

Residue-Selective Protein C-Formylation via Sequential Difluoroalkylation-Hydrolysis

Mateusz Imiolek, Patrick G. Isenegger, Wai-Lung Ng, Aziz Khan, Véronique Gouverneur, and Benjamin G. Davis*



Cite This: *ACS Cent. Sci.* 2021, 7, 145–155



Read Online

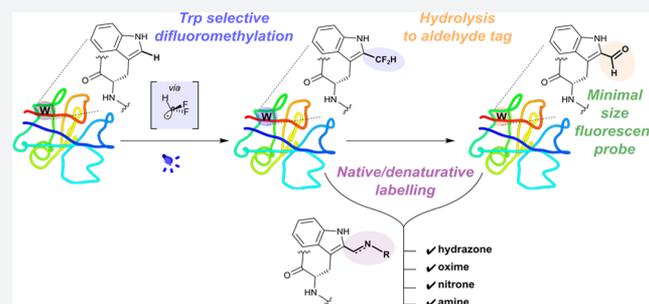
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The carbonyl group is now a widely useful, nonproteinogenic functional group in chemical biology, yet methods for its generation in proteins have relied upon either cotranslational incorporation of unnatural amino acids bearing carbonyls or oxidative conversion (chemical or enzymatic) of existing natural amino acids. If available, alternative strategies for directly adding the C=O group through C–C bond-forming C-carboxylation, particularly at currently inaccessible amino acid sites, would provide a powerful method for adding valuable reactivity and expanding possible function in proteins. Here, following a survey of methods for HCF₂· generation, we show that reductive photoredox catalysis enables mild radical-mediated difluoromethylation-hydrolysis of native protein residues as an effective method for carbonylation. Inherent selectivity of HCF₂· allowed preferential modification of Trp residues. The resulting C-2-difluoromethylated Trp undergoes Reimer-Tiemann-type dehalogenation providing highly effective spontaneous hydrolytic collapse in proteins to carbonylated HC(O)-Trp (C-formyl-Trp = CfW) residues. This new, unnatural protein residue CfW not only was found to be effective in bioconjugation, ligation, and labeling reactions but also displayed strong “red-shifting” of its absorption and fluorescent emission maxima, allowing direct use of Trp sites as UV-visualized fluorophores in proteins and even cells. In this way, this method for the effective generation of masked formyl-radical “HC(O)·” equivalents enables first examples of C–C bond-forming carbonylation in proteins, thereby expanding the chemical reactivity and spectroscopic function that may be selectively and post-translationally “edited” into biology.



INTRODUCTION

Carbonyl groups are not typically found in proteins, so when introduced, they create additional reactivity that has proven widely useful.¹ For example, amber codon suppression has allowed relatively widespread use of *para*-acetylphenylalanine (*pAcF*) as a “tag” for further conjugation chemistry (via, e.g., oxime, hydrazone formation).^{2,3} Often, enhanced reactivity is seen for aldehydes: *meta*-formylphenylalanine introduced also by amber codon suppression is more reactive than *pAcF*.⁴ Such access to aldehydes may be complemented by chemical (e.g., oxidative cleavage of *N*-terminal serine (Ser))⁵ or enzymatic (e.g., fGly-generating enzyme (FGE) conversion of cysteine (Cys)) methods.^{6–8} Indeed, the latter have been elegantly exploited and engineered to provide powerful, unnatural “tags” for protein-payload conjugation.⁹ In this way, cotranslational installation (Figure 1a) or post-translational oxidation of *N*-terminal Ser or certain Cys in FGE sequons (Figure 1b,c) allows access to certain carbonyls (see the Supporting Information for an extended comparison of methods for introducing the formyl group into proteins). However, to our knowledge, there are no methods to date for direct protein C–C(O) bond-forming carbonylation (Figure 1d).

Various methods for C–C bond formation on proteins have recently been proposed exploiting radicals to overcome the problem of inappropriate reactivity of C⁺ or C[–] equivalents in water.^{10,11} These have allowed, for example, recapitulation of natural post-translational modifications^{12,13} that would be otherwise difficult to generate or installation of effective functional mimics or unnatural handles useful for bioconjugation purposes.

The strategic problems of C–C bond formation are made yet more complicated when C–C(O) bond formation/C-carboxylation is required (Figure 2). In heterolytic (2e[–]) bond-forming methods (Figure 2a1, 2a2), even if suitable electrophilic groups could be generated in proteins, corresponding nucleophilic HC(O)[–] acyl anion equivalents would be rapidly protonated under the aqueous conditions needed for truly applicable

Received: September 4, 2020

Published: January 13, 2021



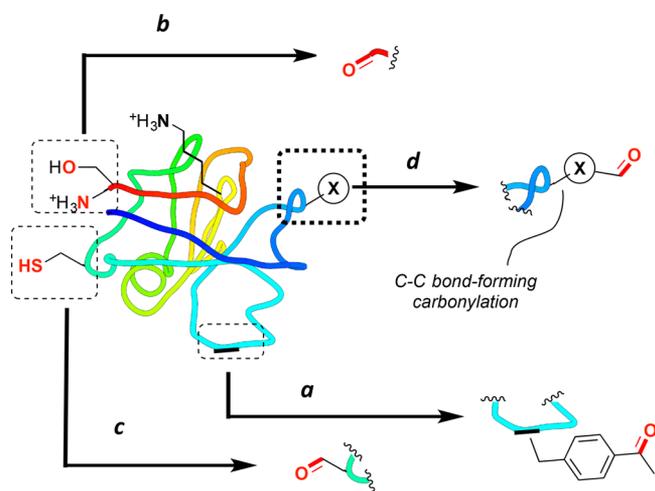


Figure 1. Comparison of methods for installation of carbonyls into proteins. Select examples of prior methods for comparison with this work. (a) Amber codon suppression allows cotranslational insertion of, e.g., *para*-acylphenylalanine. (b) Chemical periodate cleavage of N-terminal serine allows installation of N-terminal aldehyde. (c) Biocatalytic use of formyl-glycine generating enzymes (FGEs) allows conversion of cysteine to formyl glycine. (d) This work.

protein reactions (Figure 2a2). Moreover, while some electrophilic HC(O)^+ formyl cation equivalents are possible in partly hydrolytically stable form, these react preferentially with heteroatom nucleophiles such as at the N-terminus or lysine (Lys) and serine (Ser) side chains¹⁴ (Figure 2a1). Homolytic radical methods by contrast have not yet been effectively extended to useful $\text{HC(O)}\cdot$ generation;¹⁵ $\text{HC(O)}\cdot$ appears to also lack the direct reactivity (see below) needed for reaction with proteinogenic residues (Figure 2b).

We therefore considered an indirect approach (Figure 2c). We have previously demonstrated that some native protein residues will react with $\text{CF}_3\cdot$ radicals,¹⁶ which appear to sit at a cusp of reactivity suitable for certain aromatic proteinogenic residues, primarily Trp. Other C \cdot radicals require more reactive (unnatural) radical acceptor residues to be present in proteins for C–C bond formation, as native residues proved inert.¹² We

have failed to directly generate effective $\text{RC(O)}\cdot$ radicals for direct reaction on proteins (Figure 2b).

Here, during the course of our investigation into other difluoroalkyl radicals with proteins, we have discovered not only the sufficient reactivity of $\text{HCF}_2\cdot$ toward native residues when generated under mild, *reductive* conditions but also a reactivity that is coupled through their inherent chemoselectivity with a Reimer-Tiemann hydrolytic manifold in heteroaromatic residue ($X = \text{Trp} > \text{His}$, Figure 2c) adducts.^{17,18} This now allows effective use of difluoromethyl radicals as masked “ $\text{HC(O)}\cdot$ equivalents” for protein C-formylation (Figure 2c). The resulting, nonproteinogenic residue C-formyl-Trp (Cfw) thus bears a minimally sized reactive group at C2 that we show here proves useful for selective, electrophilic carbonyl-mediated modification in proteins as well acting as an enhanced fluorophore. Given the emerging modulatory role of natively, post-translationally C-modified Trp (e.g., C2-mannosylation),^{19,20} seemingly at key protein–protein interfaces, and the high propensity of Trp at protein–membrane interfaces,²¹ these “minimal size” electrophiles and spectroscopic probes could also prove to be precise and subtle probes of both protein–protein and protein–membrane interactions.

