Post-translational mutagenesis: a chemical synthetic strategy for exploration of protein side-chain diversity

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] Stefanie Gerstberger,[1] Maria E. Palm-Espling,[1] Benedikt M. Kessler,[2] Timothy D.W. Claridge,[1] Shabaz Mohammed[1] & Benjamin G. Davis[1]*

[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*

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

ψPresent address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

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

Abstract ~320 words

Site-directed mutagenesis has revolutionized biology by allowing the variation of proteinogenic amino acids at will. However, despite recent advances in genetic expansion and genome-editing technologies, routine mutagenesis is still limited to the 20 most common amino acids. In natural systems, widespread post-translational modification of proteins expands the structural and functional capabilities of proteins beyond those directly specified by the genetic code. However, the vast diversity of chemically-plausible (including those that are unnatural but functionally relevant) side-chains is not readily accessible. Taking nature's lead, we describe here a complementary chemical approach to mutagenesis that utilizes the arsenal of synthetic chemistry to alter amino acid side-chains directly on proteins. As it is applied following protein expression it is, in principle, unlimited in its scope. We demonstrate $C(sp^3)$–$C(sp^3)$ bond-forming reactions on proteins under biocompatible conditions, which exploit unusual carbon free radical chemistry, and use them to form the Cβ–Cγ side-chain bonds that are found near-universally in protein side-chains. These transformations now 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 (all alpha, alpha/beta-folds, all beta, receptor, enzyme, antibody). This 'post-translational chemical mutagenesis' has enabled access to and the study of difficult-to-access but important residues in proteins: methyl-Arg, citrulline, ornithine, methyl-Gln and even stabilized, pK_a-variable mimetics of phospho-Ser. We believe that this new mutational paradigm, where protein side-chains are not subject to the rigid constraints of the ribosome and enzymatic processing, will prove to be a general technology for accessing diverse proteins. We conceive that

further development of our strategy, which thus circumvents the 'central dogma', will provide chemical access to a theoretically vast side-chain space largely inaccessible by other current approaches, allowing the precision of synthetic chemistry to be brought to bear on the development of one form of synthetic biology.

Introduction ~700 words

Crick highlighted that current methods (in nature and in the laboratory) for changing protein structure (and hence function) are bound by, what he termed, a 'central dogma'(*1*) that essentially dictates that genetic sequence determines protein sequence. However, the hypothetical 'chemical space' accessible to proteins in nature is, in principle, unlimited and so not constrained to the canonical amino acids. Natural post-translational modifications (PTMs) to proteins partially expand the chemical groups available to proteins, modulating both structure and function (**Figure 1a**).(*2*) For example, protein glycosylation tunes both physical (e.g., solubility, stability and folding) and biological (e.g, immune response, cell adhesion events, signaling) activity.(*3*) 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.(*4*) 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.(*5*) 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.(*6*) 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'.(*7, 8*)

Site-directed mutagenesis has revolutionized the study and understanding of proteins.(*9, 10*) 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,(*11*) 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.(*12-14*) 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,(*15*) while useful, remain limited in scope of structural variation(*14*) by the tolerance of the translational machinery and hence must be optimized by on a case-by-case basis.(*13*) The total or semi-synthesis of proteins has been made possible by powerful native chemical ligation techniques.(*16*) 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.(*17*)

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.(*18*) The introduction of various chemical 'tags',(*19*) that allow selective protein modification *via* reactivity compatible with that of natural biomolecules,(*20*) has been an important step towards this goal. However, current protein modification approaches rely on unnatural carbon-heteroatom linkages(*21*) 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.(*20*) 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 present the first examples of carbon(sp3)–carbon(sp3) bond-forming protein modification protocols as an important first step towards the goal of a general post-translational mutagenesis.(*8, 18, 22*) We demonstrate that its divergent flexibility allowed the synthesis of more than 25 mutated protein sidechains from a single precursor and appears compatible with representative protein fold types. We expect that this reaction will enable 'post-translational mutagenesis' to complement and extend current methods of designed protein generation.

Results (~3500 words)

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

We envisaged a unique strategy through retrosynthetic analysis (**Figure 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, **Supplementary Figure S1**) results in consideration of reagents 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(*23*) and unreactive (and thereby compatible) with the majority of existing functionality present in biomolecules ('bioorthogonal'(*20*)). We reasoned that use of such mild, carbon-centred free radical chemistry (**Figure 1c**) 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, (*24-26*) biosynthetically or chemically(*27, 28*) and is a potent 'SOMOphile',(*29*) which, upon reaction with a suitable radical, would favourably generate a *capto*-dative stabilized Cα radical (**Figure 1d**).

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 (**Figure 1d**). 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(*30, 31*) (e.g. zinc(*32, 33*) Zn^0 or indium(*34*) 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 (**Figure 1c**, **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 (**Supplementary Figure S2** 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 (**Supplementary Figure S3**). 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 **Supplementary Figure 2** 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.(*35*)

We observed that the predicted Cα radical intermediate **1** (**Figure 1d**), 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 (**Supplementary Figure 4**). 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 cleavage or degradation from migration to give further competing reactivity(*36*). Notably, additional amounts of reagents were seemingly needed for full conversion due to such observed competing process (reduction of alkyl halide to alkane as well as disubstitution and oxidative degradation). The successful development of a radical reaction for *protein* modification (**Figure 1c**) 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(*28, 37*) 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 **Supplementary Figure 5**) in both SBL and H3. However, clear additional side-products were also detected. Careful isolation, trapping and characterization of these side-products (see **Supplementary Figures 6, 7, 8, 9**) 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 (**Supplementary Figures 6,7,8**) oxidative cleavage products – C-terminal amide **2** and di-carbonyl **3** – would arise from the termination reaction of the Cα radical with triplet oxygen (**Figure 1d**).(*38*) 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 the 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 (**Supplementary Figure 9**). 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, **Supplementary Figure 4**). These observations in proteins therefore were consistent too with radical addition at Cγ followed by termination of the Cα radical intermediate thus generated (**Figure 1d**).

