

Post-Translational Site-selective Protein Backbone α -Deuteration

Sébastien R.G. Galan,^a James R. Wickens,^a Jitka Dadova,^a Wai-Lung Ng,^a Xinglong Zhang,^a
Robert A. Simion,^a Robert Quinlan,^a Elisabete Pires,^a Robert S. Paton,^{a,c} Stephen Caddick,^b
Vijay Chudasama,^{b*} and Benjamin G. Davis^{a*}

^a Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Mansfield Road, Oxford OX1 3TA, U.K.

^b Department of Chemistry, University College London, Gordon Street, London WC1H 0AJ, U.K.

^c Department of Chemistry, Colorado State University, Fort Collins, CO 80523, USA

*To whom correspondence should be addressed:

v.chudasama@ucl.ac.uk, Ben.Davis@chem.ox.ac.uk.

Abstract 147 words

Isotopic replacement is a long-proven tool in small molecules. However, applications in proteins are largely limited to biosynthetic strategies or at exchangeable (e.g. N–H/D) labile sites only. The development of post-biosynthetic, C–¹H → C–²H/D replacement in proteins could enable probing of mechanisms amongst other uses. Here, we describe a chemical method for selective protein α -carbon-deuteration (proceeding from Cys to dehydroalanine (Dha) to deuterio-Cys) allowing overall ¹H→²H/D exchange at a non-exchangeable backbone site. It is used here to probe mechanisms of reactions used in protein bio-conjugation. This suggests, together with quantum mechanical calculations, stepwise deprotonations via on-protein carbanions and unexpected sulfonium ylids in the conversion of Cys to Dha), consistent with a ‘carba-Swern’ mechanism. The ready application on existing, intact protein constructs (without specialized culture or genetic methods) suggests this C–D labelling strategy as a possible tool in protein mechanism, structure, biotechnology and medicine.

Keywords

protein modification; protein chemistry; protein labelling; deuteration; isotopic replacement;
protein mechanism; backbone modification

Introduction

Despite their clear utility in small molecules,^{1, 2} methods for site-selective replacement of non-exchangeable hydrogen isotopes (e.g. $^1\text{H} \rightarrow ^2\text{H} = \text{D}$) into *intact proteins* are essentially unknown. Hydrogen isotopes can be used as possible probes of structure and dynamics (e.g. by NMR³ or vibrational spectroscopy^{4, 5}) as well as reactivity. Their use could therefore not only enhance study of protein mechanism, structure and function but also enable the investigation of chemical mechanism in proteins.

Current possible strategies for isotope introduction typically require isotope incorporation *prior* to assembly. One approach to building labeled proteins is the segmental isotopic labeling strategy⁶ using, for example, expressed protein ligation and *trans*-splicing starting from synthetically prepared peptides into which isotopically-labeled amino acid has been incorporated through traditional peptide synthesis.⁷ Such linear assembly approaches are complemented by incorporation during biosynthetic assembly.^{4, 7-12} The latter typically lead to global, and hence non-site-selective, labeling of an amino acid type or can require creation and use of multicomponent, engineered expression systems.

Notably, direct (convergent), site-selective incorporation of deuterium into proteins post-assembly through non-heteroatom C–D bond formation (i.e. at typically non-exchangeable sites) has not, to our knowledge, been clearly demonstrated. Early reports suggested partial exchange at C2 of His in bovine pancreatic ribonuclease A.¹³ Therefore a complementary method that could be readily applied to intact proteins and hence more directly exploited by the non-expert using existing protein samples could prove useful.

We postulated that a general site-selective backbone α -C labelling might be possible via the creation and electrophilic trapping of a suitable α -C-enolate C-nucleophile (**Figure 1a**). This, in turn, might be generated via conjugate addition of an appropriate nucleophile to an α,β -unsaturated peptidic amide. In principle, this strategy could be applied not only to small peptides but directly to intact proteins, if α,β -unsaturated protein amides could be accessed. We also reasoned that ideally, if such α,β -unsaturated protein amides could be both derived from *and* used to form the same amino acid residue then the overall process could be performed in an essentially 'traceless' manner (**Figure 1b**).

Dehydroalanine (Dha) is a potentially useful amino acid residue that can be found in proteins or peptides naturally,¹⁴ following metabolism,¹⁵ via amber-codon suppression methods¹⁶ or derived synthetically from Cys.¹⁷⁻¹⁹ Thus, in one proof-of-principle guise of this general approach (which could in principle be applied to other electrophiles also), α -C H \rightarrow D

replacement could be envisaged at Cys (and Cys derivatives) (**Figure 1c**). Indeed, the possibility of such a mechanism has been noted by others too.²⁰ The use of “D+” as a small electrophile for enolate trapping might also take advantage of concerted solvent reorganization during Michael-type addition as well as related known²¹ ‘relayed’ D-exchange mechanisms for enhanced accessibility. Conversion first of Cys to Dha could then be followed by thia-Michael addition of a suitable sulfur nucleophile (ideally with ready deprotection, if needed) in the presence of deuterium oxide, as a high-concentration source of deutron electrophile.

Here, we describe such a method that can be readily applied to create a direct and site-selective C–H → C–D isotope replacement in the backbone of intact proteins. We show the utility of this method in the study of a model chemical protein reaction where isotope replacement suggests previously unobserved mechanistic intermediates and details.

Results ~2100 words

Development of Site-Selective Protein Deuteration Method

Whilst a range of sulfur nucleophiles have been successfully reacted with Dha^{16, 17, 22-29} in proteins, these typically generate alkylated Cys variants that cannot be readily converted to Cys itself; these would not therefore readily allow traceless α -C H \rightarrow D replacement. We therefore tested a range of putative sulfur nucleophiles that might also readily generate Cys (see **Supplementary Figure 1**). Whilst use of the thioacetate (**Figure 1c**, R = Ac) proved successful and proceeded with excellent conversion (>95%, see **Supplementary Figure 1b**), subsequent deprotection proved unsuccessful. The use of unprotected S-nucleophile source Na₂S also proceeded well in some systems (see **Supplementary Figure 1c**); however, in some systems concomitant dimerization via thioether formation led to formation of unwanted byproducts. Finally, use of sodium thiophosphate (**Figure 1c**, R = PO₃²⁻) proved successful in both incorporation and deprotection steps. Importantly, high deuterium incorporation levels were achieved, enabled by careful use of fully deuterated reagents, buffer components and protein solutions and/or use of anhydrous reagents (see **Online Methods**).

Specifically, model protein histone H3 was readily H \rightarrow D exchanged, site-selectively first at site 10 (**Figure 2**). H3-Cys10 was converted to H3-Dha10 using the *bis*-alkylation-elimination reagent DBHDA¹⁷ (>95%, 3 h) then H3-Dha10 was converted to H3-[d1]Cys10-S-PO₃²⁻ (>95%, 1 h) before deprotection to H3-[d1]Cys10 was achieved readily and flexibly in either of two ways: either through use of the phosphatase enzyme PP1 (>95%, 1 h), which proved capable of hydrolyzing the P-S bond to give Cys in a range of environments, or through hydrolysis at acidic pH (>95%, 2 h) (**Figure 2b** and **Supplementary Figure 2**). Exhaustive dialysis in non-deuterated buffer then allowed full protonation of all *exchangeable* protons, without any loss of the site-selectively installed α -deuteron at the α -carbon of Cys10 and final isolation of deuteroprotein in typically 65-85% yield. Characterization by high resolution MS and tryptic-digest-MS confirmed the level and site of incorporation (**Supplementary Figures 1-8** and **Supplementary Tables 1 and 2**). The method we use here has the potential to generate epimers at the alpha position (**Figure 1**), with a stereoselectivity that is potentially influenced by the protein substrate.³⁰ Through the use of chromatographic separation and MS detection (see **Supplementary Figures 9-17**) we were able to determine a diastereomeric ratio (d.r. ~ 1:2, L/D), suggesting only slight influence by the substrate.

Next, to test breadth and application to other sites, an essentially similar process was applied to site 26 in H3 histone (see **Supplementary Figure 18**). This also proceeded

successfully (in typical overall yields of 70-75%) using the same method thereby demonstrating breadth of application. These reactions demonstrated, to the best of our knowledge, the first examples of a traceless, convergent, site-selective chemical protein modification that can be used for selective incorporation / replacement (H→D) of a deuterium atom into the α -position of an amino acid residue (here Cys) on a protein.

Using H→D Replacement to Study Protein Chemistry

Dehydroalanine has proven in recent years to be a reactively versatile and useful residue in polypeptides. It occurs in peptide natural products¹⁴ and in the biosynthetic precursors³¹ for some proteins. Selective reactions with Dha have also, for example, allowed the creation of mimics of post-translational modifications (such as peptidylation, phosphorylation, lipidation and glycosylation) of key residues of a number of proteins, including kinases and histones.^{16, 17, 22-29,}³² More recently, it has enabled a form of protein editing through the creation of native side chains via C–C bond formation.³²

Inspired by previous studies on busulfan metabolism and selective protein cleavage,^{15, 33} our group has developed the use of *bis*-alkylating agents to chemically introduce Dha residues into proteins.¹⁷ The design of the water-soluble reagent dibromohexanediamide (DBHDA) allowed ready Dha formation (**Figure 3**) from cysteine (Cys) that is unusually rapid, efficient and highly selective, when compared to typical elimination reactions under the same conditions.¹⁷ This chemical protein reaction is therefore one of potentially strong, broad interest in that has enabled (and could enable) diverse protein chemistries, yet, its mechanism remains unclear. We therefore chose this seemingly unusual reaction as one that could be studied in detail readily using the traceless deuteration strategy described above.

The postulated¹⁷ mechanism (**Figure 3a**) involves a double (*bis*) alkylation (inter- then intra-molecular) of the γ -S of the Cys residue leading to the formation of an unstable cyclic sulfonium intermediate that then gives Dha (**Figure 3**). In nearly all cases, the sulfonium ion appears to undergo direct elimination readily to yield Dha with good conversion (**Figure 3a**) and is unobservable. Interestingly, however, some of us have observed the formation of a *stable*, directly observable sulfonium intermediate upon treatment of a single variant of the green fluorescent protein (GFP) GFP-Cys147 (GFP-S147C) with DBHDA (**Figure 4a**).^{26, 34} Such an intermediate has also been observed as a side-product in the creation of a Dha-containing di-ubiquitin probe.²⁸ The efficiency of the Cys→Dha transformation is strikingly dependent on the nature of the putative sulfonium intermediate formed: reagents that would lead to 5-membered cyclic tetrahydrothiopheniums prove far more efficient than those that would give simple acyclic

or even 6-membered counterparts.¹⁷ The detailed basis of the privileged and peculiar reactivity of tetrahydrothiopheniums is therefore unclear.

Together these observations – the implied nature of the privileged sulfonium and the relative inaccessibility to solvent of the α -position of GFP residue 147 – suggested a possibly highly unusual elimination mechanism. In principle, variations on four potential mechanisms of elimination of a sulfonium ion could be considered (**Figure 3b**). Various early studies on some small molecule sulfoniums have proposed both E₂-elimination and α' - β elimination mechanisms;³⁵⁻³⁹ these and related prior observations are inconsistent with any one dominant pathway, suggesting altered manifolds dependent on e.g. sulfonium type. Moreover, no reactions of this type (or similar) have been mechanistically evaluated previously on-protein, highlighting a potentially powerful influence of substrate (i.e. the protein) upon reaction. Prompted by these intriguing observations of selectivity and putative intermediates, we set out to probe this mechanism in detail as an example of an apparently complex and mechanistically unclear chemical reaction using the deuterium-replacement method.

