Interestingly, both these residues are located in the Bateman domain (CBS) of IMPDH2, which was identified to play a major regulatory role in adenylate nucleotide biosynthesis and cellular proliferation (Figure 26b). Furthermore, the CBS domain is postulated to act as an energy sensing module, accelerating IMPDH2 activity at increased ATP concentrations.

![Figure 26: a) Identified peptides from IMPDH2 displaying a mass increase corresponding to phosphocholine. The detected masses correspond to the [M]^+ state of the peptides. The modified residues were identified using the ptmRS algorithm. The value gives the probability of the identification. b) Crystal structure of IMPDH2 from PDB (1NF7). The modified CBS domain is displayed in yellow. Phosphocholinated residues are indicated by red circles. Thr147 is located in a flexible loop region that is not resolved in the crystal structure, here indicated by dashed orange line. c) Principle of IMPDH2 activity assay. IMPDH2 catalyses the oxidation of inosine-5'-monophosphate (IMP) to xanthosine-5'-monophosphate (XMP) and parallel reduces one equivalent nicotinamide adenine dinucleotide (NAD+ to NADH). Measuring the UV absorption at 340 nm allows following the conversion of NAD+ to NADH, which correlates with the formation of XMP. d) IMPDH2 activity after phosphocholination in presence and absence of ATP. IMPDH2 was incubated with CDP-choline and AnkX for 3h to reach maximum phosphocholination. The reaction mixture was assayed for IMPDH2 activity. As control reactions IMPDH2 was incubated with only CDP-choline (ctrl a) and with only AnkX (ctrl b) before measuring the catalytic activity.]

10 LC-MS/MS experiments and site identification was performed by Dr. Clarissa Dickhut, Leibniz institute for analytical sciences Dortmund, Germany.
To test the effect of phosphocholination on the enzymatic activity of IMPDH2, we assayed the nicotinamide adenine dinucleotide (NAD\(^+\)) dependent catalytic oxidation of IMP to XMP by following the formation of reduced NAD\(^+\) (NADH) in an UV/VIS based assay. IMPDH2 (5 µM), IMP (100 µM) and NAD\(^+\) (200 µM) were incubated at 37°C and the NADH absorbance was measured at 340 nm. The activities of native and phosphocholinated IMPDH2 were compared under identical assay conditions but, as illustrated in figure 26d, no significant difference could be determined. To test for possible IMPDH2 activation by ATP and the influence of phosphocholination on this process, the assay was repeated in the presence of ATP (500 µM) but no ATP mediated increase in IMPDH2 activity could be detected nor did the phosphocholination affect the catalysed reaction under the tested conditions. This observation is in line with earlier publications documenting full catalytic activity of IMPDH2 containing mutations in the CBS domain or even having these domains deleted completely.\(^{155-157}\)

Our experimental data demonstrates that IMPDH2 is a novel substrate protein for the AnkX catalysed phosphocholination \textit{in vitro}. To further determine if IMPDH2 phosphocholination is also relevant \textit{in vivo}, the phosphocholination state of IMPDH2 isolated from infected macrophages would have to be determined in future work. However, a first hint that IMPDH2 plays a relevant role during the course of \textit{Legionella} infection is the identification of this protein from \textit{Legionella} containing vacuoles (LCVs) isolated from infected RAW264.7 macrophages.\(^{158}\) Although modifications of the CBS domains in IMPDH2 do not affect its \textit{in vitro} activity, modification or deletion of these domains has significant effects \textit{in vivo}. Deletion of the CBS domains in \textit{E.Coli} for example lead to higher ATP levels followed by a cell cycle arrest and finally cell death.\(^{159}\) Furthermore, it has been shown that targeting IMPDH2 via protein-protein interactions modulates cell growth through interaction with the CBS domain in human K562 cells.\(^{153}\) These results suggest that phosphocholination of the CBS domain targets an important regulative region in IMPDH2 that modulates cellular proliferation. The modification of this domain could enable \textit{Legionella pneumophila} to directly interfere with these mechanisms to adjust its host’s cell growth. This would be highly favorable for \textit{L. pneumophila} as it would provide time and space for the bacterial growth and save energy and resources for the bacterial replication that would normally be consumed during cellular replication. Assuming that these findings not only apply for \textit{Legionella pneumophila}, our results highlight a new molecular strategy for infectious bacteria to manipulate their host cells.
7. Summary and concluding remarks

The high interest of fundamental biological research in studying and manipulating protein function resulted in the development of various technologies for protein modification allowing protein functionalisation with a variety of synthetic reporter groups. Although powerful, the existing strategies are often limited by large tag size, lack of specificity or are restricted to a limited number of suitable labeling probes (see section 1.4). To expand the toolbox for protein labeling strategies, the here presented results describe the development of a new method for the site-specific protein functionalisation, based on the transfer of synthetically produced CDP-choline derivatives.

This thesis describes the establishment of a robust and highly flexible synthetic route that allows straightforward access to an array of functionalised CDP-choline derivatives. These analogues either carry PEG or short carbon linkers that can be conveniently functionalised with any amine-reactive probe via a terminal primary amine function. The final analogues are accessible in three to six synthesis steps in reasonable overall yields. Following this strategy, CDP-choline analogues were prepared that carry different fluorescent probes, affinity handles and convenient bioorthogonal handles, suitable for fast click reactions (see section 3.2 and 4).

It was demonstrated that the appropriately functionalised CDP-choline derivatives serve as co-substrate for the *Legionella* enzyme AnkX and could be used to functionalise AnkX's target protein Rab1b via catalytic phosphocholination. Our experiments proved AnkX's high promiscuity towards a wide range of functional CDP-choline analogues, making this approach superior to many of the existing methodologies for *in vitro* protein labeling. It was demonstrated that AnkX can be used to fluorescently label and biotinylate it's protein substrates. Furthermore, AnkX was used to install alkynes, azides, dibenzocyclooctynes, trans-cyclooctenes and tetrazines onto Rab1b, making it accessible for highly efficient and specific bioorthogonal click reactions that are useful for various bioconjugation reactions.

Furthermore, we identified the short octapeptide TITSSYYR from the switch II region of Rab1b as the minimum recognition sequence for the efficient AnkX catalysed phosphocholination. It was demonstrated that this peptide sequence could be introduced to a set of model proteins to turn them into substrates for the AnkX catalysed phosphocholination. In combination with synthetically prepared CDP-choline analogues the AnkX catalysed phosphocholination was used to fluorescently label the protein fusion TITSSYYR-SUMO, demonstrating the general applicability of the functionalisation strategy. Combining these results with the displayed promiscuity of AnkX suggest that the here presented approach can be used
to site-specifically modify proteins of interest with any conceivable functionality.

Additionally, delabeling experiments with the *Legionella* protein Lem3 established that the dephosphocholinase was able to hydrolytically cleave the previously installed modification from Rab1b and the fusion protein TITSSYYR-SUMO. The observed demodification activity of Lem3 enables the recovery of valuable protein samples and allows sequential protein labeling and delabeling to evaluate suitable labeling probes for distinct applications.

Furthermore, the contrary activities of AnkX and Lem3 were exploited to site-specifically label the model protein TITSTYYR-SUMO-Rab1b with two different fluorescent dyes on two distinct positions in the protein. This approach could be used in future applications to install FRET pairs in protein of interest and analyse distances or conformational changes.

In the attempt to expand the AnkX catalysed protein functionalisation from pure *in vitro* applications to more complex experimental settings, the protein labeling of the transiently expressed membrane protein TITSSYYR-hY₂R was investigated. The preliminary results suggest the need of further optimisation of the experimental setup as the utilised fluorescein functionalised CDP-choline derivative 26 exhibited strong unspecific staining of the cellular membrane in absence of AnkX. In future experiments the transmembrane protein will be modified with non-fluorescent trans-cyclooctene (TCO) functionalised CDP-choline 46 or 54 and subsequently visualised with a suitable fluorogenic tetrazine dye. These dyes are essentially nonfluorescent in solution and are switched on upon reaction with the installed TCO handle, which reduces background labeling and enhances the signal-to-noise ratio in labeling reactions.

In summary, the here reported results serve as promising basis to establish the presented labeling strategy as valid alternative to existing protein functionalisation strategies. The advantages of the AnkX/Lem3 catalysed protein labeling are the short recognition tag that can be introduced to terminii and internal loop regions of protein of interest; the formation of stable phosphodiester bonds between protein and label that can be hydrolytically cleaved when required; the possibility to modify POIs with a wide range of functionalised CDP-choline analogues; the reasonable fast transfer rates between $2.4 \times 10^2$ M$^{-1}$s$^{-1}$ and $1.8 \times 10^4$ M$^{-1}$s$^{-1}$ of the labeling reaction and the possibility of dual labeling. Furthermore, it should be considered that all results were achieved using native AnkX$_{1-800}$ that did not undergo any protein engineering of its active site. Thus, for future applications higher transfer rates could be achieved when AnkX is evolved to better recognise the functionalised CDP-choline derivatives and its substrate proteins.

However, the approach as it is presented in this thesis also displays clear limitations. So far the strategy is restricted to the functionalisation of
partially purified protein samples \textit{in vitro}. To enable the applicability of the presented protein functionalisation strategy \textit{for in vivo} labeling, the enzyme AnkX has to be adjusted in future projects to not recognise and functionalise its native substrate proteins. To allow a rational mutagenesis of AnkX active site we designed an experimental approach to covalent trap the phosphocholine transfer in a ternary complex between AnkX, a phosphocholine cross-linker and the substrate protein Rab1b. This complex is intended to facilitate the generation of protein crystal structures in future experiments, which will give valuable information about the interacting amino acid residues of AnkX and Rab1b. This information should allow a rational approach to mutate AnkX’s active site and to adjust the recognition tag accordingly.