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 (**Figure 1d**) 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 **Supplementary Figure 10**). 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(*39, 40*) 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 **Supplementary Figure 11** 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 (**Figure 1d**) 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 (**Figure 1d**) 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}) (**Figure 1d**). 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(*41-43*) and related aluminiumhydrides(*44*) have previously suggested possible radical behaviour under certain circumstances.(*45*) 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 **Supplementary Figure 12**). 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 **Supplementary Figure 13**) but also through the use of radical trapping; the reaction was fully inhibited by sub-stoichiometric 4-hydroxy-TEMPO or acrylamide (see **Supplementary Figure 14** 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 **Supplementary Figure 15, 16**) 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 **Supplementary Figure 17**).

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 (**Table 1**). 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 corresponding appropriate, typically readily available, organic iodides allowed wide-ranging systematic variations in various ways. Thus, not only Leu could be incorporated but so could a series of systematic variants: Leu minus methyl →demethyl-Leu (smaller); Leu plus methyl →*tert*-Leu (bulkier); Leu with migrated methyl →*nor*-Leu (slimmer); Leu cyclized →'cyclo'-Leu (conformationally restricted). 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→CF_3 (e.g. in demethyl-Leu) or CH_3→$^{13}CH_3$ (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,(*46-48*) 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.(*49*) Indeed, use of ^{19}F NMR allowed us to further confirm both the regioselectivity and stereoselectivity for chemical mutations (**Supplementary Figure 18**), 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(*3*) (in *O*- and *N*- linked form), Lys-methylation(*50*) (in all three states: mono-, di- and tri-, as well as labeled tri-^{13}C), Arg-methylation(*51*) (mono- and di-), Gln-methylation.(*52*) 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(*53*) and di- / tri-methylated-Lys that cannot yet be incorporated into proteins (even indirectly, as mono-methyl-Lys currently is(*54*)) by cellular stop-codon suppression.(*14*)

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(*55*); p38α mammalian

mitogen-activated MAP kinase(*56*); mammalian antibody cAbLys3(*57*); apoptosis marker binding protein annexinV(*58*); and pentapeptide-repeat protein Npβ.(*59*) 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). 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), structure (e.g, secondary structural content by circular dichroism) and function (e.g. enzymatic activity, Ab-binding function and biomarker recognition), (see **Supplementary Figures 19, 20 and Supplementary Methods**) confirmed not only the site-selectivity of the C–C-bond-forming chemical mutation but also its benign nature (no secondary structure disruption or abrogation of function was observed).

$C(sp^3)$–$C(sp^3)$ 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 novel 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,(*2, 60*) yet has been prominently absent as a readily accessible motif in proteins via chemical or genetic mutagenesis methods.(*61*) Until now no general chemical method for convergently installing *N*- and *O*-linked glycans has been possible.(*60, 61*) 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 2**). Despite the unnatural nature 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.(*2, 60*) We discovered that despite an apparent plasticity with respect to protein scaffold(*62*), widely-used(*63*) *N*-glycosidase PNGase did not cleave synthetic variants with extended side-chain length (**Figure 2a** and **Supplementary Figure 21**). 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 (**Figure 2b** and **Supplementary Figure 22**), including the human *O*-GlcNAc-ase (hOGA) enzyme. The latter, which is implicated in diabetes,(*64*) dementia(*65*) and cancer(*66*) has, until now, been presumed selective given that it is the sole encoded protein *O*-GlcNAc-ase in the human genome;(*67*) 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 $Man_3GlcNAc_2$ (**Figure 2c** and **Supplementary Figure 23**).

Aurora B kinase, which is overexpressed in cancers,(*68*) phosphorylates Ser10 of histone H3 during mitosis(*69*) 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.(*70*) 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 and *cf2*pSer (**Figure 3a**); MS analysis revealed that this could be achieved with a purity that is not possible with current biological methods for phosphorylating H3 (**Figure 3b**). Antibodies and appropriate binding proteins (such as the 'reader' protein MORC3(*71*)) not only recognized the resulting phosphomimic proteins (**Figures 3c,f** and **Supplementary Figure 24**), but proved stable to either chemical or enzyme-catalyzed dephosphorylation (**Figure 3d,e** and **Supplementary Figure 25**), 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 (**Supplementary Figure 26**). Whilst *c*pSer10 proved a functioning mimic of pSer10, it has been argued(*72, 73*) that fluorophosphonates may act as more effective phosphate mimics by virtue of pKa,(*74*) 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(*72-75*) difluorophosphonate mimicry of phosphates in proteins (see **Supplementary Figure 3f**).

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 (e.g. 'writers'-'readers'-'erasers') with the 'methylproteome' remains a vital 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 (**Figure 4a**) and a di-methylated-Arg residue site-selectively installed into in an intact nucleosomal particle (**Figure 4b**). ^{13}C-labelled-H3 H3-[^{13}C-Me]$_3$-Lys9 was created (**Figure 4a**) 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 (**Figure 4a** and **Supplementary Figure 27a**).(*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 (**Figure 4a,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 (**Supplementary Figure 27b**). 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 (**Figure 4b**) 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 (**Figure 4b** and **Supplementary Data Table 1**). 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 (**Supplementary Figure 28**). 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.

