Post-translational mutagenesis: a chemical strategy for exploring protein side-chain diversity

Authors: Tom H. Wright,[1] Ben J. Bower,[1] Justin M. Chalker,[1ξ] Gonçalo J. L. Bernardes,[1ψ] Rafal Wiewiora,[1χ] Wai-Lung Ng,[1] Ritu Raj,[1] Sarah Faulkner,[1] M. Robert J. Vallée,[1] Anuchit Phanumartwiwath,[1] Oliver D. Coleman,[1] Marie-Laëtitia Thézénas,[2] Maola Khan,[1] Sébastien R. G. Galan,[1] Lukas Lercher,[1†] Matthew W. Schombs,[1] Stefanie Gerstberger,[1] Maria E. Palm-Espling,[1] Andrew J. Baldwin,[1] Benedikt M. Kessler,[2] Timothy D.W. Claridge,[1] Shabaz Mohammed[1] & Benjamin G. Davis[1*]

Affiliations:

[1]Department of Chemistry, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK

[2]Target Discovery Institute, Nuffield Department of Medicine, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Headington, Oxford OX3 7FZ

*Correspondence to: Ben.Davis@chem.ox.ac.uk.

ξPresent address: Flinders University, School of Chemical and Physical Sciences, Sturt Road, Bedford Park, SA 5042, Australia.

ψPresent addresses: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK & Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028, Lisboa, Portugal

χPresent address: Sloan Kettering Institute, New York, NY, USA

†Present address: BMWZ and Institute of Organic Chemistry, Leibniz University Hannover, Schneiderberg 38, 30167, Hannover, Germany.

Abstract: Post-translational modification of proteins expands their structural and functional capabilities beyond those directly specified by the genetic code. However, the vast diversity of chemically-plausible (including unnatural but functionally relevant) side-chains is not readily accessible. We describe $C(sp^3)$–$C(sp^3)$ bond-forming reactions on proteins under biocompatible conditions, which exploit unusual carbon free radical chemistry, and use them to form Cβ–Cγ bonds with altered side chains. We demonstrate how these transformations enable a wide-diversity of natural, unnatural, post-translationally-modified (methylated, glycosylated, phosphorylated, hydroxylated) and labeled (fluorinated, isotopically-labeled) side-chains to be added to a common, readily-accessible dehydroalanine precursor in a range of representative protein types and scaffolds. This approach, outside of the rigid constraints of the ribosome and enzymatic processing, may be modified more generally for accessing diverse proteins.

Introduction

Natural post-translational modifications (PTMs) to proteins partially expand the chemical groups available to proteins, modulating both structure and function (Fig. 1a).(*1*) For example, protein glycosylation tunes both physical (e.g., solubility, stability and folding) and biological (e.g, immune response, cell adhesion events, signaling) activity.(*2*) Phosphorylation, one of the most frequently occurring posttranslational modifications, is widely used in nature as a powerful functional activation mechanism (as an "on switch") for proteins.(*3*) Even relatively small modifications such as methylation have been shown to be critical in a range of pathways with diverse biological effects, such as the transcriptional regulation mediated by histone proteins.(*4*) The ability to expand post-translational functional group diversity in an unbounded manner could therefore, in principle, allow exploration and understanding of even greater and more diverse effects in modulation of biological function.

The vast majority of all known natural PTMs feature bonds to heteroatoms (non-carbon) made at the γ (Cys *S*γ, Thr *O*γ, Ser *O*γ) or ω (Lys *N*ω, Tyr *O*ω) positions of sidechains.(*5*) Yet, one of the central features of living 'organic' matter is that it exploits carbon's ability as an element to catenate (typically through $C(sp^3)$–$C(sp^3)$ bond formation) — providing one of nature's most important structural motifs. As all amino acid side-chains contain this bond, mastering its construction on proteins could allow free-ranging structural alteration of residues in proteins (both natural and unnatural) and thence functional re-programming. Such extension of the chemical space accessible to protein engineering, could be considered a near unlimited form of synthetic biology, a form of 'chemical mutagenesis'.(*6, 7*)

Site-directed mutagenesis has revolutionized the study and understanding of proteins.(*8, 9*) This now long-standing technique, however, is generally restricted to the 20 natural amino acid building blocks by the high selectivity of natural aminoacyl tRNA synthetases and the limited plasticity of the ribosome,(*10*) which creates an effective 'filter' to translation. The incorporation of unnatural amino acids expands a protein's functional capacity and can provide insight into biochemical mechanisms.(*11-13*) Some strategies for the incorporation of non-canonical residues have emerged as a powerful route to unnatural mutant proteins. Biological techniques such as amber codon suppression,(*14*) while useful, remain limited in scope of structural variation(*13*) by the tolerance of the translational machinery and hence must be optimized by on a case-by-case basis.(*12*) The total or semi-synthesis of proteins has been made possible by powerful native chemical ligation techniques.(*15*) However, there remain restrictions on the size of the proteins that can be readily synthesized, such syntheses typically require many steps followed by correct refolding and hence expertise is required for all but the simplest protein targets.(*16*)

An alternative, divergent and potentially unlimited approach would be to incorporate a single amino acid that can act as a general chemical precursor for any desired side-chain, whether natural or unnatural.(*17*) The introduction of various chemical 'tags',(*18*) that allow selective protein modification via reactivity compatible with that of natural biomolecules,(*19*) has been an important step towards this goal. However, current protein modification approaches rely on unnatural carbon-heteroatom linkages(*20*) that do not construct the C–C framework found in biology and thus cannot be used for the site-selective introduction of natural posttranslational modifications and their modified variants.(*19*) Thus far, formation of carbon(sp3)–carbon(sp3) bonds for protein modification has remained out of reach, despite the ubiquity of the C-C bond in amino acid side chains. Access to such reactivity would enable the

rapid and divergent exploration of both natural and unnatural 'side-chain' space from a readily accessible precursor.

Here we demonstrate that such carbon(sp3)–carbon(sp3) bond-formation is possible in a protein-compatible manner by exploiting the selectivity of carbon-centered radicals to allow a form of general post-translational mutagenesis.(*7, 17, 21*)

Design of a biocompatible C(sp^3)–C(sp^3) bond-forming reaction

We envisaged a unique strategy through retrosynthetic analysis (Fig. 1b). In principle, carbon(sp3)-carbon(sp3) disconnections at the β,γ C–C bond would allow the chemical installation of not only natural amino acid residues (AAs) but also their post-translationally modified variants (ptmAAs) and a wide range of unnatural amino acids (uAAs). Traditional two-electron chemistry (employing nucleophiles and electrophiles, Figure 1b) results in consideration of reagents likely incompatible with biological substrates by virtue of their reactivity with water and/or the functional groups found in natural biomolecules. We considered that single-electron chemistry might prove more compatible with proteins, since suitable free radicals are tolerant of aqueous conditions(*22*) and unreactive (and thereby compatible) with the majority of existing functionality present in biomolecules ('bioorthogonal'(*19*)). We reasoned that use of such mild, carbon-centred free radical chemistry (Fig. 1b,c) would be enabled by matching free-radical reactivity with a suitable, uniquely-reactive functional group partner that possesses a chemical affinity for such singly-occupied molecular orbitals (SOMOs). The amino acid residue dehydrolanine (Dha) can be readily introduced in a site-selective manner genetically,(*23, 24*) biosynthetically(*25*) or chemically(*26, 27*) and is a potent 'SOMOphile',(*28*) which, upon reaction with a suitable radical, would favourably generate a *capto*-dative stabilized Cα radical (Figs 1c and 2).

Development of a peptide-compatible C(sp^3)–C(sp^3) bond-forming reaction

Our attention focused on methods for the ready generation of suitable carbon-centred free radicals and the suitable productive 'quenching' of the central Cα radical intermediate **1** generated after formation of the critical C–C bond (Fig. 2b). We considered that alkyl radicals might be derived from corresponding alkyl halides (R–Hal) through processes that would generate single electron species either through direct homolytic bond-fission (e.g. of the C–Hal bond) or through single-electron transfer from metals with suitable redox potentials in their low valence states(*29, 30*) (e.g. zinc(*31, 32*) Zn^0 or indium(*33*) In^0) followed by halide anion loss.