RESULTS

Design and Generation of a Potential Protein-Reactive $\text{RC(O)}\cdot$ or $\text{HC(O)}\cdot$ Equivalent. The carbonyl group has often been used (due to its accessibility) in small-molecule organic chemistry as a precursor for *gem*-difluorination,²² establishing a useful synthetic strategic equivalence (synthon “ CF_2 ” = “ C(O) ”, Figure 2c). We reasoned therefore that in protein chemistry, this essentially near-absent functional group (“ C(O) ”) might be accessed by a strategically reversed process (from “ CF_2 ”: $\text{RCF}_2 \rightarrow \text{RC(O)}$ or $\text{HCF}_2 \rightarrow \text{HC(O)}$) by appropriate positioning of *gem*-difluoromethylene under aqueous conditions and in an electronic environment (e.g., benzylic) that would favor hydrolysis. While the CF_2 moiety is typically stable at saturated sites, haloalkyl moieties found instead at benzylic or equivalent sites in (hetero)aromatics may be more readily hydrolyzed through Reimer-Tiemann-type manifolds.^{23,24}

We started our investigation by testing suitable difluoroalkyl $\text{RCF}_2\cdot$ radical sources and explored both *oxidative* and *reductive*

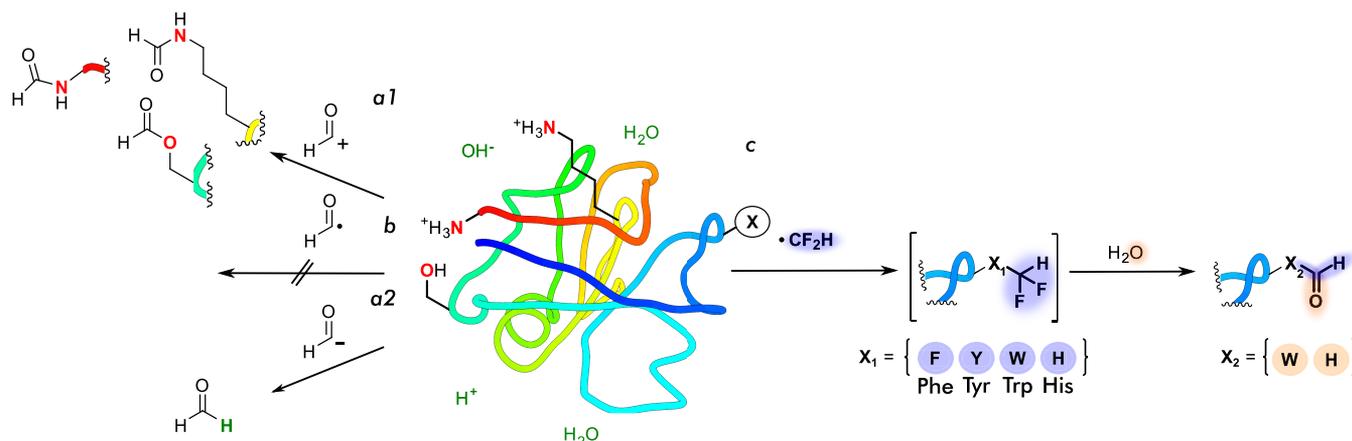


Figure 2. Analysis of potential chemical strategies toward protein C-carbonylation. (a) Heterolytic $2e^-$ approaches using HC(O)^+ or HC(O)^- equivalents (a1 and a2, respectively) are not compatible with the need for an aqueous environment and/or prevalent side reactivity with heteroatom-containing functional groups (e.g., Lys, Ser). (b) Direct radical approaches currently fail. (c) A masked $\text{HC(O)}\cdot$ equivalent strategy, where a reactive $\text{HCF}_2\cdot$ (targeting (hetero)aromatic residues) could serve as a hydrolytically, revealed formyl group for a suitable acceptor residue X_2 (Trp or His).

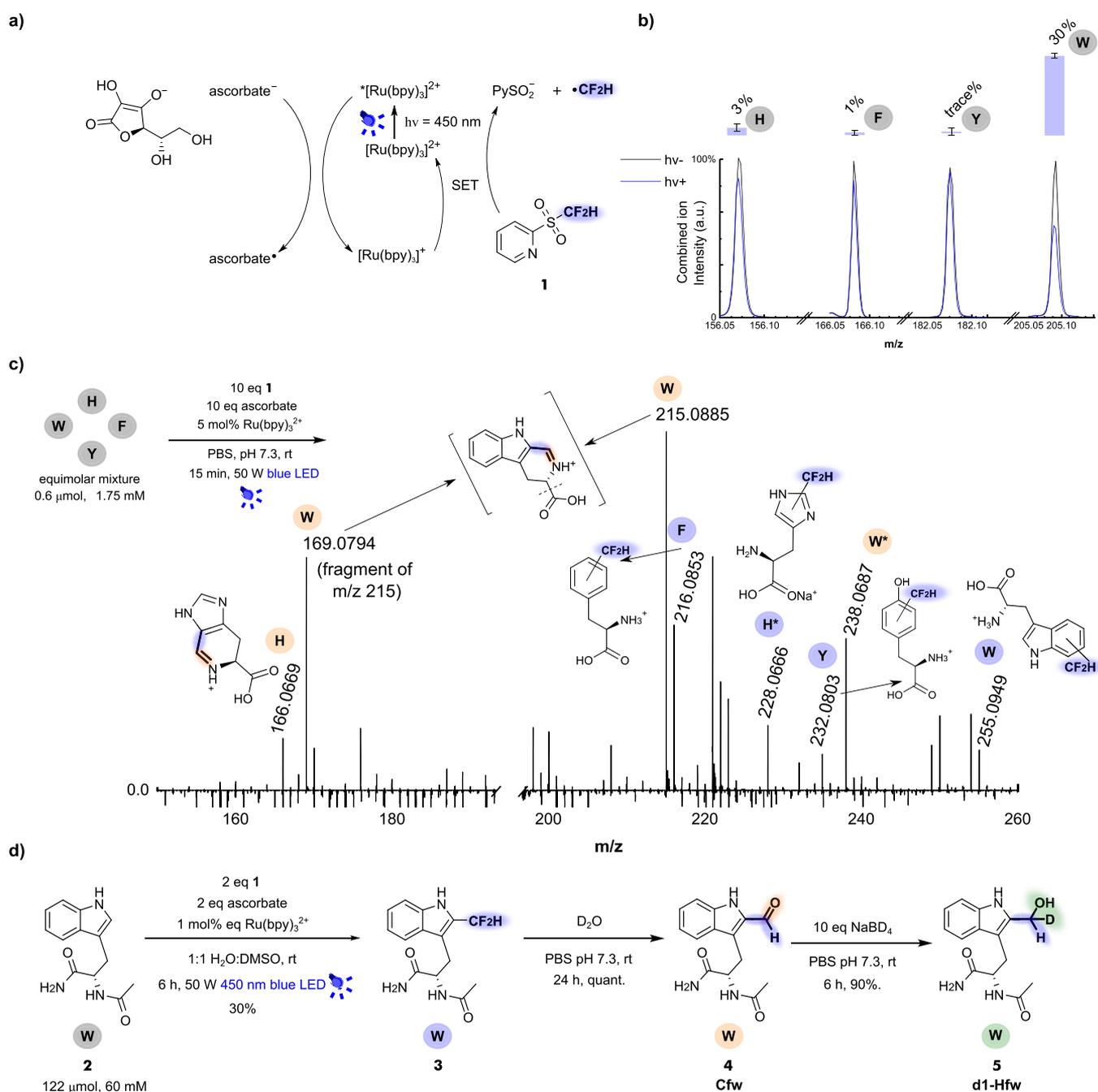


Figure 3. Scoping of reductively-initiated difluoromethylation. a) Proposed mechanism for difluoromethyl radical formation under photoredox catalysis with ascorbate as reductant. b) Reactivity of tested amino acids in a competitive setting as quantified via conversion of the starting material in an LC-MS assay for shown single parent ions (normalized using ^{15}N labeled standards, $n = 2$). c) Identification of main products of the difluoromethylation reaction by looking at differential TIC spectra (see [Supplementary Methods](#)). Identity was assigned via HRMS masses (<20 ppm) of singly protonated or Na^+ -adducted ions (marked *) and subsequent MS/MS fragmentation. d) Scaled reaction with a model tryptophan substrate under semiaqueous conditions revealed C2-difluoromethylated species as the main intermediate product, which then upon hydrolysis in aqueous buffer yielded reducible aldehyde, consistent with design and observations under fully hydrolytic conditions.