The initial design of cell permeable CDP-choline derivative 68 carrying a polyamine linker to partially compensate the negative charged diphosphate of CDP-choline at physiological pH have been investigated in preliminary experiments. These first results revealed that the conjugation of PEG functionalised CDP-choline analogues with tertiary amine containing fluorescent dyes like TAMRA exhibit a similar cell permeability promoting effect as the use of the polyamine linker. Future experiments have to prove if the polyamine linker also promotes cell permeability when CDP-choline 68 is conjugated to neutral probes. An alternative approach to generate cell permeable CDP-choline analogues is to mask the diphosphate with suitable protecting groups like 4-acetoxy benzyl. After passage over the cell membrane, the endogenous esterase activity will remove the phenolic acetates of the protecting groups, leading to a release of CDP-choline analogues inside the cell that would be available for AnkX catalysed functionalisation.

Apart from the development of AnkX catalysed phosphocholination as a protein functionalisation strategy, the results presented in this thesis demonstrate that the biotin functionalised CDP-choline derivative 35 proved as valuable molecular tool to investigate the phosphocholination reaction of AnkX in complex protein mixtures. It has been shown that this probe is applicable to isolate and identify potential substrate proteins for this unique post-translational modification. Next to the known targets for the AnkX catalysed phosphocholination, we isolated the protein IMPDH2 from HeLa cell lysate and demonstrated its phosphocholination \textit{in vitro}. Furthermore, we identified two residues in the CBS domain as side of modification and showed that a phosphocholination does not interfere with IMPDH2’s catalytic activity.

As several studies on IMPDH2 highlight its role in nucleotide pool balancing and regulation of cell proliferation, an \textit{in vivo} phosphocholination of this protein in the course of \textit{Legionella} infection might have major impact on cellular processes.\textsuperscript{155–157} By phosphocholination of the CBS domain in
IMPDH2, AnkX targets a regulative region similar to the modified switch II region in Rab1b. As the phosphocholination of the switch II region in Rab1b blocks its interaction with several effector proteins, the modification of the CBS domain may disturb similar protein-protein interactions of IMPDH2 that are involved in the control of the cell cycle. These results can provide a significant advance in the understanding of Legionella infection and give rise to the new hypothesis that the phosphocholination of IMPDH2 provides L. pneumophila with a control of the hosts’ cell cycle.

Thus, in future experiments the impact of IMPDH2 phosphocholination has to be elucidated. Therefore the comparison of ATP/GTP levels in uninfected cells versus cells that are infected with Legionella and AnkX deficient Legionella strains would give insight whether the AnkX catalysed phosphocholination interferes with IMPDH2 regulated nucleotide balancing that consequently determines cellular proliferation. As IMPDH2 was found to localise at the Legionella containing vacuole (LCV) in infected macrophages, the in vivo phosphocholination of IMPDH2 could also regulate this subcellular localisation. To further investigate a possible AnkX/Lem3 regulated recruitment of IMPDH2 to the LCV, future investigation will focus on isolating LCVs from macrophages that are infected with Legionella, AnkX deficient and Lem3 deficient Legionella strains, and IMPDH2 localisation will be quantified and compared using immunohistochemistry.
8. Acknowledgements

In the end, after an eventful 5-year scientific journey, including a literal journey of 1,948 kilometers from Dortmund to Umeå, it remains to me to thank my companions.

First of all and most importantly I thank my supervisor Christian Hedberg for accepting me as a PhD student and giving me the opportunity to work on this highly interdisciplinary project that formed the basis of this thesis. From the beginning Christian taught me the fundamentals of laboratory work and greatly improved my skills in organic chemistry. Christian’s advice and ideas always pushed forward scientific discussions and helped to solve the various problems that pop up in the laboratory. Here, his supply of good homebrew contributed significantly.

Furthermore, I would like to thank my assistant supervisor Mikael Elofson and the members of my advisory committee, Åsa Gylfe and Uwe Sauer, for accompanying my studies during the last three years and helping me to adapt to the Swedish PhD program, which allows me to stand here today and defend my thesis.

Next, I would like to thank our cooperation partner Aymelt Itzen and his group at the Technische Universität München for the close collaboration and for providing the necessary expertise on protein chemistry and biochemistry. I’d like to address special thanks to Katharina Heller and Stefan Ernst, who contributed greatly to the results that are presented and discussed in this thesis. Moreover, I like to thank the priority program (SPP1623) of the German Research Foundation (DFG) and the Knut and Alice Wallenberg Foundation (KAW 2013.0187) for founding.

Naturally, I like to thank all former group members, Martijn, Bart, Samy and Julia in Dortmund; Nasibeh, Marie and "the annoying summer student" Marianne in Umeå; as well as the current group members, Lakshmi, Saeed, Nadja, Christian, Lindon and Michael; for being good company and helpful discussion partners in recent years. A special thank belongs to Lindon for proofreading my thesis and for the less scientific discussions in the rare hours of sunshine at Rött.

Thanks to Michael, who not only shared lab- and office space but also living space with me for two years. Thank you for all the help in the laboratory, especially for introducing me to the field of biochemical experimentation and pull-downs.

Lab life would have been much less joyful without all the colleagues on the corridor. Thanks for making the chemistry department a great place to study and work. A special thanks goes to the people who regularly joined the beer corner for one or two or three or more after work refreshments. Here, I
would like to highlight Papa Beer's commitment and thank him for keeping up the high standard of our beer selection.

Spending the first two years of my PhD studies at the MPI in Dortmund, I highly benefited from the great scientific environment in the institute and would like to thank Herbert Waldmann for his generous support. I like to give special thanks to Petra Janning for the support in MS analysis and protein identification experiments. Furthermore, I want to thank all group members, with special reference to Guillaume, Vincent, Melanie, Küchler, Cromm, Kirsten, Gunther, Jochen, Janine, Julia and Kyra who were great company, not only on Schnappstour. Thanks to the Chemiewixxas, including Caro, for awesome 7 years in Dortmund. The friendship that grew from our studies at the TU Dortmund is probably more important than the knowledge we gained. Although unrelated to university life, I would like to thank the BoBs for all the awesome nights at Tobi's and everything else that we better keep confidential.

Thanks to Clarissa Dickhut and Rene Zahedi at the ISAS in Dortmund as well as Leonie Swart and Hubert Hilbi at the university Zurich for support and collaboration.

Finally, I want to close this thesis by thanking my family and my friends in Germany and Sweden for joining me on this journey and making the hard times easier and the good times even more joyful.
References

8. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins.


31. Howard C. Hang †, Ernst-Jan Geutjes †, Gijsbert Grotenbreg ‡, Annette M. Pollington †, Marie Jose Bijlmakers ‡ and, Hidde L. Ploegh* †. Chemical Probes for the Rapid Detection of Fatty-Acylated Proteins in Mammalian Cells. 2007.


81


122. Medintz I. FRET — Förster Resonance Energy Transfer: From Theory to Applications Ed. by Igor Medintz ... Wiley-VCH; 2014.


124. Svendsen A. Understanding Enzymes.


Appendix

**Rab1b functionalisation via AnkX catalysed phosphocholination**

**Labeling of Rab1b**
The functionalisation of Rab1b with various CDP-choline analogues was followed in a time dependent manner to determine the kinetics of the labeling reaction. Rab1b (50 µM) was incubated with AnkX_{1-800} (1 µM) and the respective cosubstrate (200 µM) in Tris-buffer (20 mM Tris, 100 mM NaCl, 1 mM MgCl₂, 1 mM DTT, 100 µM GDP, pH 7) at room temperature for 5h. Samples (1.5 µL) were taken at distinct time points and immediately pipetted into Laemmli-buffer (28.5 µL) to quench the reaction at 95°C for 3 min.

**In-gel fluorescence analysis**
In case of non-fluorescent bioorthogonal CDP-choline derivatives (azide 43, DBCO 45 and 53, TCO 46 and 54, tetrazine 47 and 55) the quenched reaction samples were incubated with corresponding Cy5 dyes (10 µM) for 1h. For the labeling of alkyne 39 with Cy5 azide, CuSO₄ (5 mM) and TCEP (10 mM) were added. Subsequently samples were applied to SDS-PAGE in 100 ng protein aliquots and separated using a 20% acrylamide gel. Fluorescent bands were detected with Typhoon FLA 9500 Variable Mode Imager (GE Healthcare, Amersham Pharmacia Biotech).

**Western blot analysis**
Functionalisation with biotin 35 was analysed after western blot. Therefore proteins were transferred onto nitrocellulose after electrophoresis using a Trans-Blot® TurboTM Blotting System (Bio-Rad). After blotting, the nitrocellulose membrane was blocked in Tris buffered saline containing 0.1% Tween20 (TBS-T) and 1xRoti®-Block (Roth) for 0.5 h. Subsequently, a streptavidin-HRP conjugate (Pierce™ Streptavidin Poly-HRP, Thermo Fisher) was pipetted into the blocking solution (1:1000) and incubated overnight at 4°C. The membrane was washed three times with TBS-T for 10 minutes and developed with chemiluminescent HRP substrate (SuperSignal West Pico, Thermo Fisher). Subsequently HRP-signals were detected with a luminescence imaging system (Odyssey, LI-COR Biosciences).
Determination of catalytic efficiencies

For the determination of catalytic efficiencies ($k_{\text{cat}}/K_M$), band intensities were plotted against time. These plots were then fitted to a single exponential function according to equation (1) using the software OriginPro (OriginLab). The $k_{\text{cat}}/K_M$ was obtained by dividing $k_{\text{obs}}$ by $E_0$:

$$I(t) = I_0 + I_A \times e^{-k_{\text{obs}}/t} \quad \text{Equation (1)}$$

with $I(t)$: time dependent gel band intensity, $I_0$: minimal gel band intensity, $I_A$: amplitude of intensity (i.e. $I_{\text{max}} - I_0$, with $I_{\text{max}}$: maximum intensity), $k_{\text{obs}}$: observed rate constant.