We chose leucine (Leu), an amino acid residue that has a widespread occurrence in proteins, as an initial test sidechain system, which we could potentially generate from the readily available simple organic compound isopropyl iodide as a precursor (Fig. 2a, **R** = iPr). Pleasingly, reaction of dehydroalanine (Dha)-containing derivatives and peptides as small-molecule models under Zn^0-mediated conditions in aqueous buffer (pH 5-6, NH_4Cl (aq)); afforded the corresponding Leu derivatives directly (Fig. S1 and Supplementary Methods). Generation of Dha from cysteine (Cys) followed by conversion to Leu thus allowed the overall 'chemical mutation' of a residue from Cys→Leu inside an intact peptide backbone; under optimized conditions (see Supplementary Methods) this could be achieved in >90% yield and in less than 30 min.

Extension of this methodology to a range of alkyl halides (**R**-Hal) enabled the synthesis of a variety of natural and unnatural amino acids residues, including unnatural aliphatic and cyclic structures (Fig. S2). Notably, not only were primary, secondary and tertiary alkyl halides all tolerated, allowing installation of the natural simple and hydrophobic residue sidechains, so too was the presence of polar protic (e.g. hydroxyl and amine) functionality common in amino acid sidechains. Importantly, the use of these sidechain reagents proved possible even without protection, thus highlighting not only exquisite chemoselectivity but also compatibility with common biological functional groups and hence biological compatibility (orthogonality). Pleasingly, full characterization of the adducts (see Fig. S1 and Supplementary Methods) confirmed absolute regioselectivity (> 98% with d.r. 55:45) for the radical addition, consistent with the designed, matched polarities of the radicals and the corresponding radical acceptor Dha, respectively.(*34*)

We observed that the predicted Cα radical intermediate **1** (Fig. 2b), as well as displaying advantageous stability that would favour initial reaction, was sufficiently long-lived to allow further reaction with other radicals in unproductive and unwanted termination reactions (Fig. S3). For example, reaction with a second alkyl radical afforded di-substituted ('dialkylated') products whilst reaction with molecular oxygen (which in its natively abundant state is the triplet form $^3O_2$ that may react with radicals) led to apparent oxidative protein cleavage (*35*). Notably, additional amounts of reagents were seemingly needed for full conversion due to such observed competing processes (reduction of alkyl halide to alkane as well as disubstitution and oxidative degradation). The successful development of a radical reaction for protein modification (Fig. 2a) therefore necessitated a means of eliminating and controlling these undesired pathways.

Mechanism-guided reaction development allowed optimization of protein-compatible C(sp^3)–C(sp^3) bond-formation

Next, two initial model proteins were selected to test radical reactivity on more complex extended polypeptides: a highly ordered three-layer α/β-Rossman-fold serine protease (subtilisin from *Bacillus lentus* (SBL)) and a three-alpha-helix protein representative of the histone fold that contains both ordered and disordered motifs (histone H3). Dha was installed(*27, 36*) site-selectively from corresponding single cysteine variants (see Supplementary Methods) to generate radical acceptor sites at position 156 in SBL (SBL-Dha156) and at three separate sites (9, 27 and 64) in H3 (H3-Dha9, H3-Dha27, H3-Dha64). These allowed us to test both altered protein scaffold and also variation of reaction site within the same protein scaffold.

Pleasingly, under essentially identical conditions to those used on small-molecule systems (isopropyl iodide, Zn^0 aqueous ammonium acetate buffer pH 6), direct 'chemical mutation' conversion Dha→Leu was observed (see Fig. S4) in both SBL and H3. However, additional side-products were also detected. Careful isolation, trapping and characterization of these side-products (see Figs S5-8) revealed that they were the product of two competing pathways, both of which are consistent with the long-lived intermediacy of the Cα *capto*-dative radical formed after addition of **R•** to Dha. Observed (Figs S5-7) oxidative cleavage products – C-terminal amide **2** and di-carbonyl **3** – would arise from the termination reaction of the Cα radical with triplet oxygen (Fig. 2b).(*37*) Observed di-substituted ('di-alkylated) products **4** arise from the termination reaction of the Cα radical with **R•**. Consistent with these analyses, 'peptide mapping' (tryptic digest-MSMS) confirmed clean site-selectivity of both the desired mutations

and these side reactions: no residues other than that determined by the Dha 'tag' site were identified (Fig. S8). Notably, these results were also wholly consistent with the corresponding Cγ,Cα-di-substituted products observed under comparable conditions from small molecule models (*vide* supra, Fig. S3). These observations in proteins therefore were consistent too with radical addition at Cγ followed by termination of the Cα radical intermediate thus generated (Fig. 2b).

The generality of these parallel processes – *chemical mutation* with competing *oxidative cleavage* and *di-substitution* – was confirmed by its observation in both scaffolds (SBL and H3) and at several sites (e.g. within H3 – K9, K27 and K64). Importantly, these results not only highlighted the need for an improved reaction with better control of radical addition, but together they also provided compelling chemical evidence for the intermediacy of the proposed Cα radical intermediate (Fig. 2b) and hence the (partial and initial) success of our designed C–C radical forming reaction in proteins.

Next, these methodological observations and mechanistic rationalization allowed us, in turn, to optimize the balance between the desired radical sidechain addition and the unwanted competing side reactions (oxidative cleavage and di-substitution). In our initial reaction systems, our use of metal-mediated single-electron transfer exploited a system that is reliant upon the redox potentials of the metals that were used not only in the initiation step but also in the subsequent second electron transfer that creates an enolate that is quenched by protonation (likely from solvent) (see Fig. S9). In principle, more effective second electron transfer to enhance enolate formation would allow more rapid formation of desired product at the expense of side-reactions (quenching of the intermediate Cα radical before side reaction). Survey of redox potentials(*38, 39*) suggested various metal potentials that might prove useful; of those that are compatible with water, indium suggested itself as a strong alternative candidate to zinc (E°/V, $Zn^{(2/0)}$ -0.76**;** $In^{(1/0)}$ -0.14**;** $In^{(2/1)}$ -0.40**;** $In^{(3/2)}$ -0.49). This tuning of the single electron donor, through the use of In^0 instead of Zn^0, proved partially successful yielding cleaner and more effective chemical mutation through radical addition (see Fig. S10 and Supplementary Methods) on several but not all substrates. In particular, certain primary iodides (e.g. the sidechains of MeArg and MeLys) were so reactive that unwanted di-substitution side-products remained.

Next, we considered alternative methods for 'quenching' of the intermediate Cα radical. In principle, direct hydrogen atom ('radical hydride') transfer (Fig. 2b) would not only enhance desired product formation but, by building a suitable chain reaction, could prove more efficient and sustainable. However, analysis of the putative chain cycle (Fig. 2b) highlighted importantly that this would require selective increase of the rate of this step (encompassed by k_{1app}) to a greater extent than that of not only the di-substitution and oxidative side-reactions (k_{3app}; k_{4app}, respectively) but also over that of the direct reduction of alkyl iodide (R–I → R–H, k_{2app}) (Fig. 2b). A range of traditional 'radical hydride' sources were screened (e.g. R_3SnH, R_3SiH, RSH), yet none proved useful and all appeared preferentially to favour k_{2app} over k_{1app}. We reasoned that these bulkier hydride sources preferentially transfer hydride H• to less bulky radicals (such as direct transfer to **R•** thereby favouring k_{2app}) and so we next tested less hindered hydride sources (that might be able to access the more hindered intermediate Cα radical **1**, as desired). Although, borohydrides RBH_3^- are traditionally viewed as nucleophilic hydride "H-" sources, rare studies on these(*40-42*) and related aluminiumhydrides(*43*) have previously suggested possible radical behaviour under certain circumstances.(*44*) Strikingly, we found that $NaBH_4$ in

aqueous solution proved to be a highly effective reagent, allowing improved efficiency (as judged by the need for reduced equivalents of iodides) and by the quality of protein chemical mutation products (see Fig. S11). The radical nature of this controlled, clean and efficient reaction was confirmed not only by direct observation of radicals by electron paramagnetic resonance (EPR) (see Fig. 12) but also through the use of radical trapping; the reaction was fully inhibited by sub-stoichiometric 4-hydroxy-TEMPO or acrylamide (see Fig. 13 and Supplementary Methods).