Exploring the Chemistry of a Stable Protein Sulfonium

First, before use of this stable-deuteration strategy, we set out to further confirm and explore the reactivity of the proposed sulfonium intermediate under potentially *exchangeable C-H* (i.e. basic) conditions. GFP-S147C is the only system in which a stable sulfonium has been observed (**Figure 4a**): GFP-S147Sul, resulting from reaction with DBHDA. Its stability can be attributed to local microenvironment of the protein; in GFP the structured protein β -barrel appears to shield the α -proton of residue 147, thus rendering it inaccessible and preventing elimination to dehydroalanine.²⁶ Although this sulfonium species is stable at low temperatures (ca. 4 °C) over a protracted period, at 37 °C at pH 8 we observed evidence of decomposition (see **Supplementary Figure 19**). In the absence of a pathway for the formation of Dha, apparently direct fragmentation was observed that appeared to proceed through [3+2] cycloreversion (**Figure 4a**) to form a corresponding protein vinyl sulfide with concomitant elimination of acrylamide. This was consistent with both the mass differences observed and prior observations in small molecules.^{38, 40}

Moreover, computation revealed the most stable TS barrier for this reaction. With conformational sampling, a total of 10 transition structures were found. These ranged in energy from 20.3 – 25.1 kcal mol⁻¹ (or 15.9 – 20.7 kcal mol⁻¹ with respect to the lowest energy sulfonium ylid, at 4.4 kcal mol⁻¹), indicative of a kinetically feasible process at 37 °C. The associated

intrinsic reaction coordinate confirmed that this is a concerted *albeit asynchronous* [3+2] retro-cycloaddition / cycloreversion (see **Supplementary Figure 20**)

The observation of such a retro-cycloaddition / cycloreversion implied the formation of a sulfonium ylid from sulfonium GFP-S147Sul (**Figure 4a**). We considered that this intermediate might be sufficiently long-lived to allow ‘wash in’ of label from solvent that could be observed in fragmented acrylamide (see **Supplementary Figures 21,22**). Incubation of GFP-S147Sul in deuterated buffer released β -deutero-acrylamide, which was unambiguously identified by MS and MS/MS, consistent with wash-in into GFP-S147Sul followed by fragmentation. Next, we explored ‘wash in’ in systems that lead to *productive* elimination to Dha might carry such a ‘washed-in’ label in their eliminated THT leaving groups (as well as in any long-lived carbene in the protein product). Therefore, two proteins that, unlike GFP-S147C, are known to productively produce Dha, subtilisin SBL-S156C and histone H3-S10C, were chosen as models for this ‘wash-in’ strategy (**Figure 4b-d**). In both cases, reaction via treatment with DBHDA in deuterated buffer gave by-product tetrahydrothiophene carbodiimide (THTCD) that showed deuterium incorporation (**Figure 4b,c**). MSMS analysis (see **Supplementary Figures 23,24**) unequivocally located the deuteration at the C α ’ carbon of THTCD, not on exchangeable protons. Thus, product ion spectra were collected with a normalised collision energy for each m/z corresponding to d_n -THTCD; the precursor isolation window for MSMS scans was 0.4 Th, sufficient to isolate each isobar. Control experiments showed that neither DBHDA nor THTCD were themselves directly deuterated under the reaction conditions (see **Supplementary Figure 25**), eliminating the possibility of direct H-D side-exchange of either during reaction or analysis. Together these data confirmed the formation of a sulfonium ylid and a site of deprotonation consistent with intermediacy during elimination. Moreover, the detection of doubly-deuterated THTCD-d₂ suggested that the equilibrium between ylid and sulfonium is faster than any subsequent elimination process. Furthermore, no deuteration was observed in Dha product from reactions leading us to tentatively discard α -elimination *via* long-lived carbene formation as a potential mechanism.

Next, having used such ‘wash-in’ experiments, we next chose to probe the mechanism more deeply using the protein variants bearing *pre-installed* deuterium labels at predetermined non-exchangeable sites by exploiting the α -C-deuteration chemistry developed above for this purpose.

Use of Deuteration to Probe Dha Formation

With stable C α -Cys-deuterium-labelled proteins in hand (**Figure 2**), we tested whether internal deuteron transfer (e.g. from C α to ylid) might be observed during formation of Dha from Cys (Figure 4). Deuteration of THTCD was not observed for either H3-dC10 (**Figure 4d**) or H3-dC26 (see **Supplementary Figure 26**). Together these data suggested that if the sulfonium ylid plays the role of an intramolecular base (**Figure 3b**) then any abstraction is rapid on the timescale of elimination, leading to solvent exchange ('wash-out'). To test this hypothesis, we set out to restrict solvent exchange effects through the use of aprotic solvents of similar polarity. Unfortunately, none of the protein substrates proved sufficiently soluble to allow unambiguous experiments. However, application of the C α -deuteration methodology (which was readily extended to small molecules) allowed preparation of small molecule models containing Cys-residues with sufficient solubility in DMSO. Consistent with reactivity on proteins, formation of α -deuterated AcNHCysOMe (AcNHCysOMe-d1) (**Figure 5a**) was best achieved in two steps by reaction of the corresponding Dha-residue (AcNHDhaOMe) with Na₃SPO₃ in D₂O followed by treatment with HCl (aq.) up to pH 3 to afford the desired AcNHCysOMe-d1 with >95% deuterium incorporation in good yield. Use of KSAC in D₂O followed by deacetylation using K₂CO₃ in MeOH also proved possible; whilst the overall recovered yield for this alternative two-step conversion was not outstanding (38%), it also allowed high levels of label incorporation (>95%). Upon treatment with DBHDA, the thus-labelled Dha-model AcNHCysOMe-d1 behaved essentially identically to protein systems (**Figure 5b,c,d** and **Supplementary Figures 27,28**) in 'protic' (aqueous buffer or D₂O) solvents, thereby suggesting its validity as a model in aprotic systems.

Thus, to test the mechanistic feasibility of intramolecular proton transfer in the absence of solvent exchange, AcNHCysOMe-d1 was treated with DBHDA in the presence of base (K₂CO₃) in non-exchangeable solvent DMSO-d₆. Both LC-MS and NMR revealed the formation of THTCD-d1 (**Figure 6a**). Its formation can be explained by the formation of a sulfonium ylid that acts as an intramolecular base to abstract the α -deuterium of the residue in AcNHCysOMe-d1. To further probe the intra- vs inter- molecular nature of this abstraction a series of dilution experiments (**Supplementary Figure 29**) were conducted that confirmed that at higher concentrations (beyond those that may be seen in proteins) deuterium incorporation in THTCD-d1 is lost, consistent with an intra-molecular mechanism via pre-equilibrating ylid formation at lower concentration. It should be noted that such inter-molecular (and hence protein-to-protein) abstraction is much less likely in proteins given the likely more crowded nature as compared with small molecule model AcNHCysOMe.

Computational Analysis of Dha Formation Mechanism

Together, these results on both proteins and amino acid models are consistent with the formation of a sulfonium ylid through deprotonation of sulfonium at C α '. The lifetime of this ylid appears sufficient to allow 'wash-in' from exchangeable solvent. This ylid is also capable of relayed deprotonation of the C α ' position of the Cys residue resulting in formation of site-specifically deuterated THT in non-exchangeable solvent; concentration dependency implicates intra-molecular deprotonation and this deuteration appears to be 'washed-out' in exchangeable solvent (**Figure 6a**). This implicated E1cB elimination with 'wash-out/in' at rates greater than those for elimination.

Density functional theory (DFT) calculations (**Figure 6b**) were used to characterize pathways for a concerted E2 elimination and intramolecular deprotonation via the ylid as internal base. Conformations were sampled systematically and incrementally generate possible rotamers by varying a combination of key dihedral angles. Two key dihedral angles along C α -C β and C β -S of cysteine residue were used. These structures were then used as starting structures to find all possible transition states (TSs). In total, 13 TSs were found for the intermolecular α -deprotonation to form a sulfonium ylid, with six structures having energies within 5 kcal mol⁻¹ of the lowest energy structure (at 15.3 kcal mol⁻¹). A total of sixteen TSs were found for the intramolecular deprotonation by the sulfonium ylid, with ten structures having energies within 5 kcal mol⁻¹ of the lowest energy structure (at 16.6 kcal mol⁻¹). A total of 10 TSs were found for the competing E2 pathway, with 5 structures having energies within 5 kcal mol⁻¹ of the lowest energy structure (at 21.0 kcal mol⁻¹). The final product formations in each pathways were found to be as low as -23.0 to -27.0 kcal mol⁻¹, indicating that the formation of final products was highly exergonic, and that the reaction is irreversible and under kinetic control.

The computed sulfonium ylid formation was found to be relatively facile and marginally endergonic (i.e. uphill, at 4.4 kcal mol⁻¹), making its formation reversible, consistent with the rapid equilibration observed in the labeling experiments. The subsequent step, intramolecular deprotonation (formally a retro-ene reaction) is thermally feasible and is the overall rate-determining step. A concerted transition structure was confirmed by an intrinsic reaction coordinate analysis (HOMO plotted in **Supplementary Figure 30**). By contrast the E2-elimination barrier is ca. 4.4 kcal mol⁻¹ less favourable than that computed for the intramolecular process making this pathway substantially kinetically disfavoured (> 1600 times selectivity for intramolecular deprotonation over E2 deprotonation using simple transition state theory). Together this supported the potential for a Swern-like intramolecular deprotonation event in the elimination.

Discussion

In summary, we have developed a ready, direct, site-selective method to replace hydrogen with deuterium at the α -position of a Cys residue in both amino acid models and proteins. Such α -deuterated protein substrates now prove to be useful mechanistic tools to elucidate detailed protein chemistry. Here, strikingly, they suggest that an unusual sulfonium ylid intermediate is involved in the formation of Dha from Cys (formed using bis-alkylating reagent DBHDA) and that this ylid acts as an internal base to abstract a proton from the protein backbone, which in turn leads to elimination. This is consistent with the lower pK_a s of α -protons of sulfoniums as compared to those of backbone α CH-protons of amino acid residues.⁴¹ Observation of both 'wash-out' and 'wash-in' suggests that the resulting C α carbanion is sufficiently long-lived, this in turn suggests a step-wise E1cB (perhaps E1cB_{anion}) mechanism (**Supplementary Figures 31**). Ylid intermediacy is also supported by the fragmentation of a longer-lived sulfonium in GFP and quantum mechanical calculations. Taken together this mechanism thus appears to exploit the relaying of increasingly effective kinetic basicity, using a sulfonium as an intramolecular base akin to the widely-known and used classical "Swern reaction" oxidative elimination that forms carbonyls from alcohols,⁴² although other mechanisms cannot be discounted. Such a reaction has never before been implicated in a biological system.

The development of site-selective protein labelling (here D) and MSⁿ methodology for tracking that label has therefore proven here to be a powerful analytical combination for unpicking complex, previously unknown, protein chemistry. These have not only allowed study of a useful chemical protein modification reaction, they also reveal a more general approach to studying complex mechanistic analyses in protein chemistry. It is notable that whilst hydrogen isotope methods are very common in small molecule chemistry, they have not been applied to protein chemistry previously. Thus these methods may, in turn, allow better understanding of parameters that determine selectivities and efficiencies of many other key protein reactions.

