Check for updates

Posttranslational, site-directed photochemical fluorine editing of protein sidechains to probe residue oxidation state via ^{19}F-nuclear magnetic resonance

Patrick G. Isenegger[1,5], Brian Josephson[1,5], Ben Gaunt [2], Matthew J. Davy[2,3], Veronique Gouverneur [1], Andrew J. Baldwin[1,2,4✉] and Benjamin G. Davis [1,2,3✉]

The fluorination of amino acid residues represents a near-isosteric alteration with the potential to report on biological pathways, yet the site-directed editing of carbon–hydrogen (C–H) bonds in complex biomolecules to carbon–fluorine (C–F) bonds is challenging, resulting in its limited exploitation. Here, we describe a protocol for the posttranslational and site-directed alteration of native γCH_2 to γCF_2 in protein sidechains. This alteration allows the installation of difluorinated sidechain analogs of proteinogenic amino acids, in both native and modified states. This chemical editing is robust, mild, fast and highly efficient, exploiting photochemical- and radical-mediated C–C bonds grafted onto easy-to-access cysteine-derived dehydroalanine-containing proteins as starting materials. The heteroaryl–sulfonyl reagent required for generating the key carbon-centered C• radicals that install the sidechain can be synthesized in two to six steps from commercially available precursors. This workflow allows the nonexpert to create fluorinated proteins within 24 h, starting from a corresponding purified cysteine-containing protein precursor, without the need for bespoke biological systems. As an example, we readily introduce three γCF_2-containing methionines in all three progressive oxidation states (sulfide, sulfoxide and sulfone) as D-/L- forms into histone eH3.1 at site 4 (a relevant lysine to methionine oncomutation site), and each can be detected by ^{19}F-nuclear magnetic resonance of the γCF_2 group, as well as the two diastereomers of the sulfoxide, even when found in a complex protein mixture of all three. The site-directed editing of C–H→C–F enables the use of γCF_2 as a highly sensitive, 'zero-size-zero-background' label in protein sidechains, which may be used to probe biological phenomena, protein structures and/or protein–ligand interactions by ^{19}F-based detection methods.

Introduction

Posttranslational modifications (PTMs) add diversity to protein structure and function beyond the chemical and physical properties of the canonical, proteinogenic amino acid sidechains[1]. As archetypes of substrates that are altered in this way, histones—the protein components of chromatin—can be heavily and diversely modified, with new modifications being discovered and characterized every year[2,3]. The associated in vivo installation ('writing'), interaction ('reading') and removal ('erasing') of a plethora of chemical functional groups in PTMs onto native histone sidechains is ubiquitous and crucial in regulating gene expression in response to intracellular signalling cascades, and for mitigating the effects of altering environment, such as DNA damage or oxidative stress[2].

Unsurprisingly, therefore, the field of chemical biology has developed several strategies for introducing desired PTMs site specifically into relevant full-length proteins, such as histones, providing valuable tools and materials for deciphering the roles of specific PTMs[4]. One strategy, amber codon suppression, exploits the direct introduction of noncanonical amino acids (ncAAs) in place of a reassigned stop codon[5]. This has several advantages, such as the capability of creating the desired modified proteins in living systems that can take up the required ncAAs (or precursors) and that carry the required corresponding aminoacyl-tRNA synthetase/tRNA pairs to process them. However, in some cases protein yields and incorporation efficiencies can be low, and these bespoke systems can be limited in their plasticity and therefore generality. Several relevant histone PTMs have proven

[1]Department of Chemistry, University of Oxford, Oxford, UK. [2]The Rosalind Franklin Institute, Oxfordshire, UK. [3]Department of Pharmacology, University of Oxford, Oxford, UK. [4]Kavli Institute for Nanoscience Discovery, University of Oxford, Oxford, UK. [5]These authors contributed equally: Patrick G. Isenegger, Brian Josephson. ✉e-mail: andrew.baldwin@chem.ox.ac.uk; ben.davis@rfi.ac.uk

challenging[6]. For example, Lys methylation has yet to be directly installed but can be accessed via subsequent chemical means[7], and Arg methylation has only been accessed at a low level for the case of N_ω-methyl-L-arginine[8]. Other PTMs such as Asn, Ser or Thr glycosylation, have yet to be realized with this technique[5]. In addition, many ncAAs of interest are not commercially available and therefore often a multistep synthetic process is required to yield the desired ncAAs beforehand. Another strategy, exemplified primarily by thioester-mediated native chemical ligation or expressed protein ligation, can produce full-length histone proteins with a wider range of relevant PTMs[9,10] when paired with peptide synthesis to introduce a site-specifically modified segment via 'linear', backbone–amide ligation. However, production of full-length proteins in this way may be time intensive and necessitates peptide chemistry knowledge.

A third strategy is via 'convergent', sidechain ligation using the chemical modification of an expressed (e.g., recombinant) full-length histone substrate. This may be achieved, in principle, either through selective reactions on specific native amino acids of interest (akin to posttranslational modification in a residue- and/or regio-selective manner) or through a 'tag-and modify' approach, which exploits the unique reactivity (i.e., chemoselectivity) of a site-specific chemical 'tag'. When chosen correctly, this 'tag' can be chemoselectively 'modified' in the presence of other canonical amino acids[4]. Ideally, even if the latter does not exploit the same bond-forming process as natural PTMs, this will either generate a native PTM architecture or a functional mimic[4,11,12]. Given this requirement and the often small size of observed PTMs (methylation, acetylation, oxidation), many typical bioconjugation methods that use bulky linkers (including most 'CLICK' reactions)[13] fall short of the (near) 'scarless' ligation needed to recapitulate constitution or function.

Our group, as well as others, have exploited the use of the electrophile and SOMOphile dehydroalanine (Dha) as a 'tag' to introduce various natural, unnatural and modified sidechains into proteins[11,14–16]. The Dha residue is essentially a Cα=Cβ-double bond sidechain 'stump' that may be 'grafted onto' when it is located within proteins using various methods that make Cβ–Xγ bonds (where X can represent B, C, N, O, P, S, Se, etc.)[17]. While multiple complementary methods exist for site-specific Dha incorporation into proteins (and are reviewed elsewhere[12]), histones are especially amenable to Dha chemistry as many (from various species) lack native Cys residues (except in the case of histone H3, where Cys96 and Cys110 can be mutated to Ala or Ser with little apparent consequence). This allows for the operationally simplistic formation of Dha in proteins via bis-alkylation/elimination with reagents such 2,5-dibromohexanediamide (DBHDA; Sigma-Aldrich, cat. no. 900607) on a Cys introduced via conventional mutagenesis to the desired 'tag/stump' site of interest. Recent work[18] reduces the level of 'scarring' (e.g., via Cβ–Sγ creation of thia–lysine analogs)[19] in previous Dha-modification strategies to create instead native Cβ–Cγ sidechain bonds using two complementary, mild, light-driven approaches for off-protein alkyl C• radical generation. These now allow >50 unique sidechains to be installed by grafting onto Dha residues in histones.

One of the light-driven methods utilizes fluoroalkylpyridylsulfone derivatives (termed pySOOF), in combination with photo-stimulated, outer-sphere single-electron transfer catalysts, such as Ru(bpy)₃Cl₂. This is potentiated by $FeSO_4$ to generate difluoroalkyl RCF_2• radicals in water that, once reacted with Dha, create novel γCF_2-containing protein sidechains via Cβ–CγF_2 bond-forming grafting. In comparison with previous methods[14–16], this process is more rapid (taking just 15 min) and widely tolerant of chemically-diverse sidechains containing functional groups such as halogens, azides, organosulfates, amides and esters (Fig. 1). It is also materially efficient, with the reaction cleanly proceeding to full conversion with only two to five equivalents (equiv.) of pySOOF reagent (in comparison with the greater than hundreds or thousands of equivalents required for some protein chemistries)[15] using 100–250 equiv. of $FeSO_4$ in combination with substoichiometric amounts of the Ru(bpy)₃Cl₂ 'photocatalyst'. Here, we describe a step-by-step workflow for its implementation.

Using this method, we have been able to readily generate γCF_2-containing sidechain analogs of relevant histone PTMs such as methylation, acetylation and oxidation on residues such as Lys, Arg and Met. However, whilst such generation of PTM-containing histones (using this method or others, see above) is valuable, it does not address a major problem in the field. In nature, the PTM status of individual histone sidechains is dynamic[20], with PTMs being installed and removed as needed to modulate the transcriptional landscape required for the cell. Tools for unambiguously detecting these minute molecular changes (in the case of Met oxidation, as small as a single oxygen atom to give Met(O)) in solution and in real time are extremely lacking. For example, the arguably two most common methods for detecting the PTM status of specific sidechain on histones (or any protein), liquid chromatography–tandem mass spectrometry (LC–MS/MS)[21] and antibody
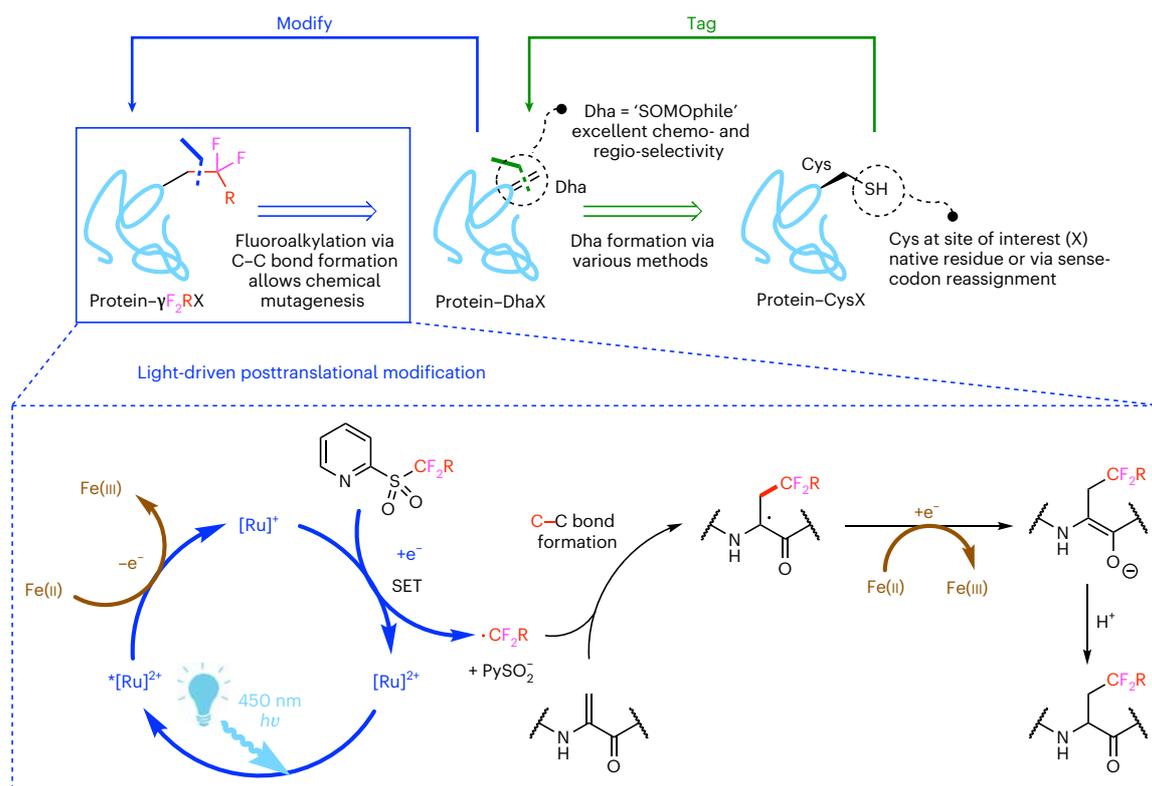
Fig. 1 | Strategy for the insertion of 'zero-size-zero-background' sidechain labels into proteins. When considered retrosynthetically (top, left-to-right arrows), γCF_2-containing sidechains can be installed (top, right-to-left arrows) site specifically into proteins by first mutating the desired residue to Cys (or by using a native Cys residue) in the protein scaffold and converting it to Dha 'tag' (green arrows) as a radical-acceptor SOMOphile. Dha has excellent chemoselectivity for the C•-centered difluoroalkylradical formed during light-driven generation. The resulting C–C bond creates the target protein product (blue arrows). The proposed mechanism (bottom) of the γCF_2R sidechain installation is shown. The C•-centered difluoroalkyl radical is formed via reductive initiation following the single-electron 'quenching' of the photocatalyst [Ru]$^+$. After C• radical addition to Dha, the corresponding α-carbon radical intermediate is reduced by iron(II) to form, after protonation, the desired fluorinated protein. Figure adapted with permission from ref. [18], Springer Nature Ltd.

