



Cite this: *RSC Chem. Biol.*, 2021, **2**, 1441

Segmental and site-specific isotope labelling strategies for structural analysis of posttranslationally modified proteins

Dominik P. Vogl,^a Anne C. Conibear ^b and Christian F. W. Becker ^{a*}

Posttranslational modifications can alter protein structures, functions and locations, and are important cellular regulatory and signalling mechanisms. Spectroscopic techniques such as nuclear magnetic resonance, infrared and Raman spectroscopy, as well as small-angle scattering, can provide insights into the structural and dynamic effects of protein posttranslational modifications and their impact on interactions with binding partners. However, heterogeneity of modified proteins from natural sources and spectral complexity often hinder analyses, especially for large proteins and macromolecular assemblies. Selective labelling of proteins with stable isotopes can greatly simplify spectra, as one can focus on labelled residues or segments of interest. Employing chemical biology tools for modifying and isotopically labelling proteins with atomic precision provides access to unique protein samples for structural biology and spectroscopy. Here, we review site-specific and segmental isotope labelling methods that are employed in combination with chemical and enzymatic tools to access posttranslationally modified proteins. We discuss illustrative examples in which these methods have been used to facilitate spectroscopic studies of posttranslationally modified proteins, providing new insights into biology.

Received 7th March 2021,
Accepted 11th August 2021

DOI: 10.1039/d1cb00045d

rsc.li/rsc-chembio

^a University of Vienna, Faculty of Chemistry, Institute of Biological Chemistry, Währinger Straße 38, 1090 Vienna, Austria. E-mail: christian.becker@univie.ac.at; Fax: +43-1-4277-870510; Tel: +43-1-4277-70510

^b The University of Queensland, School of Biomedical Sciences, St Lucia, Brisbane, 4072 QLD, Australia



Dominik P. Vogl

Dominik P. Vogl obtained his BSc degree in Biology from the University of Regensburg in 2015 and then moved to the Technische Universität München (TUM) to study Molecular Biotechnology. After two industrial placements at Vaximm (2016) and Hoffmann-La Roche (2017/2018), he carried out his Master's Thesis work at the Division of Peptide Biochemistry under the supervision of Prof. Aphrodite Kapurniotu at TUM.

After receiving his MSc, he joined

the Institute of Biological Chemistry at the University of Vienna for his PhD studies in 2019 under the supervision of Prof. Christian F. W. Becker and Dr Anne C. Conibear (The University of Queensland). His research interests encompass protein semisynthesis, segmental isotope labelling and NMR spectroscopy to study posttranslationally modified biomolecules at atomic resolution.



Anne C. Conibear

Anne C. Conibear completed her BSc (Hons) and MSc in Chemistry at Rhodes University, South Africa. She then moved to The University of Queensland, Australia for her PhD (2014), focusing on the structural characterisation and applications of cyclic disulfide-rich peptides from mammals, the theta-defensins. In 2014, she was awarded an Interdisciplinary Cancer Research (INDICAR) postdoctoral fellowship at the University of Vienna and worked on targeted immune-stimulating molecules for cancer therapy. She returned to The University of Queensland in 2019 with a UQ Development Fellowship to start an independent project aiming to understand how posttranslational modifications modulate the structure and activity of proteins, using predominantly the tools of protein chemistry and NMR spectroscopy.



Introduction

The number of protein variants (proteoforms) in eukaryotes often vastly exceeds the number of protein-coding genes. The main source of this proteoform diversity is posttranslational modification (PTM) of proteins, which can occur on backbone and side-chain functional groups and include addition of functional groups, cyclization, splicing and cleavage.¹ These PTMs are introduced either enzymatically or non-enzymatically and functional groups added vary considerably in complexity and size.² Nearly all eukaryotic proteins are posttranslationally modified during their lifetime, with diverse effects on protein structure, dynamics and function. The large number of PTMs, their chemical diversity and overall heterogeneity makes the study of their individual effects challenging but also essential to understanding the molecular mechanisms that control life.^{3–5}

Understanding the structural effects of PTMs and how these influence protein function, dynamics and interaction with binding partners is enabled by a variety of structural biology techniques. Spectroscopic techniques such as Nuclear Magnetic Resonance (NMR), infrared (IR) and Raman spectroscopy allow for the study of protein structure at atomic resolution and can provide insights into protein dynamics even in highly flexible regions. These spectroscopic techniques combined with X-ray crystallography, which has been most extensively used for protein structure determination, are now complemented by cryo-electron microscopy, which has revolutionised structural biology by enabling structural studies of large complexes at increasing resolution.⁶ Integrated structural biology approaches take advantage of the strengths of multiple techniques and show great promise for elucidating the mechanisms of protein regulation by PTMs.



Christian F. W. Becker

at the Technische Universität München in 2007. In 2011, he became Professor and Head of the Institute of Biological Chemistry at the University of Vienna and in 2020, founding head of the Vienna Doctoral School in Chemistry (DoSChem). His group develops and uses chemical as well as biochemical means to generate peptides and proteins with otherwise unattainable (posttranslational) modifications to address biomedical and biotechnological challenges.

Spectroscopic techniques applied to understanding the effects of PTMs on protein structures and mechanisms each have inherent advantages and limitations. Biomolecular NMR spectroscopy, for example, can provide information on conformational preferences and changes in flexible regions that are not visible in rigid crystal structures.⁷ PTMs can cause local alterations in the chemical environments of protein nuclei, resulting in characteristic chemical shift changes that can be used to identify the PTM type and site of modification if residue assignments are available.⁸ Solution NMR also offers the possibility to monitor enzymatic introduction of PTMs and their subsequent removal in a quantitative and non-disruptive manner under native conditions.⁹ However, protein size can be a limitation for solution NMR due to long correlation times and spectral complexity for large proteins. Combining NMR with scattering techniques such as SAXS (small-angle X-ray scattering) or SANS (small-angle neutron scattering) opens up the possibility of studying PTM-induced conformational changes even in macromolecular complexes or multidomain proteins.¹⁰ Vibrational spectroscopy, such as IR and Raman, can deliver insights into covalent bonds in a molecule based on vibrational energy transitions and therefore can report on PTM-induced structural changes (IR) or can provide valuable information about the overall secondary structure (Raman).^{11,12} Other spectroscopic techniques frequently applied to studying proteins (and the effects of PTMs) such as UV/vis- and fluorescence spectroscopy are beyond the scope of this article, as they do not rely on specific isotope labelling patterns.¹³ Mass spectrometry-based proteomics strategies also benefit greatly from isotope labels in certain quantitative applications such as SILAC (stable isotope labelling by/with amino acids in cell culture), however labelling is typically not site-selective or segmental in these applications and therefore not covered here. Please refer to recent reviews on mass spectrometry-based proteomics for further details.¹⁴

Here we focus on spectroscopic applications that require or greatly benefit from the incorporation of specific stable isotopes at non-natural abundance – namely NMR, IR, Raman and small-angle scattering (SAS) technologies to study defined PTMs in proteins. We provide an overview of the current chemical and enzymatic toolbox for segmental and site-specific isotope labelling strategies for proteins. After summarizing the available methodologies, we highlight their strengths and weaknesses using selected examples of proteins carrying a variety of PTMs.

Isotope labelling of proteins for spectroscopy

Biomolecular NMR spectroscopy relies on the spin-1/2 of ¹H, ¹³C and ¹⁵N nuclei.¹⁵ Homonuclear ¹H-NMR takes advantage of the 99.98% natural abundance of ¹H but suffers from signal overlap for proteins larger than ~50 residues because of the large number of protons in biological molecules and relatively narrow range of proton resonances.¹⁶ Multidimensional



spectra that correlate protons with nitrogen and carbon in proteins can reduce the number of signals and increase resolution, but are limited by the low natural abundance of NMR-active carbon and nitrogen isotopes (1.1% for ^{13}C and 0.366% for ^{15}N).¹⁷ Larger proteins can be deuterated in order to reduce disadvantageous dipolar interactions and adverse spin diffusion.¹⁸ For example, combining ^2H - ^{13}C - ^{15}N isotope labelling of proteins with transverse relaxation optimized spectroscopy (TROSY), an NMR experiment that reduces the effective transverse relaxation and thereby decreases linewidth and increases sensitivity of signals from large molecules with long correlation times,¹⁹ larger proteins as well as complexes typically up to 100 kDa in molar mass can be studied by NMR. In certain cases, using deuteration in combination with methyl group labelling, protein assemblies of up to 1 MDa have been successfully studied by NMR.²⁰ Alternative NMR-active nuclei that occur at 100% natural abundance, such as ^{19}F , are seeing an increasing use in NMR applications (e.g. in screening approaches), due to their sensitivity and uncrowded spectra.²¹ However, as they seldom occur naturally in proteins, we do not discuss incorporation of ^{19}F in more detail here.

Small angle scattering (SAS) and vibrational spectroscopies also rely on isotope enrichment to deduce specific information about regions of large molecules.²² SAS methods take advantage of solvent contrast variation of perdeuterated proteins by using buffers supplemented with specific $\text{D}_2\text{O}:\text{H}_2\text{O}$ ratios.²³ The resulting differences in scattering length densities between solvent and biomolecule allow the individual rendering of each (complex) part and its structural characterisation.¹⁰ For IR and Raman spectroscopies, replacement of specific bonded atoms with their isotopes allows assignment of the band arising from that specific bond in otherwise highly complex spectra. The assignment of the vibrational mode of a specific chemical group or amino acid residue enriched in ^{13}C or ^{15}N can generate valuable spatial dimensionality in the context of protein IR spectroscopy.²⁴ Similarly, isotope enrichment enables the assignment of vibrational frequencies in Raman spectra to specific residues and therefore can contribute to the determination of the secondary structure or folding state of a protein.²⁵

In order to enrich an isotope in the protein of interest, several approaches based on providing the expression system with suitable sources of the required precursor molecules have been developed, yielding proteins that differ only in their isotopic composition (isotopologues).²⁶ Typically, isotope labelling of proteins is achieved by recombinant production in a suitable expression system – typically *E. coli*, but also yeast, insect or mammalian cells.^{26,27} For the recombinant production of toxic or membrane proteins, cell-free expression can be employed.^{28,29} There are numerous experimental variations depending on the expression system and target protein, but typically minimal media for expression in bacteria are supplemented with ^{13}C -glucose and/or $^{15}\text{NH}_4\text{Cl}$ as sole carbon and nitrogen sources for the incorporation of stable carbon and nitrogen isotopes.²⁶ An alternative is Studier's glycerol-based autoinduction media.³⁰ In uniform labelling schemes, nearly all atoms of a specific element in a protein (^{12}C , ^{14}N) are replaced

with their respective isotopes (^{13}C , ^{15}N), typically leading to an isotope enrichment of more than 90% in the target protein (Fig. 1a).²⁶ Deuteration (typically for larger proteins >25 kDa) can be achieved by replacing water with D_2O and additionally by growing the expressing cells in media supplemented with ^2H - ^{13}C -labelled glucose.^{31,32} However, deuteration is only applicable in prokaryotic cells (*E. coli* and *P. Pastoris*), as eukaryotes suffer from toxic isotope effects. As an alternative, one can use deuterated amino acid hydrolysates in combination with cell-free expression systems. Cell-free expression systems are particularly useful for amino-acid selective labelling,³³ labelling of specific methyl groups,³⁴ and stereo-array isotope labelling (SAIL).³⁵ However, signal overlap in NMR spectra of large proteins can still prove challenging. Likewise, interpretation of IR and Raman spectra as well as SAS data is complicated by the number of signals arising from large proteins.^{11,22,25} In order to reduce spectral crowding and focus on particular regions of interest, it is sometimes necessary to apply more elaborate, tailored spectral simplification strategies, which will be discussed in more detail in the following sections.¹⁷

Selection of a specific isotope labelling strategy depends on the desired information, protein size and means of production. A reduction in the number of signals observed, for example in NMR spectra, can be achieved by either labelling specific groups (e.g. methyl groups), selected segments of a protein, or single amino acid residues at essential functional sites (Fig. 1). Such labelling results in more intense signals at the labelling site compared to their natural-abundance counterparts and therefore allows information on specific parts of the target protein to be obtained.

Residue-specific isotope labelling is achieved by supplying the expression host (e.g. *E. coli* grown on minimal media) with the respective labelled amino acid or its metabolic precursors (Fig. 1b). Labelled precursors for the desired amino acid are introduced into suitable amino acid synthesis pathways, resulting in labelling of all amino acid residues diverging from that point in the pathway.²⁶ Methyl ($^{-13}\text{C}^1\text{H}_3$) groups such as in valine, leucine, methionine, etc. are ideal candidates for the residue-specific incorporation of NMR-active isotopes.³⁶ Methyl groups carry three protons and their symmetric arrangement yields intense and well-resolved NMR signals.³⁷ Furthermore, the abundance of methyl functionalities in hydrophobic protein core regions and interaction sites makes them an attractive target for substitution with the respective isotopologues. Alternatively, Takeuchi *et al.* have demonstrated the usefulness of amino acids ^{13}C -labelled at the carbonyl position against a ^2H - ^{15}N background, which allows easy identification of these amino acids and shows much less scrambling compared to ^{15}N labelling.³⁸ Methyl-specific labelling takes advantage of the high sensitivity of the ^1H nucleus at certain key sites in ^1H - ^{13}C correlation spectra.³⁹ The original ^{13}C -labelling strategy for Ile, Leu and Val methyl groups was established by the Kay laboratory⁴⁰ and developed further by the Boisbouvier group to allow differentiation between the respective methyl groups in NMR spectra by adding stereo-specifically labelled methyl groups.⁴¹ Other labelled biosynthetic precursors such as precursors for labelling the indole in Trp have been explored and are valuable tools for residue-specific isotope labelling.⁴² However,