Discussion (~800 words)

We have revealed here highly unusual (and strikingly simple) reaction conditions (organic halide, $NaBH_4$ in deoxygenated aqueous buffer) not only for radical chemistry but certainly for protein modification. The striking compatibility that we have observed here for the use of radicals as effective and yet benign reactive intermediates for protein modification suggests that other radical-based methods(*36*) may prove very powerful in the field of protein chemistry (which has almost exclusively been dominated by two-electron (heterolytic) bond forming processes). 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(*36*) 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 new protein radical chemistry is likely to require new reagents, new combinations of reagents and their study; notably here we used highly familiar reagents to the organic chemist but combined and applied to achieve a highly unusual reaction process. In that context the potential of boron reagents in radical chemistry has been noted and resurfaced over many years(*85, 86*) and associated mechanisms remain a topic of active debate;(*44, 45*) whilst mechanisms other than the one we propose here cannot be discounted, their role here appeared critical.

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')(*16*) 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 **Supplementary Figure 29** and **Supplementary Methods**).

The great freedom to explore natural and unnatural side chains that is enabled by this method also suggests many experiments to probe biology with atomic-level precision. Since simple choice of the correct reagent allows ready variation of even single atom substituents (e.g. O → CH_2 → CF_2 in pSer, *c*pSer and *cf2*pSer, as we have shown here) then many experiments to enable a truly molecular understanding of mechanism in biology and a fine-tuning of function can now be planned. Thus, 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) that we were able to incorporate chemically sometimes prove difficult to incorporate by such genetic methods due to their strong structural resemblance to their unmodified counterparts and hence difficulty in differentiation for the translational machinery.

In turn, this improved access to relevant protein architectures will likely reveal surprising observations associated with their action as the primary, workhorse

biomolecules. For instance, our data has 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. Our ready elucidation here too of new binding partners for nucleosomes in previously inaccessible methylation states (e.g. asymmetrically dimethylated Argme2a) suggests that many other new interactions will be discovered via C–C chemical mutagenesis. These, in turn, will then allow us to identify and synthetically 'programme' 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 shown here with possible sugar ligands for the Fcγ-receptor raises the possibility of new cell killing strategies (*89*) mediated now by synthetic Ab-fragments. In this way, we can envisage truly free ranging access, via chemistry, to synthetic proteins that will allow not only discovery of new function but also application of those newly elucidated functions in, for example, new protein drugs, 'synthetic biologics'.(*90*)

Author Contributions

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

Acknowledgments

We thank the Felix Foundation (RR), the Croucher Foundation (WLN), the Rutherford Foundation (THW, SF), Cancer Research UK (SF), BBSRC (SRGG, BJB), AstraZeneca (SRGG), UCB (BJB) for funding. BGD was in receipt of a Royal Society Wolfson Research Merit Award during part of this project. We thank Andrew Baldwin, William Myers, Thomas Parsons, James Wickens, Hanna Hailu, Prashant Mori, Alistair Henry, K. Phin Chooi, Tom Fleming for useful discussions and / or technical assistance.

References

1. F. Crick, Central Dogma of Molecular Biology. *Nature* **227**, 561-563 (1970).
2. 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).
3. K. Ohtsubo, J. D. Marth, Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).
4. P. Cohen, The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-E130 (2002).
5. E. I. Campos, D. Reinberg, Histones: Annotating Chromatin. *Annual Review of Genetics* **43**, 559-599 (2009).
6. 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 Engl* **55**, 5896-5903 (2016).
7. C. Zioudrou, M. Wilchek, A. Patchornik, Conversion of the L-Serine Residue to an L-Cysteine Residue in Peptides*. *Biochemistry* **4**, 1811-1822 (1965).
8. 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).
9. C. A. Hutchison, III *et al.*, Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551-6560 (1978).
10. J. A. Brannigan, A. J. Wilkinson, Protein engineering 20 years on. *Nat Rev Mol Cell Biol* **3**, 964-970 (2002).
11. 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).
12. L. Wang, P. G. Schultz, Expanding the Genetic Code. *Angewandte Chemie International Edition* **44**, 34-66 (2005).
13. J. W. Chin, Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annual Review of Biochemistry* **83**, 379-408 (2014).
14. A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, Designing logical codon reassignment - Expanding the chemistry in biology. *Chemical Science* **6**, 50-69 (2015).
15. 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).
16. P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779 (1994).
17. L. R. Malins, R. J. Payne, Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Current Opinion in Chemical Biology* **22**, 70-78 (2014).

18. 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).
19. J. M. Chalker, G. J. L. Bernardes, B. G. Davis, A "Tag-and-Modify" Approach to Site-Selective Protein Modification. *Accounts of Chemical Research* **44**, 730-741 (2011).
20. E. M. Sletten, C. R. Bertozzi, Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* **48**, 6974-6998 (2009).
21. C. D. Spicer, B. G. Davis, Selective chemical protein modification. *Nat Commun* **5**, (2014).
22. 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).
23. T. Perchyonok, *Radical Reactions in Aqueous Media*. (Royal Society of Chemistry, Cambridge, 2009).
24. J. Wang, S. M. Schiller, P. G. Schultz, A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed.* **46**, 6849-6851 (2007).
25. F. P. Seebeck, J. W. Szostak, Ribosomal synthesis of dehydroalanine-containing peptides. *J Am Chem Soc* **128**, 7150-7151 (2006).
26. E. Brustad *et al.*, A genetically encoded boronate-containing amino acid. *Angew. Chem. Int. Ed.* **47**, 8220-8223 (2008).
27. 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).
28. J. M. Chalker *et al.*, Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chemical Science* **2**, 1666-1676 (2011).
29. B. Giese, *Radicals in Organic Synthesis: Formation of Carbon-Carbon Bonds*. J. E. Baldwin, Ed., (Pergamon Press, Oxford, 1986).
30. C. J. Li, Organic reactions in aqueous media - with a focus on carbon-carbon bond formation. *Chemical Reviews* **93**, 2023-2035 (1993).
31. 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).
32. C. Petrier, C. Dupuy, J. L. Luche, Conjugate additions to α,β-unsaturated carbonyl compounds in aqueous media. *Tetrahedron Letters* **27**, 3149-3152 (1986).
33. 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. *Chemical Communications*, 2440-2441 (2002).
34. H. Miyabe, M. Ueda, A. Nishimura, T. Naito, Indium-mediated intermolecular alkyl radical addition to electron-deficient CN bond and CC bond in water. *Organic letters* **4**, 131-134 (2002).
35. J. M. Tedder, Which factors determine the reactivity and regioselectivity of free radical substitution and addition reactions? *Angewandte Chemie International Edition in English* **21**, 401-410 (1982).