Finally, having suppressed competing di-substitution we were able to efficiently suppress competing oxidative cleavage simply through the removal of molecular oxygen from the buffer solutions in which we conducted 'chemical mutagenesis' reactions; controlled equilibration experiments at a variety of oxygen partial pressures (see Figs S14,15) revealed that incubation at < 6 ppm O_2 for 6 h prior to reaction proved generally sufficient. Application of the combined optimized conditions led directly to clean 'chemically mutated' proteins without side reactions that, when sequenced by MSMS, were interpreted directly as the intended mutation (Fig. S16). Vitally, both whole protein LC-MS and post-digestion MSMS (Figs 17,18) confirmed that there was no non-selective alkylation or over-alkylation by the halide reagents on other residues in proteins substrates. Additionally, control experiments in which $NaBH_4$ was omitted from the reaction mixture showed that the halide reagents alone did not react with proteins under the standard conditions.

Biocompatible C(sp^3)–C(sp^3) bond-formation enables a 'toolbox' for building natural and unnatural proteins

With optimized conditions for C–C bond formation enabling chemical mutation in hand, we next explored the breadth of both sidechains that could be introduced (and hence mutations that would be accessible) and the protein scaffolds that they could be introduced into (Fig. 3). Representative proteinogenic, modified and non-proteinogenic, polar, non-polar, aromatic, ionized and modified amino acid residues, bearing both natural and unnatural motifs were chosen and all readily incorporated. These importantly demonstrated tolerance of the reaction to many of the most common amino acid (and biological) functional groups: hydroxyl (OH), aminyl (NH), guanidine, amide, thereby highlighting its excellent biocompatibility (orthogonality). Choice of the corresponding organic iodide, typically readily available commercially or through chemical synthesis, allowed wide-ranging and systematic variations. Thus, not only Leu could be incorporated but so could a series of systematic variants of Leu: demethyl-Leu (smaller); *tert*-Leu (bulkier); *nor*-Leu (slimmer); 'cyclo'-Leu (conformationally restricted) (Fig. 3). Similarly, systematic variation of side chain length, methylene unit by methylene unit, was also possible e.g. Ala, ethyl-Gly, demethyl-Leu, *nor*-Leu. We were also able to strategically replace, with atomic precision, methyl groups in residues with their labelled or precisely-altered variants: thus, $CH_3 \rightarrow CF_3$ (e.g. in demethyl-Leu) or $CH_3 \rightarrow {}^{13}CH_3$ (e.g. in trimethyl-Lys). Current methods for isotopic labeling based on 'feeding' experiments result in universal incorporation at every codon-determined site; here, now isotopic labels can be installed at a single site. Such precisely fluorinated or isotopically-labelled amino acids are not only powerful biophysical reporters,(*45-47*) with use particularly in protein (e.g., ^{19}F and ^{13}C) NMR methods, but can also act as modulators of protein structure and binding.(*48*) Indeed, use of ^{19}F NMR allowed us to further confirm both the regioselectivity and stereoselectivity for chemical mutations (Fig. S19), which proved to be essentially identical to that found on peptide models (>98% and d.r. ~1:1). Vitally,

key post-translational modifications proved accessible too: glycosylation(*2*) (in *O*- and *N*- linked form), Lys-methylation(*49*) (in all three states: mono-, di- and tri-, as well as labeled tri-^{13}C), Arg-methylation(*50*) (mono- and di-), Gln-methylation.(*51*) Notably, no other chemical methods exist (barring total protein synthesis) for the installation of the majority of these residues. Moreover, several residues that have been previously biologically inaccessible in proteins were also readily introduced. These included ornithine (Orn), which by virtue of intramolecular cyclization chemistry cannot be loaded onto tRNAs and hence is incompatible with ribosomal incorporation(*52*) and di- / tri-methylated-Lys that cannot yet be incorporated into proteins (even indirectly, as mono-methyl-Lys currently is(*53*)) by cellular stop-codon suppression.(*13*)

Finally, we surveyed the introduction of sidechains to representative examples of protein functions (structural, channels, enzymes, glycoproteins) from differing protein folds (with varying levels of α, β and unstructured secondary motifs) and species types. Thus, as well as SBL and H3, we also surveyed the variously structured histone protein H4; the transmembrane bacterial efflux component protein AcrA(*54*); p38α mammalian mitogen-activated MAP kinase(*55*); mammalian antibody cAbLys3(*56*); apoptosis marker binding protein annexinV(*57*); and pentapeptide-repeat protein Npβ.(*58*) These also represent proteins that are variously associated with localization in different cellular environments (nuclear (H3, H4), cytosolic (p38α, annexinV), transmembrane (AcrA, Npβ), extracellular (cAbLys3, annexinV)). They also allowed us to survey highly diverse architectures: α-helix-coiled-coil (AcrA); histone fold (mixed unfolded and α-helix, H3, H4); mixed-α/β-fold (SBL, p38α); variable domain immunoglobulin fold (4-strand-β-sheet plus 5-strand-β-sheet, cAbLys); α-helix-rich globular annexin fold (four × 5-α-helix domains in a 'super-helix', annexinV); and even a right-handed quadrilateral β-helix (Npβ). Sites of modifications sat within various feature types including helix, sheets and loose loops (see Table S2). Notably, all proteins proved compatible with sidechain attachment at all targeted sites (see Supplementary Methods for full characterization details). Multiple sites in the same proteins were also surveyed: five different sites in H3 (sites 9, 10, 26, 27, and 64) and two in H4 (sites 16 and 17). Together, these cumulative variations of sidechains in different protein substrates at different sites allowed access to >50 'chemical mutants'. Analysis of sequence (e.g., by tryptic-MSMS peptide mapping, see Fig. S20), structure (e.g, secondary structural content by circular dichroism, amide- and aliphatic-resonance analysis and diffusional analysis by protein ^{1}H-NMR, see Figs S21-23) and retention of function (e.g. enzymatic activity, Ab-binding function and biomarker recognition, protein complex assembly), confirmed not only the site-selectivity of the C–C-bond-forming chemical mutation but also) suggested no gross alteration of global structure (see Supplementary Methods), although it should be noted that localized structural changes cannot be discounted. Notably, too, test experiments in both model systems and proteins containing disulfides suggested good compatibility of reagents (including DBHDA, consistent with prior results,(*27*) and $NaBH_4$, Supplementary Methods and Figs S24,25) with disulfides.

C(sp^3)–C(sp^3) bond-formation Chemical Mutation enables diverse techniques for the study of protein methylation, glycosylation and phosphorylation

With this ability to directly insert sidechain alterations and hence perform chemical mutagenesis on proteins with wide variation in protein and site, we chose to test differing proof-of-principle strategies that would allow insight into the biological function of post-translationally

modified residues (and their mimics) that are ordinarily difficult to install with fidelity into proteins by other means. To this end we chose three of the most important PTMs: glycosylation, phosphorylation and methylation.

Glycosylation is the most diverse of the post-translational modifications,(*1, 59*) yet has been prominently absent as a readily accessible motif in proteins via chemical or genetic mutagenesis methods.(*60*) Until now no general chemical method for convergently installing *N*- and *O*-linked glycans has been possible.(*59, 60*) We used C–C bond forming mutagenesis along with corresponding (entirely unprotected) *N*- and *O*-linked glycosidic iodides to install *N*-acetylglucosamine (GlcNAc), a glycan which is found naturally in both *N*- and *O*-linked form, to create an unnatural glycosylation site at position 27 of H3 (using the same, common divergent H3-Dha27 protein intermediate, Figure 4). Despite the unnatural constitution of this site, enzymatic extension with either glycosyltransferase or endoglycosidase allowed the overall installation of more complex glycans onto both *N*- and *O*-linked GlcNAc; even up to the *N*-linked core pentasaccharide that is found in all *N*-linked glycoproteins.(*1, 59*) We discovered that despite an apparent plasticity with respect to protein scaffold(*61*), widely-used(*62*) *N*-glycosidase PNGase did not cleave synthetic variants with extended side-chain length (Fig. 4a and Fig. S26). In striking contrast, we discovered that a variety of synthetic *O*-GlcNAc-ylated glycoproteins were cleaved quite readily by *O*-GlcNAc-ases from different sources (Fig. 2b and Fig. S27), including the human *O*-GlcNAc-ase (hOGA) enzyme. The latter, which is implicated in diabetes,(*63*) dementia(*64*) and cancer(*65*) has, until now, been presumed selective given that it is the sole encoded protein *O*-GlcNAc-ase in the human genome;(*66*) our results suggest a previously unappreciated and surprising plasticity. In addition to the *N*- and *O*-glycosylation of H3, we were able to chemically *N*-glycosylate and *O*-glycosylate other sites and proteins, including the naturally glycosylated protein AcrA, the pentapeptide repeat protein Nβ and notably the heavy-chain antibody cAbLys with the putative Fcγ-receptor ligand glycan Man3GlcNAc2 (Fig. 4c and Figure S28).

