nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-018-0128-y

Post-translational site-selective protein backbone α-deuteration

Sébastien R. G. Galan[1], James R. Wickens[1], Jitka Dadova[1], Wai-Lung Ng[1], Xinglong Zhang[1], Robert A. Simion[1], Robert Quinlan[1], Elisabete Pires[1], Robert S. Paton[1,2], Stephen Caddick[3], Vijay Chudasama[3]* and Benjamin G. Davis[1]*

Isotopic replacement has long-proven applications in small molecules. However, applications in proteins are largely limited to biosynthetic strategies or exchangeable (for example, N–H/D) labile sites only. The development of postbiosynthetic, C–^{1}H → C–^{2}H/D replacement in proteins could enable probing of mechanisms, among other uses. Here we describe a chemical method for selective protein α-carbon deuteration (proceeding from Cys to dehydroalanine (Dha) to deutero-Cys) allowing overall ^{1}H→ ^{2}H/D exchange at a nonexchangeable backbone site. It is used here to probe mechanisms of reactions used in protein bioconjugation. This analysis suggests, together with quantum mechanical calculations, stepwise deprotonations via on-protein carbanions and unexpected sulfonium ylides 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 labeling strategy as a possible tool in protein mechanism, structure, biotechnology and medicine.

Despite their clear utility in small molecules[1,2], methods for site-selective replacement of nonexchangeable hydrogen isotopes (for example, ^{1}H → ^{2}H = D) intact proteins are essentially unknown. Hydrogen isotopes can be used as possible probes of structure and dynamics (for example, by NMR[3] 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 mechanisms in proteins.

Current possible strategies for isotope introduction typically require isotope incorporation before assembly. One approach to building labeled proteins is the segmental isotopic-labeling strategy[6] 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 leads to global, and hence non-site-selective, labeling of an amino acid type and can require creation and use of multicomponent, engineered expression systems.

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

We postulated that a general site-selective backbone α-C labeling might be possible via the creation and electrophilic trapping of a suitable α-C-enolate *C*-nucleophile (Fig. 1a). This enolate, 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 also 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 (Fig. 1b).

Dehydroalanine (Dha) is a potentially useful amino acid residue that can be found in proteins or peptides naturally[14], following metabolism[15] or via amber-codon suppression methods[16], and can be 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→D replacement could be envisaged at Cys (and Cys derivatives) (Fig. 1c). Indeed, the possibility of such a mechanism has also been noted by others[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 known[21] 'relayed' D-exchange mechanisms for enhanced accessibility. Conversion 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 deuteron 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 in which isotope replacement suggests previously unobserved mechanistic intermediates and details.

Results

Development of site-selective protein deuteration method. Although 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 and thus would not readily allow traceless α-C H→D replacement. We therefore tested a range of putative sulfur nucleophiles that might also readily generate Cys (Supplementary Fig. 1). Though use of

[1]Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Oxford, UK. [2]Department of Chemistry, Colorado State University, Fort Collins, CO, USA. [3]Department of Chemistry, University College London, London, UK. *e-mail: v.chudasama@ucl.ac.uk; Ben.Davis@chem.ox.ac.uk

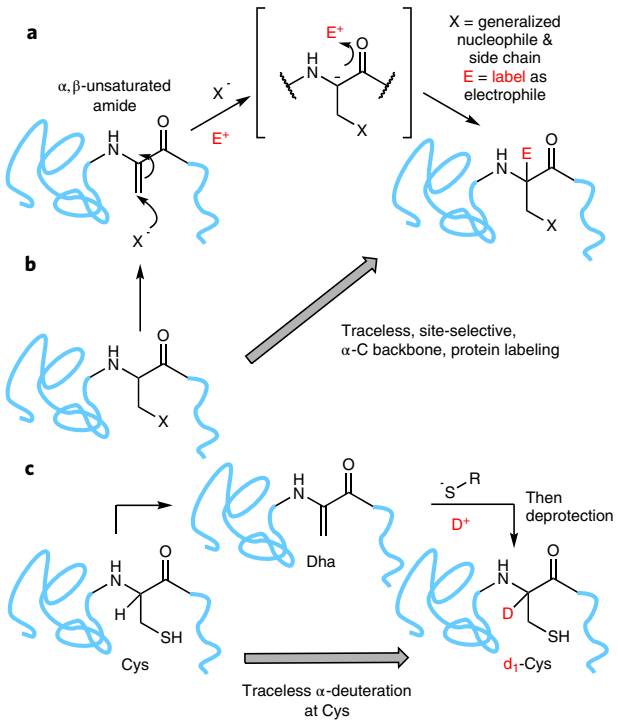

Fig. 1 | Traceless post-translational site-selective protein α-deuteration strategy. a, Strategy for a generalized, site-selective α-C labeling approach with any electrophile E in an intact protein (light blue ribbon cartoon) backbone. **b**, By combining this strategy with a method for both deriving the key α,β-unsaturated amide intermediate from side chain X and reforming the same side chain X, the process can be rendered traceless. **c**, Proof-of-principle H→D replacement process at the α-carbon of Cys demonstrated in this work, which exemplifies the general strategy.

the thioacetate (Fig. 1c, R = Ac) proved successful and proceeded with excellent conversion (>95%, Supplementary Fig. 1b), subsequent deprotection proved unsuccessful. The use of the unprotected *S*-nucleophile source Na_2S also proceeded well in some systems (Supplementary Fig. 1c); however, in some systems concomitant dimerization via thioether formation led to formation of unwanted byproducts. Finally, use of sodium thiophosphate (Fig. 1c, R = PO_3^{2-}) 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 Methods).

Specifically, model protein histone H3 was readily H→D exchanged site-selectively first at site 10 (Fig. 2). H3-Cys10 was converted to H3-Dha10 using the bis-alkylation-elimination reagent DBHDA[17] (**1**;>95%, 3 h), and H3-Dha10 was then converted to H3-[d1]Cys10–S–PO_3^{2-} (>95%, 1 h). Deprotection to H3-[d1]Cys10 was achieved readily and flexibly in one 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; Fig. 2b and Supplementary Fig. 2). Exhaustive dialysis in nondeuterated 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 deutero-protein in typically 65–85% yield. Characterization by high-resolution MS and tryptic-digest-MS confirmed the level and site of incorporation (Supplementary Figs. 1–8; Supplementary Tables 1 and 2). The method we use here has the potential to generate epimers at the α-position (Fig. 1) with a stereoselectivity that is potentially influenced by the protein substrate[30]. Through use of chromatographic separation and MS detection (see Supplementary Figs. 9-17), we determined a diastereomeric ratio (d.r.) of approximately 1:2 L:D, suggesting only a slight influence by the substrate.