binding[22] (e.g., chromatin immunoprecipitation or western blot analysis), are either destructive (in the case of LC–MS/MS) or static (i.e., trap/occlude the relevant PTM site in the case of antibodies). These therefore cannot be readily applied to proper dynamic study of epigenetic cycles. Furthermore, in both methods, when multivariate PTM complexity is present (such as when proximal residues Lys9 and Ser10 on histone H3 are trimethylated and phosphorylated, respectively—a strongly implicated combination), the modification of one residue can restrict or alter detection of the other[23].

An ideal chemical tool for detecting the PTM status of a specific sidechain would fulfil a series of requirements, such as being easy to access from, and introduce into, native histones, being amenable to a wide selection of native and modified protein sidechains, high sensitivity to changing local chemical environments (as caused by PTM processing), 'small' enough to avoid interfering with native PTM function (e.g., 'zero-size'), nondestructive, nonstatic detection in real time and with high signal-to-noise ratio (e.g., 'zero-background') in the presence of other biomolecules.

Here, we cover the use of γCF_2-bearing sidechains to fulfil these requirements when paired with solution ^{19}F nuclear magnetic resonance (^{19}F-NMR). This protocol describes their rapid installation onto Dha-containing histones in an operationally simple manner, utilizing minimal amounts of readily accessible pySOOF sidechain precursor reagents that encompass and access a plethora of relevant histone PTMs. ^{19}F is the third most receptive nuclear magnetic resonance (NMR) nucleus (only surpassed by ^{1}H and ^{3}H), highly sensitive to its environment with a large chemical shift dispersion and of a comparable size to a hydrogen atom[24]. Additionally, its near absence as an element in biology ensures an exemplary signal-to-noise ratio, even in complex biological environments. This approach, therefore, also leverages these qualities to unambiguously detect the modification states of various modified γCF_2-containing sidechains.

Development of the protocol

Strategies for the creation of the $C(sp^3)\beta$–$C(sp^3)\gamma$ bond in proteins, although proposed as a method for potential posttranslational mutagenesis[25], were not disclosed until 2016 when independent back-to-back studies identified C• radicals as biocompatible reagents that could react with Dha in proteins[14–16]. These were derived via reductive initiation from precursor alkyl halides, yet with limitations on substrate scope both with respect to sidechain and protein. They also necessitated hundreds or thousands of equivalents of sidechain precursor reagents[15] and could lead to by-product formation and/or protein degradation under certain conditions. For example, protein degradation can occur when using metals[16] or low-level reduction in some cases when using borohydride[15].

This protocol describes a recent, light-driven method that exploits photo-stimulated reductive initiation of heteroarylsufonyl pySOOF precursors using outer-sphere single electron transfer catalyst systems, potentiated by Fe(II) (ref. [18]). The resulting efficiencies and scope expand on prior C•-based methods, allowing unprecedented compatibilities and stoichiometries (2–5 equiv. pySOOF, sub-stoichiometric catalyst) exemplified in the protocols generating RCF_2• given here. Specifically, these protocols describe the installation of three different pySOOF-derived sidechains onto a Dha-containing histone H3 precursor protein (human variant histone H3.1, with Cys 96 and 110 mutated to Ala, and with dual FLAG-HA C-terminal epitope tags, called here histone eH3.1) to create systematically-varied oxidation states of Met that differ in only a single atom, thereby exemplifying both 'atom-precise' chemical protein editing and monitoring by ^{19}F-NMR.

Applications of the method

The pySOOF method has so far allowed for the installation of 30 γCF_2-containing sidechains (Fig. 2)[18]. We have shown that the γCF_2 group does not inhibit enzyme-mediated processing of corresponding modified sidechains (e.g., deacetylation of γF_2-AcLys, see below), and does not inhibit protein complex formation (e.g., in the case of histone octamer assembly). The wide range of available sidechains (Fig. 3) with the minimally perturbing but highly sensitive γCF_2 detector therefore suggests a powerful platform for assessing a plethora of PTM reactions and binding events.

Specifically, unmodified, acetylated and trimethylated γCF_2-Lys variants gave distinct ^{19}F-NMR spectra that suggested promise. The deacetylation of histone H3-γF_2-AcLys18 by Sirt2 was uninhibited, showing that the γCF_2 group could precisely report the PTM state of a specific residue while it was being processed by epigenetic enzymes. Furthermore, the γCF_2 group was sensitive enough to track the formation of submilligram amounts of a histone octamer containing histone H3 difluoroethylglycine 9 (H3-γF_2-EtGly9 = H3-DfeGly9) (ref. [18]).

For the additional examples given in this protocol, we have tested the ability to easily differentiate between the oxidation states of methionine using the distinct resonances that are observed in ^{19}F-NMR, even variations that differ by only a single atom Met→Met(O)→Met(O)$_2$. The method allows oxidation states to be precisely and clearly observed, and even distinguishes the different configurations of Met(O) sulfoxide. Given that Met oxidation is a well-known PTM that reports on and/or mediates redox processes in cells, with dedicated enzymes for example of the reduction of each distinct sulfoxide enantiomer[26], reporters for oxidized methionines provide a tool for the study of (sometimes dynamic) oxidative alterations in proteins. Furthermore, Lys to Met mutations are common in histones and can lead to cancer, earning them the nickname of 'oncohistone' mutations[27]. The site-specific H3-Lys4 to H3-Met4 mutational example used in this protocol is itself an oncohistone mutation[28]. Relatively little research[29] has gone into studying whether the oxidation state of these important cancer-inducing Met mutations has a functional consequence, perhaps due to lack of tools to study it.

The resulting γCF_2-containing Met sidechains, installed at residue 4 (natively a Lys residue) on human histone eH3.1, simply bear γCF_2 instead of $\gamma CH2$ and so are analogs of three different oxidation states of methionine (sulfide Met, sulfoxide Met(O) and sulfone Met(O)$_2$). Details are given on how to characterize and troubleshoot the pySOOF reaction, and prepare the resulting products for ^{19}F-NMR measurements (Fig. 4). Spectra for each γCF_2-Met analog are provided along with a method for analyzing and fitting the measured spectra (Fig. 5 and Table 1).

Since our initial suggestion of the potential of γCF_2-amino acid analogs in proteins[18], these have been exploited in peptides to detect lysine methylation and acetylation[30], but have yet to be utilized in full-length proteins. We anticipate their increasing use for directly detecting the dynamic nature of histone PTMs as they are written, read and erased by various epigenetic processes, particularly in cases where the regions are disordered.
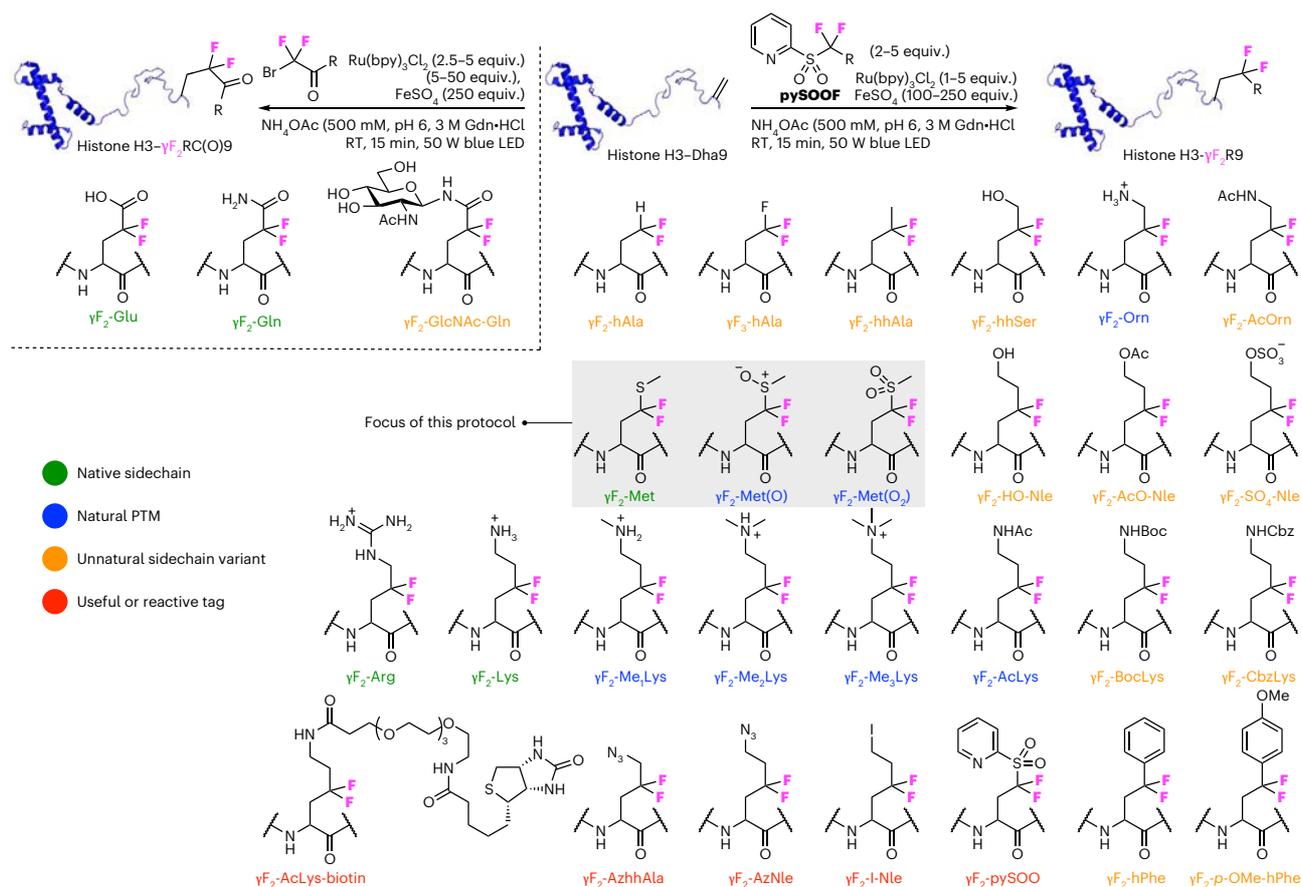
Fig. 2 | The range of γCF₂-containing sidechains that can be introduced by C–C-bond formation. After optimization of the reaction conditions, the range of the light-driven posttranslational installation method was investigated using histone H3-Dha9 as model protein[18]. During the course of this investigation, 30 γCF₂-containing sidechains were successfully installed. For the installation of Glu, Gln and GlcNAc–Gln sidechains, bromo-difluoro-carbonyl reagents were applied, with slightly higher reagent loading. By contrast, only 2–5 equiv. of pySOOF reagent were sufficient to yield the fluorinated protein with excellent conversion (>90%). Installed γCF₂-containing sidechains include native (green labels), posttranslationally modified (blue labels), unnatural (orange labels) and reactive chemical group-containing (red labels) sidechains, thereby allowing for a variety of site-specific studies into protein function. The current study focuses on Met pySOOF variants (highlighted in gray). Figure adapted with permission from ref. [18], Springer Nature Ltd.