Aurora B kinase, which is overexpressed in cancers,(*67*) phosphorylates Ser10 of histone H3 during mitosis(*68*) and is hence implicated as a controlling factor in cell division and proper distribution of genetic information. The lability of phosphorylation and the mixtures of phospho-proteins often formed from natural enzymatic phosphorylation greatly complicates the study of phosphoryl groups on given sites, such as H3-pSer10.(*69*) We used C–C bond forming mutagenesis along with corresponding (notably unprotected) iodophosphonates to create stable analogues of H3-pSer10 in which a single oxygen atom was replaced by methylene or difluoromethylene units to create *carba*-phosphoSer variants *c*pSer(*70*) and *cf2*pSer(*71*) (Figure 5a); MS analysis revealed that this could be achieved with a purity that is not possible with current biological methods for phosphorylating H3 (Fig. 5b). Antibodies and appropriate binding proteins (such as the 'reader' protein MORC3(*72*)) not only recognized the resulting phosphomimic proteins (Figs 5c,f and Fig. S29), but proved stable to either chemical or enzyme-catalyzed dephosphorylation (Fig. 5d,e and Fig. S30), even with phosphatases that readily processed naturally phosphorylated H3 (e.g. protein phosphatases 1 or 2A). It also proved possible to readily install *carba*-pSer into other proteins (Fig. S31). Whilst *c*pSer10 proved a functioning mimic of pSer10, it has been argued(*73, 74*) that fluorophosphonates may act as more effective phosphate mimics by virtue of pKa,(*71*) polarity and shape.(*75*) The C–C bond-forming mutagenesis allowed us to also install a difluorophosphonate-Ser variant (difluoro-*carba*-pSer, *cf2*pSer) at the same site of H3 to create H3-*cf2*pSer10. Consistent with improved

mimicry, this variant showed enhanced binding to MORC3, validating proposed(*71, 73-75*) difluorophosphonate mimicry of phosphates in proteins (see Fig. 5f).

Protein methylation(*76*) is a central biological process (e.g. in epigenetic regulation and cell signal transduction). Yet, the precise elucidation of the functional mechanistic role of methylation at the molecular level and the ready delineation of proteins associated with the 'methylproteome' (e.g. 'writers'-'readers'-'erasers') remains a grand challenge in biological science. We chose two methylated protein targets that have not been possible to create through other protein generation methods and created these through the site-selective C–C bond forming mutagenesis reaction: a site-selectively ^{13}C-labelled variant of tri-methylated-Lys in histone protein H3 (Fig. 6a) and a di-methylated-Arg residue site-selectively installed into in an intact nucleosomal particle (Fig. 6b). ^{13}C-labelled-H3 H3-[^{13}C-Me]$_3$-Lys9 was created (Fig. 6a) with precise tri-methylation on a Lys residue (Lys9) that has been previously been observed and implicated in direct transcriptional regulation.(*77, 78*) The structure and function of this chemical mutant of H3 was explored by protein MS and NMR. These revealed ready observation of the ^{13}C-labelled protein, by virtue of its isotopic label, and the readily observable processing and release of the ^{13}C-labeled methyl groups (^{13}C-Me) from [^{13}C-Me]$_3$-Lys9 in H3 by the known demethylation enzyme JMJD2A/KDM4a (Fig. 6a and Fig. S32a).(*79, 80*)

Having installed and explored precise Lys-methylation in an isolated histone protein, we explored precise Lys- and Arg-methylation in the context of an entire, intact, nucleosomal particle (Fig. 6a,b); both Lysme3 and Argme2a were readily introduced. Both anti-Lysme3 antibody recognition and JMJD2a/KDM4a-catalyzed demethylation were readily demonstrated using synthetically methylated nucleosome (Fig. S32b). Arg-methylation, and especially asymmetric-dimethylation (Argme2a), remains only a partially understood alteration. Moreover, given the variant isomeric forms of Arg methylation and dimethylation, precise control of the installation of methylation to create representatively methylated proteins has also not proved possible. Using the C–C chemical mutagenesis method (Fig. 6b) we precisely installed Argme2a into site 26 of H3 (H3-Arg26me$_2$a) (a site implicated(*81*) in so-called 'crosstalk' epigenetic modifications with an order that is not clearly understood) in intact nucleosomal particles.

These synthetic nucleosomal probes bearing asymmetric-dimethylation at H3-Arg26me$_2$a allowed the identification of key partner proteins through affinity enrichment proteomics in human cells (Fig. 6b and Table S1). Notably, two of the strongest interacting partners, BEND3 and BANP, contain BEN domains, a recently characterized α-helical module found in chromatin-associated proteins(*82*); BEND3,(*83*), a novel rDNA transcription repressor, is the most enriched interacting protein partner. Interestingly, BEND3 can recruit PRC2 complex that promotes H3K27Me3 modification, a known transcription repression marker.(*84*); this suggests potential cross-talk between these two modifications leading to generation of a repressive chromatin state. Intriguingly, a majority of the remaining significant interactors (SMARCAL1, RECQL, DDB1, DDB2, TOP3A) are annotated as being involved in DNA replication/repair. Although a link between histone arginine methylation and DNA repair has not been previously reported, the results of this experiment suggest increased accessibility of nucleosome-bound DNA to a range of relevant DNA-binding effectors; such a loss of DNA-to-arginine hydrogen bonding would be anticipated upon methylation. Such is the flexibility of the C–C chemical mutagenesis method, that it also proved possible to readily install methylated Lys (Lysme, Lysme$_2$, Lysme$_3$), methylated Arg (Argme$_2$a, Argme$_1$) and methylated Gln into a wide range of other sites and proteins (Fig. S33). Based on the ready discovery of previously unanticipated interacting partners

for Arg26me$_2$a (see above) we anticipate that these too will prove to be powerfully precise probes of methylation function and direct 'methylproteome' interactions.

Finally, we tested the scalability of the C–C chemical mutagenesis method; applications such as translational trials of biologics and structural biological studies can necessitate larger (multimilligram) amounts of proteins. Pleasingly, reaction of 10 milligrams of Dha-protein H3S10Dha with the relevant bromide and sodium borohydride allowed the installation of difluoromethylenephosphohomoalanine (*cf2*pSer) in useful 'synthetic' (isolated, 65-70%) yield after desalting thereby suggesting utility even for larger-scale synthetic applications in protein chemistry (Fig. S34).

Discussion

The compatibility in the use of radicals as effective and yet benign reactive intermediates for protein modification suggests that other radical-based methods(*35*) may prove powerful in the field of protein chemistry. Indeed, the key implicated propagating intermediate in our C–C bond forming mutagenesis, the *capto*-datively stabilized Cα backbone radical, is similar to other stabilized radicals suggested in natural processes(*35*) and it may well be that not only has nature long been taking advantage of such methods but that other routes of access to such intermediates could allow similarly powerful bond-forming strategies. This protein radical chemistry is likely to require new reagents such as, boron-based compounds, which have played an important role in radical chemistry(*85, 86*). The associated mechanisms of these reagents remain a topic of active debate(*43, 44*) and mechanisms other than the one we propose here cannot be discounted.