Given the continuing popularity of Cys as a handle,⁴³ the targeting of Cys (**Figure 1c**) is a useful first proof-of-principle of a potentially more general method (**Figure 1a,b**), which could in principle be applied to other electrophiles and nucleophiles, as well as be combined with either further modification or alteration to other amino acids.^{43, 44} Such access to various labeled variants of amino acids may also not be necessarily restricted to Dha as an intermediate and other variants, such as dehydroaminobutyric acid (Dhb) may also be considered, potentially expanding that range yet further.

Notably, although in this mechanistic study we have largely focused here on the use of this novel, site-selective deuteration method for the creation of alpha-deutero-Cys residues in proteins, we were also able to demonstrate in proof-of-principle experiments (see **Supplementary Figure 1**) that a variety of protected-Cys derivatives could be installed in proteins via parallel thia-Michael reactions. These included, for example photo-caged (and releasable) variants, thereby highlighting potential strategies for use of mechanistic probes with temporal control (e.g. through photo-‘uncaging’).

It is important to note that the method we use here generates epimers at the alpha position that may limit some applications.³⁰ Therefore, to be clear, the formation of some D-epimer renders this process traceless only with regard to the constitution of the bonds formed and not completely traceless in configuration. However, such deutero-epimers, as we show here, can be usefully used to probe mechanism. It can also be anticipated that the greatly reduced radius of gyration that arises from backbone modification as compared to strategies based on more mobile side-chain or heteroatom-D labels will allow structural interrogation with potentially reduced ambiguity in distance constraints, even despite the presence of epimers, and in the context of a low-background spectral window allowing good sensitivity^{45, 46} and/or ‘isotopic editing’.⁴⁷

Finally, other speculative applications can also be envisaged. Alpha-C-deuterated amino acids such as Merck drug candidate MK0641 *Fludalanine*⁴⁸ have shown clear benefits and there is a recent resurgence in the use of such deuteration in drug candidates.⁴⁹ Through the methods we describe here, directly analogous alpha-C-deuterated *proteins* as candidate ‘biologics’ can now be considered and potentially readily explored.

Author Contributions

SRGG, JD, WLN, VC, RQ conducted chemical experiments; XZ, RAS, RSP conducted computational experiments; SRGG, JRW, EP conducted mass-spectrometric experiments; SRGG, JRW, RSP, SC, VC, BGD designed the experiments and analysed the data; SRGG, VC, BGD wrote the paper; all authors read and commented on the paper.

Acknowledgements

We thank BBSRC/AstraZeneca (S.G.), EU Horizon 2020 programme under the Marie Skłodowska-Curie (700124, J.D.), the EPSRC Centres for Doctoral Training in Theory and Modelling in Chemical Sciences (EP/L015722/1) and in Synthesis for Biology and Medicine (EP/L015838/1), Leverhulme Trust (RPG-2017-288, A/N: 176274, V.C.), A*STAR Singapore (X.Z.) and the Croucher Foundation (W.L.N.) for funding. We thank Gogulan Karunanithy, Simon Nadal, Drs Ritu Raj, Ramiz Nathani, Jens Willwacher, Garry Pairaudeau, Jon Read and Profs Alex Breeze, Andrew Baldwin for useful discussions.

Notes

The authors declare no competing financial interests.

Data Availability Statement

All key MS data supporting Figures are given in the SI, and all related raw data are available on request. All primary numerical data for graphical plots in Figures will be deposited as spreadsheets in the Oxford open access depository 'ORA-data'; DOI: 10.5287/bodleian:NG0gbEEzP.

References

1. Westheimer, F.H. The Magnitude of the Primary Kinetic Isotope Effect for Compounds of Hydrogen and Deuterium. *Chemical Reviews* **61**, 265-273 (1961).
2. Cook, P.F. & Cleland, W.W. Mechanistic deductions from isotope effects in multireactant enzyme mechanisms. *Biochemistry* **20**, 1790-1796 (1981).
3. Lemaster, D.M. Uniform and Selective Deuteration in 2-Dimensional Nmr of Proteins. *Annu. Rev. Biophys. Bio.* **19**, 243-266 (1990).
4. Groff, D., Thielges, M.C., Cellitti, S., Schultz, P.G. & Romesberg, F.E. Efforts Toward the Direct Experimental Characterization of Enzyme Microenvironments: Tyrosine100 in Dihydrofolate Reductase. *Angew. Chem. Intl Ed.* **48**, 3478-3481 (2009).
5. Adhikary, R., Zimmermann, J., Dawson, P.E. & Romesberg, F.E. IR Probes of Protein Microenvironments: Utility and Potential for Perturbation. *ChemPhysChem* **15**, 849-853 (2014).
6. Liu, D.S., Xu, R. & Cowburn, D. Segmental Isotopic Labeling of Proteins for Nuclear Magnetic Resonance. *Methods Enzymol.* **462**, 151-175 (2009).
7. Cremeens, M.E., Zimmermann, J., Yu, W., Dawson, P.E. & Romesberg, F.E. Direct Observation of Structural Heterogeneity in a beta-Sheet. *J. Am. Chem. Soc.* **131**, 5726-5727 (2009).
8. Markley, J.L., Putter, I. & Jardetzky, O. High-Resolution Nuclear Magnetic Resonance Spectra of Selectively Deuterated Staphylococcal Nuclease. *Science* **161**, 1249-1251 (1968).
9. Feeney, J. et al. ¹H nuclear magnetic resonance studies of the tyrosine residues of selectively deuterated *Lactobacillus casei* dihydrofolate reductase. *Proc R Soc Lond B Biol Sci* **196**, 267-290 (1977).
10. Feeney, J., Birdsall, B., Ostler, G., Carr, M.D. & Kairi, M. A Novel Method of Preparing Totally Alpha-Deuterated Amino-Acids for Selective Incorporation into Proteins - Application to Assignment of H-1 Resonances of Valine Residues in Dihydrofolate-Reductase. *FEBS Lett.* **272**, 197-199 (1990).
11. Chin, J.K., Jimenez, R. & Romesberg, F.E. Direct observation of protein vibrations by selective incorporation of spectroscopically observable carbon-deuterium bonds in cytochrome c. *J. Am. Chem. Soc.* **123**, 2426-2427 (2001).
12. Xia, B., Jenk, D., LeMaster, D.M., Westler, W.M. & Markley, J.L. Electron-Nuclear Interactions in Two Prototypical [2Fe-2S] Proteins: Selective (Chiral) Deuteration and Analysis of ¹H and ²H NMR Signals from the Alpha and Beta Hydrogens of CysteinyI Residues That Ligate the Iron in the Active Sites of Human Ferredoxin and Anabaena 7120 Vegetative Ferredoxin. *Arch Biochem. Biophys.* **373**, 328-334 (2000).
13. Markley, J.L. Correlation Proton Magnetic-Resonance Studies at 250-Mhz of Bovine Pancreatic Ribonuclease .1. Reinvestigation of Histidine Peak Assignments. *Biochemistry* **14**, 3546-3554 (1975).
14. Ortega, M.A. & van der Donk, W.A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **23**, 31-44 (2016).
15. Marchand, D.H., Rummel, R.P. & Abdel-Monem, M.M. Biliary excretion of a glutathione conjugate of busulfan and 1,4-diiodobutane in the rat. *Drug Metabolism and Disposition* **16**, 85-92 (1988).
16. Wang, J., Schiller, S.M. & Schultz, P.G. A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angewandte Chemie International Edition* **46**, 6849-6851 (2007).
17. Chalker, J.M. et al. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2**, 1666-1676 (2011).

18. Morrison, P.M., Foley, P.J., Warriner, S.L. & Webb, M.E. Chemical generation and modification of peptides containing multiple dehydroalanines. *Chem. Commun.* **51**, 13470-13473 (2015).
19. Wever, W.J., Bogart, J.W. & Bowers, A.A. Identification of Pyridine Synthase Recognition Sequences Allows a Modular Solid-Phase Route to Thiopeptide Variants. *J. Am. Chem. Soc.* **138**, 13461–13464 (2016).
20. Metanis, N., Keinan, E. & Dawson, P.E. Traceless Ligation of Cysteine Peptides Using Selective Deselenization. *Angew. Chem. Int. Ed.* **49**, 7049-7053 (2010).
21. Campbell, S., Rodgers, M.T., Marzluff, E.M. & Beauchamp, J.L. Deuterium Exchange Reactions as a Probe of Biomolecule Structure. Fundamental Studies of Gas Phase H/D Exchange Reactions of Protonated Glycine Oligomers with D₂O, CD₃OD, CD₃CO₂D, and ND₃. *Journal of the American Chemical Society* **117**, 12840-12854 (1995).
22. Bernardes, G.J., Chalker, J.M., Errey, J.C. & Davis, B.G. Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: versatile and switchable access to functionalized proteins. *J Am Chem Soc* **130**, 5052-5053 (2008).
23. Chalker, J.M., Lercher, L., Rose, N.R., Schofield, C.J. & Davis, B.G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chem. Int. Ed.* **51**, 1835-1839 (2012).
24. Wang, Z.U. et al. A Facile Method to Synthesize Histones with Posttranslational Modification Mimics. *Biochemistry* **51**, 5232-5234 (2012).
25. Timms, N. et al. Structural Insights into the Recovery of Aldolase Activity in N-Acetylneuraminic Acid Lyase by Replacement of the Catalytically Active Lysine with γ -Thialysine by Using a Chemical Mutagenesis Strategy. *ChemBioChem* **14**, 474-481 (2013).
26. Nathani, R.I. et al. A novel approach to the site-selective dual labelling of a protein via chemoselective cysteine modification. *Chem. Sci.* **4**, 3455-3458 (2013).
27. Chooi, K.P. et al. Synthetic Phosphorylation of p38 α Recapitulates Protein Kinase Activity. *J. Am. Chem. Soc.* **136**, 1698-1701 (2014).
28. Haj-Yahya, N. et al. Dehydroalanine-Based Diubiquitin Activity Probes. *Org. Lett.* **16**, 540-543 (2014).
29. Rowan, F., Richards, M., Widya, M., Bayliss, R. & Blagg, J. Diverse Functionalization of Aurora-A Kinase at Specified Surface and Buried Sites by Native Chemical Modification. *PLoS One* **9** (2014).
30. Meledin, R., Mali, S.M., Singh, S.K. & Brik, A. Protein ubiquitination via dehydroalanine: development and insights into the diastereoselective 1,4-addition step. *Organic & Biomolecular Chemistry* **14**, 4817-4823 (2016).
31. Palioura, S., Sherrer, R.L., Steitz, T.A., Soll, D. & Simonovic, M. The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. *Science* **325**, 321-325 (2009).
32. Wright, T.H. et al. Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, 597 (2016).
33. Holmes, T.J. & Lawton, R.G. Cysteine Modification and Cleavage of Proteins with 2-Methyl-N1-Benzenesulfonyl-N4-Bromoacetylquinonediimide. *J. Am. Chem. Soc.* **99**, 1984-1986 (1977).
34. Nathani, R., Moody, P., Smith, M.E.B., Fitzmaurice, R.J. & Caddick, S. Bioconjugation of Green Fluorescent Protein via an Unexpectedly Stable Cyclic Sulfonium Intermediate. *ChemBioChem* **13**, 1283-1285 (2012).
35. Franzen, V. & Mertz, C. Zum Mechanismus Der Hofmann-Eliminierung Bei Sulfoniumsalzen. *Chem. Ber.* **93**, 2819-2824 (1960).