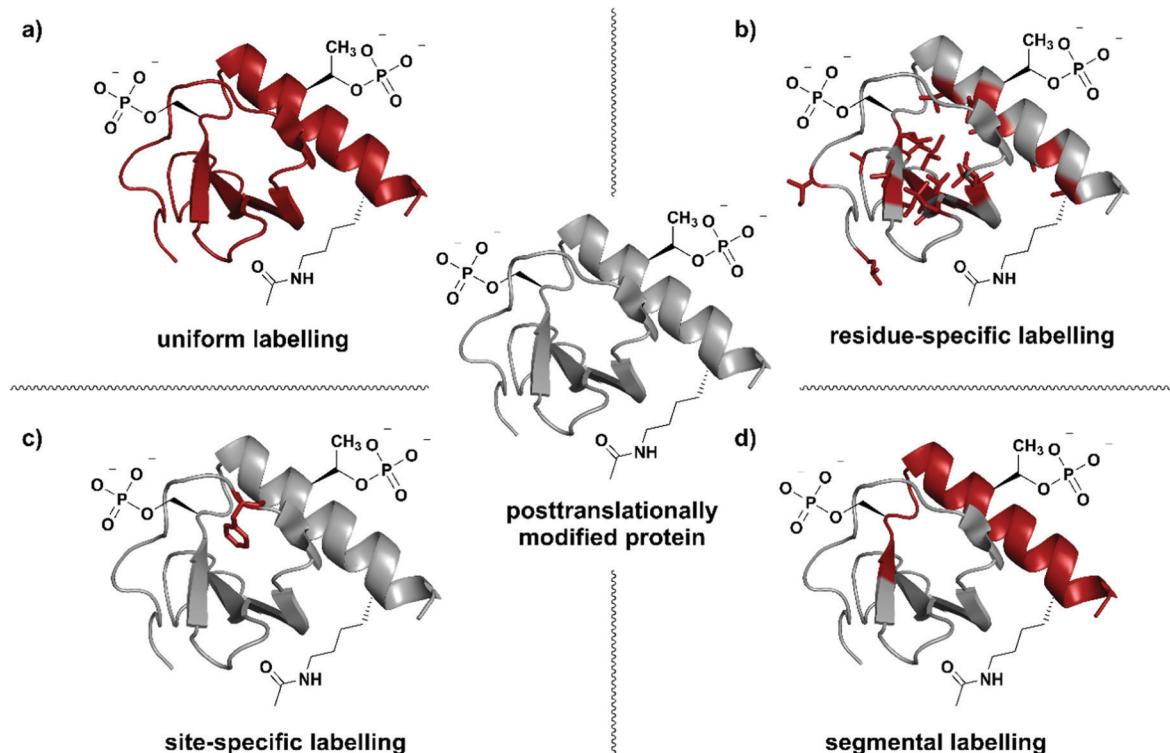


Fig. 1 Isotope labelling strategies for posttranslationally modified proteins. (a) Uniform isotope labelling of a target protein in which either no or all protons, carbons and/or nitrogens are replaced by ^2H , ^{13}C or ^{15}N (shown in red). (b) Residue-specific isotope labelling in which particular amino acid residues (e.g. Leu, Tyr, Met) are replaced by their isotope labelled counterparts, whereas the other residues remain unlabelled. (c) Site-specific labelling is the incorporation of one or more isotope labelled amino acid residues at certain key position(s) in an otherwise unlabelled polypeptide chain. (d) Segmental isotope labelling enables the incorporation of an isotope-labelled protein segment (peptide) into a larger protein. | The structure of Interleukin 8 (PDB: 3IL8) is used to illustrate the different labelling schemes. Protein parts with natural isotope abundance are shown in grey and isotope-enriched residues/segments in red.

metabolic labelling processes used for residue specific labelling might cause potential scrambling of isotope labels due to metabolic interconversion of the amino acids.^{43,44}

Labelling of single amino acid residues is a potential strategy when focusing on one or very few key positions in a protein. This strategy allows very simplified signal assignments and easy monitoring of reactions involving the labelled key residue(s), but comes with the disadvantage of ignoring the larger protein context (Fig. 1c). In this approach, isotope-labelled natural or unnatural amino acids are introduced site-specifically at the desired positions by protein (semi-)synthesis or genetic code expansion.^{45,46}

Segmental isotope labelling is a strategy for reduction of spectral complexity that is applied to small and large proteins (Fig. 1d). This labelling strategy became possible with the introduction of expressed protein ligation (EPL) and protein *trans*-splicing (PTS), as well as by the use of enzyme-mediated ligation approaches.⁴⁷ Labelled and unlabelled segments of the protein of interest are produced separately, with different uniform or residue-specific labelling patterns, and are subsequently chemically or chemoenzymatically ligated.^{48–50}

Protein semisynthesis by EPL

The incorporation of isotopologues into one or more protein segments, or at selected positions within a protein, requires

connecting two (or more) separate polypeptides, ideally with a native amide bond so that the protein structure remains unaltered. Native Chemical Ligation (NCL, Fig. 2a) was developed in 1994 by Dawson *et al.* for linking unprotected peptide segments in aqueous conditions and has provided a route to access many modified proteins.⁵¹ Alternative amide-forming ligations include KAHA⁵² and Ser/Thr⁵³ ligations; however, comparatively few proteins have been prepared using these strategies. As shown in Fig. 2a, an NCL reaction requires a C-terminal thioester functionality on one peptide segment. The other peptide segment bears an N-terminal cysteine residue. Upon ligation, a native amide bond is formed at the ligation site. The toolbox of extensions to NCL available today allows the rapid and efficient synthesis of proteins, and we direct readers to recent reviews on these topics.^{47,54–56}

The high precision and scope of solid-phase peptide synthesis (SPPS) in combination with NCL enables the site-specific introduction of a broad range of modifications (including PTMs), non-natural amino acids, tags and isotope-enriched amino acids.⁴⁷ In an early example, Kochendoerfer *et al.* demonstrated site-specific incorporation of isotope-enriched amino acids in 81-residue viral protein U (Vpu) by chemical synthesis and NCL (Fig. 2b).⁵⁷ This integral membrane protein is crucial for virus particle release after human immunodeficiency virus-1 (HIV-1) infection through its modulation of viral protein trafficking and



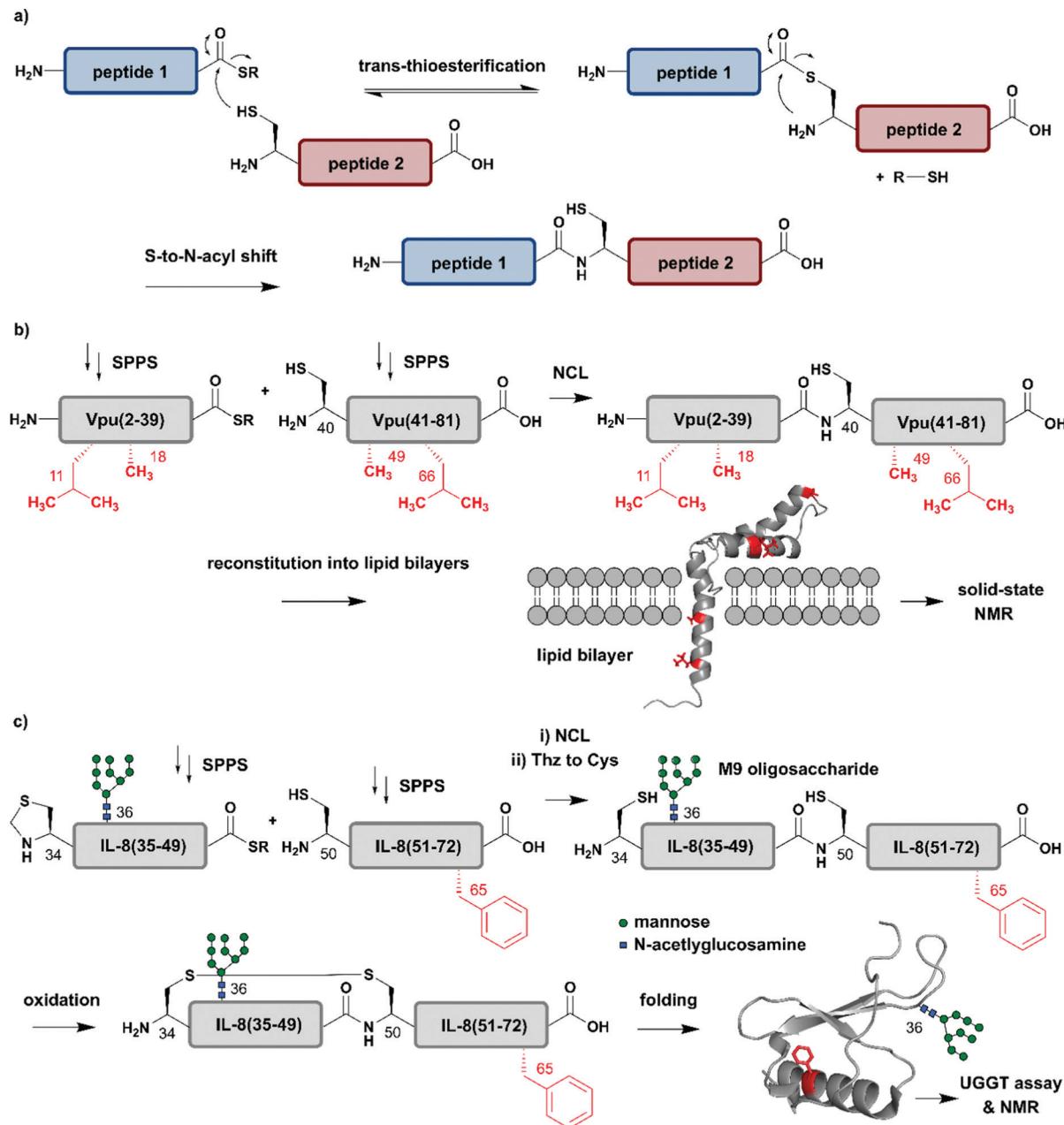


Fig. 2 Chemical protein synthesis based on Native Chemical Ligation (NCL) for site-specific isotope labelling. All peptide segments were prepared by solid phase peptide synthesis (SPPS). (a) General mechanism of the NCL reaction. (b) Site-specific ^{15}N -isotope labelling of four residues (red) of Vpu(2–81) (PDB: 2N28).⁵⁷ (c) Synthesis of a IL-8(34–72) glycopeptide library varying Phe65 (red) with 10 different ^{15}N -labelled amino acids by NCL. Site-specific isotope labelling with varying residues at the predicted core UDP-glucose:glycoprotein glucosyltransferase (UGGT) recognition site (Phe65) confirmed the crucial role of hydrophobic residues by NMR spectroscopy (PDB: 3IL8).⁵⁸ | Grey: unlabelled protein parts | red: ^{15}N isotope-labelled amino acid residues.

membrane permeability.⁵⁹ Kochendoerfer *et al.* generated a ^{15}N -labelled Vpu(2–81) analogue of this viral protein by ligating two site-specifically labelled peptide segments together. The first segment Vpu(2–39) bore ^{15}N -enriched variants (isotopologues) of Leu11 and Ala18, and the second segment Vpu(40–81) contained these isotopologues at Ala49 and Leu66. After NCL and reconstitution into micelles and lipid bilayers, Kochendoerfer *et al.* assessed the overall topology and channel activity of their site-specifically labelled synthetic Vpu. ^1H – ^{15}N HSQC

(Heteronuclear Single Quantum Coherence) and solid-state NMR (ssNMR) experiments revealed comparable characteristics between the chemically synthesised version and the recombinantly expressed, uniformly labelled control protein (without site-specific labelling). ^{15}N -labelling of the protein backbone at four specific residue sites allowed very simplified signal assignments compared to the uniformly labelled Vpu variant. The site-specific labelling pattern provided a rapid fingerprint of membrane-embedded Vpu in ssNMR



measurements, in which only the four ^{15}N -labelled residues could be detected.⁵⁷ Although lacking PTMs, Vpu is an example of how site-specific incorporation of NMR-active nuclei can be used to report on key positions in protein structure and function, especially in membrane systems studied by ssNMR.

Incorporation of both isotope enriched building blocks and posttranslationally modified residues was demonstrated for glycosylated interleukin 8 (IL-8) (Fig. 2c).⁵⁸ In general, glycosylation is a PTM involving covalent attachment of specific oligosaccharide units varying in length and composition to amino acid sidechains by N- and O-linkages, in the endoplasmic reticulum (ER) and Golgi.⁶⁰ Glycans themselves can also be characterised by NMR spectroscopy, even without specific labelling or elaborate sample preparation.⁶¹ The example selected here by Izumi *et al.* focuses on the enzyme UDP-glucose:glycoprotein glucosyltransferase (UGGT) that acts as a quality-control system for protein folding in the ER. To study the mechanism by which UGGT recognizes misfolded species of N-glycosylated IL-8, Izumi *et al.* constructed an IL-8 glycopeptide library using NCL.⁵⁸ IL-8 was split into two segments, with the N-terminal segment IL-8(34–49) bearing a Man9GlcNAc2 (M9) oligosaccharide at Asn36 and the C-terminal segment IL-8(50–72). To identify possible hydrophobic patches recognized by UGGT, Izumi *et al.* selected specific residues and incorporated ^{15}N -labelled variants at the respective positions. After NCL and subsequent air oxidation, the resulting folded IL-8(34–72) was analysed by ^1H – ^{15}N HSQC NMR and revealed a significant chemical shift perturbation of Phe65 upon binding to UGGT. To further investigate the hydrophobic interaction patch at this position, the authors generated a glycopeptide library by substituting Phe65 with ten labelled natural amino acids (Glu, Lys, Gln, Ser, Tyr, Gly, Ala, Pro, Leu and Val). Based on this library generated using NCL, the authors concluded that glycopeptide variants bearing residues with greater hydrophobicity are glycosylated and folded faster.⁵⁸ These chemically synthesised glycopeptides bearing a ^{15}N label at specific positions are an excellent example of combining site-specific isotope-labelling with installation of complex PTMs. As only selected residues are labelled, and therefore observed, assignment of signals in NMR spectra can be simpler than assigning signals from all residues. The obvious drawback is that interactions or conformational changes involving unlabelled residues cannot be assessed easily. Furthermore, the total chemical synthesis of proteins for spectroscopic applications suffers from restrictions in length of the accessible proteins and prohibitively expensive isotope-labelled amino acid building blocks.⁶² Therefore, SPPS is more widely used to incorporate site-specific PTMs than site-specific isotope labels.

Incorporation of site-specific isotope labels can also be achieved by the use of auxotrophic bacterial strains, which require the specific supplement of the respective labelled amino acid in their growth media. Although this methodology can be applied to almost all of the spectroscopic techniques described in this review, it is predominantly used for studying complex systems such as membrane proteins by IR spectroscopy.⁶³ To this end, bacteriorhodopsin was generated

with all 7 Arg residues labelled with ^{15}N in an auxotrophic *Halobacterium salinarium* strain.⁶⁴ Comparison of the vibrational differences in IR spectra between unlabelled and ^{15}N -labelled samples delivered valuable insights and helped to identify Arg82 as a key residue in the light-induced H^+ transport mechanism. Although residue-specific labelling *via* auxotrophic bacteria strains can be performed with relative ease, the approach lacks the precision of protein semisynthesis or genetic code expansion and therefore is only applied occasionally.