Comparison with other methods

pySOOF allows the posttranslational installation of a wide variety of γF₂-labeled sidechains into proteins. A recent article describes a method to incorporate γF₂-labeled lysine sidechain variants into small peptides, with a similar goal of monitoring the lysine PTM status via ¹⁹F-NMR[30]. This could realistically be extended to the complete or semisynthesis of full-length proteins, but such endeavours can take months or years, and require expert peptide synthesis expertise and equipment[19]. pySOOF is fast, taking only hours or days, can be performed by a nonchemist and provides access to a wide range of native and modified γCF₂-containing sidechains. Compared with previous posttranslational Dha-modification chemistries[15,16], this method uses less material (2–5 equiv. pySOOF reagent versus hundreds or thousands) and is widely tolerant of a range of useful reactive and sensitive chemical groups due to its mild photochemical radical generation. Other methods such as amber codon suppression have successfully incorporated various unnatural or modified amino acids of choice, but such systems can struggle with lower protein yields or subtle alterations, such as H→F[5]. Furthermore, for the synthesis of fluorinated amino acids, complex synthetic methods or multistep synthesis may be required beforehand[31]. It is yet to be seen whether a requisite synthetase could be directed to selectively choose fluorinated amino acids on the basis of a small γCH₂→γCF₂ change, as may be required for specific introduction into protein sidechains.

In terms of methods to detect the PTM status of specific protein sidechains, the ¹⁹F-NMR method used here has many distinct advantages. Common methods such as western blotting or mass spectrometry can identify PTMs of interest, but are destructive techniques, and do not measure PTM
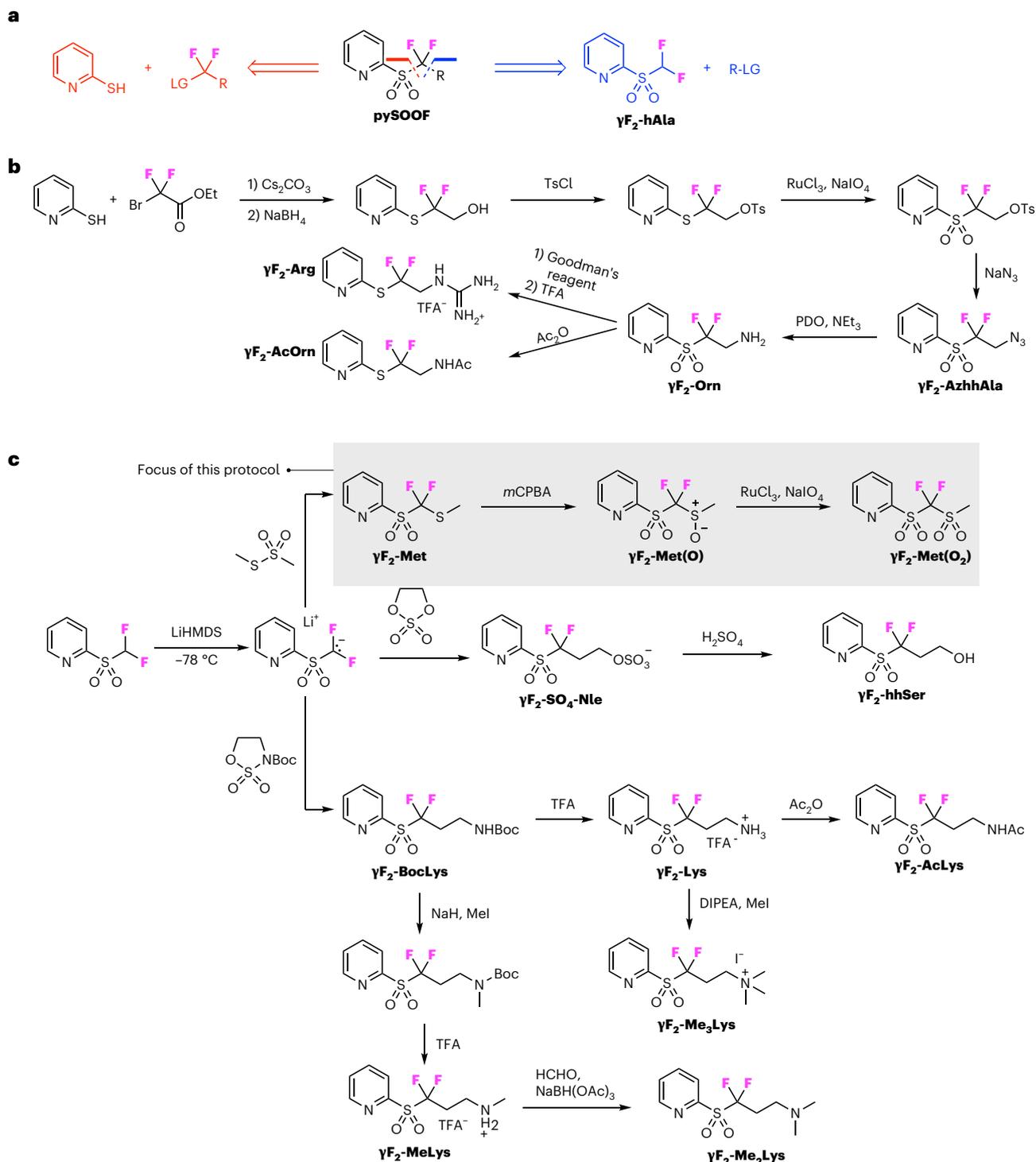
Fig. 3 | The synthesis of pySOOF reagents for C• radical generation in protein mutagenesis. a, Retrosynthetic analysis of pySOOF reagent that allows the synthesis of diverse sidechain precursor reagents. **b**, Examples of pySOOF syntheses that may be exploited via a 2-mercaptopyridine alkylation sequence. **c**, Examples of pySOOF syntheses via the alternative difluoromethyl 2-pyridyl sulfone alkylation sequence. The focus of this protocol is the creation of Met and modified-Met variants. The corresponding reagents for these processes are highlighted by the gray box and labels.

status in real time or in the desired native biological context[21,22]. Furthermore, antibodies used to detect PTMs can be unreliable in certain contexts, such as when nearby PTMs occlude the antibody's ability to bind and therefore detect its target. NMR, by contrast, is nondestructive, and can make measurements in real time and in complex solutions[24]. The use of ^{19}F-NMR in particular has the
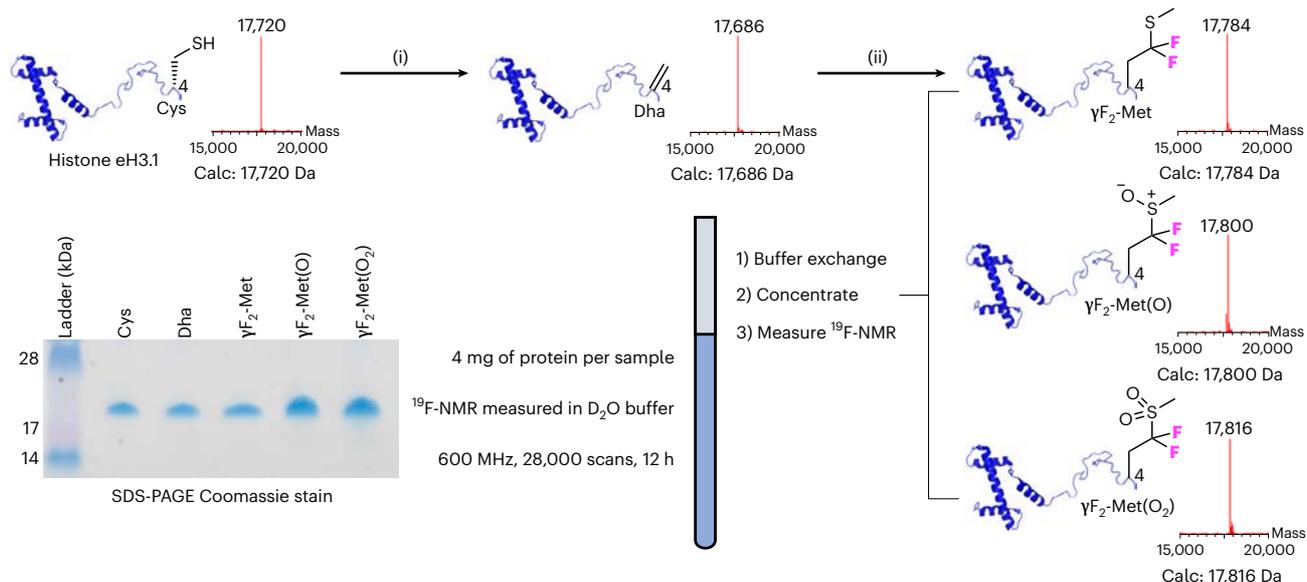
Fig. 4 | Production of histone eH3.1 with γCF₂-tagged Met variants (Met sulfide, Met(O) sulfoxide and Met(O₂) sulfone) for ¹⁹F-NMR studies. This was achieved through first mutating the site of interest (site 4) to Cys, converting Cys to Dha (i), then photocatalytically installing the γCF₂-containing sidechain (ii). Intact protein LC–MS analysis and SDS-PAGE reveal the expected masses and gel mobility in all cases, with each reaction going to near completion and without side products or protein damage. Reagents and conditions: (i) (1) DTT (3× wt/wt), 30 min, RT, (2) DBHDA (61 equiv.), RT 30 min then 37 °C 2 h; (ii) γCF₂Met-, γCF₂Met(O)- or γCF₂Met(O₂)-pySOOF reagent (2 equiv.), Ru(bpy)₃Cl₂ (1 equiv.), FeSO₄ (100 equiv.), blue LED (450 nm, 50 W), <6 ppm O₂, 15 min, RT. NMR experiments used ~4 mg of protein, equivalent to ~200 μM.

added benefit that fluorine is rare in biology, which combined with the advantageous NMR properties of the ^{19}F nuclei makes measurements of γCF$_2$ sidechains sensitive and with low background[24].