Dual labeling of TITSTYR_MBP-Rab1b

MBP-Rab1b (40 µM) was incubated with AnkX1-800 (4 µM) and CDP-choline 26 (5 mM) in Tris-buffer (20 mM Tris, 50 mM NaCl, 1 mM MgCl$_2$, 1 mM DTT, 10 µM GDP, pH 7.5) at room temperature over night. Excess of nucleotide 26 was removed by gel filtration using a Superdex 75 column, 16/600 (GE). The protein was eluted with Tris-buffer and delabled for 3h at room temperature (15 µM MBP-Rab and 0.5 µM Lem3) and subsequently purified from Lem3 by affinity chromatography using amylose-resin. For the second labeling MBP-Rab1b (5 µM) was incubated with AnkX (0.25 µM) and CDP-choline 34 in Tris-buffer for 1h at room temperature and excess of nucleotide 34 was removed by gel filtration. Samples were taken after each step and digested (5 µM MBP-Rab1b and 1 µM TEV) for 2h at 30°C. Subsequently samples were applied to SDS-PAGE in 100 ng protein aliquots and separated using a 20% acrylamide gel. Fluorescent bands were detected with Typhoon FLA 9500 Variable Mode Imager (GE Healthcare, Amersham Pharmacia Biotech).

Labeling and enrichment of protein targets for AnkX catalysed phosphocholination from cell lysate

Cell lysis

HeLa cells were seeded and cultured until confluence. Subsequently cells were collected, transferred to an Eppendorf tube, centrifuged and washed with PBS. The cell pellets were suspended in 1 mL lysis buffer (0.1% Tween20, 0.1% NP40, 0.1% Triton X100, 50 mM PIPES, 50 mM NaCl, 5 mM MgCl$_2$, 5 mM EGTA, pH 7.4, protease and phosphatase inhibitor cocktail) and incubated for 1 h on ice. After 30 min the sample was vortexed every 10 min. Afterwards, the lysate was clarified by centrifugation at 13000
rpm for 30 min at 4°C. Approximate total protein concentration was determined by Bradford assay.

**Labeling of cell lysate**

The protein mixture from HeLa cell lysate (500 µg total protein) was diluted with lysis buffer to 250 µL reaction volume and CDP-choline biotin 35 (100 µM) was added. To start the labeling reaction AnkX (1 µM) was added and the reaction was incubated for 24h. As a negative control reaction, lysate was incubated with CDP-choline 35 without addition of AnkX.

**Western blot analysis of biotinylated proteins**

Samples from the labeling reaction were taken after distinct time points and quenched in Laemmli buffer at 95°C. Subsequently samples were separated by SDS-PAGE and transferred onto a nitrocellulose membrane using a Trans-Blot® TurboTM Blotting system (Bio-Rad). The membrane was blocked in Tris buffered saline, containing 0.1% Tween20 (TBS-T) and 1xRoti®-Block (Roth) for 0.5 h. Subsequently, a streptavidin-fluorescein conjugate (Thermo Fisher) was pipetted into the blocking solution (1:1000) and incubated overnight at 4°C. The membrane was washed three times with TBS-T for 10 minutes and fluorescein-signals were detected on a fluorescence imaging system (Odyssey, LI–COR Biosciences).

**Isolation and identification of biotinylated proteins**

Proteins from the labeling reaction were precipitated by addition of acetone (approx. 1 mL, 30 min at -20°C, centrifuge 13000 rpm, 10 min, 4°C). The supernatant was discarded and the protein pallet was washed three times with ice cold MeOH (resuspended by sonication followed by centrifugation (13000 rpm, 4°C) between each washing step). Proteins were resuspended in PBS containing 0.2% SDS (500 µL) by sonication and incubated with streptavidin magnetic beads (New England Biolabs, #S1420S, 250 µL per sample) for 1h under rotatory stirring. Subsequently, unbound proteins were removed by washing the beads with PBS containing 0.2 % SDS (3 × 1 mL), 6 M aqueous urea (2 × 1 mL) and PBS (3 × 1 mL). Bound proteins were eluted by incubation with Laemmli buffer (50 µL) for 5 min at 95°C and subsequently applied to SDS-PAGE. Protein containing bands were stained with coomassie (QC colloidal coomassie stain, Bio-Rad) and cut from the gel. Gel pieces were washed with 25 mM ammonium bicarbonate (ABC) buffer / acetonitrile (3:1 followed by 1:1) to destain the pieces. Subsequently the proteins were reduced for 45 min at 37°C with ABC buffer containing 50 mM DTT and alkylated for 1 h at room temperature in the dark with 55 mM iodoacetamide. Gel pieces were washed with ABC buffer/acetonitrile (1:1), shrunk with acetonitrile, followed by drying for 10 min in air. The pieces were covered with 10 µg/mL trypsin and incubated for 12 h at 30°C.
Samples were sonicated for 30 min at 0°C, centrifuged and the supernatant was transferred to a new tube. After evaporation of solvent under centrifugation in vacuum, the samples were submitted for MS analysis.

**IMPDH2 phosphocholination assay**

The AnkX catalysed phosphocholination of IMPDH2 was followed by western blot. Recombinantly expressed IMPDH2 (5 µM) was incubated with AnkX (500 nM) and CDP-choline (100 µM) in phosphocholination buffer (20 mM Tris, 100 mM NaCl, 1 mM MgCl₂, 1 mM DTT, pH 7). His₆-MBP-PC-Rab1b₃₋₁₇₄ (2 µM) was added to the reaction as loading control. After starting the reaction by adding AnkX, samples were taken at different time points, mixed with boiling Laemmli-buffer to quench the reaction and subsequently applied to SDS-PAGE (100 ng protein). After electrophoresis the proteins were transferred onto a nitrocellulose membrane using a Trans-Blot® TurboTM Blotting system (Bio-Rad). The membrane was blocked in Tris-buffered saline, containing 0.1% Tween20 (TBS-T) and 1xRoti®-Block (Roth) for 0.5 h. Subsequently, the anti-PC-Antibody (TEPC15, sigma) was pipetted into the blocking solution in a 1:1000 dilution and incubated overnight at 4°C. The membrane was washed three times with TBS-T for 10 minutes and afterwards incubated with the secondary antibody-HRP conjugate (goat anti-mouse IgG, IgA, IgM (H+L), horseradish peroxidase conjugate, life technologies) in TBS-T for 45 min. The membrane was washed three times with TBS-T for 10 minutes and developed with chemiluminescent HRP substrate (SuperSignal West Pico, Thermo Fisher). Subsequently HRP-signals were detected with a luminescence imaging system (Odyssey, LI–COR Biosciences).

**IMPDH2 activity assay**

The catalytic activity of IMPDH2 was assayed by following the increase in absorbance at 340 nm due to formation of NADH. Reactions were started by addition of IMPDH2 (5 µM) to the assay solution (50 mM Tris, 100 mM KCl, 3 mM EDTA, 5 mM MgCl₂) containing 100 µM inosine-5'-monophosphate (IMP) and 200 µM nicotinamide adenine dinucleotide (NAD⁺). Each reaction was assayed in triplicates in 96 well-plate formate and the absorbance at 340 nm was measured over 1h with a micro plate reader (Synergy H4, BioTek).
Experimental procedures for previously unpublished compounds presented in Sections 4 and 5

**Propargylcholine bromide (37).** 37 was synthesised following a procedure by Jao et al. Propargyl-bromide (1 g, 8.4 mmol) was added to dimethyl-ethanolamine (0.8 g, 9.24 mmol, 1.1 eq.) in 3 mL dry THF, while stirring on ice. The mix was allowed to reach room temperature and stirring was continued overnight. The resulting white solid was filtered and washed with ice cold THF (3 × 5 mL), to afford propargylcholine bromide 37 (1.5 g, 85%) as a white solid.

\[
\text{H NMR (400 MHz, D}_2\text{O)} \delta 4.36 (s, 2H), 4.08 (t, J = 6.3, 3.5 Hz, 2H), 3.70 - 3.59 (m, 2H), 3.27 (s, 6H); \text{C NMR (100 MHz, D}_2\text{O)} \delta 81.42, 70.62, 64.93, 55.48, 55.35, 51.29; \text{ESI-HRMS calculated mass [M]+ = 128.1070, observed mass [M]+ = 128.1079.}
\]

**Propargyl-phosphocholine (38).** POCl₃ (1.53 g, 10 mmol, 10 eq.) was dissolved in dry acetonitrile (1 mL) under an argon atmosphere and the solution was cooled with an ice bath before 2,6-lutidine (3.2 g, 30 mmol, 30 eq.) was added to the stirred solution. 37 (208 mg, 1 mmol) was dissolved in dry acetonitrile (2 mL) and added drop wise to the cooled reaction mixture. The reaction was stirred for 15 min. on ice and subsequently quenched with ice-cold water (10 mL). The acetonitrile was removed under reduced pressure and the residual aqueous solution was basified (pH 13-14) with Ca(OH)₂. The resulting slurry was filtered and evaporated to dryness yielding the title compound as a colorless oil (141 mg, 0.68 mmol, 68%).

\[
\text{H NMR (400 MHz, D}_2\text{O)} \delta 4.12 - 3.98 (m, 4H), 3.51 (m, 2H), 3.01 (t, 1H), 2.99 (s, 6H); \text{C NMR (100 MHz, D}_2\text{O)} \delta 81.71, 70.68, 63.46, 63.37, 59.03, 58.99, 55.34, 51.16; \text{P NMR (162 MHz, D}_2\text{O)} \delta -0.29; \text{ESI-HRMS calculated mass [M]+ = 208.0733, observed mass [M]+ = 208.0748.}
\]

**Propargyl cytidine diphosphocholine (39).** 38 (44 mg, 213 µmol) and cytidin-5'-phosphomorpholidate 4-morpholine-N,N'-dicyclohexylcarboxamidium (175 mg, 255 µmol, 1.2 eq.) were suspended in benzene (3 mL), frozen in liquid nitrogen and lyophilised to remove traces of water. The dried residue was dissolved in dry pyridine (6 mL) under an argon atmosphere. \text{H-tetrazole (18 mg, 255 µmol, 1.2 eq.), evaporated from acetonitrile (566 µL, 0.45 M),} was dissolved in dry pyridine (1 mL) under an argon atmosphere and subsequently dropped to the stirring reaction mixture. The reaction was heated to 60°C and stirred under argon overnight. The solvent was evaporated under reduced pressure and the residue was purified by HPLC to yield CDP-choline analogue 39 (33 mg, 64 µmol, 30%).