Next, to test breadth and application to other sites, an essentially similar process was applied to site 26 in histone H3 (Supplementary Fig. 18). This also proceeded successfully (in typical overall yields of 70–75%) using the same method, demonstrating breadth of the application. These reactions demonstrated, to the best of our knowledge, the first examples of a constitutionally traceless, convergent, site-selective chemical protein modification that can be used for selective incorporation and/or replacement (H→D) of a deuterium atom at the α-position of an amino acid residue (here Cys) in 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 products[14] and in the biosynthetic precursors[31] 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 (Fig. 3) from cysteine (Cys), which is especially 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 it 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 postulated[17] mechanism (Fig. 3a) involves a double (bis) alkylation (inter- followed by intramolecular) of the γ-S of the Cys residue, leading to the formation of an unstable cyclic sulfonium intermediate that then gives Dha (Fig. 3). In nearly all cases, the sulfonium ion appears to undergo direct elimination readily to yield Dha with good conversion (Fig. 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 green fluorescent protein (GFP) GFP-Cys147 (GFP-S147C) with DBHDA (Fig. 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 six-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 (Fig. 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

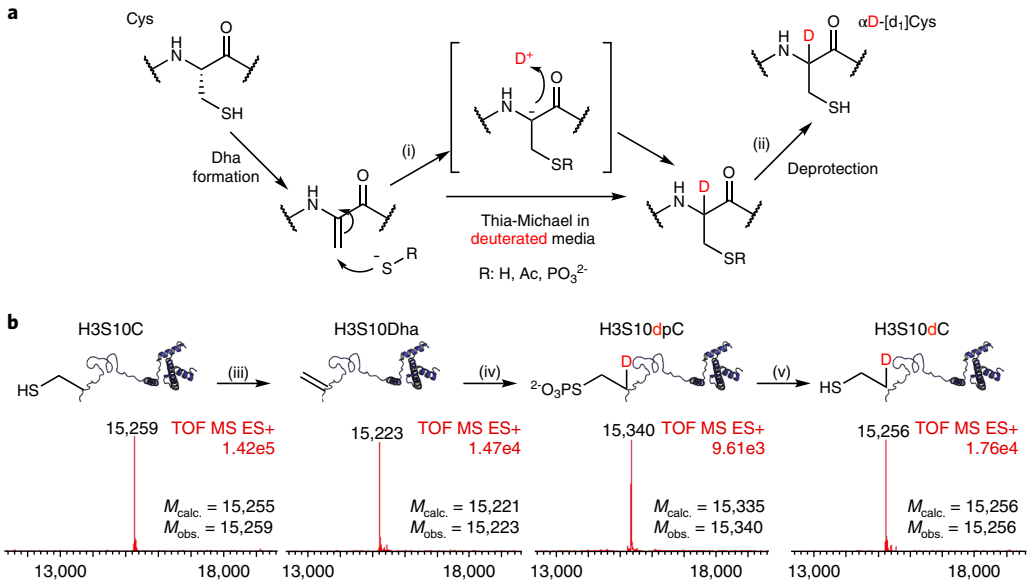

Fig. 2 | A method for site-selective deuteration of intact proteins. a, Formation of α-deuterated cysteine with two key intervening steps: (i) thia-Michael reaction with thiol in deuterated solvent followed by (ii) deprotection and removal of the general 'protecting' group R. **b**, Formation of deutero-H3 protein H3-dC10 (R = PO_3^{2-} shown), monitored with ESI–MS. This experiment was repeated independently three times. (iii) DBHDA in deuterated buffer, pH 8, 37 °C, 3 h; (iv) 0.4 M Na_3SPO_3 in deuterated buffer, pH 8.6, 37 °C, 1 h; (v) pH 4, 37 °C, 2 h or PP1 phosphatase, 30 °C, 1 h. M_{calc}, calculated total protein molecular mass; M_{obs}, observed total protein molecular mass.

on sulfonium type, for example. Moreover, no reactions of this type (or similar) have been mechanistically evaluated previously on-protein, highlighting a potentially powerful influence of substrate (that is, the protein) on the 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 conditions for potential exchange of C–H (that is basic). GFP-S147C is the only system in which a stable sulfonium has been observed (Fig. 4a), which is 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 (Supplementary Fig. 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 (Fig. 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 transition state barrier for this reaction. With conformational sampling, a total of ten transition structures were found. These ranged in energy from 20.3 to 25.1 kcal mol^{-1} (or 15.9 to 20.7 kcal mol^{-1} with respect to the lowest energy sulfonium ylide 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 (Supplementary Fig. 20).

The observation of such a retro-cycloaddition/cycloreversion implied the formation of a sulfonium ylide from sulfonium GFP-S147Sul (Fig. 4a). We considered that this intermediate might be sufficiently long lived to allow 'wash in' of a label from solvent, which could be observed in fragmented acrylamide (Supplementary Figs. 21 and 22). Incubation of GFP-S147Sul in deuterated buffer released β-deutero-acylamide, 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, which 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 (Fig. 4b–d). In both cases, reaction via treatment with DBHDA in deuterated buffer gave the byproduct tetrahydrothiophene carbodiamide (THTCD; **2**), which showed deuterium incorporation (Fig. 4b,c). MS/MS analysis (Supplementary Figs. 23 and 24) unequivocally located the deuteration at the Cα″ carbon of THTCD, not on exchangeable protons. Thus, product ion spectra were collected with normalized collision energy for each *m/z* corresponding to d_n-THTCD; the precursor isolation window for MS/MS 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 (Supplementary Fig. 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 ylide and a site of deprotonation consistent with intermediacy during elimination. Moreover, the detection of doubly deuterated THTCD-d2 suggested that the equilibrium between ylide 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.

Use of deuteration to probe Dha formation. Having used such 'wash-in' experiments, we next chose to probe the mechanism more deeply with the protein variants bearing pre-installed deuterium labels at predetermined nonexchangeable sites by exploiting the α-C deuteration chemistry developed above for this purpose.