initiation manifolds. Difluoroalkylsulfonates have been used by us previously for *oxidatively* initiated protein fluorination through trifluoromethylation¹⁶ (via $\text{CF}_3\cdot$). In our hands,^{16,17} however, such Minisci-type oxidative initiation turned out to be unsuitable due to its low efficiency in $\text{RCF}_2\cdot$ radical generation and, more pertinently, the instability of the sulfonates under the aqueous conditions required ([Supplementary Figure S1](#)). These generated instead acyl fluorides due to their ready *preinitiation* hydrolysis under oxidative conditions ([Supplementary Figure](#)

[S1](#)), leading to confounding unwanted mixed amidation and acylation that we examined in detail by proteolytic-MS/MS analyses ([Supplementary Figure S2](#)).^{25,18}

It has been known for more than 50 years²⁶ that the formal precursors to sulfonates, (hetero)aryl sulfones, can also act as a source of alkyl radicals, albeit under different initiation manifolds. Their utility in *reductive* initiation has proven successful,^{27,28} although largely untested in aqueous media until their recent use in modification of biomolecules bearing

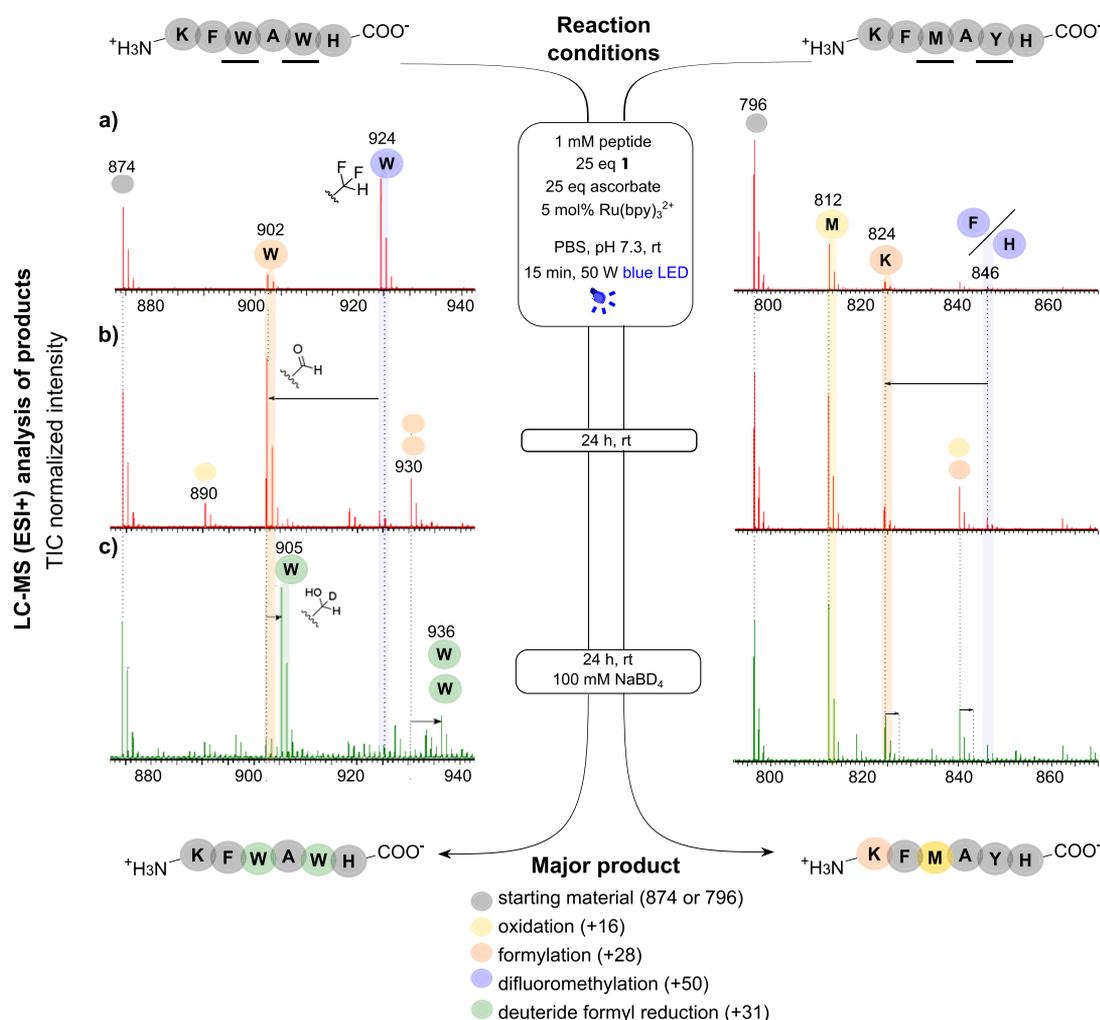


Figure 4. Verification of the difluoroalkylation/hydrolysis sequence from Trp to Cfw in peptides. Comparative analysis of combined TIC-MS spectra of products. a) After each step, reaction mixtures were analyzed by LC-MS/MS, which confirmed difluoromethylation (+50 Da) for Trp-containing peptides, while Trp-free mutants showed low levels of Met oxidation (+16 Da) and Lys formylation (+28 Da). b) Over 24 h hydrolysis occurred. c) Borodeuteride-mediated detection was performed for both samples which indicated almost complete conversion for Trp peptide and essentially no reactivity for Trp-free mutant. For full HRMS spectra and MS/MS spectra, see [Supplementary Figures S9 and S10](#).

unnatural amino acid residues.²⁹ Critically, of the sulfones tested we found that the 2-pyridylsulfone scaffolds RCF₂SO₂py **1** (Figure 3a) displayed useful solubility and good stability in aqueous reaction buffers typically used for protein manipulation and reaction.

Radical initiation was driven by photoredox-mediated reduction of sulfones using the mild and water-soluble photocatalyst Ru(bpy)₃²⁺.^{28–30} First, we tested putative (hetero)aromatic amino acids (Trp, His, Phe, Tyr) for radical acceptor reactivity (Figure 3b,c and [Supplementary Figure S3](#)). From a range of possible electron donors, sodium ascorbate^{28,30} in aqueous buffer proved most effective for these reactions. We also tested other possible biologically compatible reductants including thiols ([Supplementary Figures S6 and S33](#)), but these only led to reduced efficiencies ([Supplementary Figure S33](#)). Interestingly, however, in cellular application, use of non-ascorbate conditions did lead to partial success, suggesting that certain cell-present, endogenous bioreductants could also potentially function instead of ascorbate (see below). Under conditions mimicking typical protein modification protocols (low scale/concentrations, an amino acid residue as the limiting substrate with an excess of other reactants), monitoring via

quantitative LC-MS using ¹⁵N-labeled standards revealed higher reactivity of the Trp-substrate (30% conversion under limiting conditions), whereas for all others >94% of the unreacted starting material (<6% reacted) could be detected ([Supplementary Figure S3](#)), thus suggesting at least a >5-fold preference for Trp. This was further confirmed when tested in direct competition using an equimolar mixture of Trp, His, Phe, and Tyr under the same reagent excess (Figure 3b) where clear Trp chemoselectivity was maintained (~30% conversion) with only trace amounts of products from other residues. The results of these assays suggesting high Trp reactivity were also verified via quantitative LC-UV detection, yielding essentially the same result at an even higher excess of reagents and suggesting photon flux as the likely limiting “reagent” under these conditions ([Supplementary Figure S4](#)).