\[
\text{H NMR (600 MHz, D}_2\text{O)} \delta 8.13 (d, J = 8.0 Hz, 1H), 6.23 (d, J = 7.9 Hz, 1H), 5.86 (d, J = 3.6 Hz, 1H), 4.34 (m, 2H), 4.30 - 4.21 (m, 6H), 4.12 (m, 1H), 3.75 - 3.70
\]

\[
\]
(m, 2H), 3.21 (s, 6H), 3.19 (t, \( J = 2.5 \) Hz, 1H); \(^{13}\)C NMR (151 MHz, D\(_2\)O) \( \delta \) 162.80, 159.06, 148.38, 144.09, 95.20, 89.58, 83.22, 83.16, 81.81, 74.29, 70.91, 69.01, 64.42, 64.38, 63.68, 63.63, 63.55, 59.78, 59.74, 55.45, 51.33, 43.11; \(^{31}\)P NMR (243 MHz, D\(_2\)O) \( \delta \) -11.39 (d, \( J = 20.9 \) Hz), -12.23 (d, \( J = 20.9 \) Hz); ESI-HRMS calculated mass [M]\(^+\) = 513.1146, observed mass [M]\(^+\) = 513.1176.

3-bromopropane choline (bromide salt) (40). \( N,N \)-dimethyl-ethanolamine (1.1 g, 1.3 mL, 12.6 mmol) was added to 1,3-dibromopropane (39.6 g, 20 mL, 196 mmol, 15.5 eq.) under an argon atmosphere and allowed to stir overnight at room temperature. After full conversion of \( N,N \)-dimethyl-ethanolamine, ice cold diethyl ether (20 mL) was added and the formed precipitate was filtered, washed with ice cold diethyl ether (3 x 5 mL) and dried to yield the title compound as white solid (3.3 g, 11.3 mmol, 91%). \(^1\)H NMR (400 MHz, D\(_2\)O) \( \delta \) 4.09 – 4.04 (m, 2H), 3.62 – 3.50 (m, 6H), 3.18 (s, 6H), 2.40 (m, 2H); \(^{13}\)C NMR (100 MHz, D\(_2\)O) \( \delta \) 65.15, 63.98, 55.32, 51.64, 51.56, 29.00, 25.11; ESI-HRMS calculated mass [M]\(^+\) = 210.0488, observed mass [M]\(^+\) = 210.0504.

3-azidopropane choline (bromide salt) (41). 40 (2 g, 6.9 mmol) was dissolved in DMF (20 mL) and NaN\(_3\) (0.9 g, 13.7 mmol, 2 eq.) was added to the stirring solution. The reaction mixture was heated to 70°C and stirred overnight. The formed suspension was allowed to cool to room temperature and filtered. The filtrate was evaporated under reduced pressure and residual DMF was removed by coevaporation with toluene (3 x 5 mL). Final drying under high vacuum yielded the title compound as yellowish solid (1.5 g, 5.9 mmol, 86%). \(^1\)H NMR (400 MHz, D\(_2\)O) \( \delta \) 4.09 – 4.04 (m, 2H), 3.55 – 3.48 (m, 6H), 3.18 (s, 6H), 2.18 – 2.03 (m, 2H); \(^{13}\)C NMR (100 MHz, D\(_2\)O) \( \delta \) 65.15, 63.98, 55.32, 51.64, 51.56, 29.00, 25.11; ESI-HRMS calculated mass [M]\(^+\) = 173.1397, observed mass [M]\(^+\) = 173.1412.

3-azidopropane phosphocholine (42). POCl\(_3\) (472 mg, 3 mmol, 1.3 eq.) was dissolved in dry acetonitrile (8 mL) under an argon atmosphere and the solution was cooled with an ice bath before 2,6-lutidine (1 g, 9.5 mmol, 4 eq.) was added to the stirred solution. 41 (600 mg, 2.4 mmol) was dissolved in dry acetonitrile (2 mL) and added drop wise to the cooled reaction mixture. The reaction was stirred for 30 min on ice and subsequently quenched with ice-cold water. The acetonitrile was removed under reduced pressure and the residual aqueous solution was basified (pH 13-14) with Ca(OH)\(_2\). The resulting slurry was filtered and evaporated to dryness yielding the title compound as a colorless oil (369 mg, 1.46 mmol, 61%). \(^1\)H NMR (600 MHz, D\(_2\)O) \( \delta \) 4.30 (m, 2H), 3.71 – 3.61 (m, 2H), 3.51 – 3.44 (m, 4H), 3.17 (s, 6H), 2.12 – 2.05 (m, 2H); \(^{13}\)C NMR (151 MHz, D\(_2\)O) \( \delta \) 63.61, 62.96, 59.01, 58.98,
3-azidopropane CDP-choline (43). 42 (117 mg, 464 µmol) and cytidin-5'-phosphoromorpholidate 4-morpholine-N,N'-dicyclohexylcarboxamidium (635 mg, 928 µmol, 2 eq.) were suspended in benzene (6 mL), frozen in liquid nitrogen and lyophilised to remove traces of water. The dried residue was dissolved in dry pyridine (6 mL) under an argon atmosphere. 1H-tetrazole (64 mg, 928 µmol, 2eq.), evaporated from acetonitrile (2,1 mL, 0.45 M), was dissolved in dry pyridine (2 mL) under an argon atmosphere and subsequently dropped to the stirring reaction mixture. The reaction was heated to 60°C and stirred under argon overnight. The solvent was evaporated under reduced pressure and the residue was purified by HPLC to yield CDP-choline analogue 43 (133 mg, 239 µmol, 52%). 1H NMR (400 MHz, D2O) δ 8.22 (d, J = 8.0 Hz, 1H), 6.32 (d, J = 8.0 Hz, 1H), 5.95 (d, J = 3.4 Hz, 1H), 4.41 (m, 2H), 4.38 – 4.23 (m, 4H), 4.20 (m, 1H), 3.76 – 3.65 (m, 2H), 3.60 – 3.42 (m, 4H), 3.21 (s, 6H), 2.16 – 2.05 (m, 2H); 13C NMR (100 MHz, D2O) δ 159.08, 148.41, 144.09, 95.20, 89.56, 83.25, 83.16, 74.29, 69.01, 64.39, 63.03, 62.69, 59.60, 51.49, 47.79, 21.95; 31P NMR (162 MHz, D2O) δ -11.38 (d, J = 21.2 Hz), -12.18 (d, J = 21.2 Hz); ESI-HRMS calculated mass [M]⁺ = 558.1473, observed mass [M]⁺ = 558.1467.

3-aminopropane CDP-choline (44). Pd/C (20 mg) was suspended in MeOH (4 mL) under an argon atmosphere. Subsequently 43 (130 mg, 233 µmol) was dissolved in MeOH/H2O (2 mL, 1:1 v/v) and added to the stirring reaction mixture. The reaction was flushed with hydrogen and stirred for 4 h at room temperature. After completion the suspension was filtered over celite in a sintered funnel and was washed with several portions of methanol. The solvent was removed under reduced pressure to yield the title compound as colorless solid (68 mg, 124 µmol, 54%). 1H NMR (600 MHz, D2O) δ 8.16 (d, J = 8.0 Hz, 1H), 6.29 (d, J = 8.0 Hz, 1H), 5.91 (d, J = 3.8 Hz, 1H), 4.40 (m, 2H), 4.31 (m, 4H), 4.18 (m, 1H), 3.76 – 3.71 (m, 2H), 3.58 – 3.52 (m, 2H), 3.21 (s, 6H), 3.10 (t, J = 7.5 Hz, 2H), 2.28 – 2.20 (m, 2H); 13C NMR (151 MHz, D2O) δ 159.05, 148.31, 144.01, 95.18, 89.68, 83.06, 74.22, 69.03, 64.49, 63.38, 61.36, 59.70, 51.93, 48.84, 36.22, 20.67; 31P NMR (162 MHz, D2O) δ -11.27 (d, J = 20.5 Hz), -12.14 (d, J = 20.5 Hz); ESI-HRMS calculated mass [M]⁺ = 532.1568, observed mass [M]⁺ = 532.1564.

Cbz-amino-PEG4 choline (bromide salt) (49). 1-bromo-carboxybenzylamine -PEG4 (1 g, 2.6 mmol) was dried under high vacuum and subsequently dissolved in dimethylaminoethanol (2.3 g, 2.6 mL, 26 mmol, 10 equiv.). The solution was stirred overnight at room temperature,
subsequently diluted with DMF (5 mL) and evaporated to dryness. The residue was taken up in 5 mL of water, frozen in liquid nitrogen and lyophilised to give spacer 49 as yellow oil (1.2 g, 2.5 mmol, 98%). \(^1\)H NMR (400 MHz, D$_2$O) \(\delta 7.50 - 7.36\) (m, 5H), 5.13 (s, 2H), 4.04 – 4.00 (m, 2H), 3.93 (m, 2H), 2.70 – 2.59 (m, 12H), 5.35 – 5.31 (m, 2H), 3.4 (m, 2H), 1.16 (s, 6H); \(^{13}\)C NMR (151 MHz, D$_2$O) \(\delta 128.81, 128.78, 128.37, 127.69, 69.57, 69.46, 69.39, 69.18, 66.85, 66.20, 66.18, 64.13, 64.07, 55.35, 52.15, 52.11, 40.03;\) ESI-HRMS calculated mass [M+H]$^+$ = 400.2562, observed mass [M+H]$^+$ = 400.2593.