36. Cristol, S.J. & Stermitz, F.R. Mechanisms of elimination reactions. XXIL. Some cis- and trans-2-phenylcyclohexyl derivatives. The Hoffmann elimination. *J. Am. Chem. Soc.* **82**, 4692-4699 (1960).
37. Franzen, V. & Schmidt, H.J. Zum Mechanismus Der Hofmann-Eliminierung Bei Sulfoniumsalzen .2. *Chem. Ber.* **94**, 2937-2942 (1961).
38. Weygand, F. & Daniel, H. Fragmentierung Von S-Methyl-Thiolanium-Jodid Mit Phenyllithium Zu Athylen Und Methyl-Vinyl-Sulfid. *Chem. Ber.* **94**, 3145-3146 (1961).
39. Banait, N.S. & Jencks, W.P. Elimination-Reactions - Experimental Confirmation of the Predicted Elimination of (Beta-Cyanoethyl)Sulfonium Ions through a Concerted, E2 Mechanism. *J. Am. Chem. Soc.* **112**, 6950-6958 (1990).
40. Grob, C.A. & Schiess, P.W. Heterolytic Fragmentation . A Class of Organic Reactions. *Angew. Chem. Intl Ed.* **6**, 1-& (1967).
41. Bordwell, F.G. Equilibrium Acidities in Dimethyl-Sulfoxide Solution. *Acc. Chem. Res.* **21**, 456-463 (1988).
42. Omura, K. & Swern, D. Oxidation of Alcohols by Activated Dimethyl-Sulfoxide - Preparative, Steric and Mechanistic Study. *Tetrahedron* **34**, 1651-1660 (1978).
43. Chalker, J.M., Bernardes, G.J.L., Lin, Y.A. & Davis, B.G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. Asian J.* **4**, 630-640 (2009).
44. Dawson, P.E. Native Chemical Ligation Combined with Desulfurization and Deselenization: A General Strategy for Chemical Protein Synthesis. *Israel Journal of Chemistry* **51**, 862-867 (2011).
45. Kinnaman, C.S., Cremeens, M.E., Romesberg, F.E. & Corcelli, S.A. Infrared Line Shape of an α -Carbon Deuterium-Labeled Amino Acid. *Journal of the American Chemical Society* **128**, 13334-13335 (2006).
46. Adhikary, R., Zimmermann, J. & Romesberg, F.E. Transparent Window Vibrational Probes for the Characterization of Proteins With High Structural and Temporal Resolution. *Chemical Reviews* **117**, 1927-1969 (2017).
47. Kay, L.E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance* **89**, 496-514 (1990).
48. Darland, G.K. et al. Oxidative and defluorinative metabolism of fludalanine, 2-2H-3-fluoro-D-alanine. *Drug Metabolism and Disposition* **14**, 668-673 (1986).
49. Timmins, G.S. Deuterated drugs; where are we now? *Expert opinion on therapeutic patents* **24**, 1067-1075 (2014).

Online Methods

Analytical tools for protein reaction monitoring and characterisation

For SBL and H3, liquid chromatography-mass spectrometry (LC-MS) was performed on a Micromass LCT (ESI-TOF-MS) coupled to a Shimadzu HPLC using a Thermo Proswift (250 x 4.6 mm x 5 μ m). Water:acetonitrile, 95:5 (solvent A) and acetonitrile (solvent B), each containing 0.1% formic acid, were used as the mobile phase at a flow rate of 0.4 mL min⁻¹. The gradient was programmed as follows: 95% A (5 min isocratic) to 100% B after 15 min then isocratic for 5 min. The electrospray source of LCT was operated with a capillary voltage of 3.2 kV and a cone voltage of 25 V. Nitrogen was used as the nebulizer and desolvation gas at a total flow of 600 L h⁻¹. Spectra were calibrated using a calibration curve constructed from a minimum of 17 matched peaks from the multiply charged ion series of equine myoglobin, which was also obtained at a cone voltage of 25 V. Total mass spectra were reconstructed from the ion series using the MaxEnt algorithm preinstalled on MassLynx software (v. 4.0 from Waters) according to manufacturer's instructions.

For GFP, LC-MS was performed on protein samples using a Thermo Scientific uPLC connected to MSQ Plus Single Quad Detector (SQD). Column: Hypersil Gold C4, 1.9 μ m, 2.1 x 50 mm. Wavelength: 254 nm. Mobile Phase: 99:1 Water (0.1% formic acid): MeCN (0.1% formic acid) to 1:9 Water (0.1% formic acid): MeCN (0.1% formic acid) gradient over 4 min. Flow Rate: 0.3 mL/min. MS Mode: ES+. Scan Range: m/z = 500-2000. Scan time: 1.5 s. Data obtained in continuum mode. The electrospray source of the MS was operated with a capillary voltage of 3.5 kV and a cone voltage of 50 V. Nitrogen was used as the nebulizer and desolvation gas at a total flow of 600 L/h. Ion series were generated by integration of the total ion chromatogram (TIC) over the 3.4-4.8 min range. Total mass spectra for protein samples were reconstructed from the ion series using the pre-installed ProMass 2.8 software (Thermo) using default settings for large proteins in m/z range 500-1500.

Characterisation of intact protein was performed using a Xevo G2-S Q-ToF high resolution mass spectrometer (Waters, Wilmslow, UK) coupled with an Acquity UPLC system. The column was a Chromolith RP18 2x50mm (Merck). Elution solvents were: (A) deionised water (18M Ω resistivity) with 0.1% (v/v) formic acid (analytical grade, Sigma UK), and (B) Acetonitrile (HPLC grade, Sigma UK), isocratic 95% A for 0.5 minutes followed by gradient to 5% A for 2.5 minutes, isocratic for 0.5 minutes, then gradient for 0.05 minutes to 95% A followed by isocratic for 1.45 minutes. Instrument control, data acquisition, and data processing were performed using Masslynx 4.1 software. The mass spectrometer was operated in positive electrospray mode,

source conditions were adjusted to maximise sensitivity. Lockspray™ was used during analysis to maintain mass accuracy. The data was processed using Masslynx and deconvoluted using MaxEnt.

Tryptic peptides were analysed by an Orbitrap Elite coupled with nanoLC. The data was processed using Maxquant for site determination and Xcalibur for determination of deuterium incorporation.

Expression and purification of proteins

The gene for histone H3S10C and H3R26C was generated and expressed as described previously.²³ The gene for GFP(S147C) in the vector pNIC28-Bsa4 was generated and expressed as described previously.³⁴

Formation of GFPSul147 and Identification of Fragmentation Products

Formation of GFPSul147 (Supplementary Figure 19): Dithiothreitol (10 µL, 340 mM as a solution in H₂O, 100 equiv.) was added to a solution of GFPS147C (100 µL, 1 mg/mL) in sodium phosphate buffer (100 mM, pH 8.0), and the solution incubated at 21 °C for 1 hour. The excess reducing agent was removed by repeated diafiltration into fresh buffer using VivaSpin sample concentrators (GE Healthcare, 10000 MWCO). 2,5-dibromohexanediamide (DBHDA, 10 µL, 17 mM as a solution in DMF, 50 equiv.) was added to the reduced protein solution and the mixture vortexed for 1s before being incubated with shaking (2 hours, 37 °C, 300 rpm). LCMS analysis confirmed the formation of GFPSul147.

Fragmentation of GFP-Sul147 at pH 8 (sodium phosphate) (Supplementary Figure 19):

GFPSul147 (100 µL, 1 mg/mL) in sodium phosphate buffer (100 mM, pH 8.0) was incubated with shaking (24 hours, 37 °C, 300 rpm). Fragmented products were observed using LCMS.

GFP-Sul147 “Wash-in” Studies (Supplementary Figure 21,22): Deuterated sodium phosphate buffer (100 mM, pH 8.0) was prepared by dissolving sodium phosphate monobasic monohydrate (0.096 g) and sodium phosphate dibasic (1.32 g) in deuterium oxide (50 mL). The solution was then freeze-dried for 24 h, and the resulted salts were re-dissolved in deuterium oxide (50 mL). This operation was repeated 5 times. Dithiothreitol (10 µL, 340 mM as a solution in D₂O, 100 equiv.) was added to a solution of GFP-S147C (100 µL, 1 mg/mL) in deuterated sodium phosphate buffer (100 mM, pH 8.0), and the solution incubated at 21 °C for 1 hour. The excess reducing agent was removed by repeated diafiltration into fresh deuterated buffer using VivaSpin sample concentrators (GE Healthcare, 10000 MWCO). 2,5-Dibromohexanediamide (DBHDA, 10 µL, 17 mM as a solution in DMSO-d₆, 50 equiv.) was added to the reduced protein

solution and the mixture vortexed for 1 s before being incubated (24 hours, 37 °C, 300 rpm). LCMS analysis revealed a mix of deuterated GFPSul147 with fragment on-protein and deuterated small-molecule fragments.

Formation of dehydroalanine in deuterated buffer

Preparation of deuterated buffers: The deuterated sodium phosphate buffer was prepared by dissolving sodium phosphate monobasic monohydrate (0.24 g) and sodium phosphate dibasic (3.31 g) in deuterium oxide (50 mL). The solution was then freeze-dried for 24 h, and the resulted salts were re-dissolved in deuterium oxide (50 mL). This operation was repeated 5 times.

Formation of SBLDha156 in deuterated buffer (Supplementary Figure 23): Lyophilised SBLS156C (5 mg) in deuterated 50 mM sodium phosphate buffer pH 8 (reaction buffer) was treated with a 1 M solution of DTT in reaction buffer. The mixture was shaken at 600 rpm at room temperature. The mixture was desalted using a PD-Miditrap™ G25 preequilibrated with reaction buffer to afford 1.5 mL of a 2 mg/mL solution. The solution was treated with DBHDA in solution in DMF (200 Eq. in 1 µL of DMF). The mixture was shaken at 600 rpm at room temperature for 30 minutes and at 37 °C. The reaction was monitored by LCMS. The reaction was typically complete after 2 hours. The mixture was analysed by LC-MSMS.

General Method for Formation of 'H3DhaX' in deuterated buffer (Supplementary Figure 24): A solution of lyophilised H3CX was prepared at a concentration of 2.0 mg/mL in deuterium oxide. A 250 µL aliquot of the protein solution was diluted with 250 µL of pH 8.0 sodium phosphate buffer (50 mM). To this diluted protein solution was then added DTT (10 mg). The solution was shaken at room temperature for 15 minutes to reduce any contaminant disulfide. After this time, the protein solution was passed through a PD Minitrap G25, previously equilibrated with deuterium oxide and eluted with 1.0 mL of deuterium oxide. The histone solution was stored on ice until needed. A stock solution of DBHDA was prepared by dissolving 84 mg in 564 µL of DMF. A 250 µL aliquot of the reduced protein was diluted with 250 µL of 50 mM sodium phosphate pH 8.0 and warmed to room temperature. Once the protein was at room temperature, a 50 µL aliquot of the DBHDA solution was added. The reaction was vortexed and then shaken at room temperature for 30 minutes and then 37 °C for 60 minutes. The reaction mixture was then cooled to room temperature and precipitated DBHDA was removed by centrifugation (2 min, 14K g). The reaction was monitored by LCMS and analysed by LCMSMS.