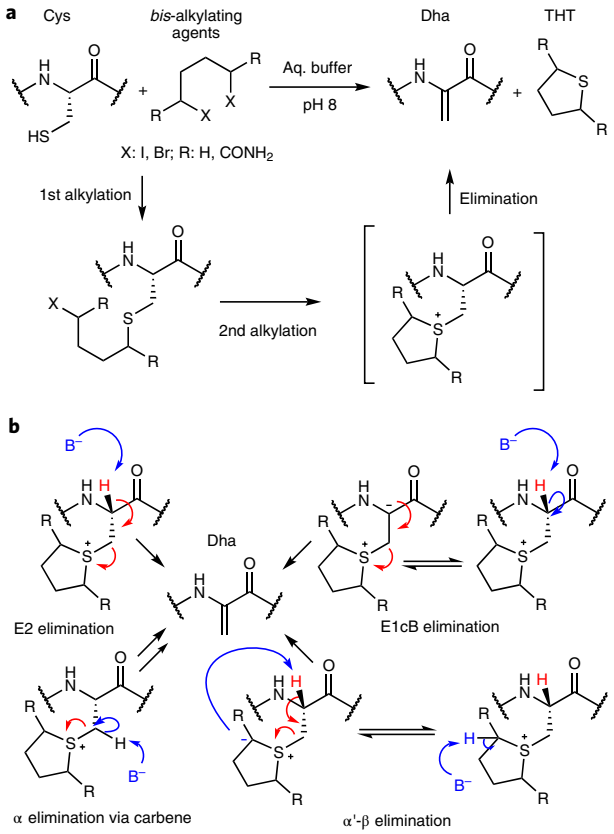

Fig. 3 | The Cys→Dha protein chemistry reaction to be mechanistically probed by deuteration. a, Formation of Dha from Cys via double alkylation and elimination of a sulfonium intermediate creates tetrahydrothiophene (THT) byproducts. **b**, Four broad mechanisms of elimination were considered in this study: anti-E2 elimination; two-step E1cB elimination via carbanion; α′-β elimination via sulfonium ylide; and α-elimination carbene formation (leading to Dha). B represents a generalized base (intra- or intermolecular).

With stable Cα-Cys-deuterium-labeled proteins in hand (Fig. 2), we tested whether internal deuteron transfer (for example, from Cα to ylide) might be observed during formation of Dha from Cys (Fig. 4). Deuteration of THTCD was not observed for either H3-dC10 (Fig. 4d) or H3-dC26 (Supplementary Fig. 26). Together, these data suggested that if the sulfonium ylide plays the role of an intramolecular base (Fig. 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 the effects of solvent exchange through the use of aprotic solvents of similar polarity. Unfortunately, none of the protein substrates were sufficiently soluble to allow unambiguous experiments. However, application of the Cα-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; **4**) (Fig. 5a) was best achieved in two steps by reaction of the corresponding Dha-residue (AcNHDhaOMe) with Na_3SPO_3 in D_2O 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 potassium thioacetate (KSAc) in D_2O followed by deacetylation using K_2CO_3 in MeOH was also successful. Though the overall recovered yield for this alternative two-step conversion was not outstanding (38%), it nonetheless allowed high levels of label incorporation (>95%). Upon treatment with DBHDA, the thus-labeled Dha-model AcNHCysOMe-d1 behaved essentially identically to protein systems (Fig. 5–bd and Supplementary Figs. 27 and 28) in protic (aqueous buffer or D_2O) solvents, thereby suggesting its validity as a model in aprotic systems.

To test the mechanistic feasibility of intramolecular proton transfer in the absence of solvent exchange, we treated AcNHCysOMe-d1 with DBHDA in the presence of base (K_2CO_3) in nonexchangeable solvent DMSO-d6. Both LC–MS and NMR revealed the formation of THTCD-d1 (Fig. 6a). Its formation can be explained by the formation of a sulfonium ylide that acts as an intramolecular base to abstract the α-deuterium of the residue in AcNHCysOMe-d1. To further probe the intra- versus intermolecular nature of this abstraction, we conducted a series of dilution experiments (Supplementary Fig. 29) confirming that at higher concentrations (beyond those that may be seen in proteins) deuterium incorporation in THTCD-d1 is lost, consistent with an intramolecular mechanism via pre-equilibrating ylide formation at lower concentration. It should be noted that such intermolecular (and hence protein-to-protein) abstraction is much less probable in proteins given the (likely) more crowded nature as compared to the 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 ylide through deprotonation of sulfonium at Cα′. The lifetime of this ylide appears sufficient to allow 'wash-in' from exchangeable solvent. This ylide is also capable of relayed deprotonation of the Cα′ position of the Cys residue, resulting in formation of site-specifically deuterated THT in nonexchangeable solvent; concentration dependency implicates intramolecular deprotonation, and this deuteration appears to be 'washed-out' in exchangeable solvent (Fig. 6a). This implicated E1cB elimination with 'wash-out/in' at rates greater than those for elimination.

Density functional theory (DFT) calculations (Fig. 6b) were used to characterize pathways for a concerted E2 elimination and intramolecular deprotonation via the ylide as internal base. Conformations were sampled systematically, and possible rotamers were generated incrementally by varying a combination of key dihedral angles. Two key dihedral angles along C_α-C_β and C_β-S of the 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 ylide, 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 ylide, with ten structures having energies within 5 kcal mol^{-1} of the lowest energy structure (at 16.6 kcal mol^{-1}). A total of ten TSs were found for the competing E2 pathway, with five 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 ylide formation was found to be relatively facile and marginally endergonic (that is, 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 Fig. 30). By contrast, the E2-elimination barrier is about 4.4 kcal mol^{-1} less favorable than that computed for the intramolecular process, making this pathway substantially kinetically disfavored (>1,600 times selectivity for intramolecular deprotonation over E2 deprotonation using simple transition state theory). Together, this supported the

