Post-translational mutagenesis for installation of natural and unnatural amino acids into recombinant proteins

Tom H. Wright, Benjamin G. Davis*

Department of Chemistry, University of Oxford, Oxford OX1 3TA, UK.

* = Corresponding author. Email: ben.davis@chem.ox.ac.uk

KEYWORDS Protein, radical, dehydroalanine, mutagenesis, synthetic biology, post-translational modification, Dha, PTM, histone, chemical biology, phosphorylation, methylation

EDITRIAL SUMMARY Producing homogenously modified protein side-chains is crucial for many biochemical and biophysical assays. This protocol describes how to perform a two-step 'post-translational mutagenesis' for the site-specific installation of protein side-chains.

TWEET Post-translational modification of protein side-chains

COVER TEASER Post-translational amino-acid modification

Up to four primary research articles where the protocol has been used and/or developed:

1. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

2. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chemical Science* **2**, 1666-1676 (2011).

3.
4.

ABSTRACT

Methods for installing natural and unnatural amino acids and their modifications into proteins in a benign and precise manner are highly sought-after in protein science. Here we describe a protocol for 'post-translational mutagenesis' that enables the programed installation of protein side-chains through the use of rapid, mild and operationally simple radical chemistry performed on recombinantly expressed and purified protein constructs. By introduction of protein dehydroalanine residues (in this instance, from a unique cysteine residue installed by site-directed mutagenesis) as free-radical trapping 'tags' for downstream modification, exquisite control over the site of subsequent modification is achieved. Using readily available alkyl halide precursors and simple borohydride salts, alkyl radicals can be generated in aqueous solution. These alkyl radicals react rapidly with protein-bound dehydroalanine residues to give functionalised protein products with new carbon-carbon bonds. Once the dehydroalanine is installed, the introduction of desired functionality is limited only by the requirement for polarity matching of the generated radical with the dehydroalanine 'acceptor', the solubility of the alkyl halide precursors in aqueous solution and the kinetics of the reaction itself. For example, methylated derivatives of lysine, arginine and glutamine are readily accessible. Further, since the side-chains are constructed chemically, many unnatural modifications can also be directly introduced as part of the side-chain, including isotope reporters (19F, 13C) that can be used in biophysical experiments such as protein NMR. From a suitable cysteine mutant of the target protein, the entire procedure for this chemical post-translational mutation takes two days and is readily performed by non-chemists.

INTRODUCTION

Post-translational modifications (PTMs) expand the chemical diversity available for protein function beyond ribosomally-processed amino acids[1]. The installation, removal and recognition of protein PTMs appears to underlie many of the central processes in biology. Yet, the precise functional consequences of many PTMs are only poorly understood since to the availability of homogeneous samples of site-specifically modified proteins is still limited. A straightforward protocol would be to install these modified residues, in a manner analogous to the pre-expression site-directed, gene mutagenesis of canonical amino acids. This would enable the use of, for example, intact, more relevant, fully-structured modified proteins as substrates in biological assays that currently use only peptide substrates (such as in *in vitro* binding assays, structural analyses or proteomic 'pull-down' technologies to identify interacting partners).

To date, chemical total or semi-synthesis of modified proteins and codon reassignment approaches have emerged as useful tools for access to proteins incorporating site-specific, defined PTMs and unnatural, 'designer' amino acids. The step-wise chemical synthesis of proteins enables precise control of the PTM site and chemical identity via linear assembly of the final polypeptide chain. However, as the routine synthesis of peptides is still limited to 40-50 amino acids, assembly of whole proteins involves the synthesis of multiple peptide fragments followed by chemical ligation and final refolding[2,3]. These are non-trivial procedures involving a significant investment in time and expertise not available to all laboratories.

Purely biological routes to modified proteins offer the advantage of requiring less chemical expertise. Amber codon suppression enables the programmed introduction of non-canonical amino acids into proteins in response to a reassigned stop-codon, via evolution of an orthogonal aminoacyl-tRNA synthetase/tRNA pair specific for the amino acid of interest[4,5]. However, despite impressive progress, the range of biologically relevant PTMs that have been incorporated remains limited. This is likely due to the difficulty in generating sufficient selectivity or tolerance in synthetase enzymes for either very small (methylation) or very large (glycosylation) modifications in amino acid substrates. Furthermore, expression yields and fidelity of the mutant proteins generated this way can show strong dependence on the amino acid incorporated, the synthetase system used for incorporation, and even the site and sequence of the target protein itself.[6]