Site selective α -deuteration of proteins

Preparation of protein and reagents to minimize the presence of exchangeable protons: Protein was lyophilised 3 times in deuterium oxide. Deuterated buffer: For 50 mL solution, desired amount of buffer components were dissolved in the desired volume of deuterium oxide. After adjusting the pH and volume, the solutions were lyophilised and resolubilised with fresh deuterium oxide.

Exploratory Reactions with Sodium Sulfide (Supplementary Figure 1c): To a solution of SBLS156Dha in 0.1 M deuterated sodium phosphate buffer (pH 7, 1.43 mg/mL, 0.15 mL) was added an aqueous solution of Na₂S (6 µL, 50 µg/µL, dissolved in pH 7 deuterated sodium phosphate buffer). The solution was briefly vortexed and shaken at 15 °C for 10 mins. At 10, 20, 30, 40, and 50 mins, 6 µL of the Na₂S solution was added. The solution was kept shaking in between each addition. After 10 mins of the last addition, the reaction mixture was analysed by protein MS to ensure complete conversion. Na₂S was removed by filtering through a Micro Bio-Spin™ 6 Columns (Bio-Rad) (pre-equilibrated with 50 mM sodium phosphate buffer, pH 8). The filtrate was filtered again using another pre-equilibrated Micro Bio-Spin™ 6 Columns. For protein MS analysis, 3 µL of the reaction mixture was taken and transferred to a MS vial. 50 µL of MQ-water and 3 µL of 1 M HCl solution were then added. The mixture was then analyzed by LC-MS (Remark: addition of dil. HCl (aq) to acidify before MS analysis can improve the quality of MS spectra).

Exploratory Reactions with Sodium Thioacetate (Supplementary Figure 1c):

Monitoring of SBLS156Dha treated with KSAc at pH 8 (A. t=0 h, B. t=1 h) showed formation of SBLS156Acc. Attempted deprotection of SBLS156CysAc to SBLS156C using NH₂OMe (500 equiv.) in 50 mM NaPi pH 8 returned only starting material after 4 h.⁵¹

Typical procedure for Protein Deuteration via Thiophosphate: Anhydrous sodium thiophosphate was prepared as described previously from commercially available sodium thiophosphate hydrate.⁵⁰

Formation of H3DhaX (Supplementary Figures 2a,18): Lyophilized histone H3CX (5 mg) was solubilised in reaction buffer, deuterated 50 mM Tris, 5 M GdmCl pH 8 buffer (500 µL). DTT solid (1 mg) was added to the solution to reduce the dimer and the β-mercaptoethanol adduct formed during purification and storage. The mixture was shaken at 600 rpm at room temperature for 30 minutes. The mixture was desalted using PD-Minitrap™ G25 preequilibrated with reaction buffer to afford 1 mL of solution. A stock solution of DBHDA in DMF was added. The mixture was shaken at 600 rpm at room temperature for 30 minutes before heating up to 37 °C for typically 3 hours. The reaction was monitored using LCMS.

Formation of H3dpCX (Supplementary Figure 2a): The reaction was monitored by LCMS to avoid overalkylation. One aliquot of the mixture (500 μ L) was desalted and buffer exchanged using PD-Minitrap G25 preequilibrated with a 0.4 M Na₃SPO₃, 5 M GdmCl pH 8.5 buffer to afford 1 mL of solution. The mixture was shaken at 600 rpm at 37 °C for typically 2 hours. One aliquot of the mixture (500 μ L) was desalted and buffer exchanged using PD-Minitrap™ G25 preequilibrated with a 50 mM Tris, 5 M GdmCl pH 8 buffer to afford 1 mL of solution. The reaction was monitored using LCMS. Observed differences in mass during synthesis were also due to slow exchangeable deuterons inside the protein.

Formation of H3dCX (Supplementary Figures 2b,2c,18):

Method A. The mixture was acidified to pH 4 with addition of glutathione. The mixture was shaken at 600 rpm at 37 °C for typically 1 hour to afford the desired H3dCX. The reaction was monitored by LCMS. The solution was a mixture of free and glutathione (Glu) adduct (+305) H3dCX. Treatment with a 10 mM DTT aqueous solution afforded the desired H3dCX.

Method B. The mixture was buffer exchanged using a PD-Spintrap G25 equilibrated with 50 mM Tris, 100 mM NaCl, pH 8 buffer. The enzymatic reaction was set up according to PP1 provider (Biolabs, P0754S): 4 μ L of H3dpCX (1mg/mL)+ 4 μ L of NE buffer (x10)+ 4 μ L of MnCl₂ buffer (x10)+0.4 μ L PP1 (2500 U/mL). The volume was adjusted to 40 μ L with Milli-Q® water and incubated at 30 °C. The reaction was monitored by LCMS. The reaction was completed after 30 minutes for H3dC10 and 2 h for H3dC26. An adduct was observed by LCMS corresponding to thiophosphate (+112). Treatment with a 10 mM DTT aqueous solution afforded the desired H3dCX. In both cases, the mixtures were then dialysed using refolding buffer 20 mM Tris, 250 mM KCl, 1 mM EDTA to afford the desired H3dCX and DTT was added to afford a 20 mM Tris, 250 mM KCl, 1 mM EDTA, 5 mM DTT solution.

Exploratory Reactions with Sodium Thioacetate (Supplementary Figure 1c):

Monitoring of SBLS156Dha treated with KSAc at pH 8 (A. t=0 h, B. t=1 h) showed formation of SBLS156Acc. Attempted deprotection of SBLS156CysAc to SBLS156C using NH₂OMe (500 equiv.) in 50 mM NaPi pH 8 returned only starting material after 4 h.⁵¹

Deuterium incorporation detection (Supplementary Figure 3)

Deuterium incorporation estimation using LCMS on intact protein (Supplementary Figure 4 and Supplementary Tables 1,2):

Aliquot of 0.5 mg/mL aliquots of the desired protein were prepared in water. The aliquots were analysed using Xevo G2-S Q-Tof. The data was process using MaxLynx v4.1. Deconvolution was set up: output mass, ranges 10000:20000, resolution 0.04 Da/channel; damage model,

uniform Gaussian width at half height 0.5 Da, minimum intensity ratios left 33% and right 33% and completion options, iterate to convergence. Two independent runs were performed on each sample.

Percentage deuterium incorporation (%D) on intact protein was determined as follow:

$$M_{\text{obs.}} = M_{\text{deut.}} \times \%D + M_{\text{non-deut.}} \times (1 - \%D)$$

with $M_{\text{obs.}}$ = Mass observed, $M_{\text{deut.}}$ = Mass of deuterated protein, %D = percentage of deuterated protein or percentage deuterium incorporation, $M_{\text{non-deut.}}$ = mass of non-deuterated protein.

This equation was equivalent to:

$$\%D = (M_{\text{obs.}} - M_{\text{non-deut.}}) / (M_{\text{deut.}} - M_{\text{non-deut.}})$$

As $M_{\text{deut.}} - M_{\text{non-deut.}} = 1$, the percentage incorporation could be estimated as $\%D = M_{\text{obs.}} - M_{\text{non-deut.}}$.

Control isotopic pattern determination (Supplementary Figure 4):

The deconvolution was performed with a uniform Gaussian at half height of 0.04 Da to allow the observation of the isotopic pattern of the desired mass, which confirmed that the variation in mass was due to a deuterium addition. A typical spectra is shown in **Supplementary Figure 4**.

Validation of deuterium incorporation levels and conformation of location of deuterium by

LCMSMS of tryptic peptide: An aliquot of H3 (1 mg/mL, 25 μ L) was diluted with a 8 M aqueous solution of urea (75 μ L). The sample was incubated at room temperature for 10 minutes with occasional vortexing. The protein was reduced with DTT (200 mM solution in water) at 56 °C for 25 minutes and alkylated with iodoacetamide (400 mM solution in water) at room temperature for 30 minutes in the dark. The sample was diluted with ammonium bicarbonate buffer and trypsin added (final enzyme:protein 1:50 (w/w) ratio). The sample was incubated overnight at 37 °C. The digested protein was diluted to a final concentration of 100 fmol/ μ L with MilliQ water. The sample was analysed by LCMSMS.

Data processing as confirmation of location of deuterium (Supplementary Figure 4 and

Supplementary Tables 1,2): The data was processed using Mascot on own database containing H3 variants with variable modifications: custom-made deuterated carbamidomethyl on cysteine residue; peptide mass tolerance: ± 20 ppm; fragment mass tolerance: ± 0.5 Da; max missed cleavages: 3; Instrument type: default.

For H3dC10

Protein sequence coverage: 68%; Matched peptides shown in **bold**.

1 MARTKQTARK **CTGGKAPRKQ** LATKAARKSA PATGGVKKPH RYRPGTVALR
51 EIRRYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV MALQEASEAY
101 LVALFEDTNL AAIAHAKRVTI **MPKDIQLARR** IRGERA

For H3dC26: Protein sequence coverage: 90%; Matched peptides shown in **bold**.

1 MARTK**QTARK** **STGGKAPRKQ** **LATKAACKSA** **PATGGVKKPH** RYRPGTVALR
51 EIRRYQKSTE LLIRKL**PFQR** LVREIAQDFK TDLRFQSSAV MALQEASEAY
101 LVALFEDTNL AAIHAKRVTI MPKDIQLARR IRGERA

Deuterium incorporation estimation using LCMS on tryptic peptide (Supplementary Figures 6,7,8): The data generated by LCMSMS of tryptic digest were processed using Thermo Xcalibur Qual browser. MS1 was processed using chromatogram ranges: base peak, scan filter full ms, range: m/z value of peptide with automatic processing (mass tolerance 10 ppm and 4 decimals 4). The m/z values used were determined in the mascot analysis. The full desired peak was integrated to afford an average of the mass of the desired peptide. 3 different peptides were used as triplicates. For example, MS1 analysis of the peptide 4-18 TKQTARKC_{CAM}TGGKAPR, m/z 415.7320 from the tryptic digests of H3dC10 batches prepared using conditions 1, 2, 3, 4 and 5 (see table S2) showed different mixture of TKQTARKC_{CAM}TGGKAPR and TKQTARKdC_{CAM}TGGKAPR. The comparison between isotopic patterns of non-deuterated peptide and the sample peptide allowed the calculation of the percentage deuteration incorporation.

Statistics and Data Analysis

Bar graphs represent the mean values of the relative intensities (normalized with respect to the most intense peak) of the mass of unlabelled and labelled THTCD, in which the ¹³C and ³⁴S isotopologue influence has been subtracted based on the theoretical unlabelled THTCD isotopic pattern (100% for 0; 6.5% for +1, and 4.5% for +2). Mass spectra were generated from n=2 independent experiments. The individual data points are overlaid as circles on the bars. Standard deviations were calculated to generate the error bars.