Gratifyingly, clear evidence was seen not only for radical addition (Figure 3c) but also for concomitant Reimer-Tiemann-type hydrolysis, as aldehyde formation was detected for Trp and His—both were observed as their internal cyclic aldimines (HRMS and fragmentation analysis, [Supplementary Figure S5](#)), in agreement with previous observations.³¹

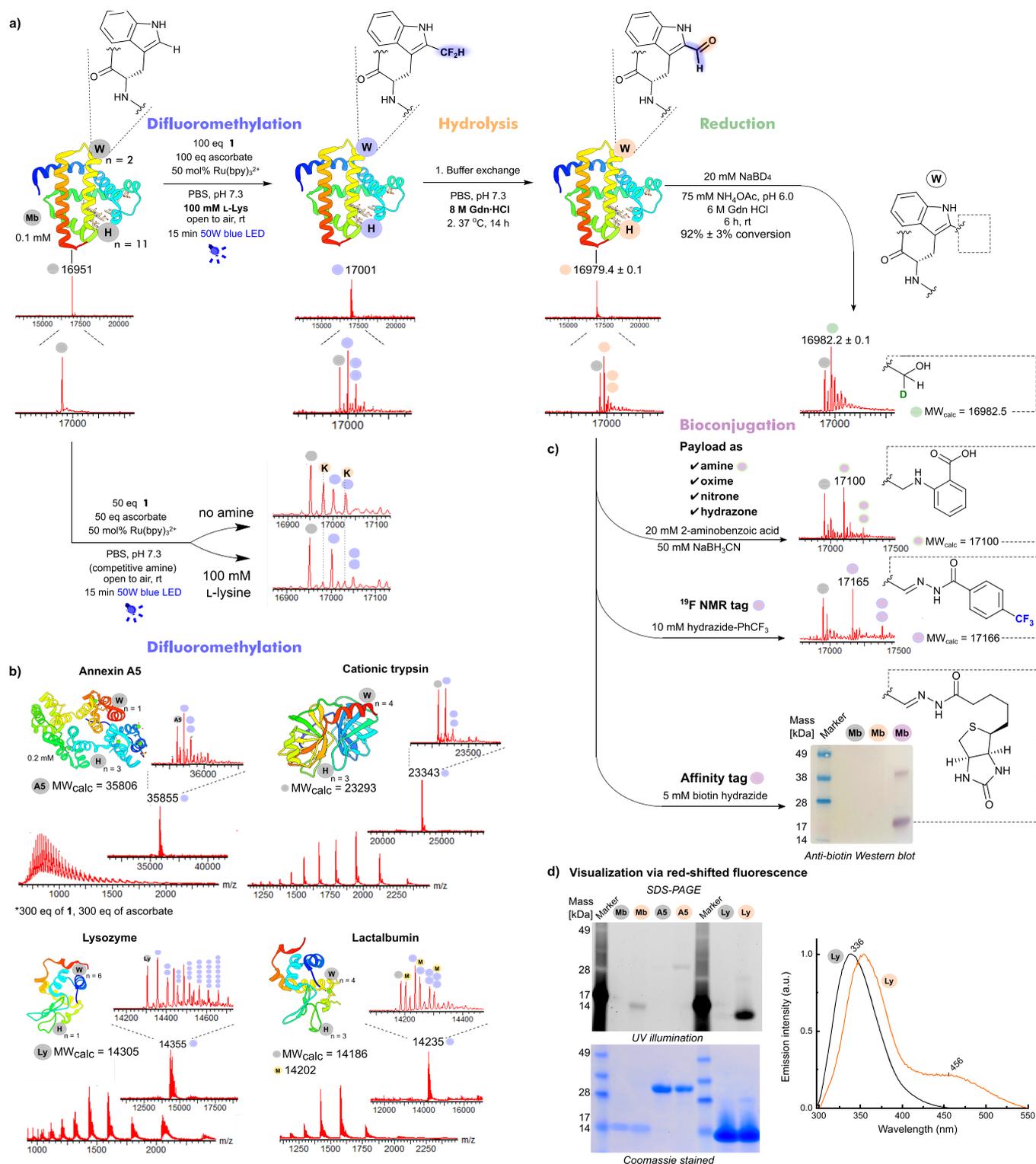


Figure 5. Sequential difluoroalkylation-hydrolysis allows access to Cfw in proteins. (a) Verification of the difluoromethylation/hydrolysis/reduction sequence confirming results observed for peptide. Identity of protein was verified by LC-MS at each stage. (b) Scope of proteins that underwent difluoromethylation under standard conditions used in (a). Color key as for Figure 4. (c) Bioconjugation: example reaction applications of Trp-aldehyde Cfw-tagged proteins in bioconjugations and labeling via carbonyl chemistry. See also Supplementary Figures S22–S27 and the main text for further payload examples and details. (d) Proteins containing Cfw display clearly red-shifted fluorescence emission (right) that allow their ready selective visualization over native, unmodified proteins.

Next, to fully confirm the product identity of this putative, desired difluoromethylation-hydrolysis sequence, we used the Trp-containing model AcNH-Trp-NH₂ (2) under scaled and semiaqueous conditions (Figure 3d). As intended, these

semiaqueous conditions slowed hydrolysis and so not only allowed isolation and full characterization of the difluoromethylated radical-adduct intermediate 3 but also critically allowed subsequent use of controlled hydrolysis to the observed formyl-

Trp 4 without cyclization. These experiments therefore importantly established not only the sequencing of reaction steps but also at the same time regioselectivity: 3 is C2-difluoromethylated and was isolated as the major product (along with a minor <20% complex mixture of multiply difluoromethylated, oxidized species and unreacted Trp substrate). When incubated in an aqueous buffer at room temperature, 3 readily gave the corresponding aldehyde, C2-formyl-Trp derivative Ac-Cfw-NH₂ (4). Furthermore, the reactivity and putative role of Cfw as an electrophilic trap and as a site for labeling was confirmed by subsequent reduction with borodeuteride to give site-selectively *deutero*-hydroxymethyl-Trp d1-Hmw (5) in 90% yield. As well as allowing labeling, the resulting observed +3.022 Da mass shift between Cfw and Hmw provided a useful diagnostic method for Cfw identification by mass spectrometry (MS) not only here but also subsequently in more complex substrates (see below).

Testing of a HC(O)· Equivalent in Peptides. Having identified a potentially productive and selective manifold for the creation of Cfw from Trp in model residues, we probed this potential reactivity in peptidic contexts. We constructed two short model peptide substrates: one containing representative proteinogenic aromatic and/or nucleophilic residues as potential, putative competing reactive sites⁹ (Phe, His, and Lys as well as Trp) in KFWAWH as well as a corresponding Trp-free (Trp → Met, Tyr) “mutant” KFMAYH found to have the most similar solubility needed for reactive comparison. These were chosen both for their utility as collision-induced-dissociation (CID)-fragmentable probe substrates for MS/MS as well as the presence of the WxW motif in inferred transmembrane sequences (*Desulfovibrio sp.* and *Pseudomonas sp.* cation-efflux proteins [ORFs SAMN05660337_0579, BA-863_11320, SAMN06295933_2365, AO242_06185]). To analyze the reaction mixtures, we employed LC-MS and LC-MS/MS monitoring supported by data-dependent acquisition (DDA)^{32,33} to ensure unbiased analyses via fragmentation of evolving intermediates and products.

The need to scavenge nonspecific oxidation was quickly realized. The use of ascorbate that had proven successful in small-molecule systems proved important here also. In the reductive photocatalytic manifold where recycling oxidative “backswing” is also required, a careful balance must be struck; ascorbate proved superior among a range of putative electron transfer mediators and redox scavengers (Supplementary Figure S6). In its absence, oxidative damage as indicated by LC-MS was observed via sequential oxidation ladders, whereas its presence allowed clean conversion (Supplementary Figure S7). LC-MS also revealed that the difluoromethylated peptide underwent a facile, gradual loss of the difluoromethyl group ($t_{1/2} \approx 5.5$ h) with appearance of the hydrolysis product when simply incubated in reaction mixtures. Variation of pH (4–8) did not significantly affect hydrolysis rates (Supplementary Figure S8).

Verification of the utility of this difluoroalkylation/hydrolysis sequence from Trp to Cfw (Figure 4) exploited the MS-detectable chemical derivatization method devised on single residues of Cfw (see above). Thus, borodeuteride-mediated detection on model peptides revealed an expected mass shift in >95% of the peaks corresponding to both singly and doubly formylated species (Figure 4c). This was complemented by LC-MS/MS fragmentation analyses that revealed Trp residues in KFWAWH as the main site of difluoromethyl radical addition, along with a minor population localized on His or Phe (Supplementary Figure S9). Notably, no detectable mass shift

was observed for the Trp-free “mutant” peptide KFMAYH further confirming Trp-dominant selectivity (under prolonged conditions this peptide, in the absence of other reactive sites, showed small amounts of nonspecific oxidation and N-formylation, Supplementary Figure S10). We also obtained essentially identical results with another peptide pair (Trp-containing and Trp-free, respectively) YPWF and YPFF, further highlighting Trp as the primary reactive residue (Supplementary Figure S11).