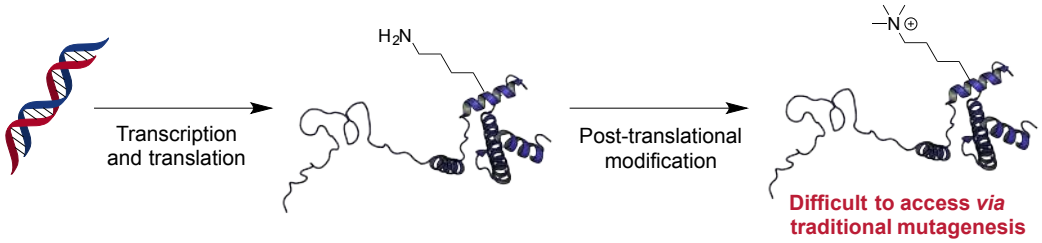
36. C. L. Hawkins, M. J. Davies, Generation and propagation of radical reactions on proteins. *Biochim Biophys Acta* **1504**, 196-219 (2001).
37. J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angewandte Chemie International Edition* **51**, 1835-1839 (2012).
38. W. M. Garrison, Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chemical Reviews* **87**, 381-398 (1987).
39. G. Milazzo, S. Caroli, V. K. Sharma, *Tables of Standard Electrode Potentials*. (Wiley, Chichester, 1978).
40. A. J. Bard, R. Parsons, J. Jordan, *Standard Potentials in Aqueous Solutions*. (Marcel Dekker, New York, 1985).
41. J. T. Groves, K. W. Ma, Dehalogenations with sodium borohydride. Evidence for a free radical reaction. *Journal of the American Chemical Society* **96**, 6527-6529 (1974).
42. E. J. Corey, J. W. Suggs, Method for catalytic dehalogenations via trialkyltin hydrides. *The Journal of Organic Chemistry* **40**, 2554-2555 (1975).
43. G. Stork, N. H. Baine, Cyclization of vinyl radicals: a versatile method for the construction of five- and six-membered rings. *Journal of the American Chemical Society* **104**, 2321-2323 (1982).
44. 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. *The Journal of Organic Chemistry* **49**, 3545-3556 (1984).
45. J. A. Barltrop, D. Bradbury, Chain photoreaction of sodium borohydride with halogenated aromatic hydrocarbons. Initiation by aryl radicals. *Journal of the American Chemical Society* **95**, 5085-5086 (1973).
46. 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).
47. 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. *Journal of the American Chemical Society* **136**, 6037-6048 (2014).
48. 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. *Journal of the American Chemical Society* **137**, 15378-15381 (2015).
49. N. C. Yoder, K. Kumar, Fluorinated amino acids in protein design and engineering. *Chemical Society Reviews* **31**, 335-341 (2002).
50. C. Martin, Y. Zhang, The diverse functions of histone lysine methylation. *Nature reviews Molecular cell biology* **6**, 838-849 (2005).
51. M. T. Bedford, S. G. Clarke, Protein arginine methylation in mammals: who, what, and why. *Molecular cell* **33**, 1-13 (2009).
52. P. Tessarz *et al.*, Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505**, 564-568 (2014).

53. H. Jakubowski, Misacylation of tRNALys with noncognate amino acids by lysyl-tRNA synthetase. *Biochemistry* **38**, 8088-8093 (1999).
54. 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. *Journal of the American Chemical Society* **131**, 14194-14195 (2009).
55. H. I. Zgurskaya, H. Nikaido, AcrA is a highly asymmetric protein capable of spanning the periplasm. *Journal of molecular biology* **285**, 409-420 (1999).
56. J. C. Lee *et al.*, A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature* **372**, 739-746 (1994).
57. 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).
58. C. E. Creutz, The annexins and exocytosis. *Science* **258**, 924-931 (1992).
59. 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 Science : A Publication of the Protein Society* **16**, 755-760 (2007).
60. D. P. Gamblin, E. M. Scanlan, B. G. Davis, Glycoprotein Synthesis: An Update. *Chem. Rev.* **109**, 131-163 (2009).
61. L.-X. Wang, B. G. Davis, Realizing the promise of chemical glycobiology. *Chemical Science* **4**, 3381-3394 (2013).
62. F. Altmann, S. Schweiszer, C. Weber, Kinetic comparison of peptide: N-glycosidases F and A reveals several differences in substrate specificity. *Glycoconjugate journal* **12**, 84-93 (1995).
63. 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).
64. S. Hardivillé, Gerald W. Hart, Nutrient Regulation of Signaling, Transcription, and Cell Physiology by O-GlcNAcylation. *Cell Metabolism* **20**, 208-213 (2014).
65. Y. Zhu, X. Shan, S. A. Yuzwa, D. J. Vocadlo, The emerging link between O-GlcNAc and Alzheimer disease. *The Journal of biological chemistry* **289**, 34472-34481 (2014).
66. C. Slawson, G. W. Hart, O-GlcNAc signalling: implications for cancer cell biology. *Nature reviews. Cancer* **11**, 678-684 (2011).
67. R. Hurtado-Guerrero, H. C. Dorfmueller, D. M. F. van Aalten, Molecular mechanisms of O-GlcNAcylation. *Current Opinion in Structural Biology* **18**, 551-557 (2008).
68. J. R. Bischoff *et al.*, A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. *The EMBO Journal* **17**, 3052-3065 (1998).
69. 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).