Experimental design

This protocol combines and exploits two complementary technologies in chemical biology (Fig. 1): recombinant protein expression (to obtain a target protein with a Cys-residue at the site of interest[32]) and posttranslational chemical protein mutagenesis (to modify the Cys and install the chosen γF$_2$-difluorinated sidechain[18]). With this procedure in hand, a formal mutation sequence of Xxx → Cys → Dha → F$_2$A (where Xxx denotes the wild-type residue at a given site and F$_2$A denotes any γF$_2$-difluorinated residue) is achieved, via a 'tag-and-modify' approach[33].

Therefore, it is important to chemo-selectively control the initial Cys site and maintain the fidelity for that site (via reactions of both Cys and Dha that do not perturb any other canonical residues). Therefore, the unique chemical properties of the noncanonical amino acid Dha prove important[12]. Once formed via the elimination of Cys, the olefinic carbon Cα=Cβ-double-bond 'stump' sidechain of Dha acts as a suitable radical acceptor for generated •CF$_2$R radicals to form a new hydrodifluorinated sidechain CHαCβH$_2$–CγF$_2$R after protonation of an ensuing enolate intermediate. For the generation of radical •CF$_2$R, a dual catalytic system is generated using an outer-sphere single-electron catalyst (a so-called photoredox catalyst) and iron(II), which are photostimulated by blue light (450 nm); this is a chemically mild process, compatible with aqueous conditions and protein stability. In combination with water-(semi)soluble pySOOF reagents and the •CF$_2$R precursor, this chemical transformation can be completed in essentially fully aqueous reaction media (≤1% dimethyl sulfoxide (DMSO)) and therefore can be applied to a wide variety of proteins.

Protein site selection

While here we exploit the use of a precursor Cys residue, the positioning of the Dha site in the protein scaffold can be achieved with various methods that are under genetic, chemical or biosynthetic control[12]. The strategic repositioning of the cysteine residue in a codon (Cys → Ser, enabling the use of the Cys codon for Dha positioning) then offers an opportunity for the installation of difluoro-labeled sites. It should also be noted that the method for Dha generation from Cys requires a free Cys (and so is compatible with disulfide-linked residues, which remain untouched). It can also be applied regioselectively to one of several Cys[34]. Certain requirements such as solvent accessibility of the site of interest and protein stability after site-directed mutagenesis may also influence the desired 'chemical mutagenesis'
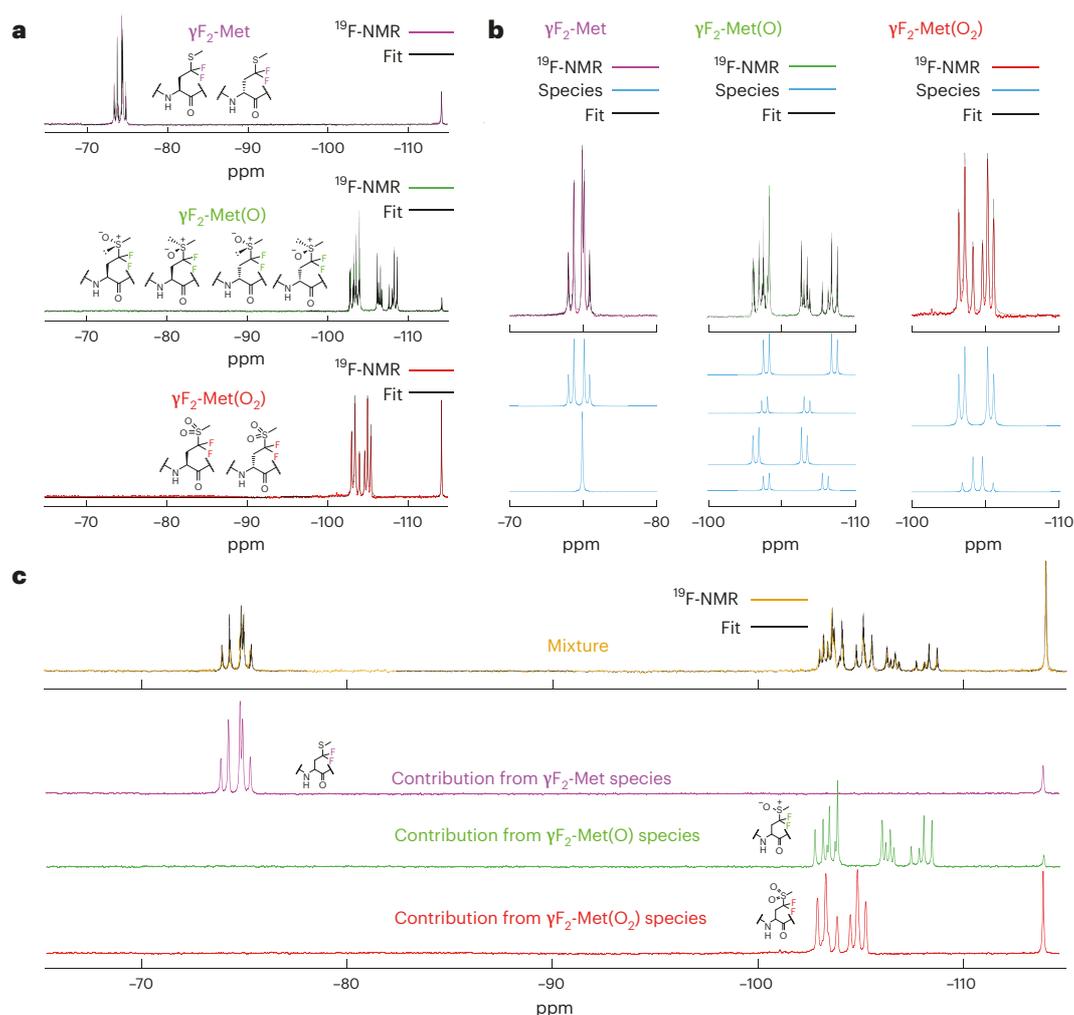
Fig. 5 | ^{19}F-NMR allows the distinction of labeled methionine diastereoisomers in the context of an intact protein. ^{19}F-NMR spectra of the three different oxidation states of γF$_2$-Met at site 4, here also obtained in the presence of Gd•HCl as a denaturant, with the structures of different species contributing to each spectrum shown. A fit is also indicated, with each spectrum having been fitted to a specified number of singlets and second-order doublets, with defined chemical shift, linewidth, intensity and $^2J_{FF}$. **a**, The spectra over a wide chemical shift range, and the expected species for γF$_2$-Met, γF$_2$-Met(O) and γF$_2$-Met(O$_2$), together with the corresponding modeled spectrum from our analysis. The singlet from fluoride at ~ −114 ppm can be seen in all spectra and was used to reference the spectra. We find that fluoride present in very small amounts can be used as this reference; we have used either residual from the reaction and/or added and/or spiked, giving rise to different concentrations as required. Top: the γF$_2$-Met spectrum was fitted to one pair of second-order doublets and one singlet (other than fluoride), corresponding to the two stereoisomers present. Middle: the γF$_2$-Met(O) spectrum was fitted to four pairs of second-order doublets, corresponding to the four stereoisomers and one singlet, describing fluoride. Bottom: the γF$_2$-Met(O$_2$) spectrum was fitted to two pairs of second-order doublets, corresponding to the two stereoisomers and one singlet, describing fluoride. **b**, The individual fitted contributions that compose the fit to the complete spectrum (**a**). While the number of species can be robustly identified via the fitting process, and these are shown to be consistent with the expected chemistry, we cannot assign a specific species to each fitted species in the spectrum. Full fitting parameters are available in Table 1. **c**, A complex mixture of the three different γF$_2$-Met oxidation states, with a fit to seven pairs of second-order doublets and one singlet (other than fluoride). Knowledge from **b** of the contributions of different stereoisomers of different γF$_2$-Met oxidation states allows determination of the relative contributions of each of the different oxidation states of γF$_2$-Met.

sequence Cys → Dha → F$_2$A on an intact and/or folded protein structure. Here, we use DBHDA as a commonly available reagent for generating Dha[35], but others can allow additional scope and control[34,36].

We have tested this method[18] by installing fluorinated sidechains at various positions in the following proteins: histones H3 and H4, single-domain antibody cAbLys3 and transmembrane bacterial efflux component protein AcrA, among others. Whilst these sites were chosen on the basis of their biological relevance, the method has also been tested and applied to sites in a coiled-coil domain of AcrA (site 123) and in a β-sheet region of the pentapeptide-repeat protein Npβ (site 61)[18]. It should also be noted that the 'largest'/most hindered reagents in this protocol (the Ru(bpy)$_3$ photocatalyst and pySOOF radical precursor) react in solution ('off-protein') to generate a much smaller C•-centered radical sidechain precursor that then reacts with Dha. Suitable reaction buffers at various pHs were ammonium acetate and sodium phosphate (pHs ~3–8) with or without denaturing

Table 1 | Full fitting parameters for the NMR spectra in Fig. 5, describing chemical shift (Ω) in ppm, linewidth at half height (LWHH) in Hz (and R2 in rads/s), normalized intensity (*I*) and *J* coupling for the second-order doublets in Hz

	Singlet				Second-order doublets					
	Ω (ppm)	LWHH (Hz)	R2 (rad/s)	*I*	Ω_1 (ppm)	Ω_2 (ppm)	$^2J_{FF}$ (Hz)	LWHH (Hz)	R2 (rad/s)	*I*
γF_2-Met	−74.93	39.2	123	0.40	−75.21	−74.21	211	39.8	125	1.00
	−114.22	47.4	149	0.15						
γF_2-Met(O)	−114.2	43.0	135	0.08	−108.52	−103.92	223	29.9	223	1.00
					−106.65	−103.81	224	33.4	224	0.43
					−106.48	−103.22	224	36.9	224	1.07
					−107.90	−103.92	226	32.8	226	0.45
γF_2-Met(O$_2$)	−114.22	36.0	113	0.19	−105.32	−103.41	234	57.3	180	1.00
					−105.04	−103.88	410	49.0	154	0.29
Mixture	−74.92	33.4	105	0.39	−75.20	−74.20	210	33.7	106	1.00
	−113.98	38.2	120	0.77	−105.32	−103.41	234	43.6	137	1.45
					−104.96	−103.95	235	42.0	132	0.56
					−108.5	−103.9	223	29.0	91	0.54
					−106.65	−103.79	229	28.0	88	0.20
					−106.46	−103.20	224	41.7	131	0.69
					−107.89	−103.97	230	31.5	99	0.22

Spectra recorded were referenced to the lock signal.

reagent such as guanidine hydrochloride (Gdn·HCl). This process can be used to modify Dha tags in folded proteins; the maintained tertiary structure of the protein can then be confirmed via circular dichroism measurement. However, if the site of interest is located in a highly hindered position in a protein, denaturing conditions can facilitate Dha formation and/or subsequent reaction. In this case, a suitable protocol for the refolding of the protein may need to be investigated.