Cbz-amine-PEG$_4$ phosphocholine (50). POCl$_3$ (574 mg, 3.8 mmol, 1.5 eq.) was dissolved in dry acetonitrile (5 mL) under an argon atmosphere and the solution was cooled with an ice bath before 2,6-lutidine (1.2 g, 11.3 mmol, 4.5 eq.) was added to the stirred solution. Choline spacer 49 (1.2 g, 2.5 mmol) was dissolved in dry acetonitrile (5 mL) and added dropwise to the cooled reaction mixture. The reaction was stirred for 30 min on ice and subsequently quenched with ice-cold water. The acetonitrile was removed under reduced pressure and the residual aqueous solution was basified (pH 13-14) with Ca(OH)$_2$. The resulting slurry was filtered and evaporated to dryness and the residue was purified by HPLC yielding the title compound as a colorless oil (370 mg, 775 µmol, 31%). \(^1\)H NMR (400 MHz, D$_2$O) \(\delta 7.48 - 7.37\) (m, 5H), 5.12 (s, 2H), 4.27 (s, 2H), 3.93 (s, 6H); \(^{13}\)C NMR (151 MHz, D$_2$O) \(\delta 158.38, 128.77, 128.37, 127.74, 127.69, 69.56, 69.45, 69.38, 69.18, 66.5, 64.79, 64.30, 64.06, 63.43, 63.08, 59.00, 58.96, 52.23;\) \(^{31}\)P NMR (162 MHz, D$_2$O) \(\delta -0.58;\) ESI-HRMS calculated mass [M]$^+$ = 479.2153, observed mass [M]$^+$ = 479.2197.

CDP-choline-PEG$_4$ Cbz amine (51). Phosphocholine spacer 50 (0.91 mg, 190 µmol) and cytidin-5’-phosphomorpholidate 4-morpholine-N,N’-dicyclohexylcarboxamidium (635 mg, 928 µmol, 2 eq.) were suspended in benzene (6 mL), frozen in liquid nitrogen and lyophilised to remove traces of water. The dried residue was dissolved in dry pyridine (6 mL) under an argon atmosphere. \(^1\)H-tetrazole (40 mg, 570 µmol, 3 eq.), evaporated from acetonitrile (1.3 mL, 0.45 M), was dissolved in dry pyridine (1 mL) under an argon atmosphere and subsequently dropped to the stirring reaction mixture. The reaction was heated to 60°C and stirred under argon overnight. The solvent was evaporated under reduced pressure and the residue was purified by HPLC to yield CDP-choline analogue 51 (50 mg, 65 µmol, 34%). \(^1\)H NMR (600 MHz, D$_2$O) \(\delta 8.09\) (d, \(J = 8.0\) Hz, 1H), 7.36 – 7.31 (m, 5H), 6.20 (d, \(J = 8.0\) Hz, 1H), 5.88 – 5.80 (m, 1H), 5.03 (s, 2H), 4.33 (q, \(J = 5.3\) Hz, 2H), 4.27 – 4.23 (m, 2H), 4.15 (m, 2H), 3.86 (t, \(J = 4.5\) Hz, 2H), 3.64 (s, 2H), 3.57 (m, 10H), 3.51 (t, \(J = 5.4\) Hz, 2H), 3.25 (t, \(J = 5.3\) Hz, 2H),
3.13 (s, 6H); $^3$C NMR (151 MHz, D$_2$O) δ 158.94, 158.25, 148.15, 143.99, 136.43, 128.73, 128.30, 127.61, 95.11, 89.63, 83.05, 82.99, 74.20, 69.54, 69.51, 69.41, 69.35, 69.17, 68.96, 66.75, 64.69, 64.64, 64.52, 64.26, 64.04, 59.83, 59.80, 52.22, 40.03; $^{31}$P NMR (162 MHz, D$_2$O) δ -11.47 (d, $J = 21.0$ Hz), -12.28 (d, $J = 21.0$ Hz); ESI-HRMS calculated mass [M]$^+$ = 784.2566, observed mass [M]$^+$ = 784.2564.

**CDP-choline-PEG$_4$-amine (52).** Pd/C (10 mg) was suspended in MeOH (4 mL) under an argon atmosphere. Subsequently CDP-choline 51 (51 mg, 65 µmol) was dissolved in MeOH (2 mL) and added to the stirring reaction mixture. The reaction was flushed with hydrogen and stirred for 4 h at room temperature. After completion the suspension was filtered over celite in a sintered funnel and was washed with several portions of methanol. The solvent was removed under reduced pressure to yield the title compound as colorless oil (38 mg, 58 µmol, 90%).

$^1$H NMR (400 MHz, D$_2$O) δ 8.19 (d, $J = 7.9$ Hz, 1H), 6.30 (d, $J = 7.9$ Hz, 1H), 5.95 (d, $J = 3.3$ Hz, 1H), 4.41 (m, 2H), 4.38 – 4.27 (m, 4H), 4.20 (m, 1H), 4.00 (s, 2H), 3.80 – 3.75 (m, 4H), 3.73 (m, 10H), 3.28 (s, 6H), 3.22 (t, $J = 5.1$ Hz, 2H); $^13$C NMR (151 MHz, D$_2$O) δ = 95.31, 89.57, 84.02, 83.03, 74.25, 69.65, 69.61, 69.49, 69.49, 69.01, 68.71, 66.33, 64.29, 64.10, 63.85, 63.54, 61.16, 59.76, 51.37, 39.03; $^{31}$P NMR (162 MHz, D$_2$O) δ = -11.36 (d, $J = 21.3$ Hz), -12.17 (d, $J = 21.3$ Hz); ESI-HRMS calculated mass [M]$^+$ = 650.2198, observed mass [M]$^+$ = 650.2250.

**General procedure for derivatisation of CDP-choline analogues with NHS-ester**

CDP-choline analogue 44 or 52 was dissolved in sat. aq. NaHCO$_3$ (50 µL/mg CDP-choline) and the corresponding NHS-ester (1.1 eq.), dissolved in little DMF, was added to the stirring reaction. After completion the solvents were removed under reduced pressure and the residue was purified by reverse phase column chromatography (Sep-Pak C18) applying an H$_2$O/MeCN gradient.

**DBCO-C$_3$-CDP-choline (45).** Yield: 16.8 mg, 21%, white solid. $^1$H NMR (600 MHz, D$_2$O) δ 8.14 (d, $J = 8.0$ Hz, 1H), 7.62 – 7.31 (m, 8H), 6.24 (d, $J = 7.4$ Hz, 1H), 5.90 – 5.85 (m, 1H), 5.03 (d, $J = 14.5$ Hz, 1H), 4.37 (s, 2H), 4.30 – 4.14 (m, 4H), 3.74 (d, $J = 14.4$ Hz, 1H), 3.61 (d, $J = 4.5$ Hz, 2H), 3.33 – 3.29 (m, 2H), 3.13 – 3.09 (m, 6H), 3.06 – 2.99 (m, 2H), 2.52 (s, 1H), 2.23 – 2.16 (m, 2H), 2.14 – 2.06 (m, 1H), 2.04 (s, 1H), 1.85 (s, 2H); $^3$C NMR (151 MHz, D$_2$O) δ 174.35, 158.95, 150.53, 148.25, 147.86, 144.02, 131.74, 129.14, 129.08, 128.91, 128.41, 128.04, 127.08, 125.74, 122.32, 121.70, 120.75, 95.18, 89.63, 89.59, 83.04, 74.24, 68.95, 64.46, 64.42, 63.65, 63.24, 59.67, 59.64, 55.49, 51.36, 35.95, 30.73, 29.94, 22.07; $^{31}$P NMR (162 MHz, D$_2$O)
\( \delta = -11.47 \) (d, \( J = 20.6 \) Hz), -12.24 (d, \( J = 20.6 \) Hz); ESI-HRMS calculated mass \([M]^+ = 819.2514\), observed mass \([M]^+ = 819.2505\).

**TCO-C₃-CDP-choline (46).** Yield: 7.8 mg, 34%, white solid. \(^1\)H NMR (600 MHz, D₂O) \( \delta = 8.16 \) (d, \( J = 8.0 \) Hz, 1H), 6.27 (d, \( J = 7.9 \) Hz, 1H), 5.90 (d, \( J = 3.7 \) Hz, 1H), 5.74 – 5.47 (m, 2H), 4.36 – 4.14 (m, 8H), 3.64 (m, 2H), 3.42 – 3.36 (m, 2H), 3.17 (t, \( J = 6.6 \) Hz, 2H), 3.14 (s, 6H), 2.29 (m, 3H), 2.00 – 1.83 (m, 6H), 1.73 – 1.57 (m, 3H); \(^{13}\)C NMR (151 MHz, D₂O) \( \delta = 159.04, 158.22, 148.31, 144.07, 135.58, 133.33, 95.16, 89.65, 83.10, 74.23, 68.97, 64.49, 63.62, 63.28, 59.73, 51.41, 40.34, 37.78, 37.20, 33.64, 33.32, 32.00, 30.52, 22.73, 21.54;\(^3\)P NMR (162 MHz, D₂O) \( \delta = -11.52 \) (d, \( J = 20.7 \) Hz), -12.29 (d, \( J = 20.7 \) Hz); ESI-HRMS calculated mass \([M]^+ = 684.2405\), observed mass \([M]^+ = 684.2399\).