Given their benign application and compatibility we can also envisage the ready combination of the C–C mutagenesis reaction with other protein chemistries or assembly methods. For example, thioester-mediated backbone assembly methods ('native chemical ligation')(*15*) typically utilize peptide fragments with *N*-terminal Cys residues that remain at the junction point after 'ligation'. Using the C–C mutagenesis reaction these Cys residues could be readily converted to Dha and thence to almost any residue of choice. Indeed, in a proof-of-principle of such a combined strategy we have been able to generate Thr–Leu containing peptides in which the Thr–Leu moiety is derived from thioester-mediated amide ligation with Cys followed by C–C chemical mutagenesis to Leu (see Fig. S35 and Supplementary Methods).

Because the simple choice of reagent allows for the variation of single atom substituents (e.g. O $\rightarrow$ CH$_2$ $\rightarrow$ CF$_2$ in pSer, *c*pSer and *cf2*pSer, as we have shown here) our approach may enable a more molecular understanding of mechanism of proteins and a fine-tuning of function. The use of stable phosphoryl mimics (such as *c*pSer or *cf2*pSer) suggests itself as a promising way to 'fish' not only for phospho-binders ('phospho-readers', as we have shown here) but even enzyme partners that would process their natural modification counterparts (e.g. phosphatases that would cleave pSer but, as we have shown here cannot cleave *c*pSer). It should be noted in this context that whilst amber-codon suppression ('genetic code expansion') methods have proven highly powerful in certain cases, many useful and biologically relevant structures elude current approaches. For example, the residue types (^{13}C-trimethyl-Lys, dimethyl-Arg, *O*- and *N*-linked glycosylated residues, *difluorocarba*- or *carba*-pSer) installed in the proof-of-principle studies (see above) have proven intractable to direct installation into biologically relevant sites by amber-codon suppression.(*87*) Indeed, residues bearing only small differences from their natural counterparts (methylated Lys, Arg, Gln), which we were able to incorporate chemically,

sometimes prove difficult to incorporate by such genetic methods due to their strong structural resemblance in the translational machinery to their unmodified counterparts.

It is important to note that although the C–C bond-forming mutagenesis method described here allows the creation of the relevant constitution (connectivity) found in proteins, our analysis of the stereoselectivity in both peptides and proteins (see above) suggests approximately equal formation of both D- / L- configuration at the mutation site. This would result in yet more complex mixtures of protein diastereomers for multiple mutation sites.

The improved access to relevant protein architectures will potentially reveal surprising biological functions. For instance, our data have revealed that human enzyme hOGA more plastic in its cleavage activity of *O*-GlcNAc-ylated proteins than had been previously anticipated. Coupled with inferred plasticity of the corresponding glycosyltransferase (OGT) that was recently suggested by studies on peptides,(*88*) this suggests that any *O*-GlcNAc-ylation-associated regulatory mechanism may be much looser than previously realized. The elucidation of new binding partners for nucleosomes in previously inaccessible methylation states (e.g. asymmetrically dimethylated Argme2a) suggests that many other new interactions may be discoverable via C–C chemical mutagenesis. These, in turn, will then allow us to identify and synthetically 'program' into proteins exactly those residues that engender wanted functional (e.g. pharmaceutical) benefit in a truly broad manner. For instance, the chemical glycosylation (via C–C bond mutagenesis) of an Ab fragment with possible sugar ligands for the Fcγ-receptor raises the possibility of new cell killing strategies (*89*) mediated by synthetic Ab-fragments. In this way, we envisage access to synthetic proteins that will allow application of newly elucidated functions in, for example, new protein drugs or 'synthetic biologics'.(*90*)

General Methods

General Dha Formation on Protein Substrate. The relevant protein cysteine mutant (10 mg of lyophilized powder) was dissolved in 500 μL of reaction buffer (3M guanidinium hydrochloride, 100 mM sodium phosphate, pH 8.0) and 30 mg DTT was added. The resultant protein solution was shaken at room temperature for 30 minutes. Excess DTT was removed by passing the solution through a PD MiniTrap G25 (equilibrated with the same buffer). Eluting with 1 mL of the reaction buffer gave the protein at a final concentration of 10 mg/mL. To the reduced protein mutant was added 50 μL of a 2,5-dibromohexane diacetamide (DBHDA) stock solution (0.5 mM, 100 mg in 610 μL of DMF). The reaction was gently shaken for 45 minutes at room temperature and then 90 minutes at 37 °C. The reaction mixture was concentrated to 500 μL using a Vivaspin 500 concentrator (MWCO 10 kDa) and excess reagent was removed by passing the solution through a PD MiniTrap G25 (equilibrated with chemical mutagenesis reaction buffer). Eluting with 1 mL of the same buffer provided the dehydroalanine-containing protein.

General Chemical Mutagenesis Method. A 2 mL eppendorf tube containing a solution of protein (typically 0.2- 5 mg/mL, 200 μL, ~ 10-200 μM) in reaction buffer (pH 4-8, most common buffers tolerated) was ported into a N_2 atmosphere glovebox (Belle Technologies) and equilibrated to anaerobic conditions (cap open) overnight at 4°C in a standard benchtop eppendorf cooler. In parallel, a stock solution of the bromide or iodide of interest (prepared by dissolving solids in water and adjusting the pH to that of the reaction buffer), or the solid reagent itself, was also

ported into the glovebox. In the case of liquid reagents, either a stock solution was prepared as above, or the neat reagent was ported into the glovebox and stored at the appropriate temperature. Solid sodium borohydride (0.3 – 1 mg, smaller amounts of sodium borohydride can be difficult to weigh accurately) was ported into the glovebox in a 2 mL eppendorf tube and placed in the cooler kept at 4°C. Following a minimum of 8 hours degassing, inside the glovebox, the volume of the protein solution was ascertained by pipette and 'topped up' with reaction buffer to the desired volume (e.g. 200 μL). 1-20 μL of the previously prepared iodide or bromide stock solution (typically 100 – 2000 molar equivalents relative to protein) was then pipetted into the protein solution, and the resultant solution gently mixed by pipette. The entire volume (~200 μL) of the protein solution was then added to the 2 mL eppendorf tube kept at 4°C and containing solid NaBH4. The tube was gently shaken by hand, causing effervescence/ foaming to spread through the solution. Any discolouration from the iodide reagent should rapidly disappear (In some cases, particularly at high pH where the hydrolysis of borohydride is retarded, sustained effervescence is observed, whereas for most iodides a colourless and still solution is obtained after ~20 minutes incubation at 4°C). The reaction was incubated at 4°C for 30 minutes (cap open) before being removed from the glovebox. Shaking is not necessary but may assist in removing evolved gases. The eppendorf tube was then capped and removed from the glovebox. Outside the glovebox, the reaction solution was loaded onto SpinTrap™ G-25 desalting column, pre-equilibrated with the desired storage buffer. Elution according to manufacturers instructions afforded the desired chemical mutagenesis product, which was analysed by LC-MS. For downstream applications, an additional SEC may be necessary for sample desalting.

References and Notes:

1. C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, Jr., Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342-7372 (2005).
2. K. Ohtsubo, J. D. Marth, Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).
3. P. Cohen, The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-E130 (2002).
4. E. I. Campos, D. Reinberg, Histones: Annotating Chromatin. *Ann. Rev. Genet.* **43**, 559-599 (2009).
5. T. H. Wright, M. R. J. Vallée, B. G. Davis, From Chemical Mutagenesis to Post-Expression Mutagenesis – a 50 Year Odyssey. *Angew. Chem. Int. Ed.* **55**, 5896-5903 (2016).
6. C. Zioudrou, M. Wilchek, A. Patchornik, Conversion of the L-Serine Residue to an L-Cysteine Residue in Peptides. *Biochemistry* **4**, 1811-1822 (1965).
7. K. E. Neet, D. E. Koshland, Jr., The conversion of serine at the active site of subtilisin to cysteine: a "chemical mutation". *Proc. Natl. Acad. Sci. USA* **56**, 1606-1611 (1966).
8. C. A. Hutchison, III *et al.*, Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551-6560 (1978).
9. J. A. Brannigan, A. J. Wilkinson, Protein engineering 20 years on. *Nat Rev Mol Cell Biol* **3**, 964-970 (2002).
10. C. R. Woese, G. J. Olsen, M. Ibba, D. Soll, Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* **64**, 202-236 (2000).
11. L. Wang, P. G. Schultz, Expanding the Genetic Code. *Angew. Chem. Int. Ed.* **44**, 34-66 (2005).
12. J. W. Chin, Expanding and Reprogramming the Genetic Code of Cells and Animals. *Ann. Rev. Biochem.* **83**, 379-408 (2014).
13. A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **6**, 50-69 (2015).
14. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182-188 (1989).
15. P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779 (1994).
16. L. R. Malins, R. J. Payne, Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr. Opin. Chem. Biol.* **22**, 70-78 (2014).
17. J. M. Chalker, B. G. Davis, Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. *Curr Opin Chem Biol* **14**, 781-789 (2010).