A third route to modified proteins, the 'tag and modify' approach, combines the genetic introduction of a uniquely reactive 'tag' amino acid and its subsequent chemical modification[7]. Tags may be reactive natural amino acids such as cysteine or unnatural handles with reactivities (azide, alkyne, etc.)[8] not typically found in nature. The use of chemistry that is selective for the introduced tag results in site-specific chemoselective modification, avoiding the production of heterogeneous mixtures. Dehydroalanine (Dha), in particular, is a readily accessible protein electrophile that possesses chemical reactivity distinct from that of the canonical amino acids[9,10]. Dha can be installed in a site-specific manner into polypeptides through alkylation/elimination of unique cysteine or selenocysteine residues; by chemical elimination of phosphoserine or phenyl- or alkyl-modified selenocysteine under basic conditions or through the use of enzyme-mediated side-chain elimination. Our group and others[11-13] have demonstrated the use of Dha as a tag for the introduction of PTMs and their mimics via operationally simple thia-Michael (*carbon-sulfur β,γ* bond formation*)* chemistry employing thiol-functionalised reagents[14].

We have recently reported a new approach, termed 'post-translational mutagenesis'[15,16] that extends and generalises the 'tag-and-modify' approach by using the *carbon-carbon (β-γ)* bond forming reaction of alkyl radicals with dehydroalanine-tagged proteins to chemically 'construct' or 'bolt-on' the side-chain of interest. This exploits the radical acceptor reactivity of Dha. Since the β,γ carbon-carbon bond is naturally present in the side-chain of many amino acids, this chemistry enables facile access to these residues and their PTMs in a programmed manner. Notably, as our strategy relies on *in vitro* chemistry to effect the key 'chemical mutation' it is not constrained by the 'evolutionary filter' of the cell's translational machinery, allowing installation of side-chains that have not yet been incorporated by codon reassignment (such as methylated arginine and glutamine). In total, this approach allowed programmed installation of over 25 different natural, modified and unnatural side chains from a single dehydroalanine 'tag' residue into a variety of protein scaffolds[15].

Overview of the procedure

The protocol presented here describes two facile and chemoselective chemical modification reactions that, starting from a free cysteine residue, together enable a formal 'chemical mutation' to the desired PTM or designer amino acid **(Figure 1).**

The first step describes the conversion of cysteine to dehydroalanine via a mild one-pot alkylation and elimination procedure (Steps 1-8). The resulting Dha-tagged protein is then used as a substrate for modification with alkyl halide reagents, with sodium borohydride mediating the radical alkylation that installs the desired side-chain (Steps 9-21). The polarity of the radical generated from the corresponding halide is a key consideration for selection of such 'side-chain' donors, with 'nucleophilic'-type radical reagents better matched to the electrophilic Dha centre. Notably, tertiary, secondary and primary alkyl halides are all readily coupled, including those with unprotected, biologically relevant hydroxyl and amine functionality[15]. In this protocol, our approach is illustrated by two examples – installation of dimethyl-lysine at site 9 of histone H3, and installation of the phosphatase-resistant difluoromethylene analogue of phosphoserine, Pfa[17], at site 10 of histone H3 **(Figure 1)**. Both of these sites are biologically relevant and are naturally occupied by either modified or unmodified variants of K9 and S10, respectively.

Applications of the method

The downstream uses of the thus-modified proteins are varied and include both applied and fundamental applications. Access to the requisite modified proteins as substrates or reagents has greatly hindered *in-vitro* biochemical assays of PTM function, including biophysical studies, such as protein NMR, that are further enabled by the ready incorporation of isotopic labels demonstrated here. Although cellular studies of chemically modified protein conjugates are limited by the requirement for techniques for localising proteins to the cellular compartment of interest (such as microinjection), we have shown the use of modified proteins as affinity reagents for proteome-wide assay of PTM binding partners[15]. In a biomedical context, we have previously demonstrated the modification of a single-chain mammalian antibody, including with complex glycosylated structures[15], highlighting potential applications in antibody-mediated cellular and *in-vivo* imaging[18] and antibody-drug conjugates. Fundamentally, by extending the 'side-chain space' accessible to protein function, we envision wide-ranging future applications of our method, including in relatively unexplored areas such as modulation of enzyme catalysis by unnatural amino acids.