Analysis of Stereoselectivity of Deuteration

Analysis of (D-/L-)C10 ratio in histone H3 (Supplementary Figures 9-17): Histone samples (1 µg/µL, 100 µL in 3 M guanidinium chloride, 100mM sodium phosphate buffer pH 8.5) were mixed with DTT (500 mM in water, 1 µL) and incubated at 25 °C for 15 minutes. 2-Chloroacetamide (500 mM freshly prepared in 100 mM TEAB buffer, 10 µL) was added and the mixture was incubated at 25 °C for 45 minutes in a dark. The excess of 2-chloroacetamide was quenched with DTT (500 mM in water, 11 µL) at 25 °C for 30 minutes. Alkylated histone sample was buffer-exchanged to ArgC incubation buffer (50 mM Tris, 5 mM CaCl₂, 2mM EDTA,

pH 7.8) using SpinTrap G-25 column (GE Healthcare). ArgC protease (PeproTech, 4 µg in 20 µL) and ArgC activation buffer (50 mM Tris, 50 mM DTT, 2 mM EDTA; 13 µL) were added and the digestion mixture was shaken (300 rpm) at 37 °C for 4 hours. Peptides were desalted using Oasis HLB 1cc 10mg column (Waters) following manufacturer protocol. Dried peptides were re-dissolved in 0.1% formic acid, 2% acetonitrile (100 µL) and analyzed on an NanoAcquity-UPLC system (Waters) coupled to an Orbitrap Elite mass spectrometer (Thermo Fischer Scientific) possessing an EASY-Spray nano-electrospray ion source (Thermo Fischer Scientific). Peptides (2 µL injection in 0.05% TFA) were separated on an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. × 15 mm, RSLC C18, 3 µm, 100 Å) using a linear gradient (length: 60 minutes; 0-3 min 1% B, 3-33 min 35% B, 33-38 min 99% B, 38-41 min 99% B, 41-44 min 1% B, 44-47 min 99% B, 47-50 min 99% B, 50-60 min 1% B; solvent A (0.1% formic acid), B (0.1% formic acid in acetonitrile), flow rate: 300 nL/min). Data-dependent MS acquisition full-scan MS spectra were collected (scan range 350–1500 m/z, resolution 120000, AGC target 1E6, maximum injection time 100 ms). After the MS scans, the 10 most intense peaks were selected for HCD fragmentation at 35% of normalized collision energy. HCD spectra were acquired in the Iontrap (AGC target 1e3, maximum injection time 50 ms).

Reaction of H3dC in aqueous buffer (Supplementary Figure 26):

A solution of H3dCX was treated buffer exchanged to 50 mM sodium phosphate buffer pH 8. The concentration was adjusted to 0.5 mg/mL. The mixture was treated with a stock solution of DBHDA in DMF (final 200 equiv. in 5% DMF). The reaction was vortexed and then shaken at room temperature for 30 minutes and then 37 °C for 60 minutes. The reaction mixture was then cooled to room temperature. The reaction was monitored by LCMS and analysed by LCMSMS.

Small Molecule Reactions.

See **Supplementary Note** and **Supplementary Figures 27-29**.

Computational Methods

Density functional theory (DFT) calculations were performed with *Gaussian09* rev D.01 (Frisch, M.J. et al. (Gaussian, Inc., Wallingford CT, USA; 2009)). The hybrid meta-generalized gradient (GGA) M06-2X exchange-correlation functional⁵² with a fine integration grid and the diffuse/polarized 6-311++G(d,p) basis set for all elements, were used. A conductor-like polarizable continuum model (CPCM) of DMSO solvation was used during all optimizations.^{53, 54} All stationary points were verified as either minima or first-order saddle points/transition states

(TSs) by the presence of zero or a single imaginary harmonic vibrational frequency, respectively. Relative free energies were evaluated at 298.15 K applying a haptic translational entropy correction to adjust from a standard-state of 1 atm to 1 mol/L, while a quasi-harmonic approximation was used in which the treatment of vibrational entropies switches from a rigid-rotor harmonic oscillator to a free rotor at frequencies below 100 cm^{-1} as described previously.^{55, 56} All structures and molecular orbitals were visualised using *PyMOL* software (The PyMOL Molecular Graphics System v. 2.0 (Schrödinger, LLC.)). Geometries of all TS structures (in .xyz format, including their associated energies in Hartrees) from all conformational samplings are given in a separate folder named *TS_xyz*. All TS conformers were named in ascending order of energy, with *c1* being the lowest energy conformer. Absolute values (Hartrees) for SCF energy, zero-point vibrational energy (ZPE), enthalpy and free energy (at 298 K) for the lowest energy structures are given in **Supplementary Table 3**.

Methods References

50. Yasuda, S.K. & Lambert, J.L. Preparation and Properties of Anhydrous Trisodium and Tripotassium Monothiophosphates. *J. Am. Chem. Soc.* **76**, 5356-5356 (1954).
51. Hermanson, G.T. Bioconjugate techniques. (Academic Press, 1995).
52. Zhao, Y. & Truhlar, D.G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).
53. Barone, V. & Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **102**, 1995-2001 (1998).
54. Cossi, M., Rega, N., calmani, G. & Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **24**, 669-681 (2003).
55. Hodgson, D.M., Charlton, A., Paton, R.S. & Thompson, A.S. C-Alkylation of Chiral Tropane- and Homotropene-Derived Enamines. *J. Org. Chem.* **78**, 1508-1518 (2013).
56. Simón, L. & Paton, R.S. Origins of Asymmetric Phosphazene Organocatalysis: Computations Reveal a Common Mechanism for Nitro- and Phospho-Aldol Additions. *J. Org. Chem.* **80**, 2756-2766 (2015)

FIGURES

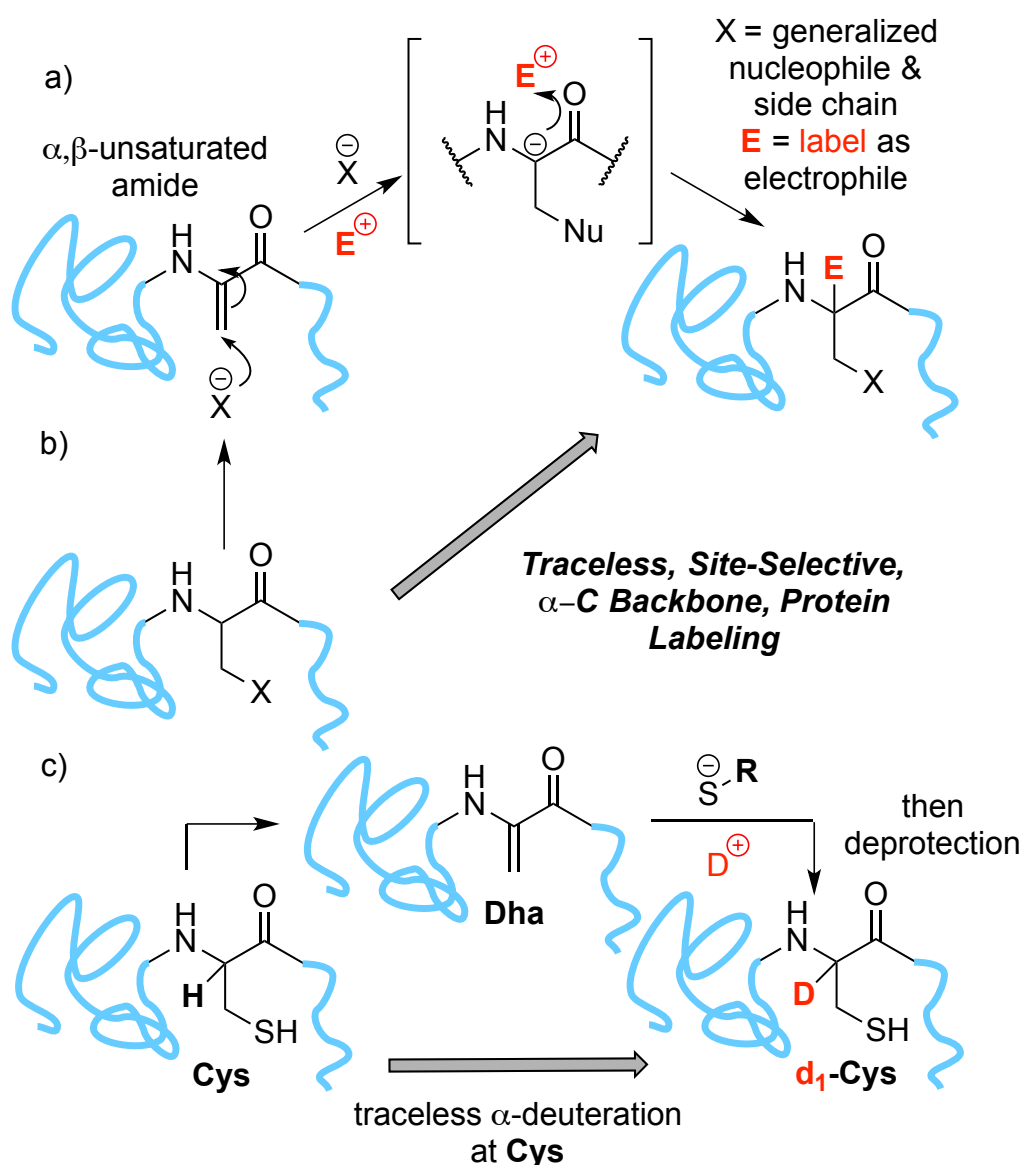


Figure 1. Traceless Post-Translational Site-selective Protein α -Deuteration Strategy. (a) Strategy for a generalized, site-selective α -C labelling approach with any electrophile E in an intact protein (light blue ribbon cartoon) backbone. (b) By combining this with a method of both deriving the key α,β -unsaturated amide intermediate from side-chain X *and* forming the same side-chain X back again then the process can be rendered traceless. (c) Proof-of-principle H \rightarrow D replacement process at the α -carbon of Cys demonstrated in this work that exemplifies the general strategy.