Two years after introducing NCL for linking unprotected synthetic peptide segments by a native amide bond, the reaction was extended to the N-terminal modification of a recombinantly expressed protein carrying an N-terminal Cys residue with a thioester-containing fluorescent dye.⁶⁵ In 1998 NCL was used for ligation of recombinantly produced α -thioesters to synthetic peptides, a process named expressed protein ligation (EPL).^{66,67} Recombinant protein expression enables access to protein segments without size limitations and various isotope patterns can be incorporated using methods discussed above. EPL can therefore be used to ligate a recombinantly expressed (and isotope-labelled) protein segment to a synthetic segment bearing a site-specific PTM.⁶⁸ Several methods to generate recombinant protein segments with N-terminal Cys residues have been used in protein semisynthesis. Direct access to such segments is possible if cleavage of the expressed protein segment by endogenous methionine aminopeptidase or signal peptidases occurs. Other strategies rely on selective cleavage of Met–Cys with cyanogen bromide or the use of proteases with highly defined recognition sequences such as factor Xa, thrombin, tobacco etch virus (TEV) protease and small ubiquitin-like modifier (SUMO) protease.⁵⁶ Access to recombinant protein segments with C-terminal α -thioesters makes use of protein splicing.⁶⁶ The underlying biochemical reaction is based on an autocatalytic process controlled by inteins (internal or intervening protein) that induce protein splicing events in which peptide bonds are cleaved and formed (Fig. 3a)^{69,70} and results in splicing of the intein from its N- and C-terminal flanking polypeptides (exteins – external proteins) that are linked by an amide bond – also termed protein splicing (*in cis*).⁵⁶

Since the discovery of inteins and their splicing mechanisms, protein engineering attempts have exploited the versatility of inteins for protein purification to generate proteins with N-terminal cysteine residues and for accessing proteins with C-terminal α -thioesters that can be used in NCL reactions (Fig. 3b). Tags such as the chitin-binding domain and/or poly-histidine tags allow for affinity purification of proteins followed by thiol-mediated intein cleavage, ideally resulting in pure protein segments carrying C-terminal α -thioesters.⁷¹ Among the large number of inteins found to date in all kingdoms of life, the DNA gyrase A (*Mxe* GyrA) intein from *Mycobacterium xenopi* has been most frequently employed for the generation of protein α -thioesters.⁷² However, inteins that show faster cleavage kinetics and higher tolerance towards varying flanking sequences have been reported.^{73,74} Combining large recombinant protein segments with tailor-made synthetic peptides has been



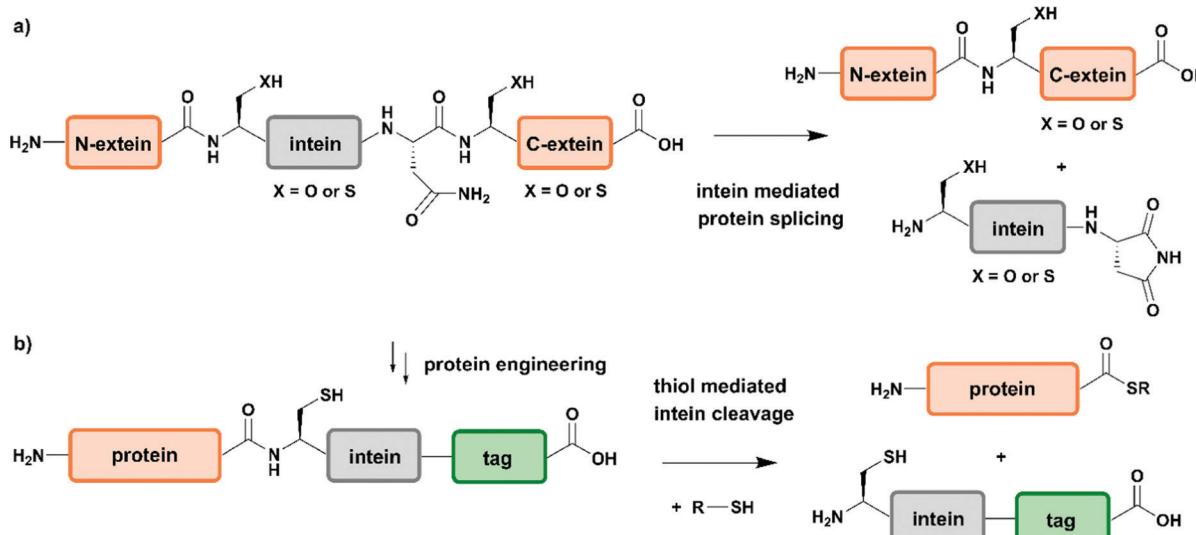


Fig. 3 Intein-mediated protein splicing and generation of protein thioesters. (a) General mechanism of protein splicing. Intein-mediated protein splicing results in linkage of the N- and C-extein (orange) by a native peptide bond after excision of the intein functionality. (b) Protein engineering can generate extein–intein fusion proteins bearing C-terminal purification tags (green), which yields protein α -thioesters after addition of a thiol. | Grey: intein | orange: N-extein and C-extein | green: affinity purification tag.

successfully exploited for segmental labelling of proteins carrying common PTMs such as phosphorylation, acetylation or ubiquitination.

Advantage was taken of EPL in the semisynthetic approach from Hejjaoui *et al.* towards α -synuclein (α Syn), in which the authors sought to introduce phosphorylation(s) site-selectively with segmental isotope labeling of the protein. More precisely, the authors investigated the site-specific phosphorylation of α Syn at Tyr125 (Fig. 4a).⁷⁵ Phosphorylation can affect protein structure and function, as has been shown for many cellular processes and also in several semisynthetic approaches for α Syn,^{75,76} which plays a central role in Parkinson's disease.⁷⁷ Phosphorylation usually occurs on sidechain hydroxyl groups of Ser, Thr and Tyr. Removal of the phosphate group is mediated by the smaller family of protein phosphatases.⁷⁸ In NMR spectra, protein phosphorylation results in a significant downfield shift of Ser (and/or Thr/Tyr) backbone amide resonances.^{79,80} In the study from Hejjaoui *et al.*, a synthetic, phosphorylated α Syn segment (residues 107–140) was ligated to its recombinantly expressed N-terminal counterpart. The recombinant segment was expressed as a α Syn(1–106)–Mxe GyrA intein–6xHis fusion protein in M9 minimal medium, supplemented with the sole nitrogen source $^{15}\text{NH}_4\text{Cl}$. The C-terminal α -thioester functionality was generated through MESNa (2-mercaptoethanol-sulfonate)-mediated intein splicing directly after Ni-NTA purification. The C-terminal α Syn segment, comprising residues 107 to 140 (containing an Ala107Cys mutation to allow EPL) and bearing either no or a single phosphorylation at Tyr125, was produced *via* SPPS. The resulting semi-synthetic, segmentally isotope labelled wildtype (wt) and phosphorylated Tyr125 (pTyr125) α Syn(1–140) variants were conjugated with a methanethiosulfonate (MTSL) spin label at the Cys residue 107 (resulting from the NCL reaction)

for NMR-based paramagnetic relaxation (PRE) experiments.^{81,82} The authors characterised both MTSN-conjugated α Syn variants in their micelle-bound state by NMR and compared biological aspects such as fibrillization. The data revealed significant differences between variants in their monomeric state and suggested a lower local compactness of pTyr125 α Syn compared to the unmodified α Syn. The semisynthetic strategy allowed Hejjaoui *et al.*, to investigate the influence of pTyr125 on α Syn structure, aggregation, membrane binding and subcellular localisation.⁷⁵ This work is an excellent example of the power of protein semisynthesis in studying proteins bearing homogeneous, site-specific PTMs and their impact on transient long-range interactions. The NMR-active ^{15}N -labelled α Syn(1–106) segment provides insight into the effects of pTyr125 on the α Syn structure in solution and in membrane proximity at atomic resolution. The synthetic, unlabelled C-terminal segment is not visible in ^1H – ^{15}N NMR spectra and therefore simplifies assignment of the signals arising from the recombinant (labelled) segment.

In another study, which combined protein semisynthesis, segmental isotope labelling and enzymatic phosphorylation, Chu *et al.* sought to characterise the molecular determinants of the (in-)activation of the AKT serine/threonine kinase 1 (AKT) at atomic resolution (Fig. 4b).⁸³ Activation of the AKT protein (480 residues) is modulated by phosphorylation events at its C-terminal domain, which lead to a relieved intramolecular interaction of its N-terminal pleckstrin homology (PH) domain from the central kinase domain.⁸⁶ To decipher the key determinants of this interaction, the authors applied a two-step NCL strategy for the generation of the three domain Akt protein kinase, which required mutation of Ser122 to Cys to enable segmental labelling *via* a chemoselective ligation reaction. The SPPS-derived C-terminal segment Akt(460–480)



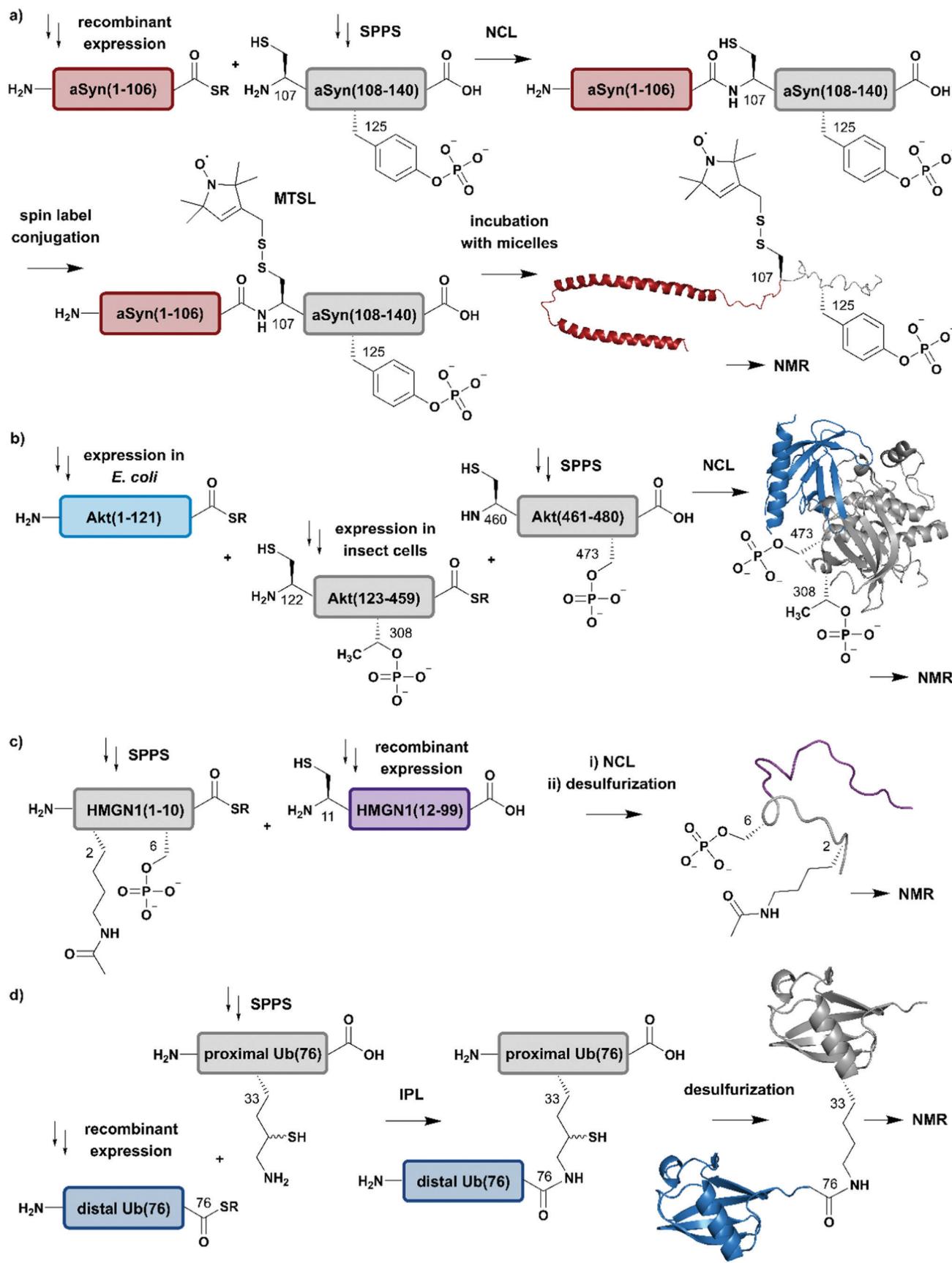


Fig. 4 Expressed Protein Ligation (EPL) for segmental isotope labelling of proteins. (a) Segmental isotope labelling of posttranslationally-modified α Syn(1–140). Recombinant expression of α Syn(1–106) (red) allowed homogenous incorporation of isotopes into the N-terminal α Syn(1–106) segment and chemical

synthesis gave access to the C-terminal peptide containing phosphorylated Tyr125. Upon ligation phosphorylated and segmentally labelled α Syn enabled the detailed study of the influence of Tyr125 phosphorylation on the protein by NMR spectroscopy (PDB: 1XQ8).⁷⁵ (b) Segmental labelling of posttranslationally modified Akt(1–480) by protein semisynthesis. The three-segment assembly strategy is worth mentioning as the two recombinantly produced segments are generated in two different expression systems (bacteria and insect cells). Isotope labelling of Akt(1–121) (light blue) was achieved in *E. coli* (PDB: 3O96).⁸³ (c) Segmental labelling and site-specific modification of HMGN1(1–99) by protein semisynthesis. SPSS allowed the site-specific introduction of selected PTMs (acetylation and phosphorylation) in HMGN1(1–10) (grey) and recombinant isotope-labelled HMGN1(11–99) (purple) segment revealed long-range chemical shift perturbation in subsequent NMR studies (PDB of HMGB: 1LWM).⁸⁴ (d) Isopeptide chemical ligation (ICL) approach for the isotope labelling of the distal Ub(1–76) subunit (blue) in a dimeric Ub chain. The synthetically-derived proximal Ub (grey) was equipped with δ -mercaptoplysine at Lys33 and ligated to the recombinant distal Ub subunit, which enabled Castañeda *et al.* to describe the characteristics of a distal Ub unit in a dimeric Ub assembly using NMR spectroscopy (PDB: 1UBQ).⁸⁵ | Grey: unlabelled protein parts | red/light blue/purple/blue: homogenous, stable isotope-labelled protein segments.

allowed the generation of the respective phospho-forms, yielding peptides bearing phosphate groups at pSer473 or pSer477/pThr479. Unlabelled, non-deuterated Akt(122–459) kinase domain was obtained recombinantly from *Sf9* insect cells, which were phosphorylated *in vitro* by GST-PDK1 at pThr308 and then ligated to the C-terminal Akt(460–480) segment. Isotope labelling of the N-terminal PH domain Akt(1–121) with 2 H, 13 C and 15 N isotopes was achieved by expression in M9 minimal media using an *E. coli* expression system, with 2 H- 13 C glucose and 15 NH₄Cl as sole carbon and nitrogen sources in D₂O-supplemented media. In a second NCL reaction, labelled Akt(1–121) was ligated to the previously generated Akt(122–480) and characterised by 1 H- 15 N HSQC NMR. By combining chemical shift perturbations with peak broadening in the resulting NMR spectra, the data revealed strong interaction patterns between the PH domain and the unlabelled Akt(122–480) segment. Segmental labelling of Akt(1–121) greatly reduced spectral crowding in these experiments and therefore allowed resonance assignment of 66% of non-proline backbone residues in the PH domain, enabling the detailed study of long-range intramolecular interactions between these Akt domains at atomic resolution. These interactions were highly dependent on the respective phosphorylation state of the C-terminal Akt(460–480) segment, which were additionally characterised by cross saturation transfer (CST) NMR measurements.⁸⁷ This CST experiment monitors the selective transfer of magnetization from the protonated Akt(123–459) kinase domain to the perdeuterated N-terminal Akt(1–122). Using this measurement, the authors were able to distinguish between direct (transfer of magnetization) and indirect (no transfer of magnetization) interaction patterns within those two Akt domains. The CST data revealed an influence of pSer473 pSer477/pThr479 on Akt kinase activity and additionally highlighted the importance of a particular linker region spacing the PH domain from the kinase domain, which was already identified in a previous study from the same group⁸⁸ and characterised in detail.⁸³

In a further example of segmental labelling by protein semisynthesis, the effects of site-specific acetylation and phosphorylation on the nucleosome-binding protein HMGN1 were recently investigated (Fig. 4c).⁸⁴ Lys acetylation is catalysed by lysine acetyltransferases (KATs) and reversed by lysine deacetylases (KDACs).⁸⁹ Both enzymatic reactions can be monitored by NMR in real time, as they exhibit characteristic features on the Lys N α and N ϵ in 1 H- 15 N correlation experiments.⁹⁰