Limitations of the method

Our protocol is, of course, not without limitations. The most severe of these is the requirement for dehydroalanine as an orthogonal 'handle' for modification. While providing the requisite chemoselectivity[15] for the subsequent radical modification, the requirement for introduction of dehydroalanine limits the method to protein targets where this residue can be installed in a facile and site-specific manner. Interest in dehydroalanine as a useful 'tag' for chemical modification has led to a number of methods for its site-specific installation into expressed proteins[11,19,20]; while the chemical method described here is the most straightforward, it is not necessarily the only or ideal route to consider. For both installation of dehydroalanine from cysteine, and its subsequent radical modification, a degree of reagent (and hence solvent) accessibility is required for the targeted site, which can be determined empirically. That said, accessibility may be increased for even buried sites by the use of denaturing (or partially-denaturing) conditions; if the latter approach is adopted then secondary structure determination (e.g. through circular dichroism) and / or a protein refolding protocol, if necessary, should be considered. For proteins containing many free cysteine residues, for example, alternate routes to dehydroalanine based on ribosomal incorporation of non-canonical amino acids and subsequent chemical modification[19] may be employed to circumvent the requirement for chemical modification of a specific cysteine.

In conjunction with other approaches, we believe that protocols for such benign protein side-chain manipulation by chemical mutation will enable a new approach to direct ‘protein editing’ to supplement and extend genetic manipulations.

Experimental design

Choice of modification strategy Our protocol involves recombinant expression of a suitably designed cysteine mutant of the protein-of-interest, followed by two facile and chemoselective chemical modification reactions that together enable a formal 'chemical mutation' to the desired PTM or designer amino acid **(Figure 1).** The protocol is dependent on the introduction of the non-canonical amino acid dehydroalanine (Dha) as a 'tag' or uniquely 'radical-reactive' handle that guides the subsequent key chemical modification that constructs the target side-chain. Dehydroalanine can be introduced to proteins via a number of routes, typically from free cysteine[10], phosphoserine[19] or phenyselenocysteine[11,20]. Of these routes, one of the most straightforward is genetic incorporation of cysteine at the desired site of modification (using standard site-directed mutagenesis), followed by one-pot chemical bisalkylation and elimination to give dehydroalanine. Undesired native cysteines can be transformed by mutagenesis to near-isosteric Ser if needed; although this is a conservative mutation, this procedure will require validation of the physical and biological effect of such a change in the resulting construct. Chemical conversion of phosphoserine and phenylselenocysteine to dehydroalanine requires highly basic or oxidative conditions that may be not compatible with sensitive protein substrates. Furthermore, as these latter precursors are non-ribosomal amino acids, additional genetic engineering steps are required to incorporate these residues by codon reassignment, which can drastically impact yields of expressed protein. However, in cases where a target protein contains essential free cysteines, the use of alternative precursor residues might in principle enable chemoselective (or potentially sequential) dehydroalanine formation. Notably, the key modification of dehydroalanine by alkyl radicals is possible in the presence of free cysteine residues[15], potentially enabling further subsequent chemical mutations. In this work, cysteine introduction to histone proteins is performed by site-directed mutagenesis, according to manufacturer's instructions. Expression and purification of histone proteins is a robust procedure outlined elsewhere[21]

Relative reactivity: protein Both dehydroalanine formation and subsequent radical modification have been demonstrated on a number of protein sites and scaffolds, under both 'native' and denaturing conditions[11]. When using folded proteins as substrates for chemical reactions, a critical variable that can affect rates and even the possibility of reaction is the reactivity of the targeted amino acid in the context of the protein tertiary structure. Solvent accessibility provides one crude estimate that can correlate with reactivity (a form of 'reactive accessibility'). In the case of multiple cysteines, relative reactivity can be gauged (for example, by titration with Ellman's reagent or iodoacetamide and tryptic peptide mapping) ~~and even tuned~~ to enable regioselective dehydroalanine formation on folded proteins. Undesired reactive cysteines can ~~also~~ be 'mutated out' (e.g. to Ser) to further guide the chemoselectivity of the dehydroalanine introduction[22]. Alternatively, other methods of Dha introduction may be considered (see above). For both of the reactions discussed in this protocol, if a particularly hindered site is targeted, partial or full denaturation of the protein may simplify chemical modification, in which case consideration should be given to suitable procedures for refolding the protein following reaction.

Relative reactivity: alkyl halide In this protocol the installation of dimethyllysine, via radical generation from 3-(iodopropyl)dimethylamine hydroiodide, and the installation of Pfa, via radical generation from (bromodifluoromethyl)phosphonic acid, are highlighted. Reactivity of alkyl halide 'side-chain donor' reagents follows the expected trend based on the radical mechanism: tertiary > secondary > primary. Primary alkyl halides are still competent in the reaction, and, depending on aqueous solubility, may be very reactive. Iodides are preferred, although in certain cases the bromide congener may prove useful if the iodide is very reactive and results in high levels of di-alkylation side-products (see below). In general, salts are preferred due to greater water solubility, although a number of commercially available short chain alkyl iodides (*isopropyl, tert-butyl* iodide etc.) can be used successfully as neat liquids[15,19]. As radical polarity effects may dominate reactivity, the most useful donors generate nucleophilic radicals following halogen abstraction, which can be productively trapped by protein dehydroalanine residues[15]. For a given alkyl halide, reactivity can be readily determined by initial 'titration', in a series of reactions, of the molar equivalents of reagent, in the presence of a fixed concentration of protein-Dha and sodium borohydride, and adjusted accordingly for full conversion to the desired mono-alkylation product.