The generated •CF$_2$R displays apparently high chemoselectivity; no cross-reactivity with other potential acceptors such as tryptophan or tyrosine (which require more stringent conditions to show reactivity[37]) is observed under these conditions.

Although redox-sensitive residues could in principle cause concern (e.g., E$_{ox}$[Tyr/Tyr•+] = +1.01 V (ref. [38])) in an oxidative quenching cycle (E$_{ox}$[Ru^{2+}/Ru^{3+}] = +1.26 V), this is in practice controlled by the use of excess iron(II) favoring a reductive quenching cycle via the corresponding Ru(+) species (E$_{ox}$ = +0.78 V) and single electron transfer to pySOOF. In all reactions tested so far, we have not observed evidence of Tyr cross-linking.

Sidechain reagent design

In our previous work, we performed a reactivity study to evaluate the minimal structural requirements of the pySOOF reagent that enables the fluoroalkylation reaction and 'brings in' the grafted residue sidechain. These results indicated the importance of the difluoromethyl unit in terms of reactivity—little or no product formation was observed for simple mono-fluoro substituted pySOOF reagents[18]. Therefore, an additional electron-withdrawing functional group linked to the fluorinated carbon appears important to achieve photoredox-mediated radical generation, probably by lowering the reduction potential and therefore 'matching' the redox thresholds of the photostimulated single-electron transfer cycle. Other electron-withdrawing groups such as carboxylates or acetamide can be used as an alternative to fluorine[18]. However, compared with the difluoromethyl 2-pyridyl-sulfone-derived reagents, more equivalents of reagent (5–25 equiv.), photocatalyst (5 equiv.) and iron salt (250 equiv.) are needed to accomplish higher conversions.

Overall, the synthesis of pySOOF reagents starts from commercially available starting materials such as difluoromethyl-(2-pyridyl)-sulfone or 2-mercaptopyridine and require two to six synthetic steps to prepare (Fig. 3a). In our previous work[18], we disclosed the synthesis of more than 30 variations of the pySOOF reagent (for selected examples, see Fig. 3b,c), that allow access to analogs of both native amino acid residues as well as posttranslationally modified counterparts. Owing to the mild and redox-selective reaction conditions, sidechains containing reactive handles and/or potentially redox-labile functional groups (such as iodide, azide, esters, biotin, amides or sulfate) can also be easily installed with this method.

This scope is not limited to these examples, as the diversity oriented synthesis (see below) of pySOOF reagent allows access in principle to further •CF$_2$R precursors with a wide variety in R, depending on the aim of the chemical–biological study. In the current study, we focused our interest on pySOOF-Met-derived reagents, which allowed us to install varied oxidation states of methionine (Met, Met(O), Met(O)$_2$) into an epitope-tagged target human protein, histone eH3.1 at site 4 (Fig. 4). While these syntheses are beyond the scope of this protocol, these and other sidechains are readily accessible in a diverse manner from common intermediates.

Reaction scale

Due to the high efficiency of this chemical mutagenesis, excellent conversions are detected at various scales (100 µg to 5 mg protein). In addition, the reaction proceeds at various protein concentrations (0.1 to 5 mg/mL). Sometimes, higher reagent (photocatalyst, pySOOF, iron sulfate) loadings are needed to obtain high conversions when the photochemical reaction is performed at lower protein concentrations. In the current study we report the use of 5 mg of histone eH3-Dha4 in each reaction, with a working concentration of 5 mg/mL (280 µM).

Reaction monitoring

Intact-protein electrospray mass spectrometry with accurate resolution in combination with a high-performance liquid chromatography (LC–MS) is recommended for analyzing both reaction progress and final product. Conversions are determined by comparing the relative total-ion-count intensity of the deconvoluted peak of precursor (e.g., Dha) and product (e.g., fluorinated) protein. The small changes in protein structure caused by installing a (difluorinated) sidechain has essentially no impact on comparative global ionization response and therefore relative intensities of different intact-protein m/z species can be compared; this has been verified by corresponding calibration curves with excellent correlations in numerous studies for varied reactions by us[18,39] and others. It even allows the kinetic parameters of on-protein reaction to be determined[40].

Side reactions

During the first reaction, a free Cys residue is converted to Dha via a three-step, bisalkylation-elimination sequence using 2,5-dibromohexanedimide (DBHDA) as alkylation reagent. Although this reaction is conducted in aqueous basic buffer (pH ≥8.0), other nucleophilic residues could in principle get alkylated; in practice these side products are rarely observed. Moreover, the formation of these overalkylated protein adducts can be easily inhibited by optimizing the reaction conditions (pH, buffer, reaction time, temperature profile, equivalents of DBHDA) or using an alternative alkylation reagent such as methyl 2,5-dibromopentanoate or others[34]. Very rarely, bis-alkylation generates a sulfonium that does not readily eliminate;[41] again, variation of reaction conditions (e.g., increase of temperature from room temperature (RT, ~18–25 °C) to 37 °C) can facilitate elimination of a successfully formed sulfonium.

If buffers or protein stocks are not properly degassed, residual oxygen can lead to oxidative damage of methionine residues during the photochemical step (observed as a series of +16 adducts for each Met in the protein sequence). Therefore, low levels of oxygen in the buffer (<6.0 ppm O$_2$) may prove important but the exclusion of oxygen from the reaction is not a strict requirement.

In rare examples, 'double •CF$_2$R addition' products are formed as minor side products. Often the formation of these side products is lowered by increasing the iron(II) concentration to promote the reductive quenching of on-protein α-C• radical intermediates to enolate. Otherwise, lowering the equivalents of the pySOOF reagent can also improve the mono/di-addition ratio.

Residual Dha-tagged protein can be observed after chemical mutagenesis in the final purified protein sample due to incomplete conversion. To reduce the amount of Dha-modified protein, the photochemical reaction needs to be monitored appropriately (see above) and optimized by changing the loadings of photocatalyst, pySOOF reagent, iron sulfate, reaction time or protein concentration. In most cases, conversions >90% are realistic. Of these, we typically find that the most important reagent concentration is iron sulfate as it is probably involved in the photoredox cycle and reductive quenching of the on-protein radical to give the corresponding product.

NMR studies

The rarity of fluorine in biology and its small size makes it a near-perfect 'zero size- zero background' label for F-dependent imaging and biophysical methods such as NMR[24]. In particular, due to the high gyromagnetic ratio of the ^{19}F isotope, this nucleus has a high sensitivity to NMR measurement,

allowing detection in principle of small quantities of fluorinated species[24]. Furthermore, because of its large chemical shift dispersion, even relatively small changes in the chemical environment of the ^{19}F nucleus can have an observable impact on the chemical shift—this allows one to readily distinguish between different species or to observe noncovalent (ligand–protein, protein–protein or protein–lipid) interactions[42], as well as aggregation and fibrillation[43]. Other application areas include the study of protein folding/unfolding[44] or enzymatic action[45]. During the course of our investigations, several different NMR magnets (400, 500 and 600 MHz) were tested, and for all instruments corresponding spectra were obtained. Sometimes, longer experiment times (more scans) were needed to improve signal-to-noise ratio and resolution.

Each protein scaffold and site is different and it is, as for any form of protein NMR, possible that there will be variation in how resolved each resonance is, depending on local motion. For solution NMR, line widths will largely be dependent on the rate of motion at the spin site of interest. In this regard, this will generally give rise to smaller line widths at flexible sites and in smaller proteins. Sites were chosen here on the basis that they reflect biologically relevant sites in a potentially relevant protein. We found that the lowest observed signal-to-noise ratio in the processed spectra was the resonance at δ_F −108.10 ppm, with a signal-to-noise ratio nonetheless greater than sixfold, most other signals gave much improved signal to noise. Thus, even in the case of substantial line broadening, sufficient signal should be able to be acquired to adequately analyze the spectra. It should also be noted that, since the ^{19}F spectra are relatively sparse, multiplication with an exponential window function corresponding to stronger linebroadening can also be used to further improve signal to noise before the reduction in resolution becomes problematic.

In the current protocol, we showcased the utility of this method by showing the clear differences in chemical shifts for proteins bearing sidechains that differ only in a single oxygen atom corresponding to the varied oxidation states of Met, Met(O) and Met(O)$_2$ at site 4 of human histone eH3.1 under denaturing conditons (Fig. 5). These experiments were performed on a Bruker AVANCE NEO 600 MHz NMR with a CPRHe-QR-1H/19F/13C/15N-5mm-Z helium-cooled cryo-probe with 28,000 scans, a 12 h experiment.

Strikingly, within a mixture containing all three modified histones together, each single modification state could be observed and easily distinguished. Each pair of ^{19}F substituents in any of the diastereomers will give rise to a coupled multiplet. In the limit where the chemical shift difference between the two fluoride (Δv) is greater than the scalar coupling constant (J), the two spins will be 'weakly coupled' and appear as a 'doublet of doublets', with each of the four multiplets having equal intensity. When J and Δv become close in value, the multiplet will shift to a 'roofed' form where the 'outer' multiplets are less intense than the 'inner' multiplets. Eventually when $J \gg \Delta v$, both resonances will appear as a singlet (see also Supplementary Methods)[46]. We have provided Python code[47] to fit these spectra that requires a user to specify 'rough' initial positions of the various multiplets, which are then optimized, revealing both a fitted spectrum and the subspectra of the individual contributing species (see, for example, Fig. 5). This procedure allows potential dissection of an initially complex spectrum into its individual discrete chemical components.

Assignment of the individual components, for example, reliably distinguishing D- and L- forms, requires a detailed calculation that is beyond the scope of this work. In brief, density functional theory can be used to calculate the chemical shift of substituents (e.g., the two ^{19}F substituents) in a given conformation[48,49]. Since the calculated chemical shifts will depend strongly on torsion angle, a calculation of this type must be averaged over an ensemble that accounts for the microscopic populations of all conformers to be accurate[50]. To perform such a calculation requires an accurate force field that will correctly reproduce the population distribution. In this case, test calculations of individual conformers reveal, because of the high electronegativity and anisotropy of the ^{19}F chemical shift caused by its proximity to the S(O) moiety, the calculated chemical shifts can vary by hundreds of ppm following relatively small changes in torsion angle, making ensemble determinations involving such $S(O)_nCF_2$ systems groups particularly challenging.

While unique assignment therefore remains challenging and computationally intensive, simple analysis of the spectrum reveals the composition of the solution. The differences can be striking and further allude to the challenge of a computational method for assignment. For example, the variation in ^{19}F chemical shift of the D- versus L- forms of γF_2Met (Fig. 5, top left): in one form $\Delta v \ll J$ and so at 600 MHz appears as a singlet, while the second form, $\Delta v = 221$ Hz, and the species contribute a 'roofed AB-quartet' to the spectrum. While it is not immediately obvious why the chemical shifts of the D- and L- form should vary in this way, identifying the presence of both species both from inspection and using our analysis software is straightforward.

Speculatively, we anticipate from initial inspection that a D- amino acid residue, when flanked by L- amino acids, will have access to a wider swathe of conformational space than the corresponding L- form[51]. From this analysis, we would anticipate D-Met variants will display more conformational freedom and will then probably attain more complete conformational averaging of chemical shift—this would lead to a full averaging of the chemical shifts and hence the appearance of a 'singlet' for this form.