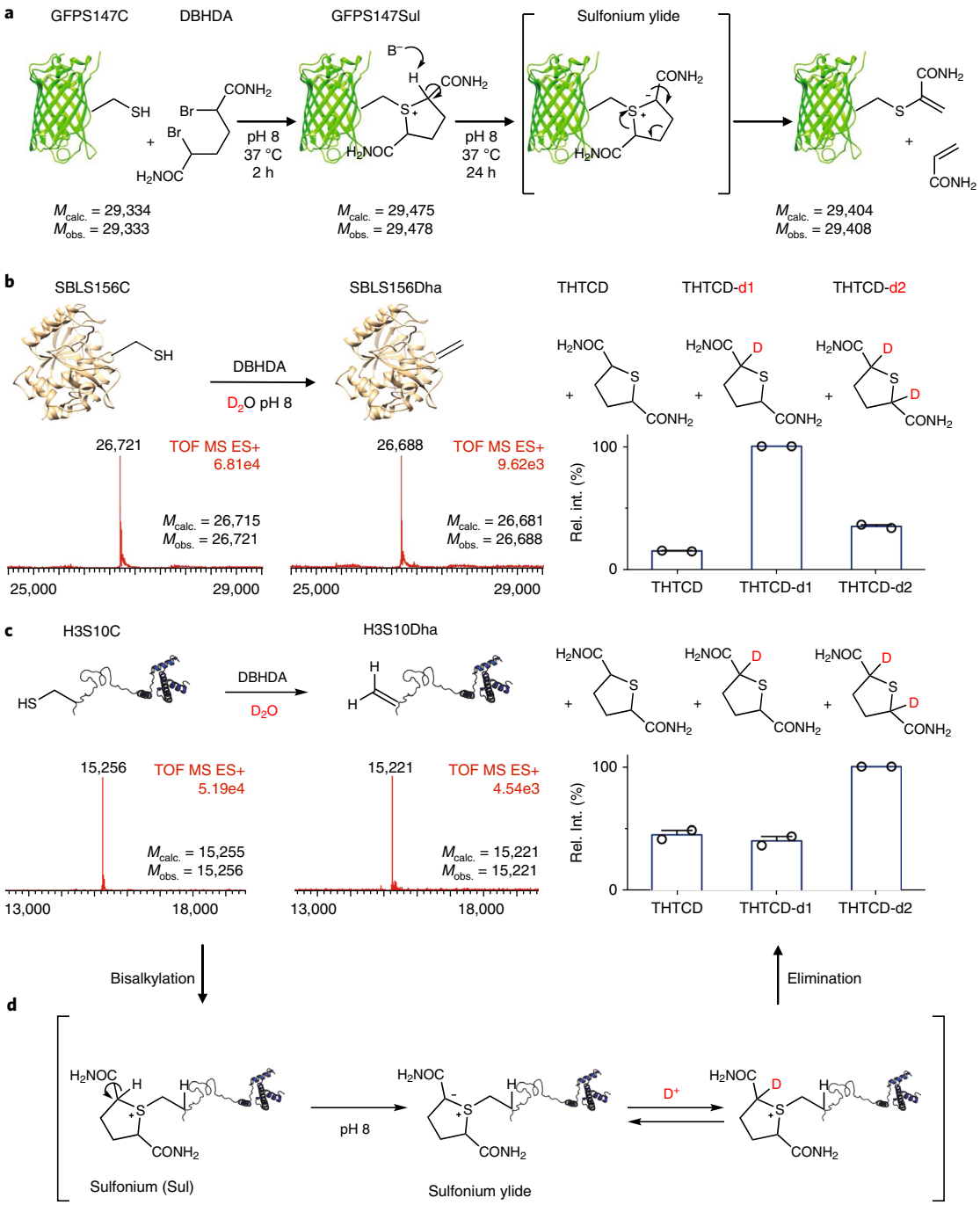

Fig. 4 | Exploring the chemistry of stable and unstable protein sulfoniums with wash-in. a, Observable GFP-S147Sul sulfonium undergoes [3+2] cycloreversion to vinyl sulfide. **b**,**c**, Mono- and bis-deuteration of byproduct THTCD observed during formation of SBLS156Dha (**b**) and H3S10Dha (**c**) in deuterated buffer using DBHDA confirmed a sulfonium ylide intermediate. Deuteration was determined by MS and site confirmed by MS/MS. Bar graphs represent the mean values of the relative intensities (normalized with respect to the most intense peak) of the masses of unlabeled and labeled THTCD in which the ^{13}C and ^{34}S isotopologue influence has been subtracted based on the theoretical unlabeled 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. s.d. values were calculated to generate the error bars. These experiments were repeated independently twice. **d**, Proposed mechanism of D incorporation via sulfonium ylide.

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 could now prove to be useful mechanistic tools to elucidate detailed protein chemistry. Here, strikingly, they suggest that an unusual sulfonium ylide intermediate is involved in the formation of Dha from Cys (formed using bis-alkylating reagent DBHDA) and that this ylide acts as an internal base to abstract a