**Tetrazine-C₃-CDP-choline (47).** Yield: 5.9 mg, 25%, red solid. \(^1\)H NMR (600 MHz, D₂O) \( \delta = 10.35 \) (s, 1H), 8.40 (d, \( J = 8.4 \) Hz, 2H), 8.12 (d, \( J = 8.0 \) Hz, 1H), 7.59 – 7.53 (m, 2H), 6.22 (d, \( J = 7.9 \) Hz, 1H), 5.85 (d, \( J = 3.2 \) Hz, 1H), 4.37 (s, 2H), 4.32 – 4.21 (m, 4H), 4.18 (m, 1H), 3.74 (s, 2H), 3.65 (s, 2H), 3.45 – 3.37 (m, 2H), 3.33 (t, \( J = 6.5 \) Hz, 2H), 3.13 (s, 6H), 2.10 – 1.98 (m, 2H); \(^{13}\)C NMR (151 MHz, D₂O) \( \delta = 174.05, 166.27, 158.93, 157.32, 148.17, 143.99, 140.61, 130.24, 130.21, 128.55, 95.11, 89.61, 83.08, 83.02, 74.24, 68.93, 64.45, 64.42, 63.59, 63.53, 62.91, 59.69, 51.60, 42.31, 36.22, 22.15;\(^3\)P NMR (162 MHz, D₂O) \( \delta = -11.47 \) (d, \( J = 20.7 \) Hz), -12.22 (d, \( J = 20.7 \) Hz); ESI-HRMS calculated mass \([M]^+ = 730.2110\), observed mass \([M]^+ = 730.2107\).

**Chloroacetamide-C₃-CDP-choline (56).** Yield: 30 mg, 63%, white wax. \(^1\)H NMR (600 MHz, D₂O) \( \delta = 8.08 \) (d, \( J = 8.0 \) Hz, 1H), 6.19 (d, \( J = 7.9 \) Hz, 1H), 5.82 (d, \( J = 3.6 \) Hz, 1H), 4.31 – 4.19 (m, 6H), 4.12 – 4.07 (m, 1H), 4.03 (s, 2H), 3.61 – 3.55 (m, 2H), 3.38 – 3.31 (m, 2H), 3.26 (t, \( J = 6.6 \) Hz, 2H), 3.08 (s, 6H), 2.01 – 1.93 (m, 2H); \(^{13}\)C NMR (151 MHz, D₂O) \( \delta = 169.94, 159.04, 148.31, 144.05, 95.17, 89.64, 83.12, 74.24, 68.99, 64.51, 63.58, 63.02, 59.72, 59.69, 51.52, 42.19, 36.47, 22.05;\(^3\)P NMR (243 MHz, D₂O) \( \delta = -11.45 \) (d, \( J = 20.4 \) Hz), -12.23 (d, \( J = 20.4 \) Hz); ESI-HRMS calculated mass \([M]^+ = 608.1284\), observed mass \([M]^+ = 608.1338\).

**NBD-p-aminobenzyl-PEG₄-CDP-choline (31).** Yield: 11.3 mg, 36%, red solid. \(^1\)H NMR (400 MHz, D₂O) \( \delta = 8.16 \) (d, \( J = 8.0 \) Hz, 1H), 8.04 (d, \( J = 8.6 \) Hz, 1H), 7.24 (d, \( J = 8.2 \) Hz, 2H), 7.19 (d, \( J = 7.7 \) Hz, 2H), 6.44 (d, \( J = 8.4 \) Hz, 1H), 6.28 (d, \( J = 8.0 \) Hz, 1H), 5.87 (d, \( J = 2.7 \) Hz, 1H), 4.43 (t, \( J = 6.9 \) Hz, 2H), 4.32 (t, \( J = 3.8 \) Hz, 2H), 4.27 (d, \( J = 3.2 \) Hz, 2H), 4.20 (dd, \( J = 9.7, 5.2 \) Hz, 1H), 4.01 – 3.97 (m, 2H), 3.79 – 3.76 (m, 2H), 3.72 – 3.64 (m, 10H), 3.63 – 3.60 (m, 2H), 3.52 (s, 2H), 3.40 (t, \( J = 5.2 \) Hz, 2H), 2.00 – 1.83 (m, 6H), 1.73 – 1.57 (m, 3H).
\[ \text{3.26 (s, 6H); } ^{13}\text{C NMR (101 MHz, D}_2\text{O) } \delta \text{ 173.85, 159.18, 148.37, 144.60, 144.24, 143.57, 142.00, 137.86, 135.79, 133.90, 130.49, 123.10, 122.49, 115.72, 95.35, 89.76, 83.43, 83.34, 74.49, 69.85, 69.80, 69.72, 69.66, 69.24, 68.99, 65.03, 64.94, 64.63, 64.55, 64.36, 59.98, 52.52, 42.02, 39.33, 25.36; } ^{3}\text{P NMR (162 MHz, D}_2\text{O) } \delta \text{ -10.38 (d, } J = 20.8 \text{ Hz), -11.15 (d, } J = 20.8 \text{ Hz); ESI-HRMS calculated mass [M] }^+\text{ = 946.2744, observed mass [M] }^+\text{ = 946.2777.}
\]

\[ \text{7-(dimethylamino)coumarin-4-amido-PEG}_4\text{-CDP-choline (32). Yield: 17 mg, 59%, yellow solid. } ^{1}\text{H NMR (500 MHz, D}_2\text{O) } \delta \text{ 8.08 (d, } J = 7.9 \text{ Hz, 1H), 7.50 (d, } J = 9.0 \text{ Hz, 1H), 6.81 (d, } J = 9.0 \text{ Hz, 1H), 6.63 (d, } J = 2.2 \text{ Hz, 1H), 6.17 (d, } J = 7.9 \text{ Hz, 1H), 6.08 (s, 1H), 5.79 (d, } J = 2.4 \text{ Hz, 1H), 4.36 (s, 2H), 4.24 (d, } J = 8.5 \text{ Hz, 4H), 4.13 (dd, } J = 10.7, 4.7 \text{ Hz, 1H), 3.92 (s, 2H), 3.73 (d, } J = 26.8 \text{ Hz, 4H), 3.66 - 3.57 (m, 6H), 3.55 (d, } J = 7.6 \text{ Hz, 6H), 3.37 (t, } J = 4.9 \text{ Hz, 2H), 3.19 (s, 6H), 3.01 (s, 6H); } ^{13}\text{C NMR (151 MHz, D}_2\text{O) } \delta \text{ 176.65, 171.08, 165.12, 163.24, 163.10, 162.88, 159.94, 144.26, 127.36, 117.93, 115.95, 115.86, 115.72, 115.04, 95.38, 89.87, 74.46, 69.82, 69.69, 69.57, 69.19, 68.90, 66.55, 64.55, 64.33, 52.50, 45.10, 39.41, 39.28, 38.89, 37.10, 31.58, 25.36, 22.42; } ^{3}\text{P NMR (162 MHz, D}_2\text{O) } \delta \text{ -10.55 (d, } J = 21.1 \text{ Hz), -11.29 (d, } J = 21.1 \text{ Hz); ESI-HRMS calculated mass [M] }^+\text{ = 879.2937, observed mass [M] }^+\text{ = 879.2938.}
\]

\[ \text{4-pentyloic-amide-PEG}_4\text{-CDP-choline (33). Yield: 13 mg, 85%, white waxy solid. } ^{1}\text{H NMR (600 MHz, D}_2\text{O) } \delta \text{ 8.23 (d, } J = 7.9 \text{ Hz, 1H), 6.32 (d, } J = 8.0 \text{ Hz, 1H), 5.95 (d, } J = 3.6 \text{ Hz, 1H), 4.42 (m, 2H), 4.36 (m, 2H), 4.32 (m, 2H), 4.27 - 4.18 (m, 1H), 4.01 (t, } J = 4.5 \text{ Hz, 2H), 3.78 (m, 2H), 3.72 (m, 10H), 3.64 (t, } J = 5.4 \text{ Hz, 2H), 3.42 (t, } J = 5.4 \text{ Hz, 2H), 3.26 (s, 6H), 2.51 - 2.46 (m, 4H), 2.38 (t, } J = 2.5 \text{ Hz, 1H); } ^{13}\text{C NMR (151 MHz, D}_2\text{O) } \delta \text{ 174.77, 159.08, 148.40, 147.20, 144.12, 141.02, 127.40, 95.21, 89.57, 83.33, 83.25, 83.19, 74.30, 70.10, 69.62, 69.58, 69.49, 69.42, 69.02, 68.83, 64.35, 64.14, 59.76, 59.73, 52.29, 42.89, 38.90, 34.37, 14.51; } ^{3}\text{P NMR (243 MHz, D}_2\text{O) } \delta \text{ -11.38 (d, } J = 25.8 \text{ Hz), -12.18 (d, } J = 25.8 \text{ Hz); ESI-MS calculated mass [M] }^+\text{ = 730.2460, observed mass [M] }^+\text{ = 730.2559.}
\]

\[ \text{biotin-PEG}_4\text{-CDP-choline (36). Yield: 4 mg, 50%, white solid. } ^{1}\text{H NMR (600 MHz, D}_2\text{O) } \delta \text{ 8.23 (d, } J = 8.0 \text{ Hz, 1H), 6.32 (d, } J = 8.0 \text{ Hz, 1H), 5.95 (d, } J = 3.1 \text{ Hz, 1H), 4.61 (m, 2H), 4.52 - 4.23 (m, 6H), 4.20 (d, } J = 12.5 \text{ Hz, 1H), 4.00 (m, 2H), 3.88 - 3.61 (m, 10H), 3.61 (m, 2H), 3.39 (t, } J = 5.4 \text{ Hz, 2H), 3.26 (s, 6H), 3.04 - 2.96 (m, 1H), 2.78 (m, 2H), 2.28 (t, } J = 7.2 \text{ Hz, 2H), 1.65 (m, 4H), 1.42 (q, } J = 7.7 \text{ Hz, 2H), 1.29 (m, 2H); } ^{13}\text{C NMR (151 MHz, D}_2\text{O) } \delta \text{ 113.61, 99.98, 95.24, 89.57, 83.25, 74.31, 69.63, 69.59, 69.50, 69.40, 69.03, 68.84, 64.36, 64.15, 62.06, 60.24, 59.72, 55.33, 54.34, 52.30, 42.52, 39.66, 38.84, 35.42, 33.13, 27.82, 27.66, 25.10, 17.68, 16.21, 12.10; } ^{3}\text{P NMR}
\]
(162 MHz, D₂O) δ -11.39 (d, J = 21.1 Hz), -12.20 (d, J = 21.1 Hz); ESI-HRMS calculated mass [M]+ = 876.2974, observed mass [M]+ = 876.3063.