Additional methods could also be considered to complement these approaches: D- versus L- diastereomer mixture ratios of peptide standards containing a D- or L- sidechain can be compared via HPLC with digested peptide fragments derived from modified protein[39].

Expertise needed to implement the protocol

For the synthesis of pySOOF reagents, standard knowledge and skills in organic synthesis, handling of toxic chemicals and purification methods, as used in a synthetic laboratory, are required. Some synthetic steps may need special equipment to create an inert (nitrogen or argon) atmosphere; in some steps, cryogenic conditions are necessary. The syntheses of these reagents are beyond the scope of this protocol but examples of the reagents that may be accessed and general strategies and schemes for their syntheses are given (Fig. 3).

To construct the plasmids and express the desired proteins, standard knowledge and skills in protein production and biology are needed. These experiments should take place in suitable bio-classified laboratories (level 1). Key equipment includes: autoclaves, incubators, laminar flow hoods, centrifuges and instruments to analyze cell culture 'optical density' values and plasmid/protein concentrations (UV/Vis absorbance). For plasmid sequencing, a suitable service partner should be identified.

The key steps: Dha-formation and photochemical mutagenesis reactions can be performed by nonexperts with basic laboratory skills. For these steps a thermomixer, glovebox and photobox with blue light emitting diodes (~50 W total, 450 nm) are needed.

To perform NMR studies, basic knowledge in NMR spectroscopy is required. To record the spectra, an NMR spectroscope with a ^{19}F probe is needed.

Limitations

The protocol necessitates the use of a Dha-containing protein (histone) precursor. While this is readily generated on histones, applying this methodology more generally to other proteins can be more involved if there are other free Cys residues required (for more detail, see discussion above). Other Dha-generation techniques are possible but can be more laborious.

Reactions with Dha unavoidably generate a mixture of L- and D- amino acids, often in close to equal amounts. Advantageously, the use of ^{19}F-NMR studied here also allows for the easy disambiguation of the two diastereomers in the spectra. In this way, not only can these mixtures of epimers be accurately assessed (diastereomeric ratio (d.r.) of D versus L in product protein) but also, potentially, so can the diastereomer (D versus L) selectivity of any processing of that product (e.g., PTM processing[18]), if desired. Detecting ^{19}F-containing residues by solution ^{19}F NMR requires there to be sufficient local dynamics to give rise to detectable resonances. It will be advantageous if the underlying protein is small (<30 kDa), the residue is in a region with rapid local motions (such as in the tails of histones), or if local motion is induced by chemical denaturation.

While the reactions themselves are rapid and tolerant enough not to require specialist conditions, the photochemical C• radical 'grafting' step requires semianaerobic buffer conditions that can be achieved by 'degassing' equilibration in a low-oxygen environment such as a glovebox (e.g., with <6 ppm O_2)—this may be difficult for some labs to access.

The analysis software used here requires some basic familiarity with Python, and the software suite nmrPipe needs to be installed; the software reads raw Bruker NMR data in FID format and a series of peak positions need to be estimated and provided to the code by a user, and then executed.

Materials

Reagents
- DBHDA, >95% (Kerafast, cat. no. EOX103, Sigma-Aldrich cat. no. 900607 or Tocris cat. no. 6175)
- Recombinant protein with one cysteine residue at the site of interest. In this study we utilize human histone eH3.1-Cys4 as model protein, which is expressed and purified following a literature-known procedure[18]

- Iron(II) sulfate heptahydrate, 99+% ($FeSO_4 \cdot 7H_2O$; Strem-Chemicals, cat. no. 93-2639)
- Tris(2,2′-bipyridyl)dichlororuthenium(II) hexahydrate, 99.95% ($Ru(bpy)_3Cl_2 \cdot 6H_2O$; Sigma-Aldrich, cat. no. 544981)
- Water (deionized water: >15 MΩ/cm resistance), filtered through a 0.2 μM disc filter)
- Dimethyl sulfoxide, 99% (Fluorochem, cat. no. 046777)
- Acetonitrile, HPLC grade ≥99.9% (MeCN; Sigma-Aldrich, cat. no. 34851)
- Formic acid, ≥95% (Sigma-Aldrich, cat. no. F0507)
- DL-dithiothreitol, >99% (DTT; Sigma-Aldrich, cat. no. D9779)
- 2-Mercaptoethanol, 99% (BME; Sigma-Aldrich, cat. no. M3148)
- Sodium phosphate monobasic, ≥98% (NaH_2PO_4; Sigma-Aldrich, cat. no. S3139)
- Sodium phosphate dibasic, ≥98.5% (Na_2HPO_4; Sigma-Aldrich, cat. no. S3264)
- Ammonium acetate, ≥98% (NH_4OAc; Sigma-Aldrich, cat. no. A1542)
- Gdn·HCl, 99% (Fluorochem, cat. no. 044943-1kg)
- Deuterium oxide, 99.9 atom %D (D_2O; Sigma-Aldrich, cat. no. 151882)

Equipment
- PD MiniTrap G-25 (GE Healthcare Life Sciences, cat. no. 28-9180-07)
- PD MidiTrap G-25 (GE Healthcare Life Sciences, cat. no. 28-9180-08)
- Sartorius Vivaspin 6 centrifugal concentrator, MWCO 5,000 Da (Sartorius, cat. no. 10767461)
- Eppendorf ThermoMixer (Grant-bio, cat. no. PHMT-PSC18, (20 × 0.5 mL + 12 × 1.5 mL)
- Positive-pressure inert gas glovebox (Belle Technology glovebox (http://www.belletechnology.co.uk/glovebox.php)) equipped with the BASF R3-11G catalyst
- LC–MS (Waters Xevo G2-S QTof coupled to Water Acquity ultra-performance liquid chromatography).
- ProSwift RP-2H 4.6 × 50 mm column (Thermo Fisher Scientific, cat. no. 064296)
- NanoPhotometer (Implen NanoPhotometer NP80 UV/Vis spectrophotometer)
- Benchtop centrifuge (Eppendorf, cat. no. 5424 R)
- AV600 NMR (Bruker AVANCE NEO 600 MHz spectrometer) with CPRHe-QR-1H/19F/13C/15N-5mm-Z helium-cooled cryo-probe
- pH meter (Orion Star A111 Benchtop pH Meter)
- Standard adjustable volume, repeat dispensing 'biological' pipettes
- UltaPoint graduated TipOne tip, 10 μL (Starlab, cat. no. S1111-3800)
- UltaPoint graduated TipOne tip, 200 μL (Starlab, cat. no. S1113-1700)
- UltaPoint graduated TipOne tip, 1,000 μL (Starlab, cat. no. S1111-6701)
- Eppendorf tubes, 1.5 mL (Eppendorf, cat. no. 10509691)
- Eppendorf tubes, 2 mL (Eppendorf, cat. no. 10038760)
- Centrifuge tube, 50 mL (Starlab, cat. no. E1450-0200)
- Centrifuge tube, 15 mL (Starlab, cat. no. E1415-0200)
- Vials for LC–MS (Superco, cat. no. SU861132)
- Vials for reaction (Scientific Glass Laboratories Ltd, cat. no. T103/V3, 14 mL snap top vial, rolled rim)
- NMR tubes, high-throughput, thin walled, 5 mm outer diameter (Wilmad-LabGlass, cat. no. 634-0871)

Reagent setup
▲ CRITICAL For the photochemical mutagenesis reaction, the reaction buffer, water and DMSO for the reagent stock solutions must be transferred inside the glovebox the day before the experiment and must degas via diffusion overnight (<6.0 ppm O_2).

Buffer for Dha formation
1. Dissolve 2.7 g Na_2HPO_4, 127 mg NaH_2PO_4 and 57.3 g Gdn·HCl in water to get a final volume of 200 mL (~140 mL of water).
2. Adjust the pH of the buffer to pH 8 by adding 3 M NaOH (prepared by dissolving NaOH pellets) under permanent pH control with a pH meter (freshly calibrated using a three-point calibration at pH 4, 7 and 10).
   ■ PAUSE POINT Typically, this buffer can be stored for at least for 2 weeks at 25 °C. For longer storage time, please check the pH before use.

Buffer for photochemical mutagenesis

1. Dissolve 7.7 g NH₄OAc and 57.3 g Gdn·HCl in water to a final volume of 200 mL (~140 mL of water).
2. Adjust the pH of the buffer to pH 6 by adding 3 M HCl (prepared by diluting 37% HCl with water) under permanent pH control with a pH meter (freshly calibrated using a three-point calibration at pH 4, 7 and 10).

   ■ PAUSE POINT Typically, this buffer can be stored for at least for 2 weeks at 25 °C. For longer storage time, please check the pH before use.

DBHDA stock solution ! CAUTION No toxicity information for DBHDA reagent is available, therefore assume it is toxic. When handling the reagent use appropriate personal protective equipment (gloves, laboratory coat and safety glasses) in a ventilated environment.

1. Before the experiment, charge a 1 mL Eppendorf tube with 10.0 mg DBHDA;
2. Dilute DBHDA with 66.2 μL DMSO (0.5 M) before the experiment.

   ■ PAUSE POINT This stock solution can be stored for 1 d at 25 °C.

FeSO₄·7H₂O stock solution

1. Before the experiment, charge a 1 mL Eppendorf tube with 5.0 mg FeSO₄·7H₂O, transfer the uncapped Eppendorf inside the glovebox.
2. Dilute the iron salt with 165 μL water (typically 50 μL water for 100 equiv. iron(II) salt) before the experiment.

   ■ PAUSE POINT This stock solution can be stored for 1 d inside the glovebox at 25 °C.

Ru(bpy)₃Cl₂·6H₂O stock solution

1. Before the experiment, charge a 1 mL Eppendorf tube with 5.0 mg Ru(bpy)₃Cl₂·6H₂O, transfer the uncapped Eppendorf inside the glovebox.
2. Dilute the photocatalyst with 165 μL water (typically 7 μL water for 1 equiv. photocatalyst) before the experiment.

   ■ PAUSE POINT This stock solution can be stored for 1 d inside the glovebox at 25 °C.

pySOOF-Met, -Met(O), -Met(O₂) reagent stock solutions ! CAUTION No toxicity information for pySOOF reagents are available, therefore assume it is toxic. When handling the reagent use appropriate personal protective equipment (gloves, laboratory coat and safety glasses) in a ventilated environment.

1. Before the experiment, charge a 1 mL Eppendorf tube with 5.0 mg pySOOF reagent, transfer the uncapped Eppendorf inside the glovebox.
2. Dilute the pySOOF reagent with DMSO (to 0.1 M) before the experiment.

   ■ PAUSE POINT This stock solution can be stored for 1 d inside the glovebox at 25 °C.

Mobile phases for LC–MS

1. Eluent A: mix 1 L of water with 1 mL formic acid (0.1% vol/vol). This eluent can be stored for 1 month at 25 °C.
2. Eluent B: mix 1 L of MeCN with 1 mL formic acid (0.1% vol/vol). This eluent can be stored for 1 month at 25 °C.

   ■ PAUSE POINT If needed, the buffer can be stored for 2 weeks at 25 °C.