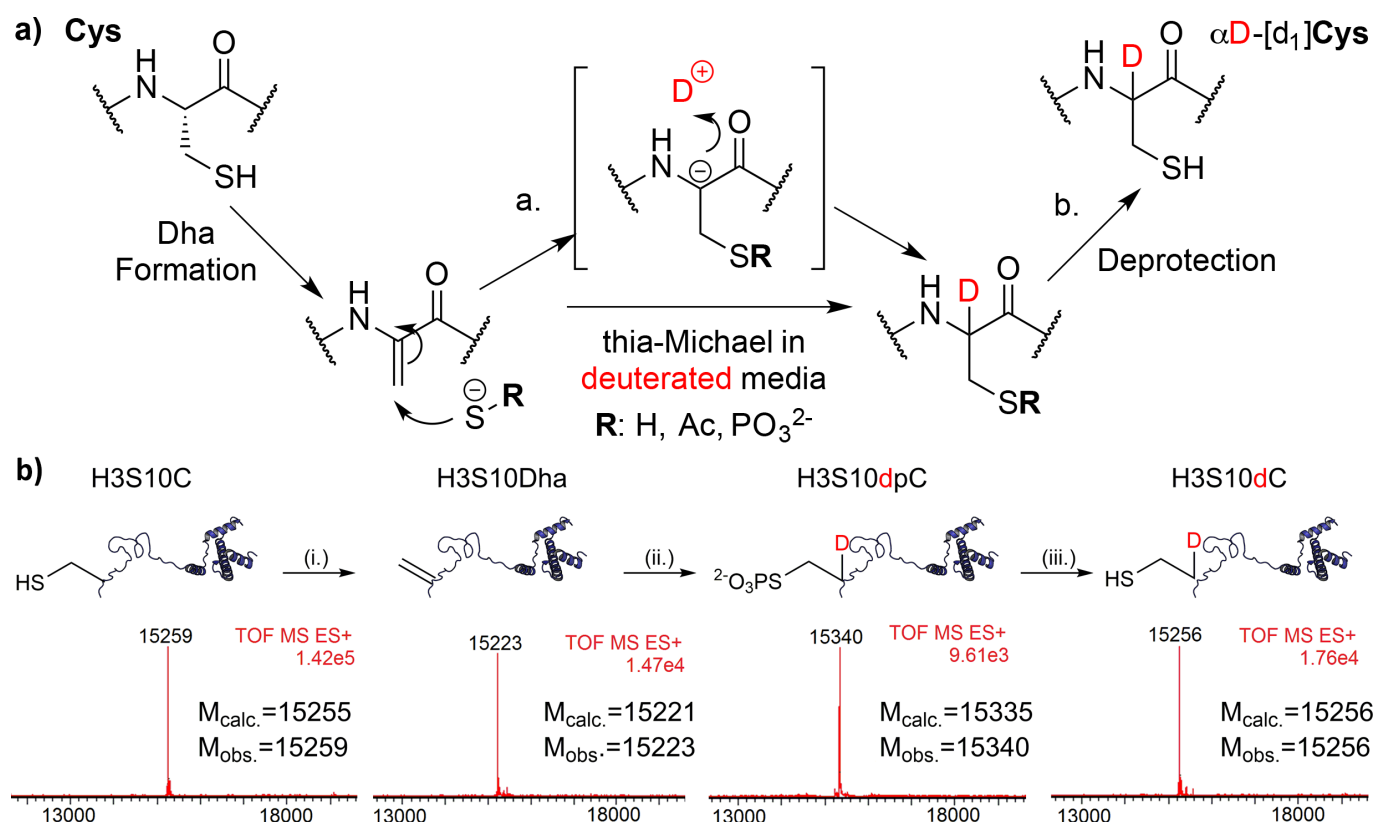


Figure 2. A Method for Site-Selective Deuteration of Intact Proteins (a) Formation of α -deuterated cysteine with two key intervening steps: a. thia-Michael reaction with thiol in deuterated solvent followed by b. deprotection / removal of the general 'protecting' group **R**. (b) Formation of deuterio-H3 protein H3-dC10 ($R=\text{PO}_3^{2-}$ shown), monitored with ESI-MS. This experiment was repeated independently 3 times. (i.) DBHDA in deuterated buffer, pH 8, 37 °C, 3 h; (ii.) 0.4 M Na_3SPO_3 in deuterated buffer, pH 8.6, 37 °C, 1 h; (iii.) pH 4, 37 °C, 2 h or PP_1 phosphatase, 30 °C, 1 h.

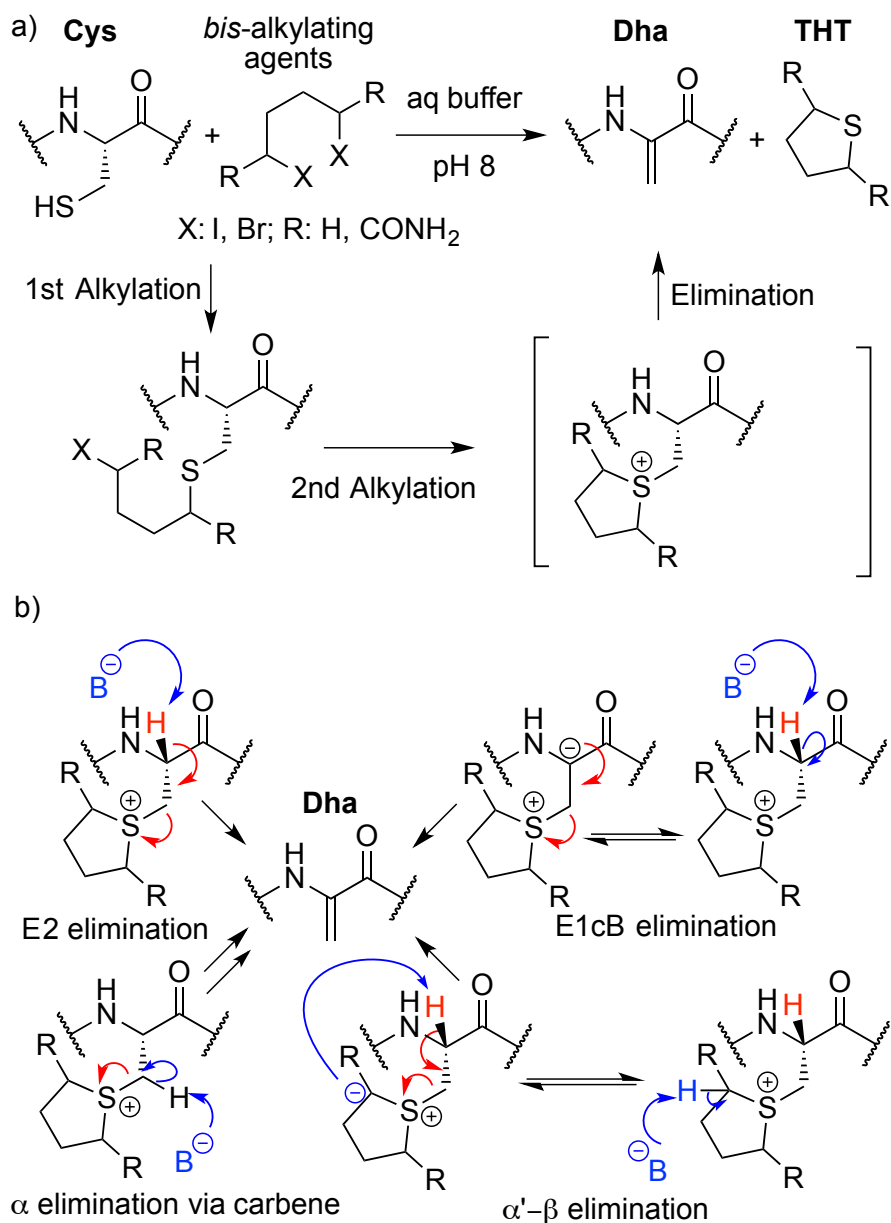


Figure 3. The Cys→Dha Protein Chemistry Reaction to be Mechanistically Probed by Deuteration. (a) Formation of **Dha** from **Cys** via double alkylation & elimination of a sulfonium intermediate and creates tetrahydrothiophene (**THT**) by-products. (b) Four broad mechanisms of elimination were considered in this study: anti-E2-elimination; two-step E1cB elimination via carbanion; α' - β elimination via sulfonium ylid; and α -elimination carbene formation (leading to Dha). B represents a generalized base (intra- or inter-molecular).

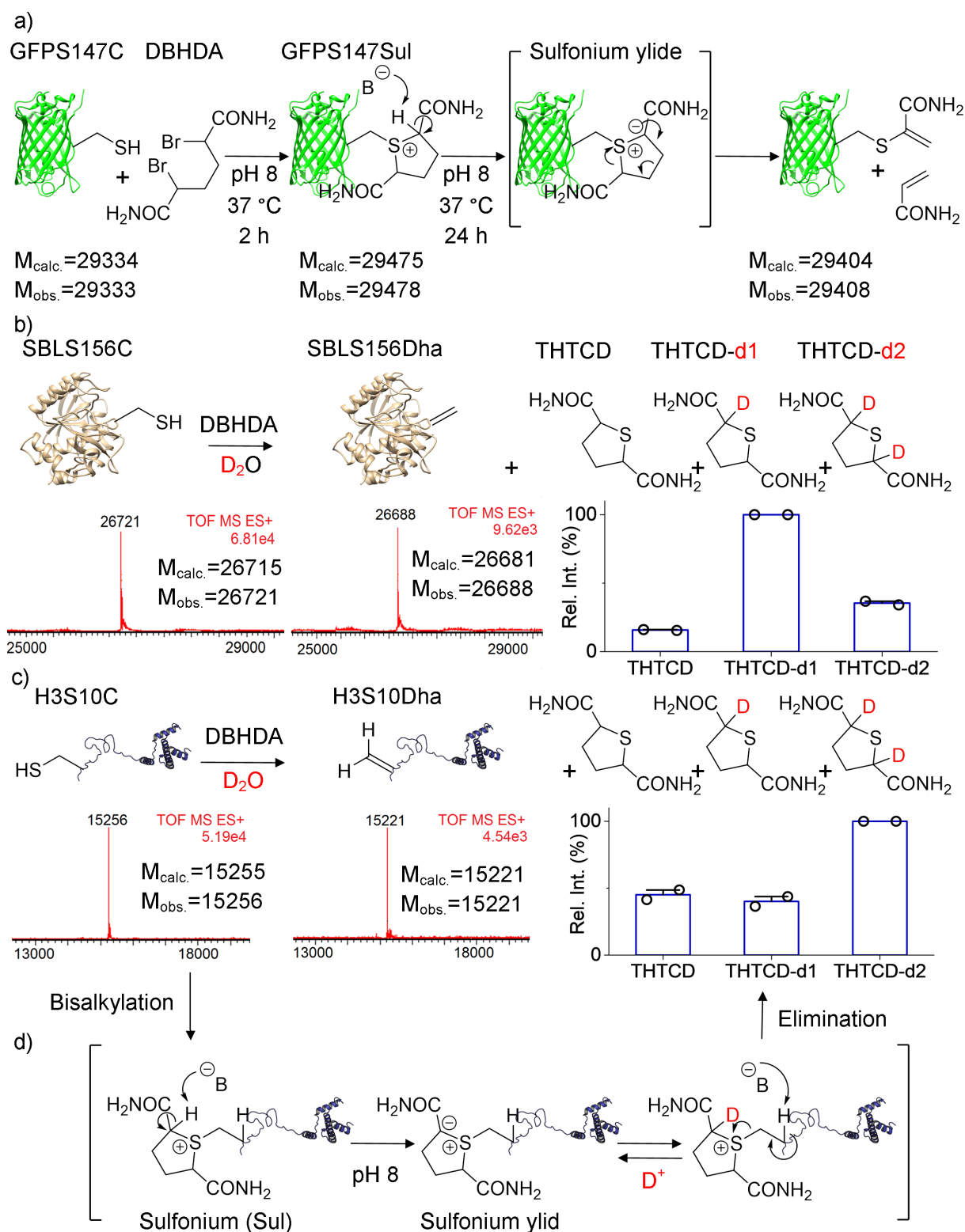


Figure 4. Exploring the Chemistry of Stable and Unstable Protein Sulfoniums with ‘Wash-in’. (a) Observable GFP-S147Sul sulfonium undergoes [3+2] cycloreversion to vinyl sulfide. (b-d) Mono- & bis-deuteration of by-product THTCD observed during formation of SBLS156Dha (b) & H3S10Dha (c) in deuterated buffer using DBHDA confirmed a sulfonium ylid intermediate. Deuteration was determined by MS & site confirmed by MSMS. Bar graphs represent the mean values of the relative intensities (normalized with respect to the most intense peak) of the mass

of unlabelled and labelled THTCD, in which the ^{13}C and ^{34}S isotopologue influence has been subtracted based on the theoretical unlabelled THTCD isotopic pattern (100% for 0; 6.5% for +1, and 4.5% for +2). Mass spectra were generated from $n=2$ independent experiments. The individual data points are overlaid as circles on the bars. Standard deviations were calculated to generate the error bars. These experiments were repeated independently twice. (d) Proposed mechanism of D incorporation *via* sulfonium ylid.