Niederacher *et al.* used EPL for the generation of several segmentally isotope-labelled HMGN1 variants bearing site-specific PTMs within the N- and C-terminal parts of the protein. N-terminal acetylation at Lys2 (acLys2) and phosphorylation at Ser6 (pSer6) were introduced in a peptide comprising residues 1–10 of HMGN1(1–10).⁸⁴ After ligation of the N-terminal segment to the isotope-labelled, recombinant HMGN1(11–99) segment, the product was characterised by NMR spectroscopy. 1 H- 15 N HSQC NMR experiments revealed long-range perturbations in the chemical environment of the backbone residues, even in this intrinsically disordered protein (IDP). Furthermore, segmental labelling here helped to minimize signal overlap in the “random coil” region of the 1 H- 15 N correlation spectra, which is often crowded for uniformly labelled IDPs.⁹¹ In particular, the variant bearing both modifications showed significant shifts in the nucleosome-binding domain, which suggests an influence on the binding affinity of HMGN1 to nucleosomes even for small PTMs like acetylation or phosphorylation. This study is an example of how protein semisynthesis can flexibly provide a set of isotope-labelled and PTM-carrying protein variants (modified at their N- or C-termini) that offer valuable insights into the effects of selected PTMs on protein structure and function at the atomic level. The segmental labelling approach facilitated resonance assignment and led to insights into PTM-modulated structural changes and functions.⁸⁴

Another frequent PTM characterised using a combination of ligation chemistry and spectroscopic investigations is ubiquitination. Ubiquitination is the covalent attachment of ubiquitin (Ub) or polyubiquitin (polyUb) chains to the Lys ϵ -amino group of a target protein. Several synthesis strategies have been used for the generation of (poly-)ubiquitin and its installation on target proteins,⁹² but we focus here on Ub attachment combined with NMR spectroscopic analysis. The repetitive nature of polyUb molecules makes it almost impossible to resolve the respective NMR signals from individual subunits, which remained a challenge until 2011.⁹³ This drawback was addressed first by Castañeda *et al.* by applying an isopeptide chemical ligation approach in order to generate segmentally isotope-labelled, dimeric Ub (diUb) (Fig. 4d).⁸⁵ In this study, the proximal Ub was chemically synthesised with a δ -mercaptoplysine at Lys33.⁹⁴ The incorporation of NMR-active isotopes (15 N) in the distal Ub molecules was achieved by recombinant expression in auto-inducing minimal media. After a successful ligation reaction and the following rearrangement *via* S-to-N



acyl shift, the native diUb was desulfurised and characterised by NMR spectroscopy. This allowed Castañeda *et al.* to completely remove signal overlap from the individual diUb elements and revealed features about the structure, conformation and ligand-binding of Lys33-linked diUb. The experiments provided insight onto the monomer-specific inter- and intramolecular interactions of the respective Ub subunits.⁸⁵ Although this approach involves isotope-labelling of the PTM (ubiquitin) rather than the substrate protein, it illustrates the power of protein semisynthesis for studying PTMs by NMR. Without the combination of peptide synthesis, which allowed the site-specific incorporation of the required δ -mercaptoplysine, with isotope labelling by recombinant expression, such a detailed characterisation would have not been possible. Furthermore, the principle of segmental labelling of individual Ub subunits paved the way for non-natively linked Lys11-, Lys27- or Lys63 polyUb chains, as discussed later in this review.

In summary, the currently-available toolbox for generating N-terminal cysteinyl and C-terminal α -thioester peptide/protein segments *via* chemical synthesis and recombinant expression allows the detailed study of posttranslationally modified proteins by NMR spectroscopy. Although other spectroscopic techniques such as Raman or IR spectroscopy greatly benefit from segmental labelling including deuteration by EPL, no relevant examples in the context of PTMs have been published so far. However, we direct the reader to examples of segmental labelling for other spectroscopic techniques,^{95–97} which suggest potential future applications.

Protein trans-splicing (PTS)

Intein-mediated protein splicing in *cis* relies on a continuous protein segment flanked by the N-extein and C-extein sequences, however there are also discontinuous intein variants (so-called split inteins), which can be exploited for segmental labelling of proteins by semisynthesis.⁹⁸ These split inteins are encoded on different gene loci and associate with high affinities in the nanomolar range. Upon association of the split intein segments, a splicing reaction of the intein through a “capture and collapse” mechanism is initiated, named protein *trans*-splicing (PTS) (Fig. 5a).⁹⁹ Natural split inteins are somewhat restricted in use due to their strict sequence requirements. Extensive split intein engineering has led to several shorter intein sequences, novel PTS triggers, faster splicing reactions and also increasing amino acid tolerance in the terminal parts of the exteins.¹⁰⁰ Whereas a broad range of C-terminal split intein domains with an appropriate sequence length (~ 36 – 50 amino acid residues) for SPPS are naturally available and can be used for the incorporation of PTMs and other modifications including specific isotope labels, recent engineering attempts have focused on the reduction of the sequence length of the N-terminal intein domain.¹⁰¹ Whereas the sequence requirements for PTS are typically more restrictive than for EPL methodology, PTS can be carried out at

significantly lower concentrations due to the high affinities of the split intein segments.⁴⁷ Furthermore, split inteins can also be made dependent on various external stimuli and exploited for conditional protein splicing *in vitro* as well as *in vivo* reactions with dimerizing molecules, pH or light.¹⁰²

Application of PTS to the study of structural and functional effects of phosphorylation is demonstrated by a study on the G-protein-coupled receptor (GPCR) β 2AR (β 2-ardrenoreceptor). The binding of agonists to the extracellular domain of GPCRs, which constitute the largest group of cell surface receptors, leads to phosphorylation of their C-terminal cytoplasmic region. Shiraishi *et al.* investigated conformational changes and altered membrane interactions due to phosphorylation of the β 2-ardrenoreceptor (β 2AR) (Fig. 5b).¹⁰³ The authors created a β 2AR variant bearing a segmentally labelled C-terminal tail, which was subsequently studied by NMR spectroscopy. Shiraishi *et al.* achieved this by using an intein-mediated PTS approach, which allowed them to conjugate unlabelled β 2AR(1–348) and ^2H - ^{13}C - ^{15}N -labelled β 2AR(349–413) with $\sim 80\%$ efficiency. Both fusion constructs bearing DnaE split inteins, β 2AR(1–348)-N-intein as well as C-intein- β 2AR(349–413), originated from recombinant expression systems and provide the possibility for isotope labelling using amino acids or metabolic precursors. To suppress signal overlap, they selectively labelled C-terminal Thr and Ile methyl groups using growth media with [α, β - $^2\text{H}_2$, γ - ^{13}C - $^{1}\text{H}_3$] Thr, [$^2\text{H}_2$] Gly and [methyl- ^{13}C , 3,3- ^2H] ketobutyric acid as appropriate precursor molecules. The resulting segmentally labelled, full-length β 2AR(1–413) was embedded into lipid nanodiscs and phosphorylated by GPCR kinase 2. This is one of the few cases in which a specific kinase is available to introduce multiple phosphorylations site-specifically into a large protein. NMR studies revealed a phosphorylation-dependent interaction of the C-terminal tail of β 2AR with intracellular membrane compartments. Following on from a previous study on labelling of Met residues,¹⁰⁴ Shiraishi *et al.* further investigated their findings by specific labelling of four Met residues in a 4Met β 2AR mutant with [^2H -9AA, $\alpha\beta\gamma$ - ^2H -, methyl- ^{13}C -Met] precursor molecules. ^{13}C -labelling of methionine methyl groups is a commonly used methodology in protein NMR and allowed them to assess the structural conformation of the transmembrane helices of the phosphorylated species. Shiraishi *et al.* proposed that phosphorylation might result in an allosteric rearrangement of the transmembrane helices and therefore lead to formation of the structural motif for β -arrestin binding, which offers a possible conserved structural pattern for GPCR recognition.¹⁰³ Here, the authors impressively demonstrate how to overcome the challenges of generating and folding an integral membrane protein after PTS and how to combine PTS with a specific kinase and reconstitution into lipid nanodiscs to collect NMR data that further improves our understanding of signalling events at the molecular level.

To summarize, the discovery, engineering and application of intein-mediated PTS has led to an extensive repertoire of methods for protein assembly. The PTS toolbox allows for the design, synthesis/expression and assembly of

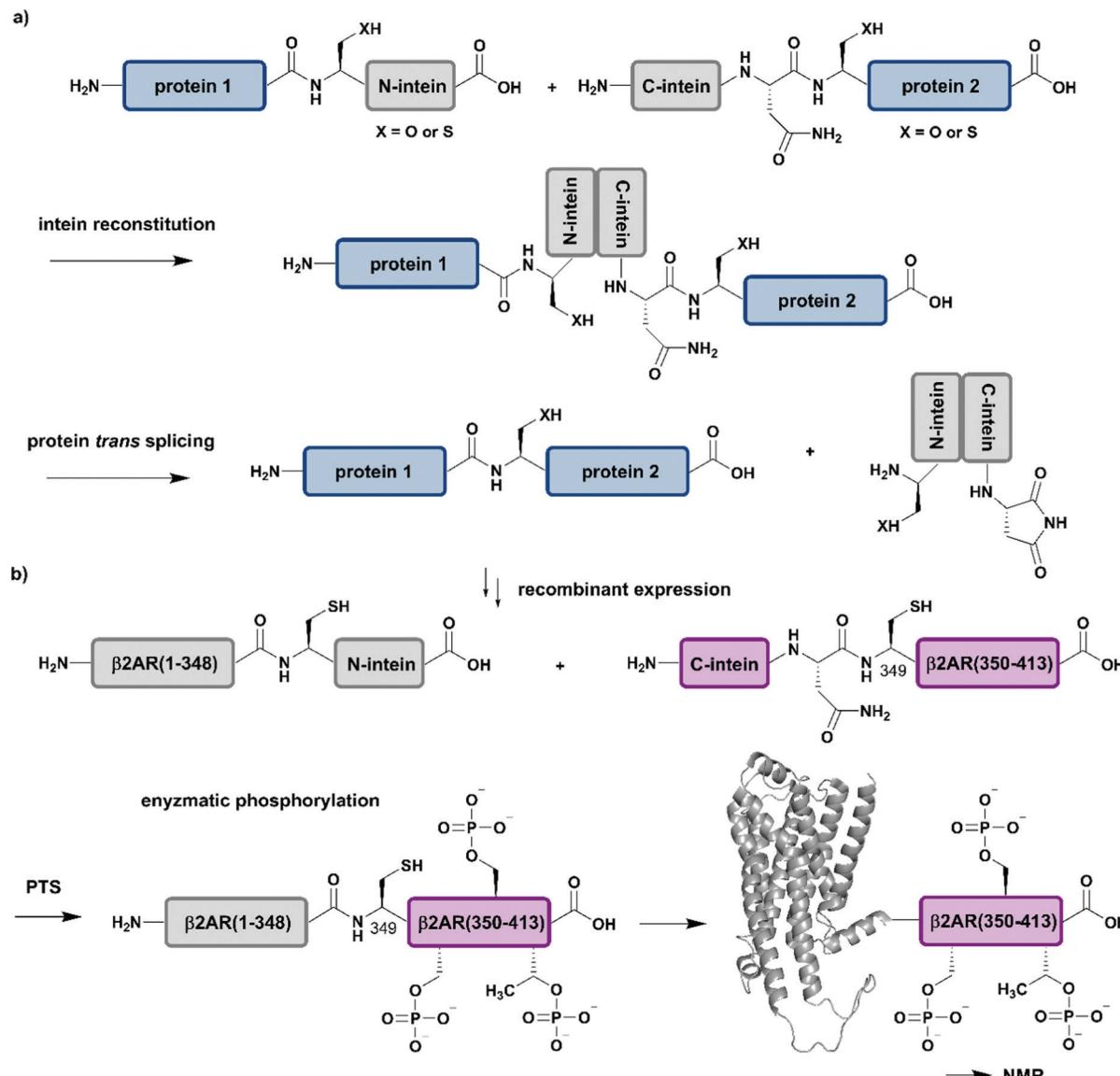


Fig. 5 Protein trans-splicing (PTS) for segmental isotope labelling. All protein segments were prepared by recombinant expression. (a) General mechanism for the PTS reaction. (b) Segmental isotope labelling of enzymatically phosphorylated β 2AR(1–413) by PTS. The unlabelled β 2AR(1–348)–N-intein fusion (grey) construct and its isotope labelled counterpart C-intein– β 2AR(349–413) (purple) were recombinantly expressed, ligated by PTS, enzymatically phosphorylated and investigated by NMR spectroscopy (PDB: 6KR8).¹⁰³ | Grey: unlabelled protein parts | purple: homogenous, stable isotope-labelled protein segments.

synthetically-derived peptides and recombinantly expressed proteins at much lower concentrations than are typically possible in NCL, for example. This allows for segmental isotope labelling and site-specific incorporation of PTMs, fluorescent tags or other modifications for structural studies including those on challenging protein targets such as membrane proteins. PTS can easily be extended to *in vivo* cell experiments, as the PTS reaction is highly biorthogonal.^{105,106} However, PTS systems are still dependent on the flanking sequence of the C-extein and suffer from cross-labelling in *in vivo* segmental isotope labelling approaches,¹⁰⁷ which restricts their application. Possibilities to circumvent these drawbacks are the use of orthogonal methods such as chemoselective ligations other than

NCL or the use of specific enzymes such as Sortase A, which we will discuss below.

Enzyme-mediated protein ligation

Naturally occurring and engineered enzymes that ligate two protein segments together provide an alternative methodology for ligating protein segments bearing PTMs and/or isotope labels. One of the most widely used, the transpeptidase Sortase A from *Staphylococcus aureus*,¹⁰⁸ has been exploited for cyclization, bioconjugation and ligation of polypeptides.¹⁰⁹ In the transpeptidation mechanism of this enzyme, the thiol nucleophile of a catalytic Cys residue attacks the Thr-Gly