Buffer for NMR studies

1. Dissolve 1.9 g NH₄OAc and 28.66 g of Gdn•HCl in D₂O to a final volume of 100 mL.
2. Adjust the pH of the buffer to pH 7 by adding 3 M HCl (prepared by diluting 37% HCl with water) under permanent pH control with a pH meter (freshly calibrated using a three-point calibration at pH 4, 7 and 10).

   ■ PAUSE POINT Typically, this buffer can be stored for at least for 2 weeks at 25 °C. For longer storage time, please check the pH before use.

Equipment setup
Photobox
A photobox can be constructed from easily accessible materials as shown in our previous work[18]. Other blue LED light sources of different intensities may be used but reaction times will vary.

PD Mini and MidiTrap G-25 equilibration

Desalting columns are equilibrated using the gravity protocol with the desired buffer following the instructions of the supplier. Afterward, the column is ready for buffer exchange or purification.

1 Remove all caps from the column and discard the storage buffer.
2 Place the desalting column in a centrifuge tube (15 mL for MiniTrap and 50 mL for MidiTrap) using the plastic holder from the supplier.
3 Charge the column with the desired buffer and let the buffer pass through the column via gravity.
4 Discard the flow-through and repeat Steps 1–4 twice for MiniTrap and once for MidiTrap.

LC–MS/MS settings

Intact protein analysis is performed on a Waters Xevo G2-S QTof coupled to Water Acquity ultra-performance liquid chromatography using a Thermo Proswift (250 mm × 4.6 mm × 5 µM) column.

Liquid chromatography parameters:

Time (min)	Eluent B (vol/vol) (%)
0	5
10	95

Mass spectrometer parameters:

Method parameter	Value
Capillary voltage	3,000 V
Cone volatege	160 V
Lock-spray analysis	Leucine enkephalin standard solution

Nitrogen with a total flow of 600 L/h is used as desolvation and nebulizer gas.

For analysis, MassLynx (Waters) and its maximum entropy (MaxEnt1) deconvolution algorithm (resolution: 1.00 Da/channel; width at half height: ion series/protein dependent; minimum intensity ratios: 33% left and right) is used. Spectra are deconvoluted between 10,000 and 25,000 Da. Any reaction conversions are calculated from relative peak intensities in the deconvoluted spectra.

Procedure

Formation of histone eH3.1-Dha4 ● Timing 21 h (including 16 h degassing)

1 Charge a 1 mL Eppendorf tube with 10 mg lyophilized Histone eH3.1-Cys4 and add 30 mg DTT (3 mg DTT/mg protein).
2 Dissolve the mixture in 500 µL NaPi (100 mM, pH 8, 3 M Gdn·HCl) buffer and vortex the solution until it is homogeneous.
   ? TROUBLESHOOTING
3 Incubate the crude mixture for 30 min at 25 °C at 500 rpm on an Eppendorf ThermoMixer.
4 Take up the crude mixture with a biological pipette and transfer the solution to an equilibrated G-25 miniTrap column (Cytiva) (with NaPi (100 mM, pH 8, 3 M Gdn·HCl)).
5 Let the protein solution pass into the column bed and place the miniTrap with the plastic holder in a new 15 mL centrifuge tube.
6 Elute off the protein with 1 mL NaPi (100 mM, pH 8, 3 M Gdn·HCl).
   ▲ **CRITICAL STEP** Do not use more than 1 mL buffer to elute off the protein otherwise DTT could co-elute and cause problems during the Dha formation reaction (inhibition of Dha formation and/or react with Dha-modified histone via thio-Michael addition).
7 Transfer the protein solution to a fresh 1 mL Eppendorf tube.
8 Check the protein concentration on a NanoPhotometer at 280 nm (8.5 mg/mL; 85% protein recovery).

9   Add 58 μL of the DBHDA stock solution in DMSO (0.5 M, 61 equiv.) to the protein solution, vortex the closed Eppendorf for 10 s and place the tube inside an Eppendorf ThermoMixer.

10 Incubate the reaction mixture for 45 min at 25 °C followed by further 2 h at 37 °C at 500 rpm.
    ▲ CRITICAL STEP Longer incubation times may result in formation of *N*-alkylated protein side products.

11 Take up the crude mixture with a biological pipette and transfer the solution to the equilibrated G-25 midiTrap column (with NH_4OAc (500 mM, pH 6, 3 M Gdn·HCl)).

12 Let the protein solution pass into the column bed and place the midiTrap with the plastic holder in a new 50 mL centrifuge tube.

13 Elute off the protein with 1.5 mL NH_4OAc (500 mM, pH 6, 3 M Gdn·HCl) and transfer the protein solution to a fresh 2 mL Eppendorf tube.
    ▲ CRITICAL STEP Do not use more than 1.5 mL buffer to elute off the protein otherwise excess of DBHDA or phosphate could co-elute and cause problems during the photochemical mutagenesis.

14 Check the protein concentration on a NanoPhotometer at 280 nm (5 mg/mL; 88% protein recovery from previous step). Dilute as necessary to desired protein Dha stock concentration (in this protocol, the Dha stock solution used was 5 mg/mL).

15 Analyze an aliquot of the Dha-modified histone (inject 2 μL of 0.04 mg/mL of protein in water with 0.1% (vol/vol) formic acid) by LC–MS (intact protein analysis).
    ? TROUBLESHOOTING

16 Transfer the protein stock solution inside the glovebox and degas the sample via diffusion for at least 8 h.
    ■ PAUSE POINT The resulting Dha-modified histone can be stored in the glovebox for 3 months at ambient temperature without decomposition of the Dha tag. Storage of other proteins will depend on their own stability in various temperatures or buffers. For longer storage times, check the quality of the stock solution by analysing an aliquot by LC–MS (intact protein analysis).

Photochemical mutagenesis—formation of histone eH3.1-γF₂Met4 ● Timing 13 h (including 12 h NMR experiment)

▲ CRITICAL STEP Reaction buffer, water and DMSO for reagent stock solutions and protein stock solutions must be transferred inside the glovebox the day before the experiment and must degas via diffusion overnight (<6.0 ppm O_2). Moreover, prepare all reagent stock solutions freshly before the experiment inside the glove box ('Reagent setup'). Improperly degassed solutions will cause oxidative damage to the protein structure during the photochemical reaction.

17 Transfer the open reaction vial with the corresponding cap inside the glovebox.

18 Add 1 mL of the histone eH3-Dha4 stock solution (5 mg protein, 280 nmol) to the reaction vial.

19 Add 7 μL of the Ru(bpy)₃ stock solution (280 nmol, 1 equiv.), 4.5 μL of the pySOOF-Met stock solution (560 nmol, 2 equiv.) and 50 μL of the $FeSO_4$ stock solution (28 μmol, 100 equiv.) to the protein solution.

20 Close the reaction vial with the corresponding cap, shake it carefully and transfer the vial outside the glovebox.
    ▲ CRITICAL STEP Ensure that the vial is closed properly to avoid oxidative damage during the photochemical reaction.

21 Place the reaction vial in the center on the plexiglass plate above the LED probe, close the door of the photobox and switch on the light source to maximum power (50 W) for 15 min.
    ! CAUTION Never use the photobox with an open door nor look at the switched on light source as high-energetic blue light may cause eye damage.

22 Switch off the light source, open the door of the photobox, remove the vial, open the cap and directly add 20 μL 2-mercaptoethanol then vortex the crude solution for 30 s.
    ! CAUTION Never use the photobox with an open door nor look at the switched on light source as high-energetic blue light may cause eye damage.
    ▲ CRITICAL STEP The addition of 2-mercaptoethanol is crucial to avoid oxidative damage after removing the cap and to complex and remove the excess iron.

23 Take up the crude mixture (~1 mL) with a biological pipette and transfer 0.5 mL of the solution twice to two separate equilibrated G-25 MiniTrap columns (equilibrated with NH_4OAc (250 mM), Gdn•HCl (3M) buffer, pH 7, in D_2O)).

24 Let the protein solutions pass into the column bed and place the MiniTrap with the plastic holder in a new 50 mL centrifuge tube.

25 Elute, then discard the first 200 μL, using the same buffer (NH_4OAc (250 mM), Gdn•HCl (3M) buffer, pH 7, in D_2O),

26 Collect the next 0.5 mL from each MiniTrap, pool and transfer the cumulative 1 mL of buffer exchanged protein solution to a fresh 2 mL Eppendorf tube.
▲ **CRITICAL STEP** This elution strategy removes the vast majority of the photocatalyst while retaining the bulk of the desired modified protein product, higher elution volumes may increase product recovery but will result in the accumulation of undesired photocatalytic side products that may interfere with accurate modified protein product quantification.

27 Check the protein concentration on a NanoPhotometer at 280 nm (3 mg/mL; 75% protein recovery).

28 Analyze an aliquot of the modified histone (inject 2 μL of 0.04 mg/mL of protein in water with 0.1% (vol/vol) formic acid) by LC–MS (intact protein analysis).
? **TROUBLESHOOTING**

29 Transfer the protein solution to a Vivaspin 6 concentrator (MWCO 5000 Da) and concentrate the sample to 500 μL in a benchtop centrifuge at 10,000g, 4 °C for 15 min. A volume equivalent to the storage buffer that is held in the concentrator is discarded in line with standard use.

30 After concentrating, transfer the protein solution to a 2 mL Eppendorf vial and transport it on ice to the NMR instrument, then pipette into the NMR tube.
■ **PAUSE POINT** If the measurement is not immediately taken, flash freeze protein samples and store at −80 °C.

31 Record a ^{19}F-NMR spectrum. In our case, using a 600 MHz NMR spectrometer equipped with a helium cryoprobe, spectra were recorded for ~12 h, using a transmitter offset of –100 ppm, a spectral width of 230 ppm, an acquisition time of 0.5 s and a 1 s relaxation delay. Spectra can be recorded either with or without ^{1}H decoupling which narrows linewidth by refocussing unresolved $^nJ_{HF}$, providing a commensurate improvement in intensities.

32 Fit the NMR spectra to user-defined numbers of those species contributing to singlets and those contributing to second-order doublet pairs. For species contributing to singlets, an initial intensity, linewidth and chemical shift need to be provided. For species contributing to a second-order doublet pair, an initial intensity, linewidth, a pair of chemical shifts and scalar coupling constant need to be provided.
▲ **CRITICAL STEP** In the protein systems described here, Gdn•HCl was used in the NMR samples to prevent aggregation and improve the quality of associated NMR spectra.

? **TROUBLESHOOTING**

Timing

Steps 1–10, preparation of Dha-tagged protein: 4 h
Steps 11–15, purification of crude mixture and analysis: 1 h
Step 16, transfer into glovebox and degassing: 16 h
Steps 17–20, preparation of reaction mixture: 15 min
Step 21, photochemical mutagenesis: 15 min
Steps 22–28, purification of the crude mixture and analysis: 30 min
Steps 29–32, NMR sample preparation and NMR experiment: 12 h

Troubleshooting

Troubleshooting advice can be found in Table 2.