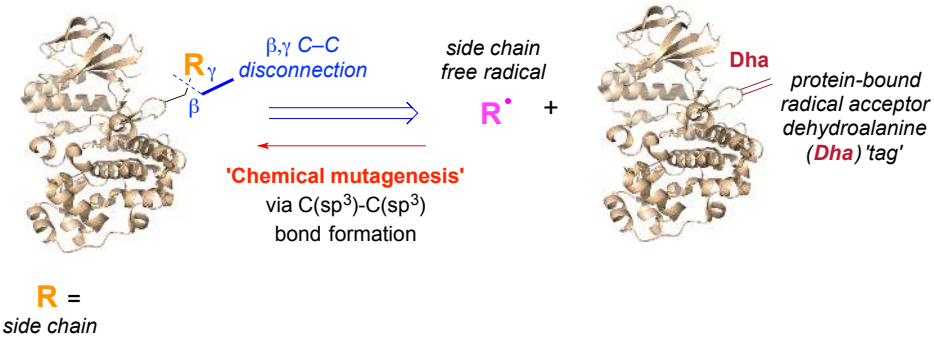
70. R. Klingberg *et al.*, Analysis of Phosphorylation-Dependent Protein–Protein Interactions of Histone H3. *ACS chemical biology* **10**, 138-145 (2014).
71. N. Kunowska, M. Rotival, L. Yu, J. Choudhary, N. Dillon, Identification of protein complexes that bind to histone H3 combinatorial modifications using super-SILAC and weighted correlation network analysis. *Nucleic acids research*, gku1350 (2015).
72. G. M. Blackburn, *Chem. Ind.* , 134-138 (1981).
73. G. M. Blackburn, D. E. Kent, F. Kolkmann, *J. Chem. Soc., Perkin Trans. 1*, 1119-1125 (1984).
74. W. Zheng *et al.*, Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *The Journal of biological chemistry* **280**, 10462-10467 (2005).
75. D. J. Burton, Z.-Y. Yang, W. Qiu, *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 & development* **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. *Nature structural & molecular biology* **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. *Proceedings of the National Academy of Sciences of the United States of America* **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. *Chemical Reviews* **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. *Journal of Biological Chemistry* **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.

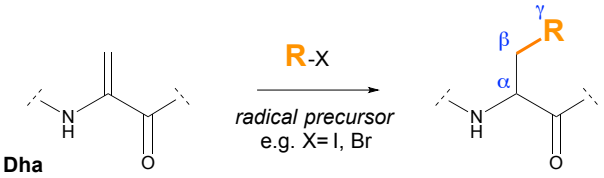
a) Post-translational modification in nature



b) Retrosynthetic logic behind chemical mutagenesis



c) Csp^3-Csp^3 bond-forming reaction for chemical protein modification



d)

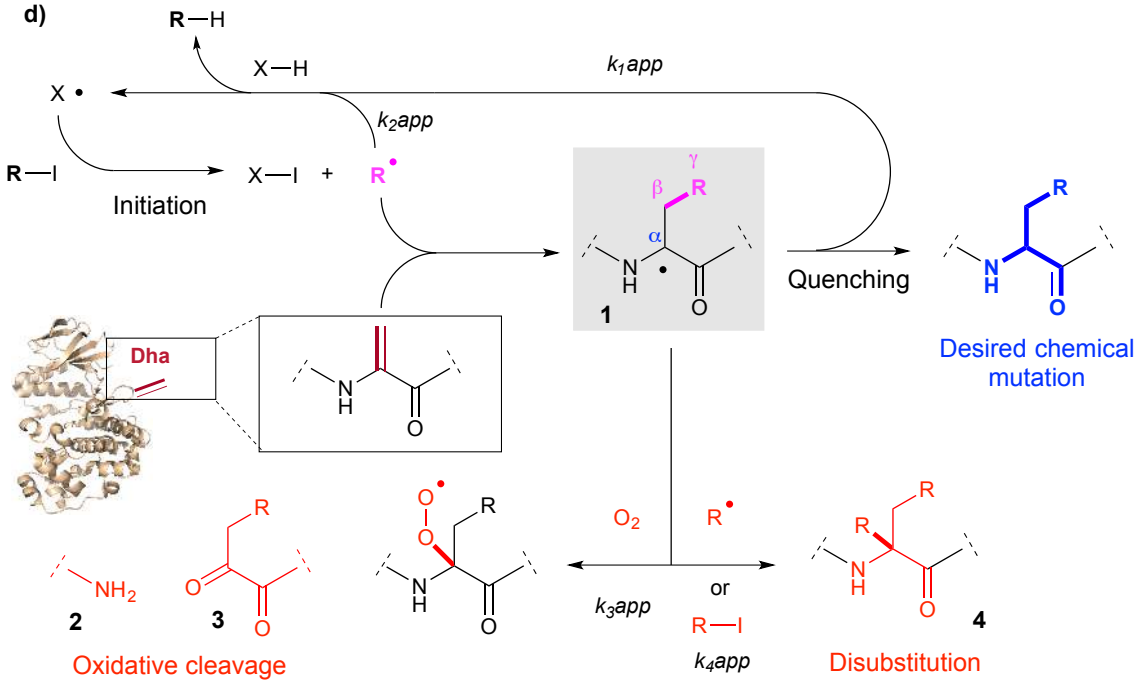


Figure 1. Design of a Bio-compatible C(sp^3)–C(sp^3) bond-forming reaction as part of a General Strategy for Chemical Mutagenesis. (a) Natural post-translational modification exploits C-heteroatom bond formation, such as C–N bond formation in Lysine (Lys) methylation, to diversify protein structure and 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' (see **Supplementary Figure 1** for others), one of which is the free radical equivalent **R•** that arises from homolytic disconnection. **(c)** Such a 'side chain free radical' synthon 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'. **(d)** This would generate the *capto*-dative stabilized intermediate **1**. The proposed mechanism of free radical **R•** generation illustrates the paths efficient C–C bond forming chain reaction and competing side reactions.