Reaction scale The post-translational mutagenesis procedure can be performed on a variety of scales, with our laboratory having successfully performed the key radical C-C bond formation on individual protein samples ranging from 50 μg to 10 mg. In all cases examined, higher working concentration of protein results in more efficient reaction. The examples reported in this protocol make use of 200 μg of histone-Dha in each reaction, with a working concentration of ~ 1 mg/mL (66 μM for the H3-Dha employed).

Reaction monitoring We recommend the use of LC-MS for monitoring of reactions on protein substrates. SDS-PAGE analysis typically does not have the requisite resolution for cleanly distinguishing the <300 Da mass shifts typical of side-chain modification during chemical mutation. We typically use an electrospray (+)-time of flight [LC-ES(+)-TOF] MS instrument coupled to an HPLC with a short protein desalting column (ProSwift® or Chromolith®). Protein content can be observed by UV/Vis plus total-ion-count (TIC) chromatograms. It is important that to be truly representative during such monitoring that all protein peaks are utilized (integrated to observe protein ion series) for subsequent deconvolution methods yielding total protein masses **(Figures 2-4)**. Typically variation of ionizability is small between product and reactant proteins, thereby allowing near-quantitative use of ion count chromatograms; such quantification is readily verified by plotting concentration *vs* TIC.

Side-reactions Dehydroalanine is readily formed from free cysteines by bisalkylation with the dibromide compound 2,5-dibromohexane-diacetamide (DBHDA). As mildly basic conditions (pH 8.0 or higher) promote this reaction, alkylation at non-target nucleophilic residues, such as lysines, is a potentially possible but rarely observed, side reaction that could be avoided if seen by optimization of buffer pH or reagent amount for cysteine selectivity. More generally, temperature, reagent concentration and reaction time can all be optimized to effectively eliminate competing side-reactions. Radical modification is largely insensitive to pH, enabling the use of neutral or slightly acidic conditions that disfavour possible alkylation side reactions with the alkyl halides employed as radical precursors. However, radical-mediated dual-modification (‘di-alkylation’) at dehydroalanine is a competing side-reaction that is inherent to the reaction mechanism and may be observed as a ‘second addition’ of the expected side-chain mass (**Figure 3)**[15], especially with methods proposed to generate radicals via metals (e.g. Zn(0) or In(0)). Use of borohydride and, if needed, titration of the side-chain precursor reagent concentration while fixing the molar equivalents of borohydride relative to protein should reveal conditions where mono-addition is the sole product. Under anaerobic conditions, the only other side-reaction potentially observed is reduction of dehydroalanine to alanine; although rare this may be controlled typically through temperature variation (e.g. cooling). Under aerobic conditions, oxidative protein cleavage at the introduced dehydroalanine is observed as a side-reaction[15]. While it may prove possible to perform size-exclusion chromatography or HPLC to remove unwanted cleaved fragments, yields of the desired protein are inevitably lowered. We recommend use of a positive-pressure inert atmosphere glove box or similar low oxygen environment.

<u>**MATERIALS**</u>

Reagents

Full procedures for all non-commercially available reagents can be found at www.sciencemag.org/cgi/content/full/science.aag1465/DC1

- Cysteine-modified Histone H3, produced and purified as described above[15]
- DL-dithiothreitol (ThermoFisher, cat. no. R0861)
- 2,5-dibromohexane diacetamide (DBHDA, 2,5-dibromohexane diacetamide (DBHDA) is readily synthesized from adipic acid[14]. This reagent is also available to the community through our laboratory and can be stored for many years as a bench-stable solid.)
  CAUTION: Toxicity information for this compound is not yet available: assume toxic. Handle with gloves at all times and work in an appropriately ventilated environment.
- Sodium borohydride (Sigma-Aldrich, cat. no. 213462)
- 3-(iodopropyl)dimethylamine hydroiodide (Readily synthesized in a single step from commercial starting materials as per Wright et al[15]. Solid crystals or powder thus obtained can be stored for many years as a bench-stable solid.)
  CAUTION: Toxicity information for this compound is not yet available: assume toxic. Handle with gloves at all times and work in an appropriately ventilated environment.
- (bromodifluoromethyl)phosphonic acid (Readily synthesized by refluxing diethyl (bromodifluoromethyl)phosphonate (Sigma-Aldrich, cat. no. 411361) overnight in conc. HCl, as per Wright et al[15]. The oil thus obtained should be stored at -20°C and should be stable for months at this temperature.)
  CAUTION: Toxicity information for this compound is not yet available: assume toxic. Handle with gloves at all times and work in an appropriately ventilated environment.
- Formic acid (LC-MS grade)
- Water (Deionized water: (>15 MΩ cm^{-1} resistance), filtered through 0.2-μM disc filter)
- Acetonitrile (LC-MS grade)