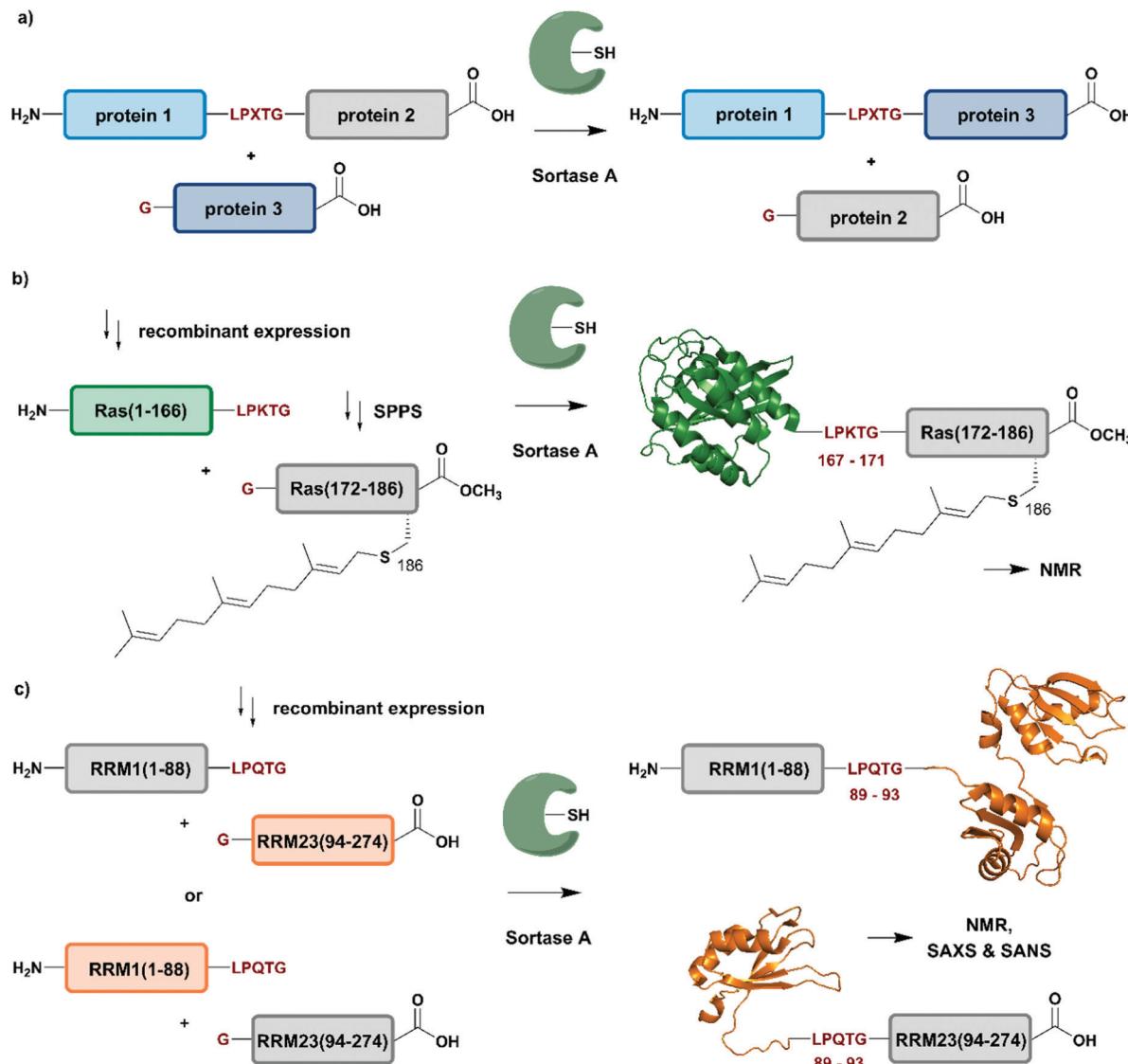


Fig. 6 Sortase A-mediated protein ligation ("sortagging") for segmental isotope labelling. (a) The transpeptidase Sortase A recognizes an "LPXTG" motif (in red) and catalyses an acyl transfer reaction *via* an acyl-enzyme intermediate. (b) Segmental isotope labelling of farnesylated Ras(1-186) (green) by employing Sortase A. Stable isotope labelled Ras(1-166) was ligated to the C-terminal Ras(172-186) (grey) peptide and subsequently characterised by NMR (PDB: 5X9S).¹¹⁷ (c) Segmental deuteration and isotope labelling of RRM1(1-274) by sortagging for NMR, SAXS and SANS experiments. Sortase A allowed the incorporation of stable isotopes either in the N-terminal RRM1(1-88) or the C-terminal RRM23(94-274) domains. (PDB: 5O2V for RRM1; 5O3J for RRM2; 1X4G for RRM3).¹¹⁸ | Olive: schematic Sortase A enzyme with its catalytic Cys residue | grey: unlabelled protein parts | green/orange: homogenous, stable isotope-labelled protein segments | red: amino acid substitutions at the ligation junction.

peptide bond in a substrate containing the P_1 -LPXTG- P_2 motif ($P_1/P_2/P_3$ = residues flanking the ligation site; X = any amino acid) (Fig. 6a and Table 1). The resulting acyl-thioester-enzyme complex is attacked by a nucleophilic polypeptide (G- P_3) carrying one or more N-terminal Gly residues. This reaction, also referred to as 'sortagging', results in the formation of the P_1 -LPXTG- P_3 ligation product, which bears a native peptide bond at the ligation site.¹⁰⁹ Sortagging is an efficient tool for bioconjugation and has been used for segmental isotope labelling of multi-domain proteins such as vinculin¹¹⁰ or the attachment of solubility tags.¹¹¹ However, the ligated products bear the LPXTG

motif as a ligation "scar" if this is not part of the native sequence at the ligation site. Further drawbacks are the poor catalytic efficiency of the enzyme and reversibility of the transpeptidation reaction as it generates a Sortase recognition site upon ligation, which can reduce the reaction yield.¹¹² Therefore, Sortase A homologs and engineered variants have been investigated to overcome these sortagging scars and the limited catalytic efficiency.^{113,114} Overall, Sortase A-catalyzed reactions can provide milligram quantities of ligated product and allow the introduction of PTMs, which makes them an attractive tool for chemoenzymatic segmental labelling of proteins.^{115,116}



Table 1 Peptidyl transferases and peptide ligases

Enzyme and recognition site	Advantages	Disadvantages
Sortase A: 	<ul style="list-style-type: none"> • Robust to a broad range of conditions • High yields possible • Recombinant expression possible 	<ul style="list-style-type: none"> • Not traceless (LPXTG scar) • Low catalytic efficiency • Reversible reaction and hydrolysis
Butelase-1: 	<ul style="list-style-type: none"> • Smaller recognition site than Sortase-A • High yields and efficiency • Broad sequence acyl acceptor • Recombinant expression possible (only for OaAEP1b) 	<ul style="list-style-type: none"> • Not traceless (D/N scar) • Recombinant expression not possible • Reversible reaction and hydrolysis • Intermediate yields and poor efficiency (only for OaAEP1b)
Trypsiligase: 	<ul style="list-style-type: none"> • Fast reaction (few minutes) • High substrate selectivity • N- and C-terminal modifications possible • Recombinant expression possible 	<ul style="list-style-type: none"> • Not traceless (Y-RH scar) • Strict sequence dependency • Reversible reaction and hydrolysis • Activated ester required
N-terminal modification 		
C-terminal modification 		
Subtiligase: 	<ul style="list-style-type: none"> • Broad sequence compatibility • Reduced hydrolysis compared to other enzymes • High catalytic efficiency • Recombinant expression possible 	<ul style="list-style-type: none"> • Reversible reaction with some variants • Activated ester required

The advantages of segmental labelling of proteins by sortagging were demonstrated in a study by Ke *et al.*, where the strategy was combined with lipidation of the rat sarcoma (Ras) GTPase (Fig. 6b).¹¹⁷ Lipidations play central roles in controlling protein localisation, in protein secretion, signal transduction or apoptosis and are therefore an important, but also challenging target (due to handling problems) for structural investigations.^{119,120} Prenylation and palmitoylation are found on a broad range of target proteins, with the rat sarcoma (Ras) GTPase family being a prominent substrate for these modifications.¹²¹ Lipidation of the Ras protein mainly occurs in its C-terminal hypervariable region Ras(172–186) and triggers several modification processes, which result in a farnesylated Ras(1–186) variant bearing a C-terminal carboxymethylation. This process leads to localisation of Ras at the cell membrane, a prerequisite for subsequent activation of downstream kinases such as c-Raf-1.¹²² To decipher the molecular key determinants of this interaction by NMR, beyond the analysis based on fluorescence developed earlier,^{123,124} Ke *et al.* utilized a versatile Sortase-A based ligation strategy for the preparation of

C-terminally modified Ras (Fig. 6b).¹¹⁷ The authors recombinantly expressed the catalytic domain of Ras(1–166) in M9 minimal medium, which allowed them to label this segment with ¹³C and ¹⁵N. The Ras(1–166) fusion protein was N-terminally conjugated to a 6xHis-tag for affinity purification and bore the required LPKTG motif for sortagging at its C-terminus. The synthetic Ras(172–186) peptide variants carried a farnesylation at Cys186 and an N-terminal Gly nucleophile, which is required for the Sortase-A mediated ligation. The ligation reaction was performed with an excess of Sortase A and the peptide-terminal segment to accommodate for the limited catalytic efficiency of the enzyme and to address further drawbacks such as product hydrolysis and undesired reverse reactions. Furthermore, the reaction required mutation of residues 167–171 in the Ras sequence to the non-native LPKTG sequence, which might alter the flexibility of the protein despite its non-native origin.¹²⁵ With the prenylated Ras variants carrying different PTM patterns in hand, Ke *et al.* characterised the intramolecular interaction between Ras(172–186) and its catalytic domains by NMR. ¹H–¹⁵N



HSQC experiments revealed several signals with significant chemical shift perturbations and led to the identification of intramolecular interaction sites between the catalytic Ras(1–166) domain and the C-terminal Ras(172–186) segment. Further experiments combined NMR spectroscopy with surface plasmon resonance and an increased binding affinity was observed between the catalytic Ras domain and the C-terminal Ras(172–186) segment upon farnesylation,¹¹⁷ in accordance with previously published results from others.¹²⁶ This work showcases an application of Sortase A-mediated protein ligations with challenging peptide substrates. The relatively small modification to the Ras sequence (compared to fusion constructs with inteins in EPL and PTS approaches) is straightforward and typically does not require extensive optimization, making it quite popular within the NMR community for segmental labeling approaches.

Segmental labelling of proteins also plays an important role when studying the shape of macromolecular assemblies using SAS technologies. By combining perdeuteration of certain segments with contrast variation with the solvent, the position of each individual protein subunit in such assemblies can be revealed, as demonstrated by Sonntag *et al.* (Fig. 6c).¹¹⁸ The authors applied a segmental, domain-selective perdeuteration approach for a structural study of the RNA-binding protein T-cell-restricted intracellular antigen-1 (TIA-1) using Sortase A-mediated ligations. TIA-1 comprises three RNA recognition motifs (RRM), which were split into two RRM segments (RRM1(1–92) and RRM2(93–274)).¹²⁷ The segments are connected *via* a flexible linker region, which was chosen as a ligation junction for sortagging. In addition to an N-terminal 6xHis-tag and TEV protease cleavage site, RRM1(1–92) carried the LPQTG motif required for sortagging at its C-terminus. RRM2(93–274) bore a Gly residue at its N-terminus. Due to the possible drawbacks of this sequence manipulation, the authors carefully analysed the properties of native and ligated RRM123 upon RNA binding by comparing both variants using ¹H–¹⁵N HSQC NMR spectroscopy, SAXS and static light scattering. As no differences were observed between the variants, the sortagging protocol was applied to segmental deuteration of either RRM1(1–92) or RRM2(93–274) (Fig. 6c). The initial structural analysis and computational modelling of both segmentally deuterated RRM123 versions revealed approximately 5000 different structural models of the protein–RNA-complex. After the use of SAXS and contrast-matched SANS filters following the initial NMR experiments, the structure ensemble could be refined to only 5 models of the RRM123–RNA complex.¹¹⁸ Although this study did not include any PTM, it represents an important example of how segmental isotope labelling of proteins can provide insight into individual subunits of multi-domain proteins and ligand-induced changes in the overall conformation analysed by scattering techniques; segmental deuteration of specific protein domains allowed for drastic reduction of the number of possible structures. Furthermore, this study illustrates how to evaluate and how to choose possible ligation junctions for sortagging that do not interfere with protein function.

An alternative enzyme-based ligation can be achieved with the Asx-specific (Asx = Asn or Asp) ligase family usually referred to as asparaginyl endoproteases (AEP).¹²⁸ The most prominent member, butelase-1, was isolated from the tropical plant *Clitoria ternatea*.¹²⁹ Butelase-1 requires only three amino acids (D/N-HV) for substrate recognition, provides a higher catalytic efficiency and has a broader sequence tolerance than Sortase A (Table 1).¹³⁰ Many applications of butelase-1 ligation of synthetic and recombinant segments in protein synthesis and bioconjugation have been demonstrated.^{131,132} So far, butelase-1 has been applied mainly for peptide cyclization reactions; its application in ligation chemistry is restricted by the fact that it is not recombinantly available and must be isolated from plants.¹³² However, recent advances in protein engineering yielded a butelase-1-like ligase, which can be recombinantly expressed in *Sf9* insect cells.¹³³ Another less active homolog of butelase-1 from *Oldenlandia affinis*, OaAEP, has been successfully produced in *E. coli*.¹³⁴ Its catalytic efficiency is ~90 times lower than butelase-1, but it has been used in ligations leading to site-specific¹³⁵ or segmentally isotope labelled proteins such as maltose-binding protein (MBP), which were subsequently characterised by NMR spectroscopy.^{136,137} Advantages and disadvantages of enzymes used for protein ligation are summarised in Table 1.

Protein segments can be ligated using proteases in reverse, in addition to peptidyl transferases, which catalyse peptide bond formation through transacylation of amines. The reaction is based on engineered proteases such as subtiligase and trypsiligase. The latter enzyme is a trypsin variant capable of ligating two protein segments together if one segment carries an activated ester moiety at its C-terminus (Table 1). Efficient catalysis is dependent on a Y-RH recognition sequence and allows for introduction of modifications at either the N- or C-terminus of a protein under native conditions.¹³⁸ For introduction of N-terminal PTMs, the released N-terminus of the RH-protein precursor can be conjugated with activated acyl donor substrate mimetics.¹³⁹ Trypsiligase can also be exploited for the regioselective C-terminal modification of proteins through a transpeptidation reaction between a Y-RH recognition motif and a nucleophilic acyl moiety (Table 1). One major advantage of trypsiligase-mediated ligation reactions is the rapid reaction (usually within minutes). However, limitations for its use in protein semisynthesis are due to its high sequence specificity, which theoretically only allows traceless ligation of 0.5% of all known protein sequences, but suppresses undesired side reactions. Additionally, yields are lowered by hydrolysis and reversibility of the reaction.¹³⁰

Subtiligase is a protease engineered for bioconjugation from the protease subtilisin isolated from *Bacillus amyloliquefaciens* (Table 1).¹⁴⁰ Subtiligase variants have high catalytic activity and a broad sequence tolerance, but reversibility of the peptide bond formation remains a significant drawback.¹⁴¹ Subtiligases catalyse ligation between a peptide ester or thioester and the N-terminal α -amine of a protein or peptide, without the strict sequence specificity required by sortases, AEPs or trypsiligases.



Ideally, the ligation site is flanked by hydrophobic amino acids (residues P_1 and P_4) at the N-terminal positions, whereas the C-terminal site is occupied by Arg, Met or small residues at P_1' and large, hydrophobic residues at P_2' .¹⁴¹ Further engineering approaches have led to an enzyme called peptiligase, a Ca^{2+} -independent enzyme for transpeptidation with higher stability than non-engineered subtiligase variants. Peptiligase has the ability to conjugate a C-terminal carboxyamidomethyl ester and a nucleophilic acyl acceptor without significant hydrolysis in an aqueous environment.¹⁴² Recent refinement attempts resulted in the generation of another subtiligase variant termed omniligase, which has a broad tolerance for substrates at the P_1' and P_2' positions.¹⁴³ Together, subtiligase variants offer several advantages and a broad spectrum of possible applications (Table 1). For application in protein synthesis, substrates for the ligation reaction can be prepared *via* SPPS, which allows the site-specific incorporation of PTMs or other modifications. Alternatively, either or both substrates can be recombinantly expressed, allowing for isotope labelling. In comparison to native chemical ligation methodologies, enzyme-mediated ligations do not require a cysteine residue at the ligation site.^{144,145}

In addition to ligating protein segments, enzyme-mediated protein conjugations can also be used to attach posttranslational modifications such as glycosylation or ubiquitination to a labelled protein. For instance, Slyntko *et al.* employed a glycosyl transferase to attach a (unlabelled) glycan part enzymatically to the isotope labelled protein.¹⁴⁶ This also applies to other PTMs such as acetylation or phosphorylation, where acetyltransferases¹⁴⁷ or kinases¹⁴⁸ can be employed for post-translational modification of an isotope-labelled protein *in vitro*.