Finally, no conversion was observed in any system in the absence of light or photocatalyst verifying the key role of photoredox. Similarly, the radical nature of the reaction was confirmed by reacting the peptide in the presence of persistent, water-soluble nitroxide radical 4-hydroxy-TEMPO, which inhibited any product formation instead generating the 4-hydroxy-TEMPO-CF₂H adduct (Supplementary Figure S12).

C–C(O) Bond-Forming Protein C-Formylation via Difluoromethylation/Hydrolysis. Having demonstrated desired reactivity, site-selectivity, and hydrolytic Reimer-Tiemann “maturation” in peptide systems capable of allowing selective conversion of Trp to Cfw, we moved next to protein substrates (Figure 5). Horse-heart myoglobin (equine Myoglobin, eqMyo) contains two Trp residues (Trp7 and Trp14) among 11 His, 7 Phe, and 2 Tyr residues and so served as a good test for assessing chemoselectivity of the C-formylation sequence (difluoromethylation-hydrolysis, Figure 5a, left-middle). Myoglobins are proposed to play a role in intracellular (sarcoplasmic for eqMyo) oxygen transport, in part, through mitochondrial membrane interactions.^{34,35}

Intact protein LC-MS revealed successful formation of products with masses corresponding to difluoromethylation and, in part, seemingly hydrolyzed species (Figure 5a, left). To assess the selectivity of the reaction, we performed tryptic digestion and analyzed resulting peptides via LC-MS/MS. This revealed not only desired difluoromethylated (hetero)aromatic products but also some nonspecific N-formylation (corresponding to mass of apparent hydrolysis and as had been seen in peptide models (see above)) (Supplementary Figure S13). A survey of several sacrificial amines as buffer components (Supplementary Figure S14), including several soluble alkylamines and aromatic anilines that exhibit higher nucleophilicity,³⁶ allowed essentially complete suppression of this nonspecific formylation (Figure 5a, left) as seen by intact protein MS with only difluoromethylation. However, anilines also lowered the efficiency of difluoromethylation, presumably via competitive trapping aromatic addition of ·CF₂H, and so lysine was chosen as a preferred component. Pleasingly, the resulting standard conditions (Figure 5a, 100 equiv of the radical precursor and ascorbate with 50 mol % of the photocatalyst at ~100 μM protein and 100 mM lysine) allowed >70% conversion to difluoromethylated eqMyo (Figure 5a, left). Notably, such was the efficiency that photocatalyst Ru(bpy)₃²⁺ could be used in substoichiometric levels (50 mol % down to as little as 5 mol %, Supplementary Figure S15). Such catalytic systems in protein chemistry remain rare.³⁷

Next we tested the protein substrate scope. Importantly, these standard conditions could be universally applied to other protein types with varying Trp content (1–6 residues) and fold state (predominantly α structured protein myoglobin; two mixed α/β domain proteins lysozyme³⁸ and lactalbumin³⁹ as well as predominantly β structured protein trypsin⁴⁰) leading to an essentially-similar, good level of difluoromethylation (>60% conversion for all, Figure 5b). In cases of lower conversions, it

was possible to increase reagent concentrations to drive greater conversion. For example in AnnexinA5, which has a single Trp residue implicated in binding to phospholipid membranes⁴¹ upon Ca²⁺ sequestration, the use of 300 equiv of radical precursor then also allowed conversions of >60%.

Finally, as for peptides, difluoromethylation did not proceed in the absence of either photocatalyst or light. We also found, interestingly, that representative reactions proceeded essentially identically with or without a denaturant (Supplementary Figure S16), suggesting reactivity irrespective of the folding state and consistent with the minimal size of ·CF₂H as a reactant. This valuably suggested good accessibility even into interfacial sites. The minimal steric hindrance to Trp residues was further confirmed by LC-MS/MS of a high Trp content protein lysozyme, which revealed a Trp-selective, statistical distribution (Supplementary Figure S17).

Next, we studied hydrolysis of these difluoromethylated Trp residues to Cfw (Figure 5a, middle). Notably, unlike radical addition, the hydrolysis rate was dependent on the protein fold state and was slower under native conditions (Supplementary Figure S18). Thus, consistent with an expected requirement for solvent (water) accessibility, solvent inaccessible difluoromethylated residues did not react—for example, Trp7 and Trp14 in eqMyo showed only low accessibility (Supplementary Figure S19). However, under mildly denaturing conditions (preferentially with guanidine hydrochloride rather than urea, Supplementary Figure S20) hydrolysis proceeded to completion and could be accelerated by heating to 37 °C (Figure 5a, middle).

Borodeuteride-mediated trapping to d1-Hmw gave a mass shift corresponding to 92 ± 3% conversion (Figure 5a, right and Supplementary Figure S21), further confirming the presence of Cfw.

Exploitation of Cfw in Proteins—Chemical Reactivity.

The installation of Cfw as a novel residue into potentially reactive sites raised the possibility of multiple electrophilic “trapping” modes via its aldehydic formyl group that could both complement and expand other modes of reactivity.¹ First, Cfw’s apparent and clear in-protein reactivity had already been demonstrated through deuteride addition (using NaBD₄) with efficiencies of >90% (see Figure 5a, right and Supplementary Figure S21 and discussion above); notably this allowed a ready, additional method for deuteride installation in proteins through C–D bond formation.

Second, alpha-nucleophiles that are widely and effectively used were tested. Varied hydrazide (*p*Tol-sulfonyl-, *p*CF₃-benz-formyl-, (+)-biotinyl-) and hydroxylamine (*O*-pentafluoroBn-, *O*-3-chloroallyl-, *N*-Bn-, *N*-tBu) partners gave their respective hydrazone, oxime, and nitron products in up to 95% conversions (Figure 5c and Supplementary Figures S22–S26). Moreover, this could be usefully telescoped with hydrolytic processing to allow direct one-pot conversion from the difluoromethylated protein (in denaturative buffers and incubation overnight at 37 °C, Supplementary Figure S22). Cfw displayed apparently strong inherent reactivity, and only in a few cases were ligation catalysts (e.g., anthranilic acid⁴²) required to ensure a more complete reaction (Supplementary Figure S23). In this way, proteins bearing Cfw could be readily labeled with ¹⁹F markers suitable for ¹⁹F NMR (Figure 5c, upper application and Supplementary Figure S24), appended with a putative payload and equipped with affinity tags such as D-biotin as confirmed by a biotin-affinity-mediated Western blot (Figure 5c, third application) and ¹⁹F NMR, respectively (Supplementary Figure S24). The scope of these reactions allowed

application to representative proteins: difluoromethylation-hydrolysis-biotinylation sequences were applied to a range of additional proteins—enzymes (thrombin, chymotrypsin, and horseradish peroxidase) as well as binding proteins (globulin lactalbumin, annexin A5, and SRP19 (protein subunit of the nucleoprotein signal recognition particle complex)⁴³) (Supplementary Figure S26).

Finally, we tested a third mode of aldehyde reactivity in Cfw. eqMyo bearing Cfw underwent ready reductive amination with a range of both anilines and alkylamines in unoptimized 85% conversions from Cfw using NaBH₃CN as part of difluoromethylation-hydrolysis-labeling sequences (Supplementary Figure S27) that, as for the reaction of alpha nucleophiles, could be conducted in a ready two-pot, telescoped process.

Exploitation of Cfw in Proteins—Spectroscopic. Inherent tryptophan fluorescence has long been exploited in protein spectroscopy and photobiology;⁴⁴ C-2 modified 3*H*-indoles also form the core of several powerful fluorophores (e.g., so-called “Cy” indocyanines and indolenine dyes), and C-2 modified 1*H*-indoles are the cores of key fluorescent indicators such as “Indo-1”.⁴⁵ “Indo-1” shows red-shifted absorption and emission over comparable unsubstituted indoles⁴⁵ (to λ_{max}(em) ~ 480 nm), and recent photophysical studies suggest such red-shifting is also seen in formylindoles.⁴⁶ We therefore reasoned that Cfw might act as a readily generated fluorophore in proteins with valuably altered (“autofluorescent”) properties over Trp.