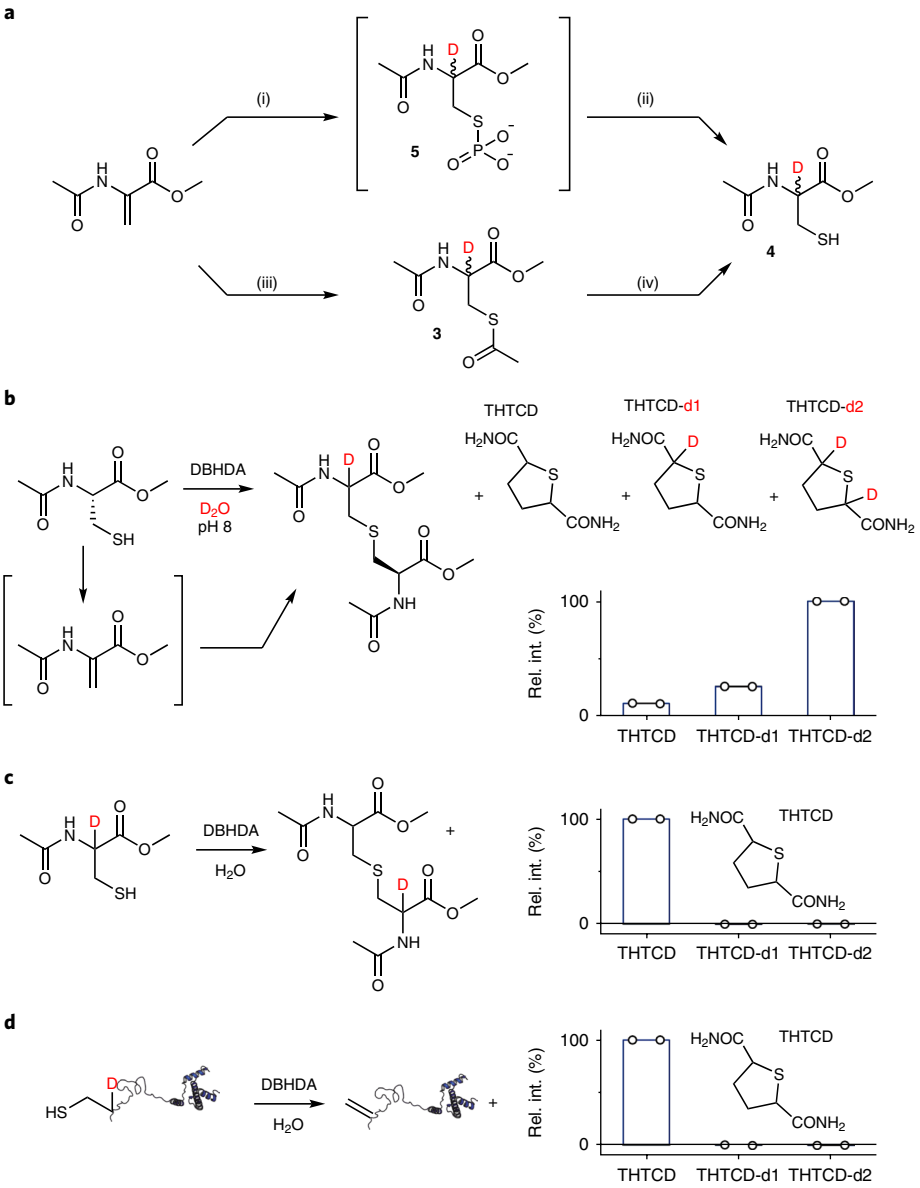

Fig. 5 | Application of the post translational α-C-deuterium method to explore the chemistry of Dha formation. a, Alternative labeling routes to the dC-containing model. (i) 0.4 M Na_3SPO_3 in D_2O, pH 8.7, rt, 2 h; (ii) pH 2–3, 37 °C, 4 h, 73% yield over 2 steps; (iii) KSAc in D_2O, pH 7, 20 h; (iv) K_2CO_3, MeOH, rt, 15 min. 38% yield over 2 steps. **b**,**c**, Consistent with reactions on proteins, both wash-in implicating an ylide (**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 labeled 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 unlabeled and labeled THTCD, in which the ^{13}C and ^{34}S isotopologue influence has been subtracted based on the theoretical unlabeled 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. Error bars, s.d.

proton from the protein backbone, which in turn leads to elimination. This is consistent with the lower pK_as of α-protons of sulfoniums as compared to those of backbone αCH-protons of amino acid residues[41]. Observation of both 'wash-out' and 'wash-in' suggests that the resulting Cα carbanion is sufficiently long-lived, which in turn suggests a step-wise E1cB (perhaps $E1cB_{anion}$) mechanism (Supplementary Fig. 31). Ylide intermediacy is also supported by the fragmentation of a longer-lived sulfonium in GFP and quantum mechanical calculations. Taken together, this mechanism 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 not previously been implicated in a biological system.

The development of site-selective protein labeling (here D) and MS^n methodology for tracking that label has proven here to be a powerful analytical combination for unpicking complex, previously unknown, protein chemistry. These techniques have not only allowed the study of a chemical protein modification reaction, but also reveal a more general approach to studying complex mechanistic analyses in protein chemistry. 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 modification handle[43], the targeting of Cys (Fig. 1c) is a useful first proof of principle of a