Reagent setup

'Dha reaction buffer': 50 mM sodium phosphate buffer (pH 8.0), supplemented with 3 M guanidine hydrochloride. Once made up, this buffer can be stored and used for up to a month at room temperature.

'Chemical mutation buffer': The radical chemical mutation reaction is largely insensitive to the buffer selected. For work with histone substrates, we have found 500 mM ammonium acetate (pH 6.0), supplemented with 5 M guanidine hydrochloride, to be a highly effective choice. Once made up, this buffer can be stored and used for up to a month at room temperature.

DBHDA: Prepare a stock solution (0.5 M) of DBHDA by dissolving 100 mg in 610 μL of DMF. Stock solutions of DBHDA should be prepared fresh for each experiment. **CAUTION:** Toxicity information for this compound is not yet available: assume toxic. Handle with gloves at all times and work in an appropriately ventilated environment.

3-(iodopropyl)dimethylamine hydroiodide (1): Stock solutions (0.8 M) of this reagent should be prepared fresh on the day of the experiment. **CAUTION:** Toxicity information for this compound is not yet available: assume toxic. Handle with gloves at all times and work in an appropriately ventilated environment.

(bromodifluoromethyl)phosphonic acid (2): Stock solutions (1 M) of this reagent should be prepared fresh on the day of the experiment.**CAUTION:** Toxicity information for this compound is not yet available: assume toxic. Handle with gloves at all times and work in an appropriately ventilated environment.

LC/MS solvent A: 0.1% (vol/vol) formic acid in water
LC/MS solvent B: 0.1% (vol/vol) formic acid and 95% (vol/vol) acetonitrile in water

For best results, prepare all HPLC solvents fresh and purge the LC-MS system before analysis of protein samples.

Equipment

PD-minitrap G25 (GE Healthcare Life Sciences cat. no. 28-9180-07)

Vivaspin 500 centrifugal concentrator MWCO 5 kDa (GE Healthcare Life Sciences cat. no. 28932223)

SpinTrap™ G-25 desalting column(GE Healthcare Life Sciences cat. no. 28-9180-04)

Positive-pressure inert gas glovebox: We use a Belle Technology glovebox (http://www.belletechnology.co.uk/glovebox.php) equipped with the BASF R3-11G catalyst

LC/MS: Waters LCT Classic coupled to a Shimadzu 20 Series HPLC using a Thermo Proswift (250 x 4.6 mm x 5 μm) column.

Equipment setup

Liquid Chromatograpy: Water (solvent A) and 95% (vol/vol) acetonitrile (solvent B), each containing 0.1% formic acid, were used as mobile phase for gradient elution. The gradient program was as below.

Time interval (mins.)	% Solvent B

0	5
5	5
15	100
18	100

Mass spectrometry analysis
Mass spectrometry was performed using a Waters LCT Classic, with parameters as displayed in box below.
Nitrogen was used as the desolvation gas and nebuliser at a total flow of 600 l/hr.

Using Masslynx (Waters), mass spectra for the total protein content of the total ion count (TIC) chromatogram were combined to display and subsequently deconvolute the multiply charged ion series for all protein present.
Spectra were calibrated by matching at least 17 of the multiply charged ion series of equine myoglobin, run under the same gradient conditions.

Method Parameter	Value
Capillary voltage	3000 V
Cone voltage	250 V

PROCEDURE

Dehydroalanine formation on histone substrates Timing: 5 h

1. Prepare protein substrate solution by dissolving 12 mg lyophilized powder of the relevant histone cysteine mutant (for dimethyllysine modification, H3K9C; for 'Pfa' modification, H3S10C) in 500 µL of 'Dha reaction buffer'. Add 30 mg of DTT and vortex or sonicate the protein-DTT solution until all material is dissolved.
2. Incubate the solution for 30 minutes at room temperature (15-25 °C) to ensure complete reduction of the cysteine residue.
3. Carefully desalt the reduced protein by passing the protein solution through a PD MiniTrap G25 column pre-equilibrated with Dha reaction buffer. After elution, the protein solution should now be 1 mL with a concentration of approximately 10 mg/mL.
   CRITICAL STEP: This desalting step is essential to remove excess DTT. Excess DTT will react with and inhibit the dehydroalanine formed in the next step.
4. Initiate the formation of dehydroalanine by adding 100 µL of the DBHDA stock solution to the reduced protein. Gently shake the reaction at 25°C and monitor every 30 minutes by removing a 5 µL aliquot using a standard biological pipette and dissolving in 95 µL Milli-Q® H_2O (+ 0.1% formic acid) for LC-MS.