Joining protein segments with non-native linkages

All (semi-)synthesis strategies described above are valuable tools for generating proteins from segments linked by native peptide bonds, which makes them indistinguishable from their native counterparts. In some cases, however, a non-native linkage between two protein segments might facilitate access to (modified) proteins or confer desired properties to the target protein (*e.g.* increased stability). To avoid undesired functional consequences caused by non-native connections, one needs to carefully select the site(s). Reactions used to create non-native linkages need to proceed quickly and be fully bioorthogonal, not affecting other functional groups in the protein of interest. Bio-orthogonal reactions typically take place on functional groups not found in proteins such as azide, alkyne, ketone or aldehyde moieties. Several excellent reviews of chemoselective and bioorthogonal reactions applied for protein modification and conjugation are available.^{149–151} Therefore, this review will remain focused on the use of such reactions in the context of specific isotope labelling.

A popular strategy for bioconjugation is the copper(i)-catalyzed azide–alkyne cycloaddition (CuAAC), sometimes

referred to as the “click reaction” (Fig. 7a).^{152,153} Both functionalities required for this reaction, azides and alkynes, can be easily incorporated into synthetic peptides or included in recombinant proteins using genetic code expansion tools to install non-proteinogenic amino acid building blocks *via* stop-codon suppression.¹⁵⁴

The advantages of click-chemistry based approaches for the generation of artificially conjugated diUb were recently demonstrated by Schneider *et al.* (Fig. 7b).¹⁵⁵ They generated artificially Lys11-, Lys27- or Lys63-linked diUb variants and assessed their conformational and functional properties by NMR spectroscopy. The unlabelled, distal Ub carried the unnatural amino acid L-azidohomoalanine in place of the C-terminal Gly75 and Gly76 residues. The labelled Ub variants, which bore single Lys-to-Cys mutations at the required positions, were reacted site-specifically with propargyl-acrylate and subsequently incubated with the distal Ub molecules. The following click CuAAC reaction resulted in an artificial triazole-linked diUb, which Schneider *et al.* used for the acquisition of NMR data, supported by molecular dynamics simulations. Although the derived molecules were artificially linked, their structural, dynamic and functional properties were assumed to mirror the behaviour of native isopeptide-linked Ub dimers. A comparison of the associated chemical shift perturbations by ^1H – ^{15}N NMR correlation spectra between the triazole-linked variants¹⁵⁵ and previously published isopeptide-linked diUb^{85,156} confirmed this hypothesis. The example described here nicely reflects the application of click chemistry to quickly link isotope labelled protein domains in a complex assembly. Segmental labelling increased the resolution of NMR signals of each individual Ub subunit, which would not have been possible by uniform labelling of the complete diUb due to multiple overlapping signals. However, the application of this labelling strategy is restricted to systems, such as that by Schneider *et al.*,¹⁵⁵ in which a disadvantageous influence of the non-native linkage can be excluded.

Several strategies generating non-native linkages have also been applied for the conjugation of Ub to substrate proteins. For example, Ub attachment *via* disulfide linkages was described first by Chatterjee *et al.* and Chen *et al.* in 2010.^{157,158} This approach was further extended by Debelouchina *et al.* to site-specific conjugation of isotope-labelled Ub to unlabelled histone H2B.¹⁵⁹ In this work, Ub was fused with an *Ava* DnaE intein and subsequently cleaved with cysteamine to generate Ub with a C-terminal thiol group (Fig. 7c). The single Cys residue (Lys120Cys) of the H2B mutant was activated with 2,2'-dithiobis(5-nitropyridine) (DTNP) after recombinant expression. Co-incubation of H2B- with Ub bearing a C-terminal thiol yielded a disulfide analogue of Ub-H2B, which then was incorporated into nucleosomes. By using a ^1H – ^2H exchange strategy in combination with NMR spectroscopy, Debelouchina *et al.* characterised the effects of ubiquitylation on chromatin modulation and structure. These experiments revealed an acidic surface patch comprising two Glu residues on Ub, which are essential for decompaction of the nucleosome. This strategy significantly accelerated the generation of protein variants carrying isotope labelled ubiquitin, however the instability of the introduced disulfide linkage to reducing conditions should be noted.¹⁵⁹

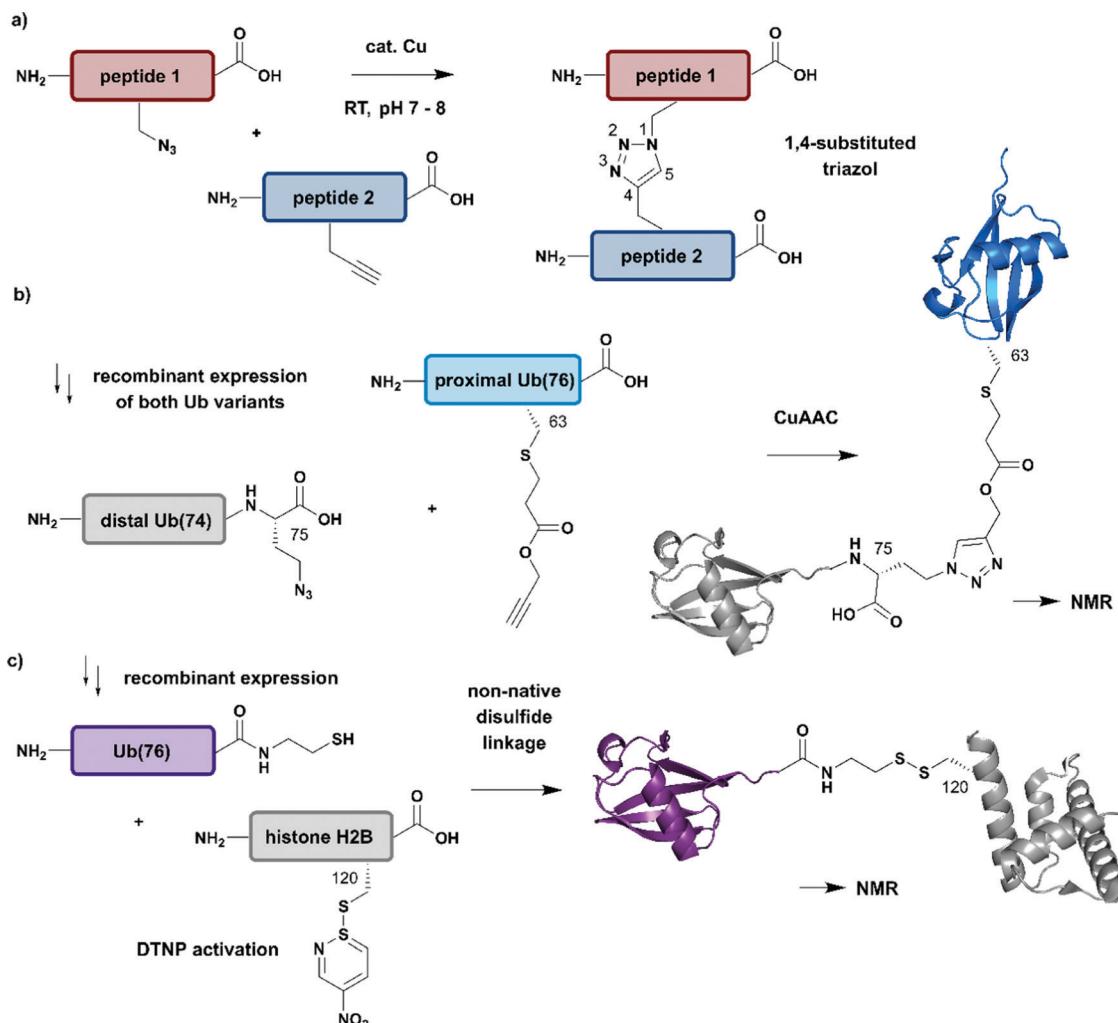


Fig. 7 Non-native linkages for joining differently isotope-labelled protein segments. (a) General mechanism of copper(I)-catalyzed formation of a 1,4-substituted triazole linkage between two peptides ('click reaction, CuAAC'). (b) Segmental isotope labelling of the proximal Ub(76) subunit (cyan) by CuAAC. Both Ub segments were derived recombinantly (PDB: 1UBQ).¹⁵⁵ (c) Site-specific conjugation of isotope-labelled Ub(76) to unlabelled histone H2B (both proteins were prepared recombinantly) by mimicking an isopeptide bond by a disulfide linkage. The obtained Ub-H2B complex analogue was characterised by NMR and revealed an acidic surface patch modulating its chromatin interaction (PDB: 6A7U).⁴⁶ | Grey: unlabelled protein parts | cyan/purple: homogenous, stable isotope-labelled ubiquitin.

Beyond the scope of segmental and site-specific isotope labelling

So far, we have discussed several methodologies for incorporation of isotope labeling patterns into segments of proteins. Alternative strategies, such as the chemical modification of lysine sidechains by ¹³C-methylation,¹⁶⁰ can introduce stable isotopes into proteins generated by conventional methods and they can deliver valuable insight into protein-protein interactions without the need to deal with elaborate expression conditions and/or bioconjugation approaches. Furthermore, a discussion of segmental labelling would be incomplete without mentioning approaches that address multi-protein complexes and methods to control their assembly with individually labelled protein components. Incorporation of stable isotopes into a single protein in a larger complex can deliver valuable insights into its role and behavior in a complex environment. A prime example of protein-specific

isotope labeling in a larger complex has been provided by Liokatis *et al.*¹⁶¹ They focussed on differential isotope-labelling and site-selective incorporation of PTMs (acetylation and phosphorylation) in an octameric histone complex (Fig. 8). For their experiment, the authors recombinantly expressed histone H3 variants in M9 minimal media containing ¹⁵NH₄Cl or ¹³C-glucose either fused to a 6xHis-tag (for ¹⁵N-labelled H3) or a Strep-tag (for ¹³C-labelled H3), which allowed for purification by a two-step affinity process of all histone H3 pools after enzymatic phosphorylation and reconstitution with histone H4. Afterwards, the phosphorylated and purified H3/H4 tetramers were incubated with unlabelled histone H2A and H2B and DNA, resulting in fully reconstituted, asymmetrically modified nucleosomes (asNucs). In order to assess the reciprocal PTM crosstalk between acetylation and phosphorylation in these segmentally labelled asNucs, they were incubated with the acetyltransferase Gcn5. Due to the sophisticated labelling approach it was possible to follow the

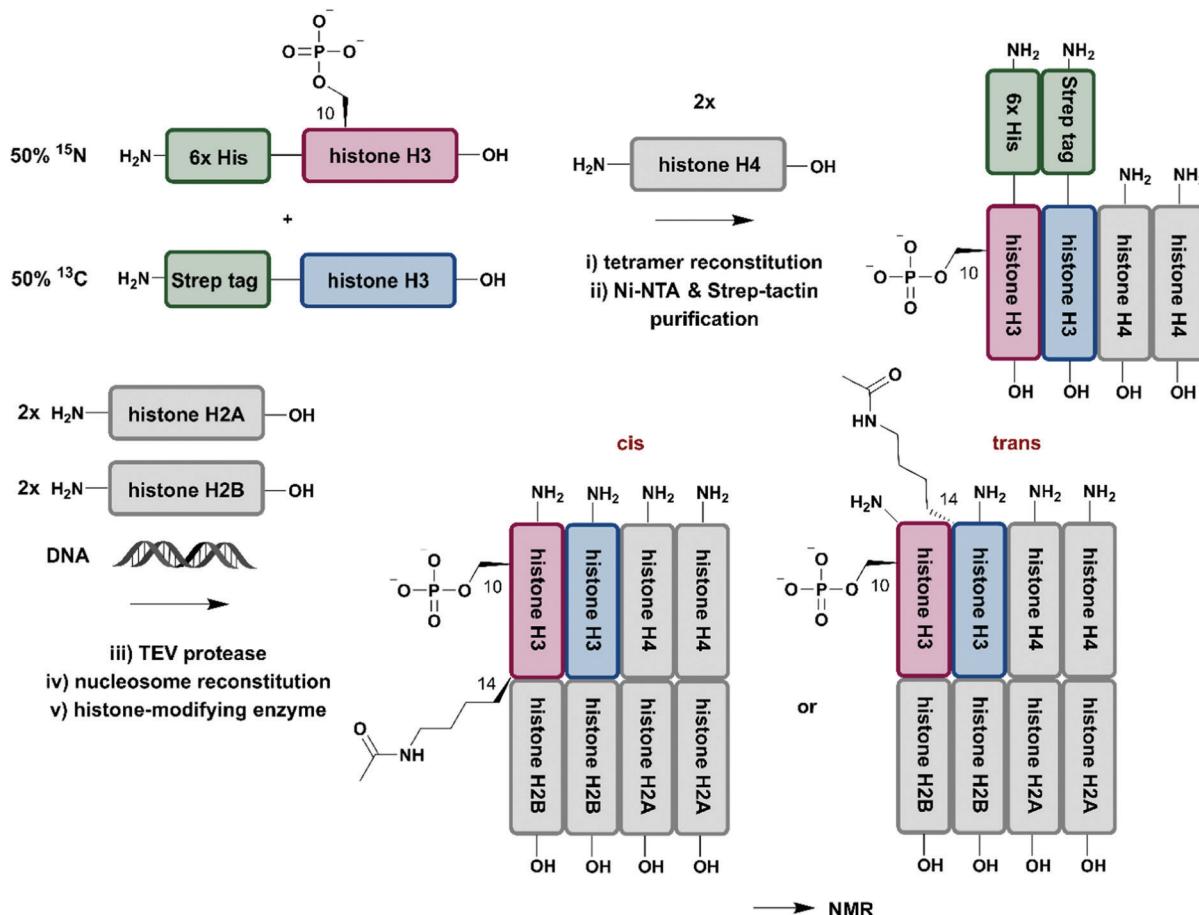


Fig. 8 Isotope labelling of individual subunits in a protein complex. All protein segments were prepared by recombinant expression. A sequential Ni-NTA and Strep-tactin separation workflow yielded asymmetrically phosphorylated and ^{13}C / ^{15}N (blue/red) labelled H3/H4 tetramers, which were employed for reconstitution of functional nucleosomes by addition of unlabelled H2A, H2B and DNA. Upon incubation of reconstituted nucleosomes with acetyltransferase Gcn5, PTM-induced crosstalk in *cis* and in *trans* could be studied by time-resolved NMR spectroscopy.¹⁶¹ | Grey: unlabelled protein segments | red: homogenous, ^{15}N stable isotope-labelled protein segment | blue: homogenous, ^{13}C stable isotope-labelled protein segment | green: affinity purification tags.

acetylation of H3K14 in a time-resolved manner on both H3 tails independently. The recorded ^1H - ^{15}N and ^1H - ^{13}C NMR spectra displayed a stimulatory effect of Ser10 phosphorylation. This influence on histone modification was described previously by the same group,¹⁶² on the acetylation of Lys14 in *cis*, whereas the reaction was delayed in *trans*. The procedure was further exploited in experiments for studying PTM crosstalk between Lys methylation and Thr phosphorylation. Liokatis *et al.* incubated the asNucs in a first step with a Lys methyltransferase, an enzyme whose activity can be monitored in a time-resolved manner by NMR,¹⁶³ and then monitored the H3 Thr3 phosphorylation by NMR. These experiments revealed an inhibitory effect of monomethylated Lys residues on the phosphorylation activity of the Haspin kinase.¹⁶¹ The approach described by Liokatis *et al.* offers an opportunity to specifically detect and monitor enzymatic PTM changes of individual histones in a nucleosome assembly and can be interfaced with other controlled assembly routes for nucleosomes,¹⁶⁴ and to study the dynamic nucleosome landscape.¹⁶⁵

Conclusions

The methodologies and examples reviewed here demonstrate strategies for segmental and site-specific labelling of proteins, facilitating their detailed characterisation by spectroscopic applications. The application of NCL (or alternative ligation reactions) to stitch together synthetic peptides allows for precise incorporation of labelled amino acids as well as other modifications such as PTMs. However, high costs and limited availability of building blocks together with size limitations of synthetic peptides restrict its broader use in the field of isotope labelling – a drawback that might be overcome by the development of cost-effective methods to generate isotope labelled amino acids suitably protected for solid phase peptide synthesis.^{62,166} EPL circumvents these challenges by accessing isotope labelled protein segments using recombinant expression, which also removes size limitations. In combination with SPPS, posttranslational modifications can be incorporated into segmentally labelled proteins. Such semisynthetic proteins offer the possibility for detailed structural and functional

characterisation, even allowing for monitoring of weak and dynamic interactions.¹⁶⁷ A major disadvantage, however is that the synthetic segment bearing the PTMs is typically unlabelled and therefore short-range structural changes caused by the PTM(s) are not observed. EPL also suffers from drawbacks such as limitations based on sequence and protein properties that might prevent expression at the required scale and/or need time-consuming optimization of workflows, which often require extensive expertise in expression and peptide synthesis not readily available in many laboratories.