Table 2 | Troubleshooting table

Step	Problem	Possible reason	Solution
2	Protein not completely in solution	Limited solubility of chosen protein substrate in that buffer solution	Sonicate the sample for 5 min. If the sample is still heterogeneous, centrifuge the sample, and take the supernatant for the DTT treatment
			Table continued

Table 2 (continued)

Step	Problem	Possible reason	Solution
15	Detection of over alkylated protein products or incomplete conversion	Too much alkylation reagent used	Repeat the Dha formation with less reagent
		Buffer pH not correct	Check pH of buffer
		Reaction temperature too high or low or reaction time too long	Optimize reaction temperature and time
28	Detection of oxidative damaged protein or incomplete conversion	Buffer, protein stock solutions not properly degassed. High oxygen levels in the glovebox. Reaction vial not sealed properly	Check oxygen levels in the glovebox. Repeat the experiments with freshly degassed solutions and seal the vial with a Teflon tape
		Nonoptimized reaction conditions	Investigate the reaction on a smaller scale, screen different reaction parameters
31	Poor signal-to-noise ratio. Complex spectra	Low protein concentration	Increase the protein concentration using appropriate vivaspin concentrator. Increase scan number
		Formation of protein adducts	Use a denaturing deuterated buffer for the NMR experiment
31	No lock or only a weak lock can be achieved by the spectrometer	The use of Gdn•HCl in the NMR samples	Perform setup such as tuning, shimming and matching on a similar sample, but without the Gdn•HCl. For example, it may be possible to lock sufficiently well to reference to the lock signal, but with need for shimming on a sample without Gdn•HCl

Anticipated results

In the current study, we successfully installed three different difluorinated amino acid mimics (γCF_2Met, γCF_2Met(O), γCF_2Met(O)$_2$) using our photochemical mutagenesis approach, starting from histone eH3.1-Dha4. For all reactions we typically observe the desired sidechain incorporation with >90% conversion to the desired labeled protein, and manage to produce, for each fluorinated mutant, multiple milligrams (e.g., ~4 mg) of product. Each modification shows a characteristic multiplicity and chemical shift in the NMR; these allow the distinction of different oxidation states of Met by NMR analysis.

Data availability

Raw MS and ^{19}F-NMR data have been deposited with the following identifier: https://doi.org/10.5281/zenodo.6836127.

References

1. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342–7372 (2005).
2. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
3. Farrelly, L. A. et al. Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **567**, 535–539 (2019).
4. Nadal, S., Raj, R., Mohammed, S. & Davis, B. G. Synthetic post-translational modification of histones. *Curr. Opin. Chem. Biol.* **45**, 35–47 (2018).
5. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
6. Chen, H., Venkat, S., McGuire, P., Gan, Q. & Fan, C. Recent development of genetic code expansion for posttranslational modification studies. *Molecules* **23**, 1662 (2018).
7. Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H. & Chin, J. W. Genetically Encoding Nϵ-Methyl-l-lysine in recombinant histones. *J. Am. Chem. Soc.* **131**, 14194–14195 (2009).
8. Akahoshi, A., Suzue, Y., Kitamatsu, M., Sisido, M. & Ohtsuki, T. Site-specific incorporation of arginine analogs into proteins using arginyl-tRNA synthetase. *Biochem. Biophys. Res. Commun.* **414**, 625–630 (2011).
9. Chatterjee, C. in *Total Chemical Synthesis of Proteins* (eds Brik, A. et al.) 489–513 (Wiley-VCH, 2021).
10. Qi, Y.-K., Ai, H.-S., Li, Y.-M. & Yan, B. Total chemical synthesis of modified histones. *Front. Chem.* **6**, 19–19 (2018).
11. Wright, T. H., Vallée, M. R. J. & Davis, B. G. From chemical mutagenesis to post-expression mutagenesis: A 50 year odyssey. *Angew. Chem. Int. Ed.* **55**, 5896–5903 (2016).
12. Dadová, J., Galan, S. R. G. & Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Curr. Opin. Chem. Biol.* **46**, 71–81 (2018).
13. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009).

14. Wright, T. H. & Davis, B. G. Post-translational mutagenesis for installation of natural and unnatural amino acid side chains into recombinant proteins. *Nat. Protoc.* **12**, 2243–2250 (2017).

15. Wright, T. H. et al. Posttranslational mutagenesis: a chemical strategy for exploring protein side-chain diversity. *Science* https://doi.org/10.1126/science.aag1465 (2016).

16. Yang, A. et al. A chemical biology route to site-specific authentic protein modifications. *Science* https://doi.org/10.1126/science.aah4428 (2016).

17. Zhang, M., He, P. & Li, Y. Contemporary approaches to α,β-dehydroamino acid chemical modifications. *Chem. Res. Chin. Univ.* **37**, 1044–1054 (2021).

18. Josephson, B. et al. Light-driven post-translational installation of reactive protein side chains. *Nature* **585**, 530–537 (2020).

19. Müller, M. M. & Muir, T. W. Histones: at the crossroads of peptide and protein chemistry. *Chem. Rev.* **115**, 2296–2349 (2015).

20. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

21. Zheng, Y., Huang, X. & Kelleher, N. L. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. *Curr. Opin. Chem. Biol.* **33**, 142–150 (2016).

22. Rothbart, S. B. et al. An interactive database for the assessment of histone antibody specificity. *Mol. Cell.* **59**, 502–511 (2015).

23. Duan, Q., Chen, H., Costa, M. & Dai, W. Phosphorylation of H3S10 blocks the access of H3K9 by specific antibodies and histone methyltransferase. Implication in regulating chromatin dynamics and epigenetic inheritance during mitosis. *J. Biol. Chem.* **283**, 33585–33590 (2008).

24. Claridge, T. D. W. in *High-Resolution NMR Techniques in Organic Chemistry* (3rd edn) (ed Claridge, T. D. W.) 421–455 (Elsevier, 2016).

25. Chalker, J. M. & Davis, B. G. Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Curr. Opin. Chem. Biol.* **14**, 781–789 (2010).

26. Kim, G., Weiss, S. J. & Levine, R. L. Methionine oxidation and reduction in proteins. *Biochim. Biophys. Acta* **1840**, 901–905 (2014).

27. Bagert, J. D. et al. Oncohistone mutations enhance chromatin remodeling and alter cell fates. *Nat. Chem. Biol.* **17**, 403–411 (2021).

28. Jang, Y. et al. H3.3K4M destabilizes enhancer H3K4 methyltransferases MLL3/MLL4 and impairs adipose tissue development. *Nucleic Acids Res.* **47**, 607–620 (2019).

29. Brown, Z. Z. et al. Strategy for "detoxification" of a cancer-derived histone mutant based on mapping its interaction with the methyltransferase PRC2. *J. Am. Chem. Soc.* **136**, 13498–13501 (2014).

30. Hintzen, J. C. J. et al. γ-Difluorolysine as a 19F NMR probe for histone lysine methyltransferases and acetyltransferases. *Chem. Comm.* **57**, 6788–6791 (2021).

31. Brittain, W. D. G., Lloyd, C. M. & Cobb, S. L. Synthesis of complex unnatural fluorine-containing amino acids. *J. Fluor. Chem.* **239**, 109630 (2020).

32. Papworth, C., Bauer, J. & Braman, J. Site-directed mutagenesis in one day with >80% efficiency. *Strategies* **9**, 3–4 (1996).

33. Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A "tag-and-modify" approach to site-selective protein modification. *Acc. Chem. Res.* **44**, 730–741 (2011).

34. Lai, K.-Y. et al. LanCLs add glutathione to dehydroamino acids generated at phosphorylated sites in the proteome. *Cell* **184**, 2680–2695.e2626 (2021).

35. Chalker, J. M. et al. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2**, 1666–1676 (2011).

36. Dadová, J. et al. Precise probing of residue roles by post-translational β,γ-C,N Aza-Michael mutagenesis in enzyme active sites. *ACS Cent. Sci.* **3**, 1168–1173 (2017).

37. Imiołek, M. et al. Residue-selective protein C-formylation via sequential difluoroalkylation-hydrolysis. *ACS Cent. Sci.* **7**, 145–155 (2021).

38. Li, B. X. et al. Site-selective tyrosine bioconjugation via photoredox catalysis for native-to-bioorthogonal protein transformation. *Nat. Chem.* **13**, 902–908 (2021).

39. Mollner, T. A. et al. Post-translational insertion of boron in proteins to probe and modulate function. *Nat. Chem. Biol.* **17**, 1245–1261 (2021).

40. Lin, Y. A. et al. Rapid cross-metathesis for reversible protein modifications via chemical access to Se-Allyl-selenocysteine in proteins. *J. Am. Chem. Soc.* **135**, 12156–12159 (2013).

41. Galan, S. R. G. et al. Post-translational site-selective protein backbone α-deuteration. *Nat. Chemi. Biol.* **14**, 955–963 (2018).

42. Stadmiller, S. S., Aguilar, J. S., Waudby, C. A. & Pielak, G. J. Rapid quantification of protein-ligand binding via 19F NMR lineshape analysis. *Biophys. J.* **118**, 2537–2548 (2020).

43. Suzuki, Y. et al. Resolution of oligomeric species during the aggregation of Aβ1–40 using 19F NMR. *Biochemistry* **52**, 1903–1912 (2013).

44. Danielson, M. A. & Falke, J. J. Use of 19F NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 163–195 (1996).

45. Hattori, Y. et al. Protein (19)F-labeling using transglutaminase for the NMR study of intermolecular interactions. *J. Biomol. Nmr.* **68**, 271–279 (2017).

46. Hore, P., Jones, J. & Wimperis, S. *NMR: The Toolkit* 2nd edn (Oxford Univ. Press, 2015).

47. Baldwin, A. J. https://github.com/ajb204/FluoroStronger (2022).

48. Karunanithy, G. et al. Harnessing NMR relaxation interference effects to characterise supramolecular assemblies. *Chem. Comm.* **52**, 7450–7453 (2016).

49. Wu, J. et al. Synergy of synthesis, computation and NMR reveals correct baulamycin structures. *Nature* **547**, 436–440 (2017).

50. Gortari, I. D. et al. Time averaging of NMR chemical shifts in the MLF peptide in the solid state. *J. Am. Chem. Soc.* **132**, 5993–6000 (2010).

51. Anil, B., Song, B., Tang, Y. & Raleigh, D. P. Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability. *J. Am. Chem. Soc.* **126**, 13194–13195 (2004).

Author contributions

P.G.I. and B.G.D. conceived and designed the experiments. P.G.I. synthesised the pySOOF reagents. B.J. performed the chemical mutagenesis experiments. B.G. performed the NMR experiments. P.G.I., B.J., B.G., M.J.D., V.G., A.J.B. and B.G.D. collected and/or analyzed data. P.G.I., B.J. and B.G.D. wrote the paper. All authors read and commented on the paper.

Competing interests

A patent is being filed that might afford authors royalties were it to be licensed. The BGD group makes DBHDA available via Kerafast on a not-for-profit basis.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-022-00800-9.

Correspondence and requests for materials should be addressed to Andrew J. Baldwin or Benjamin G. Davis.

Peer review information *Nature Protocols* thanks Olugbeminiyi Fadeyi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Related links
Key references using this protocol
Josephson, B. et al. *Nature* **585**, 530–537 (2020): https://doi.org/10.1038/s41586-020-2733-7