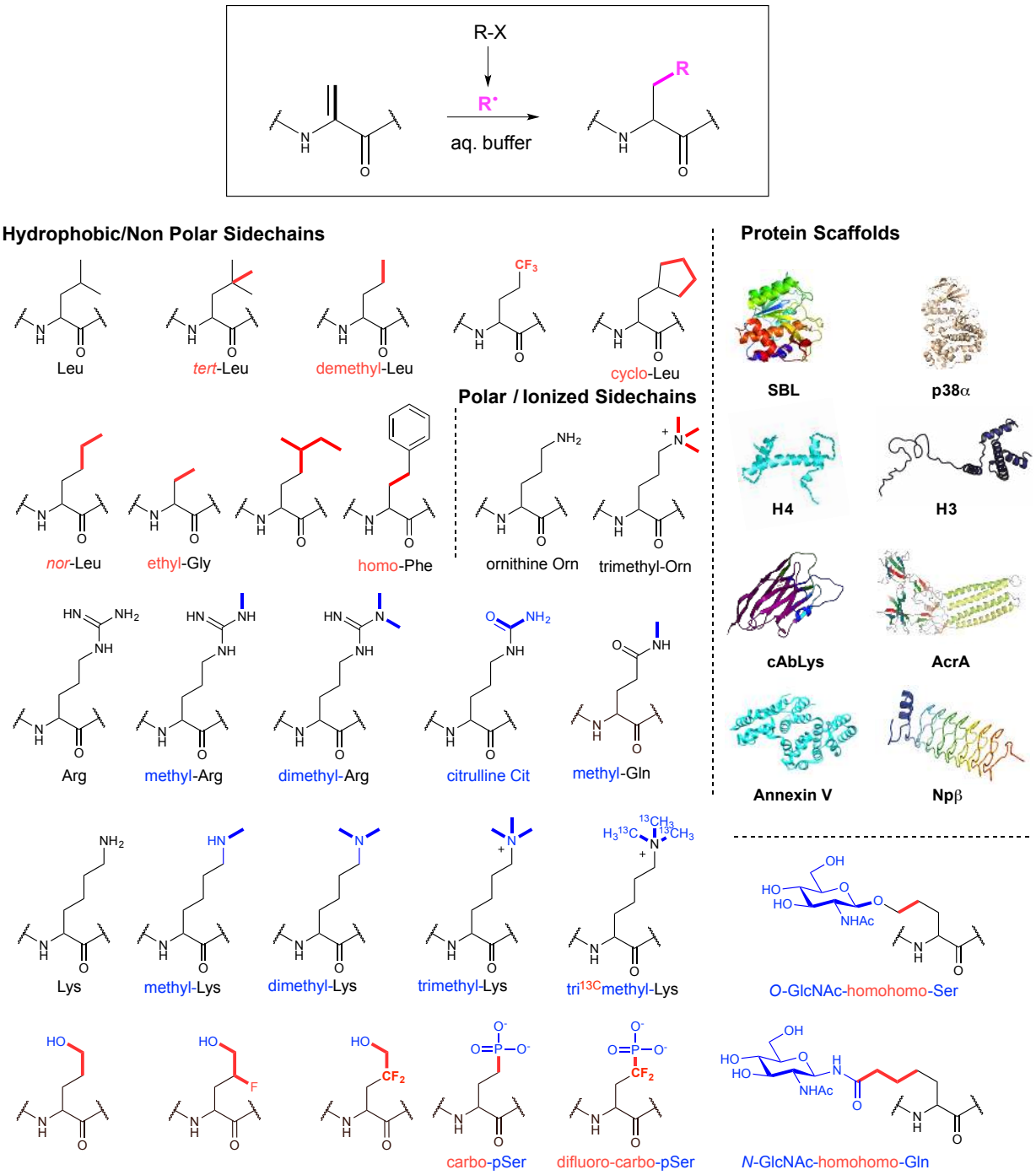


Table 1. **Bio-compatible C(sp^3)–C(sp^3) Bond-formation Allows Wide-ranging Chemical Mutagenesis**. Application of the optimized bond-forming reaction allows direct installation of sidechains onto Dha tags found in multiple, representative protein scaffolds and at different sites with in the same scaffolds. 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.