**DBCO-PEG₄-CDP-choline (53).** Yield: 9.7 mg, 59 %, white solid. ¹H NMR (600 MHz, D₂O) δ 8.13 (d, J = 8.0, 3.0 Hz, 1H), 7.58 (d, J = 7.6 Hz, 1H), 7.51–7.29 (m, 6H), 7.25 (d, J = 7.4, 1.4 Hz, 1H), 6.24 (d, J = 7.9, 2.4 Hz, 1H), 5.91 – 5.82 (m, 1H), 5.02 – 4.96 (m, 1H), 4.38 (s, 2H), 4.32 – 4.22 (m, 4H), 4.20 – 4.12 (m, 1H), 3.92 (d, J = 5.1 Hz, 2H), 3.72 (t, J = 4.6 Hz, 2H), 3.68 (d, J = 4.1 Hz, 1H), 3.64 (m, 6H), 3.58 (m, 2H), 3.55 – 3.49 (m, 2H), 3.44 – 3.37 (m, 1H), 3.33 (m,1H), 3.20 (s, 6H), 3.17 – 3.07 (m, 2H), 2.48 (m, 1H), 2.18 (m, 2H); ¹³C NMR (151 MHz, D₂O) δ 174.28, 174.22, 158.97, 150.51, 148.22, 147.75, 144.02, 131.80, 129.11, 129.04, 128.88, 128.40, 127.06, 127.06, 125.70, 122.34, 121.64, 114.19, 107.71, 95.15, 89.65, 83.00, 74.22, 69.56, 69.54, 69.44, 69.35, 68.96, 68.56, 64.74, 64.53, 64.32, 64.09, 59.84, 55.49, 52.26, 38.82, 30.89, 30.11; ³¹P NMR (162 MHz, D₂O) δ -11.51 (d, J = 20.4 Hz), -12.32 (d, J = 20.4 Hz) ; ESI-HRMS calculated mass [M]+ = 937.3144, observed mass [M]+ = 937.3136.

**TCO-PEG₄-CDP-choline (54).** Yield: 6.8 mg, 48 %, white solid. ¹H NMR (600 MHz, D₂O) δ 8.19 (d, J = 8.0 Hz, 1H), 6.30 (d, J = 8.0 Hz, 1H), 5.92 (d, J = 3.6 Hz, 1H), 5.67 (m, 1H), 4.40 (s, 3H), 4.36 – 4.27 (m, 3H), 3.75 (t, J = 4.5 Hz, 3H), 3.71 – 3.63 (m, 12H), 3.56 (t, J = 5.3 Hz, 2H), 3.27 (t, J = 5.4 Hz, 2H), 3.24 (s, 6H), 2.34 – 2.29 (m, 3H), 1.98 – 1.81 (m, 1H), 1.61 (m, 4H); ¹³C NMR (151 MHz, D₂O) δ 163.14, 162.91, 162.67, 159.06, 158.32, 148.34, 144.10, 135.60, 133.34, 95.19, 89.62, 83.17, 83.11, 74.27, 69.59, 69.48, 69.36, 69.00, 64.71, 64.47, 64.35, 64.12, 59.82, 59.78, 52.29, 40.38, 37.80, 33.66, 32.02, 30.56; ³¹P NMR (162 MHz, D₂O) δ -11.47 (d, J = 20.9 Hz), -12.28 (d, J = 20.7 Hz) ; ESI-HRMS calculated mass [M]+ = 802.3035, observed mass [M]+ = 802.2034.

**Tetrazine-PEG₄-CDP-choline (55).** Yield: 5.1 mg, 35 %, red solid. ¹H NMR (600 MHz, D₂O) δ 10.38 (s, 1H), 8.45 (d, J = 8.3 Hz, 2H), 8.18 (d, J = 8.0 Hz, 1H), 7.61 (d, J = 8.1 Hz, 2H), 6.27 (d, J = 7.9 Hz, 1H), 5.90 (d, J = 3.4 Hz, 1H), 4.40 (s, 2H), 4.33 – 4.28 (m, 4H), 3.97 (s, 2H), 3.76 (s, 2H), 3.69 – 3.62 (m, 14H), 3.43 (t, J = 5.3 Hz, 2H), 3.24 (s, 6H); ¹³C NMR (151 MHz, D₂O) δ 173.89, 166.37, 163.00, 162.77, 158.99, 157.35, 148.28, 144.06, 140.66, 130.22, 128.57, 99.98, 95.17, 89.55, 83.18, 83.12, 74.27, 69.64, 69.55, 69.47, 69.39, 68.98, 68.73, 64.76, 64.33, 64.11, 59.79, 59.75, 52.28, 42.35, 39.14; ³¹P NMR (162 MHz, D₂O) δ -11.43 (d, J = 21.0 Hz), -12.22 (d, J = 21.4 Hz); ESI-HRMS calculated mass [M]+ = 848.2740, observed mass [M]+ = 848.2730.
Di-trifluoroacetate-di-boc-spermine (61). Spermine 60 (10 g, 49.4 mmol) was dissolved in MeCN (400 mL) and ethyl trifluoroacetate (14 g, 11.76 mL, 98.8 mmol, 2 eq.) was added drop wise. After complete addition the solution was stirred for 3 h at room temperature and the reaction was cooled to 0°C. Boc₂O (21.8 g, 100 mmol, 2 eq.) was added together with Et₃N (15.3 g, 21 mL, 150 mmol, 3 eq.). The reaction was allowed to warm to room temperature and stirred overnight. Subsequently the mixture was evaporated to dryness, taken up into EtOAc (100 mL) and washed with water (3x30 mL). The aqueous fractions were re-extracted with EtOAc (3x100 mL) and the combined organic layers were dried over Na₂SO₄, filtered and concentrated under reduced pressure to yield in fully protected Spermine 61 (29.6 g 49 mmol, 99%).

**1H NMR** (400 MHz, MeOD) δ 3.27 (m, 12H), 1.81 (m, 4H), 1.54 (m, 4H), 1.48 (s, 18H); **13C NMR** (100 MHz, MeOD) δ 157.72, 156.09, 116.12 (q, J = 286.7 Hz), 79.72, 44.69, 44.14, 37.12, 27.31, 26.16, 25.23; ESI-HRMS: calculated mass [M+Na]⁺ = 617.2744, observed mass [M+Na]⁺ = 617.2791.

Cbz-di-boc-spermine (62). Fully protected spermine 61 (29 g 49 mmol) was dissolved in MeOH and NH₄OH (300 mL, 1:1). The resulting suspension was heated to 60°C and stirred for 4 h during which NH₄OH was added several times until the solution stayed clear after further addition of NH₄OH. The solvents were evaporated and the residue was dissolved in MeOH (700 mL). Et₃N (14.5 g, 20 mL, 143 mmol, 3 eq.) was added. The solution was cooled to 0°C and Cbz-Cl (8.6 g, 7.2 mL, 50 mmol, 1 eq.) was added drop wise to the stirred reaction. After complete addition the reaction was stirred at room temperature over night. The solvents were removed under reduced pressure and the residue was taken up in DCM and washed with saturated NaHCO₃, followed by Brine and dried over Na₂SO₄. After filtration and evaporation, the crude material was purified on silica (DCM / MeOH (containing 10% NH₄OH (aq)) 95:5 to 9:1) and yielded in Cbz-protected spermine 62 (14 g, 26 mmol, 52.2%).

**1H NMR** (400 MHz, CDCl₃) δ 7.39 – 7.29 (m, 5H), 5.10 (s, 2H), 4.54 (m, 4H), 3.36 – 3.09 (m, 10H), 2.90 (m, 2H), 1.78 (m, 4H), 1.45 (s, 18H); **13C NMR** (100 MHz, CDCl₃) δ 156.85, 156.54, 155.60, 136.81, 128.46, 128.04, 80.51, 76.73, 66.46, 47.00, 46.76, 46.38, 44.34, 43.56, 43.15, 37.70, 28.43, 28.41, 25.84, 25.45; ESI-HRMS calculated mass [M+Na]⁺ = 559.3472, observed mass [M+Na]⁺ = 559.3443.

Cbz-di-boc-dimethyl-spermine (63). Polyamine 62 (14 g, 26 mmol) was dissolved in EtOH (150 mL). Formaldehyde (38 mL, 37%, 42.16 g, 520 mmol, 20 eq.) and acetic acid (28.8 mL, 31.25 g, 520 mmol, 20 eq.) were added and the solution was cooled with an ice bath. NaBH₃CN (8.16 g, 130 mmol, 5 eq.) was added in portions and the reaction was stirred over night at room temperature. After complete consumption of the starting
material (checked by TLC: DCM/MeOH (10% NH₄OH) 95:5 and LC/MS) the reaction was quenched with NaOH (450 mL, aq, 5M) and evaporated to dryness. The residue was taken up in DCM, washed with saturated NaHCO₃ and Brine, dried over Na₂SO₄, filtered and evaporated to give polyamine 63 (13.4 g, 23.7 mmol, 91.2%). ¹H NMR (400 MHz, MeOD) δ 7.42 – 7.24 (m, 5H), 5.09 (s, 2H), 3.24 (m, 8H), 3.13 (t, J = 6.8 Hz, 2H), 2.46 (s, 2H), 2.32 (s, 6H), 1.84 – 1.68 (m, 4H), 1.53 (m, 4H), 1.48 (s, 9H), 1.47 (s, 9H); ¹³C NMR (100 MHz, MeOD) δ 157.40, 156.06, 155.99, 137.03, 128.06, 127.57, 127.41, 79.59, 79.55, 65.95, 63.60, 53.42, 46.48, 45.12, 44.69, 44.29, 43.87, 38.00, 28.85, 28.17, 27.38, 27.36, 25.45; ESI-HRMS calculated mass [M+Na]⁺ = 587.3785, observed mass [M+Na]⁺ = 587.3825.