Similar challenges apply to segmental isotope labelling and site-specific modification of proteins by PTS, with respect to engineering of split intein fusion constructs. However, PTS reactions can be achieved at relatively low concentrations and under physiological conditions, which has made PTS one of the most used approaches to segmentally labelled proteins.⁵⁶ Similarly popular are some enzyme-mediated ligation approaches, *e.g.* based on Sortase A, which offer a rapid and easy workflows for segmental labelling.¹³⁰ However, the resulting ligation products often carry a ligation “scar” comprising non-native amino acids at the ligation site (with the exception of subtiligase). Non-native linkages, even though more subtle, also result from the use of biorthogonal ligations such as CuAAC to generate modified proteins.

In summary, the methods discussed here are of significant value to the protein chemistry and spectroscopy communities when zooming into complex protein structures by using selective isotope labelling strategies to study (dynamic) effects of PTMs on specific residues or domains. We expect that their relevance will further increase as protein chemistry and protein engineering approaches become more easily accessible to biochemistry and structural biology groups. However, for now the few examples that document the power of selective isotope labelling to study the effects of PTMs on proteins are clear evidence for the fact that more intensive collaborations between protein chemists and structural biologists are needed as both aspects require high levels of expertise and are seldom combined. Continuing developments in protein (semi)-synthesis, such as the diselenide-selenoester ligation (DSL)¹⁶⁸ or the use of solubilization tags,¹⁶⁹ now allow protein assembly at low concentrations, at variable sites and of hydrophobic peptide segments. These advances are complemented by newly emerging flexible and traceless enzymatic ligation reactions as well as by better access to isotope labelled amino acids and metabolic precursors. Together with spectroscopic advances that allow higher-resolution data to be obtained with smaller amounts of proteins, such as dynamic nuclear polarization (DNP) NMR¹⁷⁰ or dissolution DNP methods¹⁷¹ that can also be employed in native environments such as cells,¹⁷² boundaries with respect to the complexity of systems analyzed will be pushed.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

C. F. W. B., A. C. C. and D. P. V. received funding from the Vienna Science and Technology Fund (WWTF, project LS17-008). A. C. C. is supported by a UQ Development Fellowship from the University of Queensland (project 613982).

References

- 1 R. Aebersold, J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. MacCoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schluter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlen, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young and B. Zhang, *Nat. Chem. Biol.*, 2018, **14**, 206–214.
- 2 C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto Jr., *Angew. Chem., Int. Ed.*, 2005, **44**, 7342–7372.
- 3 K. N. Chuh, A. R. Batt and M. R. Pratt, *Cell Chem. Biol.*, 2016, **23**, 86–107.
- 4 K. W. Barber and J. Rinehart, *Nat. Chem. Biol.*, 2018, **14**, 188–192.
- 5 A. C. Conibear, *Nat. Rev. Chem.*, 2020, **4**, 674–695.
- 6 E. Nogales, *Nat. Methods*, 2016, **13**, 24–27.
- 7 T. R. Alderson and L. E. Kay, *Curr. Opin. Struct. Biol.*, 2020, **60**, 39–49.
- 8 A. Kumar, V. Narayanan and A. Sekhar, *Biochemistry*, 2020, **59**, 57–73.
- 9 F.-X. Theillet, C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M.-K. Yoon, R. W. Kriwacki, I. Landrieu, G. Lippens and P. Selenko, *J. Biomol. NMR*, 2012, **54**, 217–236.
- 10 F. Delhommel, F. Gabel and M. Sattler, *J. Mol. Biol.*, 2020, **432**, 2890–2912.
- 11 A. Barth, *Biochim. Biophys. Acta*, 2007, **1767**, 1073–1101.
- 12 H. Ma, X. X. Han and B. Zhao, *TrAC, Trends Anal. Chem.*, 2020, **131**, 116019.
- 13 F. Hertel and J. Zhang, *Biopolymers*, 2014, **101**, 180–187.
- 14 R. Aebersold and M. Mann, *Nature*, 2016, **537**, 347–355.
- 15 A. H. Kwan, M. Mobli, P. R. Gooley, G. F. King and J. P. Mackay, *FEBS J.*, 2011, **278**, 687–703.
- 16 A. Bax and G. M. Clore, *J. Magn. Reson.*, 2019, **306**, 187–191.
- 17 P. Selenko, *Int. J. Mol. Sci.*, 2019, **20**, 1278.
- 18 J.-H. Ardenkjær-Larsen, G. S. Boebinger, A. Comment, S. Duckett, A. S. Edison, F. Engelke, C. Griesinger, R. G. Griffin, C. Hilty, H. Maeda, G. Parigi, T. Prisner, E. Ravera, J. van Bentum, S. Vega, A. Webb, C. Luchinat,



H. Schwalbe and L. Frydman, *Angew. Chem., Int. Ed.*, 2015, **54**, 9162–9185.

19 K. Pervushin, R. Riek, G. Wider and K. Wüthrich, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 12366–12371.

20 A. Mainz, T. L. Religa, R. Sprangers, R. Linser, L. E. Kay and B. Reif, *Angew. Chem., Int. Ed.*, 2013, **52**, 8746–8751.

21 D. Rose-Sperling, M. A. Tran, L. M. Lauth, B. Goretzki and U. A. Hellmich, *Biol. Chem.*, 2019, **400**, 1277–1288.

22 J. Hennig and M. Sattler, *Protein Sci.*, 2014, **23**, 669–682.

23 M. A. Graewert and D. I. Svergun, *Curr. Opin. Struct. Biol.*, 2013, **23**, 748–754.

24 V. A. Lorenz-Fonfria, *Chem. Rev.*, 2020, **120**, 3466–3576.

25 A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk and M. Baranska, *J. Raman Spectrosc.*, 2013, **44**, 1061–1076.

26 H. Zhang and H. van Ingen, *Curr. Opin. Struct. Biol.*, 2016, **38**, 75–82.

27 A. Dutta, K. Saxena, H. Schwalbe and J. Klein-Seetharaman, *Methods Mol. Biol.*, 2012, **831**, 55–69.

28 S. Sobhanifar, S. Reckel, F. Junge, D. Schwarz, L. Kai, M. Karbyshev, F. Lohr, F. Bernhard and V. Dotsch, *J. Biomol. NMR*, 2010, **46**, 33–43.

29 G. Rosenblum and B. S. Cooperman, *FEBS Lett.*, 2014, **588**, 261–268.

30 F. W. Studier, *Protein Expression Purif.*, 2005, **41**, 207–234.

31 A. P. Duff, K. L. Wilde, A. Rekas, V. Lake and P. J. Holden, *Methods Enzymol.*, 2015, **565**, 3–25.

32 M. Cai, Y. Huang, J. Lloyd, R. Craigie and G. M. Clore, *J. Biomol. NMR*, 2021, **75**, 83–87.

33 T. Kigawa, Y. Muto and S. Yokoyama, *J. Biomol. NMR*, 1995, **6**, 129–134.

34 R. Linser, V. Gelev, F. Hagn, H. Arthanari, S. G. Hyberts and G. Wagner, *J. Am. Chem. Soc.*, 2014, **136**, 11308–11310.

35 M. Kainoshio, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono and P. Guntert, *Nature*, 2006, **440**, 52–57.

36 S. Wiesner and R. Sprangers, *Curr. Opin. Struct. Biol.*, 2015, **35**, 60–67.

37 V. Kurauskas, P. Schanda and R. Sounier, *Methods Mol. Biol.*, 2017, **1635**, 109–123.

38 K. Takeuchi, E. Ng, T. J. Malia and G. Wagner, *J. Biomol. NMR*, 2007, **38**, 89–98.

39 R. Kerfah, M. J. Plevin, R. Sounier, P. Gans and J. Boisbouvier, *Curr. Opin. Struct. Biol.*, 2015, **32**, 113–122.

40 K. H. Gardner and L. E. Kay, *J. Am. Chem. Soc.*, 1997, **119**, 7599–7600.

41 P. Gans, O. Hamelin, R. Sounier, I. Ayala, M. A. Dura, C. D. Amero, M. Noirclerc-Savoye, B. Franzetti, M. J. Plevin and J. Boisbouvier, *Angew. Chem., Int. Ed.*, 2010, **49**, 1958–1962.

42 R. Lichtenecker, J. Schörghuber and M. Bisaccia, *Synlett*, 2015, 2611–2616.

43 T. Sugiki, K. Furuita, T. Fujiwara and C. Kojima, *Biochemistry*, 2018, **57**, 3576–3589.

44 B. Krishnarjuna, G. Jaipuria, A. Thakur, P. D'Silva and H. S. Atreya, *J. Biomol. NMR*, 2011, **49**, 39–51.

45 D. H. Jones, S. E. Cellitti, X. Hao, Q. Zhang, M. Jahnz, D. Summerer, P. G. Schultz, T. Uno and B. H. Geierstanger, *J. Biomol. NMR*, 2010, **46**, 89–100.

46 G. T. Debelouchina and T. W. Muir, *Q. Rev. Biophys.*, 2017, **50**, e7.

47 A. C. Conibear, E. E. Watson, R. J. Payne and C. F. W. Becker, *Chem. Soc. Rev.*, 2018, **47**, 9046–9068.

48 D. Liu, R. Xu and D. Cowburn, *Methods Enzymol.*, 2009, **462**, 151–175.

49 L. Skrisovska, M. Schubert and F. H. Allain, *J. Biomol. NMR*, 2010, **46**, 51–65.

50 E. Michel and F. H. Allain, *Methods Enzymol.*, 2015, **565**, 389–422.

51 P. Dawson, T. Muir, I. Clark-Lewis and S. Kent, *Science*, 1994, **266**, 776–779.

52 J. W. Bode, *Acc. Chem. Res.*, 2017, **50**, 2104–2115.

53 H. Liu and X. Li, *Acc. Chem. Res.*, 2018, **51**, 1643–1655.

54 S. Bondalapati, M. Jbara and A. Brik, *Nat. Chem.*, 2016, **8**, 407–418.

55 V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu and O. Melnyk, *Chem. Rev.*, 2019, **119**, 7328–7443.

56 R. E. Thompson and T. W. Muir, *Chem. Rev.*, 2020, **120**, 3051–3126.

57 G. G. Kochendoerfer, D. H. Jones, S. Lee, M. Oblatt-Montal, S. J. Opella and M. Montal, *J. Am. Chem. Soc.*, 2004, **126**, 2439–2446.

58 M. Izumi, R. Kuruma, R. Okamoto, A. Seko, Y. Ito and Y. Kajihara, *J. Am. Chem. Soc.*, 2017, **139**, 11421–11426.

59 M. González, *Viruses*, 2015, **7**, 4352–4368.

60 C. Unverzagt and Y. Kajihara, *Chem. Soc. Rev.*, 2013, **42**, 4408–4420.

61 M. Schubert, M. J. Walczak, M. Aebl and G. Wider, *Angew. Chem., Int. Ed.*, 2015, **54**, 7096–7100.

62 P. B. Dawadi and J. Lugtenburg, *Molecules*, 2013, **18**, 482–519.

63 K. J. Rothschild, *Biomed. Spectrosc. Imaging*, 2016, **5**, 231–267.

64 Y. Xiao, M. S. Hutson, M. Belenky, J. Herzfeld and M. S. Braiman, *Biochemistry*, 2004, **43**, 12809–12818.

65 D. A. Erlanson, M. Chytil and G. L. Verdine, *Chem. Biol.*, 1996, **3**, 981–991.

66 T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710.

67 T. C. Evans Jr, J. Benner and M.-Q. Xu, *Protein Sci.*, 1998, **7**, 2256–2264.

68 R. Xu, B. Ayers, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 388.

69 P. Kane, C. Yamashiro, D. Wolczyk, N. Neff, M. Goebel and T. Stevens, *Science*, 1990, **250**, 651–657.

70 M.-Q. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra and F. B. Perler, *Cell*, 1993, **75**, 1371–1377.

71 N. H. Shah and T. W. Muir, *Chem. Sci.*, 2014, **5**, 446–461.

72 M. W. Southworth, K. Amaya, T. C. Evans, M.-Q. Xu and F. B. Perler, *Biotechniques*, 1999, **27**, 110–120.

73 J. Zettler, V. Schutz and H. D. Mootz, *FEBS Lett.*, 2009, **583**, 909–914.

74 N. H. Shah, G. P. Dann, M. Vila-Perello, Z. Liu and T. W. Muir, *J. Am. Chem. Soc.*, 2012, **134**, 11338–11341.

75 M. Hejjaoui, S. Butterfield, B. Fauvet, F. Vercruyse, J. Cui, I. Dikiy, M. Prudent, D. Olschewski, Y. Zhang, D. Eliezer and H. A. Lashuel, *J. Am. Chem. Soc.*, 2012, **134**, 5196–5210.

76 B. Pan, J. H. Park, T. Ramlall, D. Eliezer, E. Rhoades and E. J. Petersson, *ChemBioChem*, 2021, **22**, 1440–1447.

77 H. A. Lashuel, C. R. Overk, A. Oueslati and E. Masliah, *Nat. Rev. Neurosci.*, 2013, **14**, 38–48.

78 S. Fahs, P. Lujan and M. Kohn, *ACS Chem. Biol.*, 2016, **11**, 2944–2961.

79 F. X. Theillet, H. M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuiver and P. Selenko, *Nat. Protoc.*, 2013, **8**, 1416–1432.

80 A. C. Conibear, K. J. Rosengren, C. F. W. Becker and H. Kaehlig, *J. Biomol. NMR*, 2019, **73**, 587–599.

81 C. F. Becker, K. Lausecker, M. Balog, T. Kalai, K. Hideg, H. J. Steinhoff and M. Engelhard, *Magn. Reson. Chem.*, 2005, **43**, S34–39.