5. When no dehydroalanine is visible by LC-MS within 30 minutes to 1 hour, raise the temperature to 37 °C to promote the elimination reaction. Continue to monitor the reaction every 30 minutes by LC-MS.
6. When the reaction is judged complete by LC-MS (See Figures 2 and 4 for examples of histone-Dha LC-MS spectra), concentrate the reaction mixture to 500 μL using a Vivaspin 500 concentrator (MWCO 5 kDa).
7. Desalt the sample by passing the concentrated solution through a PD MiniTrap G25 column pre-equilibrated with chemical mutation buffer. Elute the sample with 1 mL of chemical mutation buffer to provide the Dha-tagged protein, ready for the next step.
8. Check the concentration of the protein by measuring the absorbance at 280 nm. With careful desalting, concentrations greater than 80% of the starting material should be obtained.

   PAUSE POINT: The protein can be stored at -20 °C for up to 6 months.

Chemical mutation of histone proteins Timing: 16 h

9. Prepare the reagents for overnight 'degassing' in the glovebox. It is most convenient to prepare Dha-tagged protein samples for overnight incubation in the glovebox, for reaction the following morning. Prepare 5 or more eppendorf tubes (in order to 'titrate' reagent equivalents for optimal reaction) containing 200 μL samples of Dha-protein solution at a concentration of 1 mg/mL (for a 15 kDa protein, ~ 65 μM), as determined by A280 measurement or BCA assay.
10. Prepare a stock solution of the alkyl bromide or iodide 'side-chain donor' reagent. Dissolve the reagent in deionized water to a concentration of ~ 0.8 M for (3- iodopropyl)dimethylamine hydroiodide (**1**) or 1 M for (bromodifluoromethyl)phosphonic acid (**2**). Vortex if necessary to obtain a homogenous solution. Adjust the pH of the stock solution to ~6.0.
11. Carefully weigh out five or more (depending on number of reactions required) 1 mg samples of solid sodium borohydride into 2 mL eppendorf tubes. Quickly spin the tubes in a bench-top lab centrifuge to ensure the solid contents are at the bottom of the tube. The tubes containing solid borohydride will be the final 'reaction vessel' so ensure legible labeling of these tubes before placing them into the glovebox.
12. Combine the protein, reagent and borohydride tubes on a sample rack and port into the glovebox, using the rapid port pre-equilibration as per manufacturer's instructions (typically 10 minutes). It is essential that the caps of all eppendorf tubes remain open throughout the degassing procedure. In the case of sensitive proteins, storage in an eppendorf cooler kept at 4°C within the glovebox is recommended. Note that this will slow the diffusion of O_2 from the protein solution, hence extra time should be allotted for sample degassing. Robust or denatured protein solutions may be stored at room temperature overnight, within the glovebox.

    CRITICAL STEP: It is essential that the protein solution is sufficiently degassed before chemical mutation is performed. Based on our experience, for a 200 μL buffered solution, this takes a minimum of 6 hours at room temperature. Allow longer degassing times for larger volume solutions and for proteins that require storage at reduced temperature. For example, for protein solutions stored at 4°C within the glovebox, 12 hours of degassing may be more appropriate.

13. Working within the glovebox, use a pipette to ascertain the volume of the protein and stock solutions, as the positive-pressure flow of inert gas within the glovebox will cause some evaporation (typically ~ 20 μL). 'Top up' the solutions as necessary using extra buffer or deionized water. Place the eppendorf tubes containing solid sodium borohydride into an eppendorf cooler kept at 4°C (if available – room temperature reaction is also sufficient).
14. When performing a chemical mutation reaction for the first time on a given substrate/reagent pair, 'titrate' the amount of reagent required for optimal modification and protein recovery. Three protein samples are usually sufficient. Calculate 100, 500 and 2000 molar equivalents of the bromide or iodide reagent relative to protein and determine the volume of reagent stock solution required.