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

19. E. M. Sletten, C. R. Bertozzi, Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974-6998 (2009).

20. C. D. Spicer, B. G. Davis, Selective chemical protein modification. *Nat Commun* **5**, (2014).

21. L. Polgár, M. L. Bender, Simulated mutation at the active site of biologically active proteins. *Advances in Enzymology and Related Areas of Molecular Biology*, 381-400 (1970).

22. T. Perchyonok, *Radical Reactions in Aqueous Media*. (Royal Society of Chemistry, Cambridge, 2009).

23. J. Wang, S. M. Schiller, P. G. Schultz, A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed.* **46**, 6849-6851 (2007).

24. F. P. Seebeck, J. W. Szostak, Ribosomal synthesis of dehydroalanine-containing peptides. *J Am Chem Soc* **128**, 7150-7151 (2006).

25. Y. O. You, M. R. Levengood, L. A. F. Ihnken, A. K. Knowlton, W. A. van der Donk, Lacticin 481 Synthetase as a General Serine/Threonine Kinase. *ACS Chem. Biol.* **4**, 379-385 (2009).

26. G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, 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).

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

28. B. Giese, *Radicals in Organic Synthesis: Formation of Carbon-Carbon Bonds*. J. E. Baldwin, Ed., (Pergamon Press, Oxford, 1986).

29. C. J. Li, Organic reactions in aqueous media - with a focus on carbon-carbon bond formation. *Chemical Reviews* **93**, 2023-2035 (1993).

30. C.-J. Li, Organic reactions in aqueous media with a focus on carbon-carbon bond formations: A decade update. *Chem. Rev.* **105**, 3095-3166 (2005).

31. C. Petrier, C. Dupuy, J. L. Luche, Conjugate additions to α,β-unsaturated carbonyl compounds in aqueous media. *Tetrahedron Lett.* **27**, 3149-3152 (1986).

32. T. Huang, C. C. K. Keh, C.-J. Li, Synthesis of [small alpha]-amino acid derivatives and amines via activation of simple alkyl halides by zinc in water. *Chem. Commun.*, 2440-2441 (2002).

33. H. Miyabe, M. Ueda, A. Nishimura, T. Naito, Indium-mediated intermolecular alkyl radical addition to electron-deficient CN bond and CC bond in water. *Org. Lett.* **4**, 131-134 (2002).

34. J. M. Tedder, Which factors determine the reactivity and regioselectivity of free radical substitution and addition reactions? *Angew. Chem. Int. Ed. Engl.* **21**, 401-410 (1982).

35. C. L. Hawkins, M. J. Davies, Generation and propagation of radical reactions on proteins. *Biochim Biophys Acta* **1504**, 196-219 (2001).

36. J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chem. Int. Ed.* **51**, 1835-1839 (2012).

37. W. M. Garrison, Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chem. Rev.* **87**, 381-398 (1987).

38. G. Milazzo, S. Caroli, V. K. Sharma, *Tables of Standard Electrode Potentials*. (Wiley, Chichester, 1978).

39. A. J. Bard, R. Parsons, J. Jordan, *Standard Potentials in Aqueous Solutions*. (Marcel Dekker, New York, 1985).

40. J. T. Groves, K. W. Ma, Dehalogenations with sodium borohydride. Evidence for a free radical reaction. *J. Am. Chem. Soc.* **96**, 6527-6529 (1974).

41. E. J. Corey, J. W. Suggs, Method for catalytic dehalogenations via trialkyltin hydrides. *J. Org. Chem.* **40**, 2554-2555 (1975).

42. G. Stork, N. H. Baine, Cyclization of vinyl radicals: a versatile method for the construction of five- and six-membered rings. *J. Am. Chem. Soc.* **104**, 2321-2323 (1982).

43. E. C. Ashby, R. N. DePriest, A. B. Goel, B. Wenderoth, T. N. Pham, Occurrence of electron transfer in the reduction of organic halides by lithium aluminum hydride and aluminum hydride. *J. Org. Chem.* **49**, 3545-3556 (1984).

44. J. A. Barltrop, D. Bradbury, Chain photoreaction of sodium borohydride with halogenated aromatic hydrocarbons. Initiation by aryl radicals. *J. Am. Chem. Soc.* **95**, 5085-5086 (1973).

45. J. J. Liu, R. Horst, V. Katritch, R. C. Stevens, K. Wüthrich, Biased signaling pathways in β2-adrenergic receptor characterized by 19F-NMR. *Science* **335**, 1106-1110 (2012).

46. G. S. Kubelka, J. Kubelka, Site-specific thermodynamic stability and unfolding of a de novo designed protein structural motif mapped by 13C isotopically edited IR spectroscopy. *J. Am. Chem. Soc.* **136**, 6037-6048 (2014).

47. V. S. Lelyveld, A. Björkbom, E. M. Ransey, P. Sliz, J. W. Szostak, Pinpointing RNA-Protein Cross-Links with Site-Specific Stable Isotope-Labeled Oligonucleotides. *J. Am. Chem. Soc.* **137**, 15378-15381 (2015).

48. N. C. Yoder, K. Kumar, Fluorinated amino acids in protein design and engineering. *Chem. Soc. Rev.* **31**, 335-341 (2002).

49. C. Martin, Y. Zhang, The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **6**, 838-849 (2005).

50. M. T. Bedford, S. G. Clarke, Protein arginine methylation in mammals: who, what, and why. *Molecular Cell* **33**, 1-13 (2009).

51. P. Tessarz *et al.*, Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505**, 564-568 (2014).

52. H. Jakubowski, Misacylation of tRNALys with noncognate amino acids by lysyl-tRNA synthetase. *Biochemistry* **38**, 8088-8093 (1999).

53. D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, Genetically Encoding N ϵ-Methyl-l-lysine in Recombinant Histones. *J. Am. Chem. Soc.* **131**, 14194-14195 (2009).

54. H. I. Zgurskaya, H. Nikaido, AcrA is a highly asymmetric protein capable of spanning the periplasm. *J Mol Biol* **285**, 409-420 (1999).

55. J. C. Lee *et al.*, A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature* **372**, 739-746 (1994).

56. A. Desmyter *et al.*, Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat Struct Mol Biol* **3**, 803-811 (1996).

57. C. E. Creutz, The annexins and exocytosis. *Science* **258**, 924-931 (1992).

58. M. W. Vetting, S. S. Hegde, K. Z. Hazleton, J. S. Blanchard, Structural characterization of the fusion of two pentapeptide repeat proteins, Np275 and Np276, from Nostoc punctiforme: Resurrection of an ancestral protein. *Protein Sci.* **16**, 755-760 (2007).

59. D. P. Gamblin, E. M. Scanlan, B. G. Davis, Glycoprotein Synthesis: An Update. *Chem. Rev.* **109**, 131-163 (2009).

60. L.-X. Wang, B. G. Davis, Realizing the promise of chemical glycobiology. *Chem. Sci.* **4**, 3381-3394 (2013).

61. F. Altmann, S. Schweiszer, C. Weber, Kinetic comparison of peptide: N-glycosidases F and A reveals several differences in substrate specificity. *Glycoconjugate J.* **12**, 84-93 (1995).

62. A. L. Tarentino, C. M. Gomez, T. H. Plummer Jr, Deglycosylation of asparagine-linked glycans by peptide: N-glycosidase F. *Biochemistry* **24**, 4665-4671 (1985).