To test this red-shifted autofluorescence directly in Cfw-containing proteins, we used broad-range ultraviolet light transillumination (λ_{max} = 302 nm) with detection at wavelengths (>480 nm) above the emission from natural Trp fluorescence. Strikingly, only Cfw-containing samples showed light emission (Figure 5d). This observation in proteins for Cfw (2-formyl-Trp) is wholly consistent with observations in simple formylindoles; of all the formyl indoles, 2-formylindole displays the strongest absorption coefficient in the UV region (λ_{max}(abs) = 309 nm⁴⁶). Moreover, semiquantitative analysis of emission enhancement using fluorescence densitometry of Ly-Cfw, eqMyo-Cfw, and AnnexinV-Cfw samples after gel electrophoresis revealed that emission correlated with the number of Trp residues per protein (Supplementary Figure S28). To probe this unusual and novel fluorescent behavior further, we determined normalized emission spectra (λ_{ex} = 280 nm) for different Cfw-containing proteins in PBS pH 7.3 at concentrations of 10–20 μM. These all revealed the emergence of a peak at λ_{max}(em) ~ 456 ± 5 nm. As expected, this was observed most strongly for lysozyme with six potential Cfw residue sites (Figure 5d and Supplementary Figure S28) as compared to those with fewer sites.

Applicability to More Complex Biological Contexts.

The primary focus of our study was to explore the underlying factors allowing more efficient overall Trp C-formylation. As expected, the solvolytic dependency of this novel Reimer-Tiemann manifold led to enhanced efficiencies under protein conditions that allowed greater solvent access to protein cores (and hence to buried Trp) (see above). Nonetheless, we also sought to explore the applicability of this C-formylation transformation in increasingly more structured protein contexts. First, we probed the potential for application to proteins without the need for any form of increased solvent access caused by, e.g., a partial denaturant. Pleasingly, this revealed clear progression to Cfw (albeit with reduced rate) in a manner that allowed not only overall C-formylation but also allowed subsequent bioconjugation (see Supplementary Figure S34).

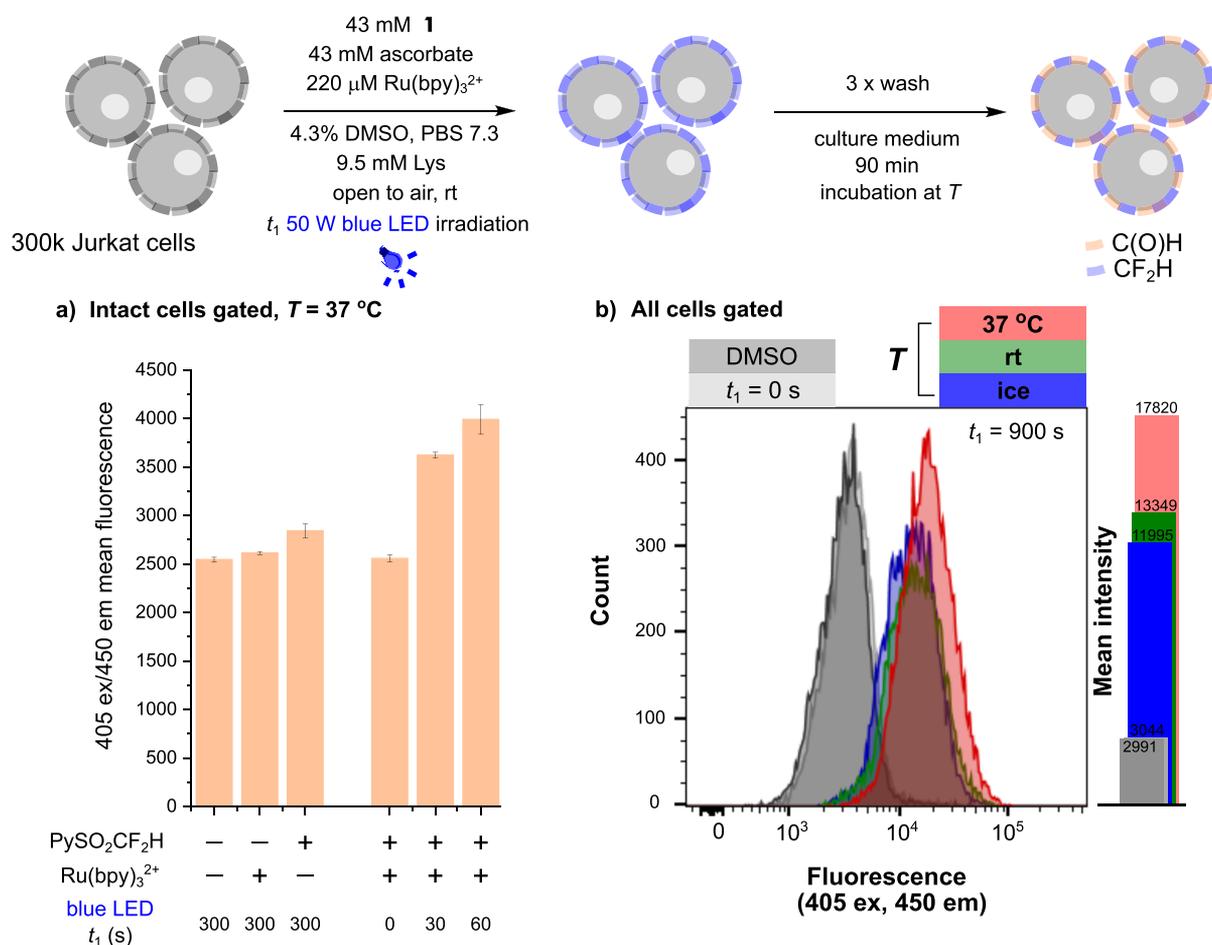


Figure 6. Difluoromethylation/hydrolysis of leukemic human T-lymphocyte (Jurkat) cells and fluorescence-activated sorting. (a) Geometrical mean of Cfw-associated 450 nm fluorescence emission intensity for cells subjected to the difluoromethylation (variable irradiation time) and hydrolysis sequence as well as controls. Controls were irradiated for 300 s, while all samples were incubated at 37 °C postirradiation to allow hydrolytic processing prior to analysis. For these analyses, only intact cells (non PI stained) were considered, see also the [Supporting Information](#). (b) Overlay of a single parameter (450 nm fluorescence) histogram for all cells (excluding debris) following hydrolytic processing at different temperatures. Associated increased levels of fluorescence following hydrolytic processing at higher temperature parallel those observed *in vitro*, consistent with Cfw formation.

Having established compatibility for functional bioconjugation under native conditions, we then tested application to a cellular context ([Figure 6](#)). We employed fluorescence associated cell sorting (FACS) analysis to identify those cells bearing enhanced fluorescence at emission frequency identified with formation of Cfw ($\sim 450\text{ nm}$, see above). Difluoromethylation/hydrolysis reactions applied to human T-lymphocyte (Jurkat) cells allowed identification of conditions under difluoromethylation reaction timecourses ($<90\text{ s}$) that allowed significant fluorescence induction in live cells upon subsequent hydrolytic processing (see [Figure 6a](#) and [Supplementary Figure S35](#)). While extended difluoromethylation reaction timecourses led to increased cell-associated fluorescence, it also led to reduced cell viability ([Supplementary Figures S35 and S36](#)). Also, consistent with Cfw formation, associated increased levels of fluorescence were observed following hydrolytic processing of reacted “formylated” cells at higher temperatures (see [Figure 6b](#)) similar to observations made on proteins *in vitro* (see above).

DISCUSSION

In summary, reductive photocatalytic protein modification in the presence of ascorbate allows effective $\text{HCF}_2\cdot$ radical generation and hence difluoromethylation with residue

selectivity for Trp; the resulting C-2-difluoromethylated Trp residues undergo Reimer-Tiemann-type dehalogenation providing hydrolytic collapse in proteins to 2-formyl-Trp (Cfw). This radical addition-hydrolysis protocol therefore now allows straightforward, robust, and scaleable (up to the $> \text{mg}$ scale) chemical access to aldehydes in proteins, allowing it to be quickly exploited in established aldehyde labeling and ligation chemistries (hydrazone-ligation, oxime-ligation, reductive amination, reductive deuteration) yet with novel site (Trp) selectivity.