15. Carefully pipette the required amounts of reagent stock into the relevant protein sample, and gently mix the resultant solution by pipette.
16. Initiate the reaction by pipetting the entire volume of the protein/reagent solution directly into the eppendorf tube containing solid sodium borohydride and gently shake the tube. Effervescence, or foaming, is observed due to the hydrolysis of borohydride in acidic media.
17. Repeat steps 15-16 for as many reactions as are being performed in parallel.
18. Incubate the reaction mixtures at 4°C with the cap open. Shaking is not necessary but may help to maintain a homogeneous solution.
19. Following 30 minutes of reaction time, cap the tubes and remove from the glovebox.
20. Analyse the reaction outcome by LC-MS. Aliquots of the protein reaction mixture can be directly used for LC-MS analysis. Dilute the reaction mixture 20-fold with H_2O containing 0.1% (vol/vol) formic acid and inject a small volume (for example, 10 μL) onto LC-MS system.
21. Optional: Depending on the downstream application envisioned and the behavior of the target protein, desalt the sample to remove excess reagent by loading the sample onto a SpinTrap™ G-25 desalting column, pre-equilibrated with the desired storage buffer. Elute the sample eluting according to the manufacturers instructions. For further downstream applications, such as biological assays, an additional size-exclusion column may be necessary for efficient sample desalting.

TIMING

Dehydroalanine formation on histone substrates.
Step 1-2: 30 minutes.
Step 3: 15 minutes.
Steps 4-5: Variable, typically 2-3 hours.
Steps 6-8: 1 hour.

Chemical mutation of histone substrates.
Steps 9-12: Preparation of reagents: 30 minutes, Degassing of samples: 6 hours to overnight.
Steps 13-17: 10 minutes.
Step 18: 30 minutes.
Step 19-21: 30 minutes.

TROUBLESHOOTING

Troubleshooting information is provided in **Table 1**

Table 1: Troubleshooting Table

Step	Problem	Reason	Solution
5	No or partial conversion to histone-Dha observed by LC-MS	Incomplete reaction	1. Add another20 μL of the DBHDA stock solution and incubate the reaction mixture for 1 h at 37 °C before re-analysis by LC-MS 2. Check if the pH of the 'Dha buffer' is >8.0. 3. Re-attempt the reaction with a fresh batch of DBHDA
17	No conversion to the desired product is observed by LC-MS, or protein products with a MW< than the starting protein are present	Reaction inhibition and oxidative protein cleavage due to presence of O_2	1. Repeat the reaction with extended period of degassing. Ensure all reaction components are thoroughly degassed. 2. Check glovebox function according to manufacturer's instructions.
18	Some, but incomplete conversion to the desired product	Incomplete reaction	1. Repeat the reaction with fresh protein-Dha, using a higher number of molar equivalents for the bromo/iodo-reagent. (Note: Reduction of Dha to Ala will occur even if no conversion to the chemical mutation product is observed- therefore it is not recommended to re-attempt reaction on 'used' Dha.) 2. Repeat the reaction with a higher number of molar equivalents of sodium borohydride.
18	*n* additions of the side-chain reagent to protein observed by	Radical di-alkylation at Dha, or alkylation of protein	1. Repeat the reaction with fresh protein Dha, using a reduced

	LC-MS	nucleophiles.	number of molar equivalents for the bromo/iodo reagent. 2. If >2 modifications are observed, check the reaction buffer pH and attempt the reaction at pH 6.0 or lower.

Anticipated results

Following the two reactions described above, proteins should be observed by LC-MS as a single species with >90% conversion to the desired ‘chemical mutation’ (**Figures 2 and 4, supplementary figures 1 and 2)**, as judged by calculation from peak intensities in the deconvoluted spectrum. No peaks below the expected starting material mass should be observed; if such peaks are visible the protein may have undergone oxidative cleavage[15]. SDS-PAGE analysis can also be used to assess protein fragmentation if oxidative cleavage is suspected. Proteolytic digest and subsequent peptide mapping by LC-MS/MS should confirm the desired site of modification and no additional, undesired modifications at other residues[15]. An unsuccessful reaction is shown in **Figure 3**; attempted installation of $H3K9me_2$ using indium powder[15] in place of $NaBH_4$ resulted in significant di-alkylation products observable in the mass spectrum.

Author contributions:

B.G.D and T.H.W designed research. T.H.W performed experiments. T.H.W and B.G.D wrote the paper.

Competing financial interests

The authors declare competing financial interests (see the HTML version of this article for details).

Acknowledgements:

T.H.W was supported by the Rutherford Foundation. B.G.D was supported by a Royal Society Wolfson Research Merit Award.

References

1 Walsh, C. T., Garneau‐Tsodikova, S. & Gatto, G. J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Intl Ed.* **44**, 7342-7372 (2005).