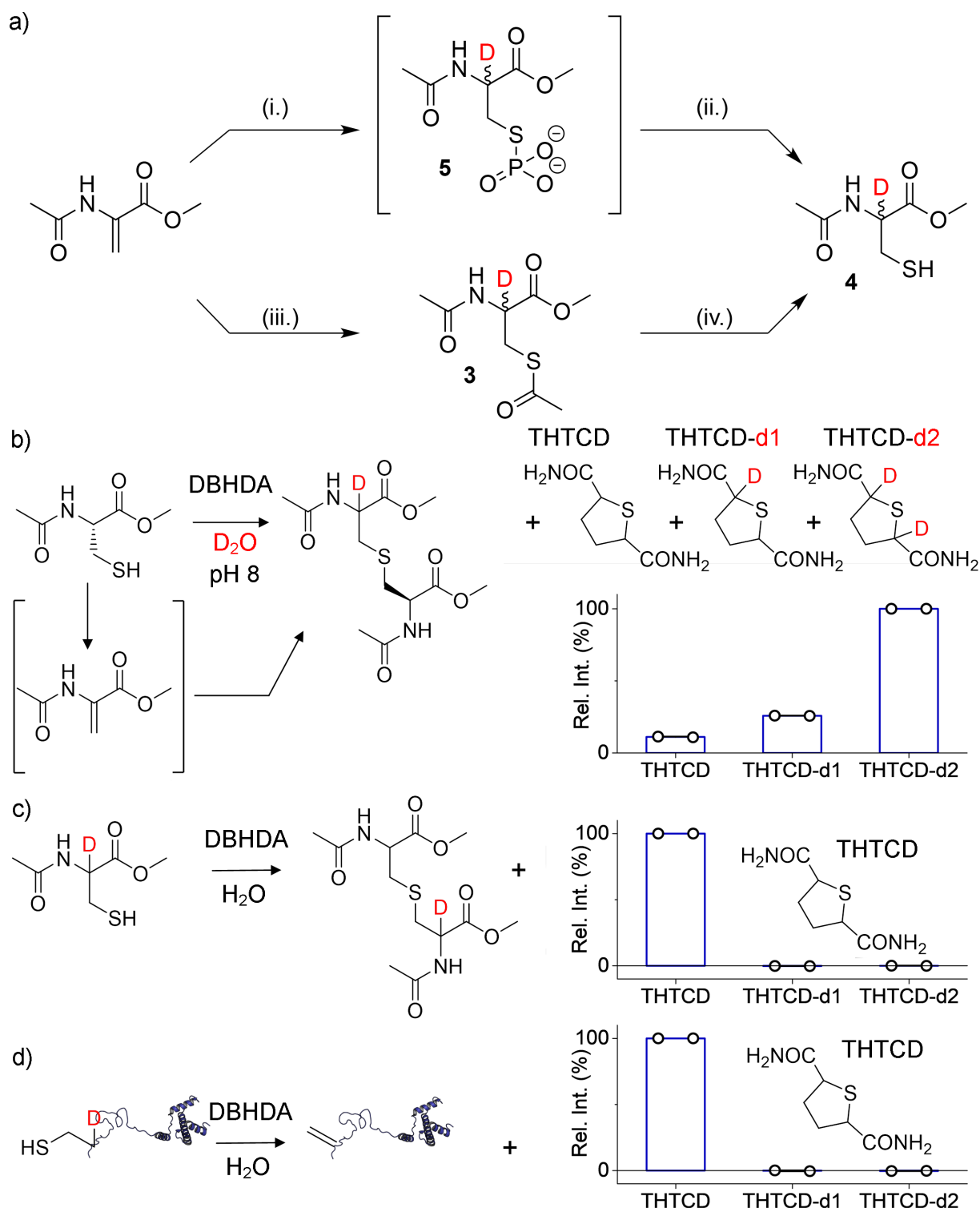


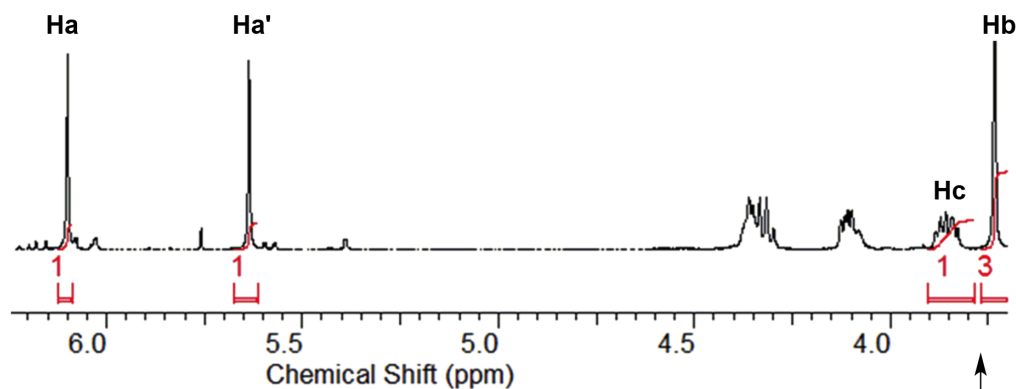
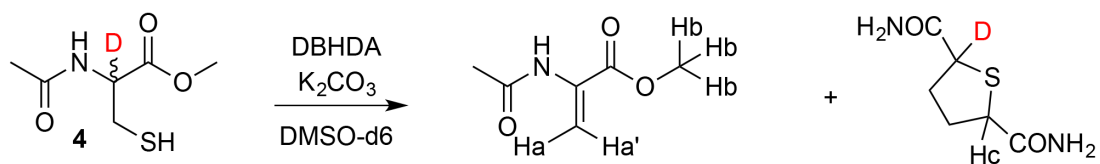
Figure 5. Application of The Post Translational α -C-deuterium Method to Explore the Chemistry of Dha Formation. (a) Alternative labelling routes to dC-containing model. (i.) 0.4 M Na_3SPO_3 in D_2O , pH 8.7, rt, 2 h; (ii.) pH 2-3, 37 °C, 4 h, 73% yield over 2 steps; (iii.) KSAc in D_2O pH 7, 20 h; (iv.) K_2CO_3 , MeOH, rt, 15 min. 38% yield over 2 steps. (b-d) Consistent with reactions on proteins, both wash-in implicating an ylid (b) and no detectable intramolecular D transfer (c) were observed upon treating AcNHCysOMe-d1 with DBHDA. (d) Analysis of

eliminated THTCD from Dha formation using DBHDA on labelled H3-dC10. These experiments were repeated independently twice. See method above. Bar graphs represent the mean values of the relative intensities (normalized with respect to the most intense peak) of the mass of unlabelled and labelled THTCD, in which the ^{13}C and ^{34}S isotopologue influence has been subtracted based on the theoretical unlabelled THTCD isotopic pattern (100% for 0; 6.5% for +1, and 4.5% for +2). Mass spectra were generated from n=2 independent experiments. The individual data points are overlaid as circles on the bars. Standard deviations were calculated to generate the error bars.

a) AcNHdCysOMe

AcNHDhaOMe

THTCD-d1



b)

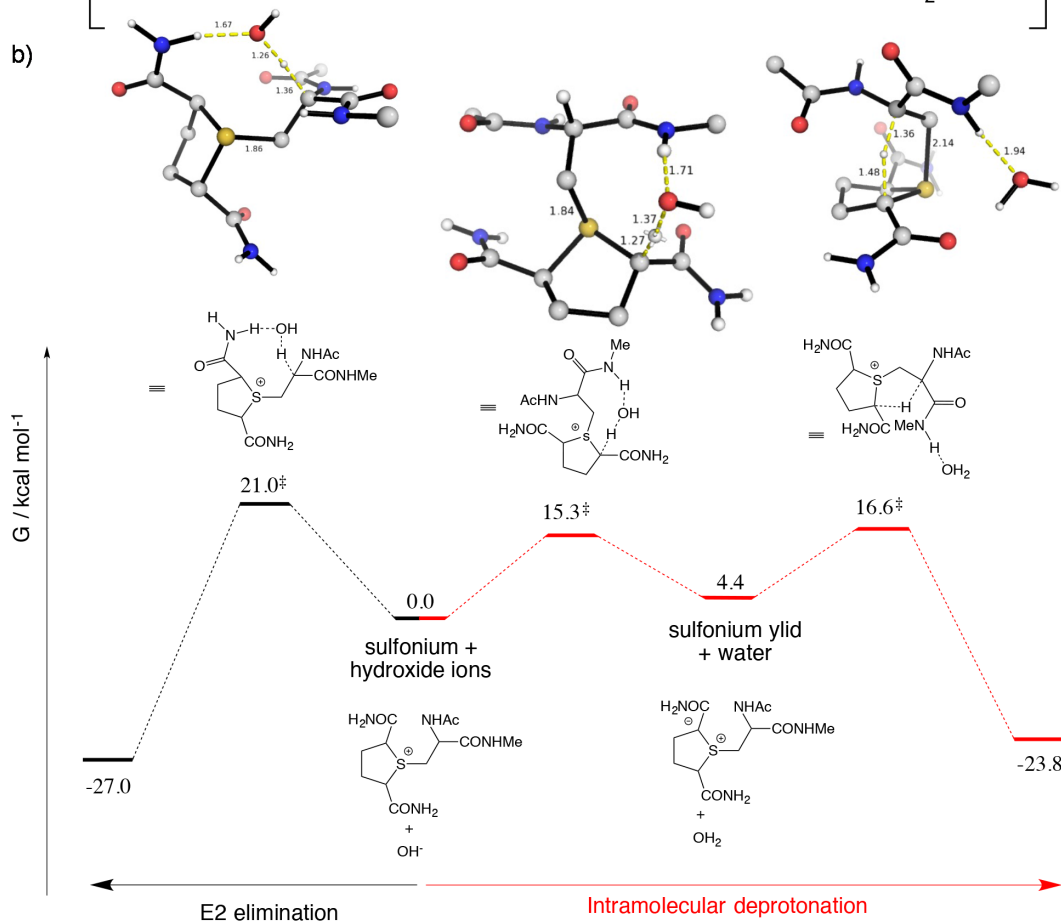
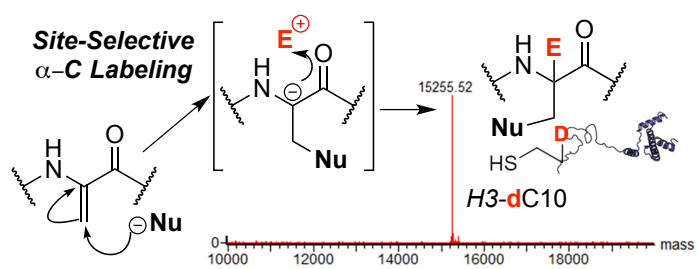


Figure 6. Mechanistic Analysis Reveals the Role of the Sulfur Ylid as an Intramolecular Base in a ‘Swern Reaction’. (a) Intramolecular D transfer during formation of Dha from Cys is observed for AcNHCysOMe-d1 in aprotic solvent by both MS and NMR. This experiment was repeated independently four times. (b) CPCM-M06-2X/6-311++G(d,p) calculations were used to characterize both E2 and intramolecular pathways; the latter passes through lower activation barriers consistent with the observed D exchange and transfer in both labelled proteins and model systems. TS structures are shown with key forming/breaking distances in Å. G_{rel} (at 298.15K) quoted in kcal/mol. OH^- was used as a base throughout ($\text{B}^- = \text{OH}^-$).



Graphical Abstract