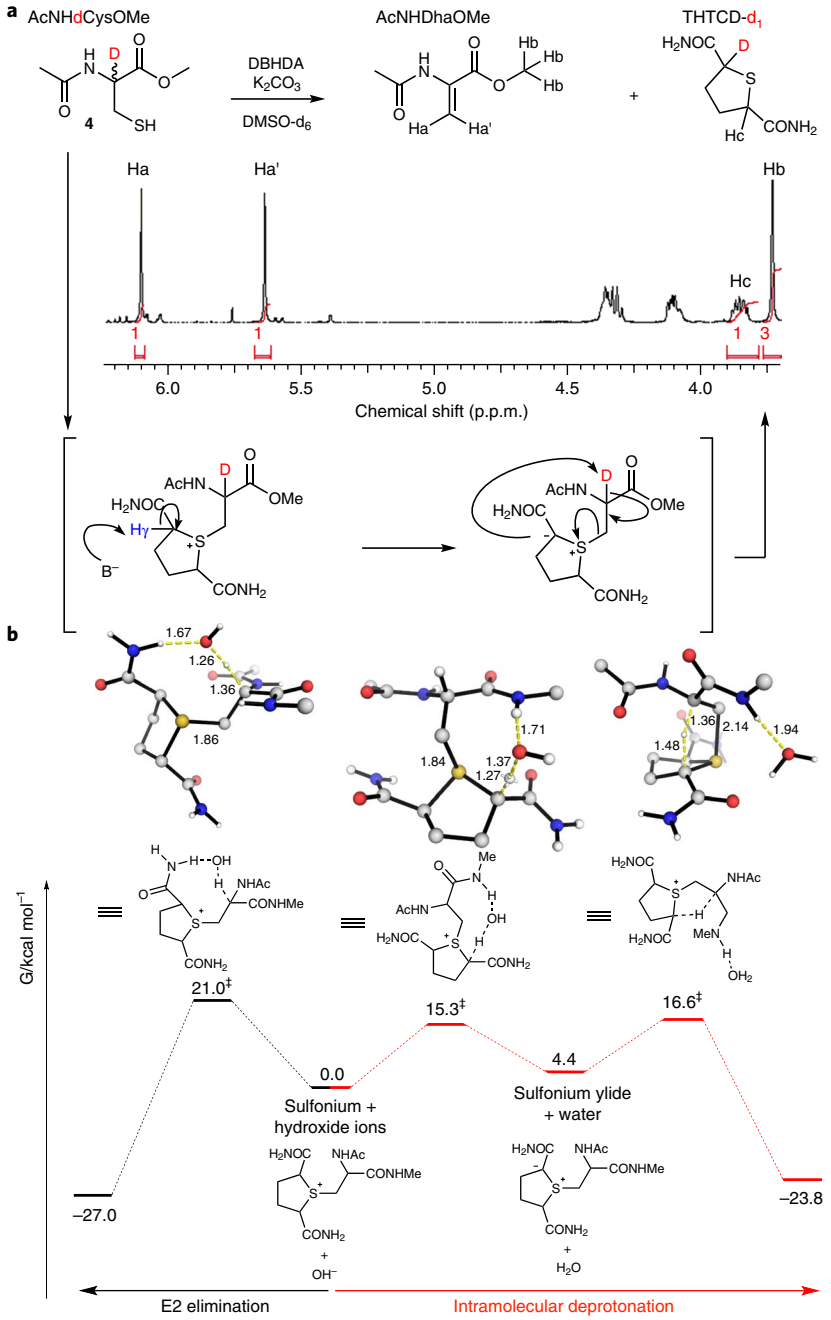

Fig. 6 | Mechanistic analysis reveals the role of the sulfur ylide as an intramolecular base in a 'Swern reaction'. a, Intramolecular D transfer during formation of Dha from Cys is observed for AcNHCysOMe-d1 in aprotic solvent by both MS and NMR. This experiment was repeated independently four times. **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 labeled proteins and model systems. Transition state structures are shown with key forming/breaking distances in Å. G_{rel} (at 298.15 K) quoted in kcal/mol. OH^- was used as a base throughout ($B^- = OH^-$).

potentially more general method (Fig. 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 further.

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

It is important to note that the method we used here generates epimers at the α-position that may limit some applications[30]. Therefore, to be clear, the formation of some D-epimers renders this process traceless only with regard to the constitution of the bonds formed and are not completely traceless in configuration. However, such deutero-epimers, as we show here, can be used to probe

mechanisms. 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. α-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 α-C-deuterated proteins as candidate 'biologics' can now be considered and potentially readily explored.

Online content

Methods (including Nature Research reporting summary), references, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41589-018-0128-y.

Received: 29 December 2017; Accepted: 19 July 2018;
Published online: 17 September 2018

References

1. Westheimer, F. H. The magnitude of the primary kinetic isotope effect for compounds of hydrogen and deuterium. *Chem. Rev.* **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 two-dimensional NMR of proteins. *Annu. Rev. Biophys. Biophys. Chem.* **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. Int. Ed. Engl.* **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., 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 1H 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 ^{1}H and ^{2}H NMR signals from the alpha and beta hydrogens of cysteinyl 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. I. Reinvestigation of the 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., Remmel, R. P. & Abdel-Monem, M. M. Biliary excretion of a glutathione conjugate of busulfan and 1,4-diiodobutane in the rat. *Drug Metab. Dispos.* **16**, 85–92 (1988).
16. Wang, J., Schiller, S. M. & Schultz, P. G. A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed. Engl.* **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. Engl.* **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. *J. Am. Chem. Soc.* **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. Engl.* **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 p38α 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**, e103935 (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. *Org. Biomol. Chem.* **14**, 4817–4823 (2016).
31. Palioura, S., Sherrer, R. L., Steitz, T. A., Söll, D. & Simonovic, M. The human SepSecS-tRNASec 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. Jr & Lawton, R. G. Cysteine modification and cleavage of proteins with 2-methyl-N1-benzenesulfony-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. XXII. 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 (β-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**, 1–15 (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. *Isr. J. Chem.* **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. *J. Am. Chem. Soc.* **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. *Chem. Rev.* **117**, 1927–1969 (2017).
47. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 496–514 (1990).
48. Darland, G. K. et al. Oxidative and defluorinative metabolism of fludalanine, 2-2H-3-fluoro-D-alanine. *Drug Metab. Dispos.* **14**, 668–673 (1986).
49. Timmins, G. S. Deuterated drugs: where are we now? *Expert Opin. Ther. Pat.* **24**, 1067–1075 (2014).

Acknowledgements

We thank BBSRC/AstraZeneca (S.G.), EU Horizon 2020 program under the Marie Sklodowska-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 G. Karunanithy, S. Nadal, R. Raj, R. Nathani, J. Willwacher, G. Pairaudeau, J. Read, A. Breeze and A. Baldwin for useful discussions.

Author Contributions

S.R.G.G., J.D., W.-L.N., V.C. and R.Q. conducted chemical experiments; X.Z., R.A.S. and R.S.P. conducted computational experiments; S.R.G.G., J.R.W. and E.P. conducted mass-spectrometric experiments; S.R.G.G., J.R.W., R.S.P., S.C., V.C. and B.G.D. designed the experiments and analyzed the data; S.R.G.G., V.C. and B.G.D. wrote the paper; all authors read and commented on the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-018-0128-y.

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

Correspondence and requests for materials should be addressed to V.C. or B.G.D.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Analytical tools for protein reaction monitoring and characterization. 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 × 4.6 mm × 5 µm). 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 liters 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 the manufacturer's instructions.

For GFP, LC–MS was performed on protein samples using a Thermo Scientific uPLC connected to an MSQ Plus Single Quad Detector (SQD). Column: Hypersil Gold C4, 1.9 µm, 2.1 × 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 = 500–2,000. Scan time: 1.5 s. Data were 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 liters/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 500–1,500.

Characterization 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 2 × 50 mm column (Merck). Elution solvents were: (A) de-ionized water (18 MΩ resistivity) with 0.1% (vol/vol) formic acid (analytical grade, Sigma UK), and (B) acetonitrile (HPLC grade, Sigma, UK), isocratic 95% A for 0.5 min followed by gradient to 5% A for 2.5 min, isocratic for 0.5 min, then gradient for 0.05 min to 95% A followed by isocratic for 1.45 min. Instrument control, data acquisition, and data processing were performed using Masslynx 4.1 software. The mass spectrometer was operated in positive electrospray mode, and source conditions were adjusted to maximize sensitivity. Lockspray was used during analysis to maintain mass accuracy. The data were processed using Masslynx and deconvoluted using MaxEnt.