While our use of a mild, photocatalyst with a low E_{ox} “backswing” coupled with an ascorbate reductant minimized associated side oxidation, nonetheless some undesired oxidation of highly sensitive residues in certain proteins (e.g., Met in lactalbumin, [Figure 5b](#)) was occasionally observed. In addition, while to our knowledge this is the first example of C–C bond-forming formylation/carbonylation in proteins, it should be noted that this two-step process shows lower overall conversions than some alternative methods for generating aldehydes in proteins (see above and also the [Supporting Information](#) for an extended comparison of methods for introducing the formyl group into proteins). This reaction with a native residue is also slower than reactions under different conditions optimized for unnatural radical acceptor residues.²⁹

Our effectively neutral conditions (pH 7–8 or near) allow application to a wide variety of proteins. Given our strong, observed Trp selectivity, we were intrigued by the recent disclosure of Minisci-type methods applied to peptides at low pH (2–3) that appeared to favor His modification.¹⁸ While such acidic conditions have not been applied to proteins, we explored pH dependence down to 5. Consistent with increased reactivity for protonated His imidazoliums (and so reaction at His), difluoromethylation started to occur in higher copy numbers (Supplementary Figure S30), to levels exceeding the total number of Trp residues. Moreover, when such overly modified difluoromethylated protein samples were subjected to hydrolysis (Supplementary Figure S31), LC-MS monitoring revealed that hydrolysis did not proceed further than in optimally treated samples. Together this suggests not only that any additional difluoromethylation under acidic conditions is at non-Trp sites (e.g., His) but also that these may be less/nonsusceptible to Reimer-Tiemann-type hydrolytic collapse to formyl, at least under these conditions.

In addition to chemical reactivity, we observed intriguing spectroscopic tuning (“red-shifting”) of Trp’s fluorescent properties through C2-formylation that are consistent with recent observations in 2-formylindole.⁴⁶ Interestingly, chloroform can be used as a nonselective reagent suggested to react with photoexcited Trp that then generates mixed products thought to also include C4- and C6-formyl-Trp^{31,47,48}—this has formed the basis of so-called “stain-free” in-gel protein imaging methods.⁴⁹

Finally, while we did not explore the full range of (physico)chemical effects of difluoromethylation-hydrolysis, numerous other possible applications may be envisaged from the protein chemistry described here. For example, it is possible that, at a sufficient copy number, high Trp-content proteins may display alterations in their phase–phase interactions upon difluoromethylation. Indeed, through reverse-phase chromatographic analysis we observed that following lysozyme (containing 5 Trp) difluoromethylation the copy number could, in fact, be readily correlated with retention time (Supplementary Figure S32), despite variation by only such a small CHF₂ group. This apparent correlation of the CHF₂ copy number with phase–phase interactions suggests a seemingly strong effect.

Additional roles of the C2-formyl in Cfw could also be envisaged using other powerful aldehyde chemistries to either bioconjugate⁵⁰ or participate in “catch-release” chemical-affinity purification. Furthermore, given the preponderance of Trp residues at protein–membrane interfaces²¹ and the small structural change in Trp → Cfw, one could perhaps envisage the aldehyde-mediated cross-linking and trapping of a range of endogenous nucleophilic partners in biological contexts as a form of protein-directed trapping or even “fixation”.⁵¹ Work to explore the latter methods in yet more complex biological contexts, as shown here by application also to cells, is highly promising (allowing already here, e.g., fluorescence sorting due to the presence of Cfw on/in cells), and exploration extended in this direction is currently underway in our laboratory.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscentsci.0c01193>.

Supplementary experimental data and figures giving details of LC-MS, MS/MS, NMR analyses and character-

ization, solvent accessibility calculations, dot-blot results, fluorescence measurements, and description of FACS experiments as well as full experimental procedures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Benjamin G. Davis – Department of Chemistry, University of Oxford, Oxford OX1 3TA, United Kingdom of Great Britain and Northern Ireland; The Rosalind Franklin Institute, Oxfordshire OX11 0FA, United Kingdom of Great Britain and Northern Ireland; orcid.org/0000-0002-5056-407X; Email: ben.davis@chem.ox.ac.uk, ben.davis@rfi.ac.uk

Authors

Mateusz Imiolek – Department of Chemistry, University of Oxford, Oxford OX1 3TA, United Kingdom of Great Britain and Northern Ireland; The Rosalind Franklin Institute, Oxfordshire OX11 0FA, United Kingdom of Great Britain and Northern Ireland

Patrick G. Isenegger – Department of Chemistry, University of Oxford, Oxford OX1 3TA, United Kingdom of Great Britain and Northern Ireland

Wai-Lung Ng – Department of Chemistry, University of Oxford, Oxford OX1 3TA, United Kingdom of Great Britain and Northern Ireland; orcid.org/0000-0003-2892-6318

Aziz Khan – Department of Chemistry, University of Oxford, Oxford OX1 3TA, United Kingdom of Great Britain and Northern Ireland

Véronique Gouverneur – Department of Chemistry, University of Oxford, Oxford OX1 3TA, United Kingdom of Great Britain and Northern Ireland; orcid.org/0000-0001-8638-5308

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acscentsci.0c01193>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the EU Horizon 2020 programme (Marie Skłodowska-Curie grant agreement No. 675071) as well as the Swiss National Foundation (Fellowship to P.G.I.), the Croucher Foundation (Fellowship to W.L.N.), and GlaxoSmithKline/BBSRC (CASE studentship to A.K.).

■ REFERENCES

- (1) Spears, R. J.; Fascione, M. A. Site-selective incorporation and ligation of protein aldehydes. *Org. Biomol. Chem.* **2016**, *14* (32), 7622–7638.
- (2) Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (1), 56–61.
- (3) Fleissner, M. R.; Brustad, E. M.; Kalai, T.; Altenbach, C.; Cascio, D.; Peters, F. B.; Hideg, K.; Peuker, S.; Schultz, P. G.; Hubbell, W. L. Site-directed spin labeling of a genetically encoded unnatural amino acid (vol 106, 21637, 2009). *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21637.
- (4) Tuley, A.; Lee, Y. J.; Wu, B.; Wang, Z. U.; Liu, W. R. A genetically encoded aldehyde for rapid protein labelling. *Chem. Commun.* **2014**, *50* (56), 7424–7426.
- (5) Dixon, H. B. F. N-Terminal Modification of Proteins - a Review. *J. Protein Chem.* **1984**, *3* (1), 99–108.