82 G. E. Fanucci and D. S. Cafiso, *Curr. Opin. Struct. Biol.*, 2006, **16**, 644–653.

83 N. Chu, T. Viennet, H. Bae, A. Salguero, A. Boeszoermenyi, H. Arthanari and P. A. Cole, *eLife*, 2020, **9**, e59151.

84 G. Niederacher, D. Urwin, Y. Dijkwel, D. J. Tremethick, K. J. Rosengren, C. F. W. Becker and A. C. Conibear, *RSC Chem. Biol.*, 2021, **2**, 537–550.

85 C. A. Castaneda, L. Spasser, S. N. Bavikar, A. Brik and D. Fushman, *Angew. Chem., Int. Ed.*, 2011, **50**, 11210–11214.

86 B. D. Manning and A. Toker, *Cell*, 2017, **169**, 381–405.

87 H. Takahashi, T. Nakanishi, K. Kami, Y. Arata and I. Shimada, *Nat. Struct. Biol.*, 2000, **7**, 220–223.

88 N. Chu, A. L. Salguero, A. Z. Liu, Z. Chen, D. R. Dempsey, S. B. Ficarro, W. M. Alexander, J. A. Marto, Y. Li, L. M. Amzel, S. B. Gabelli and P. A. Cole, *Cell*, 2018, **174**, 897–907.e814.

89 Z. A. Wang and P. A. Cole, *Cell Chem. Biol.*, 2020, **27**, 953–969.

90 A. Dose, S. Liokatis, F. X. Theillet, P. Selenko and D. Schwarzer, *ACS Chem. Biol.*, 2011, **6**, 419–424.

91 H. J. Dyson and P. E. Wright, *J. Biomol. NMR*, 2019, **73**, 651–659.

92 W. Gui, G. A. Davidson and Z. Zhuang, *RSC Chem. Biol.*, 2021, **2**, 450–467.

93 V. Ranjani, M. Assfalg and D. Fushman, *Methods in Enzymology*, Academic Press, 2005, vol. 399, pp. 177–192.

94 K. S. Kumar, L. Spasser, L. A. Erlich, S. N. Bavikar and A. Brik, *Angew. Chem., Int. Ed.*, 2010, **49**, 9126–9131.

95 S. D. Moran, A. M. Woys, L. E. Buchanan, E. Bixby, S. M. Decatur and M. T. Zanni, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 3329–3334.

96 Z. Jiang, F. Heinrich, R. P. McGlinchey, J. M. Gruschus and J. C. Lee, *J. Phys. Chem. Lett.*, 2017, **8**, 29–34.

97 J. D. Flynn, Z. Jiang and J. C. Lee, *Angew. Chem., Int. Ed.*, 2018, **57**, 17069–17072.

98 H. D. Mootz, *ChemBioChem*, 2009, **10**, 2579–2589.

99 N. H. Shah, E. Eryilmaz, D. Cowburn and T. W. Muir, *J. Am. Chem. Soc.*, 2013, **135**, 18673–18681.

100 I. V. Thiel, G. Volkmann, S. Pietrokovski and H. D. Mootz, *Angew. Chem., Int. Ed.*, 2014, **53**, 1306–1310.

101 A. S. Aranko, A. Wlodawer and H. Iwai, *Protein Eng., Des. Sel.*, 2014, **27**, 263–271.

102 B. Di Ventura and H. D. Mootz, *Biol. Chem.*, 2019, **400**, 467–475.

103 Y. Shiraishi, M. Natsume, Y. Kofuku, S. Imai, K. Nakata, T. Mizukoshi, T. Ueda, H. Iwai and I. Shimada, *Nat. Commun.*, 2018, **9**, 194.

104 Y. Kofuku, T. Ueda, J. Okude, Y. Shiraishi, K. Kondo, T. Mizumura, S. Suzuki and I. Shimada, *Angew. Chem., Int. Ed.*, 2014, **53**, 13376–13379.

105 I. Giriati and T. W. Muir, *J. Am. Chem. Soc.*, 2003, **125**, 7180–7181.

106 Y. David, M. Vila-Perello, S. Verma and T. W. Muir, *Nat. Chem.*, 2015, **7**, 394–402.

107 M. Muona, A. S. Aranko, V. Raulinaitis and H. Iwai, *Nat. Protoc.*, 2010, **5**, 574–587.

108 H. Mao, S. A. Hart, A. Schink and B. A. Pollok, *J. Am. Chem. Soc.*, 2004, **126**, 2670–2671.

109 J. M. Antos, M. C. Truttmann and H. L. Ploegh, *Curr. Opin. Struct. Biol.*, 2016, **38**, 111–118.

110 T. Aizu, T. Suzuki, A. Kido, K. Nagai, A. Kobayashi, R. Sugiura, Y. Ito and M. Mishima, *Biochim. Biophys. Acta, Gen. Subj.*, 2020, **1864**, 129419.

111 Y. Kobashigawa, H. Kumeta, K. Ogura and F. Inagaki, *J. Biomol. NMR*, 2009, **43**, 145.

112 L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. Zou and M. Sattler, *J. Biomol. NMR*, 2015, **63**, 1–8.

113 B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof and D. R. Liu, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 13343–13348.

114 R. L. Policarpo, H. Kang, X. Liao, A. E. Rabideau, M. D. Simon and B. L. Pentelute, *Angew. Chem., Int. Ed.*, 2014, **53**, 9203–9208.

115 M. Schmidt, A. Toplak, P. J. Quaedflieg and T. Nuijens, *Curr. Opin. Chem. Biol.*, 2017, **38**, 1–7.

116 C. Freund and D. Schwarzer, *ChemBioChem*, 2021, **22**, 1347–1356.

117 H. Ke, S. Matsumoto, Y. Murashima, H. Taniguchi-Tamura, R. Miyamoto, Y. Yoshikawa, C. Tsuda, T. Kumasaka, E. Mizohata, H. Edamatsu and T. Kataoka, *FEBS Lett.*, 2017, **591**, 2470–2481.

118 M. Sonntag, P. K. A. Jagtap, B. Simon, M. S. Appavou, A. Geerlof, R. Stehle, F. Gabel, J. Hennig and M. Sattler, *Angew. Chem., Int. Ed.*, 2017, **56**, 9322–9325.

119 H. Jiang, X. Zhang, X. Chen, P. Aramsangtienchai, Z. Tong and H. Lin, *Chem. Rev.*, 2018, **118**, 919–988.

120 K. Tripsianes, U. Schutz, L. Emmanouilidis, G. Gemmecker and M. Sattler, *J. Biomol. NMR*, 2019, **73**, 183–189.

121 I. Ahearn, M. Zhou and M. R. Philips, *Cold Spring Harbor Perspect. Med.*, 2018, **8**, a031484.

122 P. A. Konstantopoulos, M. V. Karamouzis and A. G. Papavassiliou, *Nat. Rev. Drug Discovery*, 2007, **6**, 541–555.

123 K. Kuhn, D. J. Owen, B. Bader, A. Wittinghofer, J. Kuhlmann and H. Waldmann, *J. Am. Chem. Soc.*, 2001, **123**, 1023–1035.



124 O. Rocks, A. Peyker, M. Kahms, P. J. Verveer, C. Koerner, M. Lumbierres, J. Kuhlmann, H. Waldmann, A. Wittinghofer and P. I. H. Bastiaens, *Science*, 2005, **307**, 1746.

125 J. Li, Y. Zhang, O. Soubias, D. Khago, F. A. Chao, Y. Li, K. Shaw and R. A. Byrd, *J. Biol. Chem.*, 2020, **295**, 2664–2675.

126 K. Weise, S. Kapoor, C. Denter, J. Nikolaus, N. Opitz, S. Koch, G. Triola, A. Herrmann, H. Waldmann and R. Winter, *J. Am. Chem. Soc.*, 2011, **133**, 880–887.

127 I. Wang, J. Hennig, P. K. Jagtap, M. Sonntag, J. Valcarcel and M. Sattler, *Nucleic Acids Res.*, 2014, **42**, 5949–5966.

128 G. K. Nguyen, S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam, *Nat. Chem. Biol.*, 2014, **10**, 732–738.

129 M. A. Jackson, L. T. T. Nguyen, E. K. Gilding, T. Durek and D. J. Craik, *Biotechnol. Adv.*, 2020, **45**, 107651.

130 T. Nuijens, A. Toplak, M. Schmidt, A. Ricci and W. Cabri, *Front. Chem.*, 2019, **7**, 829.

131 G. K. Nguyen, Y. Qiu, Y. Cao, X. Hemu, C. F. Liu and J. P. Tam, *Nat. Protoc.*, 2016, **11**, 1977–1988.

132 G. K. Oguis, E. K. Gilding, M. A. Jackson and D. J. Craik, *Front. Plant Sci.*, 2019, **10**, 645.

133 X. Hemu, A. El Sahili, S. Hu, X. Zhang, A. Serra, B. C. Goh, D. A. Darwis, M. W. Chen, S. K. Sze, C.-F. Liu, J. Lescar and J. P. Tam, *ACS Catal.*, 2020, **10**, 8825–8834.

134 K. S. Harris, T. Durek, Q. Kaas, A. G. Poth, E. K. Gilding, B. F. Conlan, I. Saska, N. L. Daly, N. L. van der Weerden, D. J. Craik and M. A. Anderson, *Nat. Commun.*, 2015, **6**, 10199.

135 F. B. H. Rehm, T. J. Harmand, K. Yap, T. Durek, D. J. Craik and H. L. Ploegh, *J. Am. Chem. Soc.*, 2019, **141**, 17388–17393.

136 K. M. Mikula, I. Tascon, J. J. Tommila and H. Iwai, *FEBS Lett.*, 2017, **591**, 1285–1294.

137 K. M. Mikula, L. Krumwiede, A. Pluckthun and H. Iwai, *J. Biomol. NMR*, 2018, **71**, 225–235.

138 S. Liebscher, P. Kornberger, G. Fink, E. M. Trost-Gross, E. Hoss, A. Skerra and F. Bordusa, *ChemBioChem*, 2014, **15**, 1096–1100.

139 S. Liebscher, M. Schopfel, T. Aumuller, A. Sharkhuukhen, A. Pech, E. Hoss, C. Parthier, G. Jahreis, M. T. Stubbs and F. Bordusa, *Angew. Chem., Int. Ed.*, 2014, **53**, 3024–3028.

140 J. A. Wells, E. Ferrari, D. J. Henner, D. A. Estell and E. Y. Chen, *Nucleic Acids Res.*, 1983, **11**, 7911–7925.

141 A. M. Weeks and J. A. Wells, *Nat. Chem. Biol.*, 2018, **14**, 50–57.

142 A. Toplak, T. Nuijens, P. J. L. M. Quaedflieg, B. Wu and D. B. Janssen, *Adv. Synth. Catal.*, 2016, **358**, 2140–2147.

143 M. Schmidt, A. Toplak, P. J. L. M. Quaedflieg, H. Ippel, G. J. J. Richelle, T. M. Hackeng, J. H. van Maarseveen and T. Nuijens, *Adv. Synth. Catal.*, 2017, **359**, 2050–2055.

144 S. H. Henager, N. Chu, Z. Chen, D. Bolduc, D. R. Dempsey, Y. Hwang, J. Wells and P. A. Cole, *Nat. Methods*, 2016, **13**, 925–927.

145 A. M. Weeks and J. A. Wells, *Chem. Rev.*, 2020, **120**, 3127–3160.

146 V. Slynko, M. Schubert, S. Numao, M. Kowarik, M. Aebi and F. H. T. Allain, *J. Am. Chem. Soc.*, 2009, **131**, 1274–1281.

147 S. Liokatis, A. Dose, D. Schwarzer and P. Selenko, *J. Am. Chem. Soc.*, 2010, **132**, 14704–14705.

148 N. Ait-Bouziad, A. Chiki, G. Limorenko, S. Xiao, D. Eliezer and H. A. Lashuel, *J. Biol. Chem.*, 2020, **295**, 7905–7922.

149 C. P. Hackenberger and D. Schwarzer, *Angew. Chem., Int. Ed.*, 2008, **47**, 10030–10074.

150 E. M. Sletten and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2009, **48**, 6974–6998.

151 O. Boutureira and G. J. Bernardes, *Chem. Rev.*, 2015, **115**, 2174–2195.

152 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004–2021.

153 M. Meldal and C. W. Tornøe, *Chem. Rev.*, 2008, **108**, 2952–3015.

154 K. Lang and J. W. Chin, *Chem. Rev.*, 2014, **114**, 4764–4806.

155 T. Schneider, A. Berg, Z. Ulusoy, M. Gamerdinger, C. Peter and M. Kovermann, *Sci. Rep.*, 2019, **9**, 19991.

156 C. A. Castaneda, E. K. Dixon, O. Walker, A. Chaturvedi, M. A. Nakasone, J. E. Curtis, M. R. Reed, S. Krueger, T. A. Cropp and D. Fushman, *Structure*, 2016, **24**, 423–436.

157 C. Chatterjee, R. K. McGinty, B. Fierz and T. W. Muir, *Nat. Chem. Biol.*, 2010, **6**, 267–269.

158 J. Chen, Y. Ai, J. Wang, L. Haracska and Z. Zhuang, *Nat. Chem. Biol.*, 2010, **6**, 270–272.

159 G. T. Debelouchina, K. Gerecht and T. W. Muir, *Nat. Chem. Biol.*, 2017, **13**, 105–110.

160 Y. Hattori, K. Furuita, I. Ohki, T. Ikegami, H. Fukada, M. Shirakawa, T. Fujiwara and C. Kojima, *J. Biomol. NMR*, 2013, **55**, 19–31.

161 S. Liokatis, R. Klingberg, S. Tan and D. Schwarzer, *Angew. Chem., Int. Ed.*, 2016, **55**, 8262–8265.

162 S. Liokatis, A. Stutzer, S. J. Elsasser, F. X. Theillet, R. Klingberg, B. van Rossum, D. Schwarzer, C. D. Allis, W. Fischle and P. Selenko, *Nat. Struct. Mol. Biol.*, 2012, **19**, 819–823.

163 F. X. Theillet, S. Liokatis, J. O. Jost, B. Bekei, H. M. Rose, A. Binolfi, D. Schwarzer and P. Selenko, *J. Am. Chem. Soc.*, 2012, **134**, 7616–7619.

164 C. C. Lechner, N. D. Agashe and B. Fierz, *Angew. Chem., Int. Ed.*, 2016, **55**, 2903–2906.

165 C. L. van Emmerik and H. van Ingen, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2019, **110**, 1–19.

166 J. Nicolas Carcelen, J. M. Marchante-Gayon, P. R. Gonzalez, L. Valledor, M. J. Canal and J. I. G. Alonso, *Microb. Cell Fact.*, 2017, **16**, 146.

167 K. Tripsianes, N. K. Chu, A. Friberg, M. Sattler and C. F. Becker, *ACS Chem. Biol.*, 2014, **9**, 347–352.

168 S. S. Kulkarni, E. E. Watson, B. Premdjee, K. W. Condefrietboes and R. J. Payne, *Nat. Protoc.*, 2019, **14**, 2229–2257.

169 J. B. Li, S. Tang, J. S. Zheng, C. L. Tian and L. Liu, *Acc. Chem. Res.*, 2017, **50**, 1143–1153.

170 K. K. Frederick, V. K. Michaelis, M. A. Caporini, L. B. Andreas, G. T. Debelouchina, R. G. Griffin and S. Lindquist, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 3642–3647.

171 S. Jannin, J. N. Dumez, P. Giraudeau and D. Kurzbach, *J. Magn. Reson.*, 2019, **305**, 41–50.

172 A. Singh, *Nat. Methods*, 2019, **16**, 676.