Tryptic peptides were analyzed by an Orbitrap Elite coupled with nanoLC. The data were 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.* Dithiothreitol (10 µL, 340 mM as a solution in H_2O, 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 was incubated at 21 °C for 1 h. The excess reducing agent was removed by repeated diafiltration into fresh buffer using VivaSpin sample concentrators (GE Healthcare, 10,000 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 was vortexed for 1 s before being incubated with shaking (2 h, 37 °C, 300 r.p.m.). LC–MS analysis confirmed the formation of GFPSul147. See Supplementary Fig. 19.

Fragmentation of GFP-Sul147 at pH 8 (sodium phosphate). GFPSul147 (100 µL, 1 mg/mL) in sodium phosphate buffer (100 mM, pH 8.0) was incubated with shaking (24 h, 37 °C, 300 r.p.m.). Fragmented products were observed using LC–MS. See Supplementary Fig. 19.

GFP-Sul147 wash-in studies. 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 resultant salts were re-dissolved in deuterium oxide (50 mL). This operation was repeated 5 times. Dithiothreitol (10 µL, 340 mM as a solution in D_2O, 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 was incubated at 21 °C for 1 h. The excess reducing agent was removed by repeated diafiltration into fresh deuterated buffer using VivaSpin sample concentrators (GE Healthcare, 10,000 MWCO). 2,5-dibromohexanediamide (DBHDA, 10 µL, 17 mM as a solution in DMSO-d_6, 50 equiv.) was added to the reduced protein solution and the mixture was vortexed for 1 s before being incubated (24 h, 37 °C, 300 r.p.m.). LC–MS analysis revealed a mix of deuterated GFPSul147 with fragment on-protein and deuterated small-molecule fragments. See Supplementary Figs. 21 and 22.

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 resultant salts were redissolved in deuterium oxide (50 mL). This operation was repeated five times.

Formation of SBLDha156 in deuterated buffer. Lyophilized 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 r.p.m. 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 r.p.m. at room temperature for 30 min and at 37 °C. The reaction was monitored by LC–MS. The reaction was typically complete after 2 h. The mixture was analyzed by LC–MS/MS. See Supplementary Fig. 23.

General method for formation of 'H3DhaX' in deuterated buffer. A solution of lyophilized 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 min 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 min and then at 37 °C for 60 min. The reaction mixture was then cooled to room temperature and precipitated DBHDA was removed by centrifugation (2 min, 14,000*g*). The reaction was monitored by LC–MS and analyzed by LC–MS/MS. See Supplementary Fig. 24.

Site-selective α-deuteration of proteins. *Preparation of protein and reagents to minimize the presence of exchangeable protons.* Protein was lyophilized three times in deuterium oxide. Deuterated buffer: For 50 mL of solution, the desired amounts of buffer components were dissolved in the desired volume of deuterium oxide. After adjusting the pH and volume, the solutions were lyophilized and resolubilized with fresh deuterium oxide.

Exploratory reactions with sodium sulfide. 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 Na_2S (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 min. At 10, 20, 30, 40, and 50 min, 6 µL of the Na_2S solution was added. The solution was kept shaking in between each addition. Ten minutes after the last addition, the reaction mixture was analyzed by protein MS to ensure complete conversion. Na_2S was removed by filtering through a Micro Bio-Spin 6 Column (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 Column. For protein MS analysis, 3 µL of the reaction mixture was taken and transferred to an MS vial. 50 µL of Milli-Q water and 3 µL of 1 M HCl solution were then added. The mixture was then analyzed by LC–MS (note: addition of dil. HCl (aq.) to acidify before MS analysis can improve the quality of MS spectra). See Supplementary Fig. 1c.

Exploratory reactions with sodium thioacetate. Monitoring of SBLS156Dha treated with KSAc at pH 8 (A. t = 0 h, B. t = 1 h) showed formation of SBLS156Acc. Attempted deprotection of SBLS156Acc to SBLS156C using NH_2OMe (500 equiv.) in 50 mM NaPi pH 8 returned only starting material after 4 h[50]. See Supplementary Fig. 1c.

Typical procedure for protein deuteration via thiophosphate. Anhydrous sodium thiophosphate was prepared as described previously from commercially available sodium thiophosphate hydrate[51].

Formation of H3DhaX. Lyophilized histone H3CX (5 mg) was solubilized in reaction buffer, deuterated 50 mM Tris, 5 M GdmCl pH 8 buffer (500 uL). 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 r.p.m. at room temperature for 30 min. The mixture was desalted using a 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 r.p.m. at room temperature for 30 min before heating up to 37 °C

for typically 3 h. The reaction was monitored using LC–MS. See Supplementary Figs. 2a and 18.

Formation of H3dpCX. The reaction was monitored by LC–MS to avoid overalkylation. One aliquot of the mixture (500 µL) was desalted and the buffer was exchanged using a PD-Minitrap G25 preequilibrated with a 0.4 M Na_3SPO_3, 5 M GdmCl pH 8.5 buffer to afford 1 mL of solution. The mixture was shaken at 600 r.p.m. at 37 °C for typically 2 h. One aliquot of the mixture (500 µL) was desalted and the buffer was exchanged using a 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 LC–MS. Observed differences in mass during synthesis were also due to slowly exchangeable deuterons inside the protein. See Supplementary Fig. 2a

Formation of H3dCX. Method A: the mixture was acidified to pH 4 with addition of glutathione. The mixture was shaken at 600 r.p.m. at 37 °C for typically 1 h to afford the desired H3dCX. The reaction was monitored by LC–MS. 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. See Supplementary Figs. 2b,c and 18.

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 the PP1 provider (BioLabs, P0754S): 4 µL of H3dpCX (1 mg/mL) + 4 µL of NE buffer (×10) + 4 µL of $MnCl_2$ buffer (×10) + 0.4 µL PP1 (2,500 U/mL). The volume was adjusted to 40 µL with Milli-Q water and incubated at 30 °C. The reaction was monitored by LC–MS. The reaction was complete after 30 min for H3dC10 and 2 h for H3dC26. An adduct was observed by LC–MS corresponding to thiophosphate (+112). Treatment with a 10 mM DTT aqueous solution afforded the desired H3dCX. In both cases, the mixtures were then dialyzed using refolding buffer consisting of 20 mM Tris, 250 mM KCl, and 1 mM EDTA to afford the desired H3dCX and DTT was added to afford a 20 mM Tris, 250 mM KCl, 1 mM EDTA, and 5 mM DTT solution.