- (6) Schmidt, B.; Selmer, T.; Ingendoh, A.; Vonfigura, K. A Novel Amino-Acid Modification in Sulfatases That Is Defective in Multiple Sulfatase Deficiency. *Cell* **1995**, *82* (2), 271–278.
- (7) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **2007**, *3* (6), 321–322.
- (8) Rush, J. S.; Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **2008**, *130* (37), 12240–12241.
- (9) Kruger, T.; Dierks, T.; Sewald, N. Formylglycine-generating enzymes for site-specific bioconjugation. *Biol. Chem.* **2019**, *400* (3), 289–297.
- (10) Bloom, S.; Liu, C.; Kolmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; MacMillan, D. W. C. Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **2018**, *10* (2), 205–211.
- (11) Yu, Y. N.; Zhang, L. K.; Buevich, A. V.; Li, G. Q.; Tang, H. Q.; Vachal, P.; Colletti, S. L.; Shi, Z. C. Chemoselective Peptide Modification via Photocatalytic Tryptophan beta-Position Conjugation. *J. Am. Chem. Soc.* **2018**, *140* (22), 6797–6800.
- (12) Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W. L.; Raj, R.; Faulkner, S.; Vallee, M. R. J.; Phanumartwath, A.; et al. Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **2016**, *354* (6312), aag1465.
- (13) Yang, A.; Ha, S.; Ahn, J.; Kim, R.; Kim, S.; Lee, Y.; Kim, J.; Söll, D.; Lee, H.-Y.; Park, H.-S. A chemical biology route to site-specific authentic protein modifications. *Science* **2016**, *354* (6312), 623–626.
- (14) Zheng, S.; Doucette, A. A. Preventing N- and O-formylation of proteins when incubated in concentrated formic acid. *Proteomics* **2016**, *16* (7), 1059–1068.
- (15) Jensen, S. K.; Keiding, S. R.; Thogersen, J. The hunt for HCO(aq). *Phys. Chem. Chem. Phys.* **2010**, *12* (31), 8926–8933.
- (16) Imiolek, M.; Karunanithy, G.; Ng, W.-L.; Baldwin, A. J.; Gouverneur, V.; Davis, B. G. Selective Radical Trifluoromethylation of Native Residues in Proteins. *J. Am. Chem. Soc.* **2018**, *140* (5), 1568–1571.
- (17) During the course of this work, oxidative reactions of presumed difluoroalkyl radicals on peptides were disclosed. See the following citation.
- (18) Noisier, A. F. M.; Johansson, M. J.; Knerr, L.; Hayes, M. A.; Drury Iii, W. J.; Valeur, E.; Malins, L. R.; Gopalakrishnan, R. Late-Stage Functionalization of Histidine in Unprotected Peptides. *Angew. Chem., Int. Ed.* **2019**, *58* (52), 19096–19102.
- (19) Hofsteenge, J.; Muller, D. R.; de Beer, T.; Löffler, A.; Richter, W. J.; Vliegthart, J. F. New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase U. *Biochemistry* **1994**, *33* (46), 13524–13530.
- (20) Shcherbakova, A.; Preller, M.; Taft, M. H.; Pujols, J.; Ventura, S.; Tiemann, B.; Buettner, F. F. R.; Bakker, H. C-mannosylation supports folding and enhances stability of thrombospondin repeats. *eLife* **2019**, *8*, e52978.
- (21) Yau, W. M.; Wimley, W. C.; Gawrisch, K.; White, S. H. The preference of tryptophan for membrane interfaces. *Biochemistry* **1998**, *37* (42), 14713–14718.
- (22) Tozer, M. J.; Herpin, T. F. Methods for the synthesis of gem-difluoromethylene compounds. *Tetrahedron* **1996**, *52* (26), 8619–8683.
- (23) Woolridge, E. M.; Rokita, S. E. Synthesis and Reactivity of 6-(Fluoromethyl)Indole and 6-(Difluoromethyl)Indole. *Tetrahedron Lett.* **1989**, *30* (45), 6117–6120.
- (24) Melanson, J. A.; Figliola, C.; Smithen, D. A.; Kajetanowicz, A. K.; Thompson, A. Probing the hydrolytic reactivity of 2-difluoromethyl pyrroles. *Org. Biomol. Chem.* **2017**, *15* (1), 144–152.
- (25) Such isobaric product mixtures may prove confounding if determined by total peptide mass only.
- (26) Horner, L.; Neumann, H. Studien Zum Vorgang Der Wasserstoffübertragung 0.12. Hydrierende Spaltung Von Sulfonen Mit Tetramethylammonium Als Elektronenüberträger. *Chem. Ber.* **1965**, *98* (6), 1715–1721.
- (27) Merchant, R. R.; Edwards, J. T.; Qin, T.; Kruszyk, M. M.; Bi, C.; Che, G. D.; Bao, D. H.; Qiao, W. H.; Sun, L. J.; Collins, M. R.; et al. Modular radical cross-coupling with sulfones enables access to sp³-rich (fluoro)alkylated scaffolds. *Science* **2018**, *360* (6384), 75–80.
- (28) Rong, J.; Deng, L.; Tan, P.; Ni, C. F.; Gu, Y. C.; Hu, J. B. Radical Fluoroalkylation of Isocyanides with Fluorinated Sulfones by Visible-Light Photoredox Catalysis. *Angew. Chem., Int. Ed.* **2016**, *55* (8), 2743–2747.
- (29) Josephson, B.; Fehl, C.; Isenegger, P. G.; Nadal, S.; Wright, T. H.; Poh, A. W. J.; Bower, B. J.; Giltrap, A. M.; Chen, L.; Batchelor-McAuley, C.; et al. Light-Driven Posttranslational Installation of Reactive Protein Side Chains. *Nature* **2020**, *585* (7826), 530–537.
- (30) Yerien, D. E.; Cooke, M. V.; Vior, M. C. G.; Barata-Vallejo, S.; Postigo, A. Radical fluoroalkylation reactions of (hetero) arenes and sulfides under red light photocatalysis. *Org. Biomol. Chem.* **2019**, *17* (15), 3741–3746.
- (31) Ladner, C. L.; Tran, K.; Le, M.; Turner, R. J.; Edwards, R. A. Excited State Photoreaction between the Indole Side Chain of Tryptophan and Halocompounds Generates New Fluorophores and Unique Modifications. *Photochem. Photobiol.* **2014**, *90* (5), 1027–1033.
- (32) Stahl, D. C.; Swiderek, K. M.; Davis, M. T.; Lee, T. D. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. *J. Am. Soc. Mass Spectrom.* **1996**, *7* (6), 532–540.
- (33) Domon, B.; Aebersold, R. Review - Mass spectrometry and protein analysis. *Science* **2006**, *312* (5771), 212–217.
- (34) Postnikova, G. B.; Shekhovtsova, E. A. Fluorescence studies on the interaction of myoglobin with mitochondria. *Biochemistry (Moscow)* **2012**, *77* (3), 280–287.
- (35) Gros, G.; Wittenberg, B. A.; Jue, T. Myoglobin's old and new clothes: from molecular structure to function in living cells. *J. Exp. Biol.* **2010**, *213* (16), 2713–2725.
- (36) Brotzel, F.; Chu, Y. C.; Mayr, H. Nucleophilicities of primary and secondary amines in water. *J. Org. Chem.* **2007**, *72* (10), 3679–3688.
- (37) Isenegger, P. G.; Davis, B. G. Concepts of Catalysis in Site-Selective Protein Modifications. *J. Am. Chem. Soc.* **2019**, *141* (20), 8005–8013.
- (38) Blake, C. C.; Koenig, D. F.; Mair, G. A.; North, A. C.; Phillips, D. C.; Sarma, V. R. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **1965**, *206* (4986), 757–761.
- (39) Schlepckow, K.; Wirmer, J.; Bachmann, A.; Kieffhaber, T.; Schwalbe, H. Conserved folding pathways of alpha-lactalbumin and lysozyme revealed by kinetic CD, fluorescence, NMR, and interrupted refolding experiments. *J. Mol. Biol.* **2008**, *378* (3), 686–698.
- (40) Huber, R.; Bode, W. Structural Basis of the Activation, Action and Inhibition of Trypsin. *Hoppe-Seyler's Z. Physiol. Chem.* **1979**, *360* (4), 489–489.
- (41) Meers, P.; Mealy, T. Relationship between Annexin-V Tryptophan Exposure, Calcium, and Phospholipid-Binding. *Biochemistry* **1993**, *32* (20), 5411–5418.
- (42) Kolmel, D. K.; Kool, E. T. Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem. Rev.* **2017**, *117* (15), 10358–10376.
- (43) Hainzl, T.; Huang, S.; Sauer-Eriksson, A. E. Structure of the SRP19 RNA complex and implications for signal recognition particle assembly. *Nature* **2002**, *417* (6890), 767–771.
- (44) Royer, C. A. Probing Protein Folding and Conformational Transitions with Fluorescence. *Chem. Rev.* **2006**, *106* (5), 1769–1784.
- (45) Gryniewicz, G.; Poenie, M.; Tsien, R. Y. A new generation of Ca²⁺ indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **1985**, *260* (6), 3440–3450.
- (46) You, M.; Fan, H.; Wang, Y.; Zhang, W. Aldehyde-derivatized indoles as fluorescent probes for hydration environments. *Chem. Phys.* **2019**, *526*, 110438.
- (47) Casasfinet, J. R.; Wilson, S. H.; Karpel, R. L. Selective Photochemical Modification by Trichloroethanol of Tryptophan

Residues in Proteins with a High Tyrosine-to-Tryptophan Ratio. *Anal. Biochem.* **1992**, *205* (1), 27–35.

(48) Edwards, R. A.; Jickling, G.; Turner, R. J. The light-induced reactions of tryptophan with halocompounds. *Photochem. Photobiol.* **2002**, *75* (4), 362–368.

(49) Freeby, S.; Liu, N.; Mcdonald, K.; Paulus, A.; Posch, A. In *PCT WO2013163451A1*, 2013.

(50) Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (1), 46–51.

(51) Thavarajah, R.; Mudimbaimannar, V.; Elizabeth, J.; Rao, U.; Ranganathan, K. Chemical and physical basics of routine formaldehyde fixation. *Journal of Oral and Maxillofacial Pathology* **2012**, *16* (3), 400–405.