2 Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-776 (1994).

3 Malins, L. R. & Payne, R. J. Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr. Opin. Chem. Biol.* **22**, 70-78 (2014).

4 Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379-408 (2014).

5 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001).

6 Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **6**, 50-69, doi:10.1039/C4SC01534G (2015).

7 Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A "Tag-and-Modify" Approach to Site-Selective Protein Modification. *Acc. Chem. Res.* **44**, 730-741, doi:10.1021/ar200056q (2011).

8 van Kasteren, S. I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446**, 1105-1109 (2007).

9 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).

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

11 Levengood, M. R. & van der Donk, W. A. Dehydroalanine-containing peptides: preparation from phenylselenocysteine and utility in convergent ligation strategies. *Nat. Protoc.* **1**, 3001-3010 (2006).

12 Wang, J., Schiller, S. M. & Schultz, P. G. A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angew. Chem. Intl Ed.* **46**, 6849-6851 (2007).

13 Guo, J., Wang, J., Lee, J. S. & Schultz, P. G. Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angew. Chem. Intl Ed.* **47**, 6399-6401, doi:10.1002/anie.200802336 (2008).

14 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. Intl Ed.* **124**, 1871-1875 (2012).

15 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, aag1465 (2016).

16 Wright, T. H., Vallée, M. R. J. & Davis, B. G. From Chemical Mutagenesis to Post‐Expression Mutagenesis: A 50 Year Odyssey. *Angew. Chem. Intl Ed.* **55**, 5896-5903 (2016).
17 Zheng, W. *et al.* Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *J. Biol. Chem.* **280**, 10462-10467 (2005).
18 Gunnoo, S. B. *et al.* Creation of a gated antibody as a conditionally functional synthetic protein. *Nat. Commun.* **5**, 4388 (2014).
19 Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science* **354**, 623-626 (2016).
20 Wang, J., Schiller, S. M. & Schultz, P. G. A Biosynthetic Route to Dehydroalanine‐Containing Proteins. *Angewandte Chemie* **119**, 6973-6975 (2007).
21 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Chromatin Protocols*, 1-16 (1999).
22 Chooi, K. P. *et al.* Synthetic phosphorylation of p38 α recapitulates protein kinase activity. *J. Am. Chem. Soc.* **136**, 1698-1701 (2014).

FIGURE LEGENDS

Figure 1: General reaction scheme for 2-step 'chemical mutation' from Cys→Dha→R. The reaction scheme is exemplified here by the installation of dimethyllysine and difluoro-carbo-pSer (Pfa). The structures of the ~~two~~ modified residues are shown, with the segment derived from the alkyl iodide 'side-chain' donor highlighted in orange. The box highlights phosphoserine and phenylselenocysteine, alternative Dha precursors that can be installed biosynthetically via stop-codon suppression techniques.

Figure 2: Installation of dimethyllysine by reaction of histone-Dha. Dimethyllysine is installed from dehydroalanine with (3- iodopropyl)dimethylamine hydroiodide (1) mediated by sodium borohydride. (**a**) Two-step reaction scheme (**b-c**) Mass spectra showing deconvoluted spectra of H3-K9Dha calc. 15182 Da obs. 15181 Da (b) and H3-K9me$_2$ calc. 15268 Da obs. 15268 Da (c). Full spectra including multiply charged ion series can be found in the Supplementary Information Figure S1.

Figure 3: Unsuccessful installation of dimethyllysine by reaction of histone-Dha. Dimethyllysine is installed with (3- iodopropyl)dimethylamine hydroiodide (1) mediated by indium powder. Similar results are obtained with zinc powder and related metal-initiated systems. In addition to the desired Kme$_2$ product, a di-alkylated side-product is also observed. See Experimental design and Troubleshooting for details. (**a**) Reaction scheme (**b-c**) Mass spectra showing representative multiply charged ion series (**b**) and deconvoluted spectrum (**c**) of H3-K9me$_2$ calc. 15268 Da obs. 15268 Da. H3-K9me$_2$(di) calc. 15354 Da obs. 15354 Da

Figure 4: Installation of difluoro-carbo-pSer (Pfa) by reaction of histone-Dha. Pfa is installed from dehydroalanine with (bromodifluoromethyl)phosphonic acid (2) mediated by sodium borohydride. (**a**) Two-step reaction scheme (**b-c**) Mass spectra showing representative deconvoluted spectra of (**b**) H3-S10Dha calc. 15221 Da obs. 15224 Da and (**c**) of H3-S10Pfa calc. 15353 Da obs. 15354 Da. Full spectra including multiply charged ion series can be found in the Supplementary Information Figure S2.