63. S. Hardivillé, Gerald W. Hart, Nutrient Regulation of Signaling, Transcription, and Cell Physiology by O-GlcNAcylation. *Cell Metabolism* **20**, 208-213 (2014).

64. Y. Zhu, X. Shan, S. A. Yuzwa, D. J. Vocadlo, The emerging link between O-GlcNAc and Alzheimer disease. *J Biol Chem* **289**, 34472-34481 (2014).

65. C. Slawson, G. W. Hart, O-GlcNAc signalling: implications for cancer cell biology. *Nat Rev Cancer* **11**, 678-684 (2011).

66. R. Hurtado-Guerrero, H. C. Dorfmueller, D. M. F. van Aalten, Molecular mechanisms of O-GlcNAcylation. *Curr. Opin. Struct. Biol.* **18**, 551-557 (2008).

67. J. R. Bischoff *et al.*, A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J.* **17**, 3052-3065 (1998).

68. J.-Y. Hsu *et al.*, Mitotic Phosphorylation of Histone H3 Is Governed by Ipl1/aurora Kinase and Glc7/PP1 Phosphatase in Budding Yeast and Nematodes. *Cell* **102**, 279-291 (2000).

69. R. Klingberg *et al.*, Analysis of Phosphorylation-Dependent Protein–Protein Interactions of Histone H3. *ACS Chem. Biol.* **10**, 138-145 (2014).

70. D. T. Rogerson *et al.*, Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat Chem Biol* **11**, 496-503 (2015).

71. W. Zheng *et al.*, Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *J Biol Chem* **280**, 10462-10467 (2005).

72. N. Kunowska, M. Rotival, L. Yu, J. Choudhary, N. Dillon, Identification of protein complexes that bind to histone H3 combinatorial modifications using super-SILAC and weighted correlation network analysis. *Nucl. Acids Res.*, gku1350 (2015).

73. G. M. Blackburn, Phosphonates as analogues of biological phosphates. *Chem. Ind.* , 134-138 (1981).

74. G. M. Blackburn, D. E. Kent, F. Kolkmann, The synthesis and metal binding characteristics of novel, isopolar phosphonate analogues of nucleotides. *J. Chem. Soc., Perkin Trans. 1*, 1119-1125 (1984).

75. D. J. Burton, Z.-Y. Yang, W. Qiu, Fluorinated Ylides and Related Compounds. *Chem. Rev.* **96**, 1641-1715 (1996).

76. K. K. Biggar, S. S. Li, Non-histone protein methylation as a regulator of cellular signalling and function. *Nat Rev Mol Cell Biol* **16**, 5-17 (2015).

77. W. Fischle *et al.*, Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870-1881 (2003).

78. C. R. Vakoc, S. A. Mandat, B. A. Olenchock, G. A. Blobel, Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Molecular Cell* **19**, 381-391 (2005).

79. J.-F. Couture, E. Collazo, P. A. Ortiz-Tello, J. S. Brunzelle, R. C. Trievel, Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14**, 689-695 (2007).

80. S. S. Ng *et al.*, Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87-91 (2007).

81. S. Daujat *et al.*, Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Current biology : CB* **12**, 2090-2097 (2002).

82. S. Abhiman, L. M. Iyer, L. Aravind, BEN: a novel domain in chromatin factors and DNA viral proteins. *Bioinformatics* **24**, 458-461 (2008).

83. A. Khan *et al.*, BEND3 represses rDNA transcription by stabilizing a NoRC component via USP21 deubiquitinase. *Proc. Natl. Acad. Sci. USA* **112**, 8338-8343 (2015).

84. N. Saksouk *et al.*, Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. *Molecular Cell* **56**, 580-594 (2014).

85. C. Ollivier, P. Renaud, Organoboranes as a Source of Radicals. *Chem. Rev.* **101**, 3415-3434 (2001).

86. T. Kawamoto, I. Ryu, Radical reactions of borohydrides. *Organic & Biomolecular Chemistry* **12**, 9733-9742 (2014).

87. A. K. Antonczak, Z. Simova, E. M. Tippmann, A Critical Examination of Escherichia coli Esterase Activity. *J. Biol. Chem.* **284**, 28795-28800 (2009).

88. C. Lizak *et al.*, Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat Commun* **4**, (2013).

89. M. Kurogochi *et al.*, Glycoengineered Monoclonal Antibodies with Homogeneous Glycan (M3, G0, G2, and A2) Using a Chemoenzymatic Approach Have Different Affinities for Fc?RIIIa and Variable Antibody-Dependent Cellular Cytotoxicity Activities. *PLoS ONE* **10**, e0132848 (2015).

90. B. G. Davis, in *Biotherapeutics: Recent Developments using Chemical Biology*, L. H. Jones, A. J. McKnight, Eds. (Royal Society of Chemistry, Cambridge, 2013), pp. 130-144.

91. R. Fischer, B. M. Kessler, Gel-aided sample preparation (GASP)--a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics* **15**, 1224-1229 (2015).

92. D. Wessel, U. I. Flugge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* **138**, 141-143 (1984).

93. J. Zhang *et al.*, PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & cellular proteomics : MCP* **11**, M111.010587 (2012).

94. S.-W. Wu, T.-H. Pu, R. Viner, K.-H. Khoo, Novel LC-MS2 Product Dependent Parallel Data Acquisition Function and Data Analysis Workflow for Sequencing and Identification of Intact Glycopeptides. *Analytical Chemistry* **86**, 5478-5486 (2014).

95. J. Saba, S. Dutta, E. Hemenway, R. Viner, Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-Product-Dependent Electron Transfer Dissociation. *International Journal of Proteomics* **2012**, 7 (2012).

96. K. P. Chooi *et al.*, Synthetic phosphorylation of p38alpha recapitulates protein kinase activity. *J Am Chem Soc* **136**, 1698-1701 (2014).

97. S. Stoll, A. Schweiger, EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson* **178**, 42-55 (2006).

98. L. Lercher *et al.*, Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nat Commun* **6**, 7978 (2015).

99. T. Bartke *et al.*, Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell* **143**, 470-484 (2010).

100. J. R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).

101. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* **26**, 1367-1372 (2008).

102. J. Cox *et al.*, Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805 (2011).

103. J. Cox *et al.*, Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics* **13**, 2513-2526 (2014).

104. H. D. Lim *et al.*, Discovery of S-(2-Guanidylethyl)-isothiourea (VUF 8430) as a Potent Nonimidazole Histamine H4 Receptor Agonist. *J. Med. Chem.* **49**, 6650-6651 (2006).

Acknowledgments: We thank the Felix Foundation (RR), the Croucher Foundation (WLN), the Rutherford Foundation (THW, SF), Cancer Research UK (SF, LL), BBSRC (SRGG, BJB, AJB), AstraZeneca (SRGG), UCB (BJB), EPSRC (ODC, LL, MWS, MEE), Wellcome Trust (NMR equipment) for funding. BGD was in receipt of a Royal Society Wolfson Research Merit Award during part of this project. BGD is a member of the Catalent Biologics Scientific Advisory Board. A pET12a PAPI plasmid encoding the gene for annexin V (originally deposited by J. Tait) was obtained from Addgene under a material transfer agreement. We thank W. Myers, T. Parsons, J. Wickens, H. Hailu, P. Mori, A. Henry, K. Chooi, T. Fleming for useful discussions and / or technical assistance. BGD, JMC, GJB are inventors on patent WO 2009103941 filed by the University of assigned to Catalent Biologics that covers in one claim reaction of the carbon-carbon double bond of dehydroalanine residues reacted with an organohalide in the presence of elemental zinc.

Data reported in the paper are presented in the Supplementary Materials (see pdf and associated xlsx spreadsheet detailing affinity proteomic data Table S3). Associated data are also archived in the University of Oxford data archive (ORA-data) with accession number doi: 10.5287/bodleian:dp1ypwOB6.