Deuterium incorporation detection. See Supplementary Fig. 3.

Deuterium incorporation estimation using LC–MS on intact protein. Aliquots of 0.5 mg/mL of the desired protein were prepared in water. The aliquots were analyzed using Xevo G2-S Q-Tof. The data were processed using MaxLynx v4.1. Deconvolution was set up: output mass, ranges 10,000:20,000, 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. See Supplementary Fig. 4 and Supplementary Tables 1 and 2.

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

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

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

This equation is equivalent to:

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

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

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

Validation of deuterium incorporation levels and confirmation of the location of deuterium by LC–MS/MS of tryptic peptide. An aliquot of H3 (1 mg/mL, 25 µL) was diluted with a 8 M aqueous solution of urea (75 µL). The sample was incubated at room temperature for 10 min with occasional vortexing. The protein was reduced with DTT (200 mM solution in water) at 56 °C for 25 min and alkylated with iodoacetamide (400 mM solution in water) at room temperature for 30 min in the dark. The sample was diluted with ammonium bicarbonate buffer and trypsin was 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/µL with Milli-Q water. The sample was analyzed by LC–MS/MS.

Data processing as confirmation of the location of deuterium. The data were processed using Mascot on own database containing H3 variants with variable modifications: custom-made deuterated carbamidomethyl on cysteine residue; peptide mass tolerance: ±20 p.p.m.; fragment mass tolerance: ±0.5 Da; max missed cleavages: 3; instrument type: default. See Supplementary Fig. 4 and Supplementary Tables 1 and 2.

For H3dC10: Protein sequence coverage: 68%; matched peptides shown in bold.

1	MARTK**QTARK**	**CTGGKAPRKQ**	**LATKAARKSA**	**PATGGVKKPH**	**RYRPGTVALR**
51	**EIRRYQKSTE**	**LLIRKLPFQR**	**LVREIAQDFK**	**TDLR**FQSSAV	MALQEASEAY
101	LVALFEDTNL	AAIHAK**RVTI**	**MPKDIQLARR**	IRGERA	

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

1	MARTK**QTARK**	**STGGKAPRKQ**	**LATKAACKSA**	**PATGGVK**KPH	R**YRPGTVALR**
51	**EIRRYQKSTE**	**LLIRKLPFQR**	**LVREIAQDFK**	**TDLRFQSSAV**	**MALQEASEAY**
101	**LVALFEDTNL**	**AAIHAKRVTI**	**MPKDIQLARR**	**IR**GERA	

Deuterium incorporation estimation using LC–MS on tryptic peptide. The data generated by LC–MS/MS 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 p.p.m. and 4 decimals). 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. Three different peptides were used as triplicates. For example, MS1 analysis of the peptide 4–18 from the tryptic digests of H3dC10 batches prepared using conditions 1, 2, 3, 4 and 5 (see Supplementary Table 2) showed different mixture of TKQTARKC_{CAM}TGGKAPR and TKQTARKdC_{CAM}TGGKAPR. The comparison between isotopic patterns of nondeuterated peptide and sample peptide allowed calculation of the percentage of deuteration incorporation. See Supplementary Figs. 6–8.

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 unlabeled and labeled THTCD, in which the ^{13}C and ^{34}S isotopologue influence has been subtracted based on the theoretical unlabeled 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. s.d. values were calculated to generate the error bars.

Analysis of the stereoselectivity of deuteration. Analysis of (D-/L-)C10 ratio in histone H3. Histone samples (1 µg/µL, 100 µL in 3 M guanidinium chloride, 100 mM sodium phosphate buffer pH 8.5) were mixed with DTT (500 mM in water, 1 µL) and incubated at 25 °C for 15 min. 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 min in the dark. The excess 2-chloroacetamide was quenched with DTT (500 mM in water, 11 µL) at 25 °C for 30 min. Alkylated histone sample was buffer-exchanged to ArgC incubation buffer (50 mM Tris, 5 mM $CaCl_2$, 2 mM EDTA, pH 7.8) using a 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 r.p.m.) at 37 °C for 4 h. Peptides were desalted using Oasis HLB 1-cc 10-mg column (Waters) following the manufacturer's protocol. Dried peptides were redissolved in 0.1% formic acid, 2% acetonitrile (100 µL) and analyzed on a 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 min; 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 350–1,500 *m/z*, resolution 120,000, AGC target 1E6, maximum injection time 100 ms). After the MS scans, the ten 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). See Supplementary Figs. 9–17.

Reaction of H3dC in aqueous buffer. A solution of H3dCX was 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 min and then at 37 °C for 60 min. The reaction mixture was then cooled to room temperature. The reaction was monitored by LC–MS and analyzed by LC–MS/MS. See Supplementary Fig. 26.

Small-molecule reactions. See Supplementary Note 1 and Supplementary Figs. 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 functional[52] 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 visualized 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.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All key MS data supporting figures are given in the Supplementary Information, 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'; https://doi.org/10.5287/bodleian:NG0gbEEzP.

References

50. Hermanson, G. T. *Bioconjugate Techniques* (Academic Press, 1995).
51. Yasuda, S. K. & Lambert, J. L. Preparation and properties of anhydrous trisodium and tripotassium monothiophosphates. *J. Am. Chem. Soc.* **76**, 5356–5356 (1954).
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., Scalmani, 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. L. C-alkylation of chiral tropane- and homotropane-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).

Corresponding author(s): Benjamin G. Davis (ben.davis@chem.ox.ac.uk)

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes chosen based on reproducibility.

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analyses.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

All experiments were performed in experimental replicates described, where material allowed.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

There was no randomization.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

There was no blinding.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Protein LC-MS: Total mass spectra were reconstructed from the ion series using the MaxEnt algorithm preinstalled on MassLynx software (v. 4.0 or 4.1, Waters) or the presintalled ProMass software (v. 2.8, Thermo). Tryptic peptides MSMS data were processed using Mascot (Matrix Science), Maxquant (Max Planck Institute) and Thermo Xcalibur Qual (Thermo Fisher Scientific). NMR data were analysed using MestReNova (v. 11.0.2, Mestrelab Research). DFT calculations were performed with Gaussian09 rev D.01.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.	No restrictions on any unique materials. Most commonly available through protocols given or samples available on request.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	The study did not involve human research participants.