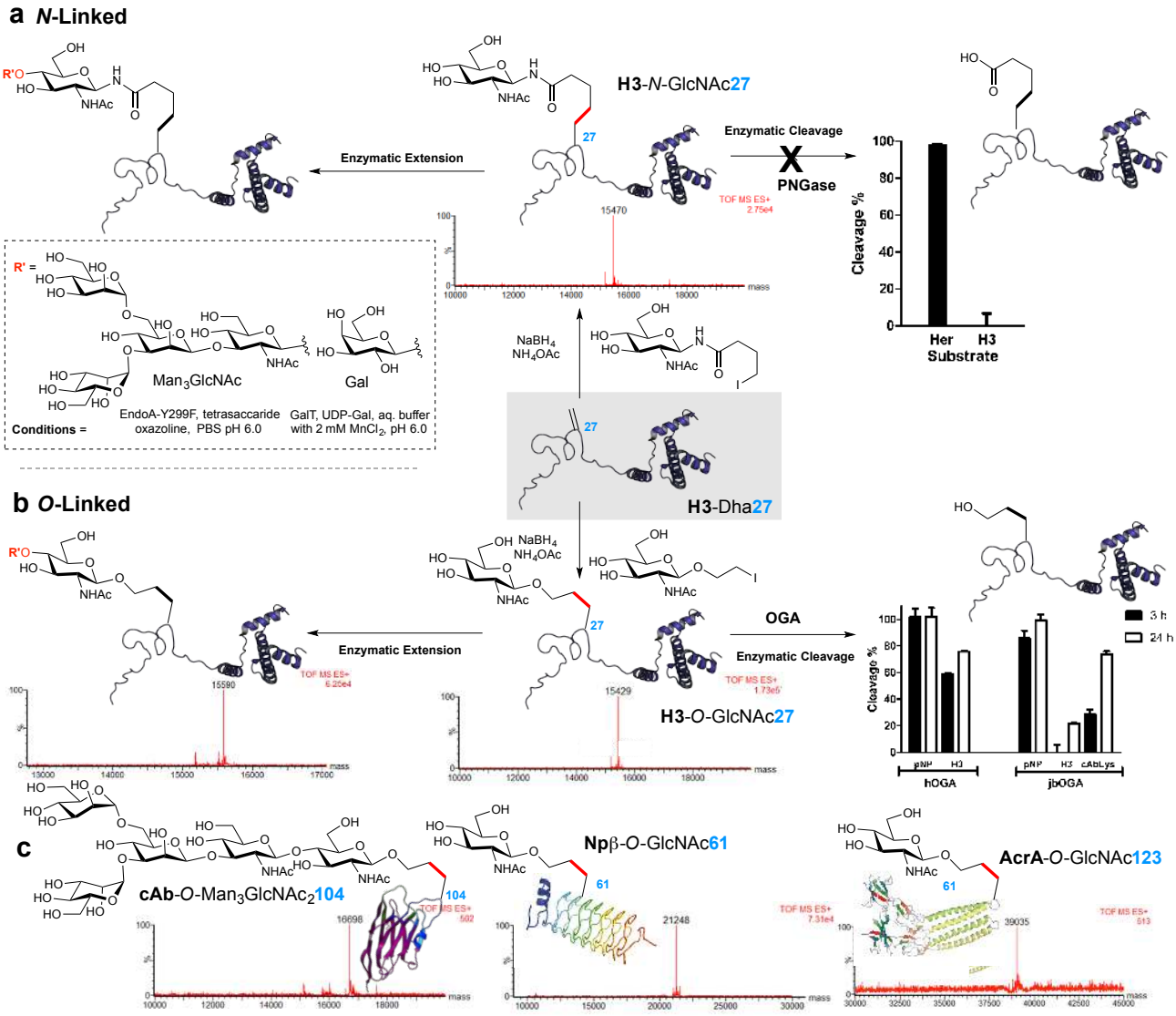


Figure 2. **Use of C–C bond forming mutagenesis to *N*- and *O*-glycosylate proteins.** Using a common intermediate H3-Dha27 both forms of GlcNAc (*O*-linked and *N*-linked) could be readily introduced and their behaviour in extension and cleavage by relevant glycan-processing enzymes tested. **(a)** 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.