Author Contributions: THW, BJB, JMC, GJB, RW, WLN, SF, MRJV, AP, ODC, MEE, SRGG, LL, MWS, conceived, designed and/or performed protein mutagenesis experiments. THW, BJB, JMC, SG performed model experiments. THW, BJB, WLN, MRJV, RR, MWS prepared proteins. THW, BJB, WLN, TDWC, RR, MRJV, MEE, M-LT, AJB, BK physically characterized protein products. THW, BJB, WLN, SF, MK, RR, SM biochemically characterized protein products. THW, BJB, JMC, GJB, RW, WLN, RR, M-LT, AJB, BK, SM, BGD analyzed data. THW, JMC, BGD wrote the paper. All authors read and commented on the paper.

Figure Captions

Figure 1. Design of a Bio-compatible C(sp^3)–C(sp^3) bond-forming reaction as part of a General Strategy for Chemical Mutagenesis. (a) Natural post-translational modification exploits C-heteroatom bond formation, such as C–N bond formation in Lysine (Lys) methylation, to diversify protein structure and hence modulate function. **(b)** Modification in a protein after translation, akin to PTM formation, but using C–C bond formation would allow *construction* of many sidechains not just the *modification* of existing natural amino acid residues. Retrosynthetic analysis (blue arrow) suggests a C–C β,γ bond disconnection and reveals several possible 'synthons'. For heterolytic two-electron reactions (shown in red) the resulting disconnection yields 'synthetic equivalents' **R+** or **R-** or reagents that would react and/or quench with the water in the necessary aqueous solvent or with example protein side chains shown. However, for the homolytic one-electron reaction (shown in blue) the resulting disconnection yields a free radical 'synthon' that would allow an equivalent **R•** for which compatibility can be envisaged. **(c)** A functional group with affinity for carbon-centred free radicals that was suitably polarized could act as the other 'synthetic equivalent' corresponding to the other protein-based 'synthon' arising from this homolytic disconnection. The residue dehydroalanine (Dha) is the functional group proposed as a radical reactive 'synthetic equivalent' that would allow corresponding C–C β,γ bond-formation.

Figure 2. Proposed Radical Reaction to Allow Post-translational Mutagenesis and Suggested Mechanism. (a) The required 'side chain free radical' synthetic equivalent could be generated from a suitable radical precursor R–X and then reacted with the radical-reactive somophile residue dehydroalanine (Dha) as a privileged unnatural amino acid 'tag' to allow site-selective 'chemical mutagenesis'. **(b)** This would generate the *capto*-dative stabilized intermediates **1**. The proposed mechanism of free radical R• generation illustrates the paths to efficient C–C bond forming chain reaction and desired product (in blue) and competing side reactions and unwanted products (in red).

Figure 3. **Bio-compatible C(sp^3)–C(sp^3) Bond-formation Allows Wide-ranging Chemical Mutagenesis**. Application of the optimized bond-forming reaction allows direct installation of sidechains onto Dha tags found in multiple, representative protein scaffolds and at different sites with in the same scaffolds. Attachment of polar, hydrophobic, non-polar, ionized and modified side-chains with natural motifs (black), with natural modifications (blue) or with unnatural motifs (red) were all possible allowing the construction of > 50 individual protein examples representative of > 25 sidechains on 8 varied, representative protein scaffolds.

Figure 4. **Use of C–C bond forming mutagenesis to *N*- and *O*-glycosylate proteins.** Using a common intermediate H3-Dha27 both forms of GlcNAc (*O*-linked and *N*-linked) could be readily introduced and their behaviour in extension and cleavage by relevant glycan-processing enzymes tested. **(a)** Despite the position in H3, which is not normally glycosylated, N-linked GlcNAc was readily extended to either a disaccharide (Gal-GlcNAc, LacNAc) or the core pentasaccharide (found in all natural *N*-linked glycans) by appropriate enzymatic systems (GalT and EndoA, respectively). However, this site proved resistant to enzymatic cleavage under those that led to full cleavage in natural *N*-glycosylated sites (conditions, PNGase, 2 M Urea pH 8.0). **(b)** As for *N*-linked glycosylation disaccharide (Gal-GlcNAc, LacNAc) and the core pentasaccharide were readily formed on *O*-linked GlcNAc, despite the unnatural site. Strikingly, *O*-GlcNAc-ase from a range of sources showed cleavage activity even at unnatural sites and notably with the human enzyme hOGA. **(c)** Glycosylation of varied protein platforms, antibody cAbLys, efflux protein AcrA and pentapeptide repeat protein Nβ, all proved possible. Experiment data reported in bar graphs represents the MFI ± SEM from three experiments or measurements.

Figure 5. **Use of C–C bond-forming mutagenesis to build phosphorylated histone proteins.** Formation of natural phospho-H3 & synthetic phospho-H3 (H3-*c*pSer10 and H3-*cf2*pSer10) via enzymatic phosphorylation & chemical mutagenesis, respectively. **(a)** Natural phosphorylation is catalyzed by Aurora B kinase and gives rise to a mixture (inset to mass spectrum) of phosphoforms. **(b)** Chemical phosphorylation via C–C bond formation gives essentially homogenous products H3-*c*pSer10 and H3-*cf2*pSer10. **(c)** Western blot analysis with anti-H3-pSer10 Ab and SDS-PAGE gel stained with Coomassie Blue (for full gel see Supplementary Figure **Error! Bookmark not defined.**) shows that the phosphomimics *carba*-pSer (*c*pSer) and *difluorocarba*-pSer installed by building a C–C bond are faithfully recognized as phosphorylated by the corresponding antibody raised to bind the natural modification (pSer). **(d, e)** A comparison of stability of H3-pSer10 and H3-*c*pSer10 under the actions of 5 different protein phosphatases shows that the C–C bond in H3-*c*pSer10 provides full resistance even to enzymes that fully and readily degrade the naturally phosphorylated H3-pSer10. Y-axis shows normalized percentage (mean ± s.d.) of phosphorylated protein taken in triplicate. **(f)** Alpha-screen binding assay using phospho-reader protein MORC3 shows binding by H3-*c*pSer10 and enhanced binding by H3-*cf2*pSer10. Measurements performed at least in triplicates and analyzed by Student's t-test at 95% confidence interval (H3-cf2Ser10 > H3-WT in binding; p-values 0.0285, 3 eq and 0.0004, 6 eq., respectively. H3-cpSer10 > H3-WT; p-value 0.0291, 6 eq.); error bars indicate mean ± SEM.

Figure 6. **Use of C–C bond-forming mutagenesis to build methylated nucleosomes.** Formation of methylLys- (H3 K9me$_3$) or methylArg- modified nucleosomes (H3 R26me$_2$a) by chemical mutagenesis enables insight into the biological functions of key histone modifications. **(a)** Chemical methylation via C-C bond formation allows installation of not only K9me$_3$ with natural isotope distribution (primarily ^{12}C) but also a ^{13}C-enriched variant precisely placed at the methyl-group carbon atoms. These reveal the time-course for demethylation by the demethylase enzyme KDM4a/JMJD2a, in both an isolated protein context (shown here, by LC/MS), and,

notably, in the context of intact nucleosomes. This suggests the use of such 'isotope-PTMs' as ~~novel~~ probes of demethylase activity, for example, by ^{13}C NMR. Time points represent the mean of three independent experiments; error bars (s.e.m) shown in SI are omitted here for clarity; curves were fitted by global least-square regression algorithm to solutions of a simplified first order model. **(b)** C-C bond forming mutagenesis also enabled the direct site-specific installation of asymmetric dimethylarginine residues into intact histones. H3R26me$_2$a was cleanly installed via C-C bond formation (see Supplementary Methods), assembled into nucleosomes. When used to probe human (HeLa) cell extracts for interaction partners (three independent biological and two technical replicates), previously unanticipated protein partners implicated in rDNA repression and recognition of DNA damage were identified amongst 797 quantified proteins in nuclear extract, suggesting that ablation of DNA-to-H3 hydrogen bonding may be critically affected by such R26 methylation. Significant interacting protein partners (denoted by their gene name) identified upon Label-free quantification (LFQ) on student's t-test analysis are shown in red in the 'volcano plot' inset (x axis, logarithmized ratio of LFQ intensity difference among two groups; y axis, logarithmized p-value from the test statistics).

Supplementary Materials:

Materials and Methods

Figures S1-S35

Tables S1-S2

Spreadsheet for Proteomic Table S3

External Databases ORA-data doi: XXXPLACEHOLDERXXX

References (*91-103*)