Cbz-di-boc-spermine choline (64). Polyamine spacer 63 (13.4 g, 23.7 mmol) was dissolved in Et₂O (500 mL) and stirred under argon atmosphere at room temperature. Iodoethanol (8.14 g, 3.66 mL, 47.4 mmol, 2eq.) was added and the reaction was stirred in the dark for 16 h. Another aliquot of iodoethanol (4 g, 23 mmol, 1 eq.) was added and stirred for 8 h until all starting material 63 was consumed (checked by LC/MS). Addition of Et₂O to the reaction crushed out the product as oily solid. The ether was decanted and the residue was taken up in water and washed with Et₂O. The aqueous layer was concentrated and the residue was purified on silica (EtOAc:MeOH:AcOH:H₂O (6:3:3:2 v/v)) to yield alkylated polyamine 64 (9.5 g, 12.4 mmol, 52.3%). ¹H NMR (400 MHz, CDCl₃) δ 7.45 – 7.30 (m, 5H), 5.10 (s, 2H), 4.14 (t, J = 4.3 Hz, 2H), 3.87 – 3.67 (m, 2H), 3.65 – 3.50 (m, 2H), 3.39 – 3.10 (m, 16H), 2.09 (t, J = 8.5 Hz, 2H), 1.71 (m, 2H), 1.46 (s, 9H), 1.45 (s, 9H); ¹³C NMR (100 MHz, CDCl₃) δ 128.50, 128.44, 128.10, 128.01, 80.12, 79.69, 79.66, 76.73, 72.69, 66.63, 66.56, 66.46, 66.37, 65.71, 63.92, 63.86, 63.84, 55.90, 52.60, 46.88, 44.16, 37.80, 28.47, 26.02, 22.42; ESI-MS calculated mass [M]+ = 609.4, observed mass [M]+ = 609.4.

Cbz-di-methyl-spermine choline (65). The Boc-protected polyamine 64 was stirred in DCM/TFA (50:50) for 30 minutes and evaporated to dryness to give the polyamine as 3x TFA salt. The deprotected polyamine (4.5 g, 5.12 mmol) was dissolved in EtOH (40 mL), formaldehyde (7.5 mL, 37%, 8.3 g, 103 mmol, 20 eq.) and acetic acid (5.7 mL, 16.6 g, 103 mmol, 20 eq.). The solution was cooled with an ice bath and NaBH₄CN (1.6 g, 25 mmol, 5eq.) was added in portions and stirred for 15 minutes. The ice bath was removed and the reaction was stirred over night. After complete conversion, the reaction was quenched with NaOH (120 mL, aq, 5M) and evaporated to dryness. The residue was taken up in water and the basic (pH 14) aqueous solution was extracted with DCM. The organic layer was washed with brine, dried over Na₂SO₄, filtered and evaporated to yield methylated polyamine 65 (1.96 g, 4.32 mmol, 84%). ¹H NMR (400 MHz,
MeOD) δ 7.46 – 7.29 (m, 5H), 5.11 (s, 2H), 4.05 (m, 2H), 3.66 – 3.52 (m, 4H), 3.49 – 3.28 (m, 6H), 3.27 (s, 6H), 3.25 – 3.09 (m, 4H), 2.97 (s, 3H), 2.89 (s, 3H), 2.39 (p, J = 8.2 Hz, 2H), 2.02 – 1.95 (m, 2H), 1.90 (m, 4H); 13C NMR (100 MHz, MeOD) δ 157.70, 136.93, 128.15, 127.69, 127.43, 66.74, 66.20, 65.55, 61.48, 55.46, 55.11, 54.99, 53.76, 52.17, 51.30, 39.49, 39.17, 37.30, 24.56, 20.99, 20.96, 18.08; ESI-MS calculated mass [M]+ = 437.3, observed mass [M]+ = 437.4.

Cbz-di-methyl-spermine phosphocholine (66). POCl₃ (1.75 g, 11.4 mmol, 10 eq.) and 2,6-lutidine (3.67 g, 33 mmol, 30 eq.) were dissolved in dry MeCN (10 mL) and cooled to 0°C with an ice bath. The solution was stirred under argon and spermine spacer 65 (500 mg, 1.14 mmol) in MeCN (25 mL) was added. The reaction was stirred for 20 minutes at 0°C before it was quenched with water (200 mL). After extraction with Et₂O (3×50 mL) the aqueous layer was evaporated to dryness. The residue was redissolved in water and the solution was basified (pH 14) with Ca(OH)₂ to precipitate inorganic phosphates. The suspension was filtered and the filtrate was evaporated to give the phosphorylated polyamine 66 (491 mg, 0.95 mmol, 83%). ¹H NMR (400 MHz, D₂O) δ 7.27 (m, 5H), 4.96 (s, 2H), 3.99 (m, 2H), 3.47 – 3.40 (m, 2H), 3.30 – 3.22 (m, 2H), 3.05 (t, J = 6.4 Hz, 2H), 3.02 (s, 6H), 2.83 (t, J = 7.9 Hz, 4H), 2.54 (s, 3H), 2.48 (m, 4H), 2.24 (s, 3H), 1.91 (m, 2H), 1.71 (m, 2H), 1.54 – 1.38 (m, 4H); ¹³C NMR (100 MHz, D₂O) δ 158.23, 136.53, 128.78, 128.36, 127.58, 115.59, 99.98, 66.90, 62.80, 57.71, 55.64, 55.63, 53.32, 52.37, 51.61, 40.46, 39.71, 37.50, 24.23, 22.49, 21.84, 18.95; ³¹P NMR (162 MHz, D₂O) δ 3.39; ESI-MS calculated mass [M]+ = 517.3, observed mass [M]+ = 517.3.

Phosphocholine-di-methyl-spermine pentynamide (67). Pd(OH)₂ on carbon (10 mg) was suspended in MeOH (2 mL) under an argon atmosphere. Subsequently Cbz-protected Spacer 66 (100 mg, 193 µmol) dissolved in MeOH (1 mL) and acetic acid (1 mL) was given to the stirring reaction mixture and flushed under hydrogen. The mixture was stirred for 1.5 h until LC/MS showed complete conversion. The reaction was filtered over celite in a sintered funnel to remove the palladium and the celite was washed with several portions of methanol. After concentration under reduced pressure the residue was dissolved in sat. aq. NaHCO₃ (1 mL). 4-pentynoic acid NHS-ester (36 mg, 193 µmol) dissolved in MeCN (200 µL) was added and the reaction was stirred for 40 minutes at room temperature until LC/MS showed complete conversion. The solvents were evaporated and the residue was purified by HPLC to obtain the title compound 67 (44 mg, 94.9 µmol, 66%). ¹H NMR (400 MHz, D₂O) δ 4.17 – 4.08 (m, 2H), 3.61 – 3.53 (m, 2H), 3.43 – 3.36 (m, 2H), 3.26 (t, J = 6.6 Hz, 2H), 3.16 (s, 6H), 2.94 – 2.75 (m, 4H), 2.66 – 2.37 (m, 8H), 2.56 (s, 3H), 2.29 (s, 3H), 2.01 (qd, J =
11.1, 10.0, 6.2 Hz, 2H), 1.90 – 1.77 (m, 2H), 1.68 – 1.46 (m, 4H); $^{13}$C NMR (100 MHz, D2O) δ 174.80, 83.37, 70.23, 64.36, 63.30, 57.68, 55.90, 55.82, 53.48, 52.47, 51.42, 48.82, 40.60, 40.05, 36.64, 34.43, 24.28, 23.05, 22.48, 19.07, 14.57; $^{31}$P NMR (162 MHz, D2O) δ 3.39; ESI-MS calculated mass [M]+ = 463.3, observed mass [M]+ = 463.4.

CDP-choline di-methyl spermine pentynamide (68). Phosphorylated spacer 67 (44 mg, 95 µmol) and cytidin-5'-phosphomorpholidate 4-morpholine-N,N'-dicyclohexylcarboxamidium (156 mg, 228 µmol, 2.4 eq.) were suspended in benzene (2 mL), frozen in liquid nitrogen and lyophilised to remove traces of water. The dried residue was dissolved in dry pyridine (2 mL) under an argon atmosphere. 1H-tetrazole (16 mg, 228 µmol, 2.4 eq.), evaporated from acetonitrile (0.5 mL, 0.45 M), was dissolved in dry pyridine (1 mL) under an argon atmosphere and subsequently dropped to the stirring reaction mixture. The reaction was heated to 60°C and stirred under argon overnight. The solvent was evaporated under reduced pressure and the residue was purified by HPLC to yield CDP-choline analogue 68 (49.5 µmol, 52.1%). $^1$H NMR (400 MHz, D2O) δ 8.07 (d, $J = 7.9$ Hz, 1H), 6.19 (d, $J = 7.9$ Hz, 1H), 5.82 (d, $J = 3.6$ Hz, 1H), 4.30 (m, 2H), 4.26 – 4.18 (m, 4H), 4.14 – 4.05 (m, 1H), 3.65 (m, 2H), 3.43 (m, 2H), 3.27 – 3.14 (m, 5H), 3.13 (s, 6H), 3.12 – 2.95 (m, 5H), 2.81 (s, 3H), 2.75 (s, 3H), 2.44 – 2.31 (m, 4H), 2.31 – 2.11 (m, 3H), 1.90 – 1.79 (m, 2H), 1.71 (m, 4H); $^{13}$C NMR (100 MHz, D2O) δ 175.12, 159.10, 148.38, 144.02, 95.17, 89.74, 83.43, 83.09, 83.00, 74.31, 74.20, 70.18, 68.99, 64.52, 63.55, 60.87, 59.66, 55.37, 55.22, 53.65, 52.29, 52.12, 43.10, 39.51, 35.92, 34.36, 23.69, 20.79, 18.14, 14.56; $^{31}$P NMR (162 MHz, D2O) δ -11.37 (d, $J = 20.7$ Hz), -12.23 (d, $J = 21.2$ Hz); ESI-HRMS calculated mass [M]+ = 768.3457, observed mass [M]+ = 768.3521.