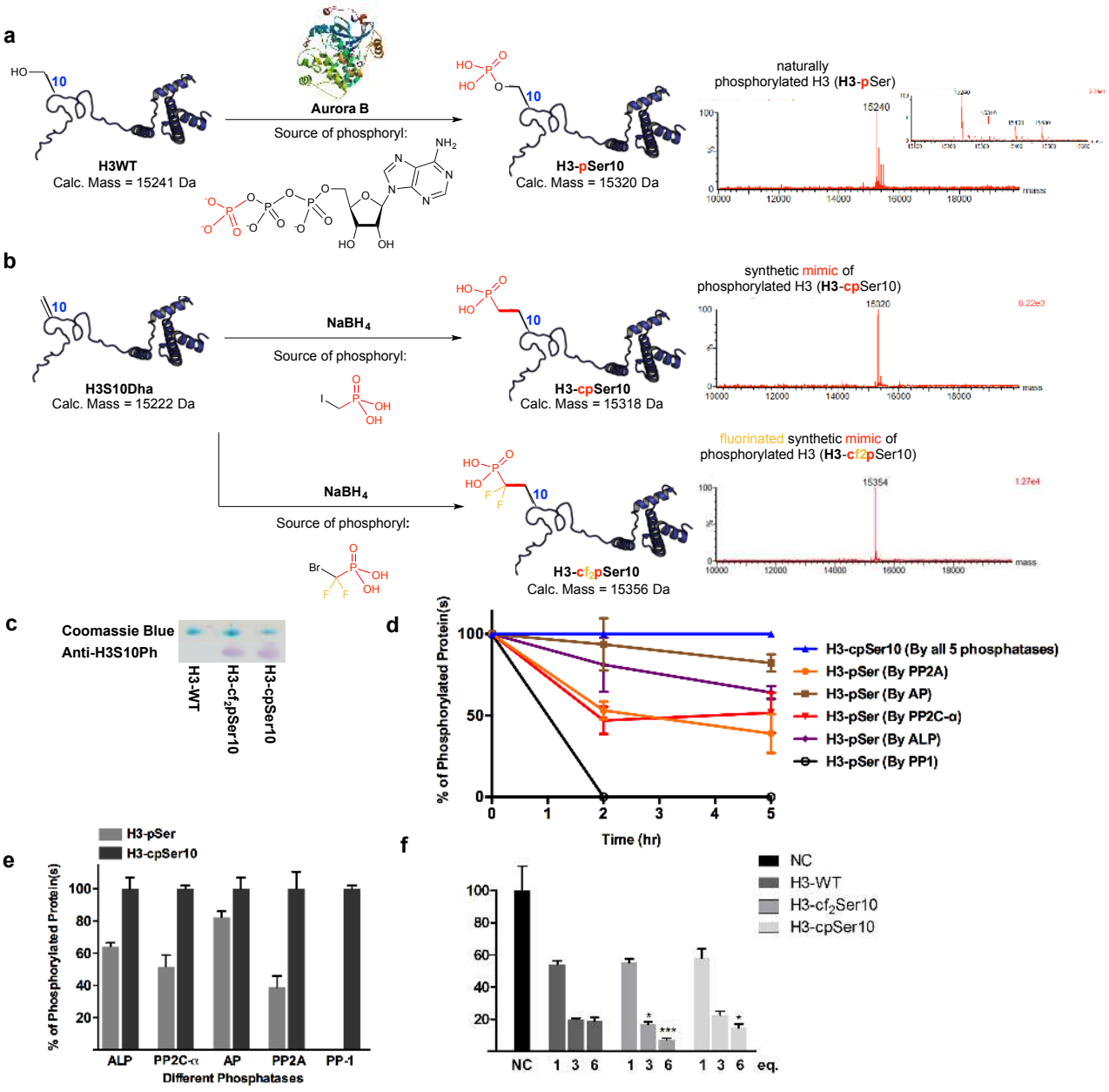


Figure 3. **Use of C–C bond-forming mutagenesis to build phosphorylated histone proteins.** Formation of natural phospho-H3 & synthetic phospho-H3 (H3-*c*pSer10 and H3-*cf2*pSer10) via enzymatic phosphorylation & chemical mutagenesis, respectively. **(a)** Natural phosphorylation is catalyzed by Aurora 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 24**) 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.

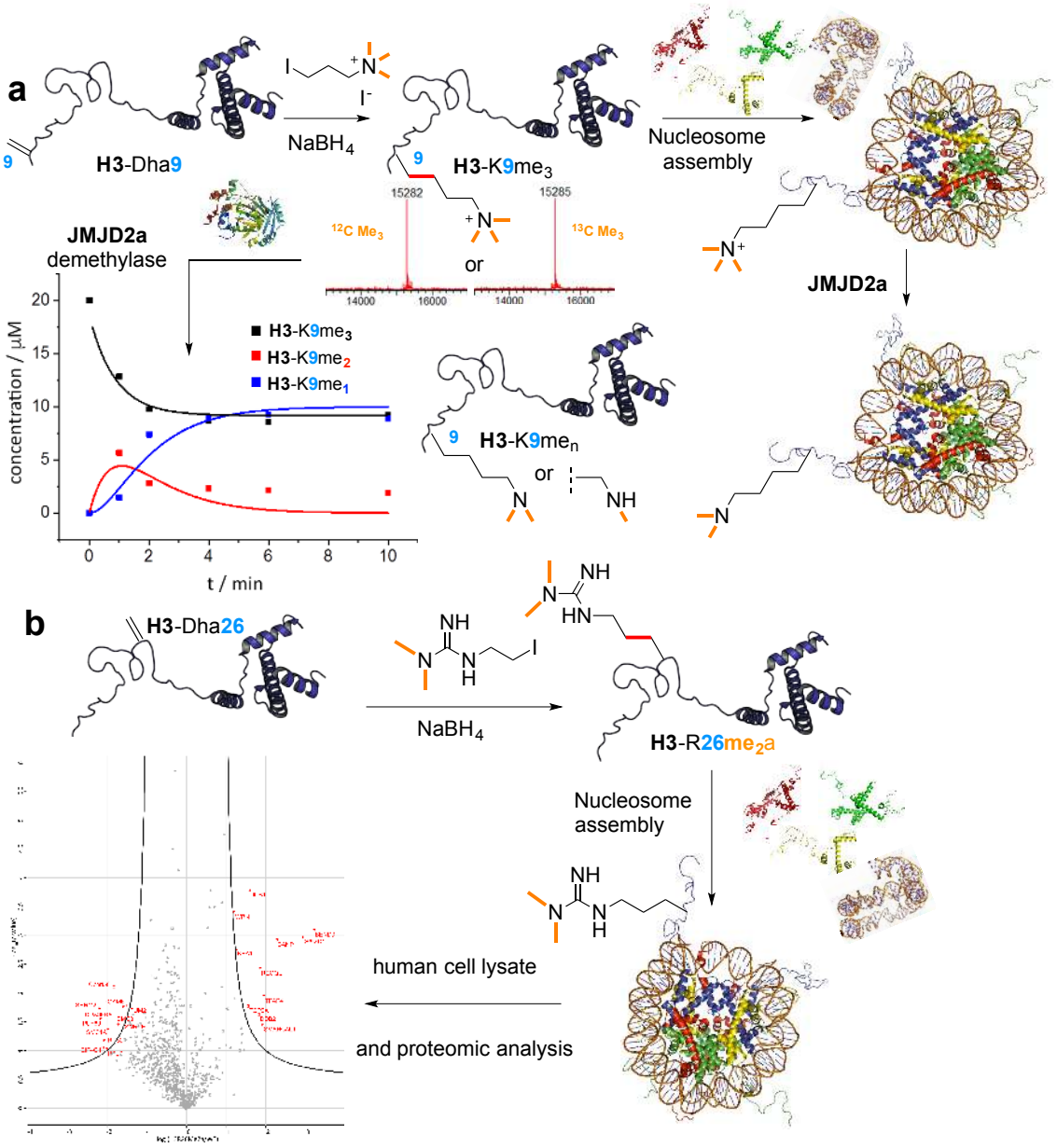


Figure 4. **Use of C–C bond-forming mutagenesis to build methylated nucleosomes.** Formation of methylLys- (H3 K9me$_3$) or methylArg- modified nucleosomes (H3 R26me$_2$a) by chemical mutagenesis enables insight into the biological functions of key histone modifications. **(a)** Chemical methylation 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_2a 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).